Full Proceedings of the 2nd International Conference on Information Management and Business (IMB2006) Sydney, Australia

www.aimb.org

13 - 16 Feb, 2006

Proceedings Highlights:
" Makes a significant and substantial contribution to the overall literature on Information Management & Business"

Proceedings Editors:
Bhuvan UNHELKAR & Yi-Chen LAN

School of Computing and Mathematics,
University of Western Sydney, Sydney, Australia
ISBN: 1 74108 122 X

Welcome to the

Proceedings of the Second International Conference on Information Management and Business (IMB2006)

Sydney, Australia, 13 – 16 Feb, 2006

Proceedings Editors: Bhuvan UNHELKAR and Yi-chen LAN
(School of Computing & Mathematics, University of Western Sydney)

Cover Design: Anant Dhume
Special Editorial Assistance: Dinesh Arunatileka & Abbass Ghanbary
Conference Website: www.aimb.org/imb2006/

Publisher & Printer: University of Western Sydney

ISBN: 1 74108 122 X

Welcome to this Second International Conference on Information Management and Business. This printed version of the proceedings contains abstracts of keynote addresses and abstracts of all papers accepted for presentation. Full length papers are available in electronic format as a separate associated publication.

Association of Information Management and Business (AIMB)

The Association of Information Management and Business (AIMB) is an international professional organisation committed to advancing information systems management in modern business environment and bridging the gap between innovation and practices.

The primary objective of AIMB is to bring the world's academics and practitioners to exchange concepts, new ideas, research results, experience and other related issues which could contribute to the rapidly changing information systems environment and benefit the corporate business community.

Information Management and Business “IMB” Conference

As such, AIMB has been organising an international conference “IMB” annually under the collaboration of University of Western Sydney, Australia, Shih Chien University, Taiwan, Fachhochschule Würzburg-Schweinfurt, Germany, South Carelia Polytechnic, Finland, and associated academic and industrial societies.

IMB provides a popular and stimulating forum that brings together academic researchers and industry professionals to share ideas, exchange knowledge, consider new advanced applications and discuss future directions on information management and business.

These are proceedings of the second annual IMB Conference, being held in Sydney, Australia, in February, 2006. This conference is providing an excellent opportunity for academics and practitioners to present their work, collaborate with each other, receive feedback and positive criticisms and thus define the future of information management and business applications. We, as editors of these proceedings, together with the organizing committee of IMB2006 are committed to making this event a success and envisage its significant influence on the business information systems and management community internationally.

We are confident you will find these proceedings an invaluable addition to your reading material and will undoubtedly provide the impetus to your research. We wish you a productive and enjoyable conference.

Bhuvan Unhelkar & Yi-chen Lan

February 2006, Sydney Australia

Organising Committee

Organised by
University of Western Sydney, Australia

In cooperation with
Shih Chien University, Taiwan
Fachhochschule Würzburg-Schweinfurt, Germany
South Carelia Polytechnic, Finland
Australian Computer Society

Conference Chair
Yi-chen Lan, University of Western Sydney, Australia

Advisory Committee
Anneli Pirttilä, South Carelia Polytechnic, Finland
Antti Lehmusvaara, South Carelia Polytechnic, Finland
Athula Ginige, University of Western Sydney, Australia
Bernd Breutmann, Fachhochschule Würzburg-Schweinfurt, Germany
Chung-Hsing Hsieh, Shih Chien University, Taiwan
Bor. Yi. Huang, Shih Chien University, Taiwan
Kari Smolander, South Carelia Polytechnic, Finland
Maria Ruey-Yuan Lee, Shih Chien University, Taiwan

Program Co-Chairs
Vijay Khandelwal, University of Western Sydney, Australia
Mahesh S. Raisinghani, Texas Woman's University, USA

Proceedings Editors
Bhuvan Unhelkar, University of Western Sydney, Australia
Yi-chen Lan, University of Western Sydney, Australia

Publicity Co-Chairs
Robyn Lawson, University of Western Sydney, Australia
Ana Hol, University of Western Sydney, Australia

Local Arrangements Chair
Joanne Curry, University of Western Sydney, Australia

Doctoral Consortium Chair
Carole Alcock, University of South Australia

Tutorials and Panels Co-Chairs
Yogesh Deshpande, University of Western Sydney, Australia
San Murugesan, Southern Cross University, Australia

Finance Chair
Ronald Schmid, University of Western Sydney, Australia

Program Committee

Program Co-Chairs

Vijay Khandelwal, University of Western Sydney, Australia
Mahesh S. Raisinghani, Texas Women's University, USA

Committee Members

Mohammad Abul Kalam Azad, International Islamic University, Bangladesh
Phillip Brook, University of Western Sydney, Australia
Bernd Breutmann, University of Applied Sciences Wuerzburg, Germany
Ian Caddy, University of Western Sydney, Australia
Meena Chavan, University of Western Sydney, Australia
Tsung Teng Chen, National Taipei University, Taiwan
Wang-Jiunn Cheng, Shih Chien University, Taiwan
Li-Pond Chin, Shih Chien University, Taiwan
Omer Eldai, Aalborg University, Denmark
Stan Glaser, Sydney Graduate School of Management, Australia
Petter Gottschalk, Norwegian School of Management BI, Norway
Ta-Wei Hung, Shih Chien University, Taiwan
Drew Hwang, California State Polytechnic University, USA
Zahir Irani, Brunel University, U.K.
Ernest Jordan, Macquarie University, Australia
Kalevi Kyläheiko, Lappeenranta University of Technology, Finland
Paul C K Kwok, The Open University of Hong Kong, China
Meng-Huang Lee, Shih Chien University, Taiwan
Antii Lehmusvaara, South Carelia Polytechnic, Finland
Chad Lin, Edith Cowan University, Australia
Heinrich C. Mayr, University of Klagenfurt, Austria
Nasrullah Memon, Aalborg University Esbjerg, Denmark
Sharon Moore, Sydney Graduate School of Management, Australia
Tomasz Müldner, Acadia University, Canada
Hossein Najafi, University of Wisconsin- River Falls, USA
John Page, Monash University, Australia
Shailendra Palvia, Long Island University, USA
Moh'd A. Radaideh, UAE University, UAE
Andreas Rinkel, Hochschule für Technik, Switzerland
Premaratne Samaranayake, University of Western Sydney, Australia
Elhadi Shakshuki, Jodrey School of Computing Science, Canada
Gilbert Silvius, University of Professional Education, The Netherlands
Kari Smolander, South Carelia Polytechnic, Finland
Zhaohao Sun, University of Wollongong, Australia
Philip Tsang, Fujian Radio and TV University, China
David Walters, Sydney Graduate School of Management, Australia
Yung-Hsin Wang, Tatung University, Taiwan

Reviewers

Reviewers have played a significant role in the success of this conference. Having volunteered their time and skills, reviewers are also required to present an unbiased and informed opinion on the suitability of the papers. The papers in this conference have reached a high quality as a result of the effort of the following individuals, whom we thank sincerely.

Ana Hol
Andreas Rinkel
Antii Lehmusvaara
Athula Ginige
Bernd Breutmann
Bhuvan Unhelkar
Carole Alcock
Chad Lin
D.M. Akbar Hussain
Drew Hwang
Elhadi Shakshuki
Ernest Jordan
G. Jimmy Chen
Hossein Najafi
Ian Caddy
Indrit Troshani
Joan Cooper
Joanne Curry
John Page
Julie Fisher
Kalevi Kyläheiko
Kari Smolander
Li-Pond Chin
Mahesh Raisinghani
Maria R. Lee
Meena Chavan
Meng-Huang Lee
Moh A. Radaideh
Nasrullah Memon
Olivera Marjanovic
Omer Eldai
Paul C K Kwok
Paula Swatman
Philip Tsang
Phillip Brook
Premaratne Samaranayake
Robyn Davidson
Robyn Lawson
San Murugesan
Shailendra Palvia
Stan Glaser
Ta-Wei Hung
Tomasz Müldner
Tsung Teng Chen
Vijay Khandelwal
Wang-Jiunn Cheng
Yi-chen Lan
Yogesh Deshpande
Yung-Hsin Wang
Zahir Irani
Zhaohao Sun

Authors' Index

Shahid Siddiqi
Shaiq A. Haq
Shu-Nu Chan
Smawati Jaffar
Stephen Weeding
Ustun Yildiz
Victor Taratoukhine
Vijay Khandelwal
Virpi Pirttimaki
Yogesh Deshpande
Yuh-Huey Chen
Yun-yi Liu
Yu-Wei Chuang

Acknowledgements

We wish to further acknowledge the effort of individuals who have played a role in putting together the conference proceedings. First and foremost, all contributing authors need a special thank for taking the time and interest in submitting their papers to this conference. The work of the many honest and eminent reviewers is also again acknowledged and thanked. Finally, the members of the organizing committee as well as volunteering higher-degree students from the school of computing and IT deserve thanks for their contributing in making this project a success. Following individuals are specifically noted:

Abbass Ghanbary (Editorial Assistance)

Dinesh Arunatileka (Editorial Assistance)

Joanne Curry (Electronic Version of Proceedings)

Anant Dhume (Cover Design)

Christine Holliday (University of Western Sydney Press)

TABLE OF CONTENTS

Humanity, Technology and Environment: A partnership for the Future

Prof. Samuel K. C. Chang
Shih Chien University, Taiwan

Abstract:
In this age of rapidly moving technology, it is imperative that a balance is struck between technological innovations and their impact of human life and the environment. This is so because without due considerations to the environment, technology is unlikely to provide benefits to humanity. In fact, it may end up being counter-productive. This keynote address looks at the technology within the context of the environment in which it exists, as well as its impact on humanity.

Keywords:
Technology, humanity, environment, innovations

From eTransformation to eCollaboration: Issues and Solutions

Professor Athula Ginige
AeIMS Research Group, University of Western Sydney
Email: a.ginige@uws.edu.au

Abstract
Since year 2000 Advanced enterprise Information Management Systems Research Group (AeIMS) has been studying how Small and Medium Enterprises (SMEs) can make use of advances in Information and Communication Technology (ICT) to become globally competitive by making the business processes efficient and effective. Our test bed has been over 72,000 business organisations in the Western Sydney region of Australia.

We learned early in our study that eTransformation cannot be done as an one off project, but it is an iterative process. Based on this finding we developed an eTransformation Roadmap and a methodology that can be used to iteratively transform the organisations.

When testing this methodology through a series of action research cycles we learned that to get optimum benefit not only we need to change the business processes, but also the skills of people, management style, and the need to introduce new policies and procedures to manage the new ICT systems. We have also observed that these changes result in a major change in the organisation culture. Thus we extended the eTransformation methodology to provide a holistic approach to eTransformation, now known as 7 Es in eTransformation. The organisations that adopted this holistic approach were able to introduce new business models. eCollaboration was a major aspect of these new business models.

The existing software development methods that require the system requirements to be fully specified at the start of a software development project were found not suitable to support the iterative eTransformation approach. As eTransformation happens in an iterative manner the system requirements also evolves in an iterative manner. Also once an application is developed it needs to evolve with changing business needs. To address these issues we adopted CBEADS; Component Based E Application Development System to develop software applications to support re-engineered business processes.

Overall our research studies have shown that the eTransformation happening among SMEs as a result of advances in ICT requires significant changes to business processes, business models, people skills, management styles, ICT systems and organisational culture. We need to take a holistic view to get optimum benefit of eTransformation. It is now necessary to further develop models, methodologies and technologies to support this eTransformation process.

Keywords:
e-transformation, e-collaboration, e-transformation methodology, component based software development, business process reengineering

INTRODUCTION

Since year 2000 Advanced enterprise Information Management Systems Research Group (AeIMS) has been studying how Small and Medium Enterprises (SMEs) can make use of advances in Information and Communication Technology (ICT) to become globally competitive by making the business processes efficient and effective (Arunatileka & Ginige, 2002, 2003a, 2003b; Ginige, 2002, 2003, 2004; Ginige, Murugesan, & Kazanis, 2001; Ginige, Murugesan, Khandelwal et al., 2001; Kapurubandara, Arunatileka, & Ginige, 2004; Kazanis, 2003; Kazanis & Ginige, 2002, 2003; Lawson, 2003; Marmaridis & Ginige, 2005).

We used a combination of research methods; surveys, action research (Kock, McQueen, & Scott, 1997) and case studies in our investigation. Our main test bed has been Western Sydney region of Australia. This is the fastest growing economic region in Australia at present. There are over 72,000 business organisations in this region; over 80% are Small to Medium enterprises (SMEs). Further 1.8 million or nearly 10% of Australian population live in this region.

In our studies we identified the following.

- eTransformation is an iterative process; it cannot be done as a one off project.
- Introduction of ICT into an organisation to enhance its business processes needs a holistic approach.
- The organisations that adopted this holistic approach were able to introduce new business models. eCollaboration was a major aspect of these new business models.

- There is a need for a new methodology for developing software applications that can evolve with changing business needs.

This paper describes above observations in detail, approaches we have developed based on above observations to support eTransformation of SMEs and solutions we have developed to address some of the issues that were identified.

An important aspect of our research is its focus on SMEs. SMEs cannot invest large amount of money up front to implement ICT solutions nor can they afford to release people for considerable amount of time to participate in activities such as requirements gathering or training sessions.

eTRANSFORMATION IS AN ITERATIVE PROCESS

In year 2000 University of Western Sydney (UWS) together with New South Wales State Government, an Industry partner (UNISYS) and a Technical College (NSW TAFF) set up a center of "Excellence in eBusiness". The aim was SMEs in the region to come up with projects to develop eBusiness systems such as eCommerce Web sites, Customer Relationship Management (CRM) Systems, Supply Chain Management (SCM) Systems etc and the Center to implement these for the SMEs. But no significant projects from the SMEs came to the Center. In our interviews with CEOs and Managing Directors of SMEs reveled the problem was that they did not know "Where to start, How to proceed, and what dollar values to be used to value the perceived business benefits to decide whether money that they need to spend make a sound investment".

We also surveyed over 500 organisations in the region to find out how they were using ICT in their business operations(Ginige, Murugesan, Khandelwal et al., 2001). Findings of this survey and the above mentioned observations lead us to perform a detail analysis of business processes in an organisation with respect to ICT.

We identified that the business processes of an organisation can be divided into internal and external processes based on whether the people who use these processes are entirely within the organisation or external to the organisation, such as customers and suppliers. We then developed a way of mapping the evolution of these processes as these gradually become IT enabled (Ginige, Murugesan, & Kazanis, 2001). This mapping is shown in Figure 1.

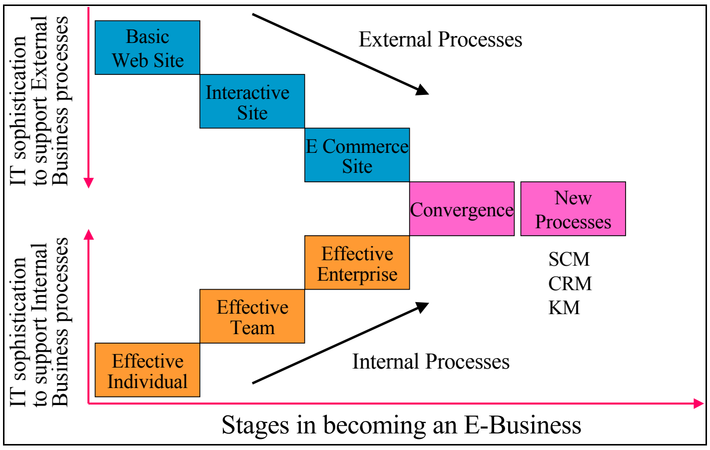


Figure 1:Roadmap to become an eBusiness

An important aspect about this diagrammatic representation was that each stage in this diagram corresponds to having specific hardware, software and networking capabilities within the organisation as shown in the table 1 below.

Table 1	
Stage	**Hardware, Software and Networking Infrastructure**
Effective Individual	Individuals with stand-alone computers having productivity software such as accounting packages spreadsheets, word processors etc installed. Possibly connection to the Internet for individual e-mail accounts.

Effective Team	Computers are networked. People can work in teams using networked applications. Providing e-mail and Intranet capabilities can enhance the productivity of a team.
Effective Organisation	Organisation now uses enterprise wide applications - a single application that supports different sections of the organisation such as purchasing, sales, accounting, manufacturing etc. thus enabling information integration and sharing across enterprise.
Basic Web Site	Organisation having its own domain name and "Brouherware" type web site hosted with an ISP.
Interactive Web Site	Organisations having Web sites that provide two way flow of information. From these Web sites users should be able to get immediate responses to structured queries such as a quotation for a particular product configuration the user has selected. Also, another feature would be to provide personalised information to frequent visitors. These types of Web sites can be hosted with an ISP, although are better hosted on site. This requires a Web server and a high speed dedicated connection to the Internet.
eCommerce Site	At this stage the organisation should have secure Web servers to facilitate financial transactions or a link to a payment gateway to obtain this facility.
Convergence	Organisation has now achieved integration of information that needs to support all its business processes. The flow of information of an organisation that has reached the level of convergence can be viewed as shown in Figure 2.
New Processes	Such organisations can now develop new processes, such as supply chain management (SCM), customer relationship management (CRM) and knowledge management (KM).

It is very important that various IT systems in an organisation can share the same information without replicating them. This we call state of convergence. At this state most of the information that is used by the organisation is in digital form and is available for all appropriate people to use and process. To achieve this state it is essential to have a very well thought out eTransformation process. Often what happens is organisations implement specific solutions using different software applications. Then it becomes impossible to share data across these different software applications. If data such as a list of employees and their details has to be duplicated, then it is not possible to achieve the state of convergence. When an organisation has to keep multiple working copies of the same data it becomes very difficult, if not impossible, to keep all copies up to date and consistent.

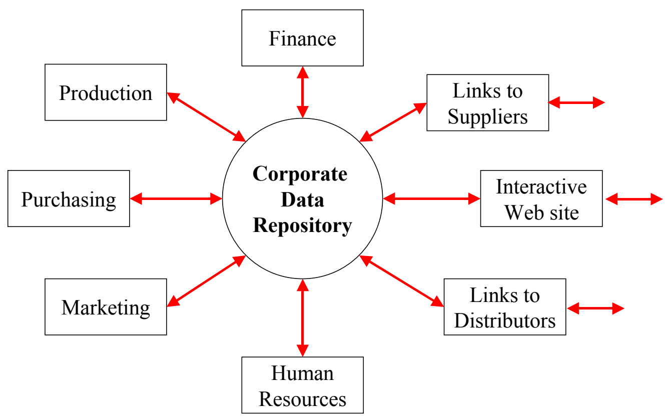


Figure 2:Flow of Information in an organisation that has reached the convergence stage

Once we developed the eTransformation Roadmap it became clear to us that eTransformation can not be done as a one-off project. It is rather a continuous, evolutionary process. When progressing along the Roadmap it is important to keep in mind the need for convergence.

We then developed an eTransformation methodology that SMEs can use to follow the eTransformation Roadmap (Ginige, Murugesan, & Kazanis, 2001). The steps in our eTransformation methodology are shown in figure 3.

First step is to analyse the business and identify how introduction of ICT can provide a benefit to the business. Once an eTransformation strategy is developed we then identify the critical business processes for that strategy to succeed. We then model the existing business processes and study how these processes can be reengineered to be more effective, develop appropriate software to optimally manage the flow of information required to carry out that process and, train the front-line people how to use the new process and the eApplication developed to support it. Then we select another business process and the cycle continues. The order in which business processes get transformed is selected based on an overall eTransformation strategy developed after carrying out a business analysis.

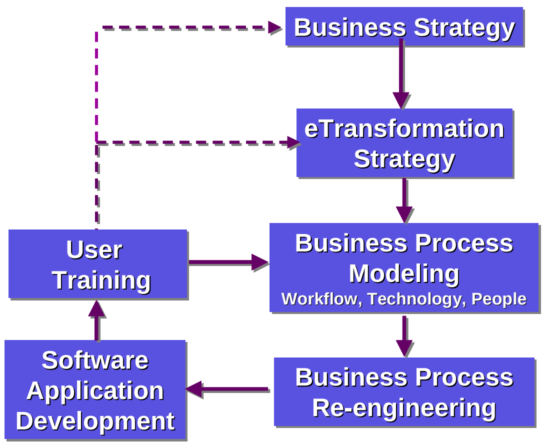


Figure 3. A Process for eBusiness Transformation

We then tested the above eTransformation model through a series of action research cycles(Arunatileka & Ginige, 2003a, 2003b; Chandrarathne & Lan, 2003; Kazanis, 2003; Kazanis & Ginige, 2002, 2003). In these studies we learned that for a SME to fully benefit from their ICT investment not only that they need to look at business processes and ICT to support these processes but also the internal and external environment, readiness of the organisation to undertake eTransformation process, leadership style, long term strategy to develop ICT skills of the employees, ICT system management procedures, overall organisational culture etc.

eTRANSFORMATION NEEDS A HOLISTIC APPROACH

From the above observations we realized that eTransformation needs a holistic approach. We then looked at existing literature on organisation transformation (Hammer & Champy, 2001; Pricewaterhouse-Coopers, 1999; Sawhney & Zabin, 2001) and developed a new approach "The 7E's in transformation" combing and extending relevant features of existing eTransformation methodologies and models (Arunatileka & Ginige, 2003b). This model is shown in figure 4 and table 2 provides the details of each step.

This model consists of seven very important aspects of eTransformation. Each stage is important in it's own right and as a part of the whole process. The seven stages, where six stages could be achieved one after the other are linked to the seventh stage 'Evolution', which deals with the crucial issues related to change management.

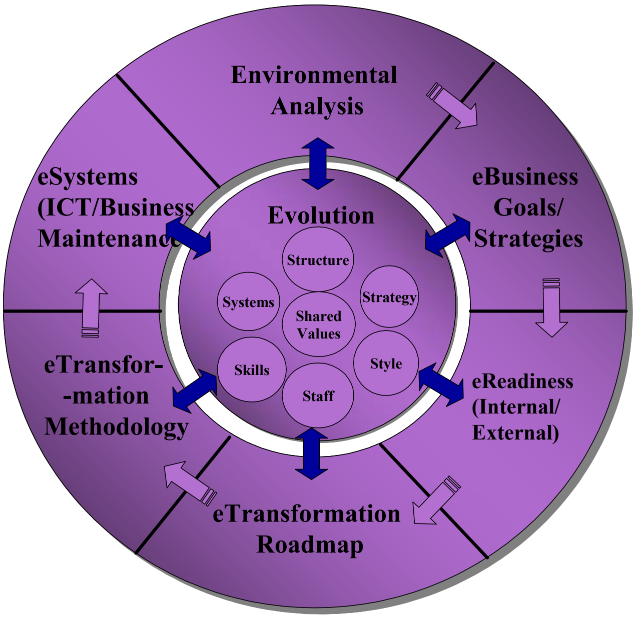


Figure 4. 7Es in eTransformation: A holistic Approach

Table 2	
Stage	**Activities**
1. Environmental Analysis	To understand the Global IT and Business Trends and the Company's Strategic Situation
2. eBusiness goals/Strategies	Develop eBusiness goals and strategies to gain the competitive advantage based on findings from environmental analysis.
3. eReadiness	To find out eReadiness of the organization in order to identify major barriers and issues related to change management
4. eTransformation Roadmap	To map the current position of the organisation on the eTransformation Roadmap and to develop the sequence of iterations to transform the business processes.
5. eTransformation Methodology	To carry out various activities as per the eTransformation methodology (Business processes modeling, re-engineering, developing software applications and user training) in an iterative manner as decided in step 4.
6. eSystems	To develop and implement maintenance policies and procedures for the implemented systems
7. Evolution – Change Management	Manage the changes in an evolutionary manner. This activity needs to happen every time when any of the other 6 activities take place.

The seven factors, which comprise the 7E's model was applied to the participating SMEs in this study. This holistic approach proved to be successful with the companies that we worked with (Arunatileka & Ginige, 2003a).

A HOLISTIC APPROACH WILL GIVE RISE TO NEW BUSINESS MODELS

We used the 7E model to eTransform 4 SMEs in the Toolmaking Industry. The Australian Toolmaking Industry is part of a wider Manufacturing Industry, and consists of about 600 organizations with over 6,000 employees. The majority of these organizations fit into the SME category (ABS, 2001) with 90% employing less than 50 employees. Over the past decade there has been a reduction in employment, which corresponds to an increase in imported tooling (Austool, 2004).

We investigated using action research how local toolmakers can respond to the threat of this diminishing market share by making their business processes efficient and cost effective with the help of ICT. Four SMEs engaged in toolmaking participated in this study. We based our study on the 7E model.

The environmental analysis reveled that the market share depends on competitive pricing, delivery time and quality. Achieving the required quality was not a problem for Australian Toolmakers. High cost of labour and longer delivery times due to limited working hours and machine availability to produce the tools immerged as major factors that need to be addressed. Also we found that there was not much marketing done to reach new customers due to lack of time and money (Lawson, Arunatileka, Ginige, & Hol, 2005).

It became clear that through collaborative strategies such as collective marketing, resources sharing, project sharing and combined purchasing some of these problems could be addressed. In the past these SMEs have been competing for the same customers. Thus "Trust", how much information they were happy to reveal to the other collaborating partners became a major issue.

A pilot collaborative manufacturing activity was carried out using email, fax and telephone to communicate with one another. It was found for a collaborative manufacturing activity among 4 partners this mode of communication was not effective. It became clear that there is a need for a better method of accessing the design drawings, and information such as cost and delivery times when partitioning the job among collaborative partners as well as tracking the progress during the manufacturing process.

This lead us to develop a strategy based on eCollaboration. In line with the Action Research nature of this project, it was identified that it would be necessary to develop the initial website before proceeding with other aspects of the project. This enabled the toolmakers to be placed on the eTransformation Roadmap, and did indeed improve the trust levels with the visual development of the website. The second stage was to develop web based applications to support business processes that they use to collaborate when manufacturing new tools for customers. The strategies for strategic marketing, resource sharing, project sharing and purchasing were agreed to by the toolmakers, which allowed work to proceed in the development of the software tools for job sharing and quoting, as well as the development of a website marketing plan (Lawson et al., 2005).

Analysis of the results to date suggests potential advantages are feasible in eCollaboration among the toolmakers. Certainly, non-technical issues such as social, behavioural and strategic management dimensions, will ultimately drive collaboration. The toolmakers have made the decision to work collaboratively and, as identified by (Boddy, MacBeth, & Wagner, 2000), successful implementation is posing some challenges.

SOFTWARE APPLICATIONS SHOULD EVOLVE WITH CHANGING BUSINESS NEEDS

We came across another challenge when we were trying to develop software applications for eTransforming organisations. As eTransformation is an iterative process it was not possible to get complete specifications for all the software requirements at the start (Ginige, 2001).

We first studied the possibility of using an Enterprise Resource Planning (ERP) package to support the reengineered business processes in an organisation. Information from an European survey of 2400 companies show that the main criterion for the selection of an information system is the best fit with current business procedures but, several studies have shown that configuring and implementing ERP solutions can be costly and may even require reengineering entire business operations (Everdingen, Hillegersberg, & Waarts, 2000). The high cost and the problems associated with configuration and implementation together with the one-off project nature of implementation of an ERP solution not supporting the evolution of the organisation makes it very unattractive to an SME.

The other option is to customer build software that supports existing and/or re-engineered business processes and the iterative nature of the eTransformation process. This approach also suffer from many draw backs such as projects not finishing on time, running over budget and many on going maintenance problems.

Also any software development approach that requires the system to be fully specified before detailed development work can start can't be used to develop eApplications. As eTransformation is an iterative process, the software application requirements will also keep evolving. Design for evolution has to be a very important requirement in any eApplication development process.

It became clear to us due to above mentioned factors that we need to find a new approach to develop eApplication software (Ginige, 2002, 2003). Armour, in his article “The case for a new business model: Is software a product or a medium?” argues that software can be viewed as a medium which can be used to store knowledge (Armour, 2000). If we take this view then what we need is a software framework, which can be set-up in an organisation that is capable of capturing and storing the organisation’s knowledge and information and provide appropriate access to relevant people.

As various eApplications get developed at different times we should be able to plug in the new applications into this framework when these get developed. This framework should provide a consistent user interface to the users for all the applications. Also it should provide a mechanism to link different applications in terms of sharing data among applications and also managing the workflow information. This is a challenging task as these different eApplications get developed at different times over several years. When developing an eApplication, developers don’t know what new eApplications will be required in years to come.

In order to implement this new paradigm for software development we developed CBEADS: Component Based E Application Deployment Shell (Ginige, 2002). This shell runs in conjunction with a Web server and provides a Web based interface to interact with it. As this is a Web based application it can easily be deployed as an Intranet within an organisation as well as an Extranet linking different organisations.

The shell itself is made of components. These components can be grouped into two major sub systems. The first sub system is CORE CBEADS that provides the overall framework to which different eApplications can be plugged in. It consists of a security module, system components, system database and a workflow component. The second is made of various applications plugged into the shell. These consist of application components and application databases. The overall architecture is shown in figure 4.

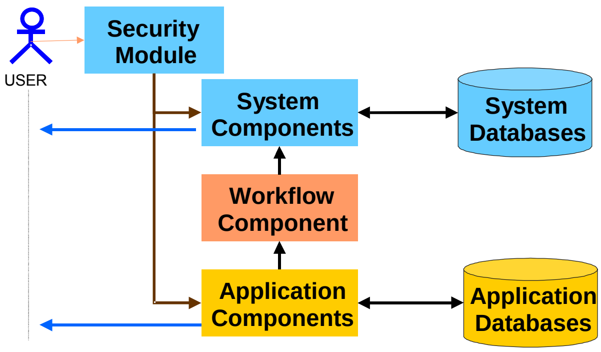


Figure 4: Overall CBEADS Architecture

At present we have deployed CBEADS in many organisations. This framework has enabled us to develop various applications at different times and plug into the CBEADS framework with out any difficulty. Also by having a framework that provides common functions such as authentication and group based access control together with smart development tools have drastically reduced the time required to develop these applications.

CONCLUSIONS

The research work that we have done in the last 6 years has shown that SMEs if properly guided can enhance their business processes to become globally competitive. Fist they will automate some of the routine tasks such as accounting and payroll by using appropriate software applications. Next they start to transform into eBusiness organisation by taking a holistic view about all the business operations and move towards the state of convergence in the eTransformation Roadmap. Then they will explore new business models to further benefit from the advances in ICT. These new business models can be based on eCollaboration among supply chain members or organisations that provide the same product or service.

These 3 phases are shown in figure 5.

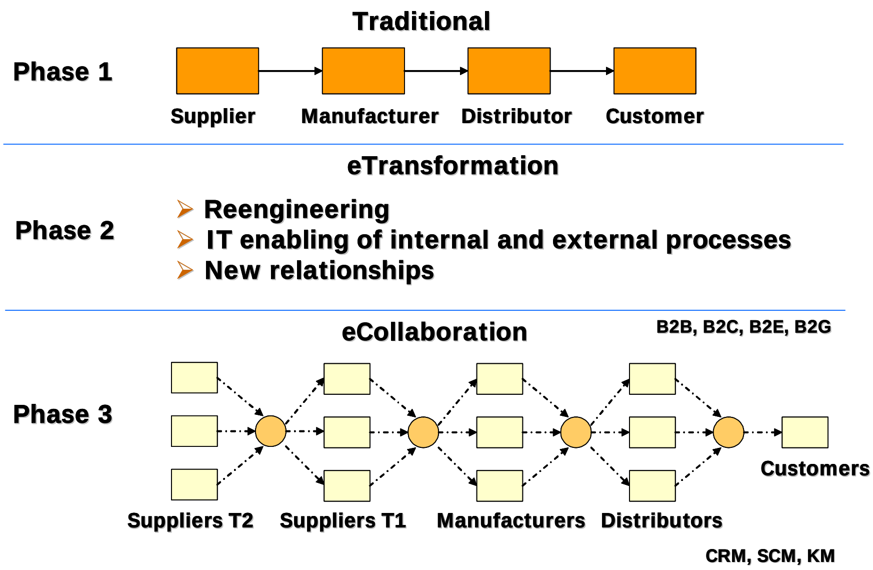


Figure 5: From eTransformation to eCollaboration

These phases are not mutually exclusive. While an organisation is in one phase it can start to investigate and trial activities in another phase.

Conceptualising the overall transformation taking place among SME sector into phases and providing models, methodologies and technologies to support this transformation has assisted many SMEs to undertake the eTransformation process and explore new collaborative business models to gain competitive advantage.

ACKNOWLEDGEMENTS

This paper brings together work done by many researchers attached to the AeIMS research group at University of Western Sydney since year 2000. They are Dr. Vijay Khandelwal, Dr. Robyn Lawson, Prof. San Murugesan, Dr. Phillip Kazanis, Nectarios Costadopolis, Tony Pollard, Shiromi Arunathilaka, Ana Hol, Makis Marmaridis, Anupama Ginige, Danny Liang, Mahesha Kapurubandara Buddhima de Silva and Dr. Yi-chen Lan. I wish to thank them for their valuable contribution.

REFERENCES

ABS. (2001). *Manufacturing Industry Report 8221.0 – 8221.5 1992-2000 (states); 1992-2001 (national)*, from www.abs.gov.au

Armour, P. G. (2000, August 2000). The Case for a New Business Model: Is software a product or a medium. *Communication of the ACM, 43*.

Arunatileka, S., & Ginige, A. (2002). *E-Transformation Strategy for the Sri Lankan Garment Industry.* Paper presented at the International Conference on Information and Communication Technologies - IICT 02, Colombo, Sri Lanka.

Arunatileka, S., & Ginige, A. (2003a). *Applying seven E's in eTransformation to the Manufacturing Sector.* Paper presented at the cChallenges, Bologna, Italy.

Arunatileka, S., & Ginige, A. (2003b, 3-6 June 2003). *The Seven E'S in eTransformation - A Strategic eTransformation Model.* Paper presented at the IADIS International Conference - e-Society 2003, Lisbon, Portugal.

Austool. (2004). *An Australian Tooling Industry Technology Roadmap: Exploratory Study 2003/2004*: Austool Limited.

Boddy, D., MacBeth, D., & Wagner, B. (2000). Implementing collaboration between organizations: an empirical study of supply chain partnering. *Journal of Management Studies, Vol. 37*, pp.1003-1027.

Chandrarathne, A., & Lan, Y. (2003, 23 -25 April 2003). *Succeeding in digital business: Investigation of hurdles for e-transiting micro companies.* Paper presented at the ISOneWorld, Las Vegas, USA.

Everdingen, Y. v., Hillegersberg, J. v., & Waarts, E. (2000). ERP adaptation by European Midsize Companies. *Communications of the ACM, Volume 43*, pages 27-- 31.

Ginige, A. (2001, 5 - 7 September 2001). *New Paradigm for Developing Software for E-Buisness.* Paper presented at the IEEE Symposia on Human-Centric Computing Languages and Environments, Stresa, Italy.
Ginige, A. (2002). New Paradigm for Developing Evolutionary Software to Support E-Business. In S. K. Chang (Ed.), *Handbbok of Software Engineering and Knowledge Engineering* (Vol. 2, pp. 711 725): World Scientific.
Ginige, A. (2003, 1 - 3 July 2003). *Re Engineering Software Development Process for eBusiness Application Development.* Paper presented at the Fifteenth International Conference on Software Engineering and Knowledge Engineering, San Francisco Bay, USA.
Ginige, A. (2004, 26-27 March 2004). *Collaborating to Win - Creating an Effective Virtual Organisation.* Paper presented at the International Workshop on Business and Information, Taipei, Taiwan.
Ginige, A., Murugesan, S., & Kazanis, P. (2001). A Road Map for Successfully Transforming SMEs into E-Businesses. *Cutter IT Journal, The Journal of Information Technology Management, 14*(5), 39 - 51.
Ginige, A., Murugesan, S., Khandelwal, V., Pollard, T., Costadopouls, N., & Kazanis, P. (2001). *Information Technology in Western Sydney: Status and Potential*. Western Sydney: University of Western Sydney.
Hammer, M., & Champy, J. (2001). *Reengineering the Corporation – A manifesto for Business Revolution*. London: Nicholas Brealey.
Kapurubandara, M., Arunatileka, S., & Ginige, A. (2004). *Application of eBusiness Strategies for SME's in Developing Countries.* Paper presented at the IEEE International Conference on e-Technology, e-Commerce and e-Service, Taipei, Taiwan.
Kazanis, P. (2003). *Methodologies and Tools for eTransformaing Small to Medium size Enterprises.* Unpublished Ph.D, University of Western Sydney.
Kazanis, P., & Ginige, A. (2002, 1-2 December 2002). *Asynchronous collaborative business process modelling through a web forum.* Paper presented at the 7th Annual CollECTeR Conference on Electronic Commerce - CollECTeR 02, Melbourne, Australia.
Kazanis, P., & Ginige, A. (2003). *Process Modelling and eTransformation for Small to Medium Size Enterprises.* Paper presented at the eChallenges, Bologna, Italy.
Kock, N., McQueen, R. J., & Scott, J. L. (1997). Can action research be made more rigorous in a positivist sense? The contribution of an iterative approach. *Journal of Systems & Information Technology, 1*(1), 1–24.
Lawson, R. (2003). *Diffusion of Electronic Business: Adoption Patterns in Manufacturing Small and Medium sized Enterprises.* University of Wollongong.
Lawson, R., Arunatileka, S., Ginige, A., & Hol, A. (2005, June 6 - 8, 2005). *A Pilot Project on eCollaboration in the Australian Toolmaking Industry.* Paper presented at the 18th Bled eConference: eIntegration in Action, Bled, Slovenia.
Marmaridis, I., & Ginige, A. (2005, 27 - 29 July 2005). *Framework for Collaborative web applications.* Paper presented at the 5th International Conference on Web Engineering, Sydney, Australia.
Pricewaterhouse-Coopers. (1999). Seven Co-business Strategy Formulation, .
Sawhney, M., & Zabin, J. (2001). *The Seven Steps to Nirvana – Strategic Insights in to eBusiness Transformation*: McGraw-Hill, USA.

A Design for an Artificial Intelligence–based Automated Optical Inspection and Reviewing System in Semiconductor Defect Detection

Alan P.C. Sun, David Ho, Yi-Chin Lai, Cheng-kuan Lin
National Kaoshiung Normal University[1], HY Neural Technologies Ltd
Kaohsiung, Taiwan[1], Taipei, Taiwan[2]
sun@nknu.edu.tw hynt, david@msa.hinet.net, vickieqo@icemail.nknu.edu.tw, ibm88540@yahoo.com.tw

Abstract

This research paper is concerned with the Taiwan semiconductor industry's ability to improve its efficiency and accuracy of manufacture. We propose a new approach to developing a system for automatic wafer inspection and defect analysis by using neural networks, fuzzy-expert and AOI technologies. This approach not only can identify the defect types efficiently but also can effectively ameliorate the manufacturing process and production management. Empirical results show that the proposed approach is efficient enough to solve practical applications and achieve good performance. Thus, the value of the proposed approach has been verified.

Keywords:

Neural Networks, Fuzzy Logic, Wafer Inspection, AOI, Defect detection

INTRODUCTION

Automated Optical Inspection (AOI) is a necessary production yield management tool (Badih, 1997). It is an important inspection technology in the semiconductor industry. This research paper is a university and industry joint undertaking. We proposed a new approach combining AI (artificial intelligence) and AOI technologies to develop a system for automatic wafer inspection. Empirical results show the proposed approach is efficient enough to solve practical applications. We hereby introduce the approach and systematic design for inspection of the bumps on wafers, as follows:

There are three major types of AOI technologies being used on the market: template matching, pattern recognition and adaptive knowledge-based processing. They can tell if there are any defects or not; however, they cannot tell the defect type accurately. Defining the defect types and accounting for the numbers of the defect type are very important, because when abound defects appear, an increase in cost and deterioration in performance will result. However, we don't have enough information to support improvement of the manufacturing process. For this reason, we must try to find out which production stage caused the defects and provide information on the causal relationship to ameliorate the manufacturing process. Hence, our research developed a system utilizing neural networks and fuzzy logic methodologies to determine the defects, and the types of defects on the bumps. Therefore, it can achieve the goal of providing the information necessary for improving the manufacturing process and production management.

There are several steps to accomplishing this analysis of the features of bump defects and to using geometry and gray levels to extract ten kinds of bump image features. After extracting the regularities and irregularities of the features of the bump defects, we can create neural network learning data sets. Moreover, we can utilize the back-propagation algorithm to train the features of defects on neural networks. After the networks are trained, the system can identify the defect type of the new sample under inspection, and thus we can know whether the inspection has resulted in a pass or fail. Especially noteworthy, we defined a fuzzy-expert system to create a knowledge database to judge the correlations between defect types and the cause of defect for improving production management. Most important, this can meet the practical needs of the semiconductor industry.

THREE MAJOR TYPES OF AOI TECHNOLOGIES ON THE MARKET

There are three major types of AOI technologies being used on the market as follows:

Template Matching

The first methodology used is template matching (see Fig. 1). Template matching is extensively used in low-level vision tasks to localize and identify patterns in images (Brunelli & Poggio, 1995). Template matching compares the image in question to a "golden" image (called the template) and uses a set of specified threshold values to determine what magnitude of difference constitutes a potential defect. Template matching is one of the oldest and most straightforward techniques used by vendors in the AOI marketplace.

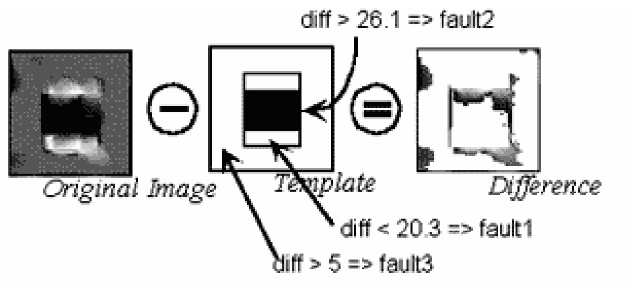


Fig. 1. Template matching technology.

Template matching uses an idealized version of the original image to compare with a captured image, along with threshold value limits to determine where a defect might occur.

Template matching is simple, the technology is very easy to implement in terms of both software and hardware. It seems reasonable that each image of a part should look very similar to a previous image of that same part.

With these benefits, there come drawbacks, however. While template matching technology is conceptually simple to understand and implement, it is very difficult to extend. There are two reasons for this: 1) the smallest feature in the pattern dictates the overall tolerance of the circuit (Thomas et al. 2002).

2) As new defects arise, functional additions must be made to the template decision method (i.e., the magnitude of the differences triggers the need to run computational functions, and the defect decision is made based on the output of the function). Prior hardware implementations of camera size and resolution may pose a problem if higher resolutions are needed to distinguish smaller defect sizes.

Pattern Recognition

Pattern recognition (see Fig. 2) is a way of pattern matching. Pattern recognition applications take many forms. In some cases, there is an underlying and quantifiable statistical basis for the generation of patterns. In other cases, the underlying structure of the pattern provides the fundamental information for PR (Schalkoff, 1997).

Pattern matching uses a set of stored models of both positive and negative aspects of the classification task and searches for those models (or patterns) in the image in question. The patterns represent a common, or prototypical construction of a type of part or defect. If the pattern is found, then the image in question is given the same classification type as the pattern that was found in it.

Among the various frameworks in which pattern recognition has been traditionally formulated, the statistical approach has been most intensively studied and used in practice (Jain et al. 2000). Statistical Pattern Matching compares the image in question to a specified statistical pattern of what the image should look like. Since the pattern is statistical, it will be found in any image to varying degrees. If the degree of match exceeds a specified threshold, the image is classified as positive. If the degree of match falls short of the threshold, the image is classified as negative.

The pattern recognition approach extends template matching by extracting a set of lines from a training image and using those extracted lines as a pattern from which to do a GR search (GR technology, 2002) in other images. There are two key aspects of the pattern recognition approach: first, selecting the lines that will form the pattern; and second, determining the algorithm that will tell how closely the image being tested matches the specified pattern. Determining these two aspects lies at the heart of the pattern recognition approach and forms the main source of both the power and the weakness of the technology. Companies like Cognex Technologies (Cognex, 2003), Accumen (Accumen, 1999) and others have used this technology for AOI.

There are several reasons why a company may choose to implement pattern recognition technology:

Data reduction: The approach has significant power. By extracting lines from the training images it significantly reduces the amount of raw data that needs to be manipulated by the AOI system. Such a system can be very fast.

Expressiveness: By varying the matching function, the system can detect a large number of defects.

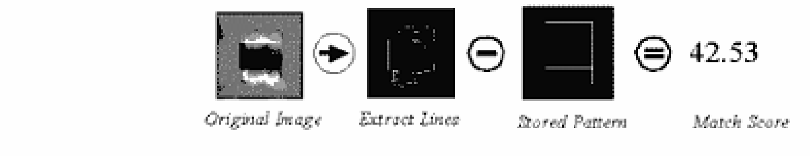


Fig. 2. Pattern recognition technology

Making the cut: One of the most difficult tasks facing those implementing pattern recognition technology is the determination of which lines to include in the patterns and which to exclude. Creating hard and fast rules for this is almost impossible. But if those rules do not exist, then it is left to the customer to determine through trial and error which parameters generate the best patterns for each different type of component, or even each individually placed component.

Match making: Similarly, there are no hard and fast rules for determining how close a match to the pattern must be found in a new image in order for the AOI system to say that a defect is present. It is, again, left up to the user to decide whether to program for each component family, or each individually placed component.

The statistical pattern recognition approach (see Fig. 3) was developed partially in response to the difficulties found with standard pattern recognition. In statistical pattern recognition, sets of lines are extracted from a number of images to form a more generally applicable pattern. By creating the pattern from a set of images rather than just one image, it can more realistically reflect the variances that will typically be seen in images coming off the production line. Relatively few companies are using the statistical pattern recognition approach, the principal one being Kestra (Kestra, 2002).

The statistical pattern recognition approach has advantages

It represents the variance in the image: By focusing on sample images that reflect the variances typically seen on the production line, the pattern recognition system ably represents those variances. In situations where there is a ready stock of images that display the variances typically seen, the statistical pattern-recognition pattern generation will do a good job of representing those differences.

Greater expressiveness: Although the algorithm to determine the degree of match between the statistical pattern and the new image is generally the same as in standard pattern recognition, the expressiveness can be greater since the statistical representation of the pattern removes some of the small anomalies that the matching algorithm would previously have had to detect.

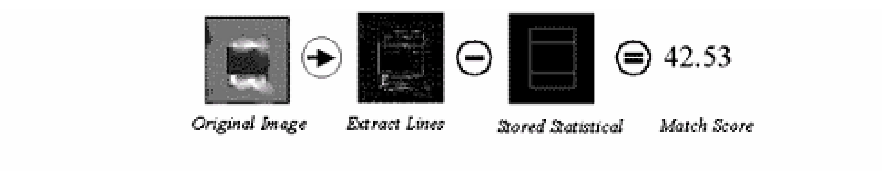


Fig. 3. Statistical pattern recognition technology.

As mentioned above, the same technological features that are the strength of the technology are also its most significant weaknesses. The disadvantages of the statistical pattern recognition approach are similar to those for standard pattern recognition:

What to include: The determination of which features to include in the pattern that represents the image is still a major problem for those implementing statistical pattern recognition systems. This is somewhat complicated by the need to determine which features match each other and which do not. There is a danger that, in the accumulation of features from different sources, the final statistical pattern will include artifacts that are not really part of the data. For example, consider the effects of combining the two sets of features shown in the Fig 4. There are two stipulations that can made. In the first case, the rule might be that the box can have either a tail on the left or a tail on the right or both. If this is the case, the pattern on the right side of the equation concisely represents that fact. If on the other hand, the rule is that the box can have either a tail on the left or a tail on the right but not both, the statistical pattern will not reflect that and additional programming will have to be done. This might include modifications to the matching procedure or splitting the pattern up into two patterns, one that represents tails on the left, and one that represents tails on the right.

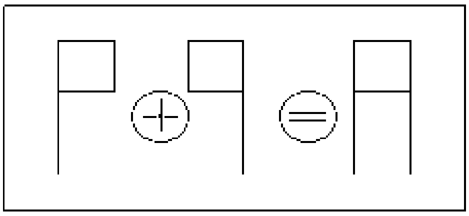


Fig. 4. Two sets of features

Match making: Even with the increased power gained by the use of statistical patterns, tuning the match algorithm to correctly classify positive and negative instances of defects is still very much a trial and error affair. In one sense, the problem is magnified in that the error in classification may be caused by incorrectly set parameters in the algorithm, or incorrectly set parameters in the pattern generation.

Adaptive Knowledge-based Processing

Adaptive Knowledge-based Processing extracts a large number of features from a set of training images to determine what features might be useful in distinguishing between positive and negative examples of an image.

Adaptive Knowledge-based Reasoning seeks to store generic inspection knowledge in a database and use that knowledge to generate high-quality defect detection and classification.

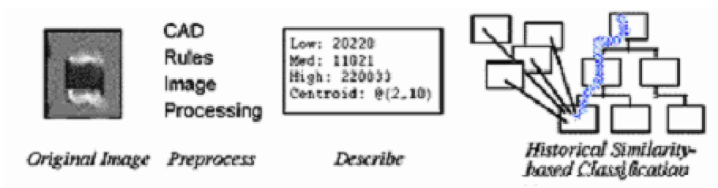


Fig 5. Knowledge-full similarity-based classification.

> The original image is converted through the use of CAD, set rules, and traditional image processing technologies to produce a set of features that describe the object in the image. The particular descriptors used complement built-in knowledge of classification and inspection in general.

Fig. 5 depicts the basic flow of operation. The image is obtained from a camera and is then transformed into a set of symbolic and real-valued features. This transformation is accomplished through the use of the board's CAD component location information, rules for determining what set of preprocessing routines need to be run on the image, as well as traditional image processing routines for generating feature values. The computation and aggregation of the feature values can be viewed as a redescription of the image that was previously represented as a set of pixels, and is now represented as a set of feature-value pairs. It may not be obvious that the set of feature-value pairs- or descriptors- is an accurate description of the image in question. In fact, it generally is not. However, it need not be a complete description of the image if the set of descriptors is chosen so that it provides the information necessary to determine if a defect exists and to which class it belongs, that is, if the descriptors reflect the semantics of inspection. Indeed, the knowledge base is engineered so that it provides a very general, very wide breadth of knowledge about the inspection and classification of board assembly components. This knowledge is stored in such a way as to employ a historical decision-making methodology of similarity-based classification. This method of classification uses a knowledge base from general inspection that is the same for all board types and is augmented with specific knowledge of how defects on different components of a given board typically look.

Summarizing the above three methods on the market, it is clear they can detect the difference between irregularities and regularities; however, they can't make the judgments called for according to the situations generated by the different defect types. Therefore, this will affect the final accuracy in judging the irregularities and regularities. In this way, accurate classification will not be achieved. What is called for is a process that can provide useful information to ameliorate the manufacturing process. This research provides a solution by combining ANN and fuzzy-expert methods to solve practical applications. The details will be discussed in what follows.

PROPOSED APPROACH

Pattern-Classification Neural Network and Fuzzy Logic

The explosive demand for electronic products that are smaller yet more powerful are driving computer chips to ever higher densities creating new challenges and placing new demands on the electronic assembly and manufacturing industries. A general methodology for inspecting any advanced packaging technology using neural network classification must show itself to be fast, accurate, flexible, and able to detect a variety of problem types.

Such a methodology uses features or signatures sampled across the object. The signatures are fed to a feature extracting subsystem to extract features selected both to facilitate classification and for translation invariance reasons. Several samples of different features are required to obtain an adequate representation, one whose number is inversely proportional to the uniformity of the object.

A pattern-classification neural network with a soft-max output and cross-entropy error term is trained on the extracted features using extended Delta-Bar-Delta learning, adaptive Gaussian noise, and optimal node pruning. The outputs are fed into a simple expert rule subsystem that decides the final classification: pass or fail. The classification system can identify a variety of problems including objects that are too high, too wide, have copper exposure, are missing, have scratches, are oxidized, deformed, or display irregularities of any kind.

Unlike other methodologies, this methodology is fast, accurate, flexible, and can handle objects of different sizes, colors, shapes, reflectance properties, and scale. An inspection system based on a neural network and fuzzy logic-based method for the automatic inspection of bumps will prove to be a highly flexible technique specially designed for advanced packaging inspection.

Six steps of the approach

This research project combines Fuzzy Logic with a Neural Network for automatic visual inspection of Bump operations. There are six steps to be identified in this workflow: Image input, Image processing, Feature extraction, ANN processing, Database preprocessing and Fuzzy-expert processing (see Fig. 6).

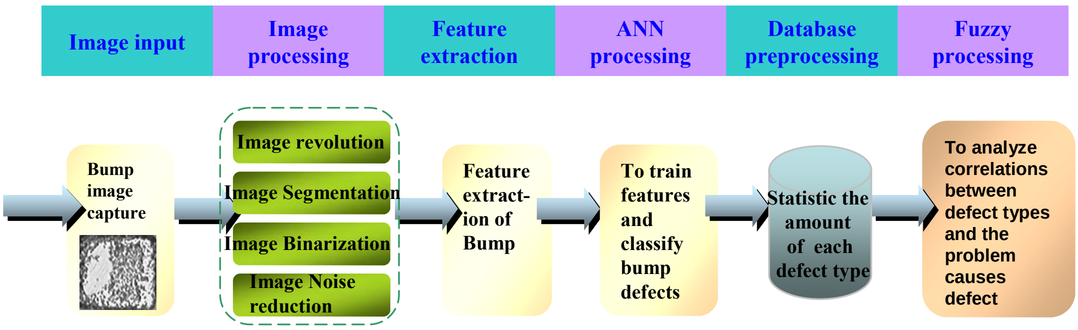


Fig. 6. The research major steps

- Image input: By use of traditional image capture technologies such as a digital camera/camcorder or scanner, the images are taken and stored in memory. The images form the basis for subsequent processing and inspection.
- Image processing: The captured images undergo further processing, such as noise reduction, image segmentation, revolution and binarization.
- Feature extraction: This step is to extract the features relevant to a Neural Network and to consider the criteria for accepting or rejecting them. Because the features of defect bumps are complex and various, we separately used geometry and gray levels to extract the specific features of the defect bump image for subsequent processing. We used “void on bump” to explain it as an example.

Let’s introduce the feature analysis of geometry first. It is clear that the features of “void on bump” can be observed by its shape. It’s usually a circular hole that can be seen through whose gray level value is close to 255 (this means white), and whose gray level value on the hole side is close to 0 (since the hole is black) (shown in Fig. 7). Therefore, we can figure out its features by using geometry as follows:

1. Form factor: we use the formula of “Form factor” to calculate if the value is 1 or not. The closer it is to 1, the closer will it approach to being a circle.

2. Radius, maximal and Radius, minimal: If the maximal value of the radius equals the minimal value of the radius, then it will be a circle.

3. Ratio: if the value is close to 1, then it will be a circle or a square.

4. Area: this can distinguish “void on bump” from “circle bump contamination” by area. It is impossible that the area of “void on bump” will be more than half the area of bumps, but the “circle bump contamination” can be more.

We not only detect defect features by geometry but also use the features of gray levels to enhance the results. Gray levels, including Contrast, Gradient std Dev, Gradient average, Gray level std Dev, Gray level average are explained as follows:

1. Contrast: contrast equals the maximal value of gray level minus the minimal value of gray level. If the defect is “void on bump” then its gray level value is white and close to 255, Therefore, it would be deemed a “void on bump”

2. Gradient std Dev.: Standard deviation of gradient of gray level image. This is suitable for identifying the defects of “void on bump”. In that the “void on bump” is a bump which can be seen through, we also use the feature to calculate the ranges of gradient.

3. Gray level std Dev.: Standard deviation of gradient inside the hole.

4. Gradient average: Average gradient inside the hole.

5. Gray level average: Average gray level inside the hole.

After repeated discussions with field engineers repeatedly, we were able to identify ten relevant features as above. Finally, they can be the inputs of a Neural Network.

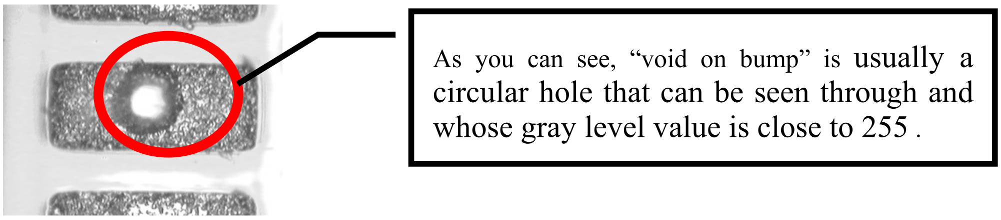


Fig. 7. The features of “void on bump”

- ANN processing: This research proposed a neural-network approach for defects on bump inspection. There are several advantages to using a neural network to solve a complex manufacturing system problem. They are: 1) high processing speed through massive parallelism; 2) effective learning and adapting by means of efficient knowledge acquisition; 3) robustness with respect to fabrication and different types of failures; and 4) compact processors for space- and power-constrained applications, etc. (Barschdorff & Monostori, 1991).

The literature of wafer defect detection has shown that ANN can achieve a better accuracy level than other machine learning algorithms, such as minimum distance algorithm, maximum likelihood classifier, Median Filter and Radon Transformation (Zoroofi, et al.2000; Huang, 2000). Among the many algorithms for ANN, we used the back-propagation algorithm to learn the defect features there. It was selected based upon prior research results that show the superior performance of back-propagation neural networks versus other supervised learning methodologies (Walczak, 1998). And they have been extremely popular due to their unique learning capability (Widrow et al., 1994) and have been shown to perform well in different applications, such as medical application (Tolle et al., 2000) and game playing (Chen et al., 1994).

A typical back-propagation neural network consists of a three-layer structure: input-layer nodes, output-layer nodes and hidden-layer nodes (Huang et al., 2003). Usually, there are one or two hidden layers (Liao et al., 1995); we used one hidden layer here, as shown in Fig. 8.

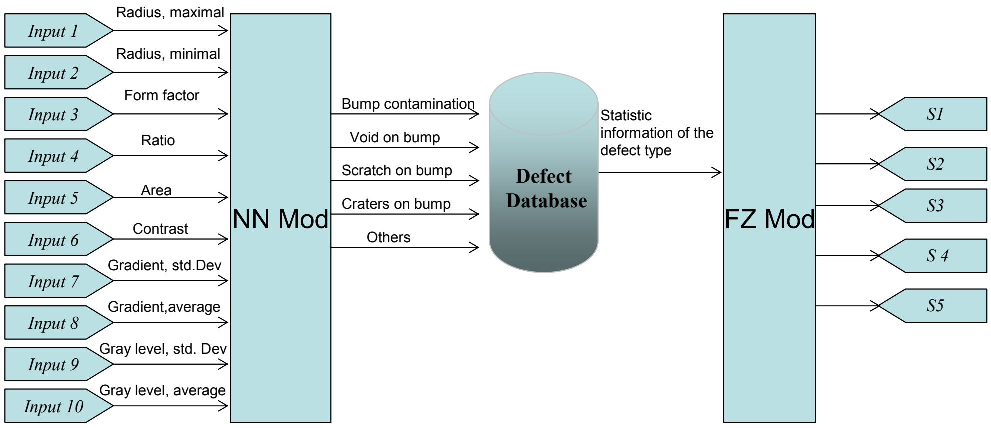


Fig. 8. The system model

- Database Preprocessing: According to the neural networks' output, the database system will save all the information on defect types for the bulk dies. There are five symbols representing the defect types: A is bump contamination, B is void on bump, C is scratch on bump, D is craters on bump and E is others.

- Fuzzy processing: We use fuzzy processing in order to find out which stage of the production line causes defects. In this step, we calculated the amount of each defect type in the database first. Unlike two-value Boolean logic, fuzzy logic is multi-valued. While classical logic operates with only two values 1(true) and 0 (false), Lukasiewicz introduced logic that extended the range of truth values to all real numbers in the interval between 0 and 1 (Negnevitsky, 2004). Fuzzy logic is determined as a set of mathematical principles for knowledge representation based on degrees of membership rather than on the crisp membership of classical binary logic (Zadeh, 1965). The concept of fuzzy logic can help us model the human reasoning process. It also allows the encoding of knowledge in a form that can be used to reflect the way humans think about a complex problem (Akhter et al. 2005) such as that of judging defects. The research endeavor was to collect engineers' experiences so as to analyze correlations between defect types and the problems that cause these defects due to fuzzification, inference and defuzzification. Thus it can not only provide information to engineers on the root-cause of defects but also ameliorate manufacturing process and production management. Fig. 8 shows the system model, whose fuzzy logic includes input elements, fuzzy-logic inferences and output elements, as follows:

  – Fuzzy logic input: According to the instructions on production experience, engineers can determine whether the amount of defects exceeds a certain value. If it does, there will be some problems. For example, given a normal amount of a specific defect type, the Fuzzy set is L. (Under normal conditions, the production line will randomly produce a few defects. However, there is no warning information.) When the amount exceeds a certain number by too much, then the fuzzy set is H. The amount between such intervals is M.

    After processing fuzzification, the crisp values are translated into linguistic variables, which are represented by fuzzy sets, and then three membership functions are used to distinguish the seriousness of the failure (shown as Fig. 9). According to the industry standard, the yield is about 0.99. For example, the amount of bulk dies is 10000 pieces and its defect amounts about 100. Therefore, we can set the numerical range of L as from 0 to 120, the numerical range of M is from 100 to 220 and the numerical range of H is over 200. Next, we need to obtain some fuzzy rules.

  – Fuzzy logic inference: In this step, we try to analyze the correlations between defect types and the root-cause of the defect. According to the engineers' experience, we can set some definitive rules, as follows. These rules consist of a precondition (IF-part) and a consequence (THEN-part). The precondition can consist of multiple conditions linked together with AND or OR conjunctions. The computation of fuzzy rules is called fuzzy inference, and we use max-min rules to do it. Finally, we use the MATLAB Fuzzy Logic Toolbox to build our fuzzy expert system. The rules of our system are as follows (see Table 1) .

Table 1 Some rules in our system

RULE1 IF A IS STRONG THEN S1 IS H

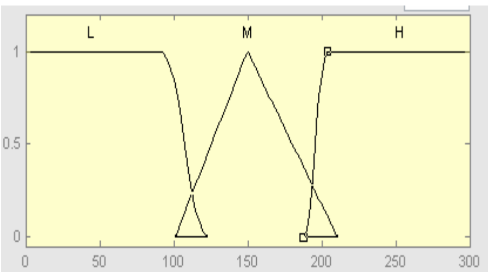
RULE2: IF B IS STRONG THEN S2 IS H
RULE3: IF C IS STRONG THEN S3 IS H
RULE4: IF D IS STRONG THEN S4 IS H
RULE5: IF E IS STRONG THEN S5 IS H
RULE6: IF (A IS STRONG) AND (B IS STRONG) AND (C IS WEAK) AND (D IS WEAK) AND (E IS WEAK) THEN Output1 IS H
………
……….

– Fuzzy logic output: In order to obtain the output value, the defuzzification step converts the fuzzy value to a single number, in other words, a crisp value. Moreover, the membership functions are presented by the fuzzy sets of L, M and H, as in Fig. 10. The numerical range of L is from 0 to 0.4, the numerical range of M is from 0.3 to 0.7 and the numerical range of H is from 0.6 to 1.

• Output: After the fuzzy inference and center defuzzification, the final output resulted. There are five stages identifying the root-cause of the defect. They are represented as S1 and as S2, S3, S4 and S5. The range of the output is from 0 to 1. After the fuzzy inference and center defuzzification, if the value of the output is close to 1, then it represents a high degree of that root-cause of defect. Since some defects are generated in the same stage of production processing, so the engineers need know the key root-cause of defect. We used a matrix to save and count the amounts of the root-causes that appeared. When the amount of the root-cause is larger, a high priority will be placed on troubleshooting the stage. The Priority Algorithm for the Root-cause of the production processing is as follows:

$$(Cont(Sx) , Max(Sx)) \{ x=1,2,3,4,5 \}$$

According to the Process Analysis System, the engineers will know the root-cause of defect and the method of debarring warning conditions. Finally, we saved and updated the information in the database, the images of the bumps, the bump features, the fuzzy logic membership functions and the defect types.



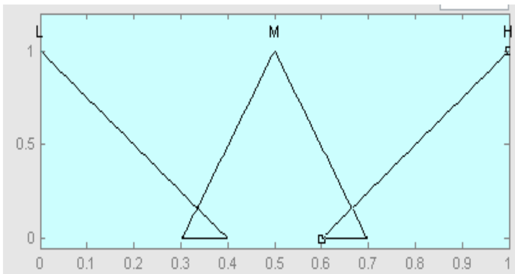


Fig. 9. Fuzzy sets of seriousness of failure of bump contamination

Fig. 10. Fuzzy sets of boundary between normal and defect

SYSTEM FUNCTIONS

Implement environment and tools

We used the Microsoft Windows XP operation system as the system platform and MatlabR14 as the development tools. And finally, we used Microsoft Windows.net 2003 to design the system interface and integer functions developed from MatlabR14.

Processing performance

From the requests of the semiconductor industry, the computer can deal with each die for 0.6 second, with 6000 die each hour at the speed of PVI3.0GHz. If we want to speed up the process, we can use more CPUs to increase the speed.

System characteristics

The system functions included six modules: the wafer inspection module, defect display module, defect analysis module, database administration module and system settings:

- Wafer Inspection Module: This can read the number of dies then inspect the bumps on the die. Moreover, it can classify the defects (see Fig. 11).
- Defect Display Module: We can inspect the defect samples of the 3D graph with different kinds of visual angles and different kinds of features; moreover, we can list the code number of the defect and classify the type of defect samples. (see Fig. 12).
- Defect analysis Module: Used to evaluate the quality of measured samples in a given period by statistics (see Fig. 13)
- System Settings: Setting arguments of the system (see Fig. 14).
- Database Administration Module: This can support the data management function by creating a database of the bump and feature values (see Fig. 15).

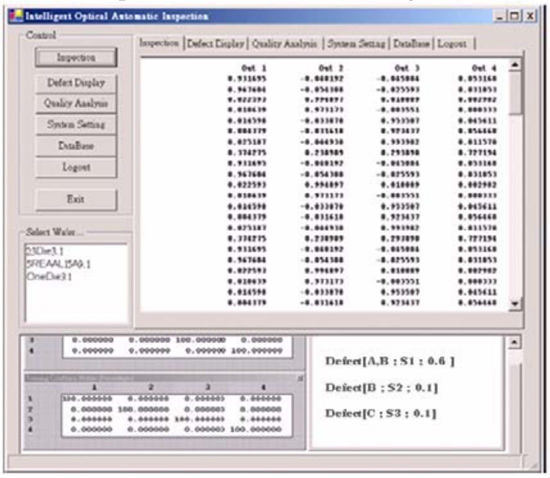


Fig 11 Wafer Inspection Module

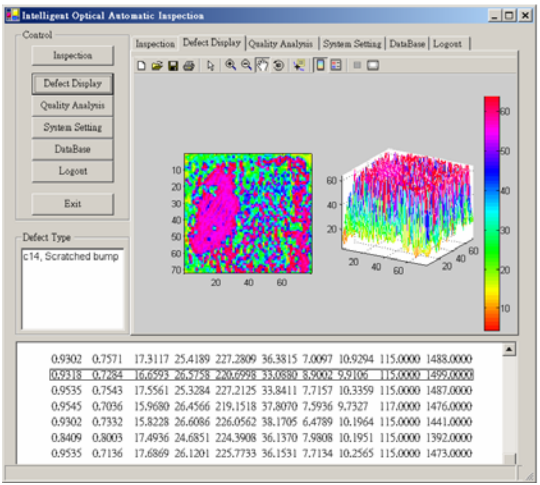


Fig 12 Defect Display Module

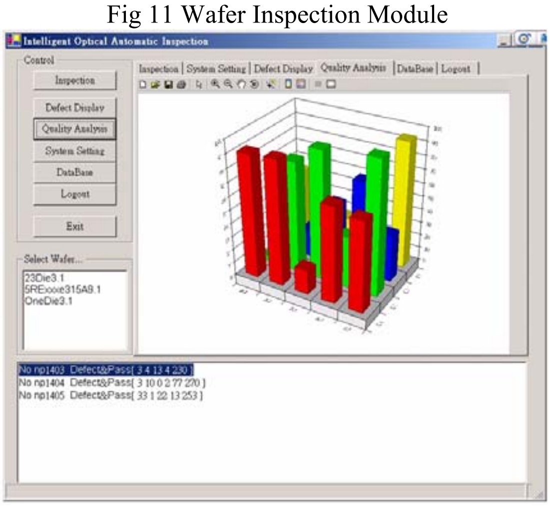


Fig. 13. Defect Analysis Module

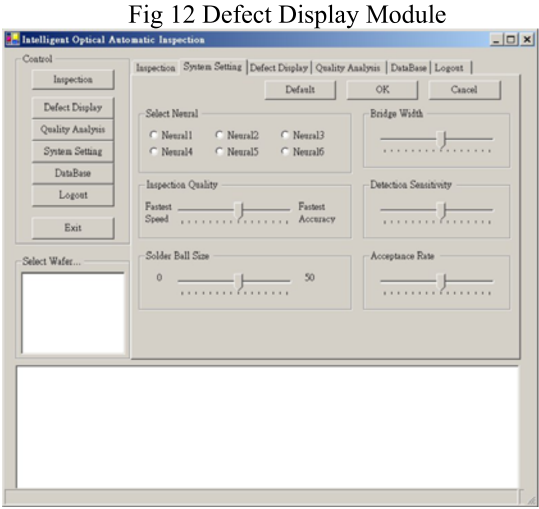


Fig. 14. System Setting

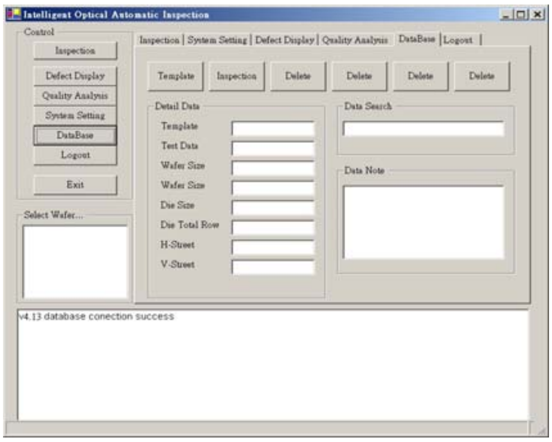


Fig. 15. Database Administration Module

DISCUSSION

The experience results showed the automatic wafer inspection system can meet the practical demands. This is the most important value of this proposed approach:

Reducing manufacturing costs

According to the quality guarantee, the system can set different defect thresholds. When one die exceeds the threshold, it will be eliminated. The higher the defect threshold, the lower the rating. So setting too high a threshold will raise the manufacturing costs. According to practical experience, defect thresholds should set for the quality and warranty of the products. So we designed a system which can adjust the threshold for the specific defect. Thus, under the quality demands, the system can attain the best rate.

Improving manufactory performance

Utilizing fuzzy expert systems to analyse the correlations between the defect type and root-cause of defect is very important for production management. When we know which defect belongs to which type, we can provide more information to the production line, and it can be adjusted to improve the performance. (ex: If the system shows “Chip Probing made the void on bump”, we know that the probe should be replaced. If a defect can’t be classified, its image will be sent to the process engineers for analysis forthwith.)

CONCLUSION

Automatic wafer defect inspection is the key process in the semiconductor industry. The industry standards in inspecting wafers during the packing process are template matching and pattern recognition. Although they can detect defects, normally they cannot tell the types of defects. Thus, the end will provide less efficient information for improving the manufacturing process.

This research was a joint project with the semiconductor industry in Taiwan. To meet the practical demands of manufacturing, we developed an automatic visual inspection system to detect the bumps on wafers using back-propagation neural networks to train on the features of the defects, then the defects could be classified. Next, according to the engineers’ experience, we used a fuzzy expert system to analyse correlations between defect types and the root-causes of the defects. According to the fuzzy-expert system, the engineers were able to quickly find the root-causes of the defects and improve the manufacturing process and production management. This is the most important value of this proposed approach.

In practical cases, there are many processes in the production line which are very complex. In our research, we just analysed the correlations between four defect types and four production process stages. We achieved good performance in our effort to improve production management, but it was unable to satisfy all the contingencies of the production line. Therefore, in future research, we will add more defect types and production process stages to achieve even more accurate results.

REFERENCES

Acumen: http://www.accumen-technology.co.uk/.

Akhter Fahim & Hobbs Dave & Maamar Zakaria (2005). A fuzzy logic-based system for assessing the level of business-to-consumer (B2C) trust in electronic commerce, Expert System with Applications.

Badih, E. K. (1997). SEMI your industry resource, Semiconductor Conference TAIWAN, pp.77-97.

Barschdorff, D & Monostori, L (1991). Neural networks: Their applications and perspectives in intelligent machining, Computers in Industry, vol. 17, pp. 101–119

Brunelli,Roberto& Poggio,Tomaso.(1995).Template matching: matched spatial filters and beyond. Massachusetts institute of technology artifical intelligence laboratory and center for biological and computational learning department of brain and cognitive sciences. A.I. Memo NO. 1549 C.B.C.L Paper No. 123

Chen, H. & Buntin, P& She, L. & Sutjahjo, S& Sommer, C & Neely, D. (1994). Expert prediction, symbolic learning, and neural networks: an experiment on greyhound racing, IEEE Expert 9 (6) 21–27.

Cognex: http://www.cognex.com.

GR technology: http://www.grtechnology.co.uk/.

Huang, Zan & Chen, Hsinchun & Hsu, Chia-Jung & Chen, Wun-Hwa & Wu, Soushan.(2003). Credit rating analysis with support vector machines and neural networks: a market comparative study, Decision Support Systems.P543-558

Huang, C.-J. (2002). Application of Neural Networks and filtered Back Projection to Wafer defect Cluster Identification, IEEE proceedings of the 4th International Symposium on Electornic Material and Packaging, pp.99-105

Jain. Anil K. & Robert P. W. Duin & Jianchang Mao (2000), Statistical Pattern Recognition: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence**,** Vol. 22, pp 4 – 37

Kestra: http://www.kstkestra.com.br/.

Liao. H. Y. & Liu .S. J & Chen .L. H., & Tyan .H. R. (1995). A bar-code recognition system using ackpropagation neural networks, Eng. Applicat.Artificial Intell., vol. 8, pp. 81–90.

Negnevitsky, Michael. (2004).Artificial Intelligence_A Guide to Intelligent Systems. Edinburgh Gate: ADDISON WESLEY.

Schalkoff, Robert J.(1997). Artificial Neural Networks. Singapore: McGRAW-HILL.

Thomas Cinque et al (2002), Automated optical inspection (AOI). A yield management solution for the high density interconnection (HDI) Industry, Nepcon West - Fiberoptic Expo.

Tolle, K.M. & Chen, H. & Chow, H. (2000). Estimating drug/plasma concentration levels by applying neural networks to pharmacokinetic data sets, Decision Support Systems, Special Issue on Decision Support for Health Care in a New Information Age 30 (2) 139– 152.

Walczak, S. (1998). Neural network models for a resource allocation problem, IEEE Transactions on Systems Man and Cybernetics, 28B276±284.

Widrow, B. & Rumelhart, D.E. &Lehr, M.A. (1994). Neural networks: applications in industry, business, and science, Communications of the ACM 37 93– 105.

Zadeh, L. (1965). Fuzzy sets, Information and Control, 8(3), 338-353.

Zoroofi, R.-A. & Tamura, H.-S. &Sato Y & Sekiya, K. (2001).Automated inspection of IC wafer contamination, Pattern Recognition, Vol.34, pp. 1307-1317

A Neural Network Based Architecture for Mobile Agents in a Security Application

Anand Kuppuswami
University of Western Sydney
a.kuppuswami@uws.edu.au

Abstract

There has been tremendous research in the field of multi-agent systems and mobile agents. Mobile devices foster a new set of applications that are receiving enormous support in the global electronic community. This is primarily because of the ability of mobile devices to connect, and reconnect with each other dynamically. The ability to create a network of devices "on the fly" can be adjudged as a major advantage of mobile gadgets as compared with the previous computer networks. The major impedances in implementing such systems on a mobile network are the latency factor, abrupt disconnection in service, lower bandwidth and minimal processing power. The mobile agent's paradigm proves to be an effective solution to various issues raised. It has received serious attention in the last decade and several system based on this paradigm has been proposed and built. All such systems have been designed for a static network, where the service providers and the requestors are connected to the server on a permanent basis. We present a new framework of managing the mobile environment and the participating nodes with active intelligent migration. The functioning of the mobile agents in such a scenario is presented. The agents have intelligence embedded into them for mimicking the human like behavior. Embedded with intelligence, agents can act autonomously, collaborate with other agents and reduce the human intervention. The agents are configured to be adaptable, learn from their experience and mature into experienced agents.

Keywords:
Mobile Agents, Neural Network, Pattern Detection, PHNN, UML Models

1. INTRODUCTION

Recently, there has been a growing interest in the field of multi-agent systems – particular in the image content analysis of large number of applications like image retrieval from database, content-based image coding and analogous applications. These applications, based on Artificial Intelligence (AI) (Jain,2003) and Distributed Artificial Intelligence (DAI) (Fox et al,2003) have been proposed and developed with varying success. Majority of the aforementioned application have a data and service repository which enables the parent application to understand and decode images. This ability to analyze and decode images has tremendous application in customizing software modules, as discussed by (Bicego et al,2003), in detecting security threats (Bicego et al,2003). However, these multi-agent applications are built in a closed architecture and as a consequence, they cannot consume other services offered over the network Tafti (1992), Sondak. Et al (1989). Furthermore, these applications also do not have the ability to adapt to the volatile environment in which they have to operate. Consolidating these applications with mobile devices offer several advantages such as the ability to connect, and reconnect with each other dynamically, able to communicate in a dynamic and moving environment and, more importantly, process information irrespective of time and place. This paper proposes a neural network based mobile architecture that would enable mobile agent systems to work effectively for a security application in a dynamic environment. To achieve this dynamicity, mobile agents have to have intelligence embedded into them. This paper discusses a framework for such an architecture containing intelligent agents and validates the framework for a face recognition application for the security domain

This paper is organised as follows. The literature review is presented in Section 2, the system models are described in Section 3 and the implementation details are described in Section 4. Finally the conclusions are summarised in Section 5.

2. LITERATURE REVIEW

Biometrics based identification has been in use for past few years. This is evident in applications with features like face, voice (Fox,2003), fingerprint and iris recognition. Face recognition forms the major research area in biometrics and the work by Zhao et al (2003) provides detailed descriptions of the underlying technologies for such solutions. The concept of person recognition and its possible applications have been identified in (Jain,2003). SVM (Bicego et al,2003) and neural networks Tafti (1992) form the major player in this arena. However, the limitations of these applications, such as mentioned in the introduction, can be handled effectively by their integration with mobile technologies. Despite the potential advantages of dynamicity with mobile technologies, there are still challenges in their integration with security applications. These challenges result from factors like low latency, physical obstruction and network connectivity. While mobile gadgets have low memory, processing capabilities and are prone to sudden failure (Cabri. et al 2002), the corresponding applications unveil the problem of client server architecture (Kendall. et al 1998). Although the novel approach of transportable programs (Lauvset. 2001), (Gray. et al 2002) offers a promising solution for some of these issues, it is important to come up with a comprehensive neural network framework for mobile applications, as has been attempted in this paper.

Transportable agents or Mobile agents, as they are called now, are autonomous programs that can migrate from one machine to another machine in the network. By migrating to the machine having the resource, the agents have the advantage of working on site where the resource is present and also use the processor's power. This eliminates all the middleware that is required for transporting the data to the client's site. Mobile agent's paradigm provides an effective solution to the problem of low latency, poor interface and bad network conditions (Gray. 2001). The middleware and the communication control mechanism form the major workload in client-server architecture and by eliminating them, the efficiency of working environment would increase. The code and state of code is migrated to another machine for resuming its execution. This also eliminates the interface required for service access. The fact that there is no need for permanent connection makes it very suitable to the mobile environment. The ad hoc client-server model is overridden by the peer-peer model which matures into grid computing, where the machines can act as client or server depending on the current situation. The programmer is swayed away from traditional multi-tier architecture to grid computing (Lauvset. 2001).

Majority of the mobile agents present in the literature is designed for static network architecture. Baring a few mobile agents, (Cabri. et al 2002), (Kendall. et al 1998) intelligence is not embedded into them. This paper presents a new set of mobile agents which work in a volatile mobile environment. Embedded with intelligence, agents can act autonomously, collaborate with other agents and reduce the human intervention. The agents are configured to be adaptable, learn from their experience and mature into experienced agents.

3. MODEL OF PATTERN DETECTION APPLICATION USING MOBILE AGENT

In order to examine the efficiency of the framework, we use a virtual case study in the security domain. Let us assume that the objective of the application is to automatically detect security threat and alarm the respective authorities. As it would be practically impossible for security personals to be present at all sites all the times, individuals with handheld gadgets and an ability to transmit images to the gateway server can be of immense help in security architecture. Such an individual can facilitate the system by providing images to the pattern detection module. Those images could provide useful information for recognizing security threats. For example he can capture a suspect's image using his mobile phone and then MMS it to the Local Gateway Server (LGS). The system should be able to identify the suspect and this process can be summarized as a face recognition application. This involves multiple activities of varying complexity and hence single thread of workflow would not suffice. We need a network of computers (each specializing in a workflow) coordinating with each other by offering and consuming services. Furthermore, if additional computers can offer their service on a non-permanent basis, it could drastically improve the system performance. This concept is embedded into the mobile environment in the new framework proposed. Continuing with the task of person identification, the gateway server coordinates with the mobile environment for successful recognition of threat. The Local Gateway Server then formulates the service required and conveys that to the mobile environment. The mobile environment acknowledges the request and begins processing it. After diligent processing with all the available nodes and their services, it returns the answer to the LGS. Subsequently, the gateway server formats the data and sends it to the mobile client or performs supplementary actions. The system architecture is as shown in Figure 1. The double headed arrow signifies the two-way communication between the blocks.

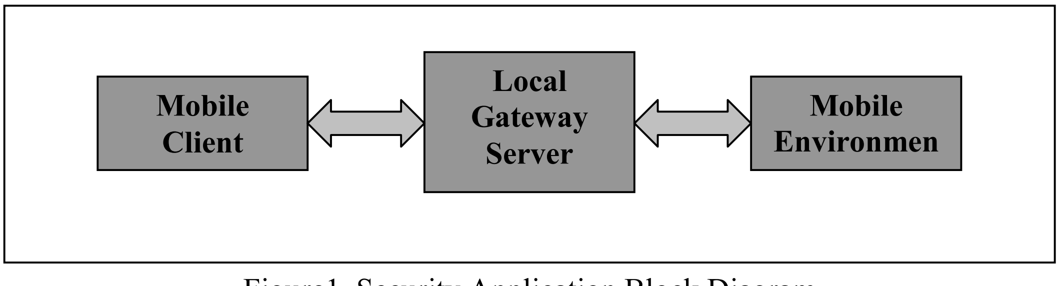


Figure1. Security Application Block Diagram

This paper concentrates on the working on the mobile environment and presents the detailed UML models only for the mobile environment modules. The description of the Local Gateway Server and the mobile client can be found in (Kuppuswami. 2006).

3.1. Mobile Environment System Models

Mobile environment forms the crucial part of the system architecture. The functioning of this environment depends on the coordination between nodes, server and mobile agents. The system has to maintain consistency in operation and needs to have a comprehensive monitoring module. In addition to these essential modules, it also requires the individual units to be working in a robust manner. The robust working of each units ensure that all the accompanying workflow does not end in a deadlock. The system development requires an open structure to cater future developments. The future developments could either be a new module developed or a new working environment. The system has been segregated into three dominant areas of development. The Node Management Module (NMM) primarily superintends the individual node's operations. The Server Management Module (SMM) administers the collection of nodes and their services. It also acts like a virtual yellow page and service directory. Finally, the Agent Management Module (AMM) governs the working of mobile agents and its migration. The Node Service Analysis (NSA), broker agent (BA) and intelligent components form an integral part of this module.

3.1.1. Node management

Node management is primarily involved in activities like adding a node, service generation and node registration. Besides, activities like XML encoding, decoding, connectivity with the server also forms the supplementary part of this module. This module is predominantly required when a node wants to enter the mobile environment and adduce its services to the service requesters. Furthermore, the general working environment being a volatile one, it needs to actively communicate with the server and other nodes for consistent working. In addition, as they need to interact with intelligent agents, they need to have a compatible open source architecture. This paper provides meticulous models for some of the core functionality.

Whenever the node wishes to offer its service, it sends a requisition to the server. This would comprise details of node, and the services offered. Upon receiving the request, sever processes it and performs subsequent activities. This process is explained in the activity diagram shown in Figure 2. The activity starts in the `Node` domain with the node migrating to the active environment. It then raises a request for a new connection. This requires a service header which is created in the activity `Create Service Header`. This information is then corresponded to the `Server`. The server receives the request and redirects the header section to the `XMLDecoder.`
To maintain the compatibility with the SAOP, the information exchange is in the XML format and hence needs an XML encoder and decoder. The decoder extracts the service description portion of the message and communicates back to the server. Sever validates the node's credibility and investigates for the presence of similar nodes. The presence of congruent node impels the server to place the incoming node's request to the Queue Management (QM), which takes further control of request. These checks ensure the consistency of the mobile environment.

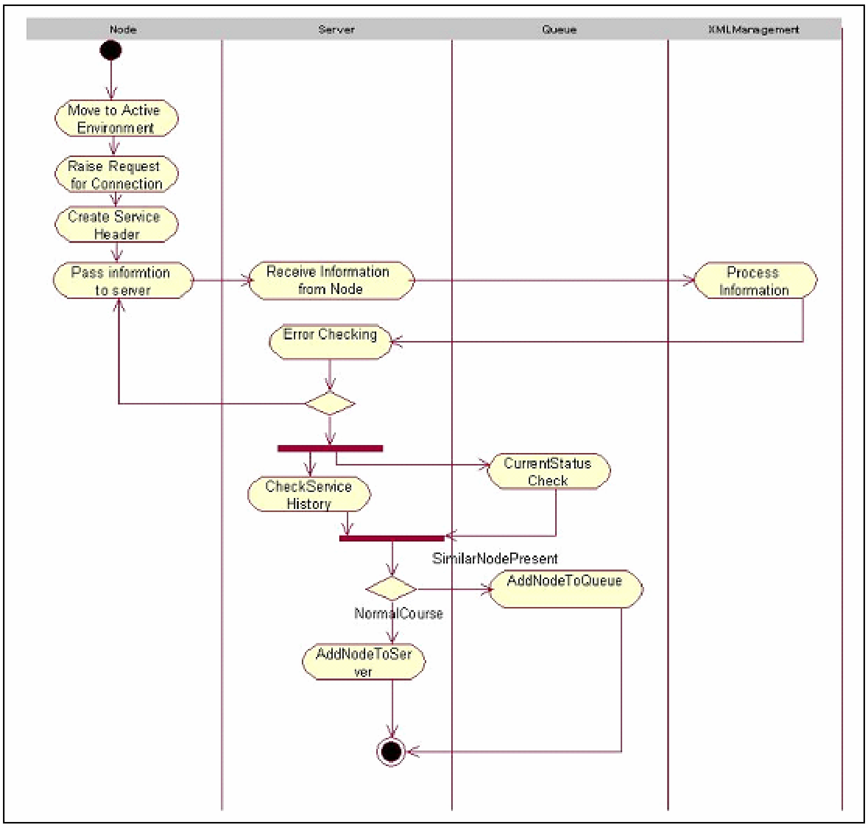
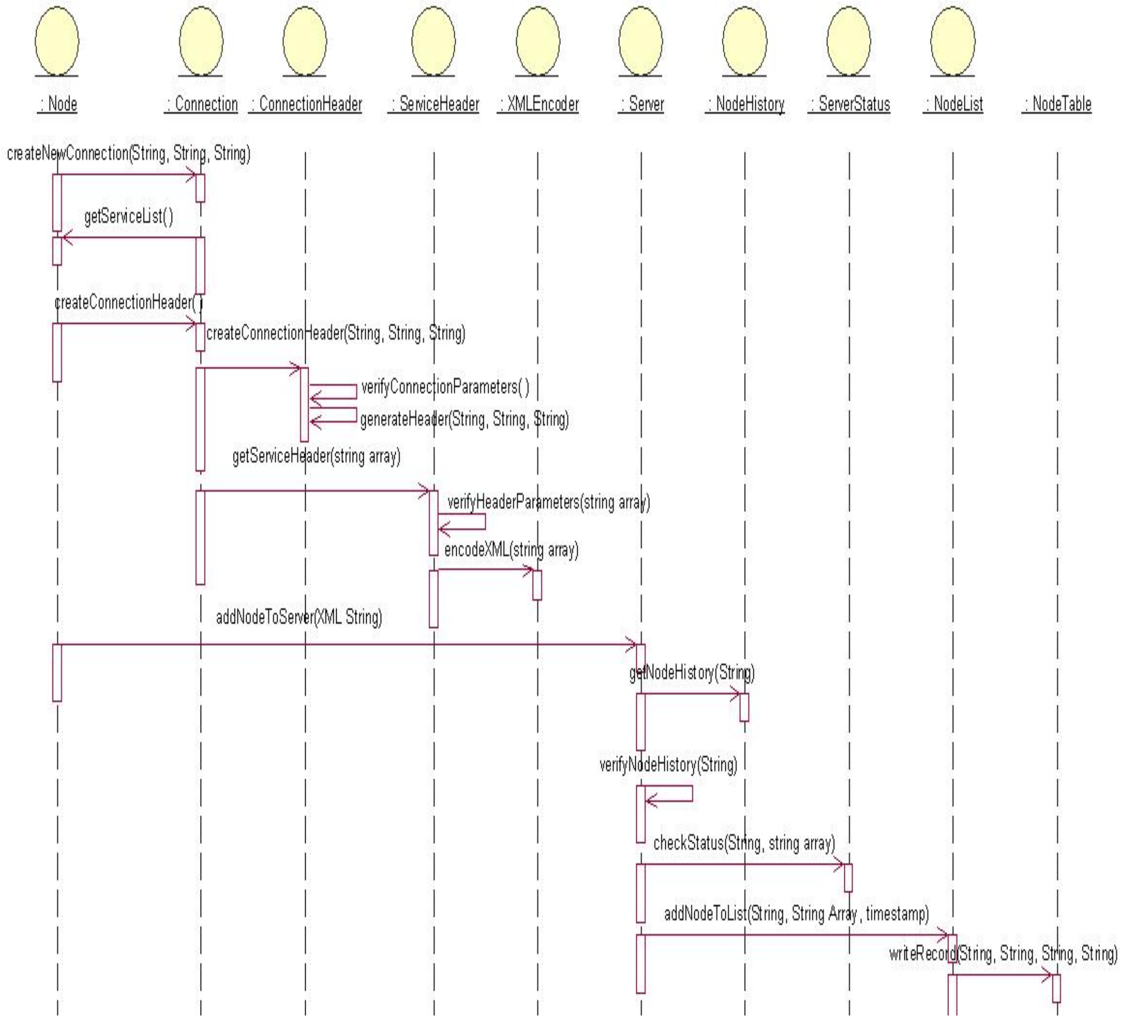


Figure 2. Activity Diagram for Adding a Node

The activity diagram provides the succinct of the tasks involved in adding a node to the mobile environment. The code level implementation demands the detailed class modeling for its unabridged working. Some of the main classes involved are `Node`, `Connection`, `ConnectionHeader`, `ServiceHeader`, `XMLEncoder` and `Server`. The implementation details of node management are explained using the sequence diagram shown in Figure 3. The :`Node` object coveys the `createNewConnection` message to the :`Connection` object. The listof services offered are passed as parameters. the. The :`Connection` object then communicates with the :`ConnectionHeader` to generate the service header with the assistance of :`ServiceHeader`. The :`XMLEncoder` acquiesces the responsibility of XML formatting. Upon the successful generation of service header, :`Node` sends `addNodeToServer` message to :`Server.` This message ensures that the process of adding the node to the environment is initiated. Server performs the customary checks before registering the node. Server communicates with :`NodeHistory` and :`ServerStatus` and evaluates the node's credibility. Upon successful validation, the node is appended to the :`NodeList`.



3.2. Agent Management Module

The mobile environment is required to perform a series of operations in order to recognize the threat. Generation of mobile agents is the principle action that needs to be performed in the Agent Management Module. The mobile agents are units which can roam in the mobile environment and perform the required work. The mobile environment being volatile, the service providers will be present on a non-permanent basis. The presence of certain nodes will be enhancing the system performance. This induces the need for components, which can automatically recognize the best nodes and redirect the mobile agents to these locations. This task is accomplished by a component called Broker Agent (BA). The BA has a Pseudo-Hierarchical Neural Network (Kuppuswami. et al 2002a) embedded into it. A PHNN is a pattern detection tool developed using neural networks. The BA's principal responsibility would be to identify the node which reconciles the mobile agent's requirement. The process of analyzing the nodes for best service is illustrated in Figure 4.

The mobile agent resides in various states during the execution cycle. This process being a complex one, the state chart diagram as shown in Figure 5 is used to describe it. It starts with pseudo `Start` state. It enters the `Instantiating` state where all the parameters are initialized. This ensures that the agent is ready to perform the required operation when required. It eventually enters the `Dormant` state and remains in this state until a request for service is received. Upon a service request, it enters the `Communicating` state wherein it talks comprehensively to the service requestor. During this state, there is active communication between the agent and the service requestor. The service required forms the core part of information exchanged. On completion of all communication customs, it enters the `Active` state. The mobile agent then starts executing in its present location. Whenever there is a need for migration, it enters the `Migrating` state. When the complier encounters the `MoveTo` command, the migration starts. . The new location is passed as a parameter to the message. After reaching the destination, it reinitiates itself and prepares for executing where it acclimatizes to the new environment. The checks are performed to ensure that the agent can survive and perform in the new environment. The agent then restarts the execution. When the mobile agent has performed all the required operationsthe result of the actions performed are passed back, the destructor is called and the resident memory is cleared.

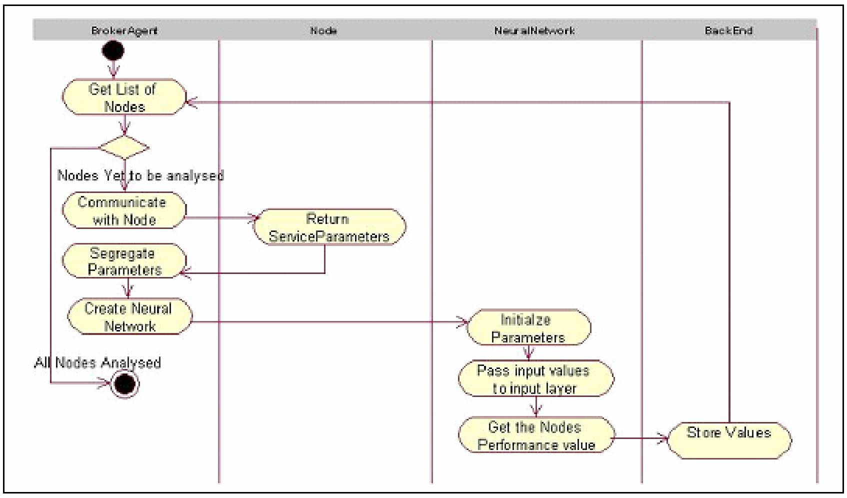


Figure 4. Activity diagram for selecting the node for service

3.3. Broker Agent Architecture and Training

The broker agent has been built on the structure of Pseudo-Hierarchical Neural Network (PHNN). PHNN (Kuppuswami. et al 2002b) is a new pattern detection tool fabricated using multiple feed-forward neural networks. The boot-strapping technique has been enforced to train the PHNN. Since the boot-strapping works on the supervised training principles, it requires an expert's advice and training set to assemble an efficacious network. This paper utilizes a three level PHNN for identifying the node for service.

The individual levels in PHNN have 16 input nodes and have 1 output node. The input for each node varies from a value 0 to 15 and the output values ranges from -1 to +1. An output of 1 signifies that the node is best suited for the particular service and -1 signifies that the node is worst suited for the service. The usefulness of PHNN depends ultimately on the training method. The number of levels in the PHNN depends on the training set, the network capacity and the classification efficiency required. The set of training data should cover a wide range of input. This ensures that the PHNN has been reasonably introduced to the variable input environment. Only after the training is complete, PHNN can handle the unknown input. Hence a large set of data is selected for training. .The Table 1 shows sample training data for PHNN.

Upon conclusion of its training phase, the PHNN reaches a stage where it is fully qualified to identify the best node for service. Thereafter the PHNN is released to the network of nodes. The PHNN commences the process of evaluating each node's service. The PHNN communicates with each node and retrieves its Service Description Header (SDH) and extracts the keywords and their respective Proximity Indexes. Depending on the keyword patterns, each node is evaluated for its service against the agent's requirements. The node providing the best correlation is opted as the service provider. In this process, the new keywords are first mapped before the agent can process them. In the mapping if there is a change in the knowledge base, the network needs to be retrained using the new training set. This is the process where the agent learns and matures. Each set of training builds the Knowledge Base (KB) and after a series of training, the KB would have saturated and the agent would have matured enough to handle the unknown keywords.

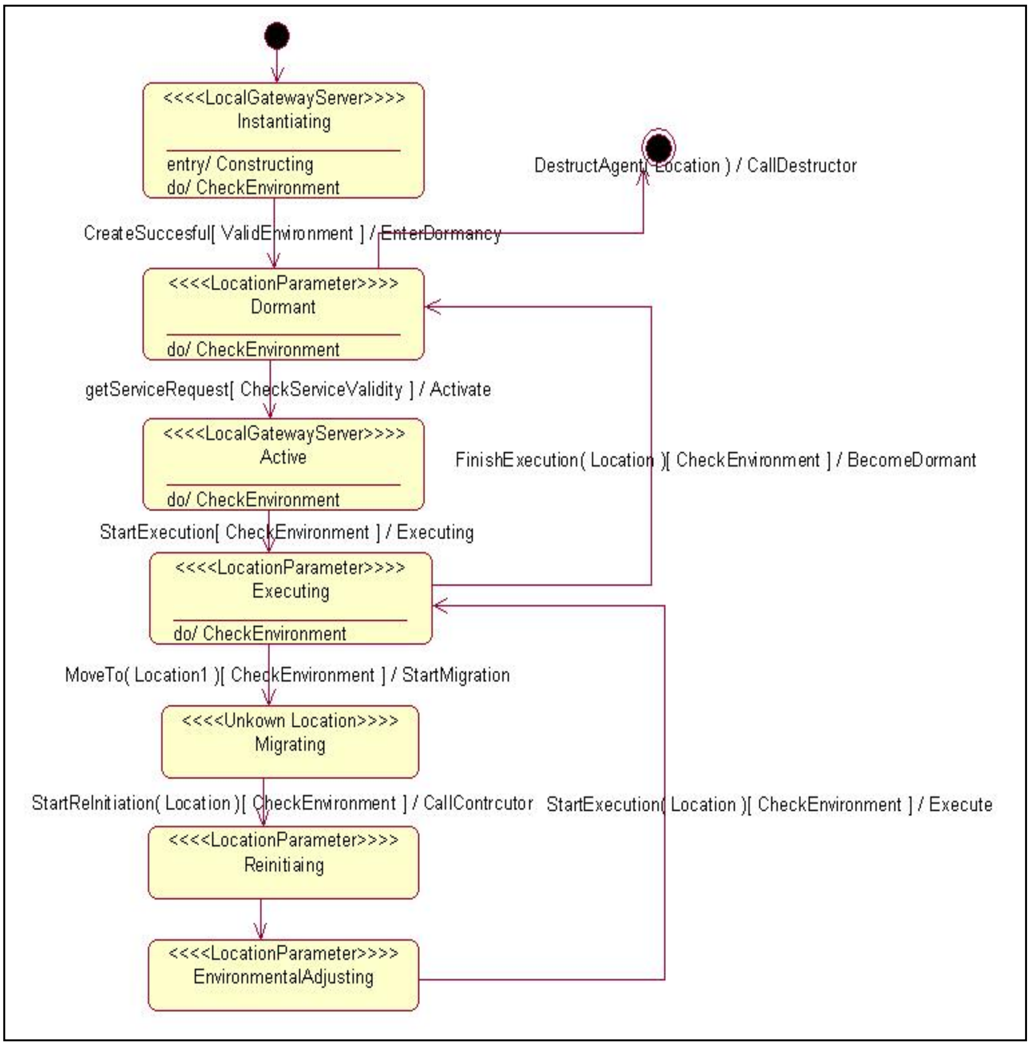


Figure 5. Agent state chart diagram

Service Provider Keyword PI	Input to Neural Network
Image 15, Processing 8	F800 0000 0000 0000
Histogram 15, Processing 8	08F0 0000 0000 0000
Histogram 15, Equalization 10	00FA 0000 0000 0000
Image 15, Processing 15, Equalization 8	FF08 0000 0000 0000
Image 15, Processing 15, Histogram 15, Equalization 15	FFFF 0000 0000 0000

4. IMPLEMENTATION

This system is being implemented in C# and .NET environment. MS SQL Server 2000 is serves as the data repository. The PHNN is constructed using the feed-forward neural network and is implemented in MATLAB 6.0 and. On completion of training, the network attributes are exported to binary files. The mobile environment code written in C#, access these files for PHNN's simulation. The system utilizes a three level PHNN for the purpose of node analysis. Each level has 16 input nodes and 1 output node. The hidden layers of Level 1, Level 2 and Level3 have 30, 25 and 20 neurons respectively. . The output node has a value ranging from -1 to +1. During the initial stage of training the PHNN is presented with 250 set of inputs. The unavailability of real training data for this application obligates a training set to be formulated with virtual expert's advice. The PHNN is being trained for maximum of 1000 epochs or gradient of less than 0.001. After the initiatory training, the agent is released into the network for testing. The performance of agent directs the next probable phase. This phase would involve testing, reconfiguration of PHNN and building Knowledge Base (KB). As the implementation of the mobile agents is not the core of the research being conducted, this paper proposes only a virtual implementation of it. The parent executable file is located at all the nodes and we pass the working memory content to the destination. It then continues the operation in the new node.

5. CONCLUSION

This paper proposes a neural network based mobile architecture that would enable mobile agent systems to work effectively for a security application in a dynamic environment. To achieve this dynamicity, mobile agents have to have intelligence embedded into them. This paper discusses a framework for such an architecture containing intelligent agents and validates the framework for a face recognition application for the security domain. Parameters like the framework's efficiency, latency factor and the network configuration are yet to be evaluated as the experiments are still in progress. This integration of two concepts opens a new area of research. The types of intelligent components that can be embedded into the framework is also being exhaustively researched.

6. REFERENCES

W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld (2003) ,"*Face recognition: A literature survey*", ACM Computing Surveys (CSUR), December 2003 Volume 35 Issue 4

Anil.K.Jain (2003), "*Multimodal user interfaces: Who's the user*?", November 2003 Proceedings of the 5th international conference on Multimodal interfaces, ACM Publications, ISBN 1-58113-621-8

Niall A. Fox, Ralph Gross, Philip de Chazal, Jeffery F. Cohn, Richard B. Reill,(2003) " *Person identification using automatic integration of speech, lip, and face experts*", Proceedings of the SIGMM, ACM workshop on Biometrics methods and applications, November 2003.

Manuele Bicego, Gianluca Iacono, Vittorio Murino (2003), "*Face recognition with Multilevel B-Splines and Support Vector Machines*", Proceedings of the SIGMM ACM workshop on Biometrics methods and applications, November 2003.

Mohammed H. A. Tafti (1992) ,"*Neural networks: a new dimension in expert systems applications*", ACM SIGMIS Database, Volume 23 Issue 1, March 1992.

N. E. Sondak, V. K. Sondak. (1989). "*Neural networks and artificial intelligence*", ACM SIGCSE Bulletin . Proceedings of the twentieth SIGCSE technical symposium on Computer science education, Volume 21 Issue 1, February 1989,

Gray, R. S, "*Agent Tcl: A transportable agent system*", Proceedings of the CIKM Workshop on Intelligent Information Agents, Fourth International Conference on Information and Knowledge Management (CIKM 95), Baltimore, Maryland, December 1995.

Lauvset, K. J. (2001). *Separating Mobility from Mobile Agents*. In HOTOS ' Proceedings of the Eighth Workshop on Hot Topics in Operating Systems (pp. 173): IEEE Computer Society, 2001

Cabri, G., Leonardi, L., & Zambonelli, F. (2002). "*Engineering Mobile Agent Applications via Context-Dependent Coordination*." IEEE Trans. Software Engineering, 28(11), Pp:1039--1055.

Kendall, E. A., Krishna, P. V. M., Pathak, C. V., & Suresh, C. B. (1998). "*Patterns of intelligent and mobile agents*." In AGENTS '98: Proceedings of the second international conference on Autonomous agents (pp. 92--99): ACM Press, 1998

Jipping, M. J. (2002). "*Symbian OS Communications Programming*." John Wiley & Sons, Ltd

Gray, R. S., Cybenko, G., Kotz, D., Peterson, R. A., & Rus, D., "*D'Agents: applications and performance of a mobile-agent system*" Experts of Software Practices, 2002, Vol 32(6), Pp 543—573

Kuppuswami, A. (2006). "*Neural network based mobile agents for mobile environment*". In B.Unhelkar (Ed.), Handbook of Research in Mobile Business, IDEA Publishing Group

Kuppuswami, A., Cheung, H., Ginige, A. (2002). "*A Pseudo-Hierarchical Neural Network for Face Detection*", International Symposium of Intelligent Processing and Communication Systems, Taiwan Nov 2002

Kuppuswami, A., Murugasen, S., Ginige, A. (2002). *"A Pseudo-Hierarchical Neural Network for Pattern Recognition*", Digital Image Computing Techniques and Application, Melbourne, Jan 2002

An Interactive Model for Fuzzy Measure Determination

Mr. Amol Wagholikar, Dr. Peter Deer,
Griffith University
Gold Coast, Australia
Email: a.wagholikar@griffith.edu.au
Email: p.deer@griffith.edu.au

Abstract

Determination of fuzzy measures is a complex task. It is difficult to determine generic values of fuzzy measures in decision problems. Various approaches have been suggested for the acquisition of fuzzy measures in the context of decision-making. However, these approaches do not handle the subjectivity associated with fuzzy measures. We propose an interactive approach for determining fuzzy measures using case-based reasoning. We propose a case-based decision support system for determining fuzzy measures and hence for evaluating the given alternatives in the new decision problem. The proposed approach is illustrated with sample a data set.

Keywords:

Multicriteria decision making, Fuzzy measures, Decision support systems, Case-based reasoning, Similarity

INTRODUCTION

Multicriteria decision making involves evaluation of set of alternatives against a set of criteria (Zeleny 1982). The criteria in a given decision problem can have an interaction or relationship between them. They can be related positively or negatively. Hence their combined importance in the overall evaluation can be different than the individual contribution. The relationship between the criteria must be considered while determining global evaluation of the alternatives. Additive measures such as weighted average are inadequate for such application. Non-additive measures such as fuzzy measures and fuzzy integrals are proven as effective tools for modeling the interaction between criteria (Grabisch, 1996). Fuzzy measures model the importance of sub-set of criteria in decision making. However, their pragmatic determination is highly complex due to 2^n-1 real coefficients, where n is the number of criteria in a given decision problem. This can be very significant especially for higher values of n. Say if n=8, then 255 fuzzy measure coefficients needs to be determined. The main problem here is not the determination of these coefficients, but the way in which they are determined. Especially, in the context of decision-making, this issue is interesting due to subjective component in decision making (Buchanan, 1997). In a given decision problem, alternatives can be evaluated differently by different decision makers. Hence the value of fuzzy measures can vary with the decision maker. Due to this, it is difficult to arrive at a definite value of fuzzy measures. We investigated different approaches suggested for practical determination of fuzzy measures. We intuitively believe that these approaches do not handle the subjectivity involved in the issue of fuzzy measures, and that interactive models would be a pragmatic approach for application of fuzzy measures in modern decision support systems.

We propose an interactive model using a direct rating method. Reasoning by analogy is the main theoretical foundation of our model. This paper discusses our model and its experimental implementation in a user-driven case-based decision support system. This paper contributes to the knowledge of fuzzy measure determination methods. The next section explains the issue of fuzzy measure determination. After this section we discuss the approaches applied to solve this problem. The following section explains the suggested approach and discusses the experimental implementation of this approach.

DETERMINATION OF FUZZY MEASURES

Fuzzy measures are monotone set functions defined over a universal set C and a non-empty family ρ of sub-sets of C. It is defined as μ: ρ→ [0,1], where $\mu(\phi) = 0 \; \& \; \mu(X) = 1$ (Grabisch 1995).

Fuzzy measures possess the property of super additivity and sub additivity. Because of this distinct property, these measures are suitable to model the phenomenon of interaction in multicriteria decision making. For more detailed description of these properties, readers are referred to Grabisch (1995). Fuzzy integrals are the incremental summation of the product of the criteria evaluation and its fuzzy measures. There are two types of commonly used fuzzy integrals viz. Choquet integral (Choquet 1953) and Sugeno integral (Sugeno 1974). We are using Choquet integral for computing the global evaluation of alternatives in the undertaken research. The Choquet integral of the function f : C → [0,1] with respect to μ is given by –

$$Ch_{\mu} = \sum_{i=1}^{n} \{f(c(i)) - f(c(i-1))\}\, \mu(A(i)) \quad \ldots\ldots(1)$$

Where $._{(i)}$ indicates that the indices, f(c(i)) have been permuted such that –

$0 \leq f(c_{(1)}) \leq f(c_{(2)}) \leq \ldots\ldots \leq f(c_{(n)}) \leq 1$ and A(i) = {c (i),………… , c (n)}.

From the definition of fuzzy measures, it can be said that fuzzy measures are the coefficients of the power set, ρ which is the set of all sub-sets of C. Hence for higher number of criteria, fuzzy measure determination becomes complex as number of fuzzy measure coefficients increase exponentially with the criteria. For application of fuzzy measures, 2^n -1 real coefficients must be determined. For such higher dimension of the data, especially for higher values of n, the way of fuzzy measure determination remains a challenging issue. There are different approaches for determining these monotonic set functions (Klir 1995). However, it is difficult to identify the best among them. Hence determination of fuzzy measures is still an interesting research issue. In the next section, we briefly discuss the approaches suggested for the acquisition of fuzzy measures.

APPROACHES FOR FUZZY MEASURE DETERMINATION

The researchers have suggested various approaches. Sugeno (1974) has suggested a λ-Measure, which can be used as fuzzy measure and subsequent fuzzy integral computation. However, this measure is inconsistent and can model only one type of interaction. Grabisch (1995) has suggested determination of fuzzy measures by interactive optimization. This approach is based on the assumption of equal importance of criteria when the decision maker is unable to provide the preferences over the criteria. This implies an equilibrium situation where there is equal contribution of each criterion in the global evaluation of an alternative. Marichal (1999) has suggested an approach based on the semantic information about the criteria using linear optimization. This approach involves direct involvement of the decision maker and modeling his/her preferences in a linear constraint satisfaction problem.

Also certain data-driven (Klir 1997) and interactive methods (Klir 1995) have been suggested. Some approaches suggest the use of pre-defined fuzzy measures to determine the new fuzzy measures or use of global evaluations from the learning data (Klir 1996). After analyzing these approaches, it is concluded that all of the above-suggested approaches follow the notion of using the same set of past precedents (Wagholikar 2005). We are extending this notion and suggesting an interactive model for fuzzy measure determination. However, acquisition of fuzzy measures can be highly subjective in the context of the decision-making. This subjectivity can be handled by human involvement. Hence we propose an interactive model for eliciting fuzzy measures from the decision maker.

SUGGESTED APPROACH

We are extending the notion of using learning data about the past decisions. Reasoning by analogy is the main theoretical foundation of our approach (Vosniadou 1989). New problems can be solved using solutions of the past similar problems. We are proposing a new method where the learning data is in the form of a case-base. Case-based reasoning allows solving new problems based on the solutions from the past similar problems (Zhaohao 2004).

Why CBR-based decision support?

The main objective of using fuzzy measures is to model the decision making process that resembles the human reasoning process. It is proven that humans routinely use similarity-based reasoning in their decision processes (Kolodner 1991). Despite of its routine usage, humans are not able to utilize prior experiences effectively. Humans are not able to consistently recall the appropriate set of past cases, distinguish between important and unimportant features, recall experiences under time pressures and deal with incomplete and uncertain information in current problems (Angehrn 1995). Hence in order to enhance the human ability of using the prior cases, CBR-based decision support would be useful. The undertaken issue of fuzzy measure determination focuses on modeling the human decision making processes. CBR can be a good tool for this purpose. There have been few models suggested for CBR. We are using model suggested by Aamodt and Plaza (1994) for developing our approach. In the next section, we propose our approach.

Proposed CBR-based Interactive model

We are proposing a 2-phase model. In the first phase, a case library of past decisions is refereed to and fuzzy measure values for the new decision are determined using the most similar case from the case library and its corresponding fuzzy measure values. In the second phase, fuzzy measure values are elicited from decision maker and agreement of these values with the system values from the first phase is determined. These phases are as follows-

Phase I

1. Let $C = \{C_1, C_2, ...C_n\}$ be the set of criteria against which a set A of alternatives $\{A_1, A_2, ..., A_n\}$ is to be evaluated. The individual weights, w_i on the criteria is elicited from the decision maker such that $\sum_{i=1}^{n} w_i = 1$.

For the above decision problem, we have case-base in the following form.

i. A set of attributes same as the new decision problem or target problem mentioned above.

ii. Absolute values of each attribute.

iii. A set of individual weights defined by the decision maker at the previous decision making instance.

iv. A set S of fuzzy measure coefficients defined previously. We assume that these coefficients are constructed using expert's feedback.

v. Global evaluation of the object obtained by using the Choquet integral and the fuzzy measure coefficients obtained from the experts.

2. After constructing the above data set, the similarity of the cases w.r.t to the new decision problem is determined using absolute value-based City-block distance metric or Manhattan distance metric (Kardi 2005). The Manhattan distance metric is defined as the absolute difference between the values of pair of attributes or criteria I.e.

$$d_{ij} = \sum_{k=1}^{n} | x_{ik} - x_{jk} | \quad \text{...} \; (2)$$

Where x_i & x_j are the values of the attribute "k". For computing the similarity, the distance values should be normalized in the range [0,1] by the maximum distance d_{max} between the corresponding attribute values in the new problem and the case base. The global similarity of a given case will be calculated by simple weighted average of the local similarities of the given attributes. Thus we have,

$$d(Case_t, Case_u) = \frac{\sum_{k=1}^{n} w_k d_{ij}}{\sum_{k=1}^{n} w_k}$$

Based on this distance, the similarity is calculated as

$$Sim(Case_t, Case_u) = 1 - d(Case_t, Case_u) \quad \text{..} (3)$$

Where $Case_t$ = Target case i.e. the new decision problem & $Case_u$= Case from the case library i.e. past decision problems.

3. After such similarity computation the case with maximum similarity would be retrieved and the values of its fuzzy measures would be considered for computing the global evaluation in the new decision problem.

Phase II

The decision maker is likely to have different importance values. Hence using a direct rating method (Bilgic 1997), we propose an Importance scale designed to elicit the new fuzzy measure coefficients from the decision maker. The importance scale is designed to ask the decision maker about the importance of sub-set of criteria. The decision maker can directly rate the sub-set of criteria in a range between 0 and 1. The designed scale is shown below.

Table 1.1: Importance Scale

Sub-set is Less important	Extremely	0
	Highly	0.1
	Very	0.2
	Strongly	0.3
	Moderately	0.4
	Equally	0.5
Sub-set is More important	Moderately	0.6
	Strongly	0.7
	Very	0.8
	Highly	0.9
	Extremely	1

Based on this scale, the new fuzzy measures can be elicited from the decision maker. Fuzzy measure values computed in the second phase indicate the agreement between the system values and the decision maker's values. The above approach can be illustrated using the following sample data set.

ILLUSTRATION

Let us consider a classic car selection problem where the following data is assumed

Table 1.2. Sample Data set for car selection problem

	Price		Power		Fuel Economy		Space		Comfort	
Weights →	0.3		0.2		0.2		0.1		0.2	
Car A	\$30,000	0.6	2000	0.8	10	0.6	Good	0.8	Good	0.8
Car B	\$25,000	0.7	1000	0.6	12	0.7	OK	0.7	OK	0.7
Car C	\$20,000	0.8	1500	0.7	12	0.8	OK	0.7	OK	0.7
Car D	\$35,000	0.5	2500	0.5	8	0.5	Very good	0.8	Very good	0.8

Suppose for the above data set, we have the following case-base.

Table 1.3. Case Base of Past Decisions

	Price		Power		Fuel Economy		Space		Comfort		Global Evaluation
Case1	\$40,000	0.8	3000	0.9	8	0.6	Very good	0.7	Very good	0.7	0.82
Case2	\$35,000	0.7	2000	0.6	10	0.5	Good	0.5	Good	0.6	0.66
Case3	\$45,000	0.7	3000	0.6	7	0.7	Very good	0.7	Very good	0.7	0.665
Case4	\$20,000	0.6	1500	0.5	12	0.6	Good	0.7	Good	0.6	0.575

For the above data set, we conducted experiments for determining similarity of the alternatives in table 2 with respect to the case-base in table 3 using SPSS 3.0. We conducted experiments with different similarity measures. Following table shows the results using City-block distance measures chosen for this approach.

Table 1.4: Proximity Matrix - SPSS output for Case similarity using City-block distance measure

	Absolute City Block Distance			
	Case 1	Case 2	Case 3	Case 4
Car A	.008	.009	.000	.008
Car B	.035	.017	.027	.035
Car C	.005	.018	.008	.000
Car D	.060	.068	.058	.050

This is a dissimilarity matrix

From the above dissimilarity matrix, we can observe that –

Car A is similar to Case 3

Car B is similar to Case 2

Car C is similar to Case 4

Car D is similar to Case 4

Based on the above similarities and corresponding fuzzy measures of the similar cases, the global evaluations of all the cars is given by the proposed case-based decision support system. We propose to use RDBMS technology for the system development. The sample output from the system is shown below.

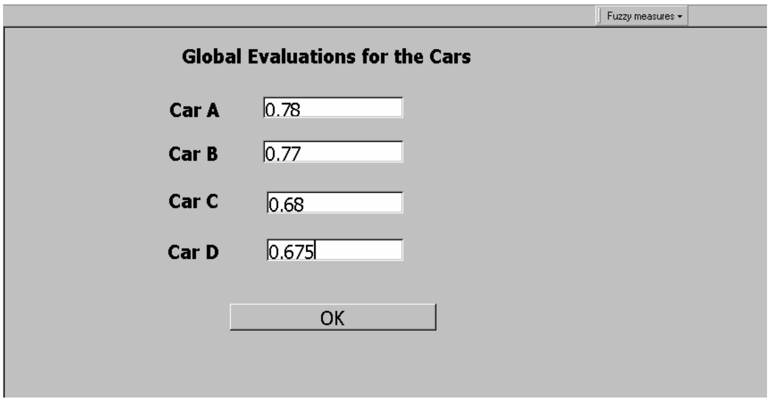


Fig. 3.1. Sample Output from the proposed case-based decision support system

The agreement of the decision maker with the above set of results is determined by the direct rating method using the proposed importance scale. The sample output for the direct rating method is shown in the following interface.

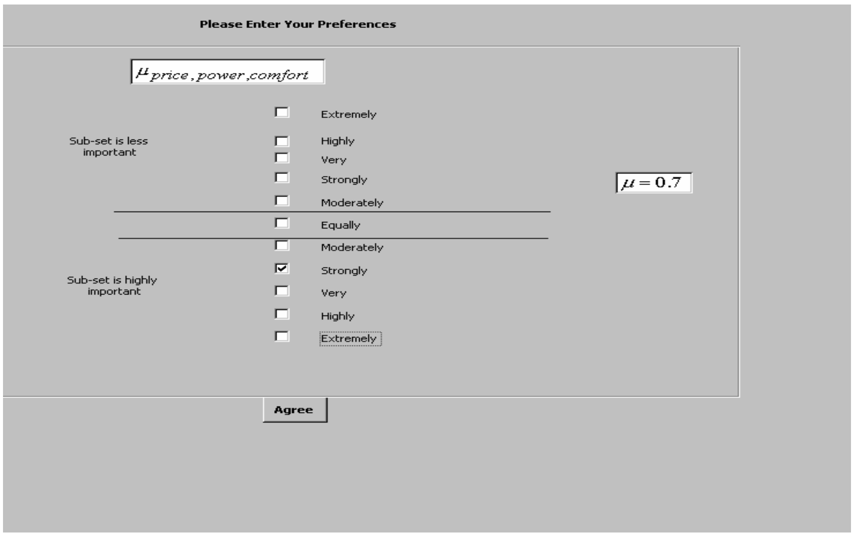


Fig. 1.2. Interactive interface for fuzzy measure determination

From the above sample result, the fuzzy measures can be elicited from the decision maker and hence alternatives can be evaluated. This result certainly indicates the usefulness of the proposed approach.

Evaluation of the proposed system

There can be various ways of evaluating a decision support system. These are as below.

i. If the system is built using data, information and knowledge from one set of situations, then it can be evaluated using an independent set of data (Priya 2001)(Rice 1984).

ii. When a data-driven model is a significant part of the decision support system, sometimes the data can be randomly separated into two parts, one for model development and one for validation (Haberlandt *et al*. 2002).

iii. When the decision support system is not data-based but rather knowledge-based, then it can be empirically evaluated using outputs from the system against a historic data set. This assumes that the logic underlying the system is constant over time.

We believe that the proposed system significantly follows a data-driven model. Hence it can be evaluated using the 2nd strategy in the above options where data in the first phase is used for model development and the data in the second phase is used for validation of the system. Since this is a case-based decision support system, its performance will depend upon the number of cases in the case library. If the number of cases is large, then the results expected from the system should be greater. However, the system performance may degrade due to higher retrieval time. Hence there may need to be a trade off between the number of cases and the system performance.

CONCLUSION

In this paper, we presented an interactive model for determining fuzzy measures in the context of multi-criteria decision-making. We have proposed a model using similarity-based reasoning. We have illustrated our model with a sample data set and through an experimental implementation. Evaluation of the proposed system with various data sets would be undertaken in future.

REFERENCES

Aamodt A, Plaza E., (1994) Case-based reasoning: Foundational issues, methodological Variations, and system approaches, Artificial Intelligence Communications, IOS Press, 7(1), pp 39-59.

Angehrn, A, A., Dutta, S.,(1995) “Case-Based Decision Support”, Communications of the ACM.

Bilgic, T.,Turksen,I.,(1997), Measurement of Membership Functions: Theoretical and Empirical Work, Handbook of Fuzzy Sets and Systems, Vol.1, foundations.

Buchanan, J.,(January,1997) Henig, M., Objectivity and Subjectivity in the decision making process, Department of Management Research seminar series.

Choquet, G.,(1953) Theory of capacities, Ann.Inst.Foruier 5, 131-195.

Grabisch, M.,(1996). The application of fuzzy integral in multi criteria decision-making, European Journal of Operational Research, 89(3), 445-456.

Grabisch, M.,(1995), Fuzzy integral in multi-criteria decision making, Fuzzy Sets and Systems, 69, 279-298.

Grabisch, M.,(1995) Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference,Kluwer Academic Publishers.

Haberlandt, U., V. Krysanova, and A. Bardossy,(2002) Assessment of nitrogen leaching from arable land in large river basins - Part II: regionalisation using fuzzy rule based modelling, Ecological Modelling, 150, 277- 294.

Kardi, T., Similarity Measurement. http:\\people.revoledu.com\kardi\tutorial\Similarity\, Accessed March 2005.

Klir G, J.,(1997) Constructing fuzzy measures in expert systems, fuzzy sets and systems 92, 251-264.

Klir G.J., Yuan, B.,(1995) Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall PTR, Upper Saddle River, NJ.

Klir,G,J.,Wang,Z.,Wang,W.,(1996) Constructing Fuzzy Measures by Transformations, J.Fuzzy Math. 4 (1) 207-215.

Kolodner, J.L.(1991) Improving human decision making through case-based decision aiding. AlMag., 52-68.

Marichal, J.,(1999) Aggregation Operators for Multicriteria decision aid, Ph.D. thesis, University, of Li`ege, Belgium.

Priya, S., and R. Shibasaki,(2001) National spatial crop yield simulation using GIS-based crop production model, Ecological Modelling, 135, 113-129.

Rice, J. A., and P. A. Cochran,(1984) Independent evaluation of a bioenergetics model for largemouth bass, Ecology, 65(3), 732-739.

Sugeno, M.,(1974) Theory of fuzzy integrals and its applications, PhD Dissertation, Tokyo Institute of Technology, Japan.

Vosniadou, Stella and Andrew Ortony (eds.), (1989) Similarity and Analogical Reasoning, Cambridge University Press, New York, USA.

Wagholikar, A., Deer, P., (January 2005) Determination of fuzzy measures using similar data set, the 4th Asia Pacific International Symposium on Information Technology, Gold Coast, Australia, 319-322.

Wagholikar A., Deer, P.,(June 2005) Methods for Fuzzy Measure Determination: A Theoretical Study, Proc. for International Conference on Artificial Intelligence (ICAI'05), Las Vegas, USA.

Zeleny,M., (1982). Multiple Criteria Decision Making, New York: McGraw-Hill.

Zhaohao, S., Finnie, G, R, (2004) Intelligent Techniques in E-Commerce: a case-based reasoning perspective, Springer-Verlag, New York.

Fuzzy Economic Production Quantity Model for Items with Imperfect Quality

Shan Huo Chen
Department of Information Management, Ching Yun University,
229, Chien-Hsin Rd., Jung-Li, 320, Taiwan.
Email: shchen_im@cyu.edu.tw

Chien-Chung Wang
Department of Statistics, National Defense Management College, National Defense University,
150, Ming-An St., Chun-Ho, 235, Taiwan.
Email: wangcc@rs590.ndmc.edu.tw

Shu Man Chang
Department of Shipping Business Management, China College of Marine Technology and Commerce,
Taipei, 111, Taiwan.
Email: smchang@mail.ccmtc.edu.tw

Abstract

In this paper, we have proposed a Fuzzy Economic Production Quantity (FEPQ) model with imperfect products that can be sold in a discount price. Compared with the traditional inventory model, we consider trapezoidal fuzzy costs and quantities in our model. Function Principle and Graded Mean Integration Representation method are applied to find optimal economic production quantity of the fuzzy production inventory model.

Keywords:

Fuzzy inventory model, Economic production quantity, Function principle, Graded mean integration representation method, Optimization, Imperfect production, Defective products.

INTRODUCTION

The fuzzy set concept has been used to treat the classical inventory model recently. Park (1987) considered fuzzy inventory cost in economic order quantity model. Chang (1999) discussed how to get the economic order quantity, when the quantity of demand is uncertain. Chen and Hsieh (2002b) established a fuzzy economic production model to treat the inventory problem with all parameters and variables are fuzzy numbers. Hsieh (2002), Lee et al. (1998), Lin et al. (2000) also wrote some articles about fuzzy production model.

In real world, imperfect products cannot be avoided in most production process. It would be interesting to discuss the models with imperfect products. Recently, Salameh et al. (2000), Mohamed (2002), Lin et al. (2003), Chung et al. (2003), and Lee (2005) have written papers about imperfect production process. The above researchers discussed fuzzy costs, but did not discuss imperfect production model. Some researchers discussed imperfect production processes, but did not discuss fuzzy costs. Therefore a Fuzzy Economic Production Quantity model with imperfect products that can be sold in a discount price would be a interesting topic to treat vague environment problems.

In viewing of production management, the cost should include explicit and implicit cost, and the calculation should be easy. We apply Function Principle to calculate the fuzzy total revenue (FTR) instead of Extension Principle, and use the Graded Mean Integration Representation Method to defuzzify the FTR. Finally, a numerical example is proposed, and the result is quite satisfied.

This paper is organized as follows. In section 2, the methodology is introduced. In section 3, fuzzy inventory model is discussed. A numerical example is shown in section 4, and section 5 concludes this paper.

METHODOLOGY

In this paper, Function Principle and Graded Mean Integration Representation method are used to find the optimal economic production quantity in a fuzzy inventory model. When the quantities are fuzzy numbers, we can use the Kuhn-Tucker conditions to solve the model. Therefore we introduce these three methodologies as following.

Graded Mean Integration Representation Method

Chen et al. (1998, 1999c, 2000b) introduced Graded Mean Integration Representation method based on the integral value of graded mean h-level of generalized fuzzy number for defuzzifing generalized fuzzy number. They also found this method is better than the methods of Adamo (1980), Campos et al. (1989), Yager (1981), Kaufmann et al. (1991), Chen (1998), Lee et al. (1998), Liou et al. (1992), and Heilpern (1997). Now, we describe generalized fuzzy number as following.

Suppose $\widetilde{A}$ is a generalized fuzzy number as shown in Figure 1. It is described as any fuzzy subset of the real line R, whose membership function μ_A satisfies the following conditions:

(1) μ_A is a continuous mapping from R to the closed interval [0, 1],

(2) $\mu_A = 0$, $-\infty < x \le a_1$,

(3) $\mu_A = L(x)$ is strictly increasing on $[a_1, a_2]$,

(4) $\mu_A = w_A$, $a_2 \le x \le a_3$,

(5) $\mu_A = R(x)$ is strictly decreasing on $[a_3, a_4]$,

(6) $\mu_A = 0$, $a_4 \le x < \infty$,

where $0 < w_A \le 1$, and a_1, a_2, a_3, and a_4 are real numbers.

Also this type of generalized fuzzy numbers be denoted as $\widetilde{A} = (a_1, a_2, a_3, a_4; w_A)_{LR}$. When $w_A = 1$, it can be simplified as $\widetilde{A} = (a_1, a_2, a_3, a_4)_{LR}$.

In Graded Mean Integration Representation method, L^{-1} and R^{-1} are the inverse functions of L and R respectively, and the graded mean h-level value of generalized fuzzy number $\widetilde{A} = (a_1, a_2, a_3, a_4; w_A)_{LR}$ is $h(L^{-1}(h) + R^{-1}(h))/2$ as shown in Figure 1. Then the graded mean integration representation of $\widetilde{A}$ is $P(\widetilde{A})$ with grade w_A where

$$P(\widetilde{A}) = \int_0^{w_A} h\left(\frac{L^{-1}(h) + R^{-1}(h)}{2}\right) dh \Big/ \int_0^{w_A} h\, dh, \tag{1}$$

with $0 < h \le w_A$ and $0 < w_A \le 1$.

Throughout this paper, only normal trapezoidal fuzzy number as the type of all fuzzy parameters are used in our proposed fuzzy production inventory models. Let $\widetilde{B}$ be a trapezoidal fuzzy number, and be denoted as $\widetilde{B} = (b_1, b_2, b_3, b_4)$. Then we can get the graded mean integration representation of $\widetilde{B}$ by formula (1) as

$$P(\widetilde{B}) = \int_0^1 h\left(\frac{b_1 + b_4 + (b_2 - b_1 - b_4 + b_3)h}{2}\right)dh \Big/ \int_0^1 h\, dh = \frac{b_1 + 2b_2 + 2b_3 + b_4}{6} \tag{2}$$

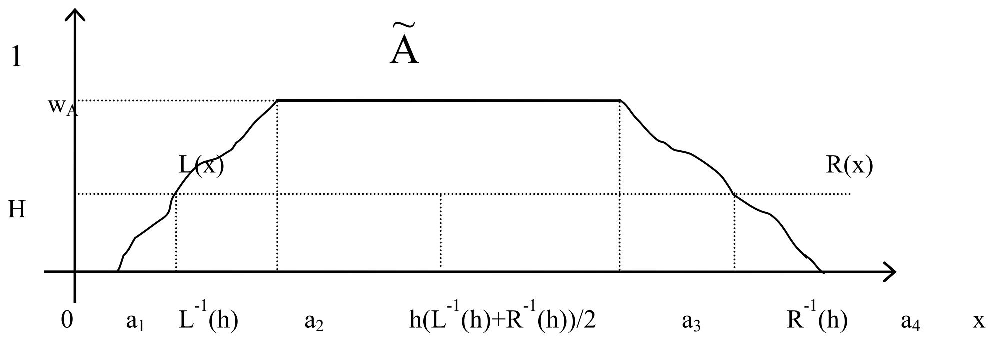


Figure 1. The graded mean h-level value of generalized fuzzy number $\widetilde{A} = (a_1, a_2, a_3, a_4; w_A)_{LR}$.

Fuzzy Arithmetical Operations under Function Principle

Function Principle is introduced by Chen (1986) to treat the fuzzy arithmetical operations by trapezoidal fuzzy numbers. We apply this principle as the operation of addition, multiplication, subtract, division of trapezoidal fuzzy numbers, because (1) Function Principle is easier to calculate than Extension Principle, (2) Function Principle will not change the shape of trapezoidal fuzzy number after the multiplication of two trapezoidal fuzzy numbers, but the multiplication of two trapezoidal fuzzy numbers will become drum like shape fuzzy number by using Extension Principle, (3) If we have to multiple more than four trapezoidal fuzzy numbers, Extension Principle can not solve the operation, but Function Principle can easy to find the result by pointwise computation. Here we describe some fuzzy arithmetical operations under Function Principle as following:

Suppose $\widetilde{A} = (a_1, a_2, a_3, a_4)$ and $\widetilde{B} = (b_1, b_2, b_3, b_4)$ are two trapezoidal fuzzy numbers. Then,

(1)The addition of $\widetilde{A}$ and $\widetilde{B}$ is

$$\widetilde{A} \oplus \widetilde{B} = (a_1+b_1, a_2+b_2, a_3+b_3, a_4+b_4),$$

where a_1, a_2, a_3, a_4, b_1, b_2, b_3, and b_4 are any real numbers.

(2)The multiplication of $\widetilde{A}$ and $\widetilde{B}$ is

$$\widetilde{A} \otimes \widetilde{B} = (c_1, c_2, c_3, c_4),$$

where $T = \{a_1b_1, a_1b_4, a_4b_1, a_4b_4\}$, $T_1 = \{a_2b_2, a_2b_3, a_3b_2, a_3b_3\}$, $c_1 = \min T$, $c_2 = \min T_1$, $c_3 = \max T_1$, $c_4 = \max T$.

If a_1, a_2, a_3, a_4, b_1, b_2, b_3, and b_4 are all nonzero positive real numbers, then

$$\widetilde{A} \otimes \widetilde{B} = (a_1b_1, a_2b_2, a_3b_3, a_4b_4).$$

(3)$-\widetilde{B} = (-b_4, -b_3, -b_2, -b_1)$, then the subtraction of $\widetilde{A}$ and $\widetilde{B}$ is

$$\widetilde{A} \ominus \widetilde{B} = (a_1-b_4, a_2-b_3, a_3-b_2, a_4-b_1),$$

where a_1, a_2, a_3, a_4, b_1, b_2, b_3, and b_4 are any real numbers.

(4)$1/\widetilde{B} = \widetilde{B}^{-1} = (1/b_4, 1/b_3, 1/b_2, 1/b_1)$, where b_1, b_2, b_3, and b_4 are all positive real numbers.

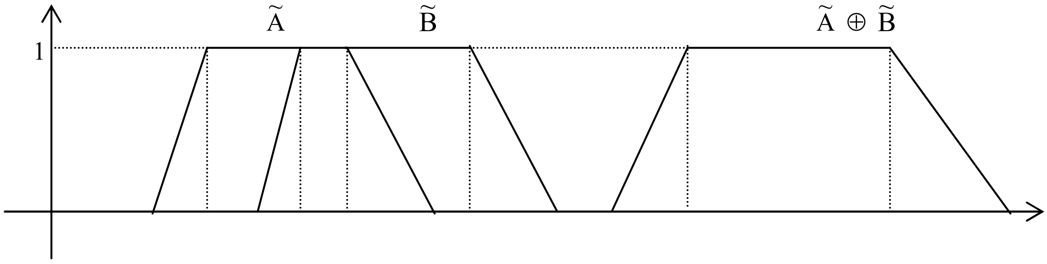
If a_1, a_2, a_3, a_4, b_1, b_2, b_3, and b_4 are all nonzero positive real numbers, then the division of $\widetilde{A}$ and $\widetilde{B}$ is

$$\widetilde{A} \oslash \widetilde{B} = (a_1/b_4, a_2/b_3, a_3/b_2, a_4/b_1).$$

(5)Let $\alpha \in R$, then

$$\begin{cases} \text{(i) } \alpha \geq 0, \alpha \otimes \widetilde{A} = (\alpha a_1, \alpha a_2, \alpha a_3, \alpha a_4), \\ \text{(ii) } \alpha < 0, \alpha \otimes \widetilde{A} = (\alpha a_4, \alpha a_3, \alpha a_2, \alpha a_1). \end{cases}$$

Note: We do not introduce a new addition symbol, as the sum under Extension Principle is the same as Function Principle in Figure 2. For a mathematically minded reader, we observe that the Extension Principle is a form of convolution (Chen, 1986) while the Function Principle is akin to a pointwise multiplication as Figure 3.



0 a_1 a_2 b_1 b_2 a_3 a_4 b_3 b_4 a_1+b_1 a_2+b_2 a_3+b_3 a_4+b_4

Figure 2. Fuzzy addition operation using Function Principle and Extension Principle.

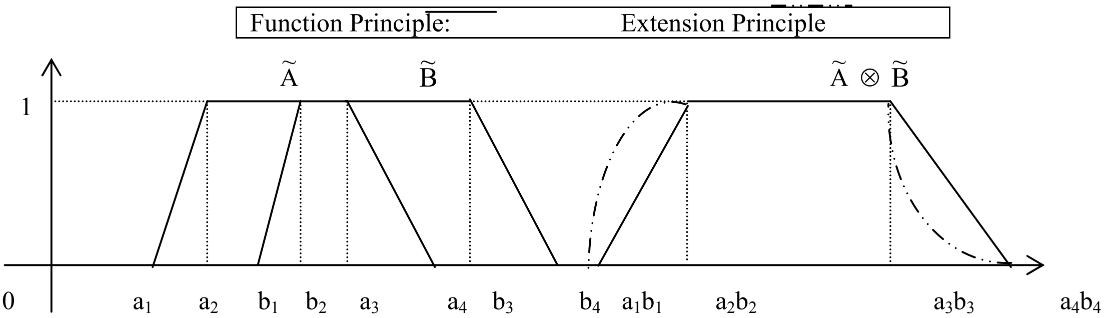


Figure 3. Comparison of fuzzy multiplication operation under Function Principle and Extension Principle.

The Kuhn-Tucker conditions

Taha (1997) discussed how to solve the optimum solution of nonlinear programming problem subject to inequality constraints by using the Kuhn-Tucker conditions. The development of the Kuhn-Tucker conditions is based on the Lagrangean method.

Suppose that the problem is given by

Maximize $y = f(x)$

Subject to $g_i(x) \le 0$, $i=1, 2, \cdots, m$.

The nonnegative constraints $x \geq 0$, if any, are included in the m constraints.

The inequality constraints may be converted into equations by using nonnegative slack variables. Let's S_i^2 be the slack quantity added to the ith constraint $g_i(x) \le 0$. Let $\lambda=(\lambda_1,\lambda_2,\cdots, \lambda_m)$, $g(x)=(g_1(x),g_2(x),\cdots,g_m(x))'$ and $S^2=(S_1^2, S_2^2, \cdots, S_m^2)'$. The Lagrangean functions is thus given by

$L(x, s, \lambda) = f(x) - \lambda[g(x)+S^2)$.

Given the constraints $g_i(x) \le 0$.

Taking the partial derivatives of L with respect to x, S, and λ, we obtain

$$\frac{\partial L}{\partial X} = \nabla f(x) - \lambda \nabla g(x) = 0 \ ,$$

$$\frac{\partial L}{\partial S_i} = -2\lambda_i S_i = 0 \ , \ i=1, 2, \cdots, m.$$

$$\frac{\partial L}{\partial \lambda} = - g(x) - S^2 = 0 \ , \ i=1, 2, \cdots, m.,$$

From the second and third sets of equations, it shows that

$\lambda_i g_i(x) = 0$, $i=1, 2, \cdots, m$.

The Kuhn-Tucker conditions need x and λ to be a stationary point of the minimization problem which can be summarized as follows:

$$\begin{cases} \lambda \geq 0, \\ \nabla f(x) - \lambda \nabla g(x) = 0, \\ \lambda_i g_i(x) = 0, \quad i = 1, 2, \ldots, m. \quad (3) \\ g_i(x) \leq 0, \quad i = 1, 2, \ldots, m. \end{cases}$$

FUZZY INVENTORY MODEL

We thereby discuss the case of imperfect products that can be sold in a discount price with fuzzy costs and fuzzy order quantity. Throughout this paper, we use the following variables in order to simplify the treatment of the fuzzy production inventory model:

$\tilde{H}$: fuzzy daily holding cost per unit,

$\tilde{K}$: fuzzy setup cost,

c : the unit production cost,

s : unit selling price of items of good quality,

v : unit selling price of defective items,(v<c),

x : screening rate, a: unit screening cost,

p : the percentage of defective items in a production lot,

$\tilde{D}$: fuzzy daily demand over the planning time period [0, 365],

$\tilde{Q}$: fuzzy production quantity,

T: cycle time.

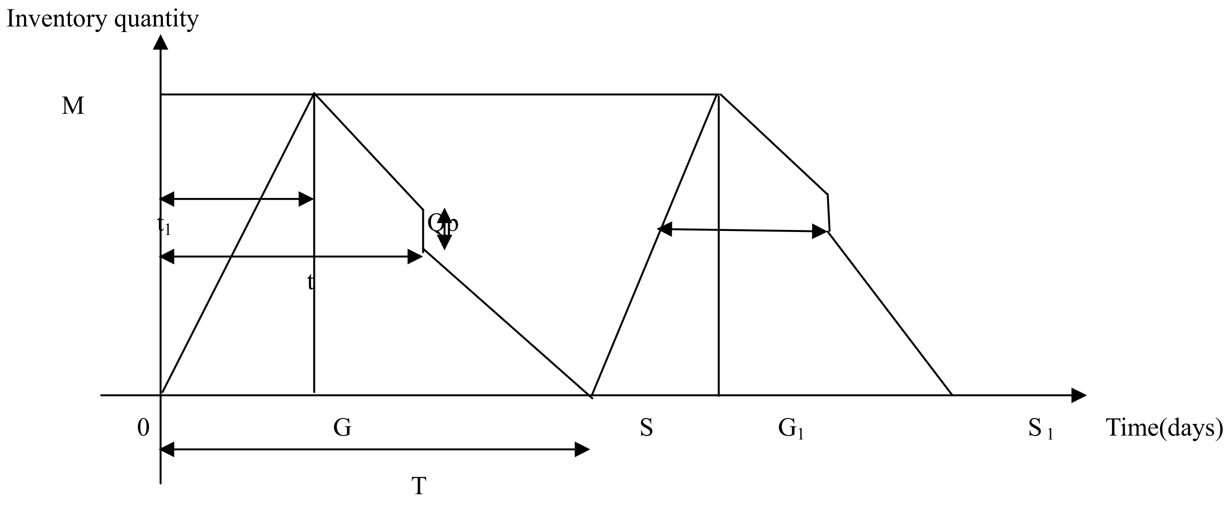


Figure 4. Inventory control and the production process.

The manufacturer produces and sells in the time OG, SG_1, …, etc., and he only sells in the time GS, G_1S_1, …,as shows in Figure 4.

Let FN($\tilde{Q}$, p) be the fuzzy number of good items, and it approximates to FN($\tilde{Q}$, p) = $\tilde{Q} \Theta p \otimes \tilde{Q} = (1 - p) \otimes \tilde{Q}$. To avoid shortage, it is assumed that number of good items is at least equal to the demand during screening time t, that is $FN(\widetilde{Q}, p) \geq \widetilde{D} \otimes t$. Replacing t by $\widetilde{Q} \oslash x$, the value of p is restricted to $p \leq 1 \ominus \widetilde{D} \oslash x$. Now the fuzzy total revenue $FTR(\widetilde{Q})$ is the approximation to $s \otimes \widetilde{Q} \otimes (1 - p) \oplus v \otimes \widetilde{Q} \otimes p$. The fuzzy total cost per cycle $FTC(\widetilde{Q})$ approximates to the sum of procurement cost per cycle, screening cost per cycle and holding cost per cycle that is $\widetilde{K} \oplus c \otimes \widetilde{Q} \oplus a \otimes \widetilde{Q} \oplus \widetilde{H} \otimes \{(\widetilde{Q} \otimes (1\text{-}p) \otimes T) \oslash 2 \oplus p \otimes \widetilde{Q}^2 \oslash x\}$. The total profit $FTP(\widetilde{Q})$ is the approximation to $s \otimes \widetilde{Q} \otimes (1\text{-}p) \oplus v \otimes \widetilde{Q} \otimes p \ominus \{\widetilde{K} \oplus c \otimes \widetilde{Q} \oplus a \otimes \widetilde{Q} \oplus \widetilde{H} \otimes \{(\widetilde{Q} \otimes (1\text{-}p) \otimes T) \oslash 2 \oplus p \otimes \widetilde{Q}^2 \oslash x\}\}$.

Firstly, the fuzzy total profit per unit time ,

$$FTPU(\widetilde{Q}) = TP(\widetilde{Q}) \oslash T \approx \text{(approximates to)}\ \widetilde{D} \otimes (s \ominus v \oplus \widetilde{H} \otimes \widetilde{Q} \oslash x)$$
$$\oplus\ \widetilde{D} \otimes (v \ominus \widetilde{H} \otimes \widetilde{Q} \oslash x \ominus c \ominus a \ominus \widetilde{K} \oslash \widetilde{Q})\,(1/(1 - p)) \ominus [\widetilde{H} \otimes \widetilde{Q} \otimes (1 - p)] \oslash 2, \quad (4)$$

where $\oplus$, $\ominus$, $\otimes$ and $\oslash$ are the fuzzy arithmetical operations under Function Principle.

Here, we suppose $\widetilde{H} = (h_1, h_2, h_3, h_4)$, $\widetilde{K} = (k_1, k_2, k_3, k_4)$, $\widetilde{D} = (d_1, d_2, d_3, d_4)$, $\widetilde{Q} = (q_1, q_2, q_3, q_4)$ are non-negative trapezoidal fuzzy numbers. Then we solve the optimal fuzzy total profit per unit time formula (4) as the following steps.

Firstly,

$$FTPU(\widetilde{Q}) = (\, d_1s+pv/(1\text{-}p)\, d_1\text{-}p\, d_4\, h_4\, q_4/(1\text{-}p)\text{-}c\, d_4/(1\text{-}p)\text{-}a\, d_4/(1\text{-}p)\text{-}\, d_4k_4/(\, q_1(1\text{-}p))\text{-}(1\text{-}p)\, h_4\, q_4/2,$$
$$d_2s+pv/(1\text{-}p)\, d_2\text{-}p\, d_3\, h_3\, q_3/(1\text{-}p)\text{-}c\, d_3/(1\text{-}p)\text{-}a\, d_3/(1\text{-}p)\text{-}\, d_3k_3/(\, q_2(1\text{-}p))\text{-}(1\text{-}p)\, h_3\, q_3/2,$$
$$d_3s+pv/(1\text{-}p)\, d_3\text{-}p\, d_2\, h_2\, q_2/(1\text{-}p)\text{-}c\, d_2/(1\text{-}p)\text{-}a\, d_2/(1\text{-}p)\text{-}\, d_2k_2/(\, q_3(1\text{-}p))\text{-}(1\text{-}p)\, h_2\, q_2/2,$$
$$d_4s+pv/(1\text{-}p)\, d_4\text{-}p\, d_1\, h_1\, q_1/(1\text{-}p)\text{-}c\, d_1/(1\text{-}p)\text{-}a\, d_1/(1\text{-}p)\text{-}\, d_1k_1/(\, q_4(1\text{-}p))\text{-}(1\text{-}p)\, h_1\, q_1/2)$$

Secondly, we defuzzify the fuzzy total profit per unit time by using graded mean integration representation method. The result is

$$P(FTPU(\widetilde{Q}))=\{s(d_1+d_4)+ pv/(1\text{-}p)(\, d_1+d_4)\text{-}p/(1\text{-}p)(d_4h_4q_4+d_1h_1q_1)\text{-}c/(1\text{-}p)(d_1+d_4)\text{-}a/(1\text{-}p)\,(d_1+d_4)\ \text{-}1/(1\text{-}p)$$
$$(d_1k_1/q_4+d_4k_4/q_1)\ \text{-}(1\text{-}p)/2(h_4q_4+h_1q_1\,)\ +2[s(d_2+d_3)+ pv/(1\text{-}p)(\, d_2+d_3)\text{-}p/(1\text{-}p)\,(d_3h_3q_3+d_2h_2q_2)\ \text{-}c/(1\text{-}$$
$$p)(d_2+d_3)\text{-}a/(1\text{-}p)\,(d_2+d_3)\ \text{-}1/(1\text{-}p)(d_2k_2/q_3+d_3k_3/q_2)\text{-}(1\text{-}p)/2(h_2q_2+h_3q_3\,)]\}/6 \quad (5)$$

where $0 < q_1 \leq q_2 \leq q_3 \leq q_4$.

It will not change the meaning of formula (5), if we replace the inequality conditions $0 < q_1 \leq q_2 \leq q_3 \leq q_4$ with the following inequality constrains:

$$q_1 - q_2 \leq 0,\ q_2 - q_3 \leq 0,\ q_3 - q_4 \leq 0, \text{ and } -q_1 < 0.$$

Thirdly, let $L(\widetilde{Q})=P(FTPU(\widetilde{Q}))\text{-}\lambda_1(q_1\text{-}q_2+S_1^2)\text{-}\lambda_2(q_2\text{-}q_3+S_2^2)\text{-}\lambda_3(q_3\text{-}q_4+S_3^2)\text{-}\lambda_4(\text{-}q_1+S_4^2)$, then the Kuhn-Tucker condition is used to find the solution of q_1, q_2, q_3, and q_4 to maximize $P(FTPU(\widetilde{Q}))\ P(\widetilde{F})$ in formula (5) , subject to $q_1 - q_2 \leq 0$, $q_2 - q_3 \leq 0$, $q_3 - q_4 \leq 0$, and $-q_1 < 0$. The Kuhn-Tucker conditions are thus given as formula (3):

$$\lambda \geq 0,$$
$$\nabla f(P(FTPU(\widetilde{Q}))) - \lambda \nabla g(Q) = 0\ ,$$
$$\lambda_i g_i(Q) = 0,$$
$$g_i(Q) \leq 0,$$

These conditions simplify to the following

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0, \quad (6\text{-}1)$$
$$1/6[\text{-}p/(1\text{-}p)d_1h_1+d_4k_4/\,(q_1^2\,(1\text{-}p))\text{-}(1\text{-}p)/2h_1)]\text{-}\lambda_1 + \lambda_4 = 0, \quad (6\text{-}2)$$
$$1/6\{2[\text{-}p/(1\text{-}p)d_2h_2+d_3k_3/\,(q_2^2\,(1\text{-}p))\text{-}(1\text{-}p)/2h_2)]\}+\lambda_1 - \lambda_2 = 0, \quad (6\text{-}3)$$

$$1/6\{2[-p/(1-p)d_3h_3+d_2k_2/(q_3^2(1-p))-(1-p)/2h_3)]\}+\lambda_2-\lambda_3=0, \quad (6\text{-}4)$$

$$1/6[-p/(1-p)d_4h_4+d_1k_1/(q_4^2(1-p))-(1-p)/2h_4)]+\lambda_3=0, \quad (6\text{-}5)$$

$$\lambda_1(q_1-q_2)=0, \quad (6\text{-}6)$$

$$\lambda_2(q_2-q_3)=0, \quad (6\text{-}7)$$

$$\lambda_3(q_3-q_4)=0, \quad (6\text{-}8)$$

$$\lambda_4 q_1=0, \quad (6\text{-}9)$$

$$q_1-q_2 \le 0, \quad (6\text{-}10)$$

$$q_2-q_3 \le 0, \quad (6\text{-}11)$$

$$q_3-q_4 \le 0, \quad (6\text{-}12)$$

$$-q_1 \le 0. \quad (6\text{-}13)$$

Because $q_1>0$, and $\lambda_4 q_1=0$, then $\lambda_4=0$. If $\lambda_1=\lambda_2=\lambda_3=0$, then $q_4<q_3<q_2<q_1$, it does not satisfy the constrains $0<q_1\le q_2\le q_3\le q_4$. Therefore $q_2=q_1$, $q_3=q_2$, and $q_4=q_3$, that is $q_1=q_2=q_3=q_4=Q$. Hence, from formula (6-2), (6-3), (6-4), and (6-5) we find the optimal production quantity Q^* by the above equation as

$$Q^*=\sqrt{\frac{2(d_1k_1+2d_2k_2+2d_3k_3+d_4k_4)}{2p(d_1h_1+2d_2h_2+2d_3h_3+d_4h_4)+(1-p)^2(h_1+2h_2+2h_3+h_4)}} \quad (7)$$

When demand, production and costs are real numbers, that is $h_1=h_2=h_3=h_4=H$, $k_1=k_2=k_3=k_4=K$, and $d_1=d_2=d_3=d_4=D$, then formula (7) can be revised as

$$Q^*=\sqrt{\frac{2KD}{(2pD+(1-p)^2)H}}$$

When p=0, then $Q^*=\sqrt{\frac{2KD}{H}}$.

EXAMPLE

ABC manufacturing company produces commercial television units in batch. The firm estimated that the fuzzy daily storage cost ($\tilde{H}$) per unit is about NT$1, the fuzzy setup cost($\tilde{K}$)is about NT$100,000, the unit production cost (c) is NT$5,000, the percentage of defective items in a lot (p) is about 1%, the screening rate (x) is 1unit/10mins, the screening cost (a) is NT$20/unit, the unit selling price of items of good quality (s)is NT$10,000, the unit selling price of defective items (v) is NT$3,000, the fuzzy total demand over the planning time period [0,365] ($\tilde{D}$) is greater or less than 20,000 units. How many television units should ABC manufacturing company produce in each batch?

[Solving]:

Here we use a general rule to transfer the linguistic data, "greater or less than X" and "about X", into trapezoidal fuzzy numbers as

"greater or less than X" = (0.9X, 0.95X, 1.05X, 1.1X), and

"about X" = (0.95X, X, X, 1.05X).

By using the above rule, the fuzzy parameters in this example can be transferred as follows: $\tilde{H}$ = (0.95, 1, 1, 1.05), $\tilde{K}$ = (95000, 100000, 100000, 105000), c=5000, $\tilde{D}$ =(18000, 19000, 21000, 22000).

Replace the above fuzzy parameters values into formula (7), we find the optimal fuzzy production quantity

$\widetilde{Q}$ = (2234.25, 2234.25, 2234.25, 2234.25) $\approx$(2234, 2234, 2234, 2234).

Then, the minimization fuzzy total production inventory cost is

FTTU($\widetilde{Q}$) = (67423065, 82667569,114246867,128233884).

The result shows that the production units per batch for ABC manufacturing company is 2234.

CONCLUSION

In real world, defective products cannot be avoided in some production processes. It may be possible and reasonable to discuss the model with defective products. This model is applicable when inventory continuously flows or builds up over a period of time after an order has been placed and units are produced and sold simultaneously. Under this circumstance, the length of the planning time period is measured in days.

REFERENCES

Adamo, J. M. (1980). Fuzzy decision trees, Fuzzy sets and systems, 4:207-219.

Campos, L. and Verdegay, J.L.,(1989). Linear programming problems and ranking of fuzzy numbers, Fussy sets and systems, 32:1-11.

Chang, S. C. (1999). Fuzzy production inventory for fuzzy product quantity with triangular fuzzy number. *Fuzzy sets and systems*, 107: 37-57.

Chen, S. H. (1985). Fuzzy linear combination of fuzzy linear functions under Extension Principle and Second Function Principle. *Tamsui Oxford Journal of Management Sciences*, 1: 11-31.

Chen, S. H. (1986). Operations on fuzzy numbers with function principle. *Tamkang Journal of Management Sciences*, 6(1): 13-25.

Chen, S. H., Wang, C. C. and Arthur Ramer, A. (1996). Backorder Fuzzy Inventory Model under Function Principle. *Information sciences*, 95(1) (2): 71-79.

Chen, S. H. and Hsieh, C. H. (1998). Graded mean integration representation of generalized fuzzy numbers. *The 6th Conference on Fuzzy Theory and Its Applications* (CD-ROM), filename: 031.wdl, Chinese Fuzzy Systems Association, Taiwan, pp. 1-6.

Chen, S. H. and Hsieh, C. H. (1999a). Optimization of Fuzzy Simple Inventory Models. *1999 IEEE International Fuzzy Systems Conference*, Seoul, Korea, Vol. 1, pp.240-244.

Chen, S. H. and Hsieh, C. H. (1999b). Optimization of Fuzzy backorder Inventory Models. *1999 IEEE International Conference on Systems*, Man and Cybernetics (SMC'99), Tokyo, Japan, pp.425 (File name of CD-ROM is Fq041).

Chen, S. H. and Hsieh, C. H., (1999c.). Graded mean integration representation of generalized fuzzy numbers. *Journal of The Chinese Fuzzy Systems Association,* 5(2): 1-7.

Chen, S. H. and Hsieh, C. H. (2000a). Optimization of Fuzzy Production Inventory Model under Fuzzy Parameters. *The 5th Joint Conference on Information Sciences(JCIS'2000)* ,Atlantic city, USA, vol. 1, pp. 68-71.

Chen, S. H. and Hsieh, C. H., (2000b). Representation, Ranking, Distance, and Similarity of L-R type fuzzy number and Application. *Australian Journal of Intelligent Processing Systems*, 6: 217-229.

Chen, S.M., (1998). A new method for tool steel material selections under fuzzy environment, Fuzzy sets and systems, 92:265-274.

Chung, K.J. and Hou, K.L., (2003). An optimal production run time with imperfect production processes and allowable shortages. Computers and operations research, 30:483-490.

Heilpern, S., (1997). Representation and application of fuzzy number, Fuzzy sets and systems, 91:259-268.

Hong, J. D. and Hayya, J. C. (1995). Joint investment in quality improvement and setup reduction. *Computers and Operations Research,* 22: 567-574.

Hsieh, C.H., (2002), Optimization of fuzzy production inventory models. Information sciences, 146:29-40.

Hwang, H., Kim, D. B. and Kim, Y. D. (1993). Multiproduct economic lot size models with investments costs for setup reduction and quality improvement. *International Journal of Production Research*, 31:691-703.

Kaufmann, A. and Gupta, M.M. (1991). Introduction to fuzzy arithmetic theory and applications, van Nostrand Reinhold.

Keller, G. and Noori, H. (1988). Impact of investing in quality improvement on the lot size model. *OMEGA International Journal of Management Sciences,* 15: 595-601.

Lee, E.S. and Li, R.J., (1988). Comparison of Fuzzy numbers based on the probability measure of fuzzy events, computer and mathematics with application, 15:887-896.

Lee, H.H., (2005). A cost/benefit model for investments in inventory and preventive maintenance in an imperfect production system. Computer and industrial engineering, 48:55-68.

Lee, H.M. and Yao, J. S., (1998). Economic production quantity for fuzzy demand quantity and fuzzy production quantity. European journal of operational research, 109:203-211.

Lin, C.S., Chen, C.H., Kroll, Dennis, E., (2003). Integrated productions-inventory models for imperfect production processes. Computers and Industrial Engineering, vol. 44, issue 4, 633-650.

Lin, D.C. and Yao, J. S.,(2000). Fuzzy economic production for production inventory. Fuzzy sets and systems, 111:465-495.

Liou, T.S., Wang, M.J.J., (1992). Ranking fuzzy numbers with integral values, Fuzzy sets and systems, 50:247-255.

Mirko, V., Dobrila, P. and Radivoj, P. (1996). EOQ formula when inventory cost is fuzzy. *International Journal of Production Economics*, 45: 499-504.

Mohamed, B.D., (2002). The economic production lot-sizing problem with imperfect production processes and imperfect maintenance. Int. J. of production economics, 76:257-264.

Moon, I. (1994). Multiproduct economic lot size models with investment costs for setup reduction and quality improvement: Review and Extensions. *International Journal of Production Research*, 32: 2795-2801.

Ouyang, L. Y. and Chang, H. C. (1999). Impact of investing in quality improvement on (Q,r,L) model involving the imperfect production process. *Production planning and Control*, 11: 598-607.

Park, K. S. (1987). Fuzzy set theoretic interpretation of economic order quantity. *IEEE Transactions on Systems*. Man, and Cybernetics SMC-17, 1082-1084.

Porteus, E. L. (1986). Optimal lot sizing process quality improvement and setup cost reduction. *Operations Research*, 34: 137-144.

Rosenblatt, M. J. and Lee, H. L. (1986). Economic production cycles with imperfect production processes. *IIE Transitations*, 18: 48-55.

Roy, T. K. and Maiti, M. (1997). A fuzzy EOQ model with demand-dependent unit cost under limited storage capacity. *European Journal of Operational Research,* 99: 425-432.

Salameh, M.K. and Jaber, M.Y. (2000). Economic production quantity model for items with imperfect quality. Int. J. of production economics, 64: 59-64.

Taha, H. A. (1997). *Operations research*. Prentice Hall, New Jersey, USA.

Yager, R.R., (1981). A procedure for ordering fuzzy subsets of the unit interval, Information sciences, 24:143-161.

Yao, J. S. and Lee, H. M. (1996). Fuzzy inventory with backorder for fuzzy order quantity. Information Sciences, 93: 283-319.

A fuzzy-neural approach for predicting literature research tendency

Yun-yi Liu, Dr. Liang Liang
University of Science and Technology of China
Hefei, China
Email: yunyiliu@mail.ustc.edu.cn
Email: lliang@ustc.edu.cn

Abstract

This paper proposes a fuzzy-neural approach for predicting literature research tendency. The system begins with the auto-feature-abstract retrieved by a programmed sampling process. A general neural-network model for fuzzy logic control is presented. This model adopts backpropagation multilayer net. A fuzzy logic control network is constructed automatically by learning the training examples itself. By combining both unsupervised and supervised learning schemes, the learning speed converges much faster than the original backpropagation learning algorithm. The structure avoids consuming rule-matching time of the inference engine in the traditional fuzzy logic system. Experimental example is presented to illustrate the performance and applicability of the proposed model.

Keywords:

Fuzzy logic, Auto-feature-abstract retrieval, Fuzzy neural networks, Classify, Predict

INTRODUCTION

Literature plays a very important part in research work. It is extremely vital for scholars to choose their research orientation. Analyzing the relationship among academic papers on all kinds of subjects and deducing the future tendency contribute to efficient research work. The purpose of feature abstracts on literature is to facilitate quick and accurate identification of the tendency of published papers. The objective is to save a prospective researcher time and effort in finding useful information about research orientation. The preparation of feature abstracts is an intellectual effort, requiring general familiarity with the subject. To bring out the salient points of related papers calls for skill and experience. Consequently a large amount of qualified manpower that could be used to advantage in other ways have to be diverted to the task of facilitating access to information. This widespread problem is being aggravated by the ever-increasing output of literature (Boger & Guterman, 1997).

The application of machine methods to literature searching and analyzing is currently receiving a great deal of attention. The system outlined here begins with the document in machine-readable form and proceeds by means of a programmed sampling process comparable to the scanning a human reader would do. However, instead of sampling at random, as a reader normally does when scanning, the new mechanical method selects those among all the sentences of a paper that are the most representative of pertinent information. These feature sentences are then enumerated to serve as clues for judging the character of the paper. Thus, citations of the author's own statements constitute the "auto-feature-abstract". The programs for creating auto-feature-abstracts must be based on writing properties ascertained by analysis of specific types of literature.

This paper is structured as follows: in the next section, a module on mutual dependence is presented and used to select features for later processes. Then the notion of equivalent radius is introduced. Next the algorithm for classification, based on the module of mutual dependence and equivalent radius, will be explained, and the classifying algorithm is used to classify Chinese-language management literature. In section 3, the construction of nonnumeric training data and the design of the fuzzy neural network will be proposed. In section 4, experimental example is presented to illustrate the performance and applicability of the proposed model. Finally, in section 5 this paper will end with the conclusion and stating future work.

PROPOSED RESEARCH METHODOLOGY

In the research of natural language processing, the specimen should be segmented for later processes. The segmenting algorithms can be classified into two categories (Huang, 2000; Boger et al, 2001). Each of them outperforms the other in different aspects. In this paper, a module on mutual dependence, constructed by fetching the virtues and discarding the shortcomings of mutual information and dependence value modules, is put forward to decide the correlation between two words and then used to segment a document. So not only is the precision of segmentation raised, but dimension is also decreased. Furthermore, according to Apriori property, the following rule is deduced and used to decrease dimension.

Definition1 (MD——Mutual Dependence)

The mutual dependence between two variables is:

$$MD(\chi,\eta)=[F(s)\times L-F(\chi)\times F(\eta)]\times \log(\frac{F(s)\times L}{F(\chi)\times F(\eta)})^{\frac{1}{L^2}} \quad \text{---(1)}$$

Where $F(\chi)$ and $F(\eta)$ are frequency of two variables, $F(s)$ is the concurrence frequency of the two variables, and L is the length of a document.

Theorem 1: The range of the mutual dependence between two variables is $[0,\frac{1}{4}\log L)$, in which L is the length of specimen.

The proof of theorem 1 is shown in publication (Wang, 2002). According to Theorem 1, the threshold of Mutual Dependence for segmentation μ_1 and μ_2, can be determined pertinently ($0<\mu_2<\mu_1<\frac{1}{4}\log L$).

The classifying algorithm

The classifying algorithm is based on the module of Mutual Dependence and the notion of equivalent radius. The classifying algorithm begins with constructing the criterion function for class ω_i . To classify the specimen $x=(x_1,x_2,...,x_d)$, $\frac{(\chi_j-Center_{ij})^2}{(R_{ij}^{Equal})^2}$ is taken as a relative distance component of the criterion function, which is the ratio between the square of the projective difference in the dimension d_j of class ω_i and the specimen, and that of R_{ij}^{Equal} of the class. Since the constructed criterion function is a relative measure deduced from the concrete distribution of the training specimen rather than an absolute quantitative one, the miscarriage of justice caused by different number of specimen and different scope in different class can be avoided, then the classifying accuracy can be improved. Besides, if the projection of the class in dimension d_j is equal to 0, R_{ij}^{Equal}=0, it will result in the division overflow. For this reason, the distance coefficient $\frac{1}{\beta^2}$ is introduced. The product of $\frac{1}{\beta^2}$ and the square of the projection of the specimen in the dimension is taken as the relative distance subsection in that dimension. The criterion function is:

$$g_i(x)=\sqrt{\sum_{j=1}^{k}\frac{(x_j-Center_{ij})^2}{(R_{ij}^{Equal})^2}+\sum_{j=k+1}^{d}\frac{\chi_j^2}{\beta^2}}\text{, (i=1, 2,…, c) --- (2)}$$

If class ω_i is the nearest to the specimen, to make the relative distance $g_i(x)$ minimal, then the specimen is classified to the class ω_i. So, the decision rule is:

If $g_j(x)=\min_i g_i(x)$, i=1, 2… c, then $x\in\omega_j$. --(3)

In the later experiment, it is instantiated that the performance of classification is not too sensitive to the distance coefficient. The reason is that, most of the dimensions, in which the projection of class ω_i is 0, are the specific features of some other class, and slight change of the distance coefficient cannot affect the performance too heavily.

Algorithm 1: A simple and efficient algorithm to classify texts

Input: Training corpora; Testing corpora; Let c be the number of classes and d be dimension respectively.

Output: Specimen are classified into proper classes.

Steps:

a) select features from training corpora by segmentation based on Mutual Dependence;

b) construct the classifying module with R_{ij}^{Equal}, $Center_{ij}$ and a_{ij} $(1\le i\le c, 1\le j\le d)$ from training corpora:

c) For i=1 to c and j =1 to d, normalize weights of specimen in class ω_i;

d) project all specimen in class ω_i to each dimension d_j and record $Range_{ij} = [R_{ij}^{-} R_{ij}^{+}]$;

e) calculate $a_{ij} = \dfrac{n_{ij}^{-}}{n_{ij}^{-} + n_{ij}^{+}}$ and R_{ij}^{Equal} in dimension d_j according to formula (1);

f) j=j+1, i=i+1;

g) fetch one untreated specimen x from testing corpora, get the weights of specimen x;

h) calculate the value of $g_i(x)$ according to formula (2);

i) If $g_j(x)$ satisfies rule (3) then specimen x is classified into class ω_j;

k) If all specimen have been treated, then end of the algorithm; Otherwise, turn to step f).

Besides, the classifying module can be easily updated. On adding or deleting training specimen, the new classifying module can be constructed by adjusting the old classifying module properly using weights of the specimen, and there is no need to use all training specimen to train the classifier again. Moreover, only the barycentre $Center_{ij}$ and the ER R_{ij}^{Equal} of the class ω_j $(i = 1,...,c)$, which the specimen belongs to, need to be adjusted, and the updating process has nothing to do with other portions of the classifying module. Since the training corpora are expected to be as large as possible when to construct the classifier, the following updating algorithm focuses on depicting the updating process to add specimen. The updating process to delete specimen can be got by analogy. When adding a training specimen $x = (x_1, x_2,..., x_d)$, the new barycentre $Center_{ij}^{(1)}$ has the following relation with the old barycentre $Center_{ij}^{(0)}$:

$$Center_{ij}^{(1)} = \frac{\mathrm{n} \times Center_{ij}^{(0)} + x_j}{n+1} \text{ --- } (4)$$

After the former processes, n_{ij}^{+} and n_{ij}^{-} should be adjusted too. When more than one specimen are to be added, the new module can be constructed by iterating the three aforementioned steps in order. So the updating process for module is quick and easy to be implemented. As the computing complexity and convenient update are concerned, the algorithm can be used to classify a large number of specimen.

THE DESIGN OF THE FUZZY NEURAL NETWORK

After obtaining feature abstracts, a degenerated fuzzy-number processing system based on artificial neural networks (ANNs) is introduced. The digital representation of fuzzy numbers is assumed, where the universe of discourse is discretized into equally divided intervals. It is proposed that fuzzy-number processing be performed in two basic stages (Freeman & Skapura, 1991; Hideyuki, 1991; Kosko, 1994). The first stage performs the retrieval of fuzzy data consisting of degenerated fuzzy numbers and the second stage performs the desired fuzzy operations on the retrieved data. The method of incomplete fuzzy-number retrieval is proposed based on an ANN structure that is trained to estimate the missing membership function values. Most of the work done thus far has been focused on incorporating neural networks into the existing classical retrieval methods. This section demonstrates how to apply neural networks with fuzzy logic for feature-abstract retrieval. In particular, the Fuzzy Neural Network (FNN) is used as an example to illustrate the versatility of fuzzy neural networks as applied to the feature-abstract-retrieval process. In this section, the training, pattern recall, ranking, relevance feedback and performance issues of the FNN feature-abstract-retrieval process are discussed.

Context-sensitive fuzzy clustering

One should become aware that the ensuing optimization occurs at the numeric level. The underlying objective of this study is to make the contribution of fuzzy sets more visible and radical. We depart from a purely numeric oriented style of the design of neural networks and revisit this very development problem at the nonnumeric level. Here nonnumeric information granules are formed as a result of some sound summarization of the original elements of the training set and made these granules directly available to neural networks for design purposes. We denote the training set as:

{(x(k), y(k))}; k =1, 2,…,n ,where x(k)□ R^n and y(k)□R.

Here, special attention is paid to the problem of a spatial summarization of detailed numeric information into the elements of lower granularity guiding the learning mechanisms. The crux of this methodological avenue pursued in neural networks is to carry out some preprocessing of the training data with an intent of enhancing the learning abilities of the original neural network and accelerating its pace of learning.

In what follows, we concentrate on the contextual Fuzzy C—means (FCM). The method was also discussed in the setting of knowledge discovery and data mining (Pedrycz,1998). We present this very idea in a more detailed fashion as it becomes geared into the learning thrust of this study. Firstly, we briefly summarize the generic fuzzy clustering method in the form as being usually encountered in the literature. This makes the material self-contained and helps understand the context-sensitive enhancement by contrasting it with the existing approach.

One should stress that standard FCM is direction free constructed. What happens is a strong effect of averaging as all variables take over the output one. This may distribute the prototypes in the output space so the ensuing model may not be able to cope with the entire range of the output. The linguistic contexts, particularly those more specific, invoke only a subset of the original data, thus accelerating the learning. Essentially, the context reduces the entire data set into its subset that is defined by the support of the context.

Our main thrust in the design of the neural network is to support processing of linguistic information. Bearing this in mind, we admit the individual inputs and outputs to be fuzzy sets. A basic notation of fuzzy sets with trapezoidal membership functions. This choice is fully justified considering the uncertain quantities of reasonable models with the ensuing straightforward computations. In particular, trapezoidal or triangular fuzzy sets have already been recognized as versatile models in the technology of fuzzy sets. In the discussed model of the network, we use a slightly different description of trapezoidal fuzzy sets. By a fuzzy neuron, we mean a processing unit with fuzzy inputs and fuzzy outputs, all being modeled as trapezoidal fuzzy sets. The weights that characterize the connections between the nodes are represented by four coordinates (b_w ; c_w ; α_w ; β_w), one for every defining point of the trapezoidal fuzzy sets.

The output of the neuron can be approximated as a trapezoidal fuzzy set; we enumerate two phases of this construction:

Phase 1: a weighted sum is formed as $T^{'}(net; b_{net}, c_{net}, \alpha_{net}, \beta_{net})$

where $b_{net} = \sum_{i=1}^{n} b_w(i) b_x(i)$, $c_{net} = \sum_{i=1}^{n} \beta_w(i)\beta_x(i) b_w c_w \alpha_w \beta_w T^{'}(y; b_y, c_y, \alpha_y, \beta_y)$

$$\alpha_{net} = \sum_{i=1}^{n} \alpha_w(i)\alpha_x(i) \quad \beta_{net} = \sum_{i=1}^{n} \beta_w(i)\beta_x(i)$$

,

Assuming that the inputs are given, the constraints translated into a number of admissible strategies regarding the changes of the connections of the network. Here we proceed with a two-stage mode starting from all coordinates being equal $(b_w = c_w = \alpha_w = \beta_w)$ and allowing them to become different afterwards.

Phase 2:

The trapezoidal fuzzy set of the weighted sum becomes transformed nonlinearly via an activation function f( ). To use only one type of fuzzy numbers, we approximate the result as a trapezoidal fuzzy set $T^{'}(y; b_y, c_y, \alpha_y, \beta_y)$ with the parameters being computed as:

$T^{'}(y; b_y, c_y, \alpha_y, \beta_y)$ =(y; f(b_{net}), f(c_{net}), f(b_{net})-f(b_{net} - α_{net}), f(c_{net} + β_{net})-f(c_{net}))

In fact, the activation function f( ) is nonlinear, and b_y , c_y are points, whereas α_y , β_y are intervals.

The learning algorithm

We first discuss the performance index and afterwards elaborate on the associated learning mechanisms. As the output of each neuron arises as a fuzzy set, there is a need to re-examine a way of comparing it with the required target. There are two possible categories of methods:

(i) The use of fuzzy arithmetic. We define a fuzzy set of error Q (not necessarily a trapezoidal fuzzy set) as:

Q = $(target_1 - Y_1)^2$ + $(target_2 - Y_2)^2$ +…+ $(target_m - Y_m)^2$

where "m" denotes the number of outputs of the fuzzy neural network. The summation as well as the square function are taken over the fuzzy sets of the outputs. Thus all computations, dealing with fuzzy sets, heavily hinge on the extension principle (Dubois & Prade, 1988). Overall, these methods are computationally expensive.

(ii) The second option is to calculate a distance (or a certain similarity index) between the target fuzzy set and the corresponding output of the network. Here the ways of capturing this problem are numerous. We select the one that does not carry a significant computational overhead. Let us minimize the squared difference between the parameters of the fuzzy sets (Tsai, 2000; Zopounidis& Doumpos 2002). There are some approximations to the learning procedure in order to speed up the learning process.

Owing to its introduced architecture along with the number of provisions made with respect to the learning mode, the fuzzy neural network exhibits several useful properties. First, and this should be again stressed, the network was developed at the higher non numeric level to handle linguistic information. The training was based on information of a certain granularity and the primordial intent was to make the network operate at this selected level of granularity of the input information.

In fact, most of the unnecessary numeric details were hidden (on purpose) from the network during its training. As a byproduct, the granulation mechanism of this pre-learning aggregation of information delivers a regularization effect. It does not only smooth out the data (that is what a regularization term in the extended performance index really enforces), but also focuses on some selected regions of data and models their characteristics at some suitable level of specificity.

The consideration of nonnumeric information in the training set has a remarkable effect on the resulting performance of the neural network. Now, if the ensuing fuzzy neural network is furnished with a detailed numeric information (that means that the trapezoidal membership functions collapse to singletons; say T(x; a; a; a; a)), then the corresponding output of the network remains, in general, a fuzzy set of some, eventually higher, granularity. Now assume that (x, target) is a testing input–output pair of numeric data.

The result produced by the fuzzy neural network is apparently a fuzzy set (Pedrycz, 1998; Pavur, 2002). The evaluation of the performance of the network is done by comparing whether the output Y of the network, Y =NN(x), includes the target value. In general, we can talk about a degree of containment of the target in Y, or, in other words, a membership value Y (target). For the trapezoidal output fuzzy set, three distinct situations may occur:

1) Y (target) =1; the target is contained in the core of the output fuzzy set,

2) 0< Y(target)<1; the target is contained in the support of the output fuzzy set,

3) Y (target)=0; the target and the output fuzzy set are disjoint.

For the fixed target, the broader Y is, the higher the value of the degree. This sheds light on the interpretation of the result generated by the fuzzy neural network. By making the training set coarser (by composing it with fuzzy sets of lower granularity), we can arrive at the network with broader, less specific outputs. The confidence in the obtained result gets higher while the granularity of the result itself gets lower. The training set with the data set of higher granularity implies higher resolution however this could come associated with the lower confidence level.

This illustrative numerical study emphasizes the main features of the fuzzy neural network. Let us start with a two-input single-output training set.

Table1. Note that the fuzzy data are represented in the notation $T^{'}(x;b,c,\alpha,\beta)$ explicitly expressing the core of the fuzzy set and its spreads. Learning set consisting of trapezoidal fuzzy sets as follows:

Input1	Input2	Target
{0.10, 0.30, 0.10, 0.10}	{0.45, 0.55, 0.10, 0.10}	{0.85, 0.95, 0.05, 0.05}
{0.85, 0.95, 0.05, 0.05}	{0.30, 0.60, 0.10, 0.10}	{0.85, 0.95, 0.05, 0.05}
{0.35, 0.40, 0.05, 0.05}	{0.15, 0.20, 0.05, 0.05}	{0.10, 0.20, 0.10, 0.10}
{0.45, 0.50, 0.05, 0.05}	{0.30, 0.35, 0.05, 0.05}	{0.15, 0.25, 0.10, 0.10}
{0.45, 0.50, 0.05, 0.05}	{0.45, 0.50, 0.05, 0.05}	{0.05, 0.15, 0.10, 0.10}
{0.60, 0.65, 0.05, 0.05}	{0.60, 0.65, 0.05, 0.05}	{0.20, 0.25, 0.10, 0.10}

Table 1. Learning set

This low number of the input variables helps interpret and visualize the results. The elements of the training set exhibit a wide range of granularity. The training exploits the introduced BP scheme with the learning rate changing linearly from 1.0 to 0.1 during the course of learning. The number of learning epochs was 100,000. The momentum was again changed linearly in the range from 0.05 to 0.10 over the period of training. The hidden layer comprises four elements. The results of learning are summarized in Table 2; the target fuzzy sets are contrasted with the outputs of the fuzzy neural network. It becomes obvious that the network has learned the training data quite well.

Input1	Input2	Target	Output FNN

{0.10, 0.30, 0.10, 0.10}	{0.45, 0.55, 0.10, 0.10}	{0.85, 0.95, 0.05, 0.05}	{0.8490, 0.9319, 0.0525, 0.0302}
{0.85, 0.95, 0.05, 0.05}	{0.30, 0.60, 0.10, 0.10}	{0.85, 0.95, 0.05, 0.05}	{0.8504, 0.9513, 0.0571, 0.0208}
{0.35, 0.40, 0.05, 0.05}	{0.15, 0.20, 0.05, 0.05}	{0.10, 0.20, 0.10, 0.10}	{0.1009, 0.2314, 0.0769, 0.1045}
{0.45, 0.50, 0.05, 0.05}	{0.30, 0.35, 0.05, 0.05}	{0.15, 0.25, 0.10, 0.10}	{0.1384, 0.1950, 0.0852, 0.0927}
{0.45, 0.50, 0.05, 0.05}	{0.45, 0.50, 0.05, 0.05}	{0.05, 0.15, 0.10, 0.10}	{0.0813, 0.1785, 0.0408, 0.0998}
{0.60, 0.65, 0.05, 0.05}	{0.60, 0.65, 0.05, 0.05}	{0.20, 0.25, 0.10, 0.10}	{0.1952, 0.2546, 0.1064, 0.1151}

Table 2. Training results

Now we analyze the behavior of the network for some selected inputs of the trained neural networks. These inputs are summarized in Table 3.

Input1	Input2	Output FNN
{0.10, 0.20, 0.10, 0.10}	{0.45, 0.55, 0.10, 0.10}	{0.8490, 0.9888, 0.0525, 0.0050}
{0.20, 0.30, 0.10, 0.10}	{0.45, 0.55, 0.10, 0.10}	{0.3735, 0.9319, 0.0770, 0.0302}
{0.85, 0.95, 0.05, 0.05}	{0.30, 0.40, 0.10, 0.10}	{0.8504, 0.9537, 0.0571, 0.0175}
{0.85, 0.95, 0.05, 0.05}	{0.40, 0.50, 0.10, 0.10}	{0.8312, 0.9495, 0.0684, 0.0205}
{0.85, 0.95, 0.05, 0.05}	{0.50, 0.60, 0.10, 0.10}	{0.8049, 0.9513, 0.0898, 0.0208}
{0.45, 0.50, 0.05, 0.05}	{0.15, 0.20, 0.05, 0.05}	{0.2089, 0.3496, 0.1587, 0.1237}
{0.45, 0.50, 0.05, 0.05}	{0.20, 0.25, 0.05, 0.05}	{0.1873, 0.2828, 0.1335, 0.1140}
{0.20, 0.25, 0.10, 0.10}	{0.35, 0.40, 0.10, 0.10}	{0.3475, 0.9049, 0.1190, 0.0370}

Table 3. Selected inputs and the corresponding outputs of the network

In this section, we combine the two essential and already discussed building concepts into a coherent design platform within which fuzzy neural networks are developed. As these were discussed so far, we concentrate on the design issues.

The contexts and the resulting prototypes in the input space are directly used towards the construction of the fuzzy neural networks. Note that the contexts are already nonnumeric quantities. We would like to summarize the input data in the same way. The simplest option to be pursued deals with trapezoidal (triangular) fuzzy sets formed around the numeric prototypes of the clusters (Moon ea al, 2001; Charalampidis, 2001; Parrado et al, 2003). As all membership values are available through the partition matrices, it is quite straightforward to approximate them as trapezoidal membership functions (Srinivasan et al, 2001; Shailendra & Lipika, 2005).

The choice of the number of clusters c(1),c(2), … c(p) is a general problem coming with any clustering method. The use of some cluster validity indexes can offer some help. Again, the number of clusters could vary depending upon the size of the information granule of the context triggering this clustering.

EXPERIMENTAL RESULTS

The effects of fuzzy-neural approach were examined by conducting a feature-analysis-type review of the Chinese-language management literature from 2003 to 2004. In this study we illustrate the behavior of the context-sensitive clustering algorithm in application to the auto-feature-abstract data. These data represent a time series consisting of a single input u(k) and the associated state variable x(k).

The data set deals with auto-feature-abstract retrieval. There are three input linguistic variables, these include factors such as introduction factor, keywords factor, conclusion factor. More importantly, by selecting some other linguistic terms, we can model a certain requirement about the problem or a way in which we would like to explore the data and focus the ensuing training of the neural networks. Some other eventual modeling request would be focus the training of the network on the broad range. All of these settings of the linguistic contexts underline a proactive role of the network's designer in the utilization of the data and the customization of the neural network. Discussing the already defined three contexts (weak, medium, strong), the conditional clustering is performed with three clusters allocated per class.

We feel that these variables could be of particular interest as meaningful factors distinguishing between the formulated categories. In fact, these fuzzy sets are far more illuminating than the single numerical values of the prototypes themselves. The findings that one can easily summarize from these fuzzy sets are intuitively appealing. The pace of training of the network (learning rate set in the range [0.5, 0.2] and the momentum set in the range [0.2, 0.1] and changed linearly over the learning period) is summarized in terms of the performance index produced in the successive learning epochs. We also find the nonnumeric results of the fuzzy neural network in relation with the numeric data of the original data set; in most cases the fuzzy output (fuzzy set) either subsumed the numeric data in its core (that is with the membership equal to 1).

CONCLUSION

Mutual Dependence is introduced in this paper. And then based on this notion, a new classifying method for auto-feature-abstract is offered. Later it is applied to classify Chinese-language management documents. It can be used to classify a large number of specimen and has good scalability, while the precision and recall of classification are kept. We have proposed a new architecture of fuzzy neural networks, delivered its complete learning scheme and offered a new type of linguistic computing in the setting of neural networks. The study supports the commonly yet not vigorously experimentally justified claim that fuzzy sets do indeed contribute to the reduction of otherwise immense computational effort. The study shows that the training of fuzzy neural networks with summarized data rather than working with the individual elements of the input–output data cuts the required training time. Obviously, the granularity of the processed information is lower than the one originally residing within the training data. This, in turn, implies that the results of such linguistic processing are genuine fuzzy sets. There is no much sense analyzing the performance of the network at the purely numeric level, as the structure has never been geared towards this particular level of processing. On the other hand, the processing at the linguistic level makes the network transparent and the patterns describing the linguistic data become readily available. Further impact of this methodology on data mining and interpretation of neural networks is quite promising.

ACKNOWLEDGEMENTS

This research is partly supported by National Outstanding Youth Science Foundation of China (Project No. 70525001), National Natural Science Foundation of China (Project No.70371023). The anonymous referees also gave useful comments and suggestions on the content and presentation of the paper.

REFERENCES

[1] Boger Z, Kuflik T, Shapira B, et al. Automatic keyword identification by artificial neural networks compared to manual identification by users of filtering systems [J]. Information Precessing and Management, (37)187-198, 2001.

[2] Boger Z, Guterman H. Knowledge extraction from artificial neural networks models [J]. In Proceedings of the IEEE International Conference on Systems Man and Cybernetics, SMC 3030-3035, 1997.

[3] Charalampidis D, Kasparis T, Georgiopoulos M. Classification of noisy signals using Fuzzy ARTMAP neural networks [J]. IEEE Transactions on Neural Networks, 12(5)1023-1036, 2001.

[4] Dubois D, Prade H. Possibility Theory: An Approach to Computerized Processing of Uncertainty[M]. Plenum Press, New York, 1988.

[5] Freeman J, Skapura D. Neural Networks: Algorithms, Applications and Programming Techniques[M]. Addison-Wesley, Reading, MA, 1991.

[6] Hideyuki T, Isao H. NN-driven fuzzy reasoning[J]. Int. J. Approx. Reason, 5(3)191-212, 1991.

[7] Huang X. A Probabilistic Approach to Chinese Information Retrieval: Theory and Experiments[J]. In Proceedings of the BCS-IRSG, 2000.

[8] Kosko B. Neural Networks and Fuzzy Systems[M]. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[9] Moon B, Jagadish H, Faloutsos C, et al. Analysis of the clustering properties of the Hilbert space-filling curve[J]. IEEE Transactions of Knowledge and Data Engineering 13(1), 2001.

[10] Pedrycz W. Fuzzy set technology in data mining and knowledge discovery[J]. Fuzzy Sets and Systems 3:P279–290, 1998.

[11] Parrado E, Gomez E & Dimitriadis, Y.A. Study of distributed learning as a solution to category proliferation in Fuzzy ARTMAP based neural systems[J]. Neural Networks (16)1039-1057, 2003.

[12] Srinivasan P, Ruiz M E, Kraft D H, et al. Vocabulary mining for information retrieval: Rough sets and fuzzy sets[J]. Information Processing and Management, (37)15-38, 2001.

[13] Shailendra S, Lipika D. A rough-fuzzy document grading system for customized text information retrieval[J]. Information Processing and Management, (41)195-216, 2005.

[14] Tsai, C Y, An iterative feature reduction algorithm for probabilistic neural networks[J]. Omega (28)513-524, 2000.

[15] Wang Jianhui, Hu Yunfa, An algorithm to select features based on mutual dependence, Fudan University, Shanghai, China, Technology Report, No.021011399, 2002.

[16] Zopounidis C, Doumpos M. Multicriteria classification and sorting methods: A literature review[J]. European Journal of Operational Research 138, P229-246, 2002.

Expert Decision Support Systems for Effort Estimation in Software Projects

Madan Kumaraswamy
Assistant Professor
Goa Institute of Management
Goa, INDIA
Email:madan@gim.ac.in

Abstract

Software development effort estimation research has matured beyond Software economics, cost estimation models, and tools. However, for any software organization the generic estimation models are of little use due to their inability to capture all software project dynamics. Research evidence suggests that estimation done by experienced project managers using their knowledge gained through expertise in their professions are closer to the actual effort required for software projects. Expert estimation is superior to parametric and non-parametric model based estimation. This is possible through selection of technologies and models that serve as knowledge repositories for estimation purposes. This paper shows how to capture, select, and preserve estimation knowledge using Expert-Decision Support Systems (EDSS). Besides, this paper demonstrates the ability of EDSS to combine models and expert opinion to arrive at reliable effort estimates and to illustrate how expert knowledge can be collated into an EDSS using CLIPS- a rule-based expert systems shell developed originally by national Aeronautics & Space Agency.

Keywords:

Expert-Decision Support Systems, Effort Estimation, Mean Magnitude of Relative Error, Round-robin reviews, Expert Systems, Rule-based Systems

INTRODUCTION

Software Effort metric is the unit of inputs (resources) for software development measured in Person Months, where one-Person Month is equivalent to 152 hours (Boehm, 1981). Software development is not governed by objective parameters for production of software. The measure of effort required for software development is not absolute and restricted to product and project factors. Human factor productivity is one of the main factors that determine the effort required to accomplish software project completion. While Decision Support Systems (DSS) combine data, and models to interactively solve structured and semi-structured problems, Expert Systems (ES) mimic the decision-making ability of a human expert. An EDSS combines the qualities of DSS and ES to recapture decision scenarios, memory, and provide explanatory power to substantiate decision-making. An EDSS is not meant to scrutinize decisions to judge expert (estimator) prowess but to aid the estimator to answer why, how, and which estimates relate to reasoning, methods, and evaluation of decision choices. Today, Software Organisations have to mature through a continuous process and people management phases, say within the Capability Maturity Model Integrated [CMMi] framework, and People Capability Maturity Model (CMMi, 2000, PCMM, 2001). It is a challenge for organisations to retain expert estimators and capture their knowledge expertise for deciding future estimates. EDSS will help in optimising and managing estimation resources through multi-criteria decision-making.

SOFTWARE EFFORT ESTIMATION

We must address a few basic questions on effort estimation practices to develop a strategy for EDSS deployment. The questions relate to what estimation approaches are applicable to a software development environment. Surveys on software effort estimation show inaccurate estimation techniques, uncertain software requirements, and lack of expert knowledge base as problem areas in effort estimation in practise (Jorgensen, 2004a, Molokken and Jorgensen, 2003, Ruhe et al. 2003). If estimation accuracy is a major concern, selection of effort estimation approach that provides an estimate with the least Mean Magnitude of Relative Error (Conte et al. 1986) must start with estimation model selection and evaluation. An estimation model that does not fit into a development strategy should not be used for estimation. Besides model selection, when to start estimation is a problem area. For example, estimations can be done at the start of requirements gathering (top-down) or done after the prototyping stage of software development (bottom-up). However, estimates are required very early in the project for billing clients, and software project costing.

MODEL EVALUATION AND SELECTION

Parametric and non-parametric models for Software effort estimation approaches are most popular among estimators (Briand and Wieczoreck, 2000a). Software effort estimation is a subset of software cost estimation. Boehm (1981) introduced the Constructive Cost Model [COCOMO] as a relationship between effort and size of a software product (in lines of code). However, the basic COCOMO model has undergone a lot of calibrations to include new effort multiplier parameters (Boehm et al 2000). Models for effort estimation have utilized Oridnary least square Regression methods (Albrecht and Gaffney, 1983), Artificial neural Network [ANN] models (Srinivasan and Fisher, 1995), Fuzzy prediction (Tian and Noore, 2003), and Estimation using project analogies (Shepperd and Schofield, 1997). Research evidence and comparative studies on estimation models report inconsistent results and high magnitude of error in the estimates (Briand et al 1999). It is in this context that estimation based on expert opinion and knowledge of past project estimates is proving to be an efficient estimation technique (Jorgensen, 2004a, Jorgensen, 2004b, Molokken and Jorgensen 2003).

Expert Judgement based estimation

Hybrid models that combine model based estimates and subjective effort estimate judgments by experts are popular due to better accuracy, and explanatory power (Briand et al. 1998). Myrtveit and Stensrud (1999) report that estimators used more than one model for effort estimation provided better estimates than with those who used a single model estimate. Jorgensen (2004a) establishes a case for use of expert judgement over model-based estimates.

It is a challenge for estimators and software project managers to choose an appropriate technique to arrive at an estimate with reasonable accuracy, where mean magnitude of relative error [MMRE] for the estimate is less than 25 percent (Conte et al. 1986). An EDSS for model evaluation will provide project managers with a choice of techniques with verification and validation of the model base. The EDSS development strategy should evolve through a collaborative design strategy.

STRATEGY FOR EDSS DEVELOPMENT

Expert Systems should emulate the decision-making capabilities of a human expert. However, the design and deployment strategy cannot be restricted to the conventional interviewing of experts. The knowledge engineer interacting with experts to gather explicit knowledge can find tremendous opportunities in formal technical reviews and walkthrough sessions. In this process, the knowledge engineer becomes the part of the structured walkthrough of the estimation process. To enhance the knowledge discovery process, the critiques of estimation models are passed through the expert participants (experts) in a Round-robin manner. Round-robin reviews enable the estimators to justify their estimation logic using their expertise. This requires the knowledge engineer to be qualified in effort estimation and expert systems domain. It is important for the estimator to be a neutral observer for an un-biased view. This approach might sound similar to the Delphi approach expert opinions are collated by a moderator (knowledge engineer). The difference between the Delphi approach and Round-Robin Reviews is that the knowledge engineer participates in the review and walkthrough sessions. Figure 1 illustrates the design and development stages of the expert system.

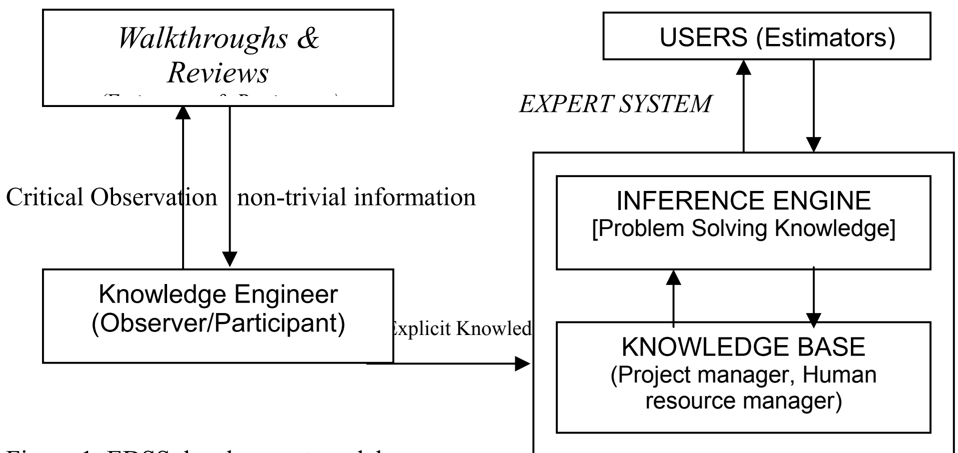


Figure 1. EDSS development model

AN EXAMPLE OF KNOWLEDGE ACQUISITION

Let us assume a situation of a software project management team developing an estimation plan for effort estimation. The estimators have reached an agreement on the software estimation models that will be used for projects as shown in Table 1. During the round-robin review, the estimators agree upon the criterion for model selection and rate them on a five-point scale, where a highly satisfying criterion is rated 5.

Evaluation Criteria	**COCOMO**	**OLS Regression**	**Analogy Estimation**	**Neural Network Model**	**Expert Judgement**
Estimation Availability (late-1 early-5)	2	1	4	3	5
Automation (low-1 high-5)	5	4	3	4	1
Generalisation (low-1 high-5)	4	2	1	2	2
Calibration (low-1 high-5)	3	2	4	1	1
MMRE (high-1 low 5)	3	3	4	2	5

Table 1. Expert Evaluation of Software Estimation techniques

The software project manager, who was present in the review meeting raised some questions of "*what-if*?" and "how" nature. *What-if* all the estimation models do not satisfy the error specification of less than 25 percent? *How* to decide on which technique to choose? The review teams of estimators interpret these questions in the form of rules for evaluation. The estimators establish the first criterion for model selection as Mean Magnitude of Relative Error (MMRE) to be less than 25 percent. A model should be selected if and only if 75 percent of the estimate fall within 25 percent of the actual effort required for projects. The following formula gives the MMRE for N number of observations.

$$MMRE_i = \frac{1}{N_i} \sum \frac{|\text{Actual Value}_i - \text{Predicted Value}_i|}{\text{Actual Value}_i} * 100$$

The software estimation review group can use past project data to obtain analogy-based parameters for estimation. The past project data sets containing actual effort and model estimate data should be used to compare the estimation model effectiveness. A Meta analysis could be done to compare and combine individual studies across different estimation datasets and models. Meta-Knowledge or The knowledge about knowledge will enrich the knowledge base for future estimation needs (Nonaka, 1994). A meta analytic study using statistical validation techniques (paired t-test) on estimation data containing average MMRE scores of a number of effort estimation attempts will give the prediction confidence at different levels of significance. This approach can be effectively used to verify and validate the outputs given by the EDSS estimation technique.

EDSS IMPLEMENTATION

The application of artificial intelligence for software effort estimation is focused on use of Artificial Neural networks Models, and Case Based reasoning tools (Finnie and Wittig, 1996). An evaluation study done on expert systems in software engineering cites the use of bottom-up approach using rule based deductive logic to give better estimates than a static frame based approach (Ramsey and Basili, 1989). Empirical studies on software estimation using AI tools and techniques show inconsistent results and poor generalization of estimation across projects. The reason for poor reception for AI based estimation is the nature of static models that cannot be changed as per user requirements or change in estimation parameters. A rule based expert system can overcome the handicaps of these static AI models and tools to provide flexible alteration of estimation rules and parameters, updating the knowledge base the system acquires with user inputs. On the larger context, this acquired knowledge based on expert judgment could be preserved in an EDSS that combines a rule-base with a knowledge base and inference engine using CLIPS (Software Technology Branch, NASA, 1998). An example of a simple rule in CLIPS is

```
(defrule Rule-1-epidemic-outbreak "Rule for Epidemic-outbreak"
(epidemic dengue)
=>
(printout t "Action 1: Activate fumigation"))
```

In the above *rule, defrule, printout* and *t* are keywords in CLIPS, and the symbol => means then. The pattern epidemic dengue on the left hand side enclosed within parenthesis is the premise and the pattern "Action-1: Activate fumigation" after the symbol => is the conclusion. The above rule can be activated by the fact: (epidemic dengue) and stored as follows.

```
(deffacts epidemic-type  "Facts about epidemics"
(epidemic dengue))
```

When a run command is activated after the CLIPS prompt as

```
CLIPS> (run)
```

The system will output: Action 1: Activate fumigation.

Using the above principles we can store facts about the a project estimation evaluation in two **deffacts** constructs as follows:

```
(deffacts Estimation-Technique
(Project Estimation Code P) )

(deffacts model-evaluation "evaluation of models based on Round-robin review-Table-1"
(P COCOMO                 2    5    4    3    3)
(P OLS-Regression         1    4    2    2    3)
(P Analogy-estimation     4    3    1    4    4)
(P Neural-Network-Model   3    4    2    1    2)
(P Expert-judgment        5    1    2    1    5))
```

The rules that can match the facts can be written as:

```
; *** Rule 1.
(defrule Estimation-Model-Evaluation
(?project ?estimation-technique ?availability ?automation ?generalization ?calibration ?MMRE)
(test (eq ?project P))
=>
(printout t ("P " ?estimation-technique ?availability ?automation ?generalization ?calibration ?MMRE crlf)
(printout t ("Do you want to know the reasons for not selecting other estimation techniques? Answer Y or N"))
```

The EDSS rule base for model selection based on MMRE values is as follows:

```
;***Rule 2.
(defrule  Model-Selection
(?project ?estimation-technique ?availability ?automation ?generalization ?calibration ?MMRE)
(test (< ?MMRE 25))
=>
(printout t (?project  ?estimation-technique  ?MMRE)))
```

The Decision Support System component of EDSS takes the form of another CLIPS rule that computes the Magnitude of Relative Error [MRE] as follows:

```
; *** Rule 3.
```

```
(defrule Magnitude-of Relative-Error
(EstimationModel   Model-Name)
=>
(printout t "Enter the desired MRE, e: ")
(bind ?e (read))
(printout t crlf "Enter the actual estimate, a: ")
(bind ?(read))
(= ?estimate  (* a (1-e)))
(printout t crlf "For desired MRE " ?e "and actual estimate " ?a)
(printout t crlf "The model estimate should be = "  ?estimate crlf))
```

The above rule can be fired by using the following assert statement and making this a part of the right hand side rule that requires value of the estimate to be run:

```
(assert(EstimationModel  Model-Name))
```

The advantage of CLIPS is the flexibility to add, delete, and modify the knowledge base without affecting the system logic, which conventional programming languages donot allow. The other advantage of the rule-based CLIPS EDSS is the design simplicity for the knowledge engineer to convert knowledge into if-then-else type rules and maintain the knowledge base without worrying about system modifications.

CONCLUSION

We learn from the simple case situations how EDSS can help in simplifying and automating effort estimation tasks without sacrificing the learning capabilities. This paper emphasizes the importance of reviews for knowledge capture and transfer. Besides, the knowledge engineer and estimators are present in the EDSS design and development stage. The paper demonstrated the use of public domain software, CLIPS to create a EDSS prototype for effort estimation. The knowledge could be captured and transferred into the system with very little effort. The EDSS implementation can be strengthened with the use of other Expert Systems shells like Visual Prolog, Java Expert System Shell (JESS).

REFERENCES

Albrecht, A. J., and Gaffney, J. E. (1983). Software Function, Source Lines of Code, and Development Effort Prediction: A Software Science Validation. *IEEE Transactions on Software Engineering*, 9(6). 639-648.

Boehm, B. W. (1981). *Software Engineering Economics*. Prentice Hall, Inc., Englewood Cliffs, NJ.

Boehm et al. (2000). *Software Cost Estimation with Cocomo II*, Prentice Hall, NJ.

Briand, C. L., El Emam, K., and Bomarius, F. (1998). COBRA: A hybrid Method for Software Cost Estimation, Benchmarking, and Risk Assessment. *Proceedings of IEEE 20th International Conference on Software Engineering*, April, 390-399.

Briand, C.L., El Emam, K., Surmann, D., Wieczorek, I., andMaxwell, K.D. (1999). An Assessment and Comparison of Common Software Cost Estimation Modeling Techniques. *Proceedings of the 1999 International Conference on Software Engineering*, 16-22 May 1999, 313 – 323.

Briand, C. L., and Wieczorek, I. (2000). Resource Estimation in Software Engineering. Technical Report, International Software engineering Research Institute, ISERN 00-05.

CMMI-SE/SW, V1.02. (2000). CMMI for Systems Engineering/Software Engineering. Version 1.02, Carnegie Mellon-Software Engineering Institute, PA: Pittsburgh, November 2000.

Conte, S.D., H.E. Dunsmore, and V. Y. Shen V.Y. (1986). *Software Engineering Metrics and Models*. Menlo Park, CA: Benjamin-Cummings.

Finnie, G.R., and Wittig, G.E. (1996). AI tools for software development effort estimation. *Proceedings of International Conference on Software Engineering: Education and Practice*, 1996. , 24-27 Jan. 1996,346 – 353.

Jorgensen, M. (2004a). A review of studies on expert estimation of software development effort. *The Journal of Systems and Software*, Elsevier Science, 70 (1-2), 37-60.

Jørgensen., M. (2004b). Top-down and bottom-up expert estimation of software development effort. *Journal of Information and Software Technology*. 46(1): 3 – 16.

Molokken., K. and Jorgensen., M. (2003). A review of software surveys on software effort estimation. *Proceedings of International Symposium on Empirical Software Engineering*, 2003, ISESE 2003., 30 Sept.-1 Oct. 2003, 223 – 230.

Myrtveit, I., and Stensrud., E. (1999). A controlled experiment to assess the benefits of estimating with analogy and regression models. *IEEE Transactions on Software Engineering*, 25(4), 510-525.

Nonaka, I. (1994), A dynamic theory of organizational knowledge creation. Organization Science 5(1), 14-37.

PCMM-Version 2.0. (2001). People Capability Maturity Model. Version 2.0, Carnegie Mellon-Software Engineering Institute, PA: Pittsburgh.

Ramsey, C. L., and Basili., V. R. (1989). An Evaluation of Expert Systems for Software Engineering Management. *IEEE Transactions on Software Engineering*, Volume 15, No. 6, June, 747-759.

Ruhe, M., Jeffrey, R., and Wieczorek, I. (2003). Cost Estimation for Web Applications". *Proceedings of IEEE 25th International Conference on Software Engineering (ICSE'03)*, May 2003.

Shepperd, M., and Schofield., C. (1997). Estimating Software Project Effort Using Analogies*. IEEE Transactions on Software Engineering*, 23(12), 736-743.

Software Technology Branch, NASA. (1998). CLIPS Reference Manual-Volume 1 Basic Programming Guide Version 6.10, August 5th 1998.

Srinivasan, K. and Fisher., D. (1995). Machine Learning Approaches to estimating software development effort*. IEEE Transactions on Software Engineering*, 14, 1743-1757.

Tian, L. and Noore., A. (2003). Multistage Software Estimation. *Proceedings of the 35th Southeastern Symposium on System Theory*, 2003, 232 – 236.

Analysis, Planning and Practice of Trading Agent Competition Supply Chain Management TAC/SCM)

Golenur Begum Huq
University of Western Sydney
Sydney, Australia
E-mail: golenur@cit.uws.edu.au

Abstract

This paper discusses Trading Agent Competition Supply Chain Management (TAC/SCM) practises, analyses with emphasis on component procurement and product market. The focus is on the problems of facing daily operations in uncertain situation and how to improve the situation. In addition, this paper also highlights lessons learnt from history of the game competition. We propose Long-run and Short-run procurement model, which will enhance efficient procurement. This paper also indicates other problems related to purchasing which hinders the development of the supply chain management in the TAC.

Keyword:
Trading agent, Supply Chain Management, Procurement

INTRODUCTION

The Trading Agent Competition (TAC) is an international forum which is designed to promote and encourage high quality research into the trading agent problems. It has been running since 2000, introduced by Wellman and Wurman (Wellman & Wurman, 1999). A new scenario involving supply chain management has been introduced in 2003 by CMU and SICS (Arunachalam et.al., 2004). The Trading Agent Competition Supply Chain Management scenario (TAC, SCM) was designed to pose a complex multi-tiered, multi-period for automated trading agents in a plausible supply chain game (Sadeh & Janson, 2003). The supply chain game presents automated trading agents with challenging decision-problems, including procurement of supplies across multiple periods using multi-attribute negotiations. The procurement process involves substantial uncertainty and competition among multiple agents (Kiekintveld et.al., 2004).

The participating agents in the TAC tournament have to maximise their profit in the strategic environment. Agents always negotiate with their suppliers to procure components and customers to sell their finished products. The agents cannot avoid making commitment i.e., to procure the components before all relevant uncertainty is resolved e.g., customer demand. They predict market demand and took their decision tactfully.

We present the practices of analysis with emphasis on procurement of components and as well as focus on the problems of day to day operations and how to improve the situation of the TAC SCM tournament 2004.

This paper has been structured as follows. At first we present a brief TAC game, then discuss about Production Plan and Performance After that we analyse Procurement Problem, Price Uncertainty and Production Decisions. Discuss about product market performance

TAC SCM GAME OVERVIEW

The TAC, SCM is a contest of six agents who are the manufacturer of personal computer (PC) in a simulated common market economy for the suppliers and end user i.e., customer orders. Each agent can manufacture 16 different types of computers, characterized by different *stock keeping units* (SKUs). Every SKU consists of different combination of components, which are of 10 different types. In each TAC day of the game customers send a set of requests for quotes (RFQs) to the agents. Each RFQ contains a SKU, a quantity, due date, a penalty rate and reserve price- the highest price that the customers are willing to pay. Each agent sends offer mentioning the price less than reserve price to the customer against the received RFQs. The agent wins the bid on the basis of lowest price offered. The winning agent delivers the entire order by due date then it is paid fully and if it is delivered within five days of due date then it incurred less penalty based on the number of differed days. Consequently, if the agent cannot deliver the entire order within five days of the due date then this order is cancelled and the maximum penalty is incurred.

On the other hand, the agents send RFQs to the suppliers for components mentioning quantity and expected delivery date. The suppliers respond to these requests by next day with partial or full offers, specifying the price per unit against received RFQs. A partial offer is made when supplier can not supply whole quantity requested by the agent i.e,. can deliver lesser quantity on that day. Full offers should either have a delivery date on the day requested, or a delivery date later than the requested day. The agent can accept or reject these offers according to their requirements and enter into an agreement with supplier that the agent will be charged for the components on their arrival.

Each agent must solve the daily problems:

i) bidding problems for customer order of PC
ii) procurement problems deals with which components need to be purchased from which supplier
iii) production problems concerns with the every days production scheduling
iv) allocation problems deals with matches SKUs in the inventory to orders.

At the end of the game the agents are awarded based on their profits.

PRODUCTION PLAN AND PERFORMANCE

How can the agent solve its problem of input choice from a number of inputs that would minimize production cost? What would be the best way to solve this problem?

Initially agents need to prepare their hypothetical production plan for the whole game. For this plan the agents firstly need to schedule the type of components (characterize by SKU) they should to order each day, quantity, due date, price of component, initial component inventory and secondly, initial product inventory, production capacity, reserve price, penalty, delivery date of product. And finally they need to find which production schedule will earn more profit. To maximize profit, the agents should decide strategically when, which market segment, what quantity, what price, which delivery date, which supplier to be chosen, what is the inventory condition. To search all these things the agents should schedule their daily operation as balanced ordering of components to maximize the bidding ability of customer orders.

The supplier pricing of components depend on the availability of the day which is most uncertain in the game. Therefore if components are more available then the price of that components are comparatively low otherwise it would offer high price which become a risk to the agents. In this type of uncertain situation, it is very much difficult to take decision whether the agents purchase components or not. For this reason the agents need to adopt machine learning to predict what quantity and type should be purchased.

Although each agent can assemble 2000 cycle per day according to TAC SCM game scenario, Zhang et.al indicates that each agent should produce at least 357 PC (Table 1) per day in equilibrium condition. Because in the competition each agent tries to assemble as many as possible (more than that equilibrium quantity) and finally utilise its full capacity. The average base price of PC is $ 2000. If agents order component on the day

Table 1. The table lists the equilibrium production in case of no more than 6 manufacturers.

Number of Manufacturers	Equilibrium Production	Aggregate Production	**Market** Price
1	2000	2000	2000
2	1000	2000	2000
3	666	1998	2000
4	500	2000	2000
5	417	2085	1830
6	357	2142	1716

Source (Zhang, et.al., 2004)

zero then the average price of PC becomes half of the base price i.e., $1000. For this reason most of the agents order for their component on day 0. The agent can not control this price because of reserve price which is specified by customer. According to TAC SCM specification, customer reserve price was set as 75% to 125% of base price of PC. Therefore, the agents should look only cheap priced components which can lead better bidding. Otherwise agents will be unable to win for bidding.

(3)

(4) Performance of TAC SCM

The agent can predict the total quantity of PC will require to produce for the game from previous game by the following ways :
To obtain the quantity of PC which the agent should produce everyday, we calculate this as backward from game data. If the aggregate market demand of PC is Dm in d days. Then each agent will produce at least q quantity of PC per day in equilibrium condition which is as follows:

$q = \gamma\,((Dm\ /n\,)\,/\,d)$ …………………………………………….(1)

where
γ = constant ($\gamma > 0$)
n = the number of manufacturer.
Dm = market demand
d = day of game

By using formula 1, we find the PC produced (Table 2) by each agent on an average is in equilibrium condition of different round of the competition of TAC SCM, 2004. First column of Table 2 depicts actual quantity won for PC by each agent on an average is in equilibrium condition and consequently second column presents quantity multiplied by γ (assume the value is 0.60). The result defers from the last row of Table 1. The result of semi-final (gr-1) is close but the others are less than the equilibrium condition. Therefore we can conclude that the actual game situation presents different result from the proposal of Table 1. PC production also depends on the situation of game. Consequently, the hypothesis can be proposed as "the agents need to change their procurement strategy according to game situation".

(5) Table 2. **PC produced by each agent on an average in equilibrium condition of different round of competition of TAC SCM, 2004**

Semi-Final Round (GR-1) ((D*m* /n) / d)	Semi-Final Round (GR-1) 0.60 ((D*m* /n) / d)	Semi-Final Round (GR-2) ((D*m* /n) / d)	Semi-Final Round (GR-2) 0.60((D*m* /n) / d)	Final Round ((D*m* /n) / d)	Final Round 0.60((D*m* /n) / d)
221	353	170	273	152	243

We see that the top scored agent usually won the customer order 25-26% of the total PC market demand (Figure:1). How does the agent improve this percentage of wining? Is it the problem of bidding to win customer order or shortage of stock -. probably both? How does the agent solve these problems? Firstly, they needed sufficient stock and then efficient bidding behaviour which led to win the bid. Any how we present economic procurement model which will enhance efficient procurement capacity of the agent.

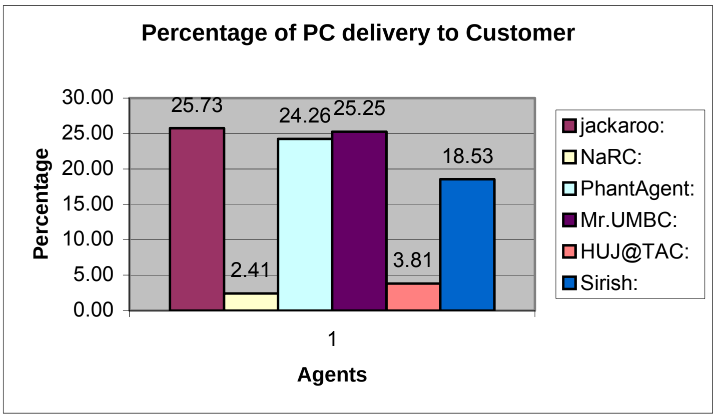


Fig. 1. Percentage of PC delivery to customer of the qualifying round.

Eventually, if the agent produce more PC with low priced on an average and sell those as soon as possible to the customer then the agent can profit more.

PROCUREMENT PROBLEM OF TAC SCM

Purchasing decision is one of the most important vital issue of supply chain management of the Trading Agent Competition (TAC SCM) in which the wining agent of the competition basically depends on this purchasing decision. Consequently, because of supplier component pricing offering process completely depend on the availability of components and at the same time, agents' ordering component for daily operation directly depends on this supplier pricing offer which is also completely uncertain.

Suppliers set prices of component based on an analysis of available capacity. The TAC/SCM component catalog (Arunachalam et al., 2004) associates every component c with a *base price*, b*c*. The correspondence between price and quantity for component supplies is defined by the suppliers' pricing formula.

The price offered by a supplier at day d for an order to be delivered on day d + i is as follows:

$$P(d,d+i) = P_{base}\left[1 - \delta p^* \left[\frac{C_{\text{available}}(d,\ d+i) - \text{qty})}{C_{\text{current}}(\text{d})\ * i}\right]\right]$$

where

- $P\ (d+i)$ is the offer price
- δp is the price discount factor and has a value 50%
- P_{base} is the baseline price of the components
- $C_{\text{current}}\ (d)$ is the suppliers capacity on day d
- $C_{\text{available}}\ (d,d+i) = C_{\text{current}}\ (d) - C_{ordered}(d)$
- *qty* is the quantity requested by the order

In SCM game, the agents purchase the components as direct materials categories i.e., components used to make finished goods (Chopra & Meindl, 2003). This direct materials further classified into bulk purchase, critical and strategic items. We are not discussing strategic items because it doesn't relate to SCM game. Most of the agents purchase their components as bulk on the first day in which suppliers tend to have the same selling price which was 50% less compared to the rest of the game. Some agents like RedAgent (Keller et.al. 2004) and PackaTAC (Dahlgren & Wurman, 2004) also use the critical items which include components with long led times and the key sourcing objective of this critical items is not low price but to ensure availability. Both agents used fairly simple offer acceptance strategies that did not reject offers on the basis of price (TAC SCM, 2003 game). The current (2004) competition differs from last years because of a few enhancements like 1) price function has been modified to better reflect demand; 2) storage costs have been introduced; and 3) customer demand has been segmented into multiple markets. For these changes the agents competed using different strategies from last years competition.

The main objective of the agent is to maximize its profit and minimize its cost of production and inventory. How is it possible? What are the strategies that should be adopted to maximize this profit? There are many factors involve to achieve profit. One of the most important factors is that, the agent needs to increase its productivity with on an average low priced components and another one is that, sell the finished product as soon as possible to clear inventory with an average high priced according to the market situation. How do they optimize these problems? Let us see how the agents perform in the TAC SCM competition.

To increase production capacity and to sell finished product, the agent needs to achieve some improvements: 'auto-activation', just-in-time etc. How does these types of improvements are possible? The agent should need to be tactful in identifying market demand of finished product, calculate how many products should need to be produced and order balanced components for PC from the suppliers and finally, bid customer order for PC.

The agent needs to maintain its competitive edge for as long as possible. This advantage allows them simultaneously to gain market share while also maintaining prices and improving production cost. Competitiveness and productivity are interrelated in the sense that former largely depends on latter. Therefore, it is very important that how many (Cournot model) components will be ordered by agent from supplier and how biding (price competition: Bartrand model) techniques will be applied for the customers order.

Economic procurement model

The main objective of the agent in the TAC game is to procure raw materials from the supplier which always need to be in reduced price and sell them to the customer as soon as possible to reduce inventory cost. We assume that the agent sell the final products efficiently. Therefore, what are the strategies need to be applied which leads the supplier's price all most all the time or in an average is low. We propose a model which will lead efficient procurement as a) Long-run procurement model and b) Short-run procurement model, and this model. will not adjust only the price like classical market model, also focus on the co-ordination of activities is based on confidence and co-operation..

The agents can order components as large quantities which will be delivered throughout the game or separately for small quantities. We consider the first day's component ordering which involves large quantities as Long-run procurement, because the average price of components are comparatively cheap. The Figure 2, Figure: 3, Figure 4 indicate that the agents order their component on average 50% to 100%. There are some risks involve in the long run procurement process, it may also increase total cost because of storage cost. On the other hand, it has an advantage which ensures the availability of resources within required time frame. Therefore long run procurement process ensures the lowest possible cost of producing any given level of output. But it also varies from game to game. For instance the performance of Final round game is completely different from other rounds (we can compare it from the Figure 5 to Figure 2, 3, & 4). By adopting Long-run procurement model, the coordination and confidence will develop among the agents and suppliers which will enhance the Short-run procurement as well.

Average percentage of order on day 0 of Seeding Round

FreeAgent – 51%
Jackaroo – 75%
MrUMBC – 77%
Boticelli – 100%
UMTac-04 – 98%

Figure 2.

Average percentage of order on day 0 of Semi-Final (Gr-1)

FreeAgent: 76%
SouthamptonSCM: 90%
Mr.UMBC: 86%
ScrAgent: 99%
KrokodilAgent: 99%
Socrates: 92%

Figure 3.

Average percentage of order on day 0 of Semi-Final (Gr-2)

jackaroo	92%
Botticelli	78%
UMTac-04	94%
Maxon	100%
deepmaize	80%
PSUTAC	90%

Figure 4.

Average percentage of order on day 0 of Final

FreeAgent	15 %
Mr.UMBC:	31 %
UMTac-04	49 %
SouthamptonSCM	61 %

Figure 5.

In the Semi-final round (GR-2), some agents dropped off the game and other agents received negative results due to extremely high price component and storage cost. In the Final round all agents dropped their component orders on day 0 and also earned many negative results due to the same reason.

The short-run procurement decision for the agent is to find the output level and the associated input levels which achieves the objective it desires, i.e., to maximize profits. In this type of procurement, the agents should either stop purchasing or purchase components to achieve positive levels of output. We assume that the agents will produce a positive output because the market price of PC is above minimum average variable cost. The optimum short-run procurement decision will not be stopped unless market price is below minimum average variable cost. The problem is that the agent always calculate its price of out put at present and probably it will sell in future when the market price of output may be comparatively cheap which will decrease its profit. On the other hand, if the agent wants to increase small amount of output then if marginal cost is above average revenue, which will also decrease of its profit.

For the Short-run procurement model, we can consider the agents order for small quantities of components except first day of the game (Table 3.). Usually we found that the price of the components are expensive than first day which are completely uneven and uncertain. For that reason this short-run procurement becomes risky to the agents in one hand but on the other hand it reduces storage cost.

Table 3. Average price of initial and late order of the component Pintel 2GHz Game 1131 of Semi-Final Round Gr-1

Agent	First day's Order (quantity)	First day's Average Price	Other day's (Late) Ordered	Late Average Price	Average Price of Initial and Late Order
FreeAgent:	11000	532	7	548	532
SouthamptonSCM:	14500	541	2335	681	561
Mr.UMBC:	10643	544	2675	703	576
ScrAgent:	10658	529	1000	658	540
KrokodilAgent:	12000	530	111	1110	536
Socrates:	12663	541	2200	998	608

From the analysis we've found that the supplier pricing formula provides a strong incentive to purchase large quantities of components on the first day of the competition. Although it has been set up the inventory cost to reduce first day's component ordering. We have also found that the price of components on the first day was almost half or more than half than all the other days. On the other hand, sometimes it causes dramatic delay due to huge order. As a result, in TAC/SCM-04 agents employed increasingly 'aggressive' first day procurement policies like TAC/SCM'03, leading to a mutually destructive overcapacity of components for the aggregate system (Daniel et.al 2003) .

Price Uncertainty and Production Decisions

Price uncertainty of component market is the most important issue in TAC SCM. How do the agents purchase components in this environment? What is the best strategy to solve this problem? This price uncertainty may result a vulnerable situation for the agent in the game in which the agent will earn low profits or negative scores.
We can assume that the agent's total costs are $C(Q)$, and the price can be either $(1 + a)P$ or $(1 - a)P$ with equal probability, $0 \le a \le 1$. The expected profit will

$$E[\pi] = \tfrac{1}{2}(1 + a)PQ + \tfrac{1}{2}(1 - a)PQ - C(Q) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where Q is quantity of finished product and P and a are given. Increasing a increases price volatility, while $a = 0$ corresponds to a riskless setting. The agent always seeks to maximize its expected profits and hence it must make its production decision before price is known. If decisions can be postponed until after the random event occurs, then substantial reduction in profit risk can be achieved, but the essence of a risky planning problem is removed. How does this uncertainty influence price, quantity demanded or cost? What is the objective of the agent? Each of these questions must be answered in a situation of more complicated problems. Maximising expected profits requires

$$\frac{\mathrm{d}\,E[\pi]}{\mathrm{d}Q} = \tfrac{1}{2}(1 + a)PQ + \tfrac{1}{2}(1 - a)PQ - C'(Q) = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where $C'(Q)$ is marginal cost and the first two terms of the right-hand side of the equation (4) are expected selling price. If the expected price is equal to P, the condition in equation (4) is satisfied at the same level of output as in a riskless setting. If the agent commits itself to this level, it will be unable to expand output to take advantage of the higher price of $(1 + a)P$, and will produce in excess of demand when the price is $(1 - a)P$. Therefore, expected profits will be less than the profit that could be achieved under certainty. In addition, as the variation in price increases (i.e., as a increases), expected profits decline, thus creating an economic value for information, which would reduce the degree of uncertainty. On the other hand, if we introduce Rubinstein's *alternate offering* model (Ariel, 1982) then it will reduce degree of risk of the agent.

Although a detailed development is beyond our scope, the major implications of price uncertainty for a purely competitive agent may be summarized as :

1. For risk averse agent, optimal output is less than under certainty.
2. Increased fixed (storage) costs will normally lead to a decrease in optimal output.
3. Increasing marginal cost is no longer a necessary condition for a finite optimal production level
4. For a risk averse agent, an optimal production plan is characterised by *average total cost* less than expected price.

Fixed cost become relevant when risk is introduced, since the agent will want to provide a margin for possible losses by keeping fixed cost down. Last point indicates that expected economic profits will be positive, not zero, in equilibrium; in fact, they will increase as the agent assumes more risk. This result corresponds to the common assumption that risk taking is a valuable economic function which should command a positive market price.

PRODUCT MARKET PERFORMANCE

As we know, a pure competitor or monopolist can simply choose its price or output policy and directly calculate the resulting gain or loss. In an oligopoly market setting, the choice of a price, output, or other marketing policy does not uniquely determine profit, because the outcome for each firm depends on what its opponents decide to do. The Cournot and Chamberlin descriptions of oligopoly suggest the kind of interdependence that arises explicitly here, but do not take into account uncertainty about opponent's decisions ((Meyer, 1976) .

The market price of PCs for all the agents depend on the quantity they produced. This means that, the profit for each agent is linked directly to the profit of the other. Consequently different agent has its own cost functions which imply different payments for inputs. Therefore each agent has its own policy to bid customer order which will enhance to win the bid.

PC market is another vital part of TAC SCM in which agents are directly involved in winning. In the competition, we find the following critical questions to resolve or improve the agents performance as price competition.

i) How does the agent bid for customer reserve price for PC.
ii) What are the strategies need to be adopted for this?
iii) How much agent should need to be reduced the price to win the bid?

To improve the performance of the agent, it is necessary to learn from the history of game. As for example, Figure 6 presents the average price of PC of the competition. The agents can learn from the chart that when the market price of PCs are high and medium, and low. Equilibrium prices arise when supply equals demand: $Q^s_i = Q^d_i$ for product *i.* If $Q^s_i \geq Q^d_i$, agents will bid price P_i lower ; if $Q^s_i \leq Q^d_i$, agents will bid price P_i higher. Usually the price of product increases at the beginning of game due to lack of supplies. Therefore the agent who supply the product at the time of low market supply in higher price and as a result that agent can earn more shares with more profits. Consequently, the agent who can adopt this strategy i.e., increase productivity and bid according to market situation then the agent will have better opportunity to maximize its profit.

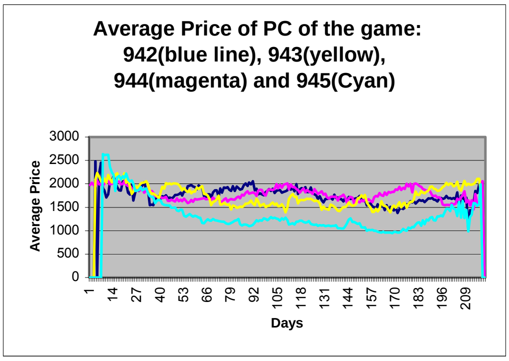


Figure 6. Market price of PC of the game 942 - 945

The average market demand of PC of Semi-Final and Final round game can be depicted in the Table 4, 5, and 6 where the second column is average PC delivered by the agents, third column is the total average market demand. Then we find the FreeAgent bids on an average with high average price and higher percentage of order.

Table 4. Average Total PC delivered by agents of Semi-Final 0f Gr-1 (Tac3 and Tac4)

Agents	Av delivey	Total AV	% of Order	Av Price
FreeAgent:	46882	291505	16	1656
SouthamptonSCM:	61759		21	1508
Mr.UMBC:	51587		18	1527
ScrAgent:	40995		14	1504
KrokodilAgent:	41551		14	1538
Socrates:	48732		17	1323

Table 5. Average Total PC delivered by agents of Semi-Final of Gr-2 (Tac5 & Tac6)

Agents	Av delivery	Total AV	% of Order	Av Price
jackaroo:	35933	224546	16	1595
Botticelli:	38821		17	1714
UMTac-04	41332		18	1455
Maxon:	45631		20	1544
deepmaize:	33780		15	

Table 6. Average Total PC Ordered by agents of Final Round

Agents	AV PC delivey	Total AV Market demand	% of Order	Av Price
FreeAgent:	41659	201227	21	1842
•Mr.UMBC:	45465		23	1670
•UMTac-04	44665			22
1481				
•Botticelli:	13765			7
1434				
•deepmaize:	24487			12

CONCLUSIONS AND FUTURE RESEARCH

We propose Long-run and Short-run procurement models, which will enhance efficient procurement and by adopting Long-run procurement model, the coordination and confidence will develop among the agents and suppliers which will enhance the Short-run procurement as well. These models provide a simple and powerful mechanism for judging the probable impacts of alternative policies or decisions. The price uncertainty of component market in a purely competitive setting emphasized the gross errors- both theoretical and practical – that can result from ignoring risk elements. If we introduce Rubinstein's *alternate offering* model (Ariel, 1982) then it might reduce degree of risk of the agent. The agent can improve its bidding performance by predicting market price of PC from game history.

The following works can also be done in future: a) Introduce new negotiation mechanism to the TAC/SCM to cope with the first day purchasing behaviour.

b) Extend the current negotiation formate (component buying) from existing three ways to five ways (introducing bargaining) communication.

References

Ariel, R. (1982). Perfect equillibrium in abargaining model. *Economitrica, 50 (1)*, 97 -109.

Arunachalam, R., Sadeh, N., Eriksson, J., Finne, N., and,, & Janson, S. (2004). *The Supply Chain Management Game for the Trading Agent Competiton 2004*: ISRI Technical Report CMU-ISRI-04-104.

Chopra, S., and, & Meindl, P. (2003). *Supply Chain Management: Strategy, Planning, and Operation* (2nd ed.).

Dahlgren, E. a., & Wurman, P. R. (2004). PackaTAC: A Conservative Trading Agent. *ACM SIGecom Exchange,, 4 no. 3.2*, 33-40.

Daniel Michael, Peter Naude, and, R. S., & Valla, J.-P. (2003). *Business-to-Business Marketing: Strategies and Implementation*: PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth avenue, New York, N.Y. 10010.

Keller, P. W., Duguay, F.-O. a., & Precup, D. (2004). RedAgent - winner of TAC SCM 2003. *ACM SIGecom Exchanges, 4, No. 3. 02*, 1-8.

Kiekintveld, C., Wellman, M. P., Singh, S., & Soni, V. (2004). Value-Driven Procurement in the TAC Supply Chain Game. *ACM SIGecom Exchange,, 4, No. 3*(February), 9-18.

Meyer, R. A. (1976). *Microeconomic Decisions*: Houghton Mifflin Company.

Sadeh, N., Arunachalam, , R., E., J., F., N., and , & Janson, S. (2003). TAC-03: A supply chain trading competition. *AI Magazine, 24,* 92-94.

Terpsidis, I., Moukas, A., Pergioudakis, B., & P, M. (1997). *The Potential of Electronic Commerce in Re-Engineering Consumer-Retail Relationships through Intelligent Agents*: IOS Press.

Wellman, M. P., & Wurman, P. R. (1999). *A trading agent competition for the research community.* Paper presented at the IJCAI-99 Workshop on Agent-Mediated Electronic Trading, Stockholm.

Zhang, D., Zhao, K., Liang, C.-M., Huq, G. B., & Haung, T.-H. (2004). Strategic Trading Agents vial Market Modelling. *ACM SIGecom Exchange,, 4 No. 3*, 46-55.

An Object-oriented Framework for Individual's Policies using Dynamic, Distributed Workflows

Ms. Indira Meshram
Indian Institute of Technology
Delhi, India
Email: indira@cse.iitd.ernet.in

Prof. S. K. Gupta
Indian Institute of Technology
Delhi, India
Email: skg@cse.iitd.ernet.in

Mr. Vikram Goyal
Indian Institute of Technology
Delhi, India
Email: vkgoyal@cse.iitd.ernet.in

Mr. Anand Gupta
Netaji Subhas Insitute of Technology
Delhi, India
Email: omaranand@cse.iitd.ernet.in

Abstract

The object-oriented paradigm has been widely accepted in software development. A similar approach may be applied for policy definition and implementation. In this case the 'objects' will be components of a policy such as actors, events and actions. It is suggested that development of such an object-based framework will be particularly useful in policy maintenance and will also simplify the process of policy definition. Perhaps the greatest advantage of such a framework is for individual policies. Unlike organisation-wide policies which are universally applicable, individual policies are meant to give flexibility to the users or customers to make their own choices. With such an object-based setup, the choices made by the user will be easily transferable to a policy that can also be deployed and enforced on-the-fly because it will simply mean picking up the relevant components and composing them in the required order. A generic framework for the entire policy lifecycle is proposed, such that it will be particularly useful for individual policies.

Keywords:

Individual policy, policy specification, policy enforcement, workflows, object-oriented

1. INTRODUCTION

Every organisation works under a set of policies. These policies are a set of rules which define how the organisation functions. Thus, each company has a set of policies for its internal functions and also a set of policies regarding how they interact with their customers or other companies. However, in today's competitive market, there is an emphasis on personalization. A lot of companies want to provide their customers with the flexibility to define their own way of dealing with the company, i.e. form their own policy. For e.g. a health insurance company may allow the customer to define his own privacy policy, a bank may allow their customers to have their own set of functions at the ATM, a publishing house may give the flexibility to the customers to decide on the frequency of their subscriptions and so on. This brings in a dynamic aspect to the policies.

There has been a lot of recent research on policy specification (Alaettinoglu et al ,1999; Damianou, Dulay, Lupu and Sloman, 2001), but not enough on policy enforcement. A generic framework for the entire policy lifecycle is proposed, beginning with specification and ending with enforcement, such that it will be particularly useful for individual policies. Unlike an organisation-wide policy, in case of individual policy, the policy cannot be hard-coded beforehand. However, since it must fall within the scope decided by the company, it is possible to define the set of choices offered to the individuals beforehand. This means that the components of the policy can be predefined but their composition must be done at run-time. For such a scenario, the object-oriented approach is most suitable (de Champeaux et al, 1993; Lea, 1994).

There is a close relationship between policies and workflow, often a policy needs to define the flow of the process. Also, any policy can be enforced using a workflow. For our purposes we require a dynamic workflow, so that the policy can be enforced on-the-fly. Also a distributed workflow is more natural for most organisations since policies often span several departments or even external agencies. A distributed architecture allows them to maintain their autonomy and has the advantages of robustness and scalability.

The framework has 4 phases: object definition, policy definition, policy deployment and policy enforcement. Object definition and policy definition can be done using a web-based tool, policy deployment requires transferring the set of rules defined using the tool to the relevant actors and policy enforcement is done using a dynamic, distributed agent-based workflow.

The remaining part of the paper is organized as follows. The next section looks at related work, sections 3 - 6 discuss each of the 4 phases, section 7 gives the framework architecture, section 8 describes an example application and the final section gives the conclusion.

2. RELATED WORK

Policy definition and implementation in databases is one of the active areas of research. There are several policy specifications languages, such as Routing Policy Specification Language (Alaettinoglu et al, 1999) and Ponder, the policy specification language for security (Damianou et al, 2001). However, these are specific to a particular domain. There is ongoing work in the area of a generic policy specification language, such as Policy RuleML ("The Policy RuleML Technical Group" [RuleML], 2004; Ross-Talbot et al, n.d.) and Rewerse WG I2 ("Rewerse WG I2" [Rewerse], 2004), but these are yet to be standardized. It is clear that specification of a policy basically boils down to specification of rules (RuleML, 2004; Ross-Talbot et al, n.d.; Rewerse, 2004). We have exploited the same in our framework.

For policy enforcement, we choose to use a dynamic, distributed workflow architecture. There are several dynamic workflow models (Meng et al, 2002; Grefen et al, 2000). However, in these the focus is on enabling different organisations to team up and operate as a virtual enterprise. It is dynamic in the sense the various actors of the workflow are not fixed and the workflow can be built on-the-fly between organisations as and when required. There is an emphasis on service discovery and dynamically binding e-service requests to e-service providers, and the need to adapt to dynamic business conditions and exception situations. For our purposes we are consider an organisation whose functions are known beforehand and various components such as the actors, the rules, etc. are also all identified. The dynamicity comes into play because each individual may have his own policy specification which is translated into his own workflow specification. So we need the workflow to be designed in such a way that it should be possible to have several definitions of the workflow, each of which can be specified as well as enacted on-the-fly. Apart from being dynamic, a distributed model is highly desirable. Webwork is one such architecture (Miller et al, 1998). For our framework we have built upon the architecture proposed by Singh and Tomlinson (1994) which suggests use of algebra of events for representing and reasoning about intertask dependencies (Singh and Tomlinson, 1994; Attie et al, 1993; Attie et al, 1996; Singh, 1996; Wan et al, 1999). This is a generic architecture which can be used in several domains to enforce the policies.

The flexibility required for individual's policies has been achieved by breaking up the policy into components or objects. A similar approach has been suggested for business processes by Orriëns et al (2003).

3. OBJECT DEFINITION

This phase requires that the organisation maps its various functions into modules. Each such module will be the responsibility of an 'actor'. The actual functionality of these modules will be the 'actions'. Defining when these actions are to be performed gives us the 'conditions' which could either be 'constraints' of the form a < 100, or 'messages' from one actor to the other. Thus a condition is a combination of such constraints and messages.

The final and most important object is the 'rule' which describes which action should be done by whom and when. There are also rules defining who should be informed on completion of a particular action. These rules help to capture the flow. A policy is then a combination of such rules. The object definition phase is limited to defining actors, actions, constraints, messages and conditions. Rules and policies are defined in the policy definition phase.

A brief (but not complete) description is as follows.

<digit> := **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9**

<number> := <digit> | <digit> <number>

<actor_id> := **A**<number>

<actor> := <actor_id> <actor_body>

<actors> := <actor>

<actors> := <actor>, <actor>

<action_id> := **AC**<number>

<action> := <action_id> <action_body>

<message_id> := **M**<number>

<message> := <message_id> <message_body>

<constraint_id> := **CS**<number>

<constraint> := <constraint_id> <constraint_body>

<condition> := <message>

<condition> := <constraint>

<condition> := <condition> <op> <condition>

<op> := **AND | OR | NOT**

<rule> := **On condition** <condition> **perform action** <action>

<rule> := **Action** <action> **gives output trigger** <message>

<rule> := **Output trigger** <message> **should be sent to** <actors>

<policy> := <rule>

<policy> := <rule>, <rule>

The components have the following attributes.

Actor

Actor is an entity capable of carrying out a task. A task is an atomic action. For e.g. in a bank's ATM machine, on withdrawal of money, updating of the customer's account is one task, printing the transaction slip is another, dispensing the cash is yet another and so on. The enforcement of policy involves execution of several such tasks when certain conditions are fulfilled. An actor has the following features

i) Id – Unique identifier

ii) Name – Unique name across the system which identifies the actor to other actors

iii) Description – General description of the actor's functionality

iv) Location – Required in case it is static and needs to be communicated to other actors desiring to communicate with it. This is redundant in case there is a directory server keeping track of actors, as is the usual practice.

v) Rule File – The rule file is the interface between policy definition and policy implementation, since the policy definition is saved in the rule file and the actor accesses this file to fetch and implement the policy. Again this could be redundant in case there is some other mechanism to define the rule, e.g. each user can have its rule file name of the form <user>_<actor>

Message

Message or trigger is synonymous with an event. It is used to inform actors that an event has occurred. A message has the following features.

i) Id – Unique identifier

ii) Name – Unique name across the system which is used for communication among actors

iii) Type – This may be implementation-specific.

iv) Content – It describes the parameters which may need to be conveyed regarding the event.

v) Receivers – It refers to the actors who will receive this message. These may be defined statically but can be easily overridden in case a flow is defined as part of the policy which requires this message to be received by a different actor than the one defined here.

vi) Senders – It refers to the actors who will send this message. These may be defined statically but can be easily overridden in case a flow is defined as part of the policy which requires this message to be send to a different actor than the one defined here.

Constraint

Constraint can be used to define predicates of the form x > y. A constraint has the following features

i) Id – Unique identifier

ii) Name – Unique name across the system which is used to form conditions

iii) Body – The predicate referenced by this constraint

iv) Parameters – Inputs required by the predicate

Condition

Condition is the 'if' part of a rule. Condition may include 2 kinds of components – messages and constraints. This means that a condition may check triggering of an event or may check constraint satisfaction. These checks can be combined using AND, OR, NOT, etc. to form the condition. A condition has the following features

i) Id – Unique identifier

ii) Name – Unique name across the system which is used by a rule to define the 'if' part

iii) Body – Contents of the condition which is composed of messages and constraints combined using AND, OR, etc.

Action

Action is the task that needs to be performed when a certain condition is satisfied. These are defined as part of the rule, and the actual implementation occurs at the actor associated with the rule. An action has the following features

i) Id – Unique identifier

ii) Name – Unique name across the system which is used by actors to identify which action needs to be performed.

iii) Body – Defines the body of the action. This may be free flow description, in case it is to be used as a guideline for the programmer responsible for developing the action. Alternatively, it could contain actual code, which can be dynamically fetched and executed by the actor at run-time.

iv) Parameters – Defines the parameters passed to the action. This may be free flow description, in case it is to be used as a guideline for the programmer responsible for developing the action. Alternatively, it could contain the actual parameter names, which can be dynamically fetched by the actor at run-time.

v) Description – Description of what the action does.

4. POLICY DEFINITION

Policy definition involves defining a set of rules which together form the policy. Rule is the main component that ties the whole framework together. A rule has the following features.

i) Id – Unique identifier

ii) Name – Name is unique within the set of rules belonging to a particular actor and a particular user, it need not be unique across the system.

iii) Type – Type may be 'normal', 'action-rule', 'output-message' and 'output-actor'.

iv) Body – The contents of the rule. This may be a free flow description

v) Syntax – It contains the rule in the syntax of the rule engine.

vi) Actor – The actor to whom the rule belongs

vii) User – This identifies the individual who has defined the rule

Rules can be of many types, such as permission and prohibition rules, duty assignment rules, and empowerment rules (RuleML, 2004). We describe few additional rule types.

a) Normal – This rule is of the following form.

On condition <condition> perform action <action>

where <condition> and <action> refer to the condition and action components previously defined in the system. This rule captures the usual if-then construct required in most situations. When <condition> is satisfied for the actor, it should perform <action>

b) Output-Message – This rule is of the following form.

Action <action> gives output trigger <message>

where <action> and <message> refer to the action and message components previously defined in the system. This rule construct, usually in conjunction with the Output-actor construct, helps to capture flow. It defines what message should be sent out when a particular action completes.

c) Output-Actor – This rule is of the following form

Output trigger <message> should be sent to actors <actor>

where <message> and <actor> refer to the message and actor components previously mentioned in the system. This is the main construct used to capture flow. It defines which actor to send the specified message to.

The above rules are used to capture flow in this way

1. A rule is constructed to define what action should be taken on an input trigger (rule type: normal)
2. A second rule is defined to state what message should be sent out when the action is complete (rule type: output-message)
3. Finally, a rule is defined to state which actor is the recipient of the message defined in the previous rule (rule type: output-actor)

After the rules are in place, policies can be defined. A policy has the following features.

i) Id – Unique identifier

ii) Name – Name must be unique within the set of policies defined for a particular user.

iii) Type – This may be domain dependent

iv) Body – It enlists the ids/names of the rules which form the policy.

v) User – It identifies the user who has defined the policy

Once the objects have been created in the Object Definition phase, the Policy Definition phase can be done using a tool through which all the components are visible. Such a tool can then be used by the policy designer to define the policy. However, in case of Individual Policy, it is not advisable to provide access to such a tool to the users, since typically an organisation wants to have some kind of control on the policies defined by the users; all the permutations and combinations will not be available to them. So another layer can be developed, in which the choices made by the user using checkboxes or radio buttons are translated into rules.

5. POLICY DEPLOYMENT

Once the policies have been defined, the deployment phase involves translating the rules of the policies into the syntax of the rule engine being used, and appending them to the rule file of the particular actor.

6. POLICY ENFORCEMENT

For policy enforcement the architecture we suggest is that of a dynamic, distributed workflow using agents interacting with a rule engine. In fact the rules definition has been designed keeping this architecture in mind, since there are rule types which capture the flow through use of messages passed between agents. The agents are the actors (these may even be human agents). The onus is on the designer to first determine the intertask dependencies and ensure that the rules are formed accordingly. For e.g. At actor a1 if event e1 cannot take place unless event e2 has already take place, then a suitable rule should be defined at actor a1 which forces it to wait for event e2, before executing event e1. It is through the use of such intertask dependencies, maintained at each actor that we do away with the need of a central scheduler. The agents interact amongst themselves using messages.

When an event occurs, the agent should interact with the rule engine using the set of rules from the relevant rule file (i.e. the rule file belonging to the particular user). If the conditions are satisfied, a rule is fired, and the agent must carry out the required action. On completion of the action the agent must again check using the rule file whether the result of the action has to be communicated to another agent and act accordingly.

This architecture gives us domain independence since it can be plugged in any domain. The functionality (the 'action' part) can be programmed as subroutines of the agents and the policy will be enforced in form of rules of the rule engine.

7. ARCHITECTURE

The architecture is as described in figure 1. The various modules are as follows.

Object Definition Tool

This is used to define the various objects such as actor, action and condition. Typically only the administrator has access to this tool. The objects are stored in a central repository.

Policy Definition Tool

This is used to define the rules and policies using a combination of the objects defined earlier.

Interface

Users are typically not allowed all permutations and combinations when defining their own policy. They have to work under a restricted set of choices. This restriction is enforced using an interface. The interface is also required to make the process of defining the policy more user-friendly. For e.g. a user may be able to choose whether he wants his pay slip printed or send by email. In the policy definition tool he will have to create a rule of the form *on condition pay_slip_available do action print*. Instead he can be provided with an interface which allows him to determine his choice by checking the specific checkbox.

Rule Interpreter

It translates the rules created in the policy definition tool to the rule engine's syntax. Further these rules are written in a file where they can be accessed by the associated actor when interacting with the rule engine for the particular user's session.

Rule Engine

The rule engine determines whether a particular rule has been fired and gives the resulting action. The rule engine also saves the 'state', in the sense it remembers which facts have been asserted so far.

Agent

Agents are the entities capable of executing the actions. On receiving a particular trigger, they check the user's session, fetch the user's rule file, assert the fact corresponding to the trigger and interact with the rule engine to determine whether a rule has been fired. If yes, then the agent carries out the required task. After the task is complete it again interacts with the rule engine to determine whether the task completion should result in sending out a message and if yes, then to whom. Typically each agent runs its own instances of the rule engine.

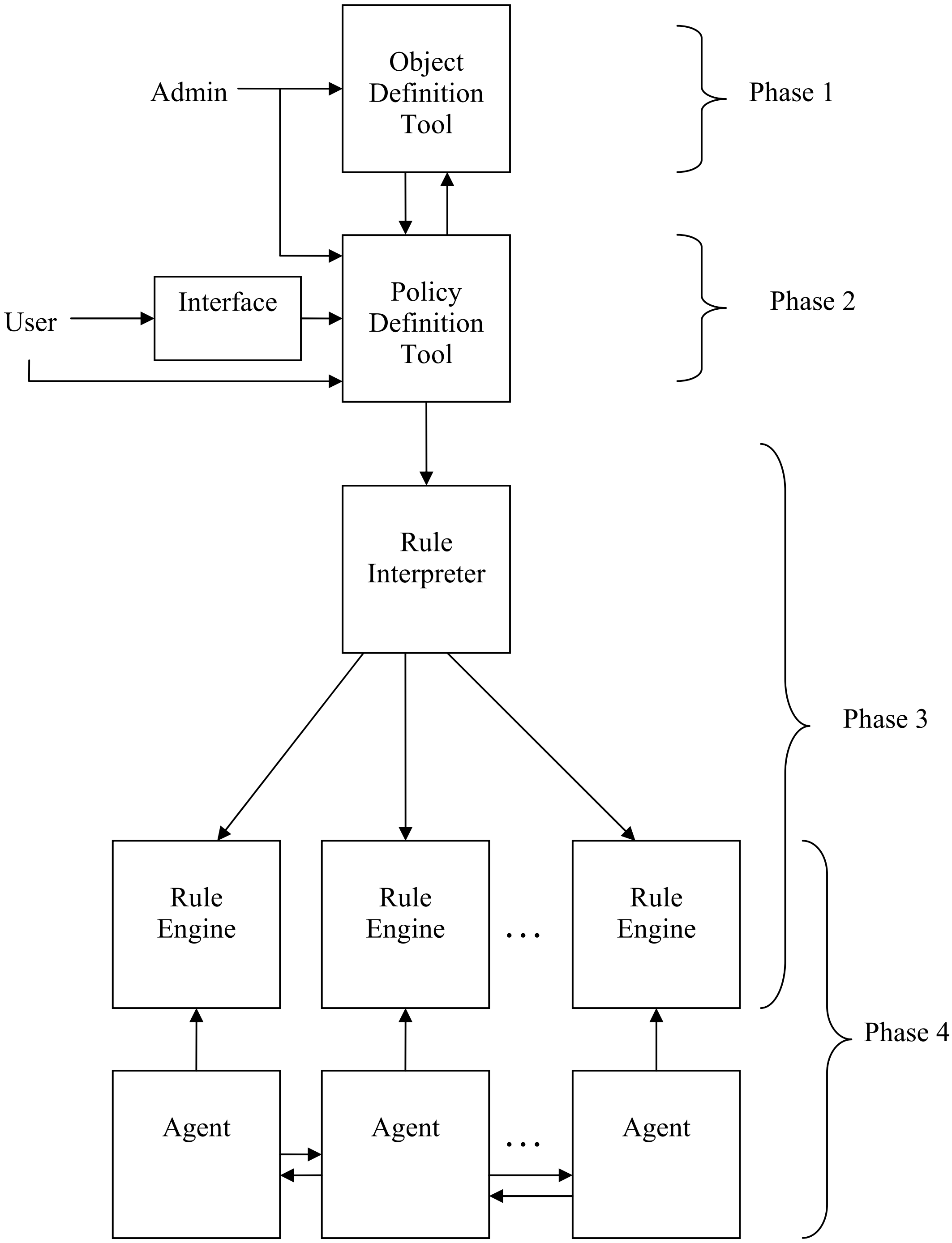


Figure 1. Architecture

8. EXAMPLE APPLICATION

Consider the privacy policy of a hospital. Suppose this hospital has an understanding with some pharmaceutical companies to share information about its patients, with the consent of the patient. So the privacy policy then becomes whether or not the information should be shared. This could be extended to any level, depending on the choices one wants to offer the customers, e.g. they could decide what portion of the information to share, or which companies would be allowed to view the information, whether the information is to be used for statistical purposes or active marketing and so on.

To implement the individual's privacy policy, consider 3 modules – information store, information filter and information distributor. Initially when a patient is admitted to the hospital, he is allowed to decide on his policy. The information store has all his details, information filter selects information which the customer has authorized for distribution (if any), and information distributor is responsible for sending this information to the companies selected by the patient.

The objects defined will be as follows. (All the features are not described, in most cases only the name is mentioned which is indicative of the function of the object, the remaining features are implementation-specific)

Actors

i) information_store

ii) information_filter

iii) information_distributor

iv) There will also be external agents to which information_distributor will communicate such as dr_reddy_actor, hll_actor, xyz_actor, etc.

Each will have the required functionality built in according to the actions it is required to do, e.g. information_store will have modules to store information in the database corresponding to the action store_information.

Messages

i) new_information

Constraints

None are required in the current scenario

Conditions

i) new_information - Its body will simply include the message new_infomation)

Actions

i) store_information - At the actor information_store, this will be a subroutine responsible for storing the new information received in the database

ii) complete_information – At the actor information_filter, this will be a subroutine responsible for fetching the complete information of the user from the database and packing it in a message to be sent to the next actor.

iii) recent_history – At the actor information_filter, this will be a subroutine responsible for fetching the recent medical history of the user from the database and packing it in a message to be sent to the next actor.

iv) personal_information – At the actor information_filter, this will be a subroutine responsible for fetching the personal details of the user from the database and packing it in a message to be sent to the next actor.

v) send_forward – At the actor information_distributor, this will be subroutine which packs the information sent into a message with additional headers or information if required. This message will then be sent to the external actors.

Rules

Rules will be as follows (to describe the rules we use the format rule_name: rule_body).

Rules for information_store actor will be as follows.

- Rule_1: On condition new_information perform action store_information
- Rule_2: Action store_information gives output trigger new_information

- Rule_3: Output trigger new_information should be sent to actors information_filter

Rules for information_filter actor will be of the form

- Rule_1: On condition new_information perform action complete_information
- Rule_2: Action complete_information gives output trigger new_information
- Rule_3: On condition new_information perform action recent_history
- Rule_4: Action recent_history gives output trigger new_information
- Rule_5: On condition new_information perform action personal_information
- Rule_6: Action personal_information gives output trigger new_information
- Rule_7: Output trigger new_information should be sent to actors information_distributor

Rules for information_distributor actor will be of the form

- Rule_1: On condition new_information perform action send_forward
- Rule_2: Action send_forward gives output trigger new_information
- Rule_3: Output trigger new_information should be sent to actors dr_reddy_actor, hll_actor, xyz_actor
- Rule_4: Output trigger new_information should be sent to actors dr_reddy_actor
- Rule_5: Output trigger new_information should be sent to actors xyz_actor

Individual Policy

Consider a patient, Rahul Sharma, who has fills up an online form allowing the hospital to send only his recent medical history to Dr. Reddy's.

His policy will be deployed by creating a new rule file for him at each of the 3 actors.

1. At information_store his rule file will contain Rule_1, Rule_2 and Rule_3.

2. At information_filter his rule file will contain Rule_3, Rule_4 and Rule_7.

3. At information_distributor his rule file will contain Rule_1, Rule_2 and Rule_4.

This can be done as soon as the online form is submitted, and the policy is immediately enforced. Tomorrow, even if Mr. Sharma wants to change any of his preferences, it will only involve changing the relevant rule in the relevant rule file. The change is also immediate if allowed online. In this way, maintenance of the policy is also effortless.

9. CONCLUSION AND FUTURE WORK

We have developed a prototype for the above suggested framework. Web-based tools were developed using Perl CGI for object definition, policy definition and user interface. JADE (Java Agent DEvelopment) toolkit for agents with Jess as the rule engine was used to develop the policy enforcement architecture ("JADE", n.d.; "Jess", 2005). We have not defined a formal language for policy specification. The definition phase is intentionally implementation-aware. Each of the objects defined can be mapped to objects in the implementation, e.g. the actors are mapped to the agents, the actions are mapped to subroutines in the agents, the messages are mapped to the FIPA compliant messages passed between agents to indicate triggering of a particular event, and the rules are mapped to the rules to be defined in the rule file, and interpreted by the rule engine. Of course our definition tool is user-friendly and allows these to be defined in plain English; the interpreter converts them to the required syntax and finally in the deployment phase the mapping is complete. The advantage is that with this framework in place, we can define policies which can be immediately enforced.

The framework is found to be most suitable for defining individual policies. Furthermore, it has the advantages of scalability and ease of maintenance. However, as is the case with most generic architectures, it may not work for all business environments as it is, e.g. some additional rule types may need to be added, such as the ones identified in Rewerse (2004) and RuleML (2004). Again since it is object-oriented, extensibility is easy.

As is the case with individual policies, they should not violate the organisation-wide policy. In case the individual is allowed to define this policy through the use of an interface, this violation can be prevented, since the user is given only restricted choices. However, if he is allowed to compose his policy using all the existing components, such violations need to be identified and prevented. The precedence of policies also needs to be taken into account, i.e. in case of a hierarchy of policies, a policy defined at a certain level must conform to the policies defined at the higher levels. These issues are however, beyond the scope of the current work.

REFERENCES

[1] Alaettinoglu, C., Villamizar, C., Gerich, E., Kessens, D., Meyer, D., Bates, T., et al (1999, June). RFC 2622 - Routing Policy Specification Langauge (RPSL). *Internet FAQ Archives*. Retrieved September 22, 2005 from http://www.faqs.org/rfcs/rfc2622.html

[2] Attie, P. C., Singh, M. P., Sheth, A., Rusinkiewicz, M. (1993). Specifying and Enforcing Intertask Dependencies. *Proceedings of the 19th Conference on Very Large Data Bases*. Dublin, Ireland: Morgan Kaufman.

[3] Attie, P. C., Singh, M. P., Emerson, E. A., Sheth, A., Rusinkiewicz, M. (1996, December) Scheduling Workflows by Enforcing Intertask Dependencies. Distributed *Systems Engineering Journal, 3(4)* (pp. 222-238).

[4] Chakravarthy, S., Mishra, D. (1994, Nov). Snoop: An Expressive Event Specification Language for Active Databases. *Data and Knowledge Engineering, 14(1)* (pp. 1-26).

[5] Damianou, N., Dulay, N., Lupu, E., Sloman, M. (2001, January). The Ponder Policy Specification Language. *Proceedings Policy 2001: Workshop on Policies for Distributed Systems and Networks* (pp. 18-39). Bristol, UK: Springer-Verlag LNCS 1995.

[6] de Champeaux, D., Lea, D., Faure, P. (1993, May). *Object Oriented System Development*: Addison-Wesley.

[7] Grefen, P., Aberer, K., Hoffner, Y., Ludwig, H. (2000). CrossFlow – Cross-organizational Workflow Managament in Dynamic Virtual Enterprises. *International Journal of Computer Systems Science & Engineering, Vol. 15, No. 5* (pp. 277-290).

[8] *JADE – Java Agent DEvelopment Framework* (n. d.) Retrieved September 22, 2005 from http://jade.tilab.com

[9] *Jess, the Rule Engine for the Java™ Platform* (2005, August) Retrieved September 22, 2005 from http://herzberg.ca.sandia.gov/jess/index.shtml

[10] Lea, D. (1994, January). Christopher Alexander, An Introduction for Object-Oriented Designers. *ACM SIGSOFT Software Engineering Notes, Volume 19, Issue 1* (pp. 39 - 46): ACM Press, New York, USA.

[11] Meng, J., Su, S. Y. W., Lam, H., Helal, A. (2002, January). Achieving Dynamic Inter-Organizational Workflow Management by Integrating Business Processes, Events and Rules. *Proceedings of the 35th Hawaii International Conference on System Sciences*. Big Island, HI, USA. IEEE Computer Society, 2002 - Track 1

[12] Miller, J. A., Palaniswami, D., Sheth, A. P., Kochut, K. J., Singh, H. (1998). WebWork: METEOR_2's Web-based Workflow Management System. *Journal of Intelligent Information Systems, Special Issue on Workflow Management Systems, Volume 10, Number 2* (pp. 185-215).

[13] Orriëns, B., Yang, J., Papazoglou, M. P. (2003, September). A Framework for Business Rule Driven Service Composition. *Proceedings of the 3rd VLDB-TES Workshop (TES03)*. Berlin, Germany.

[14] *Rewerse WG I2 – Policy language, enforcement, composition*. (2004, March). Retrieved September 22, 2005 from http://rewerse.net/I2/

[15] Ross-Talbot, S., Tabet, S., Chakravarthy, S., Brown, G. (n. d.). *A generalized RuleML-based Declarative Policy specification language for Web Services*. Retrieved September 22, 2005 from http://www.w3.org/2004/08/ws-cc/srtst-20040903

[16] Singh, M. P., Tomlinson, C. (1994). Workflow Execution Through Distributed Events. *Proceedings of the 6th International Conference on Management of Data (COMAD).*

[17] Singh, M. P. (1996, March). Synthesizing Distributed Constrained Events from Transactional Workflow Specifications. *Proceedings of the 12th International Conference on Data Engineering (ICDE)* (pp. 616-623).

[18] *The Policy RuleML Technical Group*. (2004, March). Retreived September 22, 2005 from http://policy.ruleml.org

[19] Wan, F., Rustogi, S. K., Xing, J., Singh, M. P. (1999, August). Multiagent Workflow Management. In IJCAI *Workshop on Intelligent Workflow and Process Management: The New Frontier for AI in Business*. Stockholm, Sweden.

[20] Xing, J., Wan, F., Rustogi, S. K., Singh, M. P. (2001, October). A Commitment-Based Approach for Business Process Interoperation. *IEICE Transactions on Information and Systems, E84-D(10)* (pp. 1324-1332).

Solving the Optimal Strategic Plan of Fuel Cell Industry by TOPSIS Method

Hua-Kai Chiou
Department of Statistics, National Defense University, Taiwan
Email: hkchiou2005@aptg.net

Gwo-Hshiung Tzeng
Institute of Management of Technology, National Chiao Tung University;
Department of Business Administration, Kainan University, Taiwan.
Email: ghtzeng@mail.knu.edu.tw

Chia-Chin Wan
Department of Business Administration, Kainan University, Taiwan.
Email: wancc@mail.knu.edu.tw

Abstract

The fuel cell is one of the most important energy products in the 21st century, and most industrial-advanced countries in the world are placing high expectation on its development. In Taiwan, applications of fuel cells are spread broadly in many products and it takes high global market share, such as notebook computers, PDAs and digital cameras, along with items like automobiles, motorcycles and power generation equipment. In this study we introduce a multi-objective compromise optimization called TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) to solve the optimal strategic plans of fuel cell industry, which is emerging industry in Taiwan. Firstly, fuzzy AHP (Analytic Hierarchy Process) was employed to establish the hierarchy system to determine the relative importance of evaluated criteria. Secondly, fuzzy geometric mean method was utilized to aggregate the performance score by individual judgment of evaluator, which scores express the measurement of strategic plan proposed by participated experts. Thirdly, the synthetic value of each strategic plan was integrated by criteria weights with performance score, and positive ideal solution (PIS) and negative ideal solution (NIS) were defined by Minkowski's L_p *-metric. Fourthly, the optimal compromise solution should have the shortest distance from the positive ideal solution as well as the farthest distance from the negative ideal solution. Through this research, we successfully demonstrate that TOPSIS is a good assessment for evaluation of multiple criteria decision making problems.*

Keywords:

Multi-objective Compromise Optimization, TOPSIS, Fuzzy Analytic Hierarchy Process, Fuel Cell Industry

INTRODUCTION

In recent years, the increasing problem of global warming and climate change has resulted in a big impact on the environment and economy. Therefore, the Protocol of United Nation Climatic Change Summit was signed in Kyoto, Tokyo 1977 to commit by 2012, there would be a 5% decrease of total emissions against that of 1990, thus slowing the global warming process. According to the blueprint of this protocol, each nation has started to actively push the development of green energy. The fuel cell, one of important green energy, has the advantages of being highly efficient and environmental friendly. It also can be applied to the national defense, livelihood of the people and the industry. Fuel cell development has become a main policy in suppressing a nation's future carbon dioxide emissions. The relevant research as well as its progress to commercialization can therefore be facilitated (Table 1).

Hester (2001) predicted the global market for fuel cell could reach US$ 8.5 billion in 2005, and up to US$ 23.3 billion in 2010. Among this, electrical fuel cell will have a market of US$ 3.8 billion in 2005, and up to US$10.5 billion in 2010, with a compounded growth rate of 26.6% for the years 2000-2010; Automobile powered fuel cell market predicted to be US$1.8 billion in 2005, and reaching US$3.9 billion in 2010, with compounded growth rate of 20.6% for the years 2000~2010; Portable power will have a market of US$1.2 billion, and reaching US$4.5 billion in 2010, with a compounded growth rate of 35.6% for the years 2000-2010; other applications will have an estimated market of US$1.7 billion, growing to 4.4 billion in 2010, with compounded growth rate of 21.8% for the years 2000-2010. When looking at each nation's market in the year 2000, the US lead with US$0.72 billion (29.5%); Japan in second with 0.675 billion (27.7%); Germany ranking third with 0.26 billion (10.7%). The compounded growth rate for the three countries of years 2000-2010 are estimated at 24.6%, 21.7%, and 21.4%, respectively (Table 2 & Table 3).

Table 1. Alternative Energy Development Policy from Major Countries (Tsai, 2001)

Country	Projects/Actions	Alternative Energy/New Energy Policy
Austria	Recycle energy power generation project	50% of the financial aid for the equipment used on recycle energy in some areas.
US	Technology development foundation	US$2.4 billion to the original global warming foundation, increasing 40% budget for developing the clean energy.
Japan	Global warming action plan	Enhance technology development on fuel cell, solar energy and bio energy. Japan's fuel cell development goal, 2.2 million watts by 2010.
Germany	Revision on Power Parallel Connection Bill	Revision Act to target 10% of the recycle energy by 2010 and energy ratio of 4%.
	Ecology tax act	Higher tax on petrochemical energy to reduce the total consumption of the petro fuel.
Korea	1997-2006 energy technology development program	Plan to invest Korea Won 2,047 billion to develop energy protection technology, alternative energy, and clean energy. The alternative energy includes fuel cell, solar energy, solar optic electronics and IGCC.
Korea	New and recycle energy development act	Mainly to develop 12 alternative/ clean energy technologies, including fuel cell. The incentives granted include 5% of the investment is tax free and up to 80% of the investment can enjoy the 5% low interest loan.

Table 2. Global Regional Fuel Cell Market Status and Its Forecasting (Hester, 2001)

Countries	US	Canada	Germany	UK	France	Italy
2000 Market Size in US$ million (Global Share%)	720 (29.5%)	140 (5.7%)	260 (10.7%)	130 (5.12%)	50 (2.1%)	55 (2.3%)
2010 Market Size in US$ million (Global Share%)	65	11	18	11	7	6.5
2000~2010 Compounded Growth Rate	24.61%	22.89%	21.35%	24.29%	30.20%	28.01%
Countries	**Japan**	**China**	**Korea**	**Latin America**	**Europe**	**Africa & Middle East**
2000 Market Size in US$ million (Global Share%)	675 (27.7%)	50 (2.1%)	50 (2.1%)	15 (0.6%)	225 (2.9%)	15 (0.6%)
2010 Market Size in US$ million (Global Share%)	48	13	6.5	5.5	7	3.5
2000~2010 Compounded Growth Rate	21.67%	37.97%	29.24%	43.36%	25.89%	37.02%

Table 3. 2000-2001Market Forecasting on Global Fuel Cell Applications (Hester, 2001)

Applications	2005 (US$ million)	2010 (US$ million)	2000~2010 Compounded Growth Rate
Power Generation	38	105	26.64%
Vehicle Power	18	39	20.61%
Portable Power	12	45	35.6%
Others	17	44	21.78%
Total	85	233	

Ozbek (2001) estimated that total capacity of 2003 global fuel cell generator would reach 405 million watts, while reaching at 16,608 million watts in 2011, with a compounded growth rate of 59% for the years 2003-2011.

Motomatsu (2002) estimated that global PEMFC market could reach at US$10 billion by 2010 while the global value of fuel cell of US$ 0.7 billion in 1998. As to the applied market in the industry, the electronics industry was estimated at US$1.5 billion, and with the market for UPS system, motorcycles, bicycles, and electric power generators estimated at US$1.5 billion. Residential and Commercialized non-movable batteries have an estimated market of US$6.0 billion. Other transportation applications have a market smaller than US$1.0 billion.

Makoweal et al. (2003) estimated that the global clean energy market will be at US$0.95 billion, reaching an estimated of US$89 billion in 2012. Among this market, the fuel cell market is estimated at US$0.5 billion in 2003, reaching 12.5 billion in 2012.

However, in past researches mostly applied descriptive statistics or multivariate analysis to express the industrial development. This study introduces alternative method, multi-criteria compromise optimization, to solving the optimal strategic plan.

In addition, the decision-making problems are very often uncertain or vague in a number of ways in real world systems. Due to lack of information, the future state of the system might not be known completely. This type of uncertainty has long been handled appropriately by probability theory and statistics. However, in many areas of daily life, such as engineering, medicine, meteorology, manufacturing and others, human judgment and decisions often employ natural language to express subjective perception such as “Very Important”, “Important”, “Medium Important”, “Not Important” and “Very Unimportant”. In these natural languages the meaning of words is often vague. Although the meaning of a word might be well-defined, when using the word as a label for a set, the boundaries within which objects do or do not belong to the set become fuzzy or vague.

Furthermore, human judgment of events may be significantly different based on individuals’ subjective perceptivity or personality, even using the same words. Fuzzy numbers are introduced to appropriately express linguistic variables. In this study, we first applied fuzzy hierarchical analytic process to evaluate the development strategy for nanotechnology industry, and then utilized multiobjective compromise optimization to determine the preferred order of these strategies.

In what follows, fuzzy analytic hierarchy process for multi-criteria decision making problems introduce in Section 2. The multiobjective compromise optimization method describe in Section 3. The evaluation of emerging industrialized technology and development strategies for fuel cell industry demonstrated in Section 4. Finally, some concluding remarks are summarized in Section 5.

FUZZY ANALYTIC HIERARCHY PROCESS FOR EVALUATION

Analytic Hierarchy Process (AHP) is a popular technique often used to model subjective decision-making processes based on multiple attributes (Saaty, 1977; 1980). Application of AHP in MCDM environments involves defining a common hierarchy of criteria, specifying pairwise comparisons by members of the group and aggregating those pairwise comparisons for the entire group. Saaty used the principal eigenvector of the comparison matrix to find the comparative weights among the criteria of the hierarchy systems. Here, we employ the approach proposed by Buckley (1985) to derive analytic hierarchy process by allowing fuzzy numbers for the pairwise comparisons, and find the fuzzy weights and fuzzy performance; in this section we briefly review concepts for fuzzy hierarchical evaluation model.

In real Multiple Criteria Decision Making (MCDM) problems, it is necessary to divide the process into distinct stages. Firstly, based on a general problem statement, the various stakeholders are defined, typically including the decision-makers, various interest groups affected by the decision, experts in the appropriate fields, as well as planners and analysts responsible for the preparations and managing the process. The overall objective will be set up in this stage. Secondly, based on various points of view from stakeholders, the problems can be categorized into distinct aspects. Thirdly, defining alternatives/strategies and criteria, a discrete MCDM problem consisting of a finite set of alternatives/strategies can be evaluated in terms of multicriteria. Finally, choosing a suitable method to measure the criteria can help the evaluators and analysts to process the evaluating cases.

Building Hierarchy Frame for Evaluation

The analytic hierarchy process can decompose a complicated policy decision problem from a higher hierarchy topic into many smaller items for a more quantitative/qualitative analysis. The evaluators must establish a hierarchical system for analysis and evaluation in the multiple criteria decision-making problem. Keeney and Raiffa (1976) suggest that five principles must be followed when criteria are being formulated: (1) Completeness, (2) Operationality, (3) Decomposability, (4) Nonredundancy, and (5) Minimum size.

Generally, the hierarchy system for evaluation can therefore be simply structured into three phases. The first phase is the goal, which is the overall objective. The second phase includes all aspects of sub-objective for analysis. The last phase includes all feasible strategies/alternatives for implementing the evaluated aspects.

Determining the Criteria Weights

Considering the evaluation of criteria entails diverse and meanings, we cannot assume that each evaluated criterion is of equal importance. Hwang and Yoon (1981) summarized many methods that can be employed to determine weights such as the eigenvector method, weighted least square method, entropy method, Analytic Hierarchy Process (AHP), as well as linear programming techniques for multidimension of analysis preference (LINMAP). However, the selected approach depends on the nature of the problems.

Buckley (1985) considered a fuzzy positive reciprocal matrix $\tilde{A} = [\tilde{a}_{ij}]$, extending the geometric mean technique to define the fuzzy geometric mean of each row $\tilde{r}_i$ and fuzzy weight $\tilde{w}_i$ corresponding to each criterion as follows:

$$\tilde{r}_i = (\tilde{a}_{i1} \otimes \tilde{a}_{i2} \otimes \cdots \otimes \tilde{a}_{im})^{1/m}; \quad \tilde{w}_i = \tilde{r}_i \otimes (\tilde{r}_1 \oplus \tilde{r}_2 \oplus \cdots \oplus \tilde{r}_m)^{-1} \tag{1}$$

where $\oplus$ and $\otimes$ express the addition and multiplication operation of fuzzy numbers, respectively.

According to the characteristics of triangular fuzzy numbers and the extension principle which put forward by Zadeh (1965), the operational laws of two triangular fuzzy numbers $\tilde{A} = (a_1, a_2, a_3)$ and $\tilde{B} = (b_1, b_2, b_3)$ are as follows:

(1) Addition of two fuzzy numbers $\oplus$

$$(a_1, a_2, a_3) \oplus (b_1, b_2, b_3) = (a_1 + b_1,\ a_2 + b_2,\ a_3 + b_3) \tag{2}$$

(2) Subtraction of two fuzzy numbers Θ

$$(a_1, a_2, a_3) \Theta (b_1, b_2, b_3) = (a_1 - b_3,\ a_2 - b_2,\ a_3 - b_1) \tag{3}$$

(3) Multiplication of two fuzzy numbers $\otimes$

$$(a_1, a_2, a_3) \otimes (b_1, b_2, b_3) \cong (a_1 b_1, a_2 b_2, a_3 b_3) \tag{4}$$

(4) Multiplication of any positive real number k and a fuzzy number

$$k\ (a_1, a_2, a_3) = (ka_1, ka_2, ka_3) \tag{5}$$

(5) Division of two fuzzy numbers Δ

$$(a_1, a_2, a_3) \Delta (b_1, b_2, b_3) \cong (a_1 / b_3,\ a_2 / b_2,\ a_3 / b_1) \tag{6}$$

Deriving the Synthetic Values

In addition, we request the evaluators choose a performance value for each feasible strategic plan corresponding to considered criteria based on their subjective judgments. When we determined the criteria weights and performance values of feasible strategies corresponding to criteria, the next step is to aggregate the synthetic value for each strategy. In this study, we utilize the fuzzy geometric mean method to determine the criteria weights and integrate the performance value through participated evaluators. In order to provide easy-to-follow process for MCDM problems, here we simplify the process to conduct the synthetic values employing simple additive weighted method to integrate the criteria weights with performance values for each strategy.

The result of the fuzzy synthetic decision reached by each strategy is a fuzzy number. Therefore, it needs to defuzzify the fuzzy numbers for getting the preferred order of the strategies. In previous works the procedure of defuzzification has been to locate the best nonfuzzy performance (BNP) value. Methods of such defuzzified fuzzy ranking generally include three kinds of method, mean of maximal, center of area (COA), and α -cut (Tsaur et al., 1997; Opricovic and Tzeng, 2003). Utilizing the COA method to determine the BNP is a simple and practical method, the BNP value of the triangular fuzzy number (LR_i, MR_i, UR_i) can be found by as follows:

$$BNP_i = [(UR_i - LR_i) + (MR_i - LR_i)]/3 + LR_i, \forall i \tag{7}$$

TOPSIS METHOD FOR MULTI-OBJECTIVE COMPROMISE OPTIMIZATION

With a given reference point, the MCDM problem can then be solved by locating the alternatives or decision points that are the closest to the reference point. Therefore, the problem becomes how to measure the distance to the reference point.

Generally, the global criteria method measures the distance by using Minkowski's L_p-metric. The L_p metric defines the distance between two points, f_j and f_j^* (the reference point), in n-dimensional space as:

$$L_p = \left\{ \sum_{j=1}^{k} (f_j^* - f_j)^p \right\}^{1/p}, \text{ where } p \geq 1 \tag{8}$$

Distance L_p $(p = 1, 2, ..., \infty)$ are especially operationally important, the distance L_p decreases as p increase, i.e., $L_1 \geq L_2 \geq \cdots \geq L_\infty$. Specifically, for $p = 1$, it implies equal weights for all these deviations, L_1 called the Manhattan distance; for $p = 2$, it implies that these deviations are weighted proportionately with the largest deviation having the largest weight, L_2 called the Euclidean distance. Ultimately, while $p = \infty$, it implies the largest deviation completely dominates the distance determination, L_∞ usually called the Tchebycheff metric, is the shortest distance in the numerical sense (Steuer, 1986). That is,

$$L_\infty = \max_j \left\{ \left| f_j^* - f_j \right| \middle| j = 1, ..., k \right\} \tag{9}$$

Considering the incommensurability nature among objectives or criteria, Yu and Zeleny (1975) normalized the distance family of Eq.(8) to remove the effects of the incommensurability by using the reference point. The distance family then becomes as follows:

$$L_p = \left\{ \sum_{j=1}^{k} \left(\frac{f_j^* - f_j}{f_j^*} \right)^p \right\}^{1/p}, \text{ where } p \geq 1 \tag{10}$$

Hwang and Yoon (1981) proposed TOPSIS approach to solve multiple attribute decision making problems (MADM) using the concept of optimal compromise solution. Lai et al. (1994) further extended the concept of TOPSIS for MADM problems and developed a methodology for solving multiple objective decision making (MODM) problems. In their study using the normalized distance family from Eq.(10) with the ideal solution being the reference point, the problem as Eq.(8) becomes how to solve the following auxiliary problem:

$$\min_{x \in X} \; L_p = \left\{ \sum_{j=1}^{k} \left(\frac{f_j^* - f_j}{f_j^* - f_j^-} \right)^p \right\}^{1/p} \quad \text{where } p = 1, 2, ..., \infty \tag{11}$$

where f_j^* is the best value of corresponding j-th criterion, f_j^- is the worst value of corresponding j-th criterion, and $\boldsymbol{f}^* = \left(f_1^*, ..., f_j^*, ..., f_k^* \right)$ is the vector of positive ideal solution, $\boldsymbol{f}^- = \left(f_1^-, ..., f_j^-, ..., f_k^- \right)$ is the vector of negative ideal solution, respectively. The value chosen for p reflects the way of achieving a compromise by minimizing the weighted sum of the deviations of criteria from their respective reference points.

With the concept of optimal compromise solution, the best alternative or decisions of TOPSIS method are those that have the shortest distance from the positive ideal solution as well as have the farthest distance from the negative ideal solution.

The TOPSIS procedure consists of the following steps:

(1) Calculate the normalized decision matrix. The normalized value r_{ij} is calculated as:

$$r_{ij} = f_{ij} \Big/ \sqrt{\sum_{j=1}^{J} f_{ij}^2}, \; j = 1, ..., k; \; i = 1, .., n \tag{12}$$

(2) Calculate the weighted normalized decision matrix. The weighted normalized value v_{ij} is calculated as:

$$v_{ij} = w_j r_{ij}, \; j = 1, ..., k; \; i = 1, ..., n \tag{13}$$

where w_j is the weight of the j-th attribute or criterion, and $\sum_{j=1}^{k} w_j = 1$.

(3) Determine the ideal and negative-ideal solution.

$$\begin{cases} A^* = \{v_1^*,...,v_k^*\} = \{(\max_j v_{ij} \left| i \in I^{'}\right.), (\min_j v_{ij} \left| i \in I^{''}\right.)\} \\ A^- = \{v_1^-,...,v_k^-\} = \{(\min_j v_{ij} \left| i \in I^{'}\right.), (\max_j v_{ij} \left| i \in I^{''}\right.)\} \end{cases} \tag{14}$$

where $I^{'}$ and $I^{''}$ are associated with benefit and cost criteria, respectively.

(4) Calculate the separation measures, using the k-dimensional Euclidean distance. The separation of each alternative from the positive ideal solution is given as:

$$D_i^* = \sqrt{\sum_{j=1}^{k}(v_{ij} - v_j^*)^2}\,;\ i = 1,...,n \tag{15}$$

Similarly, the separation from the negative ideal solution is given as:

$$D_i^- = \sqrt{\sum_{j=1}^{k}(v_{ij} - v_j^-)^2}\,;\ i = 1,...,n \tag{16}$$

(5) Calculate the relative closeness to the ideal solution. The relative closeness of the alternative a_i with respect to A^* is defined as:

$$C_i^- = D_i^- / (D_i^* + D_i^-);\ i = 1,...,n \tag{17}$$

(6) Rank the preference order according to closeness, the great the better of closeness.

Eq.(17) represents the basic principle in the TOPSIS method.

EMPIRICAL CASE

This section is divided into five subsections: (1) problem description, (2) determining the evaluated criteria weights, (3) calculating the performance matrix of strategies corresponding to considered criteria, (4) integrating the synthetic utilities using simple additive weighted method, (5) deriving the preferred order of strategic action plans using TOPSIS process.

Problem Description

This research took three months to complete the investigation. The participated evaluators have 35 experts who have good knowledge about fuel cell development in Taiwan. There then have 26 responded questionnaires, accounting for 74.3% responded rate. Nine experts were from some major private companies in the industry (group 1), six were from the government sectors such as Industrial Development Bureau and Technology Development Division of Ministry of Economics Affair (group 2), five were from the universities in the academia sector (group 3), and six were from the not-for-profit research institutes such as Industrial Technology Research Institute (group 4). The study indicated that the fuel cell product development priority lists are information/communication computer and consumer electronics (3C products), power equipment, motorcycle and automobile. The encountered difficulties for the industry are technology bottleneck, insufficient investment on R&D, high production cost, fuzzy unclear government policy and insufficient R&D professionals. The top five concluded strategic action plans using by fuzzy multi-criteria decision making (Fuzzy MCDM) procedure are conducted as: Government policy to regulate the required percentage of the clean energy, establish national research program, increase R&D budget, carefully select niche product and plan the operating demonstration pilot zone with financial aid.

This study using the AHP and special vector process to calculate the weightings from the nineteen criteria is as shown in Figure 1. The thirteen proposed strategic action plans are listed in the priority order based on the calculated weightings from their attributes of encountered difficulties.

Determining the Criteria Weights

After establishing the hierarchy frame of evaluation model shown in Figure 1, Fuzzy AHP technique utilized to determine the relative weights of considered criteria from evaluators' subjective judgment. The weights from each criterion in the proposed strategy can therefore be determined. We aggregate their own subjective judgments by fuzzy geometric mean method and then conduct the final fuzzy weights, we further utilize fuzzy AHP to derive the aggregated weights of industrialized technology that from subjective judgment of evaluators and then rank the development priority of these technologies, computing the defuzzified BNP values followed that proposed by Opricovic and Tzeng (2003) as shown in Table 4.

Integrating the Synthetic Values

To determine the performance value of each strategy, the evaluators can define their own individual range for the linguistic variables employed in this study according to their subjective judgments within a fuzzy scale. In order to make more clearly comprehensive in considered criteria with strategies for readers, we express the nonfuzzy performance value of strategies with respect to evaluated criteria and show as Table 5.

When we determined the criteria weights and performance values of evaluated criteria and feasible strategies, the next step is to integrate the synthetic utility for each strategy. Here we employ simple additive weighted method to integrate the nonfuzzy criteria weights with nonfuzzy performance values for each strategy, the results as shown in Table 6.

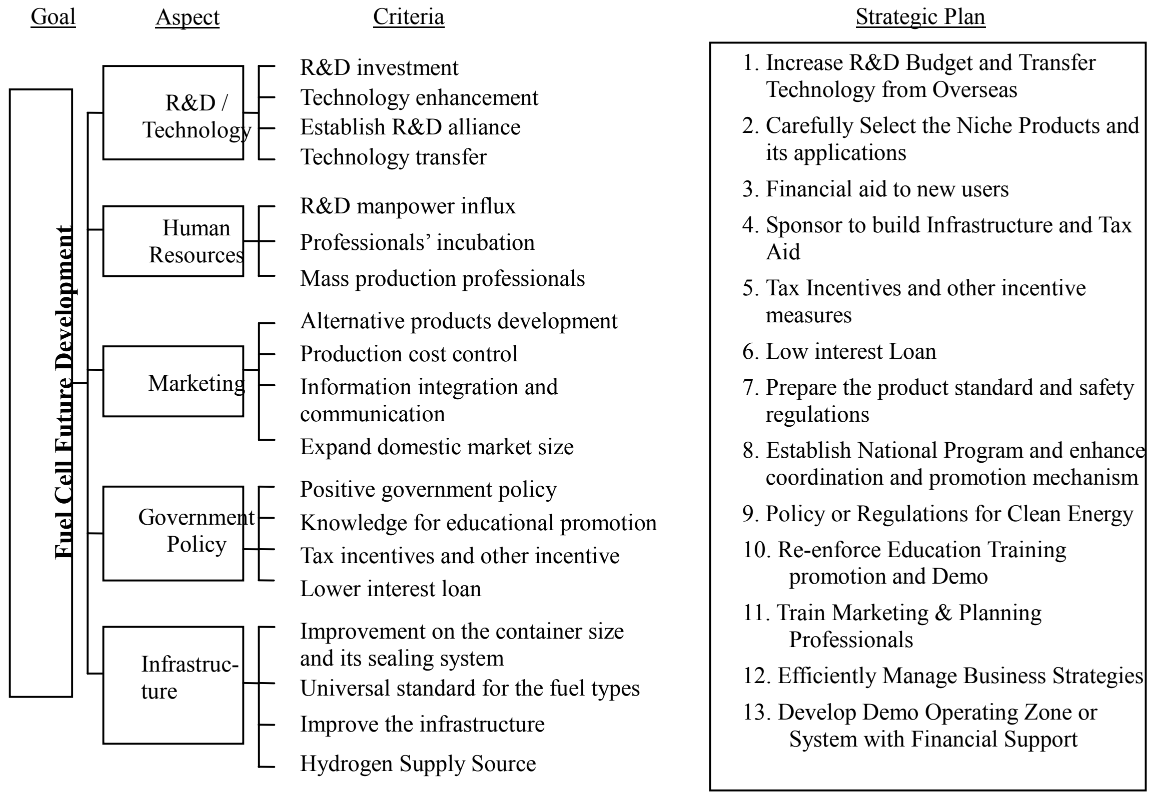


Figure 1. Evaluation Model for Fuel Cell Development in Taiwan

TABLE 4. Normalized BNP Values of Criteria

Criteria	c11	c12	c13	c14	c21	c22	c23	c31	c32	c33
Group 1	0.052	0.099	0.056	0.061	0.043	0.034	0.037	0.069	0.098	0.06
Group 2	0.067	0.118	0.045	0.033	0.096	0.055	0.092	0.040	0.045	0.01
Group 3	0.157	0.129	0.052	0.062	0.093	0.066	0.032	0.038	0.054	0.02
Group 4	0.098	0.099	0.031	0.029	0.044	0.033	0.020	0.053	0.107	0.03
Aggregate	0.094	0.111	0.046	0.046	0.069	0.047	0.045	0.050	0.076	0.03

Ranking	2	1	9	10	5	8	11	6	4	16

Criteria	c34	c41	c42	c43	c44	c51	c52	c53	c54
Group 1	0.033	0.092	0.040	0.063	0.055	0.034	0.023	0.033	0.017
Group 2	0.017	0.101	0.067	0.039	0.034	0.026	0.052	0.040	0.020
Group 3	0.015	0.051	0.018	0.050	0.026	0.044	0.016	0.044	0.033
Group 4	0.053	0.060	0.046	0.044	0.010	0.073	0.049	0.058	0.057
Aggregate	0.029	0.076	0.043	0.049	0.031	0.044	0.035	0.044	0.032
Ranking	19	3	14	7	18	12	15	13	17

TABLE 5. Normalized BNP Performance Values

Criteria	c11	c12	c13	c14	c21	c22	c23	c31	c32	c33
S1	81.549	68.761	52.527	68.656	78.166	69.975	64.190	68.402	68.616	60.21
S2	74.237	70.294	56.783	54.352	64.435	59.763	61.416	67.244	66.475	50.53
S3	62.632	41.742	37.480	43.220	50.124	39.756	48.469	64.609	69.157	46.97
S4	57.234	42.475	44.236	37.138	49.470	40.562	45.201	58.827	63.836	43.98
S5	66.000	49.842	50.863	51.819	48.103	47.656	50.964	57.897	66.692	44.01
S6	38.607	31.475	28.288	30.985	31.301	28.099	33.951	39.810	44.495	36.44
S7	55.534	52.496	54.358	54.226	60.028	56.177	54.365	49.308	47.629	46.29
S8	75.429	74.435	64.663	65.618	75.186	73.713	58.701	60.378	53.935	65.12
S9	76.680	69.505	63.839	63.818	69.633	64.402	64.039	69.104	71.344	56.66
S10	47.001	41.860	42.391	43.347	53.184	66.725	53.200	48.371	41.665	54.12
S11	32.786	25.739	30.292	29.838	35.781	50.090	36.944	36.459	35.842	36.71
S12	46.499	46.811	52.551	54.007	47.501	50.441	56.550	52.415	53.493	44.62
S13	58.776	53.701	47.379	42.641	53.956	52.428	55.829	63.880	59.573	51.39

Criteria	c34	c41	c42	c43	c44	c51	c52	c53	c54
S1	56.861	64.684	55.732	46.710	44.077	65.734	65.029	61.857	58.306
S2	61.581	64.221	60.049	54.694	45.417	38.000	43.156	44.230	42.765
S3	62.048	63.403	59.957	63.271	44.368	42.780	50.610	66.952	59.318
S4	58.247	57.550	54.081	71.404	42.986	38.968	41.786	49.889	42.992
S5	45.646	40.547	40.218	52.918	60.640	27.420	27.522	37.825	30.237
S6	39.971	58.508	45.138	38.643	34.700	45.662	60.445	57.294	45.424
S7	48.116	79.249	65.679	39.631	36.105	68.095	65.375	59.147	51.767
S8	64.158	78.463	73.453	59.920	53.038	59.375	71.624	71.455	59.256
S9	44.006	58.260	80.916	38.080	30.815	37.176	50.051	39.933	34.298
S10	37.410	34.247	52.363	33.010	27.929	23.520	22.721	29.237	24.157
S11	48.726	52.925	52.660	41.710	36.215	41.096	45.063	46.102	40.795
S12	55.416	69.614	71.118	50.796	43.218	45.387	53.880	58.950	46.038
S13	51.013	78.563	59.086	49.897	46.249	76.025	64.360	63.438	56.500

TABLE 6. Nonfuzzy Synthetic Values of Strategies

Criteria	c11	c12	c13	c14	c21	c22	c23	c31	c32	c33
S1	7.627	7.642	2.427	3.170	5.408	3.297	2.905	3.420	5.209	1.944
S2	6.943	7.813	2.624	2.510	4.458	2.816	2.779	3.362	5.047	1.631
S3	5.858	4.639	1.732	1.996	3.468	1.873	2.193	3.230	5.251	1.516
S4	5.353	4.721	2.044	1.715	3.423	1.911	2.046	2.941	4.847	1.420
S5	6.173	5.540	2.350	2.393	3.328	2.246	2.306	2.894	5.063	1.421
S6	3.611	3.498	1.307	1.431	2.166	1.324	1.536	1.990	3.378	1.176
S7	5.194	5.835	2.512	2.504	4.153	2.647	2.460	2.465	3.616	1.494
S8	7.054	8.273	2.988	3.030	5.202	3.473	2.656	3.018	4.095	2.102
S9	7.171	7.725	2.950	2.947	4.818	3.035	2.898	3.455	5.417	1.829
S10	4.396	4.652	1.959	2.002	3.680	3.144	2.407	2.418	3.163	1.747
S11	3.066	2.861	1.400	1.378	2.476	2.360	1.672	1.823	2.721	1.185
S12	4.349	5.203	2.428	2.494	3.287	2.377	2.559	2.620	4.061	1.440
S13	5.497	5.968	2.189	1.969	3.733	2.470	2.526	3.193	4.523	1.659
Criteria	c34	c41	c42	c43	c44	c51	c52	c53	c54	Grand
S1	1.501	5.970	2.523	2.438	1.454	3.346	2.248	2.788	1.802	67.119
S2	1.673	4.916	2.380	2.283	1.386	2.893	2.271	2.719	1.859	62.361
S3	1.812	4.880	2.564	2.673	1.428	1.672	1.507	1.944	1.364	51.600
S4	1.826	4.818	2.560	3.092	1.395	1.883	1.767	2.943	1.892	52.594
S5	1.714	4.373	2.309	3.489	1.351	1.715	1.459	2.193	1.371	53.689
S6	1.343	3.081	1.717	2.586	1.906	1.207	0.961	1.662	0.964	36.846
S7	1.176	4.446	1.928	1.888	1.091	2.009	2.111	2.518	1.449	51.496
S8	1.416	6.022	2.805	1.937	1.135	2.997	2.283	2.600	1.651	64.737
S9	1.888	5.963	3.137	2.928	1.667	2.613	2.501	3.141	1.890	67.971
S10	1.295	4.427	3.455	1.861	0.969	1.636	1.748	1.755	1.094	47.808
S11	1.101	2.603	2.236	1.613	0.878	1.035	0.793	1.285	0.770	33.256
S12	1.434	4.022	2.249	2.038	1.138	1.809	1.574	2.026	1.301	48.408
S13	1.631	5.290	3.037	2.482	1.359	1.997	1.882	2.591	1.468	55.466

Deriving The Preferred Order

Following the procedure as mentioned in Section 3, we derive distance from the positive ideal solution (S^{+}) and from the negative ideal solution (S^{-}) of each strategy. Finally, we further compute the relative closeness as Eq.(17) and then assign the preferred order to each strategy according to their closeness index (Table 7).

TABLE 7. Preferred Order Derived by TOPSIS

	Derived by TOPSIS				Derived by Fuzzy MCDM	
	S^{+}	S^{-}	C_i^{-}	Rankin g	Synthetic values	Ranking
S1	0.0096 5	0.04888	0.835	2	67.119	2
S2	0.0139 0	0.04333	0.757	4	62.361	4
S3	0.0304 1	0.02726	0.473	9	51.600	8

S4	0.0297 3	0.02783	0.483	8	52.594	7
S5	0.0263 0	0.03056	0.537	6	53.689	6
S6	0.0476 7	0.01064	0.182	12	36.846	12
S7	0.0276 4	0.02848	0.507	7	51.496	9
S8	0.0134 9	0.04766	0.779	3	64.737	3
S9	0.0077 8	0.04818	0.861	1	67.971	1
S10	0.0349 5	0.02251	0.392	11	47.808	11
S11	0.0522 3	0.00598	0.103	13	33.256	13
S12	0.0324 8	0.02276	0.412	10	48.408	10
S13	0.0237 1	0.03160	0.571	5	55.466	5

From Table 7, we conduct the top five preferred strategic action plans of fuel cell industry in Taiwan as follows: Government policy to regulate the required percentage of the clean energy, Establish national research program, Increase R&D budget, Carefully select niche product, and Plan the operating demonstration pilot zone with financial aid, this order is the same as mentioned above while we employ fuzzy multi-criteria decision making (Fuzzy MCDM) procedure.

CONCLUSIONS

The results of this research indicated the priority of the Taiwan's fuel cell development aspects or attributes: (1)R&D/Technology; (2)Government policy; (3)Marketing; (4)Human Resources; (5)Infrastructure.

The first five priority of the evaluated criteria are: (1)Technology Enhancement; (2) R&D investment; (3)Production Cost Control; (4)Positive Government Policy; (5)R&D Manpower Influx.

The priority of the strategic actions are: (1)Government should have the policy to regulate the required percentage for the clean energy; (2)Increase R&D Budget; (3)Establish National level Program; (4)Carefully select the niche products and their applications; (5)Support the operating demonstration or pilot zone with financial aid; (6)Prepare the product standards and safety regulations; (7)Expand tax incentives and incentive measures; (8)Build Infrastructure and Tax Aid; (9) Incentives with financial aid to the new users; (10)Flexible efficient business strategies; (11)Enhance educational training and promotion activities; (12)Low interest loan; (13)Train marketing & planning professionals.

REFERENCES

[1] Buckley, J.J. "Fuzzy Hierarchical Analysis," *Fuzzy Sets and Systems*, Vol. 17, No. 3, pp. 233-247, 1985.

[2] Chen, S.J. and Hwang, C.L. *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1992..

[3] Hester, E. D., *World Fuel Cells*, The Freedonia Group. Inc. Cleveland, Ohio,2001.

[4] Hwang, C.L. and Yoon, K. *Multiple Attribute Decision Making*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1981.

[5] Keeney R.L. and Raiffa, H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, New York: John Wiley and Sons, 1976.

[6] Lai, Y.J., Liu, T.Y. and Hwang, C.L. "TOPSIS for MODM," *European Journal of Operational Research*, 76(3), 486-500, 1994.

[7] Malowea, J., Pernick, R. and Wilder, C., *Clean Energy Trends 2003*, Clean Edge Inc. California San Francisco, 2003.

[8] Motomatsu, M., "Global Market Trend of PEM Fuel Cells and Dupont's strategy",2002 International Seminar on Fuel Cell July 2-3, 2002, Taiwan, 2002.

[9] Opricovic S. and Tzeng, G.H. "Defuzzification for a fuzzy multicriteria decision model," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* Vol.11, No.5, pp.635-652, 2003

[10] Ozbek, A., *Fuel Cell Industry Competitive Analysis*, Allied Business Communications, Inc. New York, 2001.

[11] Saaty, T.L. "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, Vol. 15, No. 3, pp. 234-281, 1977.

[12] Saaty, T.L. *Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.

[13] Steuer, R.E. *Multiple Criteria Optimization Theory, Computation, and Applications*, Wiley, New York, 1986.

[14] Tsai, M.S. Global Warming Protocol Decision Supporting Project Final Report, Industrial Technology Research Institute 2001(in Chinese).

[15] Tsaur, S.H. Tzeng G.H. and Wang, G.C. "The Application of AHP and Fuzzy MCDM on the Evaluation Study of Tourist Risk," *Annals of Tourism Research*, Vol. 24, No. 4, pp. 796-812, 1997.

[16] Yu P.L. and Zeleny, M. "The Set of All Nondominated Solutions in Linear Cases and A Multicriteria Simplex Method," *Journal of Mathematical Analysis and Applications*, Vol.49, No.3, pp.430-448, 1975.

[17] Zadeh, L.A. "Fuzzy Sets," *Information and Control*, Vol. 8, No. 2, pp. 338-353, 1965.

On Improving Estimates of User Sessions from Web Server Logs

Amithalal Caldera, Yogesh Deshpande
School of Computing and Information Technology
College of Science, Technology and Engineering
University of Western Sydney
PO Box 1797, Penrith South DC, NSW 1797, Australia
h.caldera@uws.edu.au, y.deshpande@uws.edu.au

Abstract

The standard formats of Web server logs capture the browsing behaviour of Web site users without explicitly identifying them. Individual user session, i.e. a sequence of requests to a server from a single user within a certain time window, can only be inferred by analysing the logs heuristically and under various assumptions. Some heuristics estimate the number of user sessions using a field called ***referrer****, which refers to the page containing the link that a user clicked to generate the request. These heuristics are adversely affected by Web caching and undefined referrer entries in the logs. In this paper, we propose a new heuristic to overcome these two factors. The heuristic is evaluated by analysing the logs of a university Web server that explicitly records user ids, which help to identify user sessions exactly. We also evaluate our heuristic against other existing heuristics. The evaluation has shown satisfactory result.*

Keywords:

Heuristics, user session identification, mining, log mining, web usage mining, log analysis, referrer heuristic

1. INTRODUCTION

With the transformation of the Web into major platform for delivery of applications and information [Bernstein *et al.* 1998], it has imperative for organizations and companies, who have invested millions in Internet and Intranet technologies, to track and analyse their users. Users in their Web navigation make requests to Web servers through Web browsers. A sequence of requests from a single user within a certain time window is called *a user session*. User sessions, if available, carry valuable information about the user interests [Shahabi *et al.* 1997] and patterns of usage. There is a general realization that ease of navigation can greatly determine the success of Web sites and consequently of the businesses and activities they represent [Heer and Chi 2002]. Knowledge of user sessions has numerous applications in the Web such as personalized services [Mobasher *et al.* 2000; Daniel and Rosalee 2003], adaptive Web sites [Xiong *et al.* 2000; Perkowitz and Etzioni 1997], Evaluation of Web sites [Spiliopoulou 2000] and Web cache enhancement [Yang and Zhang 2003].

At any point in time, many users, each with his or her own goals and concerns, visit Web sites and their requests are captured mainly in Web server logs. The Web Server log is a record of Web server activity and consists of details about file requests to a Web server and the server responses to those requests. A typical Web server log contains information such as the IP address of the machine that made the request, the date and time a request was made, the request method that the client used, the URL of the requested page, the status code of the response message, the size of the document transferred, the protocol used, the user agent, which is the application software used to browse the Web and the URL of the referrer page from which the request was initiated. For example, consider the following typical server log entry.

"2003-06-04 10:00:02 210.49.90.176 GET /iwsd/header.htm 200 2654 HTTP/1.0 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+98) http://tl28serv.uws.edu.au/iwsd/default.htm".

This tells us that on June 04, 2003, a request was received at 02 seconds after 10:00 AM from the machine with IP address 210.49.90.176, for the file /iwsd/header.htm using HTTP/1.0. The status code "200" indicates that the server successfully responded to the request and set 2654 bytes as the response to the client. The agent, i.e. the browser, used to view the file was MSIE 6.0 (Microsoft Internet explorer 6.0) compatible, running on Windows 98 and the request originated from a link, i.e. referrer, http://tl28serv.uws.edu.au/iwsd/default.htm.

The entries in a Web server log are explicit records of the browsing behaviour of site visitors, which, by default do not identify them. The requests of a single user intermingle in the server logs with those from other users. The difficulty of identifying user sessions from Web server logs was addressed in [Pirolli *et al.* 1996; Pitkow 1997]. Since then, many researchers have proposed various methods to resolve this problem. But, most of these methods required additional configuration of the server to store more information in the server log such as cookies [Kamdar and Joshi 2000] and logins. When logins are not desirable and cookies are blocked by browsers or firewalls, most of the approaches used to identify user sessions through the Web server logs are heuristic. The heuristic strategy only exploits the available information recorded in the server logs to assess whether the requests registered by the Web server were performed during the same visit of the individual. The referrer information is useful to determine which sites or pages generate traffic to and how users move among pages on the site. The heuristics, which exploit the referrer to infer user sessions, are directly affected by the use of Web caching and the existence of undefined referrers in logs, as explained below. In our previous work [Caldera and Deshpande 2005], we proposed a new robust heuristic to overcome the problems and investigated its performance against two existing heuristics. In this paper, we propose another new heuristic and investigate its robustness against the caching and referrer problems and accuracy against all the available heuristics.

The paper is structured as follows. Section 2 explains limitations involved in analysing the logs. Section 3 describes briefly the earlier heuristics and their approaches to address those issues. Section 4 presents our proposed heuristic, which addresses the issues of Web caching and empty referrer entries too. Section 5 describes the environment in which the Web server logs were generated. These logs were gathered from a university Web server that recorded student ids for administrative purposes, in addition to the normal details of a Web server log. Section 6 describes the methodology applied to test the accuracy of the proposed heuristic and reports the result of the experiments conducted. The analysis of the result is given in section 7. The logs used in the current study facilitate comparisons of heuristics amongst themselves and also against the direct records of user sessions based on user ids. Section 8 concludes the paper.

2. LIMITATIONS IN LOG ANALYSIS

The heuristics, which exploit the referrer to infer user sessions, are affected by the use of proxy servers, caching, stateless service model of the HTTP protocol and empty referrers.

The process of user session identification is usually performed mostly based on the IP address [Berendit and Spiliopoulou 2000]. Due to the use of proxy servers by Internet Service Providers (ISPs), true client IP addresses are not always available to the Web server when a user initiates a request through ISP. In such cases, the IP address of a proxy server will be recorded in the server logs. If several users come through the same ISP, one IP address may represent all of them. It is also not rare that the Web sites use multiple servers running on different computers to handle heavy traffic, in which case a reverse proxy server could direct the internal traffic. In fact, this was one of the techniques used in the Web site for the 1998 Winter Olympic Games in Nagano, Japan, to control the heavy traffic received around the world [Iyengar *et al.* 2000]. On the other hand, some ISPs use multiple proxy servers and requests from the same user may go through different proxy servers. As there are many users at any one time, requests made by a single user (in a single visit) may be pointed to different servers. Thus, requests made by a single user (in a single visit) may have multiple IP addresses and one IP address may 'hide' multiple users.

In order to improve performance and minimize network traffic, most Web browsers cache the pages that have been requested. People revisit a considerable number of Web pages that they have viewed previously during their Web browsing and they heavily use "Back" button provided at the top of the popular browsers to revisit pages [Catledge and Pitkow 1995; Tauscher and Greenberg 1997; Greenberg and Cockburn 1999]. People also revisit pages by clicking a hypertext link that has been traversed in the current session or directly accessing a page from recent *history* list provided by the browser. The proxy servers also cache the pages that are often requested by users. Unlike the browser caching, due to proxy caching, multiple users could actually view the results of a single request from the server. All the requests come to the pages that have been cached may be served from the cache rather than from the Web server where the document originally resided. As a result, the Web server is not aware of the cached accesses and there are no corresponding log entries for those accesses. This has made the server log incomplete for the successful implementation of any referrer based heuristic. Inferring the missing cache hits in the server log is a non trivial problem.

Web uses the HyperText Transfer Protocol (HTTP) [Fielding *et al.* 1998] for communication between the clients and the servers on the Internet. HTTP is a simple, connection-oriented, application level protocol that employs the Transmission Control Protocol/Internet Protocol (TCP/IP) [Postel 1981] to manage connections and provide reliable bi-directional communication channels between the clients and the servers. Communication is initiated by a Web client in the form of a request to the Web server. After TCP/IP has established a communication channel, the client sends a request to the server. The server parses the request received from the client, retrieves the requested information, and returns it to the client. Every request results in a response from the server. After a response is received, the TCP/IP connection is usually closed. If the same client makes a new request, the server does not know what request the browser made previously or what state the browser was left in. This means that every request from the client to the server is treated independently and information from previous connections is not maintained for use in future connection [Iyengar 1997]. Hence, HTTP is stateless and does not allow support for establishing long-term connections between the Web server and the browser/user. The requests of a single user are recorded in the server logs nested with the requests of other users. Clustering of server log entries into its sessions is not straightforward.

A referrer is the URL of the page the client was on before requesting the current page and it was designed for link backtracking and verification. It is useful in determining which sites are driving traffic to, and how users move among pages on the site. Most of the popular browsers usually send referrer information with the HTTP request only when the link or image is actually embedded in an HTML page. However, the browser is not aware of the referrer if users type in URLs directly in the address location or use bookmarks from their favourite list to retrieve a page without using a site's link structure. In this case, nothing is passed as the referrer and the empty referrer is recorded in the server log as "-". The referrer may also not be available for the following three reasons: (1) a robot performs the requests and it does not follow hyperlinks (2) the user has certain privacy features activated to block the referrer: The referrer page can come from other Web sites, which sometimes encode part of the information that the user has provided to them in the referrer. (3) the referrer field of the first entry of most of the sessions is not defined. The presence of empty referrers in the server logs has also made the exploitation of the referrer information to infer user sessions non trivial.

3. RELATED WORK

Prior to 1997, the process of identifying user sessions from Web server logs was usually carried out mostly based on the IP address. Cooley et al [Cooley *et al.* 1997; Cooley *et al.* 1999] proposed four heuristics for the attribution of requests to different users. We denoted as **h0,** the heuristic that identifies user sessions by IP address alone and the four heuristics proposed by Cooley et al as **h1**, **h2**, **h3** and **h4**.

Heuristic **h1** states that *each different user-agent type for an IP address represents a different user.* A user-agent identifies the browser version and the operating system. The rationale here is that a user rarely employs more than one browser when navigating the Web. Hence, a user session is defined by aggregating accesses on unique user agents for an IP address. However, the heuristic ignores the possibility that requests made by a single user (in a single visit) may have multiple IP addresses and the requests of different users may have the same IP address, as discussed in section 2 above. The heuristic **h2** is based on the user's navigational behaviour on the site. It states that i*f a Web page is requested and this page is not reachable from previously visited pages, then the request should be attributed to a different user. A new session is also suspected if the referrer is undefined*. The rationale behind this heuristic is that users generally follow links to reach a page. However, two users can easily be confused as a single user if they are looking at the same set of pages. Conversely, if a user types in URLs to reach pages not connected via links, then the absence of referrer can be misconstrued to represent a different user. The heuristics **h3** and **h4** are based on the time that the user spends on the site or makes his/her next request. The heuristic **h3** states that *the duration of a session must not exceed a pre-specified threshold*. The threshold is an upper bound on the time spent in the site during a visit. The duration is determined externally, based on generic observation of user behaviour or subjectively. The heuristic **h4** states that *a new session is suspected if the time spent on a page exceeds a pre-specified time threshold*. Users who do not request pages within a certain time limit are assumed to have left the site.

Researchers have used some or all of the above heuristics to identify user sessions. In [Yao *et al.* 2000], IP and user agent are used to identify unique users and a time period of 6 hours is used as an upper bound for a session (**h1** and **h3** together). In [Berendit and Spiliopoulou 2000], the IP address and the user agent are used to identify the sequences of requests and an upper bound of the session duration as a whole is used to split these sequences into sessions (**h1** and **h3** together). In [Wu *et al.* 1998], Wu et al assumed that different IP addresses or user agents indicate different user sessions, and if the timestamps indicate that two accesses are separated from by a more than a pre-specified period of time, the accesses are also considered to belong to different sessions. In addition to this, they used the referrer page and the requested page to identify user sessions (**h1, h2** and **h4**). In [Pirolli *et al.* 1996], new sessions are created for each combination of IP address and user agent when the time between two consecutive requests is greater than the pre-specified threshold. New sessions are also created when the requested page is not connected to the last page in the session (**h1**, **h2** and **h4** together).

Huang et al [Huang *et al.* 2004] proposed n-gram language model to identify sessions. In their work, an n-gram language model, which has been widely used in statistical language modelling for speech recognition, was applied to identify session's boundaries. They assumed that a user's next step depends only on his/her previous accesses and a significant decrease of the probability of a page sequence indicates a clear signal of a session boundary. However, in their approach, the problem of interleaved sessions is expected to be solved in advance based on the user's IP address. When the IP address is not reliable to identify a user, the performance of this approach will decrease.

In our previous work [Caldera and Deshpande 2005], we presented three new base heuristics termed as H1, H2 and H3 to identify user sessions from server logs and reported the efficiency of them as a one single heuristic. Although there were many existing heuristics available, the efficiency was tested against only two of them. In this paper, we are consolidating our previous work by introducing another new robust heuristics originated from our previous work and presenting more comprehensive and systematic approach in testing it.

4. PROPOSED HEURISTIC

A new heuristic proposed in this paper is a combination of three heuristics, termed as H1, H2 and h4. The first two heuristics are the results of our own analysis of Web server logs and introduced briefly in [Caldera and Deshpande 2005]. For the completeness of this paper and presenting more rigorous arguments, we introduce them here with some comprehensive illustrations. The last is the existing heuristic described in section 3.

H1: *Each different user agent for a domain of the IP addresses represents a different user sessions. A domain referred here identifies a common subnet where a group of IP addresses may belong.*

As explained in the previous section, it is both possible that a single IP address appears in multiple sessions or multiple IP addresses appear in a single session. In the former case, true IP addresses of different users are covered with a single proxy address. In this situation, instead of the domain, IP address itself can also be used in the heuristic, which is then equivalent to the heuristic h1 explained above. In the latter case, the true IP address of a single user is hidden by the multiple IP addresses of multiple proxy servers. As these IP addresses belong to the same ISP, it is assumed that all the IP addresses share the same domain. For example, consider a sample log shown in Table 1. It shows a part of log entries in one-day log, which represent the page requests coming behind the two proxy servers with IP addresses, 137.154.189.12 and 137.154.189.13, where 137.154.189 represents the common domain. Upper case letters, A,B, … are used to represent the requested and referrer pages. The heuristic identifies two user sessions represented by log entries (1,2,3,4,7,9,11,12,13) and (5,6,8,10) as each session uses a unique combination of user agent and domain. Note that h1 does not handle the problem of multiple IP addresses appearing in a single session.

Table 1. A sample of server log

#	IP address	Time	URL	Referrer	User Agent
1	137.154.189.12	03:04:41	A	-	Mozila/4.0+(IE+6.0+Win2000)
2	137.154.189.13	03:05:34	B	A	Mozila/4.0+(IE+6.0+Win2000)
3	137.154.189.12	03:05:39	L	K	Mozila/4.0+(IE+6.0+Win2000)
4	137.154.189.13	03:06:02	F	B	Mozila/4.0+(IE+6.0+Win2000)
5	137.154.189.12	03:06:58	A	-	Mozila/4.0+(IE+6.0+Win98)
6	137.154.189.13	03:07:42	B	A	Mozila/4.0+(IE+6.0+Win98)
7	137.154.189.12	03:07:55	R	L	Mozila/4.0+(IE+6.0+Win2000)
8	137.154.189.12	03:09:50	C	A	Mozila/4.0+(IE+6.0+Win98)
9	137.154.189.13	03:10:02	O	F	Mozila/4.0+(IE+6.0+Win2000)
10	137.154.189.13	03:10:45	T.	C	Mozila/4.0+(IE+6.0+Win98)
11	137.154.189.12	03:12:23	G	B	Mozila/4.0+(IE+6.0+Win2000)
12	137.154.189.13	03:15:22	A	-	Mozila/4.0+(IE+6.0+Win2000)
13	137.154.189.12	03:16:03	D	A	Mozila/4.0+(IE+6.0+Win2000)

On further analysis, it may be argued that H1 will suffer in performance as the market for both browsers and operating systems gets ever more consolidated because it is highly likely that different users coming behind the same ISP will have the same user agent. For instance, most users may use the same version of Microsoft Internet Explorer running on a Windows 2000 computer. As such, H1 is unable to decipher this situation. Heuristic H2 further checks each session identified by H1 for the validity of the membership of its entries within itself.

H2: Let p and q be two consecutive page requests in a session S identified by heuristic H1. Let also r(q) and ip(q) be the referrer and IP address of q respectively. The membership of q in S is confirmed if one of the following three conditions is satisfied.

1. If r(q) is equal to r(p) or p , or r(q) was previously invoked within S
2. If r(q) is undefined and q was previously invoked within S
3. If r(q) is undefined and ip(q) does not represent a common proxy server

Otherwise, q belongs to a new session.

Assuming that users follow hyperlinks to reach a page, each access pair of the referrer page and the requested page constitute a connected traversal path. That is, if none of the pages is brought from the cache, *r(q)* should be equal to *p*. A proxy server is said to be common if numerous overlapping requests from multiple users come behind it. In the site under our inspection, large numbers of students come behind university proxy servers within a very small period of time during a peak period such as assignment uploading. In fact, two proxies shown in Table 1 represent the university proxy servers. Consider the log entries contained in the first session (1,2,3,4,7,9,11,12,13) identified by H1. The heuristic H2 further divides this session to two sessions, (1,2,4,9,11,12,13) and (3,7) as the third entry, page L, is not directly reachable from pages A or B and the seventh entry, page R is only reachable from page L.

If *r(q)* is not equal to *p*, we assume that the user must have accessed a cached page connecting *p* and *q*. These hits are missed in the server log. But these missing cache hits must have accessed the server in the recent past. Hence, to confirm the memberships, heuristic H2 does not need the requested page to be accessible from the page immediately accessed before it. For example, consider the log entries contained in the session (1,2,4,9,11,12,13) identified above. Cache hits probably occurred between log entries 9 and 11. The referrer page of page G at log entry 11 is not equal to the requested page O at 9. The user has backtracked to a cached page B to access a new page G. The log does not show any of these cache accesses at the request page field but it has captured the last cached page at the referrer field. Thus, the heuristic has suggested that the log entry 11 still belongs to the same session. The heuristic h2 described in section 2 is similar to this but it has dealt with the problem of missing cache hits by consulting a site topology (site structure). Consultation of the site topology presupposes the availability of the site's graph in an appropriate format.

When *r(q)* is undefined, the heuristic assures the membership of *q* in the session depending on whether *q* was previously invoked or *ip(q)* does not represent a common proxy server. The requests where the referrer is undefined are often made through bookmarks or typing URLs. It can be argued that the request page of this type might have been previously visited in the recent history. A proxy server is said to be common if numerous overlapping requests from multiple users come behind it. If *ip(q)* does not represent such a common proxy server, it is assumed that q should be part of *S* as each of the requests of *S* carries the same user agent and the same domain or IP address.

5. EXPERIMENTAL ENVIRONMENTS

The Web logs used in this investigation came from a university Web server used exclusively for teaching. The server runs Microsoft Internet Information Server 5.0 on Windows 2000 advanced server platform. The students have to create a Web site each and then learn scripting for both client-side and server-side processing. Each student is given an id and Web space. Students access the server from the labs within the university or from outside, using either the university dial-up lines or some ISP. The university routes its traffic through two proxies. The university semester runs for 16 weeks during which time the students typically complete two assignments, some quizzes and a mini-project each. We chose the first semester of 2003, viz. March-June 2003 that was the latest semester when we launched our experiments. Approximately 500 students used the server. We examined all the Web server logs of the semester (109 daily logs). The server logs are created daily, starting at 10.00 AM and go on for the next 24 hours and the size of the log files ranged from over 1 MB to more than 77 MB.

6. METHODOLOGY AND RESULTS

Web server logs, in general, easily reach tens of megabytes per day. For proper analysis, they have to be 'cleaned' prior to the session identification process based on heuristics discussed above. The 'cleaning' process employed here performs the following tasks. First, the log entries referring to images are removed, based on the suffixes such as gif, jpg, jpeg, and png. Most Web pages contain images, which either serve design purposes (for example, lines and buttons) or present information (for example, graphics, maps and pictures). Whether to keep or remove the log files for these images depends much on the purpose of analysis. As this paper focuses on a reconstruction of the server logs into user sessions, the log entries referring to images are removed. Second, log entries with server response codes 4xx and 5xx are removed since they are client and server errors respectively. Third, robot accesses in the server logs are removed. In general, robot refers to any programmable software agent that does not access a site interactively. These requests can distort the process of user session identification, as they do not reflect the way human visitors navigate the site. Finally, the entries for the system administrator and tutors who mark the student assignments are eliminated, since we want to analyse only student sessions. The cleaned data is the base to which the heuristic is applied.

The performance of the new heuristic is first evaluated by comparing the number of sessions produced by it to the 'exact' number of sessions, derived from a combination of explicit user id and the heuristic h4 described in section 3 to split the students' activities. This combination is assumed to be the best approximation for the exact sessions as user id uniquely identifies every user. The 'exact' method is called M1 and the method of constructing sessions using the new heuristic M2. Then, the new heuristic is compared with existing heuristics often used in heuristic based approaches, similarly. The methods of constructing sessions using these heuristics are shown in Table 2. The base heuristics, h0, h1, h2, h3 and h4 shown in the table were described in section 3. The threshold parameters of h3 and h4 are set to 30 and 15 minutes respectively.

Table 2. The Existing Methods

Methods	Heuristic
M3	h0
M4	h0, h2
M5	h0, h3
M6	h0, h4
M7	h1
M8	h1, h2
M9	h1, h3
M10	h1, h4
M11	h1, h2, h3
M12	h1, h2, h4

Table 3 shows the statistics of number of sessions produced by each method and log entries before cleaning (Hits) and after cleaning (Views). In Table 3, some methods generated fewer sessions than were there, which leads to sessions that are, on average, longer than the real sessions and others generated more sessions, which leads to shorter sessions. We use the term underestimating for former phenomenon and the term overestimating for the latter. For example, M3 and M7, which use IP address and the combination of IP address and user agent respectively, underestimate M1 and others overestimate it. Note that M3 and M7 do not address the problem of using a single proxy server. While M3 merges all the sessions created behind the same proxy server into a single session, M7 mergers only the sessions created behind the same proxy server that have same user agent. Thus, M7 shows a lesser underestimate. While the usage of h2, h3 and h4 in other methods leads to split sessions, the tendency of splitting sessions by the referrer heuristic h2 is more severe. Thus, M4, M8, M11 and M12 use h2 and overestimate M1 by a large factor.

Table 3. Statistics on number of sessions and log entries

Item	Hits	Views	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Total	2201304	1508331	19694	23152	11705	109005	30052	21577	12756	109276	31083	22592	122617	116451
Avg/day	20196	13837	181	212	107	1000	276	198	117	1003	285	207	1125	1068
Max/day	273677	208070	1583	1904	630	14799	2915	1676	712	14827	3090	1827	16574	15713
Min/day	420	174	5	5	5	15	7	5	5	15	7	5	15	15

The impact of referrer heuristic h2 is best illustrated in Table 4, which shows the statistics for average sizes of sessions and pages. The time values shown in the table are recorded in hour:minute:second format. According to Table 4, the methods M3 and M7, which underestimate M1, have the lowest performance, as all their statistics are much higher than the others. Among the other methods, which overestimate M1, the average session lengths of the methods, M4, M8, M11 and M12, which use h2 are much lower. Indeed, they produced more than thrice as many sessions than really existed. It seems that the referrer heuristic split almost every sessions erroneously. Note that h2 doesn't address the problems of the existence of undefined referrers and the usage of caching.

Table 4: Average statistics on sizes of sessions and pages

Items	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Avg. session time	0:21:01	0:18:58	1:45:01	0:26:04	0:14:35	0:19:27	1:30:46	0:25:41	0:14:26	0:19:05	0:09:27	0:10:39
Avg. page time	0:00:26	0:00:27	0:02:40	0:01:07	0:00:38	0:00:29	0:02:45	0:01:04	0:00:38	0:00:29	0:00:29	0:00:23
Avg. session length (no. of pages)	32	28	52	9	15	28	48	9	15	27	6	7

According to table 3 and 4, the statistics for number of sessions, size of sessions and pages for M2, M6 and M10 are very much closer to M1 than others and they are within the range of 2% – 15%. However, most individual methods bear both underestimating and overestimating effects. Therefore, we cannot draw definite conclusions about the efficiency of these heuristics in approximating the exact sessions using only the basic usage statistics calculated above. We have investigated the significance of all the methods to M1 using Chi-squared test and the next section reports on it.

7. CHI-SQUARED TEST

A χ^2 test is used here to investigate the significance of the differences between two distributions based on two samples spread over k classes. The samples refer to the methods M1 to M12 and the k classes refer to the 109 daily logs under inspection. The test methods M2 to M12 are each compared to the exact method M1. The quantity χ^2 describes the magnitude of discrepancy between the exact and the test method. It is defined as:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where O refers to the observed frequencies (from the heuristics) and E refers to the expected frequencies (based on the exact method).

In order to test whether or not the two distributions are the same, we take the null hypothesis as H_0: Two samples are from same distribution and alternative hypothesis as H_1: Two samples are not from same distribution. The calculated value of χ^2 is compared with the table value for given degree of freedom (d.f) at 5% level of significance. If the calculated value of χ^2 is less than the table value at the specified significance level, the results of the test do not provide justification to reject the null hypothesis. In other words, the hypothesis that two samples are from same distribution holds good. On the other hand, if the calculated value of χ^2 is more than the table value at the specified significance level, the results of the test lead to the rejection of the null hypothesis or, in other words, two samples are not from same distribution.

χ^2 values calculated over 109 observations for each test are recorded in Table 5, which confirms that only the tests 1, 5 and 9 are insignificant. That is, only the methods M2, M6 and M10 are not significantly different to M1. Note that both M6 and M10 do not use h2. However, M2 is highly insignificant to M1 than M6 and M10 as M2 gives the smallest χ^2 value (largest P-value). Therefore, by considering the facts that M2 addresses all the limitations stated in section 2 and is highly insignificant to M1, we conclude that the new heuristic best approximates the exact method.

Table 5. Chi-squared test results

Test	Control	Test method	d.f	Observed χ^2 values	P-value	Significance at 5%
1	M1	M2	108	100.789	0.676	Insignificant
2	M1	M3	108	358.804	<0.001	Significant
3	M1	M4	108	4526.175	<0.001	Significant
4	M1	M5	108	351.106	<0.001	Significant
5	M1	M6	108	112.572	0.362	Insignificant
6	M1	M7	108	315.024	<0.001	Significant
7	M1	M8	108	4515.417	<0.001	Significant
8	M1	M9	108	397.451	<0.001	Significant
9	M1	M10	108	106.076	0.534	Insignificant
10	M1	M11	108	4067	<0.001	Significant
11	M1	M12	108	4144.802	<0.001	Significant

The interpretation of the results is thus straightforward. The new heuristic, M2, based on additional consideration of the effects of Web caching and the existence of undefined referrers in logs, leads to identification of user sessions that are not significantly different from the actual sessions in the statistical sense. In other words, the sample generated by M1 and M2 come from the same distribution.

8. Conclusions and Further Work

This paper has proposed a new heuristic that exploits the referrer information recorded in the Web server logs to infer user sessions. It has illustrated that the new heuristic is robust in dealing with the problems of caching and undefined referrers, as detailed in the statistical analysis. The analysis has confirmed by conducting a comprehensive and systematic approach that the new heuristic outperforms all the existing heuristics and best approximates the exact method under the specific circumstances of a university Web server with a controlled environment.

Further work needs to be carried out to validate the heuristic generally. We are currently investigating the possible effects of different characteristics of Web sites on our heuristics. The site used to test the heuristics in this paper identified the users explicitly and thus we were able to establish the performance levels of all the heuristics against the one objective criterion. However, the server was within a university, established for and used by a section of students and thus did not contain the kind of information and documents that other Web sites would. We plan to test these heuristics on other sites where the types of users could be quite different as also the usage and traffic pattern. Most of the log analysis tools generate reports aim at providing information on the activity of the server rather than the user. The server only tracks the pages requested through it. There is a strong need for a tool that can analyse user-oriented behaviour from the Web server logs without requiring user registrations or log-ins or the employment of cookies as many users wants to remain as anonymous as possible in their Web navigations. Our future work also includes evaluating and using the heuristic as a user-oriented log analysis tool.

REFERENCES

[Berendit, B. and M. Spiliopoulou, 2000] Berendit, B. and M. Spiliopoulou, "Analysis of Navigation Behaviour in Web Sites Integrating Multiple Information Systems", VLDB Journal, **9**(1), 56-75, 2000

[Bernstein, P., *et al.*, 1998] Bernstein, P., *et al.*, "The Asilomar Report on Database Research", ACM SIGMOD Record, **27**(4), December, 1998

[Caldera, A. and Y. Deshpande, 2005] Caldera, A. and Y. Deshpande, "Effectively Capturing User Navigation Paths in the Web Using Web Server Logs", In the Proceedings of Fifth International Conference in Web Engineering (ICWE), **3579**, p. 63-68, Lecture Notes in Computer Science, Springer-Verlag, Sydney, Australia, July 25-29, 2005

[Catledge, L. D. and J. E. Pitkow, 1995] Catledge, L. D. and J. E. Pitkow, "Characterizing Browsing Strategies in the World-Wide Web", In Proceeding of the 3rd International World Wide Web Conference on Technology, tools and Applications, 1065 -1073, Darmstadt, Germany, 1995

[Cooley, R., *et al.*, 1997] Cooley, R., *et al.*, "Grouping web page references into transactions for mining world wide web browsing patterns", In Knowledge and Data Engineering workshop, 2-9, 1997

[Cooley, R., *et al.*, 1999] Cooley, R., *et al.*, "Data preparation for mining world wide web browsing patterns", In Journal of Knowledge and Information Systems, **1**(1), 5-32, February, 1999

[Daniel, M. C. and W. Rosalee, 2003] Daniel, M. C. and W. Rosalee, "User-Centered Web Site Development: A Human Computer Interaction Approach: Chapter 14: Personalization and Trust", **ISBN: 0130411612**, Prentice Hall, 2003

[Fielding, R., *et al.*, 1998] Fielding, R., *et al.*, "Hypertext Transfer Protocol - HTTP/1.1", HTTP Working Group, Internet Draft, November, 1998

[Greenberg, S. and A. Cockburn, 1999] Greenberg, S. and A. Cockburn, "Getting Back to Back: Alternate Behaviors for a Web Browser's Back Batton", In proceedings of the 5th Annual Human Factors and the Web Conference, Gaithersburg, Maryland, USA, June 3rd, 1999

[Heer, J. and E. H. Chi, 2002] Heer, J. and E. H. Chi, "Separating the Swarm: Categorization Methods for User Sessions on the Web", Computer Human Interaction (CHI 2002), Minneapolis, Minnesota, USA, April 20-25, 2002

[Huang, X., *et al.*, 2004] Huang, X., *et al.*, "Dynamic Web log session identification with statistical language models", Journal of the American Society for Information Science and Technology (JASIST), **55**(14), 1290-1303, 2004

[Iyengar, A., 1997] Iyengar, A., "Dynamic Argument Embedding: Preserving State on the World Wide Web", IEEE Internet Computing, **1**(2), 50-56, March-April, 1997

[Iyengar, A., *et al.*, 2000] Iyengar, A., *et al.*, "High-Performance Web Site Design Techniques", IEEE Internet Computing, **4**(2), 17-26, March-April, 2000

[Kamdar, T. and A. Joshi, 2000] Kamdar, T. and A. Joshi, "On Creating Adaptive Web Servers Using Weblog Mining", Technical Report TR-CS-00-05, November, 2000

[Mobasher, B., *et al.*, 2000] Mobasher, B., *et al.*, "Automatic Personalization Based on Web Usage Mining", Communications of the ACM, **43**(8), 142 -151, August, 2000
[Perkowitz, M. and O. Etzioni, 1997] Perkowitz, M. and O. Etzioni, "Adaptive Sites: Automatically Learning from User Access Patterns", The Sixth International WWW Conference, Santa Clara, California, USA, April, 1997
[Pirolli, P., *et al.*, 1996] Pirolli, P., *et al.*, "Silk from a Sow's Ear: Extracting Usable Structures from the Web", In Proceedings of 1996 Conference on Human Factors in Computing Systems (CHI-96), 1996
[Pitkow, J., 1997] Pitkow, J., "In Search of Reliable Usage Data on the WWW", In The Sixth International World Wide Web Conference, Paper126, 1997
[Postel, J., 1981] Postel, J., "Transmission Control Protocol", RFC 793, September, 1981
[Shahabi, C., *et al.*, 1997] Shahabi, C., *et al.*, "Knowledge discovery from users web-page navigation", In Workshop on Research Issues in Data Engineering, April, 1997
[Spiliopoulou, M., 2000] Spiliopoulou, M., "Web Usage Mining for Web Site Evaluation", Communication of ACM, **43**(8), 127-134, 2000
[Tauscher, L. and S. Greenberg, 1997] Tauscher, L. and S. Greenberg, "Revisitation Patterns in World Wide Web Navigation", In ACM SIGCHI '97 Proceeding of the Conference on Human Factors in Computing System, 399-406, Atlanta, Georgia, USA, March 22-27, 1997
[Wu, K.-L., *et al.*, 1998] Wu, K.-L., *et al.*, "SpeedTracer: A Web Usage Mining and Analysis Tool", IBM Systems Journal, **37**(1), 89-105, 1998
[Xiong, H., *et al.*, 2000] Xiong, H., *et al.*, "Adapting the Right Web pages to the right users", In SPIE Proceeding of International Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology II, **4057**, 380-387, Orlando, Florida US, April, 2000
[Yang, Q. and H. H. Zhang, 2003] Yang, Q. and H. H. Zhang, "Web-Log Mining for Predictive Web Caching", IEEE Transactions on Knowledge and Data Engineering, **15**(4), IEEE Computer Society, July/August, 2003
[Yao, Y. Y., *et al.*, 2000] Yao, Y. Y., *et al.*, "PagePrompter: An Intelligent Agent for Web Navigation Created Using Data Mining Techniques", Technical Report CS-2000-08, Department of Computer Science, University of Regina, 2000

State of Business Intelligence in the Large Finnish Companies

M.Sc.Virpi Pirttimäki

Tampere University of Technology

Tampere, Finland

Email: virpi.pirttimaki@tut.fi

M.Sc. Antti Koskinen

Novintel Inc.

Helsinki, Finland

Email: antti.koskinen@novintel.com

Abstract

In the literature, there are several terms which contain the idea of the systematic collection, analysis, storing, and dissemination of information. In Finland, the term business intelligence (BI) is by far the most popular name for these kinds of activities. The main objective of the paper is to consider the state of BI in the top-50 Finnish companies. The paper also presents and analyzes the key results of two surveys on BI carried out in 2002 and 2005 respectively. The empirical data was gathered using telephone interview. The responses received showed e.g. that the most important benefits that Finnish companies have gained through BI include improved quality of information, improved internal information dissemination, and increased level of awareness. In addition, the responding companies found out that identifying critical information needs is the most important development issue in the case of their BI activities.

Keywords:

Business Intelligence, Comparison, Large Finnish Companies, Survey

INTRODUCTION

Background of the Research

Ståhle and Grönroos (1999) have stated that the lack of information capital is often the main cause for the failure of business strategies. Concurrently, the amount of information at hand is increasing apace with advances in information and communication technologies. It is not sufficient just to gather and produce large amounts of information; the information also has to be refined and analyzed effectively. For example, Thierauf (2001) has claimed that the real problem is how to filter the essential information from a stream of information to produce knowledge for identifying notable events, predicting difficulties, and perceiving opportunities.

In the literature, there are a number of terms containing the idea of systematic information collection, analysis, storing, and dissemination (see e.g. Fleisher, 2001; Fuld, 1995; Kahaner, 1996; Vibert, 2004). However, some of these terms seem to overlap (see e.g. Hannula & Pirttimäki, 2005). In reality, organizations use multiple terms for these activities, too. One reason for this inconsistency is that although these activities, in one form or another, have existed as long as there have been companies, the boundaries between different terms have not yet been very well established. In Finland, the term business intelligence (BI) is the most popular name for these kinds of activities (Pirttimäki, 2002; Hannula & Pirttimäki, 2003; Hirvensalo, 2004; Knip & Fleisher, 2005; Koskinen, 2005). Namely, most Finnish companies consider BI to be a systematic process used to monitor the entire business environment not only competitors or customers, and to handle both internal and external information. Because of this, the term BI is used hereafter in this study.

The role played by BI has recently expanded considerably in Finnish companies (Pirttimäki, 2002; Hannula & Pirttimäki, 2003; Hirvensalo, 2004; Knip & Fleisher, 2005; Koskinen, 2005). According to Hirvensalo (2004), a real increase in Finnish BI activities began in the middle of the 1990s. In 1995, Finland became a member of the European Union and conceptions regarding the business environment and the market areas of Finnish companies became more internationally oriented. At least partly because of the more international business environment, BI activities and markets have rapidly increased in Finland during the past few years (Hirvensalo, 2004; Hannula & Pirttimäki, 2003). Nowadays, there are several different firms offering BI consultancy and BI system development in Finland. The common mission of all the actors seems to be to refine and provide business information.

In addition to its place in business life, BI as a research area has gained ground in the academic world (see e.g. Dishman et al., 2003; Knip & Fleisher, 2005). In Finland, academic BI research was introduced at the Institute of Business Information Management at the Tampere University of Technology in 2000. The Institute and Business Intelligence Forum Finland carried out the first comprehensive study on BI activities in Finland in the summer of 2002 (Pirttimäki, 2002; Hannula & Pirttimäki, 2003). The study was repeated at the beginning of 2005 (Koskinen, 2005). In addition to academic research, commercial research units have carried out a few BI oriented studies in Finland (see e.g. Korhonen & Horttanainen, 2002; Market-Visio Inc., 2001; Market-Visio Inc., 2002; Vesa, 2002).

Research Objectives and Methods

The main objective of the paper is to explore how BI activities have changed and developed in major Finnish companies over three years (2002 – 2005). The paper also presents and analyzes the main results of two Finnish surveys on BI. Overall, the paper describes what BI represents for the large Finnish companies at the beginning of the third millennium.

The research paper is based on two surveys which have been carried out in Finland in 2002 and 2005 (Pirttimäki, 2002; Hannula & Pirttimäki, 2003; Koskinen, 2005). Both surveys were descriptive in nature. Telephone interview was used as the primary research method in both studies. Individuals responsible for BI in the top-50 Finnish companies by net sales were interviewed. The lists of the companies were based on the Finnish business magazine *Talouselämä*'s top 500 lists in the spring of 2002 and 2004 (Saarelainen & Vanhanen, 2002; Lilius, 2004). The top50 company list has not changed radically in three years. Namely, 90 per cent of the companies are the same, and thus the comparison between these two studies should be relevant. At each survey, the companies were categorized by the line of business: industry, trade and services, and information and communication technology (ICT).

Only one respondent per company was interviewed. Thus, the results of the studies are an aggregate of a number of individual subjective views and this paper is based on these possibly subjective results. In order to increase the response rate, every respondent was guaranteed total anonymity in both surveys. The interviewees were sent a cover letter by e-mail before the telephone interview. The questionnaire was attached to the e-mail for the respondents' advance information. The sampling plan of the studies required that the BI experts answered a multi-page questionnaire during the telephone interview. The interviewers navigated the respondents through the questionnaire, and assisted when necessary. The duration of the interviews varied between 20 and 30 minutes.

In both surveys, the questions were derived from the research questions and the theoretical framework used in the study in question (see Appendix 1). In addition, the 2005 questionnaire was heavily based on the 2002 survey to ensure comparability of the results. The questions were mainly multiple-choice questions but there were also a few open-ended questions. Obviously, the responses to the open-ended questions were not unambiguous and more difficult to quantify than the answers to the multiple-choice questions. However, the authors of the surveys have drawn some conclusions by identifying repeatedly appearing themes in these responses. Thus, the results were analyzed with qualitative analysis methods. The results of the multiple-choice questions were analyzed utilizing statistical analysis methods. In the 2002 survey, there were 16 questions while the 2005 survey had 19 questions. There is therefore a lot of empirical data and this paper presents only the main findings of these two studies.

The first survey in 2002 attracted a total of 46 participants. Thus, the response rate reached 92 per cent forming a sound basis for the survey. Four companies declaimed to respond on grounds of trade secrecy or busy schedules. The objective of the survey was to explore the situation prevailing at the specific point in time when the research was conducted. The interviews took place between June 11 and July 3, 2002.

The 2002 survey showed that most of the large Finnish companies utilized BI in their decision-making. After three years, the interest was in identifying changes of developments in BI over that period. In 2005, the number of useful answers was 41, thus giving a response answering rate of 82 per cent. Nine companies declined to answer the survey. The main reason for refusal was lack of time. The interviews of the second survey were implemented in February, 2005.

MAIN RESULTS OF THE SURVEYS

Survey in 2002

According to the 2002 survey, 80 per cent of the companies interviewed had systematically organized BI. All of the ICT companies included in the survey reported a systematic way of collecting business information. In the industry sector, 82 per cent of the companies reported having systematic BI activities. There were more companies in the trade and services sector with no systematic BI than in the other sectors. It should be noted, however, that in the trade and services sector three quarters (74 %) of the companies did in fact have systematic BI.

A need for better information to support planning and decision-making and a need to obtain intelligence about the business environment and its development in support of operative actions were the most important reasons for introducing BI in the companies interviewed. Overall, the goal of starting activities in BI had mainly been to increase the use and sharing of existing information resources. One of the options evinced – active promotions by service providers – was not chosen by any of the interviewees. Hence, external application providers or consultants in the field did not have any particular role in initiating BI in the large Finnish companies. On the other hand, only a few companies named a single reason for introducing BI. Overall, it seems there have been several reasons for each of the companies initiating their BI.

In 2002, 95 per cent of the companies gathered information about their competitors and their own field of business. Over 80 per cent reported gathering information about customers and general macro-trends. Other important subjects were customers' fields of business or those related to the customers. Also, legal information, technical specifications, and country-specific information were considered relevant. The interviewees themselves named the following as essential subjects for information gathering:

- legislature work in the European Union
- the world market situation in various fields
- specific trends in the companies' own fields
- prices of raw-materials.

Almost one third considered information related to their suppliers and their fields less important. Logistics was the least important subject of monitoring for the companies interviewed (24 %).

The companies interviewed felt that the most notable benefits of BI were better-quality information for decision-making and improved ability to anticipate threats and opportunities earlier than before (see Figure 1). Namely, 95 per cent of the companies named gaining better-quality information for decision-making either as a very significant or as a significant benefit for their operations. Other important benefits were improved sharing and dissemination of information within organizations, improved efficiency, faster decision-making, and growth of a general organizational knowledge base. More than one third of the companies considered the time saved and easier acquisition and analysis of information a fairly notable benefit gained from BI. Over half of the companies (59 %) considered cost savings gained from BI to be either insignificant or not particularly significant. This may be because the cost-saving benefits of BI are often long-term. The benefits gained are also quite intangible and therefore often difficult – even impossible – to measure. (See e.g. Gartz, 2004; Simon, 1998; Lönnqvist & Pirttimäki, 2005; Pirttimäki et al., 2005.)

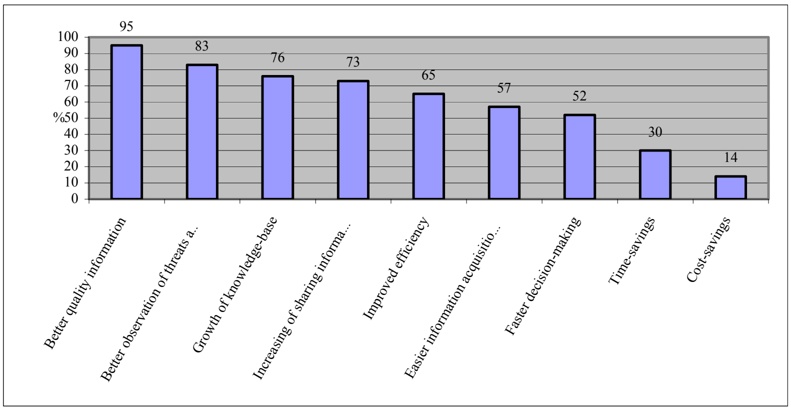


Figure 1. Benefits expected from BI activities.

Top-level management utilized the information and intelligence gained through BI in nearly all (95 %) the companies interviewed. Half of the companies named top management as the single most important user of BI. Middle management and specialists were also an important group of users in 85 per cent of the companies. In addition, 14 per cent named another user for BI, such as customers, sales and the development organization, and communications units.

Almost 80 per cent of the companies felt that crucial needs for information are not recognized at all or not well enough. Namely, identifying critical information needs was considered a notable area for improvement. The main reason for the need for development was that BI related operations were not systematic or comprehensive enough. Another crucial (70 %) issue of development was the utilization of internal information. In addition, nearly half of the companies emphasized that the diverse utilization of information was essential for competitiveness. There might be a lack of suitable integrated technologies and human resources, as well as inefficiencies in information management and acquisition. These were also considered notable targets for development. Twenty-two per cent of the companies stated their BI was in its infancy, even if the activities had been started several years ago. According to these responses, there again appeared a need for a more systematic approach, even if the basic requirements were already fulfilled. Figure 2 summarizes these targets for development.

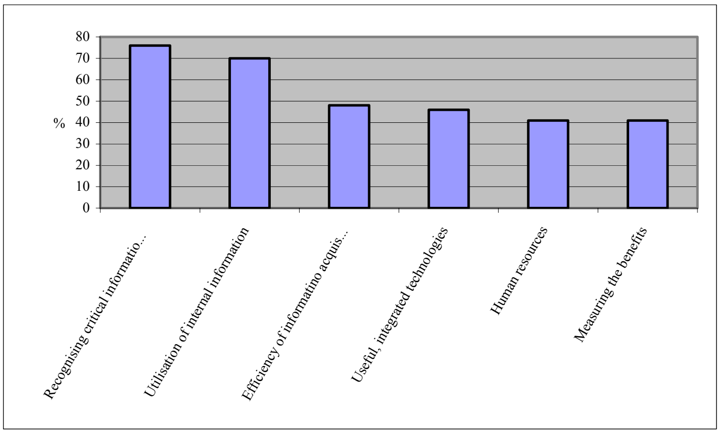


Figure 2. Most significant targets for development.

In 2002, one third of the companies believed that the emphasis on activities concerning BI would considerably increase by 2007. Almost half of the companies (44 %) suspected that the emphasis on BI would increase slightly and 20 per cent believed the emphasis would remain unchanged. Only two per cent predicted a slight decrease in emphasis, and none of the interviewees believed the emphasis would decrease significantly. The results suggest that the companies who already had BI also considered these activities important for the future. However, it was also possible that with some companies this was a sign of not having very mature techniques and methods or efficient information systems for the handling and utilization of information, and that investments were needed to improve this situation. Only the companies, who had had operations for BI for more than ten years, considered that investments in these operations would decrease in the future.

In 2002, over half of the companies believed and partly hoped that BI would become an integrated part of normal business operations by 2007. Only then would BI increasingly become an established and expanding element of the entire organization and no longer an isolated function. The interviewees saw BI in the future as an enabling activity contributing to the controlled management of businesses. On the other hand, the interviewees already called BI a crucial element in competitive business operations. Over half of the companies believed that outsourcing basic information acquisition would increase in the next five years, as there was a need to increasingly prioritize operations. The companies also expected the dissemination of information to improve when the contents of external information banks could be integrated into company portals. On the other hand, the companies questioned the increasing outsourcing of information acquisition if the price and quality -ratio of such services did not improve. Nearly 30 per cent wished the measurement of BI to develop by 2007. In 2002, measuring was not being done as either no suitable measures were identified, or the companies had no resources for such activity. With the help of appropriate measurement, BI activities could be more easily proved beneficial and valuable, for instance to a management board not yet committed to these activities.

Survey in 2005

In 2005, almost all companies (95 %) reported that they had systematically organized BI activities. The results for the ICT sector showed that all of the responding companies had systematically organized BI. In addition to the ICT sector, all of the companies in the trade and services sector had systematically organized activities in question. Only in two cases out of 41 (5 %), was BI either non-existent or could not be considered systematic. Both of these companies were industrial companies. Thus, 88 per cent of the companies in the industry sector had systematic BI.

The ranking of the most important information needs covered by BI showed that information about competitors was the most important (see Figure 3). The second most notable need was information about a company's own industry. In addition to these two, customer information was reported to be the third most essential information need. These three information needs were clearly the most important ones. Some of respondents pointed out that many of the information needs are intertwined; making it difficult to place them in order of importance. However, competitor information was seen to be important in positioning oneself in the competitive field. Some of the companies stated that it was important to be up-to-date on customers' needs and possible changes in those needs. Some companies had a customer-based strategy, making it natural to have a high regard for customer information. Other important information needs were macro-trends, country-specific information, and customers' fields of business. Technical information and information related to similar industries were considered less important.

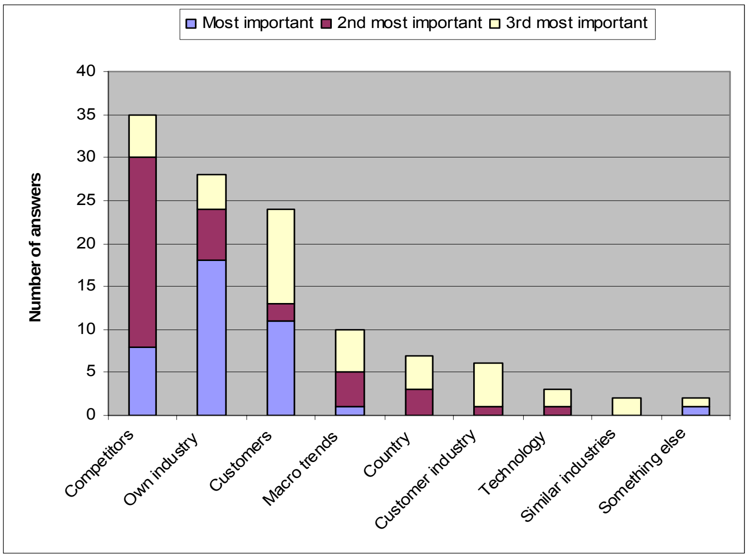


Figure 3. Illustration of the absolute value importance ranking of the most important information needs.

According to Figure 4, improved quality of information to support decision-making was the most important benefit achieved through BI. The second most important benefit was improved internal information dissemination, which is a natural consequence of using portals for example. The respondents felt that the general level of awareness had increased and that information gathering and analysis had improved through BI. They listed these respectively as the third and the fourth most important benefits. The companies interviewed *did not feel that cost savings achieved were among the top three benefits gained. It was more* surprising *that the companies did not list improved threat and opportunity detection as one of the three most important benefits gained through BI. Quite often, early signal detection is used as a strong selling point of BI activities. The low ranking may, in the estimation of the authors, indicate that in reality, companies are not able to fully utilize BI in this sense.* In the 2005 survey, the interviewees were also asked to give concrete examples of how they assessed the benefits of BI. Half of the interviewees said that they had end-user surveys in order to evaluate the benefits achieved. In addition, the interviewees mentioned that they measured time-savings, volume discounts by service providers, and utilization rates of BI tools. One in ten companies used open feedback in assessing the benefits. Around one in six companies did not measure the benefits of BI at all.

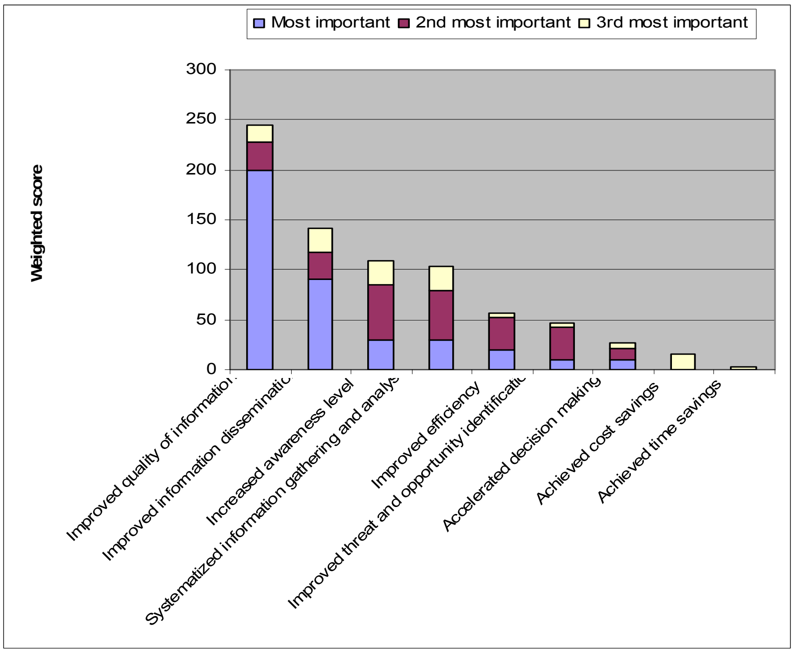


Figure 4. Most important benefits gained through the activity.

The information that companies acquire from different sources is often in a form that needs processing in order to be useful in decision-making. On average, 27 per cent of information processing was done outside the large Finnish companies and 73 per cent internally. None of the respondents admitted outsourcing all processing activities. The highest percentage reported by any company regarding outsourcing was 90 (reported by only one respondent). On the other hand, 33 per cent reported that only ten per cent or less of processing activities was outsourced. These results strongly indicate that the large Finnish companies tend to rely on their own resources for information processing. In addition, almost one in five respondents reported that 90 per cent or more of information processing was accomplished in a centralized BI unit.

Most companies (82 %) had an information system dedicated to BI activities. The companies which had a BI tool in place stated very clearly that the BI tool was used for disseminating information. A few interviewees reported that the BI tool was used as a data warehouse or as a portal of some sort. Obviously, portals can also be used to store information and distribute it. Some of the respondents reported one of their goals regarding BI to be eventually to have all the information, internal and external, in the same system instead of having several systems running simultaneously.

In the 2005 survey, the interviewees were asked to name the personnel groups utilizing information and intelligence gained through BI. In addition, the interviewees were asked to place these groups in order of importance with regard to the utilization volume of BI products. It was quite surprising that top management utilized BI products in 92 per cent of the companies, while middle management and experts did so in all of the companies interviewed. In 39 per cent of the companies, other employees also utilized BI products. However, 49 per cent regarded top management as the most important BI user group.

Most of the companies (83 %) regarded identifying critical information needs as the most important development issue. Over half of the companies identified utilizing internal knowledge and effectiveness of information gathering as belonging to the top three priorities in developing their BI. These three development issues emerged clearly above all others. From the authors' perspective, it is interesting to note that overcoming the resistance to change of personnel was not considered as a top-three development issue. Hence, it is probable that the personnel recognize the benefits of BI and are willing to utilize it. No did technology seem to be an issue in the sense that respondents seemed to know how to use it.

In 2005, 64 per cent of the companies thought that the investments for BI would increase in next five years. None of the responding companies intended any marked reduction in the level of investments. On the other hand, the trade and service sector was the only sector in which the companies predicted the investments in BI to increase significantly in the period of 2006 – 2011. The respondents were also asked to forecast how their BI activities would change by 2011. The following themes among others emerged from the investigation:

- advancing in the direction of knowledge management

- intensifying analysis further
- expanding geographic monitoring
- increasing coordination of BI activities
- outsourcing parts of BI activities
- combining internal and external information
- increasing the role of BI
- increasing utilization of BI in the formation of strategy.

The first two of the above-mentioned themes concern refining information into intelligence instead of focusing on information collection. The third one also contributes to creating more comprehensive intelligence by widening the scope of information gathering. Some companies described this trend of digging deeper by "moving from nice-to-know to need-to-know." With the increasing coordination of BI activities, many responses point towards establishing a centralized, group wide coordination of BI. Five companies mentioned considering outsourcing parts of their BI, which obviously is a way of being sure of getting the services that cannot be provided or which it is not feasible to provide internally. Five respondents also believed that during the five year period, their company would aim to combine internal and external information. The emphasis was often technological, in the sense that the interviewees wanted to see internal and external information in the same tool. By upgrading the role of BI, the respondents often meant that the expectations of BI will increase together with the volume of BI. One of the respondents reported that the need for BI had been recognized but the term itself "is confusing and does not open to anyone except BI professionals." The increasing utilization of BI in strategy formation was the theme that stood out the most; eight companies predicted that this would happen.

MAIN CHANGES IN BI ON A THREE-YEAR TIME SCALE

In 2005, almost all respondents (95 %) reported having systematically organized BI activities. In 2002, the percentage of companies answering affirmatively to the same question was lower, 80 per cent. According to Table 1, the results for the ICT sector remained unchanged; still all the responding companies had systematically organized BI. The industry section had experienced a notable increase in BI over three years. However, the trade and services sector was the one truly showing a clear difference between the results of the years of comparison. In 2005, 95 per cent of the companies reported having systematically organized BI, whereas in 2002, "only" 74 per cent of the companies had engaged such activities.

Table 1. The percentages of the large Finnish companies having systematically organized BI activities in 2005. Comparable results of the 2002 survey in the brackets.

	Industry	Trade and Services	ICT
Systematically	88 % (82 %)	100 % (74 %)	100 % (100 %)
Not systematically	12 % (18 %)	0 % (26 %)	0 % (0 %)
TOTAL	100 % (100 %)	100 % (100 %)	100 % (100 %)

These quite high percentages show that the large Finnish companies engage in BI. On the other hand, one has to remember that at the time of the surveys, Finnish companies were in general well-off and under no imminent pressure to cut costs.

Three information needs – competitor, own industry, and customer – were most often covered by BI in major Finnish companies in 2002 and 2005 and they stood out very clearly in both surveys. The authors found it somewhat unexpected that technology information and information about similar industries were ranked low in importance. This may be an indication of companies using BI strategically, but not in order to spot major, often out-of-the-blue opportunities, or threats with a potential strategic impact.

According to the 2005 survey, the most important benefits that the responding companies had been able to gain were improved quality of information, improved internal information dissemination, and improved level of awareness. In 2002, the benefits were almost the same:

- improved quality of information
- improved threat and opportunity identification
- improved level of awareness.

Surprisingly, the respondents did not list improved threat and opportunity detection as one of the three most important benefits gained through BI in 2005 though it was the second most important benefit among the respondents in 2002. It is suggested that one of the reasons for this decline could be the shift in market conditions over three years. Unsurprisingly, the companies interviewed did not feel that achieved cost savings were among the top three benefits gained in 2005. Three years earlier, this benefit was seen as the least significant of all the benefits. It is obviously very hard to measure cost savings regarding BI.

In 2002, top management utilized BI deliverables in 95 per cent of the companies, while in 2005, the number was 92 per cent. In 2005, middle management and experts used BI deliverables in all of the responding companies. In 2002, these personnel groups were seen as important users of BI deliverables in 85 per cent of the companies. There appears to have been a shift in the importance of management as a user of BI products. Nowadays, middle management and top management are almost equally important. Keeping this in mind, and adding that the functions of strategic planning and business development are often the most important functions in the utilization sense; one could conclude that BI is plainly in a strategic role in Finnish companies. The emphasis may be changing to the tactical front, however, with middle management becoming increasingly important.

Interestingly, the same issues were ranked top three in both surveys when the interviewees were asked to suggest the most notable areas for improvement in BI. The top three improvement areas were:

- identifying critical information needs
- utilizing internal knowledge
- effectiveness of information gathering and management.

Identifying critical information needs is one of the requirements proving the usefulness of BI and keeping the end-users satisfied. If the information needs are not correctly identified, it is very difficult to meet them. Producing information that is not needed makes a BI unit look useless. The results of both studies also suggest that utilizing internal knowledge will stand out as an important development issue. This is an issue that many companies struggle with.

According to Figure 5, the greatest shift in BI investments had been experienced in the ICT sector, where a vast majority expected its investments to increase only moderately in the future. In 2005, 20 per cent of the companies operating in the ICT sector expected their investments in BI to decrease moderately by the year 2011. Three years earlier, no ICT company expected its BI investments to drop in the following five-year period. In the industry sector, the growth rate of BI investments was slowing down; only 13 per cent of companies predicted notable growth in BI investments in 2005. This is a significantly lower figure than in 2002, when 41 per cent of the companies in the industry sector predicted a notable increase in BI investment. In the trade and services sector, there was an ever larger group of companies forecasting significant growth in their BI investment in 2005 than in 2002. However, the group of companies projecting the current level of BI investment in the future had also increased in the trade and services sector during the three previous years.

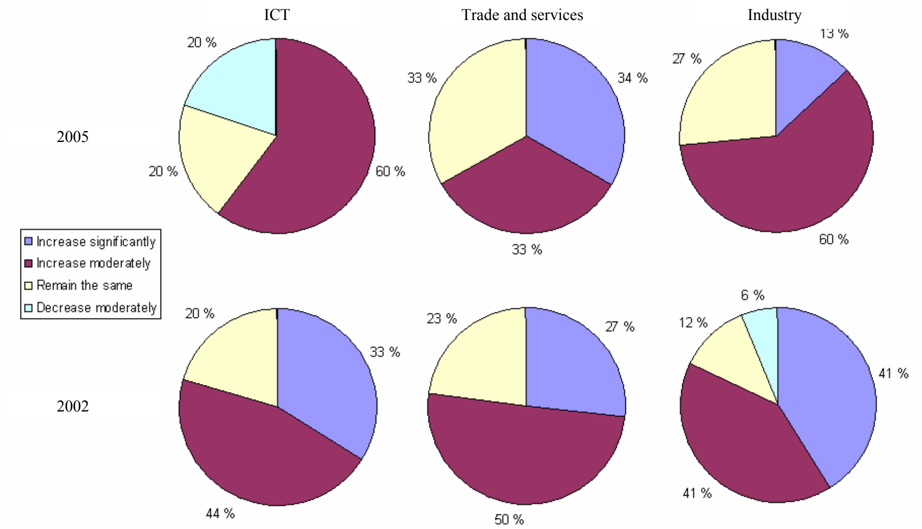


Figure 5. Comparison of estimated development of investments in BI activities in the different sectors between the 2005 and the 2002 results.

The percentage of companies planning to increase their BI investments at least moderately declined to 64 per cent from 77 per cent over three years. On the other hand, only the ICT sector had some companies (20 %) that anticipated drops in their BI investments by 2011. Therefore, it seems that the companies were still willing to invest in BI, but not as heavily as in 2002. This could be interpreted as an indication that BI is starting to be a quite or partly established activity in large Finnish companies. In addition, these companies may be in that part of the learning curve where they can start optimizing i.e. reaching the same performance with fewer resources.

CONCLUSIONS AND DISCUSSION

A major development in BI utilization is that 95 per cent of the large Finnish companies have systematically organized activity to collect and analyze information regarding the operating environment of the company in 2005. Three years earlier, the figure was 80 per cent. BI is apparently becoming an integral part of the large Finnish companies' operations.

According to both surveys, top management and middle management are the most important user groups of BI products in the large Finnish companies. These two management groups are usually involved in the strategy process. Thus strategic decision-makers seem to be an important user group of BI. In the authors' experience, it is sometimes hard for top management to understand that the information they get through discussions with middle management has often been actually provided by a BI unit. When the BI unit is invisible to top management, it is difficult to appreciate or need it. In light of this observation, it is definitely good development for BI if it is also utilized more and directly in strategic decision-making.

The general impression received from the studies is that there are only few BI measures. From the authors' perspective, measuring the benefits of BI is a difficult task and is definitely something on which Finnish and other companies need to concentrate more. In the BI literature, there are various benefits that can be derived by using BI. However, the arguments on whether these benefits actually occur in practice are inconclusive. In addition, using BI consumes resources. Therefore, there should be a way to measure whether the costs are reasonable in comparison to the profits received. Altogether, measuring is important regarding any business activity and hence should also be taken seriously with regard to BI.

According to both surveys, recognizing needs for information is quite often felt to be demanding and difficult in the large Finnish companies, as the needs are partly unarticulated. However, defining the needs for information is very important for companies, as without such effort there is no real purpose in any operations concerning business information acquisition and analysis. Generally, the large Finnish companies see BI in the future as an enabling practice which can assist in controlled management of businesses and information. On the other hand, the companies interviewed already consider BI as a crucial element for competitive business operations. The following trends regarding BI could be discerned in both surveys:

- BI will become an integral part of large companies' operations.
- Middle management will become the most important user of BI deliverables.
- BI will continue in a strategic role at least in the medium-term. The role may become more tactical if the above-mentioned trend materializes.
- The level of BI investment growth will decrease as companies approach an optimal level of BI performance.

In both surveys, the validity of the questions was confirmed by a number of BI professionals who were involved in the design and testing of the questionnaire. Because this paper is based on the results of two surveys, the reliability of respondents' answers must be also considered. In both surveys, only one individual from each of the top 50 companies was interviewed. Thus, it is unclear if an interviewee's answer represented the company itself or a subjective opinion. Understandably, this impairs the reliability of the results. The respondents were mainly people responsible for BI and they knew extremely well what they were talking about. On the other hand, they might even be too close to the activity and hence have a too favorable or biased a view of BI. In addition, some of the responses may have been based more on a gut feeling than on hard facts. This could also impair the reliability of the results. However, BI is quite a specific field; hence random respondents might not have had a thorough enough picture of BI to participate in the surveys.

The results of the surveys covered the Finnish large-scale companies belonging to the top-50 group in 2002 and 2005. Because of this, the results cannot be generalized to all Finnish companies. On the other hand, the response rate reached 92 per cent in 2002 and 82 per cent in 2005, which implies high reliability. Thus, the results are probably representation of the state of BI in the target group – the 50 largest Finnish companies. Since the surveys were quite similar, comparing the results between these two studies was possible. The reliability of this comparison should be reasonable, due to the use of similar and in some cases identical questions in 2002 and 2005. To conclude, the reliability and validity were not on the highest possible level in either survey but in the authors' estimation, they were sufficient to support the conclusions of this paper.

REFERENCES

Dishman, P., Fleisher, C. S. & Knip, V. (2003). Chronological and Categorized Bibliography of Key Competitive Intelligence Scholarship: Part 1 (1997 – Present). *Journal of Competitive Intelligence and Management,* 1(1), 13-79.

Fleisher, C. S. (2001). An Introduction to the Management and Practice of CI in C. S. Fleisher & D. L. Blenkhorn (eds.) in *Managing Frontiers in Competitive Intelligence*, Quorum Books, Westport, 77-79.

Fuld, L. M. (1995). *The New Competitor Intelligence: the Complete Resource for Finding, Analysing and Using Information about Your Competitors.* John Wiley & Sons, New York.

Gartz, U. (2004). Enterprise Information Management in M. Raisinghani (ed.) in *Business Intelligence in the Digital Economy: Opportunities, Limitations and Risks.* Idea Group Publishing, Hershey.

Hannula, M. & Pirttimäki, V. (2003). Business Intelligence – Empirical Study on the Top 50 Finnish Companies. *Journal of American Academy of Business,* 2(2), 593-599.

Hannula, M. & Pirttimäki, V. (2005). A Cube of Business Information. *Journal of Competitive Intelligence and Management,* 3(1), 34-40.

Hirvensalo, I. (2004). Competitive Intelligence in Finland. *Journal of Competitive Intelligence and Management,* 2(2), 22-37.

Kahaner, L. (1996). *Competitive Intelligence. How to Gather, Analyze, and Use Information to Move Your Business to the Top*. Simon and Schuster, New York.

Knip, V. & Fleisher, C. S. (2005). As the Globe Spins: A Benchmark Examination of CI in 15 Key Countries. Presentation at *SCIP's 20th Annual International Conference & Exhibition*, Chicago, Illinois, USA, April 6-9.

Korhonen, J. & Horttanainen, A. (2002). *Business Intelligence -tutkimus, BI-tiedosta tuotteistamiseen.* Presentation at a seminar on Business Intelligence, October 23. (In Finnish.)

Koskinen, A. (2005). *Business Intelligence in Major Corporations in Finland and Canada.* Master of Science Thesis, Industrial Engineering and Management, Tampere University of Technology.

Lilius, A.-L. (2004). 500 suurinta. Finnish Business Magazine *Talouselämä,* 21, 42-43. (In Finnish.)

Lönnqvist, A. & Pirttimäki, V. (2005). Measurement of Business Intelligence. Forthcoming in *Information Systems Management* in the Winter 2006 issue.

Market-Visio Inc. (2001). Newsletter on the Organization's Internet site. Retrieved from http://www.marketvisio.fi/itmanagement/multiclient2.cfm?tutkimus_ID=206. (In Finnish.)

Market-Visio Inc. (2002). Newsletter on the Organization's Internet site. Retrieved from http://www.marketvisio.fi/ajankohtaista/tiedote1.cfm?ID=141&paluu=../etusivu.cfm. (In Finnish.)

Pirttimäki, V. (2002). *Liiketoimintatiedon hallinta suomalaisissa suuryrityksissä.* Master of Science Thesis, Industrial Engineering and Management, Tampere University of Technology. (In Finnish.)

Pirttimäki, V., Lönnqvist, A. & Karjaluoto, A. (2005). Measurement of Business Intelligence in a Finnish Telecommunications Company. Forthcoming in *Electronic Journal of Knowledge Management*.

Saarelainen, J. & Vanhanen, A. (2002). Talouselämä 500 -taulukot. *Finnish Business Magazine Talouselämä,* 20, 88-89. (In Finnish.)

Simon, N. J. (1998). Determining Measures of Success. *Competitive Intelligence Magazine*, 1(2), 45-48.

Ståhle P. & Grönroos, M. (1999). *Knowledge Management – tietopääoma yrityksen kilpailutekijänä.* 2nd Edition, WSOY, Porvoo. (In Finnish.)

Thierauf, R. J. (2001). *Effective Business Intelligence Systems.* Quorum Books, Westport.

Vesa, J. (2002). BI tulee taas. *Tietoviikko,* 8.5.2002, 14-15. (In Finnish.)

Vibert, C. (ed.). (2004). *Competitive Intelligence. A Framework for Web-Based Analysis and Decision-Making*. Thomas South-Western.

APPENDIX 1: SAMPLE OF THE QUESTIONNAIRE USED IN THE SURVEYS

A. DEFINITION OF THE ACTIVITY

0. Is there a systematically organized activity to collect and analyze information regarding the external operating environment of your company?

a) Yes. ➔ *Move to question no. 1.* □
b) No. ➔ *Answer the sub-questions below and move to question no. 17.* □
   i) How do you keep up-to-date on the developments in your external operating environment?
   ____
   ____
   ii) Are you planning on launching an activity of this sort?
      a) Yes □
         i) When? Why? Elaborate:
         ____
         ____
      b) No □
         i) Why not? Elaborate:
         ____
         ____

1. What is the activity called?

a) Business Intelligence □
b) Competitive Intelligence □
c) Competitor Intelligence □
d) Knowledge Management □
e) Market Intelligence □
f) Market Monitoring □
g) Market Research □
h) Something else, what?____________________________________ □

2. For how long has the activity been systematically organized?

a) Over 10 years □
b) 5 – 10 years □
c) 3 – 5 years □
d) 1 – 3 years □
e) Less than 1 year □
f) Do not know □

3. List the three most important intelligence needs in order of importance (1 – 3, 1 = most important, 3= 3rd most important).

a) Intelligence regarding the industry in which your company operates □
b) Competitor intelligence □
c) Customer intelligence □
d) Macro environmental trends □
e) Customer industry intelligence □
f) Country intelligence □
g) Technological intelligence □
h) Intelligence regarding similar industries □
i) If something else, what?__________________________________ □

Why do you consider those needs important? Elaborate, please:

Contract-Driven Cross-Organizational Business Processes

Ustun Yildiz, Olivera Marjanovic, Claude Godart
[1]LORIA-INRIA
615, rue du jardin botanique
Villers-lès-Nancy, 54600, FRANCE
yildiz, godart@loria.fr

[2]The University of New South Wales
Sydney, NSW 2052, AUSTRALIA
o.marjanovic@unsw.edu.au

Abstract

Proper management of cross-organizational processes is one of the key requirements for successful collaboration among autonomous business organizations. This is a challenging task, because these processes are normatively regulated by mutually agreed contracts. Thanks to the advances in middleware IT technologies, some of the most important technical limitations of cross-organizational collaboration have been resolved. However, there are numerous open problems related to management of these business processes, caused by their distributed and heterogeneous nature. Monitoring and coordination support are certainly the most challenging aspects of cross-organizational business process management. However, they are yet to be properly addressed. In this paper, we describe a contract-driven, cross-organizational process monitoring and coordination mechanism. Compared to the existing workflow-based solutions, the proposed contract-based solution goes one step further in terms of provided flexibility. More specifically, we propose a generic contract model to formally represent cross-organizational collaboration and then we introduce an event-driven infrastructure for the enactment and monitoring of contract related processes.

Keywords:
Business Process Management, Virtual Enterprise Modeling, Cross-organizational Collaboration, E-contracting, Complex Event Processing

I.INTRODUCTION

Recently, Business Process Management (BPM) domain has gained considerable attention from researchers and industry-practitioners. From the technical point of view, the main objective of Business Process Management (BPM) is to provide a computer-based solution for management of processes that involve entities such as persons, activities and systems that all operate in tandem to fulfill (a) particular business goal(s). BPM tools have been on the market since early nineties. Currently, the market for BPM solutions is growing from the estimated $214 million in 2002 and is expected to double this value in 2008 (Wintergreen, 2003). The advances in middleware IT technology make the construction of cross-organizational processes easier than in the past. In a cross-organizational process, the intra-organizational processes of business partners are interconnected in order to satisfy the mutual business objectives. This kind of collaboration gives an organization a possibility to use the capabilities of their business partners that they do not currently have. However, cross-organizational collaborative processes are limited by two contradictory requirements: The first one is the rigorous coordination and monitoring mechanism among collaborating partners to ensure their compliance with corresponding agreements. The second one is the necessity of flexible dependencies between intra and cross-organizational processes that will enable the protection of the partner's privacy (Orlowska et al. 2005).

Workflow is an established and powerful formal concept, widely used for modeling and implementation of business processes. Although a workflow is an efficient way to model processes, it is not fully adequate for modeling of non-sequential cross-organizational collaboration. First, it tightly couples the business activities with their implementation (Schuster et al. 2000), (Ellis 2003). Within the cross-organizational context, this limitation prevents the autonomous sub-process owner to choose possible alternatives for the enactment of the same goal and obliges it to couple its internal processes with its partner's. The second important limitation is that long duration collaboration can be based on on-demand linkages that are expressed with mutual commitments which can not be modeled with current workflow solutions.

Besides modeling of cross-organizational collaboration, another important requirement is verification of consistency between the actual (run-time) and initially agreed behavior of business partners. The process instance is expected to be consistent with its model (Rinderle et al. 2005). This requires a background process that monitors the actual collaborative process. This requirement is crucial to insure that every partner sees benefits from the involved collaboration and is able to detect any contract-related problems. Due to the dynamic and unpredictable changes of the distributed environment, it is hard to prescribe all possible exceptions that can occur. But somehow they must be handled in order to allow collaboration to continue. However, monitoring and exception handling are complex tasks in workflow management systems (Hagen et al. 2000). From the perspective of business partners, we can summarize their requirements about cross-organizational collaboration as follows: (i) efficient management of complex interactions among involved business entities, (ii) conformity of partners to their agreed behavior, (iii) provision of a flexible mechanism that does not couple their intra-organizational processes with the collaboration.

The main objective of this paper is to propose a contract-based approach to enable cross-organizational collaboration of autonomous partners. The contract based solutions stand as an alternative to the existing workflow based solutions and nowadays, are receiving increasing attention from a number of researchers (Berry et al. 2005), (Xu et al. 2004), (Zdravkovic et al. 2004), (Farrell et al. 2005). Like in traditional business collaboration, the contract prescribes the constraints that all business partners are supposed to comply with, during their collaboration. Contrary to the workflows and workflow-like approaches, collaboration described by the contract should be flexible enough to allow its implementation by intra-organizational processes and enable long-term collaboration. Assuming that the contract adopts an open-policy (this means that the absence of a permission to execute an action is not taken to be the same as presence of a prohibition to execute it), monitoring of the related processes via the contract is a considerable solution to ensure the requirements are satisfied.

In this paper, we introduce a business contract model that is used to model the complex interactions of cross-organizational partners. Furthermore, we propose an event-based infrastructure for the instrumentation of the environment that will interconnect the partners and check the conformity of their behavior. The use of event driven approaches for process interconnection and monitoring is not new, due its generic and fine-grained nature; it offers a considerable support to model the run-time features of dynamic applications (Luckham, 2002). However, in this paper we apply it to contract-driven business processes to offer an innovative approach to monitoring and implementation of business processes.

The remainder of this paper is organized as follows: the next section will introduce a case study that describes a real life collaboration of two autonomous business partners. In section 3, we propose a business contract model to define this collaboration. Section 4 explains how the contract and received events are handled in order to monitor the behavior of business partners and coordinate their interactions. Section 5 gives an overview of similar approaches while in the last section we describe the main conclusion of describe our future work.

II ILLUSTRATIVE EXAMPLE

In this section, we describe an example of a cross-organizational collaboration scenario originally proposed by the CrossFlow Project (CrossFlow 2000). This scenario is derived from the real life example of co-operation between two companies. For the purposes of this paper, we have extended this scenario with new process-related requirements. The first autonomous organization *(Provider)* is a logistic provider company that has a global network of domestic and international delivery services. The second organization *(Customer)* is an on-line store company such as Amazon.com that uses the delivery services of the *Provider* to handle the orders of its clients. Each company has its own Workflow Management System *(WfMS)*. In order to fulfill their commitments to their business partners, each company relies on several business policies that govern and normatively regulate its intra-organizational processes. For example, the *Provider* aims to improve the quality of services it offers to its customers by introducing flexible delivery support. This allows a customer to change the details of their orders in order to reduce the associated costs etc. Long-term collaboration of *Provider* and *Customer* is based on on-demand services. Consequently their cross-organizational relationships are defined and regulated by a business contract that is independent of their specific internal enactment mechanism. The contract includes several features such as the declaration or legal contacts of involved parties.

In this paper, we are particularly interested in the contractual issues that define the relationships and collaboration of the *Provider* and *Customer.* The following is an informal description of the scenario and the corresponding contract that we are going to use later in the paper to illustrate our approach. The aim of *Customer* is to manage the orders of its clients that use its web portal. It may initiate the cross-organizational process placing a delivery order to *Provider*. In the delivery order, *Customer* defines the characteristics of the item and its destination. As soon as the *Customer* requests a delivery service, this information is distributed to all sub-units of *Provider*. *Provider's* pick-up unit picks up the item to be delivered from *Customer*, attaches a bar-code and then sends the item to its closest warehouse. This warehouse is a node in the global network of the *Provider*, so it transfers the item to another warehouse closer to its destination and this warehouse than proceeds with the delivery of the requested item.

The *WfMS* of *Customer* and *Provider* are interconnected on the top of a contract repository structure that tracks the execution of the contract. Consequently, both organizations have the same view of the contract related features before the beginning of its execution. However, some flexibility is allowed. For example, *Costumer* may change the delivery address before the actual delivery process starts from the warehouse of the *Provider*. *Customer* is only allowed to place an additional delivery order if the resulting total amount of its unpaid orders is not above a certain limit. Any item must be delivered to its destination within a precise time if *Customer* does not change its delivery address. As defined in the rules, *Customer* can influence the internal processes of *Provider*. The influence is restricted by the current state of their contract statements. For example, *Customer* can start another delivery if it does not violate the rule that restricts the amount of its total unpaid orders. The role of the contract is not limited to restrict the operation of its parties; it is also an intermediary for the coordination.

Let us now assume that the item to be delivered is in the warehouse of *Provider*. As long as the item stands in the warehouse, *Customer* holds the flexibility to change the delivery address. In this particular situation, an unpredicted wait time in the warehouse caused by an internal problem of *Provider* can be transformed to a situation that *Customer can* take an advantage of. In the contract model that we present in the next section, we aim to model the rules that restrict the behavior of partners and regulate their interactions. These rules gather contract entities and define their relationships. Besides the relationships included in a single rule, the rules are interconnected by special relations that express the contract as a whole.

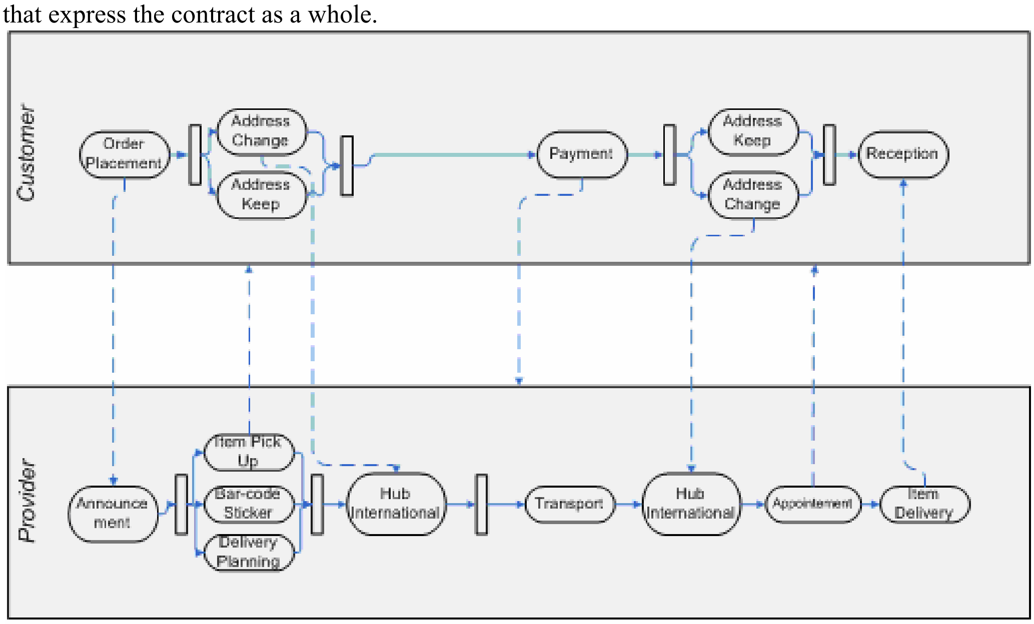


Figure 1: The Cross-Organizational Scenario

Figure 1 illustrates partially the private workflows of *Provider* and *Customer.* The dashed arrows are the interactions that can happen or must happen between them. The contract defines these interactions and the conditional, causal and temporal statements that are necessary for their existence.

III BUSINESS CONTRACT MODEL

In this section, we define a business contract model that will be used in the rest of the paper. The contract model gathers the key elements of business processes and restricts their run-time behavior. As we mentioned in the previous section, although the contract governs the internal business processes of contract partners and the cross-organizational process, its specification and operation are very different compared to the Workflow concept that is also used to represent business processes. A contract predefines all business constraints that the respective business partners are supposed to adhere to while they collaborate. Our main focus is on formal representation of the contract constraints. These constraints are the key features that should be dealt with to support collaboration enactment and monitoring. To be able to formally model a business contract and define its monitoring infrastructure, it is necessary to have a conceptual tool that can be used for: (i) expression of the complex relationships among contract entities, (ii) description of run-time behavior of the contract-regulated processes, (iii) monitoring of cross-organizational collaboration via the contract. To satisfy these modeling requirements, in this paper, we have adopted the Complex Event Processing (CEP) technology originally proposed by (Luckham 2002), (Shanahan 1999) and (Adi et al. 2004). The CEP is conceived as a middleware support for the formalization of event-driven dynamic applications. It allows definition of events that consist of the relevant occurrences in the observed environment and their aggregation to describe semantically interested situations worth observing. Here we apply the concept of CEP to a new domain in order to model the contract content and then monitor the related processes. Familiarity with CEP could help to better understand its application to the contract-monitoring process and our contribution. However, this paper is self-contained and includes all of necessary definitions.

We consider the contract as a set of ***Constraint***s that have properties and inter-relationships. A single constraint gathers atomic elements of the contract. The atomic elements of the contract are ***Parties***, ***Operation***s, ***Metric***s, ***Date***s and ***Object***s. The contract ***Parties*** perform ***Operations***. The performed operations operate on ***Object***s and ***Metric***s. A ***Metric*** is a numerical fact worth observing because its value can change during collaboration. The temporal issues are modeled by absolute and relative time values expressed as ***Date***s.

Constraint Definition

In this section, we give the formal definition of a contract ***Constraint***. The definition includes the atomic elements of the contract, their relationships and also the relationships among the constraints in order to express the contract as a whole.

Definition 1 (Constraint): *A Constraint is defined by a tuple (id, executer, operands [], operators [], lifespan,) where:*

- *id ∈ ID, ID is the set of identifiers*
- *executer ∈ Parties is the contract party that is supposed to execute or respect the activities depicted in the constraint,*
- *operands [] is the list of atomic contract entities that are mentioned in the constraint such as operations, metrics, objects etc (see next section),*
- *operators[] is the list of operators that are used to aggregate the operands and restrict their behavior(see next section),*
- *lifespan defines the temporal context in which the constraint is relevant (see Definition 1.1),*

Definition 1.1 (Constraint lifespan): *The lifespan of a constraint is defined by a tuple (initiator, terminator, initiationType, correlationType, terminationType) where:*

- *initiator defines the fact(s) that initiate(s) the temporal context during which the restriction of the constraint is relevant,*
- *termination defines the fact(s) that ends the validity of the initiated constraint ,*
- *initiationType ∈ {"basic", "dependent", "complex"},*
- *correlationType ∈ {"ignore", "add", "add(k)"| k is a condition},*
- *terminationType ⊂ {"basic", "dependent", "complex", "discard", "all"},*

The proper modeling of the temporal context of a constraint is a very critical issue. The lifespan of a constraint defines the temporal context and conditions that the purpose of the constraint becomes relevant. The lifespan begins with the occurrence of an *initiator* and ends with the occurrence of a *terminator*. The *initiator* can be a simple predicate that becomes valid or indicates a receipt of a notification message. In these cases the *initiationType* of the constraint is "basic". The constraints that have "dependent" *initiationType* are activated according to the state changes of other constraints. For example, a penalty constraint that has the purpose to compensate the damage of another not respected constraint is activated when the state of latter becomes violated. A constraint can be initiated by several "basic" initiators or by the state changes of another initiated constraint, then its *initiationType* is "complex". Another important point of the constraint lifespan is its repetition over the deployment. A constraint can be re-initiated if the same initiators happen while there is at least one initiated constraint instance. The *correlationType* of a constraint defines whether the constraint can be re-initiated while there is another initiated instance of it. If the *correlationType* is "ignore" the constraint will have a single instance. If the *correlationType* is "add" then another constraint instance can be created when the same initiators occur, finally the "add(k)" type defines the conditions that another constraint can be initiated. For example, each time the *Customer* places an order, the constraints that impose *Provider* to process the delivery are initiated.

The *terminationType* of a constraint defines the termination mode of the relevance of it. The constraints that have "basic", "dependent" and "complex" *terminationType*s end with the occurrence of terminators, the state changes related to another constraint or both of them. The relevance of the constraint can end while its own state changes during its lifespan. For example, if a constraint that depicts a relationship that must hold during its lifespan, begins to not hold then its state changes. This kind of termination is “discard”. For the cases that a constraint has “add” *inititiatonType* and “all” *termiantionType*, if there are more than one instance of the same constraint, the relevance of the initiated constraints end if one of them ends.

Operators and Operands

To represent the temporal context and dependencies of a constraint, we use separate components to reduce the complexity of the formal model. *Operators* and *Operands* are key features that are used to express the restriction purposes of the constraints. In order to model them formally, we use two techniques adopted from CEP and Deontic Logic (Lee, 1988). The *Operators* designate structures that are used to express syntactic, algebraic and behavioral restrictions on their *Operands*. The *Operands* are the atomic elements of the contract but also their properties can be subjects of restriction.

The Operators that we consider are classified into three categories. The first is *Behavioral Operators* taken from the Deontic Logic. The Deontic Logic studies the logical relationships of entities asserting that certain of them are *Obligations*, *Permissions*, and *Prohibitions*. The operation of these constraints is similar to their dictionary meanings: Obligations describe the facts that must happen or hold, Permissions characterize the facts that can occur, and Prohibitions are used for the facts that must not occur. The second type of operator is *Value Operators* that consist of arithmetical operators that restrict the numeric values of their operands. The last type of operator we consider are *Syntactic Operators.* They are used to express aggregation relationships among the restricted entities. In the following table we present the operators that we use in this paper.

Oper ID	Operator	Type	Operands	Semantic
D1	***P***	*Deontic*	*All*	*Permission*
D2	***O***	*Deontic*	*All*	*Obligation*
D3	***F***	*Deontic*	*All*	*Prohibition*
V1	$=$	*Value*	*Metric*	*Comparison*
V2	$<$	*Value*	*Metric*	*Comparison*
V3	$>$	*Value*	*Metric*	*Comparison*
V4	$\neq$	*Value*	*Metric*	*Comparison*
S1	$\neg$	*Syntactic*	*All*	*Negation*
S2	*Sequence*	*Syntactic*	*Operations*	*Aggregation*
S3	*All*	*Syntactic*	*Operations*	*Aggregation*
S4	*Atmost*	*Syntactic*	*Operations*	*Aggregation*
S5	*Each*	*Syntactic*	*Operations*	*Aggregation*
S6	$\bigcap$	*Syntactic*	*All*	*And*
S7	$\bigcup$	*Syntactic*	*All*	*Or*

Table 2. Operators and Operands

Table 1 introduces the classification of the Operators and their Operands used in this paper.

III. ENSURING PROCESS ENACTMENT

Besides the expression of the content of the contract, another important feature is its view and sharing by involved partners. Like in traditional business affaires, we assume that the contract is a set of rules that have an internal coherence and it is agreed by business partners before the beginning of their collaboration. So, every business partner has the same view of the contract as a reference document. By definition, the content of the contract can not change during collaboration. A change can be considered if all of the involved parties agree about the change. We consider the contract as an entity that is taken as a reference by business partners in order to regulate the individual behavior. The related individual processes and the monitoring mechanism can be implemented differently by involved parties. There can be a trusted third party that observes the actual states of contract constraints or the partners may prefer to set up their own monitoring mechanism that makes the same task.

In any case, there will be a different interpretation of the same instance for different partners. For example, *Customer* has to pay the delivery costs before the end of the delivery. The *Customer* will see the state of this constraint as an operation that it must execute and the *Provider* will see the same constraint instance as an expected operation of *Customer*. In order to set up a mechanism that will indicate the state of contract related processes and monitor their execution, there is a necessity for additional data structures. In this section, we introduce an enriched data model for executed contract content. The model we propose is similar to process data that are used by workflows, but it is modified for contract needs.

Constraint Instance

A constraint instance is a run-time representation of a contract constraint. It has a state for every instant during the enactment. According to *correlationType* and *initiators* of the constraint, the new instance of the same constraint can be created during the enactment. As the partners will have different interpretation of the same constraint, the constraint instance will have a different semantics for each of them.

Definition 2 (Constraint Instance): *A constraint instance is defined by a tuple (instance_id, mode, current_state, inactive_state, initiation_date, termination_date) where:*

- *<u>instance_id</u> ∈ ID, ID is the set of identifiers,*
- *<u>number</u> ∈ ℵ is the counter that illustrates the number active instance of the constraint,*
- *<u>currentState</u> ∈ {"initial", "active"," executed"," violated"," terminated" }*
- *<u>mode</u> ∈ {"request-response", "solicit-reponse"}*
- *<u>initiationDate</u> ∈ Date*
- *<u>terminationDate</u> ∈ Date*

The properties of a constraint instance can change after its initiation. When a constraint is initiated, its *currentState* becomes *active*. An active constraint instance requires to be monitored by its executer and also the partners that invoked it or by the partners related to the constraint. Depending on the purpose of the constraint, the state of the rule can be violated, terminated or executed. If the purpose of a constraint is expressed by a permission modality, the permission can be executed while the rule is *active* or the party may prefer not to execute its permission. If the permission is executed than the state of the constraint instance becomes *executed.* On the other hand, if the permission is not executed the states becomes *terminated* with the occurrence of a terminator. If the deontic modality of a rule is an obligation and if the executer can fulfill the respective obligations than the state of the constraint becomes *executed.* It the obligationcannot be fulfilled than the state becomes *violated*. For the constraints that include prohibition operators, the instance can pass to the *violated* state or to *terminated* state. At each state transition a *transitionEvent* is generated by the contract monitor. Possible *transitionEvents* include *initiator* or *terminator* of other contract constraints. At the moment, a constraint instance is initiated, it has a *initiationDate* and at its state transition, it has a *terminationDate.* Besides these conventional properties, a constraint instance has a mode property. The mode of a constraint instance characterizes its semantic for the entity that holds it. If the mode of a constraint instance is "request-response" than the constraint instance concerns its holder. This means that the activated constraint instance, according to its deontic modality, should be processed by its holder. If the mode of the constraint is " solicitresponse", this indicates that the constraint must be respected or can be executed by a party besides the holder of this instance. So, the instance holder party should respect "request-response" instance, managing its internal processes, and it has to observe the state of the "solicit-response" constraints to check if its partners behave as expected.

Business Events

We use the event paradigm to model dynamic infrastructure of the execution environment. The events correspond to the relevant occurrences that can come from different sources. The first is the contract repository that is a shared space between contract parties. The second event source are the entities are involved in collaboration, but not originally stated in the contract. For example, a bank that holds the accounts of *Provider* can produce events to *Provider* even if it is not a party in a particular contract.. Separation of event sources has a major advantage as it enables cross-validation of constraints instances. For example, *Customer* can claim that the requested order payment has been made. The corresponding event will be generated by the contract repository as the initiation and termination of constraints are made by contract partners. But at the same time, the *Provider* can check the bank that can generate an event confirming that the required payment is not received. So, the *Provider* can initiate cross-validation of the “solicit-response” constraint instance in which the *Customer* has an obligation and detect a possible inconsistency.

In this framework, the events coming from different sources and contract repository are gathered in the event middleware and forwarded to contracting parties. The contract repository also generates events when the states of the constraint instances change. These events are forwarded to contract parties to update their constraint instances. Furthermore, when an environment is modeled with events, we need to find a way to identify and describe all relevant occurrences of important events. . In order to proceed with this mapping operation, we propose the events as follows:

- *Constraint instance events:* These events are implicit events produced by constraint instance holders in order to initiate or terminate the constraint instances,
- *Metric events:* These events are produced when the value of an event changes,
- *Objects events:* Based on the presence of an object, object events are generated to describe, they describe if an object exists or does not exist
- *Operation events:* They describe events related to an operation such as its occurrence or execution ,

Due to the lack of space we do describe the content of each event in more details. However, it is important to point out that each event contains a number attributes that defines its properties such as occurrence time, detection time, the parties or fact that caused it or the data that are transferred with this event. This information enables the mapping between the events and the relevant constraint instances. Thus, the evolution of the execution can be compared to the run-time constraints.

Putting the Enactment into Practice

In this section, we illustrate the application of the proposed methodology to the example presented in the previous section. Figure 2 illustrates how the constraint instances are initiated and terminated according to the behavior of the partners. The frame 1 is initiated with the order placement, the *correlationType* of this constraint is *add* with the condition associated with another constraint: the total amount of unpaid orders is limited. The purpose of frame 1 is to illustrate the obligation of Provider to deliver the item before a precise deadline dynamically defined when the order is placed. For the constraint that prescribes the amount, another frame is initiated in order to collect the amount of unpaid orders. This frame is not depicted by Figure 2. However, it can be described as follows. Its initiator is the first placed order and its *correlationType* is *ignore because* it will terminate when it is violated. When the product order is received from *Customer*, they are permitted to change the delivery address until the order is taken from its first international hub. This permission is expressed by frame 3. The initiator of this frame is the placement of the order, defined by *Customer* and the terminator is defined by *Provider* when it takes the item out of the international hub. While the order is processed by *Provider*, *Customer* has to pay the charges of the order. It can pay the charges before the prescribed deadline which is calculated after the order placement. This obligation is initiated with the order placement; its *terminationType* is “discard” as it will finish with the payment. This obligation and similar “discard” type constraints have also an expiration date until which the constraint should be satisfied. The frame 5 depicts another constraint instance which is initiated when the item comes to its destination’s hub, *Customer* can change its delivery address again if the changed address is accessible from the hub. This constraint is terminated when *Provider* contacts the final customer that purchased the item for the final delivery. Because of the violated constraints, such as a late delivery, new constraints can be initiated in order to compensate the effects of the former. The initiations and terminations of constraint instances are driven by the events we presented in the previous section. For example, the constraints that have dependent *initiationType* are initiated by the *Constrain instance events* that occur when the state of the instance changes.

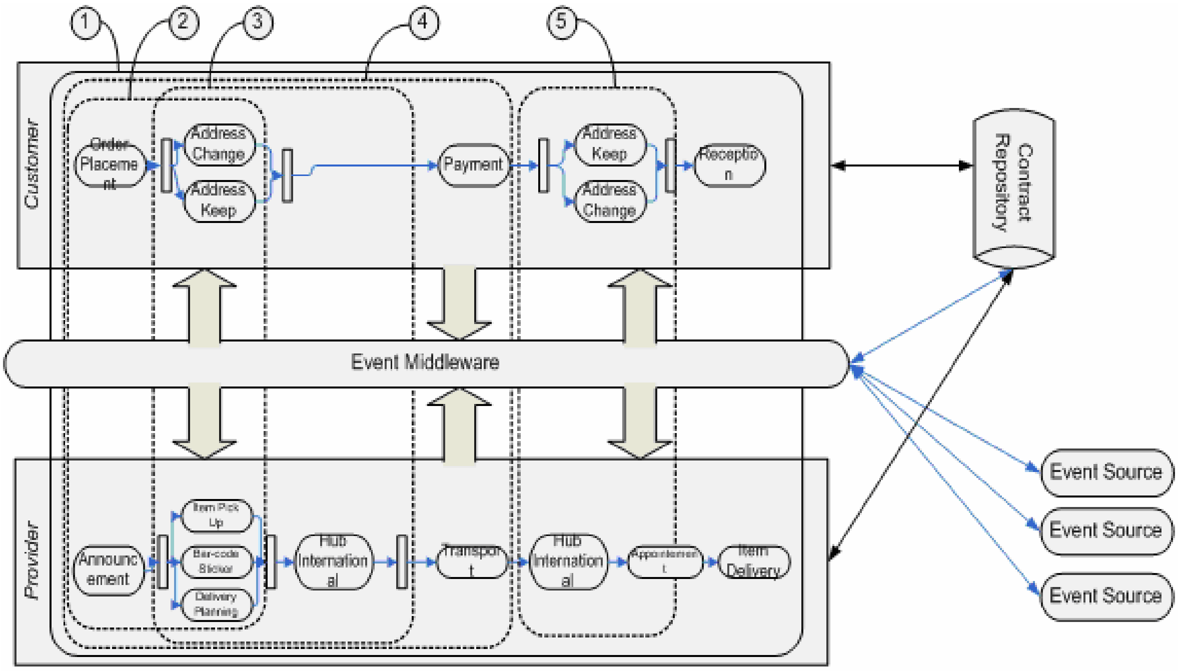


Figure 2. The environment modeled by events and constraint instances

For the constraints that depict a restriction on a shared metric, event notification mechanisms are associated to each event sources and the instance of these constraints are evaluated with the data carried with the associated event.

IV. RELATED WORK

Monitoring and coordination of collaboration of autonomous organizations is a hot topic in the recent research literature (Orlowska et al. 2005) (Aalst et al. 2005). The use of business contracts for this purpose is not a new idea. For example, (Xu et al. 2004) proposes a conceptual multi-party contract and a mechanism to monitor the execution of the contract statements. The proposed model is based on commitments graphs among contract parties. However, this particular solution does not provide flexible collaboration as required in the contract based collaboration. Furthermore, the same author is concerned by the temporal dependencies among business actions, but the model does not take into account all of the process related features. Another example is the Business Contract Language (BCL) (Milosevic et al. 1995). This example is closely related to our work. The proposed model is flexible but does not provide generic monitoring facilities. Consequently, it requires each monitor to develop its own monitoring mechanism.

The well-known application of contracts is Service Level Agreements (SLA) is also related to our work, It is used in the utility computing where SLAs are used to specify and manage the underlying network activities of service provider and service consumer. They define minimum obligations and expectations between participating actors. These types of contracts are invented for network activities and cannot be an efficient support for the expression of the high-level business objectives of partners and their complex interactions.

Business process flow languages (e.g. BPEL4WS) provide a language and a structure for describing a business process and how to define assertions. However, BPEL focuses on non functional or procedural contract and does not provide any explicit support for the business contract level of abstraction. In fact, these approaches are interesting but are dealing with recovery and failures support rather than prevention support.

V. CONCLUSION AND FUTURE WORK

The work presented in this paper is motivated by two major requirements of cross-organizational processes regulated by business contracts. The first is the necessity to provide the required flexibility and the second is the coordination and monitoring that has to be rigorous to make sure all parties comply with the contract.. In this paper, we propose a formal contract model to express the cross-organizational process enacted by autonomous organizations. More specifically, we used an event-based infrastructure to express the interactions between contract parties and the relevant changes in the business environment. The use of non-sequential contracts is an innovative solution that enables us to support flexible collaboration in which collaborating partners can protect their privacy, as they don't need to expose the details of their internal workflows to their business partners. As the contract is the key element of governance regulating this collaboration, the expression of its run-time state is an important and trusted solution for the problem of monitoring. Because, in the dynamic, heterogeneous and untrusted environments of the business market, the problems can be dealt when they are detected on time and in an objective way. Service Oriented Architectures (SOA) architectures, especially Web Services, are emerging as middleware to implement cross-organizational processes. The very recent advances in this domain such as WS-Coordination, WS-Agreement, WS-Policy are the actual agreements among involved Web Services. These technologies aim to specify the protocols supported by collaborating services such as the order in which individual services can be invoked or how the relevant message should be exchanged while they interact. Actually, to base collaboration of autonomous actors on a set of constraints is to define a business contract that regulates their interaction. The main objective of our future work is to further extend this work and apply it to the problem of dynamic composition of web services and the monitoring of their execution.

REFERENCES

Adi A. and Amit O.E. (2004), The situation manager. VLDB Journal., 13(2). pp.177–203.

Berry, A. and Milosevic, Z. (2005), “Extending choreography with business contract constraints”, International Journal of Cooperative Information Systems (IJCIS), Volume 14, Number 2-3, 131-179.

CrossFlow (2000), Cross-Organizational Workflow Support in Virtual Enterprises, ESPRIT Project 28635, available from http://www.crossflow.org, accessed October 2005.

Ellis, C. (2003), “Net models supporting human and humane behaviors”, Invited talk, in: Conference on Business Process Management (BPM), Eindhoven, The Netherlands.

Farrell, A.D.H., Sergot, M.J., Sallé M. and Bartolini, C.(2005), “Using the event calculus for tracking the normative state of contracts”, International Journal of Cooperative Information Systems (IJCIS), Volume 14, Number 2-3, pp. 99-129.

Hagen, C. and Alonso G. (2000), “Exception Handling in Workflow Management Systems”, IEEE Trans. Software Eng. 26(10): pp. 943-958.

Lee R. (1998), “Towards open electronic contracting”, Electronic Markets, Vol. 8, No. 3.

Luckham. D. (2002), The Power of Events. Addison Wesley.

Milosevic, Z., Berry, A., Bond, A. and Raymond, K. (1995), “An Architecture for Supporting Business Contracts in Open Distributed Systems”, Proc. of the 2nd International Workshop on Services in Distributed and Networked Environments (SDNE'95), Whistler, Canada.

Orlowska M.E. and Sadiq, S. (2005), “Collaborative business process technologies”, Data & Knowledge Engineering, Volume 52(1).

Rinderle, S., Reichert M. and Dadam, P. (2004), “Correctness criteria for dynamic changes in workflow systems - a survey”, Data & Knowledge. Eng. 50(1): 9-34.

Schuster, H., Georgakopoulos, D., Cichocki C. and Baker, D. (2000) “Modeling and composing service-based and reference process-based multi-enterprise processes” Proceedings of the 12th Conference on Advanced Information Systems Engineering (CAiSE), LNCS 1789, Springer Verlag, Stockholm, Sweden, pp. 247–263.

Shanahan M. (1999), "The event calculus explained", In Artificial Intelligence Today, LNCS: 1600, 409-430, Springer.

van der Aalst W.M.P, Benatallah B., Casati F. and Curbera F. (Eds) (2005) Proceedings of the 3rd International Conference of Business Process Management, LNCS 3649.

Wintergreen (2003) Business Process Management (BPM) market opportunities strategies, and forecasts, 2003 to 2008, Lexington: WinterGreen Research

Xu, L. and Jeusfeld, M. A. (2004) "Detecting Violators of Multi-party Contracts", Proc. of the CoopIS/DOA/ODBASE (1) 2004 pp. 526-543

Zdravkovic, J., and Johannesson, P. (2004), "Cooperation of Processes through Message Level Agreement. Proceedings of Advanced Information Systems Engineering", Proc. of the 16th International Conference, CAiSE 2004, Riga, Latvia, pp. 564-579.

Survey of Requirements Modeling in Software Engineering: A Research Perspective

Niu Xi
Tsinghua University
Beijing, China
Email: nx03@mails.tsinghua.edu.cn

Liu Qiang
nasrTsinghua University
Beijing, China
Email: liuqiang@mail.tsinghua.edu.cn

Abstract

Getting high-quality requirements is difficult and critical. Recent surveys have confirmed the growing recognition of RE as an area of important in software engineering. This paper presents five prevailing modeling method in software requirements engineering. Main concepts and techniques of each method are introduced first. To illustrate the detailed modeling process, case study of ' network-based project manager' is presented. Then for each method, the main characteristics are outlined and the advantages and disadvantages are analyzed. In the end, modeling as a common denominator to all RE process and the future work of requirements analysis is also discussed.

Keywords:

structured modeling, objected-oriented modeling, goal-based modeling, scenario-based modeling, formal method

1.INTRODUCTION

Software requirements have been repeatedly recognized during the past years to be an important problem. Inadequate, inconsistent, incomplete, or ambiguous requirements are numerous and have a disastrous impact on the quality of the resulting software. Recent surveys have confirmed the requirements problem. Over 50% of the failure projects are the result of poor requirements. Improving the quality of requirements is thus crucial but difficult to achieve. [1]Modeling appears to be a core process in requirements engineering. Currently, much work has been done on developing new modeling language and framework. This paper presents five typical of modeling language and evaluates these methods through a case study 'Network-based project manager'.

2.REVIEW OF MODELING METHODS

2.1Structured Modeling

Structured analysis (SA)[2] employs top-down hierarchical decomposition method to capture requirements of system. Based on structured analysis, specification could be elicited. Structured modeling is graphic framework for system specification and requirement analysis that is well adapted to describe functions and their interconnections in manufacturing systems. DFD, a typical method for structured modeling, employs a few notations to comprehensively express the flowing, processing, and storing of data in the system

2.2 Object-oriented Modeling

Object-oriented Method is initially derived from Object-oriented programming and then be popular in system analysis and design. Object-oriented (OO) languages is a kind of modeling techniques, bringing with it new concepts such as data encapsulation, inheritance, messaging, and polymorphism. The UML (Unified Modeling Language) [3]is OMG's most-used object-oriented specification, and the way the world models not only application structure, behavior, and architecture, but also business process and data structure. UML also provides a key foundation for unifying every step of development and integration from business modeling, through architectural and application modeling, to development, deployment, maintenance, and evolution.

2.3 Goal-based Modeling

In requirements engineering, a goal-oriented modeling approach has been recognized to be useful [1,4]. In general, goals describe the objectives that the system should achieve through the cooperation of actors in the software-to-be and in the environment. It captures ‘why’ the data and functions are there, and whether they are sufficient for achieving the high-level objectives that arise naturally in the requirements engineering process. The incorporation of explicit goal representations in requirement models provides a criterion for requirements completeness, i.e., the requirements can be judged as complete if they are sufficient to establish the goals that they are refining. GRL (goal-oriented requirements language)[5,6] is a language for supporting goal- and agent-oriented modeling and reasoning about requirements, with an emphasis on dealing with non-functional requirements (NFRs). GRL describes the intention and interests of stake-holders according to the environments, thus helping software designers to learn the motives of system roles and the rationale between roles (‘Why’ question).

2.4 Scenario-based Modeling

Scenarios have been used for various purposes—as means to elicit or validate system requirements, as concretization of use-oriented system descriptions, or as bases for test cases [7-9]. Scenario-based models present possible ways in which a system can be used to accomplish some desired functions or implicit purpose. Typically, it is a temporal sequence of interaction events between the intended software and its environment (composed of other systems and humans). Scenarios have also become popular in other fields, notably human–computer interaction and strategic planning [10,11]. UCM(Use Case Maps) is a typical framework of scenario-based method, which is developed by R.J.A.Buhr [12], Carleton University in the year 1992. UCM allows the behavioral aspects of the designed system to be visualized at varying degrees of abstraction and levels of detail.

2.5 Formal Method

Formal methods used in developing computer systems are mathematically based techniques for describing system properties. Such formal methods provide frameworks within which people can specify, develop, and verify systems in a systematic manner. A method is formal if it has a sound mathematical basis, typically given by a formal specification language [13]. This basis provides the means of precisely defining notions like consistency and completeness and, more relevantly, specification, implementation, and correctness. Formal methods can be used in the initial statement of a customer’s requirements, through system design, implementation, testing, debugging, maintenance, verification, and evaluation. When used early in the system development process, they can reveal design flaws that otherwise might be discovered only during costly testing and debugging phases. Well-known or commonly used formal methods include Temporal Logic, LOTOS, SDL, etc. The formal framework give rise to the KAOS methodology [14] in the generation of alternative system designs from high-level goals defined in temporal logic [1].

3.CASE STUDY

3.1 Creating model of the example with DFD

Fig.1. illustrates the example of ‘Network-based projects manager’ using DFD. Circle denotes process while arrow denotes dataflow. A lifecycle of a project from Set Up to Check&Accept is modeled, and the input and output of every step are also expressed, for example, the input of Track&Control is annual report and the output is final report.

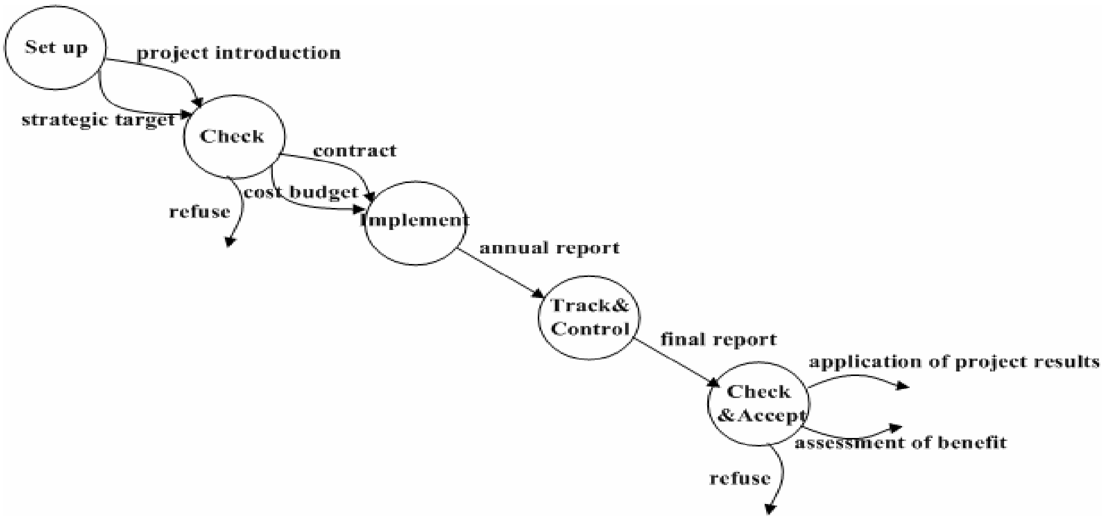


Fig.1 creating model of the example with DFD

3.2 Using UML to set up models

Two common views-use case view and sequence view are demonstrated in Fig.2 and Fig.3. In use case view, system functions are expressed as use cases with the stereotype of《functional》 such as Contract project, First check, while Non-functional requirements of system are described as use cases with the stereotype of《non-functional》,such as High quality & Low cost. In sequence view, actors interact with others through sending and receiving information, for example, Project Contactor sends information 'apply for projects' to 863 Committee.

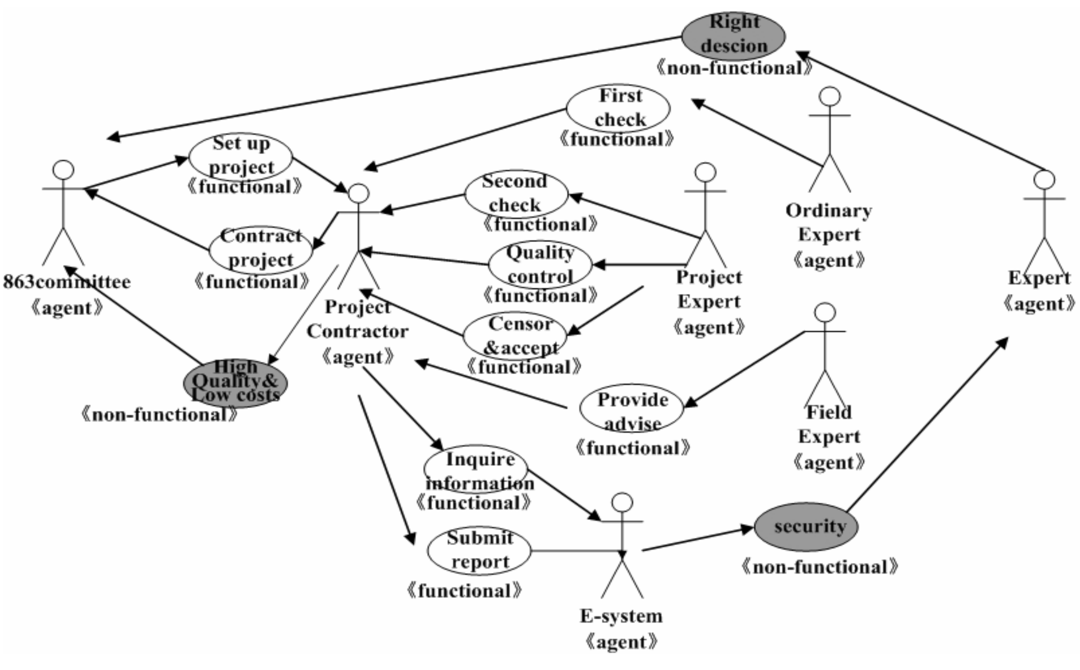


Fig.2. Use case view

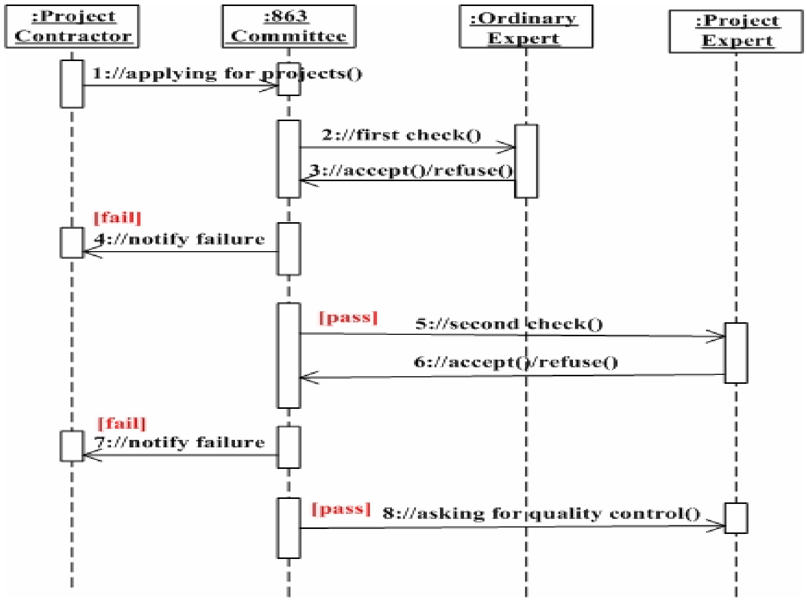
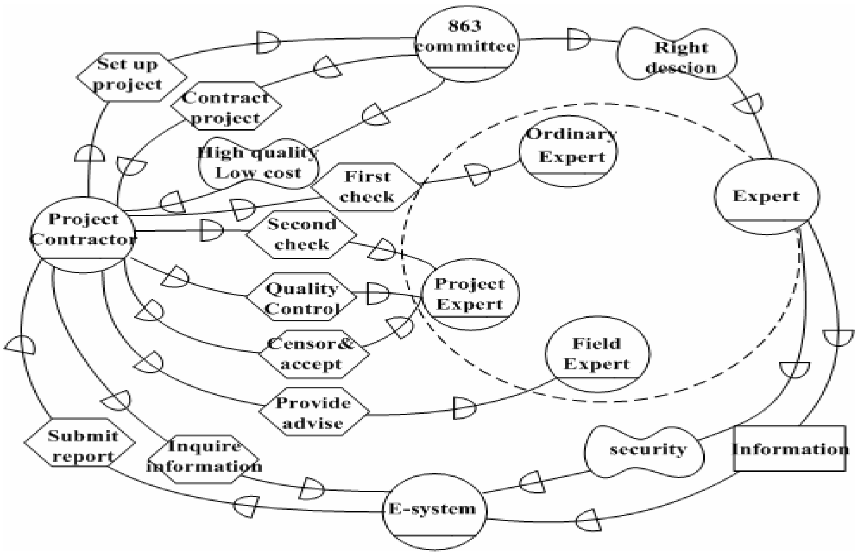


Fig.3 Sequence view

3.3 GRL models

GRL has three main categories of concepts[15]: intentional elements, intentional links, and actors. GRL elements and links are intentional in that they are used in models that answer questions about intents, motivations and rationales. The intentional elements in GRL are goal, task, softgoal, resource and belief; Intentional links in GRL include means-ends, decomposition, contribution, correlation and dependency links.



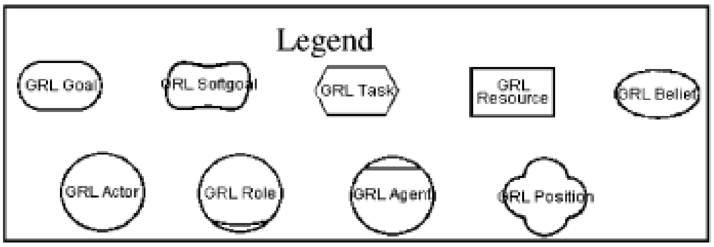


Fig.4 .Major players and role dependency relationships

In Fig.4, Major players in 'network-based project manager' domain and role dependency relationships are captured. Take 863 Committee and Project Contractor for example. 863 Committee depends on Project contractor to Contract projects; Project contractor depends on 863 committee to Set up projects. Moreover, achieving softgoal High quality & Low cost depends on Project Contractor.

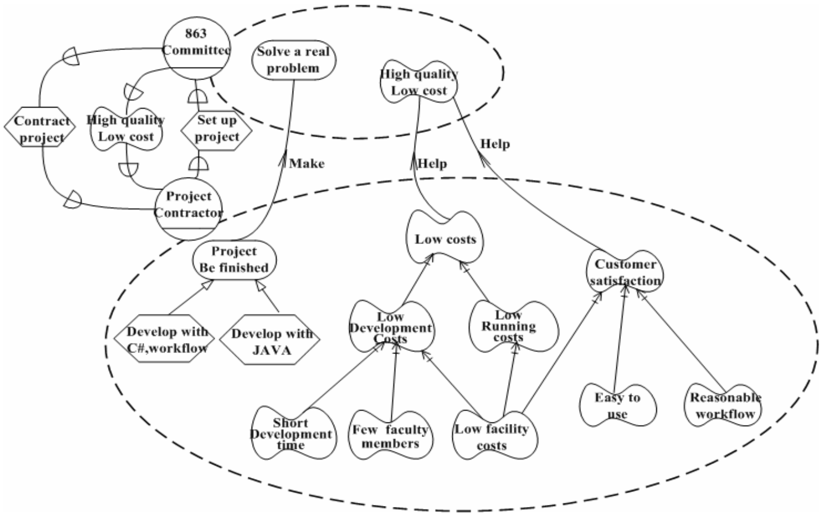


Fig.5. Explore possible designs for the future system (high level).

In Fig.5, the two goals of 863 Committee can be reduced into two general goals applicable to all Project Contractor—Low costs and Customer Satisfaction. The two softgoals, together with the goal of Project Be Fininished, are refined individually. Since the two most common choices for finishing a project are to develop with C# or JAVA. The softgoals Low Costs and Customer Satisfaction are refined in detail.

3.4 Creating model with UCM

Imagine tracing a path through a system of objects to explain a causal sequence, leaving behind a visual signature. Use Case Maps capture such sequences. [16]They are composed of:

• **start points** (filled circles representing pre-conditions or triggering causes)

• causal chains of **responsibilities** (crosses, representing actions, tasks, or functions to be performed)

• **end points** (bars representing post-conditions or resulting effects).

• **components** The responsibilities can be bound to components, which are the entities or objects composing the system.

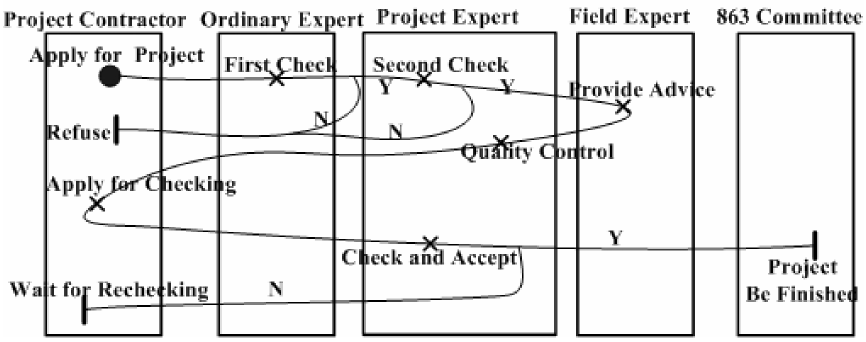


Fig.6 UCM models

'Network-based project manager' model with UCM is illustrated in Fig.6.Actors such as Project Contractor of system are captured as components, The wiggly lines are paths that connect start points, responsibilities, and end points, denoting the business workflow of a project. A responsibility is said to be bound to a component when the cross is inside the component. In this case, the component is responsible to perform the action, task, or function represented by the responsibility, such as First Check and Provide Advice.

3.5 Creating model with Temporal Logic

Temporal Logic defines the goal and relationship as follows. Taking **Goal** definition for example, the first line declares Achieve method, the parameter of which is the content of goal- ProjectBeFinished; the second line captures the related entities, Project Contractor, Project Expert, and 863 Committee; the third line refines the goal with two sub-goals- Passing the Check and Obtaining the Approval; the fourth line describes the goal with natural language while the fifth line with mathematical formula.

Goal Achieve[ProjectBeFinished]

Concerns Project Contractor, Project Expert, 863 Committee

Refined to Passing the Check, Obtaining the Approval

InformalDef The project is finished after the check of Project Expert and the approval of 863 Committee within C days.

FormalDef $\forall$ p: Project Contractor, e: Project Expert,com: 863 Committee

check(p,e) ^ approval(p,com) $\Rightarrow \Diamond_{\leq Cd}$ finished(com,p)

Relationship check

Links Project Contractor {Card:0:1}, Project Expert {Card:1:N}

InformalDef Project Expert checks the project iff Project Contractor applies 863 Committee for checking

DomInvar $\forall$ p: Project Contractor, e: Project Expert, com: 863 Committee

check(p,e) $\Leftrightarrow \blacklozenge$ apply(p, com)

($\Diamond$ denotes next operator. $\Diamond$ A means there exist a state A in future;

$\Diamond_{\leq Cd}$ A means in C days there must exist state A;

$\blacklozenge$ denotes previous operator. $\blacklozenge$ A means there exist a previous state A.

)

4.ANALYSIS AND EVALUATION OF MODELING METHODS

4.1 Structural modeling language- DFD

As illustrated by Fig 1, the DFD model provides a view for the whole business process of system and it objectively and precisely describes the flowing of data. DFD is thus easy to read and to understand, even if the reader is not a professional of computer field. Also according to Fig 1, DFD only captures the basic logic function and workflow of system, ignoring the details of implement and therefore successfully reduce the complexity and ambiguity of the whole system.

On the other hand, DFD has its flaws. As a semi-formal notation, DFD has definite syntax but lacks formal semantics. Therefore, as demonstrated by Fig1, the content and context described by DFD is incomplete and inadequate, for it only expresses the direction of dataflow and the basic function of system but other aspects such as time control, interface management, are beyond the scope of DFD. Moreover, DFD can never bridge the gap between the models and specification. In order to obtain a language that can verify the specification through models, we need provide precise semantics for the language.

4.2 OO modeling technique-UML

As illustrated by examples, use case view in UML visually expresses the interaction between actors and system. It is easy to understand, even for readers who know little about the system. Traditional development process such as DFD is somewhat incidental for it lacks early requirements analysis and thus difficult for users to express their input for the system. But use cases successfully overcome this disadvantage and facilitate users to understand the system from their own perspectives. Sequence view in Fig.3 describes the sequence of system interaction according to the time and events. Sequence view is based on use case and thus can well express the interaction of actors and system. As illustrated in Fig.3, 863 Committee corresponds to Project Contractor through sending back information and to Ordinary Expert and Project Expert. Moreover, correspondence between use cases is prohibited; therefore, the complexity of system is significantly reduced.

However, there are three limitations for UML. First, capturing of use case is difficult. Generally, listing all use cases, especially the high-level use cases, is relatively easy, but precisely express extensive use cases is rather difficult. The other limitation of UML is that the conflicts between use cases can not be described. If conflicts do exist in system, they have great impact on use cases. The third limitation of UML lies in its disability of expressing concurrency of system. In UML, a use case can be interrupted only by one another use case or event. This mechanism sometimes can not express complex information system.

4.3 Goal-based language- GRL

According to the example, GRL describes the objectives that the system should achieve through the cooperation of actors in the software-to-be and in the environment. It captures ‘why’ the data and functions are there, and whether they are sufficient for achieving the high-level objectives, supplementing the traditional ‘what’ and ‘how’ questions of system. GRL provides constructs for expressing the requirements and high level design process while ignoring the details of implement. Thus, a big view of system and its environment is easily acquired. The strength of GRL modeling is that it puts the design in a broader context, it considers from different stakeholders’ viewpoints, and seeking for a balanced solution for all. Another advantage of GRL is that not only functional requirements but also NFRs (in other words, the quality requirements) are dealt with.

While goal orientation can be highly useful for requirements engineering, goals are sometimes too abstract to capture all at once. Often they are discovered and become explicit only after a deeper understanding of the system has been achieved. In particular, users and developers often find it natural to think about operational scenarios about using the hypothetical system. More conventional requirements and design notations typically answers the ‘what’ questions such as ‘what should the system do to provide an efficient management of projects?’ or ‘what is the process of managing a project?’ Thus, GRL is always combined with UCM, as proposed by ITU, to provide a comprehensive description of system.

4.4 Scenario-based framework-UCM

The main strength of UCMs resides in their capacity to visually integrate behavior and structural components in a single view [17]. By combining behavior and structure in one view in an explicit and visual way, UCM enable the understanding of scenario paths in their context as well as high-level architectural reasoning. As UCM paths can easily be decoupled from structures and are less likely to change than underlying architectures, UCMs improve the reusability of scenarios and lead to behavioral patterns that can be utilized across a wide range of applications. If a structure is modified, path elements need only to be reallocated to appropriate components. UCMs facilitate abstract thinking at the right level of detail for architectural design since UCM descriptions abstract from messages, protocols, communication constraints, data, control, and structural evolution while focusing on general causal flows between causes, responsibilities, and effects.

UCMs are by no means a complete notation for all issues that arise in system development; scenarios are usually partial and incomplete. [18]They only deal with the issue of getting a high level view of emergent behavior of systems and are important only because this is such a difficult view. UCMs supplement with other techniques that give views of emergent behavior in more detailed terms, to give a different perspective. They neither replace such techniques nor depend on them. UCMs may be usefully integrated with other high-level techniques that have other purposes than behavior description, such as high-level class relationship diagrams in object-oriented design [19].

4.5 Formal Method (FM)

Formal Method has its mathematical and logical basis. Applying formal method helps clarify a customer's set of informally stated requirements. A specification helps crystallize the customer's vague ideas and reveals contradictions, ambiguities, and incompleteness in the requirements. Formal methods can be used in any stage of system development. When used early in the system development process, they can reveal design flaws that otherwise might be discovered only during costly testing and debugging phases. What's more, FM is well supported by many computer-based tools and thereofore facilitates the automation of completeness check and reduces the costs of late analysis.

FM users must acknowledge two important limitations between the mathematical world and the real world. First, FM excessively depends on mathematics and logics, the complex expressions and the obscure semantics of which are usually hard to understand. The second limitation is the difficulty of mapping from the real world to some abstract representation. In reality, it is impossible to formally model many environmental aspects such as unpredictable or unanticipated events, human error, etc.

5.CONCLUSIONS AND FUTURE WORK

This paper presents five prevailing modelling methods in software requiements, which respectively illustrated by an example of 'network-based project manager' . For each method, the main characteristics are outlined and the advantages and disadvantages are analyzed. In sum, according to the readability, DFD and UML are of advatages; in the logicality, GRL is better; in the dynamic and collaborative respect, UML and UCM are recommended; when exactitude is concerned, Formal Method is the best.

Recently, how to combine the use of these frameworks to capture various aspects of system requirements has been recognized as a focus of requirements engineering. The main one is URN [20], by International Telecommunication Union. This proposal currently studied combines GRL, for the description of business goals, non-functional requirements, alternatives, and rationales, with UCMs for the specification of functional requirements in the form of causal scenarios. The other combining plan is to add UCM view to UML, or extend basic UML concepts with UCM model concepts. Another major work item is the combination use of UCM and Formal Method (SDL for example). The proposed method employs UCM to capture system requirements and then converts these scenarios into SDL model, which facilitates the acquirement of code(c, java, ADA, etc).

In future, when we deal with requirements analysis, we can combine more than two modeling language to provide a comprehensive description of system requirements. A complete requirements analysis of information system includes six parts [21]- Structural Diagrams, Behavioural Diagrams, FR descriptions, NFR descriptions, Informal Requirements and Testing and Performance Languages. In Structural Diagrams, SDL, UML Class View, Component View, and Deployment View could be used; in Behavioural Diagrams, MSC/SDL or UML Sequence View and Collaborative View can be employed; on the other hand, UCM unifies the structural and behavioural aspects at the system level; functional requirements can be described by UCM while non-functional requirements by GRL; informal language can employ the use case description language; TTCN and LQN can be used as the Testing and Performance Languages for system.

REFERENCES

[1] A.V. Lamsweerde, "*Requirements engineering in the year00: a research perspective"*, in: The Proceedings of the 22nd International Conference on Software Engineering, Limerick, June 2000, ACM press, New York.

[2] D.T.Ross, "*Structured Analysis(SA): A language for Communicating Idea"s,* IEEE Transactions on Software Engineering, Vol.3, No.1,1977,16-34.

[3] Booch G, Jacobson I, Rumbaugh J. "*Unified modeling language user guide [M]"*. Addison Welsley Object Techno logy Series, 1999.

[4] E. Yu, J. Mylopoulos, "*Why goal-oriented requirements engineering",* in: E. Dubois, A.L. Opdahl, K. Pohl (Eds.),Proceedings of the Fourth International Workshopon Requirements Engineering: Foundations of Software Quality, Pisa, Italy, Presses Universitaires de Namur, Paris, June 1998, pp. 15–22.

[5] GRL web site. http://www.cs.toronto.edu/km/GRL/

[6] E. Yu, *"Towards modelling and reasoning support for early phase requirements engineering",* in: Proceedings of the Third IEEE International Symposium on Requirements Engineering (RE '97), Washington, DC, USA, January 6–8, 1997, pp. 226–235.

[7] J.G. Leite, J. Doorn Hadad, G. Kaplan, "*A scenario construction process, Requirements"* ,Eng. 5 (1) (2000) 38–61. ARTICLE IN PRESS L. Liu, E. Yu / Information Systems 29 (2004) 202 187–203

[8] A. Sutcliffe, N. Maiden, S. Minocha, D. Manual, "*Supporting scenario-based requirements engineering",* IEEE Trans. Software Eng. 24 (12) (1998) 1072–1088.

[9] K. Weidenhaupt, K. Pohl, M. Jarke, P. Haumer, *"Scenario management in software development"t:* current practice, IEEE Software (15) (1998) 34–45

[10] J.M. Carroll, Introduction: the scenario perspective onsystem development, in: J.M. Caroll (Ed.), Scenario-Based Design: *"Envisioning Work and Technology in System Development"*, Wiley, 1995, pp. 1–17.

[11] M. Jarke, X.T. Bui, J. MC arroll, *"Scenario management—an interdisciplinary approach, Requirements"* Eng. J. 3 (3–4) (1998) 155–173.

[12] R.J.A. Buhr, *"Use case maps as architectural entities for complex systems"*, in: Transactions on Software Engineering, Vol. 24, No. 12, IEEE Press, New York, December 1998, pp. 1131–1155.

[13]J. Wing, *"A Specifier's Introduction to Formal Methods"*, IEEE Computer,23(9):8-23, September 1990.

[14] A.Dardenne, A.van Lamsweerde and S. Fickas, *"Goal-directed Requirements Acquisition"*, Science of Computer Programming, Vol.20,1993,3-50

[15] Lin Liu, "*Designing information systems in social context: a goal and scenario modelling approach"*,Information Systems 29, 2004

[16] Daniel Amyot, *"Use Case Maps Quick Tutorial Version1.0"*, University of Ottawa ,September 1999

[17] Daniel Amyot, Gunter Mussbacher, *"Bridging the Requirements/Design Gap in Dynamic Systems with Use Case Maps (UCMs)",* ICSE 2001: 743-744. 7, EE

[18] R.J.A. Buhr, *"Use Case Maps as Architectural Entities for Complex Systems"*, In:
IEEE Transactions on Soft-. ware Engineering, 24 (12), Dec. 1998, 1131-1155. 8.

[19] R.J.A. Buhr, *"Understanding Macroscopic Behaviour Patterns in Object-Oriented Frameworks, with Use Case Maps"*,chapter of a forthcoming Wiley book on OO Application Frameworks, ed. Mohamed Fayad.

http://www.sce.carleton.ca/ftp/pub/UseCaseMaps/uoof.ps

[20] Z.150*: Users requirements notation,* Recommendation to ITU-T Study Group17, Available at URN web site. http://www.usecasemaps.org/urn/
[21] Daniel Amyot, Gunter Mussbacher, "URN: Towards a New Standard,for the Visual Description of Requirements"

The Team-Based, Information Technology-Enabled Organizations

Dr. Hamid Tohidi
Islamic Azad University, South Tehran Branch
Industrial Engineering Department
Tehran, Iran
Email: H_Tohidi@azad.ac.ir

Mohammad Mehdi Jabbari
Islamic Azad University, South Tehran Branch
Tehran, Iran
Electrical Engineering Department

Abstract

If teamwork is the key to effective organizations, information technology is the key to effective teamwork. The first part of this paper introduces the concept of using teams and information technology to effect organizational performance. Each element on the left side of this "equation" (the teams and the information technology) is of equal importance in bringing about the "result" (organizational performance); in fact, the synergistic relationship between teams and information technology makes them far more effective agents of change together than either could be apart. In the second part of this paper a quantitative model is presented that makes clear the significance of this relationship and presents the basic components of the process intended to exploit it: the mutual design of teams and information technology.

Keywords:

Teamwork, performance, Information Technology, organization.

1. INTRODUCTION:

Teams and information technology are two of the most important developments in organizations today. Year after year, organizations increase their investment in new information systems and use teams to do more and more of their work. Vast amount of time, money, and effort are spent with the expectation that their impact on the bottom line will eventually justify their costs. But many organizations are disappointed in the results. Few are getting the bang from the many bucks they spend each year to create teams and to develop new information systems. The challenge facing these organizations is how to fulfill the potential of these two promising and complex developments.

The premise of this paper is that each is necessary to take full advantage of the possibilities created by the other-that is, information technology can make teams more effective, and teams can help fulfill the promise of new information technology. Together, teams and new information technology can catalyze dramatic improvements in organizational performance. How to make this happen is the subject of this paper.

Research suggests that teamwork is "an integral tool aiding continuous improvement in work operation" [2]. However, the empirical evidence regarding team effectiveness is limited and often has the form of anecdotes or descriptive case studies; stories of huge cost saving and quality improvements abound [11, 20, 14, 5, 3, 9].

In several studies it is hypothesized that variation in team performance can be explained by differences in team structure [4, 12, 16, 19, 24].

Tranfield & Smith [23] examined in depth the form teamworking which take in a number of teamworking organizations across the study to ascertain their similarities and differences.

According to Doorewaard, et al. [8], the performance in team-based working also largely depends on the employee's authorities and function design; i.e. to what extent the planning, performing and controlling responsibilities are integrated in the team tasks.

Delarue, Gryp, et al. [7] investigated the impact of specific structure team types on the performance of the organization, measured by labour productivity.

When a new project starts, one of the most difficult tasks is to choose the most suitable members of the work team. The most relevant factors may be grouped into three categories: I) Individual characteristics; П) Social characteristics; Ш) Temporal and economic costs [1].

Advances in information technology have enabled new organizational forms and new ways to structuring work. In the age of the knowledge economy, most tasks accomplished as a part of one's job require some forms of communications [17].

For long, researchers have investigated organizational communications, both formal and informal. Yet, we still need to understand better how communication based tasks can be better supported to lead to efficiencies in an environment where individuals are distributed. Regardless of specific type of work environment, individuals should manage multiple relationships to work productively [20].

Team can enable a company to execute more quickly, changes are made easily, allowing the company flexibility [18].

Each member of a group adds more information, perspectives, experiences and competencies [10].

1-1. The structure of the paper is as follow: Section two introduces the concept of using teams and information technology to affect organizational performance and provides basic background information on the focus of teams, information technology and organizations. The basic assumptions for the considered model are presented in section three. The parameters, used in the considered model are introduced in section four. Section five presents a model that can be used to determine the value of teamwork performance related to information technology and team size factors. Section six includes a sensitivity analysis to the model, based on information technology, and the last Section is a conclusion part.

2. THE CONCEPT OF USING TEAMS AND INFORMATION TECHNOLOGY TO AFFECT ORGANIZATIONAL PERFORMANCE

Before describing how to initiate and conduct the model, we need to know some basics. In this section, we define, describe, and otherwise clarify what we mean by teams, technology, and organizations-the cornerstones of the mutual design of teams and information technology and the term knowledge.

1-2. Information technology: in this paper, the term information technology refers to the information and communications technologies, systems, and tools used by individuals and teams in their work. It does not refer to production technology that is controlled directly by a computer and does not require direct human involvement. As Shoshana Zuboff [21] notes in her book In the age of the Smart Machine, however, "The devices that automate by translating information into action also register data about those automated activities, thus new streams of information". This process of "informating" as she calls it, can be used to indicate potential problems, provide diagnostic data on the source of these problems, and help project future demand and production. For these uses, human operators must be brought back into the loop to analyze this information and make judgment and decisions based on these analyses.

Our definition therefore refers to information systems in which humans are in the loop, in which knowledge-based, human interactions are critical to the way the technology performs. It excludes robotics and highly automated, numerically controlled manufacturing process, but it does include the "informating" systems that help human operators troubleshoot the process.

2-2. Organizations: We use the term organization in two ways. Our general use cast a wide net and includes large and small manufacturing and service business, government agencies at all levels, and nonprofit firms of all kinds, to mention some of the most obvious categories. When we refer to an organization as a focus for mutual design of teams and information technology, however, it most often means something less than the entire firm, agency, or institution. So, organization, as we use it within the context of "Mutual Design of teams and information technology" usually refers to the level within the larger body that has the authority to implement the kinds of systemic change needed to support individual teams and to integrate the activities of many.

There is another sense in which our use of organization is unusual, a sense that reflects the ongoing evolution of organizational forms. During the hey-day of mergers and acquisitions in the 1980s, our notions of what constitute organizational boundaries began to change. The emerging era of transnationals, alliances, and metaorganizations may finish the job, assisted to a considerable degree by internal and external communications networks. As a result, organizational boundaries have grown increasingly permeable and difficult to identify. With access to networked systems, these boundaries need no longer be barriers to cross-organizational collaboration [15]. Teams made up of individuals and groups from different organizations can work together on projects that exceed the resources and capabilities of any single organization. Identifying the appropriate "organizational context" for instituting cross-organizational change through teams and technology will challenge the ingenuity of even the most resourceful and creative mutual design team.

3-2. Teams: When it comes to organizational effectiveness, no topic is more important than teams. Moreover, no topic is more popular or more widely misunderstood. We have all worked in teams, so we all figure that we know something about them. Many of us played on athletic teams in our youth, and just about all of us have been on work teams as adults. It is probably not surprising then that almost everyone thinks they know a great deal about them. In many cases this knowledge seems to boil down to the core belief that effective teamwork require little more than commitment and a willingness to make personal sacrifices in the service of team goals. Exhorting people to pull together and facilitating interpersonal interactions are the favored means for eliciting this commitment and sacrifice.

First we need to define what we mean by team. Calling a group of people a team does not necessarily make them one. The characteristics that define a team are task interdependence and a shared goal or purpose-that is, the work of each member is dependent on the work of at least some of the others. For example, in a production team one member may pass on the product of his or her work for another member to work on. Or the interdependence may be more interactive, such as in a writing team, where the members go back and forth in rapid, overlapping succession creating complex concepts and searching for ways to express them. By this definition, a word- processing pool or a department of field engineers who work on separate projects are not teams. They work independently of each other except for the occasion when one has to back up another because of work overload, absence, or poor performance. For the most part and most of the time, they are not interdependent. Therefore, they are not, for our purposes, teams.

Five types of teams can be found in organizations today, (1) work teams, (2) project and development teams, (3) parallel teams, (4) management teams, and (5) ad hoc networks [22]. These types define a continuum of sorts. At one end are the work teams, which tend to be formal, ongoing, and have permanent members. At the other end are the loose, informal collections of people we call ad hoc networks (to distinguish them from technology networks). They are less bound by time, and their membership is more fluid and diffuse. The other team types fall somewhere in between.

4-2. Knowledge: Knowledge is the key to organizational performance. Knowledge work, knowledge workers, the knowledge society-these expressions reflect the growing recognition that knowledge is the key strategic resource of the postindustrial organization. To succeed in the fast-paced, intensely competitive global market place, organizations will have to learn how to generate, organize, manage, and apply knowledge more effectively. Whether solving production problems on the shop floor, creating new products in the laboratory, or mapping strategy in the executive suite, organizations need access to knowledge that is as broad as it is deep [6].

That is why organizations are now so intensely interested in teams and information technology. The link between information technology, knowledge, and organizational performance is clear. Information technology provides access to diverse sources for specialized information and enhances our ability to analyze, manage and apply this information to our work. While the link between teams, knowledge, and organizational performance may be less obvious, it is just as important.

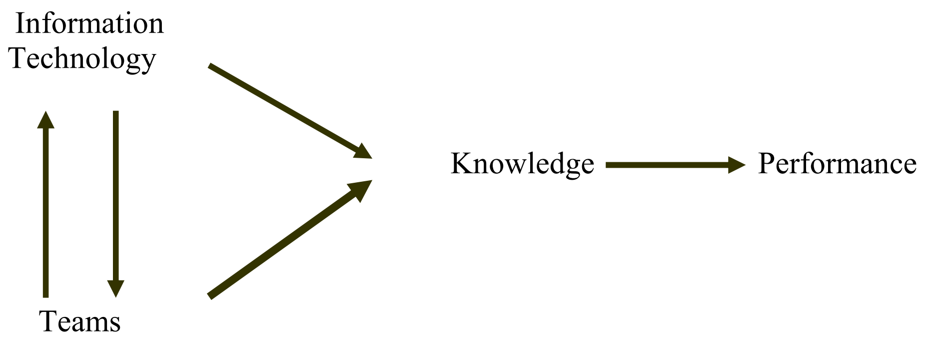


Figure 1: Synergy between Information Technology and Teams Enhances Knowledge and Improves Performance

A team brings together different individuals who know and can do different things. It is a means of pooling and using the diverse "knowledge" and skills of its members to accomplish mutual goals [6]. When there is a synergy between teams and information technology, as pictured in figure 1, the contribution of the two to knowledge and organizational performance is greater than the sum of the parts

3. THE MODEL ASSUMPTIONS:

The assumptions of the model are as follow: there is a particular team as in Figure 2. Team size is n and its members have full information interaction between each others. The team is assigned a task which requires coordinated individual effort. The individual effort is production oriented, where production has a broad interpretation. It may, for instance, be consulting services, new product development, decision making, and so on. The individual production leads to some type of output which adds value to the organization and the team. Production also generates information which is needed in the coordination of activity between team members. Coordination takes place by each individual processing the information generated by the other team members.

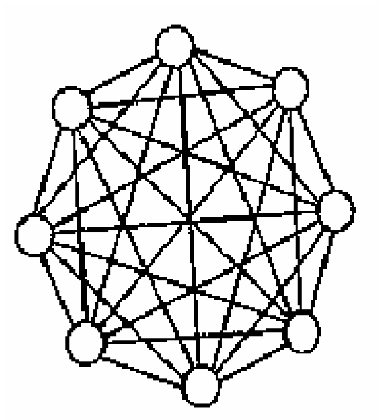


Figure2: A team with full information interaction between members

1-3. General assumptions:

In any primal model it needs to suppose some assumptions and then it can be extended to a complete model by developing the assumptions. So in this section some primal assumptions are considered that have important role in the considered model.

- An individual divides his/her time between production and information processing.
- If one unit of time is exclusively devoted to production, exactly one unit of output is generated.
- For each unit of output, there is also a unit of information generated.
- Each individual has to process all information received from the other team members in order to coordinate the team task.
- It takes less than one time unit to process one unit of information.

4. PARAMETERS USED IN THE CONSIDERED MODEL:

n : The number of team members.
α : The fraction of a time unit required for processing a unit of information provided by other team members about their production. Also, $0 < \alpha < 1$.
$\Omega(n)$: The fraction of time an individual can spend on production after processing the information received from the other members.
$P(n)$: Output of team (Quantity of production.

5. THE MODEL:

According to the considered assumptions an individual divides his/her time between production and information processing and if one unit of time is exclusively devoted to production, exactly one unit of output is generated. Also for each unit of output, there is a unit of information generated and it is assumed that it takes less than one time unit to process one unit of information. Assuming that all received information is processed, the processing of information during one time period consumes as follow:

$$\alpha.\Omega(n).(n-1) \text{ units of individual's time} \tag{1}$$

The remaining fraction of the time period which can be spent on production, is given by:

$$\Omega(n) = 1 - \alpha.\Omega(n).(n-1)$$
$$= \frac{1}{1+\alpha(n-1)} \tag{2}$$

So equation (2) is equilibrium condition on information generation and information processing. As the size of team increases, each individual will spend a larger proportion of her time processing information provided by other team members and, hence, the time left for production is reduced. In practical terms, this implies that as the team size grows, the individual team members get saturated with information and productivity drops [13]. The total production of the team during one time period is then:

$$\mathrm{P}(n) = n.\Omega(n) \tag{3}$$

$$\mathrm{P}(n) = \frac{n}{1+\alpha.(n-1)} \tag{4}$$

Theorem 1: $P(n)$ is a concave, monotonically increasing function of n for all values of $0 < \alpha < 1$

Proof:

$$\frac{dP(n)}{dn} = \frac{1+\alpha.(n-1)-\alpha.n}{[1+\alpha.(n-1)]^2} \tag{5}$$

$$= \frac{1-\alpha}{[1+\alpha(n-1)]^2} > 0 \qquad , \qquad 0 < \alpha < 1 \tag{6}$$

$$\frac{d^2P(n)}{d\ n^2} = \frac{-2\alpha 0(1-\alpha)}{[1+\alpha.(n-1)]^3} < 0 \qquad , \qquad 0 < \alpha < 1 \tag{7}$$

Hence, $P(n)$ is a concave, monotonically increasing function in n .

Theorem 1 indicates that team output can be increased by adding members to the team. However, the marginal product of team members is decreasing due to the increased coordination effort required so that for each added team member, there is a smaller and smaller increase in output.

Beyond some value of n, the marginal cost of an additional team member exceeds the marginal value of the team's production.

Theorem 2: For any non zero α , $P(n)$ is a bounded function.

Proof: Theorem 1, $P(n)$ is concave and monotonically increasing function of n . Also, $P(0) = 0$.

$$\lim_{n\to\infty} P(n) = \lim_{n\to\infty} \frac{n}{1+\alpha.(n-1)} = \frac{1}{\alpha} \tag{8}$$

Hence, $P(n)$ is a bounded function.

The practical implication of Theorem 2 is that the maximum total production of a team during one time period depends on the speed at which the team members can coordinate their activities with their peers.

To increase the team's maximum production capacity, it is necessary to change the communication and processing technology (i.e. decrease the value of α). Or, the work has to be reorganized so that each team member does not process all of the information provided by the other members.

Theorem 3: The marginal product of team size is asymptotically zero.

Proof:

$$\lim_{n\to\infty} \frac{dP(n)}{dn} = \lim_{n\to\infty} \frac{1-\alpha}{[1+\alpha.(n-1)]^2} = 0 \tag{9}$$

Theorem 1 shows that the marginal product of team size is decreasing and theorem 3 states that the marginal product of team size is asymptotically zero. These two facts imply that for a one-period production effort, there is a single optimal team size if the cost per team member is positive and marginally non-decreasing. This condition is equivalent to the well-known profit maximum condition marginal cost equals marginal revenue in economic theory.

6. SENSITIVITY ANALYSIS:

In the following the effect of changing information technology on team output is studied. An improvement in information technology implies that the time it takes to communicate and process a unit of information is reduced. Thus, as information technology improves the parameter α decreases.

Although information technology improvements are likely to occur in discrete increments, it is useful to study the first order derivative of the total team output.

Theorem 4: $P\left(n,\alpha\right)$ is monotonically decreasing function of α for all values of $0<\alpha<1$.

Proof:

$$0<\alpha<1 \quad , \quad \frac{\partial P(n,\alpha)}{\partial\alpha}=\frac{-n^2}{(1+\alpha.(n-1))^2}<0 \qquad (10)$$

Hence, $P\left(n,\alpha\right)$ is monotonically decreasing in α.

Thus, as information technology improves (α is reduced), team output increases. This result is consistent with expectation since less time spent on information processing implies more time spent on production.

Similarly, as information technology improves, so does the maximum output of the team. Let Δ be the reduction in processing time of one unit of information so that $\alpha'=\alpha(1-\Delta)$. Then, the increase in maximum team production is:

$$\frac{1}{\alpha'}-\frac{1}{\alpha}=\frac{1}{\alpha(1-\Delta)}-\frac{1}{\alpha}=\frac{\Delta}{1-\Delta}\frac{1}{\alpha} \qquad (11)$$

In marginal terms, there is a trade-off between adding manpower to a team and improving the information technology support to the team.

The following example will illustrate the concept. Consider a team with 20 members and information technology which allow team members to process information at a rate of 20 units per time period (i.e. $\alpha=0.05$). According the formula (4) the output of this team is 10.256 per time period. If the team size is increased to 30 members it's output will be 12.245.

The same output per time period can be achieved by information technology improvement with rate of information technology processing (i.e. $\alpha=0.033$).

If the cost of 10 new team members is higher than the cost of upgrading the information technology, then an information technology upgrade is the best decision. If there are a number of technology improvement options, there may be a mix of technology improvement and team size increase that will yield the most cost efficient solution to increase team output.

Similarly, if demand for the organization output is fixed, the organization can achieve a productivity increase by investing in improved communication and processing technology and reduce the number of team members.

If technology investments change the information processing rate (i.e. $\alpha = 0.033$), in the above Example, the team size can be reduced to 20 members without reducing production. Thus, by investing in communication and information processing technology, labor cost can be reduced by 33.3%. Considering the significant price reduction trends in communication and information processing technology, this explains the substantial reduction in team size, often referred to as corporate downsizing, that take place in modern post-industrial economies.

Figure 3 & 4 show team output per period for various levels of information technology factor (α)

7. CONCLUSIONS:

In this paper, a model is presented that can be used to determine the value of teamwork performance related to information technology and team size.

According to this model, team output can be increased by adding members to the team. But beyond some value of team size, the marginal cost of an additional team member exceeds the marginal value of team's production. Also to increase the team's maximum production capacity, it is necessary to change the communication and processing technology.

If the cost per team member is positive and marginally none decreasing, there is a single optimal team size.

If there are a number of technology improvement options, there may be a mix of technology improvement and team size increase that will yield the most cost efficient solution to increase team output.

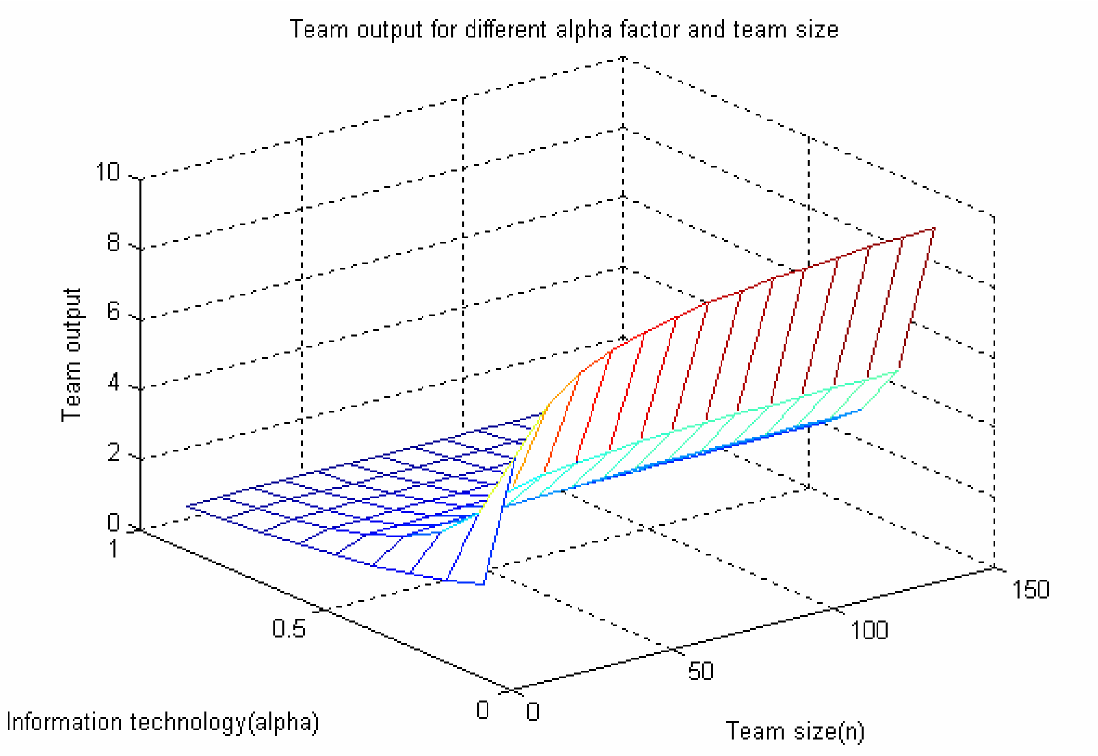


Figure 3: Team output per time period in quantity of production versus information technology and team size

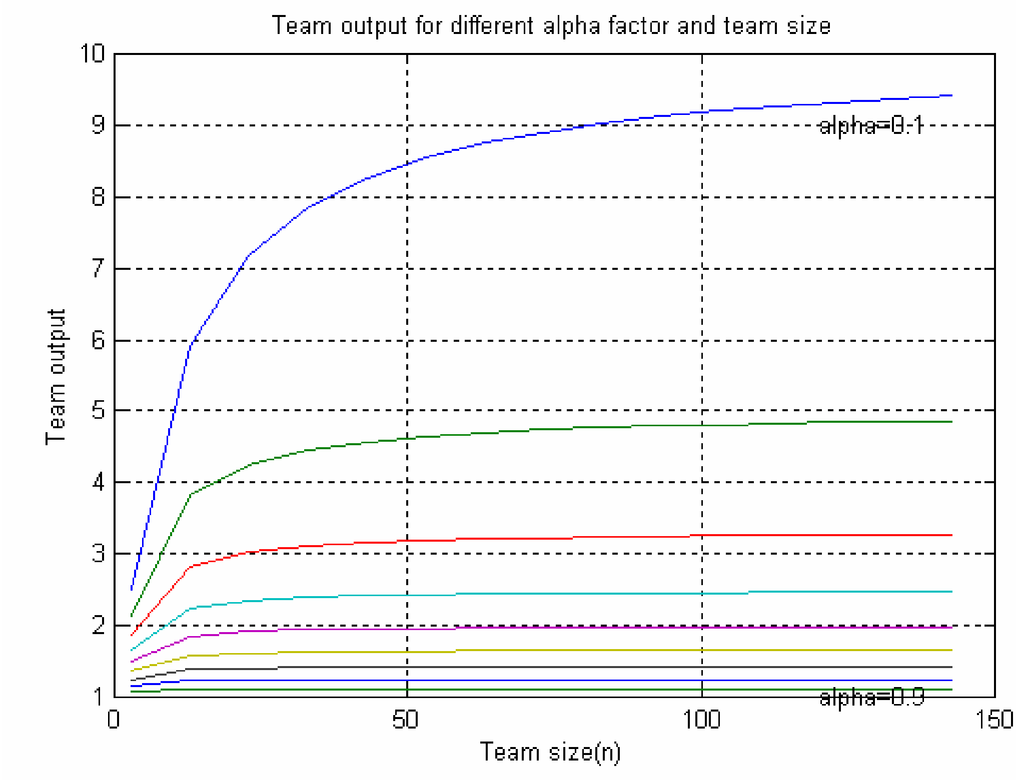


Figure 4: Team output per time period in quantity of production versus information technology and team size

REFERENCES:

[1] Antonio Moreno, Aida Valls & Mata Marin, 2003, Multi-agent simulation of work teams, CEEMAS, 281-291.
[2] Banker, R.D., Field, J.M., Schroeder, R.G. & Sinha, K.K., 1996, Impact of work teams on manufacturing performance: A longitudinal field study, Academy of Management Journal, 39(4), pp.867-890.
[3] Benders, J. & Van Hootegem, G., 1999, Teams and their content: moving the team discussion beyond existing dichotomies, Journal of Management Studies, 26, 5, pp. 609-628.
[4] Cohen, S.G. & Bailey, D.E., 1997, What makes teamwork? Group effectiveness research from the shop floor to the executive suite, Journal of Management, 23, pp. 13-43.
[5] Cohen, S.G. & Ledford, G.E., 1994, The effectiveness of self-managing teams: A quasi-experiment, Human Relations, 47, 1, pp. 13-43.
[6] Drucker, P., 1994, The age of Social Transformation, Atlantic Monthly, 53-80.
[7] Delarue, A., Gryp, S. & Van Hootegem, G., 2003, Productivity outcomes of teamwork as an effect of team structure, Paper presented at the 7th International Workshop on Teamworking, October 2nd-3rd, Prato, Italy.
[8] Doorewaard, H.; Huys, R. & Van Hootegem, G., 2002, Team responsibility structure and team performance, Personnel Review, 31(3), pp. 356-370.
[9] Glassop, L.I., 2002, The organizational benefits of teams, Human Relations, 55(2), pp. 225-249.
[10] Gmelch, W.H., 1984, Productivity Teams: Beyond Quality Circles, Toronto, Canada, John Wiley & Son, Inc.
[11] Gupta, Y.P. & Asf, D., 1994, Excellence at Rohm and Haas Kentucky: A case study of work team introduction in manufacturing, Production and Operation Management, 3(3), pp. 186-200.
[12] Hackman, J.R., 1987, The design of work teams, in J.W. Lorsch(ed.), Handbook of organizational behavior, pp. 315-452, Prentice-Hall, Englewood Cliffs.
[13] Hilz, S.R. & M. Turoff, 1985, Structuring Computer-Mediated Communication Systems to Avoid Information Overload, Communication of the ACM 28(7), 680-689.

[14] Katzenbach, J.R. & Smith, D.K., 1993, The wisdom of teams: Creating the high performance organizations, Harper Collins, New York.
[15] Mankin, D., Cohen, S. G., & Bikson, T. K., 2000, Teams and Information Technology, Harvard Business School Press, Boston, Massachusetts
[16] Manz, C.C. & Sims, H.P., 1987, Leading workers to lead themselves: The external Leadership of self-managing work teams, Administrative Science Quarterly, 32, pp. 106-128.
[17] Mary Beth Watson-Manheim & France Belanger, 2002, Exploring Communication-Based Work Processes in Virtual Work Environment, Proceeding of the 35th Annual Hawaii International Conference on System Science (HICSS-3502).
[18] Mohrman, S.A., Cohen, S.G. & Mohrman, Jr., A.M., 1995, Designing team-based organizations: New forms for knowledge work, San Francisco, CA: Jossey-Bass.
[19] Murray, R.B. & Stewart, G.L., 2000, Team structure and performance: Assessing the mediating role of intrateam process and the moderating role of task type, Academy of Management Journal, 2000, 43(2), 135-148.
[20] Neck, C., Connerly, M. Zuniga, C. & Goel, S., 1999, Family Therapy Meet Self Managing Teams: Explaining Self-Managing Team Performance Through Team Member Perceptions, The Journal of Applied Behavioral Science, 35, 2, pp. 245-259.
[21] Shoshana, Z., 1988, In the Age of the Smart Machine: The Future of Work and Power, Basic Books, New York.
[22] Sundstorm, E., DeMeuse, K.P., Futrell, D., 1990, Work Teams: Applications and Effectiveness, American Psychologist 45, no.2, 120-33.
[23] Tranfield, D. & Smith, S., 2002, Organization designs for teamworking, International Journal of Operations & Production Management, 22, 5, pp. 471-491.
[24] Wageman, R., 1995, Interdependence and group effectiveness, Administrative Science Quarterly, 40, pp. 145-180.

Business Process Integration and Automation in ERP System Environment – Integration of Applications and Workflows

Dr. Premaratne Samaranayake
University of Western Sydney
Sydney, Australia
Email: p.samaranayake@uws.edu.au

Dr. Felix T. S. Chan
The University of Hong Kong
Pokfulam Road, Hong Kong
Email: ftschan@hkucc.hku.hk

Abstract

Business process integration and automation have gained popularity among business community, in the recent past. This increased popularity can be seen as a result of the widespread use of Enterprise Resource Planning (ERP) systems as part of maximizing the benefits of Business Process Reengineering (BPR) projects. This paper presents a framework for integrating applications and workflows within an ERP system environment. The framework is based on enhanced Event-driven Process Chain (EPC) methodology, incorporating unitary structure-based models for business process applications and workflows. Main features of the proposed framework include (i) individual application and workflow models, incorporating many components, relationships and links; (ii) integrated model as a basis for common application/workflow transactions within ERP systems. The proposed framework is illustrated using individual models for sales order process and quotation approval workflow and integrated model of sales order with the quotation approval workflow. The proposed framework eliminates need for separate execution of business processes (normal transactions) and workflows involved in many process cycles within an ERP system. It also simplifies the implementation of both applications and workflows using a single implementation cycle. Further, the framework highlights the visibility, flexibility and connectivity of business processes and workflows as well as data elements and structures involved.

Keywords:

Business process, integration, automation, components, workflow, data structures

INTRODUCTION

For the last two decades, ERP systems and their effects on business process integration and automation have been the subject of considerable interests. A number of concepts/systems have been identified as very popular in the context of integration for business enterprise of the 21st century. Implementation of business process integration has resulted in a number of commercial software packages such as SAP R/3, PeopleSoft, Oracle, Baans for ERP. On the other hand, implementation of process automation is mainly confined to standalone Workflow Management Systems (WMS), using various tools and techniques, mainly driven by the information system approach.

Today's ERP systems have represented for large companies' one of the most relevant areas of investment in IT (Davenport, 1998). Despite these investments and widespread use of ERP systems, many companies are beginning to realise that the real impact of ERP systems on management styles and practices is actually well below expectations, especially on the front of organizational integration (Beretta, 2002). Need for better utilisation and performance of enterprise system resources has led to the focus on improving implementations of ERP systems.

It has been recognized that BPR plays a significant role in ERP system implementations (Sandoe *et al*. 2001; Sumner, 2005). Thus, success of ERP implementations and subsequent improvements on performance are dependent on how well benefits of BPR projects are achieved and maximized. This suggests that process integration within an ERP has a potential for bringing the maximum benefits of BPR (Sumner, 2005; Hammer, 1999). However, it is a challenge for many organizations to carry out BPR project effectively (Peppard and Rowland, 1995), even as a standalone project. With ERP systems, BPR has a dual role – one of result and the other as a prerequisite (Sandoe, *et al*., 2001). Moreover, BPR and ERP can be supportive of each other (Koch, 2001). In the case of mutual support, ERP systems can support BPR, through implementation of business process integration using applications and automation using business workflow.

Further, process automation can be achieved through workflow tools of many ERP systems although many ERP system implementations have overlooked workflows. Thus, process automation within ERP system environment can be seen as a solution not only to eliminate interfacing between systems but also to improve utilisation of system resources. Since many organizations run some kind of an ERP system for business process integration through application modules and do not use workflow tools for process automation, there is a potential for enhancing the utilization of existing ERP system resources for both applications and workflows.

Apart from benefits of individual applications and workflows, integration of applications and workflows within an ERP system environment can maximize the overall benefits of ERP systems through process automation as part of process integration. In this regard, Beretta (2002) identified the lack of organisational integration with the business processes in ERP while McAdam and McCormack (2001) identified the lack of research exploring the integration of business processes extending throughout supply chains. Further, researchers began to view the integration as one of the important issue in successful implementation of business processes in ERP (Beretta, 2002). Recently, Lau et al. (2003) proposed a system with features of ability to reconfigure and customize business workflows, allowing the integration of workflow in a flexible way within industry environment. Thus, need for an integrated system for both process integration and automation is inevitable if ultimate goals of BPR (process improvements through waste elimination to automation) are to achieve. This research aims to integrate those two aspects of BPR at structural level (a level higher than database level), using common structures based on unitary structuring technique.

This paper is organised as follows. An overview of business process and workflow modelling is presented, followed by the framework for business process applications and workflows. The framework is presented using (i) individual models of business process integration and automation with data elements and structures, (ii) integrated model for process applications and workflows using enhanced process and workflow models, based on part of sales order process and quotation approval workflow. Finally, this paper concludes with research findings and recommendations for future research directions.

OVERVIEW OF BUSINESS PROCESS AND WORKFLOW MODELING

Implementation of both business process integration and automation within an ERP system environment requires modeling of both aspects using standard methodologies. Business process modeling deals with process components and associated relationships and would result in process cycles usually packaged into application modules in ERP systems. Process integration, based on application modules within an ERP system, provides the basis for business workflow through process automation. Workflow modeling brings together many components and would result in workflow models, usually packaged into a workflow tool within ERP systems.

Business process modeling can be described as a configuration of process components (inputs, outputs, activities, resources, customers), relationships and links, so as to meet customer (internal and external) requirements in the best possible manner, in an organizational setting. Business process modeling leads to a number of process models, which form a process map, usually known as the business blueprint of the organisation. Event-driven Process Chain (EPC) methodology is widely used, as the common and standard methodology for business process modeling. Recently, Samaranayake (2003) proposed an integrated approach to process modeling at structural level, enhancing EPC methodology with added flexibility and functionality.

Many organizations develop their business blueprints as part of BPR or as a pre-requisite for ERP system implementation. Apart from business blueprints of individual organizations, ERP systems also have their own business blueprints (Curran and Ladd, 2000), which are usually claimed to be based on best business practices, using business logics, expertise knowledge and business theories. Currently, process models in business blueprint mainly focus on process orientation with loose integration of other orientations such as data, communication and information flows. For example, data inputs and outputs of a process in SAP R/3's blueprint are loosely listed against the process without direct relationships between components. Thus, EPC-based blueprints lack the integration at the structural level with a clear visibility of relationships between process and data elements.

Workflow modeling combines workflow components to represent triggering events, steps and relationships in business process automation. Further, workflow models enable implementation of process automation within an ERP system through associated links with business processes. Similar to BPR being a pre-requisite for ERP, business process integration is a pre-requisite for business process automation, and eventually business workflow. Thus, application integration within ERP systems becomes the foundation for process automation since applications within ERP systems guarantees the process integration. Further, process integration covers many cross-functional processes, where various activities/tasks are carried out across those functional boundaries. Within ERP system environment, current business workflow models are based on EPC methodology, using basic components such as events (triggering and terminating), steps (task, decision, etc.) and loops. However, these models lack structural integration of all the components involved, such as workflow agents usually involved in approvals as part of process automation.

INTEGRATED FRAMEWORK FOR BUSINESS PROCESSES AND WORKFLOWS

The integrated framework for business processes and workflows is based on enhanced EPC methodology (Samaranayake, 2003), using individual models, incorporating many components involved in both applications and workflows. These components include resource components and agents, involved in relevant areas (applications and workflows) within an ERP system environment. Thus, the integrated framework provides a basis for developing a set of transactions not only carrying out execution of process steps but also execution of relevant workflows within the process. For example, the function “quotation to be created from inquiry” within the standard sales order process can be automated through the business workflow “quotation approval”. This would eliminate separate execution of applications and business workflows within an ERP system environment.

The integrated framework can be used to represent process cycles in terms of process components, data elements and structures involved, and relationships between components. Relationships are represented by links between components with appropriate precedence (parent-component, component-component and activity precedence). Moreover, the proposed framework provides the flexibility and maintainability of many process cycles, and can be used to simulate existing processes to achieve better planning and execution schedules for all the components involved.

Workflow models are integrated with business process models, using triggering and terminating events of workflows. Triggering events are used to branch out workflows from process models while terminating events are used to branch in after completion of the workflow within the business process. Thus, proposed framework for integrating process models with workflow models is developed in three steps:

- Business process integration with data elements/structures and organizational elements
- Business workflow integration with data and organizational elements
- Integration of process and workflow models using individual models and links.

Business Process Integration with Data Elements/Structures and Organisational Elements

Standard sales order process within SAP R/3’s business blueprint was chosen for the purpose of illustrating process and workflow modeling as part of integrated framework development. The sales order process, centered in sales and distribution module of SAP R/3, is a cross-functional process and involves many events, functions/tasks, information and organizational elements. In addition to basic components of the process, it also involves a number of data elements and structures, usually modeled separately using data models based on entity-relationship diagrams (ERD). It is clear from the EPC-based process model (Sandoe, et al., 1998) that components in the process view do not include data elements and/or structures. However, the process model within an ERP system links with associated data model(s) at the database level. This means that the business process model lacks representation of data elements and structures at the time of business process modeling. On the other hand, the same process can be modeled using enhanced EPC methodology (Samaranayake, 2003), eliminating separate models for process and data components, providing an integrated model for process components as well as data elements and structures. Figure 1 shows the integrated process model for part of the sales order process, incorporating basic process components and data elements based on enhanced EPC methodology.

Figure 1 also shows that functions within the process can be expanded with a set of activities, precedence between activities and set of resources. Events in the process do remain unchanged while functions are replaced by a number of process and data elements. In the terminology of this approach, the outline of the component icons appear as “M”, “A”, “R” and “S” which represent Material, Activity, Resource and Supplier respectively. Thus, A1-A5 represents pre-sales activities. Components attached to those activities are labour and machine resources (L1-L3, MC1). Material M1 supplied by supplier S1 could be a material and/or a set of materials required for pre-sales activities. This suggests that data elements and structures could be attached to the business process at this level of business process modeling and eliminates the need for ambiguous situation of data relationships with business process components. It also provides a foundation for simultaneous planning of all process components. Simultaneous planning of components can be carried out using scheduling paths, designed specially for these models and have been discussed earlier in detail (Samaranayake, 2003).

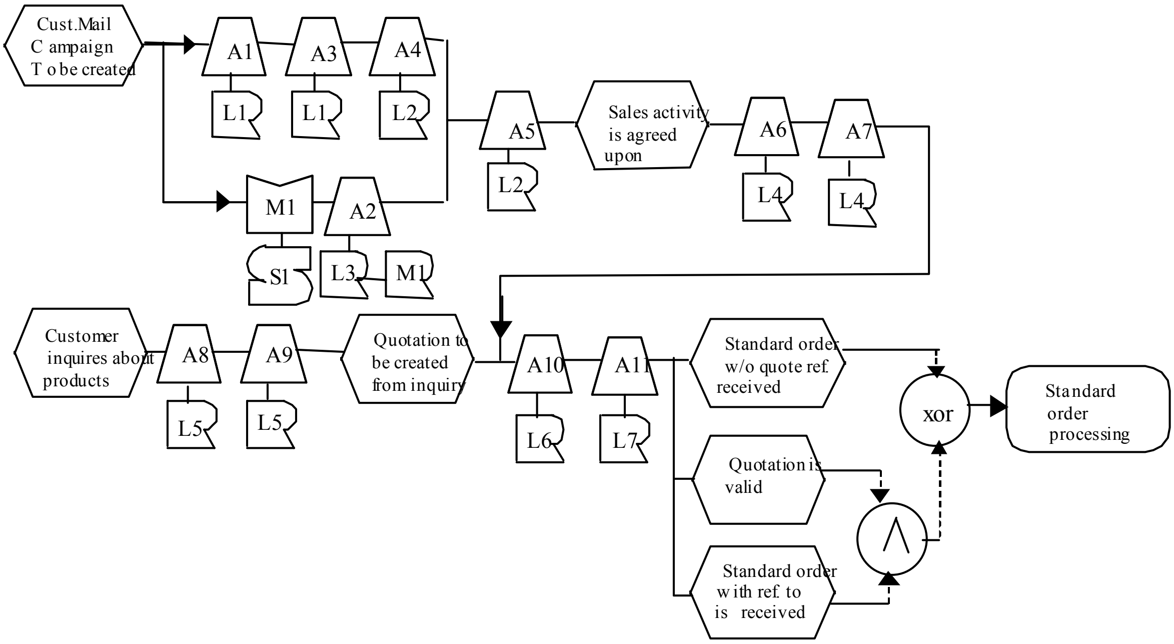


Figure 1: Part of Standard Order Process based on Enhanced EPC Methodology

Business Workflow Integration with Data and Organisational Elements

Business process automation, similar to process integration, involves a number of components such as events, agents, steps, and loops for decisions, and relationships. Modeling of process automation enables implementation of business workflows within system environments. Although agents are components of the process automation, many business workflows do not show them as components rather they are interfaced dynamically at the time of workflow execution. Similar to business process modeling outlined above, business process automation can also be modeled using workflow components with data elements and structures.

For the purpose of this research, quotation approval workflow within the sale order process is considered and extended with data and organizational elements. The workflow for approval of quotation can be considered as an automated process of approval or rejection, based on notification of quotation. Thus, the workflow model for quotation approval involves the triggering event of notification of quotation and few steps and loops before it terminates. The workflow starts immediately after the start of the triggering event of the workflow. In general, a workflow starts with initial values and involves interface specifications such as workflow and event containers. In a business workflow, binding identifies the relevant data required for the execution of the workflow. It provides the connection between the data required as input for executing a task, and the output data that has to be passed on to subsequent steps in the workflow as a result of the task. Based on the traditional workflow definition and enhanced EPC methodology for business process modeling, the notification of quotation approval business workflow is modeled as shown in Figure 2.

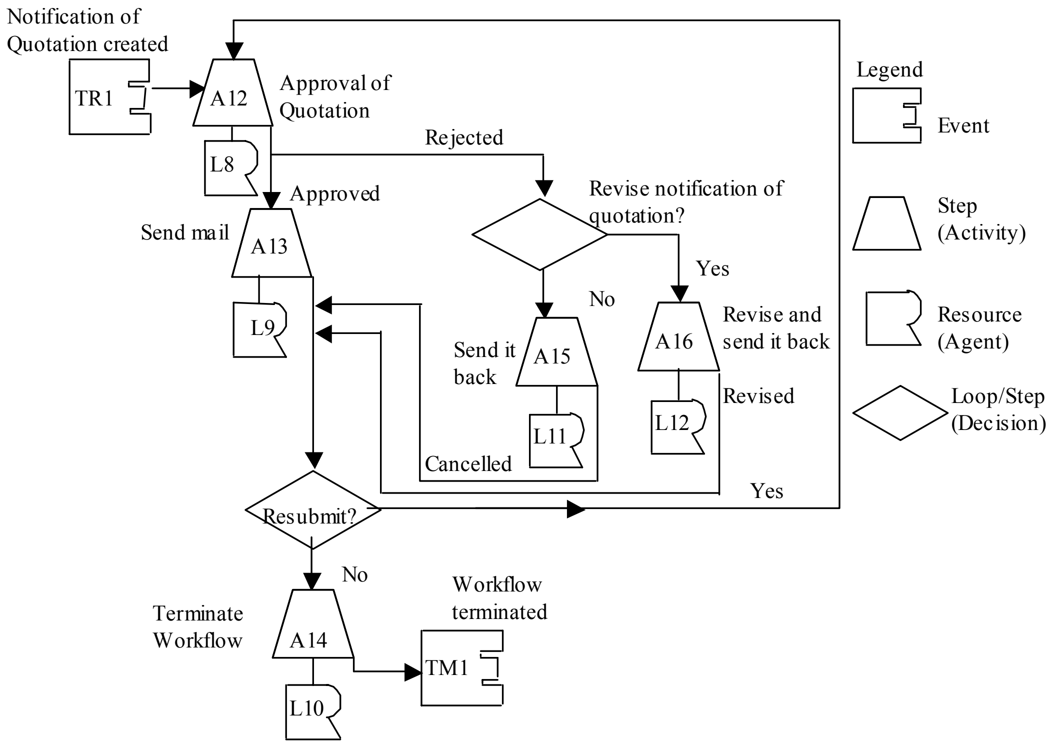


Figure 2: Enhanced Business Workflow Model for Quotation Approval Process

The workflow model above (Figure 2) represents an automation of a business process (notification of quotation approval application) using user decisions and events. This workflow includes triggering event (TR1), terminating event (TM1), a number of steps (activities A12-A16) and a number of resources (labour resource components L8-L12) attached to those steps. The workflow model is similar to business process model except that business workflow takes all aspects of process models (process integration) and adds process automation components. It also incorporates not only business workflow elements (events, loops and steps) but also agents as data elements.

Therefore, enhanced workflow models integrate many types of components including resource components at the structural level, and eliminate a need for interfacing workflows with data at the time of workflow execution. For example, an agent (resource component) associated with a decision point in the workflow is not a component in traditional workflow models at structural level. Further, these models clearly indicate branching in and out of process models. Thus, workflow models enhance the visibility of all involved and integrate all elements including data and process components. The main feature of the enhanced model is that incorporation of material and resource elements into the workflow rather than having them as background elements. While the enhanced methodology for workflow provides integration of both workflow and data elements at the structural level, workflows are still executed separately from applications. Thus, workflows need to be integrated with applications for common execution of applications and workflows.

Integration of Process and Workflow Models using Individual Models and Links

Improvements on process integration and automation, through enhanced modeling and subsequent features such as simultaneous planning of many components and elimination of interfacing, can be further extended using integration of individual models of applications and business workflows. As the first step of such integration, individual models along with links between two models are considered. The very important links between business process and workflow models are the branching out position on the process for business workflow and the branching in after the execution of the business workflow.

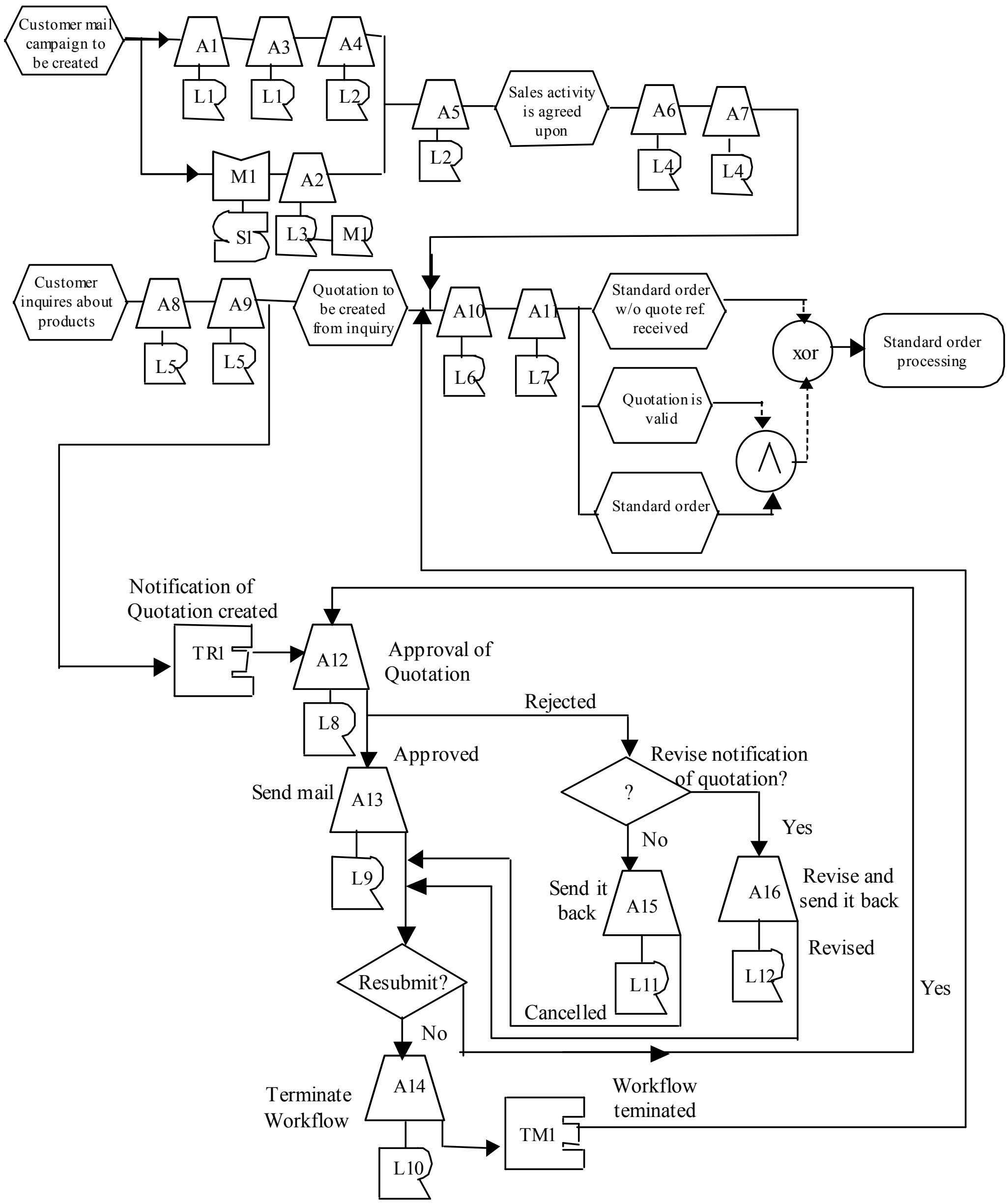


Figure 3: Integrated Model for Part of Sales Order Process and Quotation Approval Workflow

Based on the enhanced sale order process model (Figure 1), part of the process with events and functions is identified as an eligible candidate for business process automation using workflow in an ERP system environment. The main selection criterion is that the chosen process has a triggering event, terminating event and few steps. Careful analysis of the sales order process shows that the event “quotation to be created” can be treated as a triggering event for the business workflow. This means that this event may trigger an approval and/or rejection process. It could also involve various levels of approvals depending on the quantity and other information relevant to the quotation. Steps involved in automation (workflow) would be things like sending mail, revision of the quotation, approving the quotation. Once the quotation is approved or rejected, workflow could be terminated and the original process can proceed. Thus, the resulting model based on both models as well as associated links is shown in Figure 3.

It can be seen from Figure 3 that the quotation approval workflow is incorporated into the sales order process. Therefore, the resulting model is a combination of both application and workflow models within overall sales and distribution module of SAP R/3. This would provide a foundation for enhanced functionality for both applications and workflows within ERP system environments. The integration of both process and workflow models can be carried out for many functional areas within ERP, provided links between process and workflow models are identified. The workflow usually needs a triggering event to start the workflow. In this case, the triggering event selected is "quotation to be created" based on the information from enquiry. This is the point it branches out of business process and branches into the business workflow. Upon realizing of the terminating event in the business workflow, the information is sent back to business process for the reminder of the process. In order to gain maximum benefits of this integration, the integration needs to be implemented within an ERP system. This requires re-thinking of how business blueprint of an ERP system can be enhanced with all the common workflows. This means that each process cycle within ERP needs to be re-designed with relevant business workflows. Once all process cycles are redesigned, they can become the complete blueprint of the system. Based on such an enhanced business blueprint, transactions need to be extended to cover both applications and workflows within those applications.

CONCLUSIONS AND FUTURE WORK

It is evident from the literature that process integration and automation are key elements of BPR and have been in operational as two separate areas (application modules and workflow tools) within ERP systems. Thus, execution of applications and workflows have been separate and in some cases workflows have been overlooked, due to the main focus of ERP systems being process integration rather process automation. This research recognized the need for integration of both applications and workflows within ERP systems at structural levels for common transactions for business process integration and automation. This paper presented an integrated approach for both business process and workflow modeling with data structures in ERP. The approach is based on the unitary structuring technique, which has been successfully implemented in many applications such as project-based manufacturing, maintenance planning and scheduling and supply chain management (Samaranayake, 2005). It has been shown that the proposed framework integrates business process and workflow components with data elements and structures such as materials, resources, and activities. Main features of the proposed framework are (i) both business process and workflow modeling, based on enhanced EPC, incorporate many components including data elements and structures, relationships and links; (ii) integration of both business process and workflow models for extended application transactions within ERP systems. Proposed framework has been illustrated using a standard business process cycle in ERP system. Further, the resulting structure eliminates need for separate transactions for applications and business workflows within an ERP system, if both are integrated at the structural level. It also simplifies the implementation of both applications and workflows using a single implementation cycle. Further, the framework highlights the visibility, flexibility and connectivity of business processes and workflows as well as data elements and structures involved.

REFERENCES

Beretta, S. (2002), "Unleashing the Integration Potential of ERP Systems: The Role of Process-Based Performance Measurement Systems", *Business Process Management Journal*, Vol. 8, No. 3, pp. 254-277.

Curran, T. A. and Ladd, A. (2000), *SAP R/3 Business Blueprint Understanding Enterprise Supply Chain Management,* 2nd Edition, Prentice Hall, New Jersey, USA.

Davenport, T. H. (1998), "Putting the enterprise into the enterprise system", Harvard Business Review, July-August.

Hammer, M., "Re-engineering work: don't automate – obliterate", *Harvard Business Review*, July-August (1990), 104-112.

Koch, C., (2001), "BPR and ERP: realizing a vision of process with IT", *Business Process Management Journal*, Vol. 7, No. 3, pp. 258-265.

Lau, H.C.W, Lee, W.B., and Lau, P.K.H., (2003), "Flexible workflow integration: an object technology approach", *Industrial Management & Data Systems*, Vol. 103, No. 3, pp. 167-176.

McAdam, R. and McCormack, D. (2001), "Integrating Business Processes for Global Alignment and Supply Chain Management", *Business Process Management Journal*, Vol. 7, No. 2, pp. 113-130.

Mentzas, G., Halaris, C., and Kavadias, S., (2001), Modelling business processes with workflow systems: an evaluation of alternative approaches, *International Journal of Information Management*, Vol. 21, pp. 123-135.

Olson, D.L., (2004), *Managerial Issues of Enterprise Resource Planning Systems*, McGraw Hill, USA.

Peppard, J. and Rowland, P., (1995), *The Essence of Business Process Re-engineering*, Prentice Hall, UK.

Samaranayake, P., 2003, "Business Process Integration with Data Structures in Enterprise Resource Planning Environment", *British Academy of Management Annual Conference*, September 2003, Harrogate, UK.

Samaranayake, P. (2005), "A Conceptual Framework for Supply Chain Management: A Structural Integration", *Supply Chain Management: An International Journal*, Vol. 10, No. 1, pp. 47-59.
Sandoe,K, Corbitt, G., and Boykin, R., (2001), *Enterprise Integration*, John Wiley & Sons, Inc., USA.

Sumner, M, (2005), *Enterprise Resource Planning*, Pearson Prentice Hall, New Jersey, USA.

Reverse Logistics Processes: An Australian Study

Dushantha Dissanayake, Mohini Singh
School of Business Information Technology
RMIT University, PO Box 2476V
Melbourne, Victoria
Australia
dushantha.dissanayake@rmit.edu.au
mohini.singh@rmit.edu.au

Abstract

This paper presents a discussion on reverse logistics, a conceptual model for reverse logistics management, including recovery processes and how they add value to business. The processes and the flow of products included in the conceptual model are reverse logistics issues identified from a number of earlier models and literature. It also includes some preliminary findings of a research project with the Australian original machine and equipment manufacturers highlighting some reverse logistics trends in the Australian manufacturing industries. An analysis of the data determining the validity of the conceptual model in the Australian context is also discussed. It highlights how reverse logistics information is managed.

Key Words:
Reverse logistics, recovery processes, conceptual model for reverse logistics

INTRODUCTION

Reverse logistics, in spite of being an important process in manufacturing, is not widely deployed. It enables recapturing value, as well as proper disposal of returned goods. Reverse logistics also enhances customer relationship management, strategic marketing, and environmental protection supporting sustainable development of manufacturing organisations. Reverse logistics entails managing returned merchandise due to damage, salvage, recalls, seasonal excess and restocking of inventory. It also encompasses returned product management programs such as recycling, obsolete equipment disposition, hazardous material management and asset recovery. Dekker and Brito (2002) have categorised returns into commercial returns, product recalls, warranty returns, manufacturing returns, service returns and end-of-use and end-of-life returns. Direct and indirect benefits of reverse logistics according to Kokkinaki, Dekker, Lee and Pappis (2001) are regaining value, achieving a competitive advantage and a positive environmental impact. Similarly Roy (2003) and Bayles (2001) highlighted opportunities with reverse logistics bringing economic gains by recapturing value from raw materials and used components from returned products and rejects.

Therefore it is vital to manage reverse logistics efficiently and effectively to achieve benefits and capture value from returned goods. This paper discusses some preliminary findings of reverse logistics in Australia. It highlights how reverse logistics processes are managed in the Australian manufacturing organisations, who manages reverse logistics, the level of recovery processes performed and some discussion on technologies that support reverse logistics in these organisations. This is a research in progress paper.

LITERATURE REVIEW

Morrell (2001) is of the opinion that reverse logistics is "the forgotten child of the supply chain" since many organisations do not treat reverse logistics with the same care as traditional (forward) logistics. However, reverse logistics is just as important as forward logistics addressing management of returned products. Dekker and Brito (2002) categorised the returns into a number of groups based on the reasons for returning such as manufacturing returns, commercial returns, product recalls, warranty returns, service returns, end-of-use returns and end-of-life returns. Reverse logistics entail planning, implementing, controlling and cost effectively managing the flow of raw materials, inventory in process, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal (Rogers and Tibben-Lembke, 1999). Fernandez (2003) emphasized that reverse logistics can be applied to different types of items such as used products, unused products, components, parts and raw materials that are incorporated in both upstream and downstream production operations. Reverse logistics cost in the United States represents approximately 4 percent of total logistics costs (Delaney, 2001).

A number of opportunities arise from reverse logistics management. Andel (1997) emphasised profit-making opportunities from reverse logistics while Roy (2003) is of the opinion that significant economic gains from raw materials, used components, used products and rejected goods can be recaptured. On the other hand Bayles (2001) explicated reuse of discarded products from commercial returns and excess inventory of products as opportunities for reverse logistics. Reverse logistics management also includes recycling, hazardous material management, obsolete equipment disposal and asset recovery programs. Reverse logistics is still being identified as a new concept with unrealized business value. According to Roy (2003) reverse logistics is an important management issue due to significant economic gains. Kokkinaki, Dekker, Lee and Pappis (2001) are also of the opinion that direct and indirect benefits of reverse logistic include regaining value, achieving a competitive advantage and a positive impact on the environment encouraging organisations to manage it effectively.

Reasons for Reverse Logistics Management

Important reasons for managing reverse logistics identified from literature are discussed below.

- *Electronic Business*

Electronic business is sale of goods via electronic channels. As Internet customers purchase goods by virtual conceptualisation of size and quality under more flexible return policy, a high volume of returns results, forcing organisations to have an efficient reverse logistics system (Trebilock, 2002). Smith, Bailey and Brynjolfsson, (2000) and Boyer, Hallowell and Roth (2002) are of the opinion that invisibility and physical distance between customer and seller, lack of face-to-face interaction with customers in electronic businesses and an inefficient reverse logistics management could destroy customer trust in electronic businesses. Sarkis et al (2004) also emphasised that growth of electronic business has precipitated the need for efficient and effective reverse logistics management. According to Bizrate.com (1999) report 94% of online buyers commit to purchase products if a product return policy is in place. The estimated average percentage of customer returns in Internet sales is 30 and sometimes up to 50, (Nairn, 2003; Mason, 2002; Sharma, Wickramasinghe and Singh 2005, Rogers and Tibben-Lemke, 1999). The business to customer (B2C) Internet sales is predicted to be at the value of US$ 210 Billion in the Asia Pacific region by 2006. The business-to-business (B2B) electronic business is estimated to be US$ 6 trillions by 2005 by Jupiter Communication, however, in B2B e-business products come back in bulk (Brito, Flapper and Dekker, 2002).

Other reasons for increased volume of e-business returns identified by Lee (2002) are errors in ordering, picking or shipping, damages due to transportation handling or cancellation of orders. Haris and Goodman (2001) highlighted that growing frustrations among e-business customers are due to unfulfilled promises, unrealistic response time and untrained workers leading to higher returns. Pan, Ratchford and Shankar (2001) divided online sales processes into four equally important areas, including pre-sales service, transactions, order fulfilment and after sales services. Order fulfilment and sales services require effective reverse logistics management.

- *Strategic Marketing and Liberal Return Policies*

Strategic marketing and liberal return policies have lead to the growing need for efficient reverse logistics management (Krumwiede and Sheu, 2002). These policies are important in direct store, catalogue, telephone, television and electronic business customers. They need some control over the product they purchase, and also allow them to return items for a variety of reasons. Customer protection legislation allows customers to return products if they are not satisfied (Brito and Dekker, 2003 and Krumwiede and Sheu, 2002). Accordingly, customers rightfully and sometimes wrongfully take advantage of this opportunity. The recorded return rates in a retail store are 10 percent (Dekker and Van der Laan, 2002), catalogs, telemarketing and the television models returns are 35 percent (Trebilcock 2002) and Internet sales returns are 30 to 50 percent (Nairn 2003; Mason 2002; Sharma et al 2005). Customer protection regulation and competition among vendors imposes more relaxed return polices, resulting in unlimited returns (Rogers and Tibben-Lembke, 1998).

- *Commercial and Warranty Returns*

Commercial returns include products that are returned without fulfilling some or all of preceding business transactions and therefore buyers return products to the original sender with demand for refund (Lee, 2002). This type of return starts from two destinations, the retailer returns products to manufacturers and customers return products to manufacturer via a retailer. Brito et al (2002) stated that, financial risk transferability from the buyer to the seller and roles and position of different parties in the supply chain, are key factors that control volume and quality of commercial returns. Further they described that warranty returns include products that failed during use, were damaged during delivery from manufacturer to buyers, product mismatches, safety concerns and manufacturer recalls.

- *End of Life and End of Use Returns*

End of use or end of scheduled life returns also warrant reverse logistics management. End of use occurs once a product's scheduled life is complete for example, end of lease or contract and end of life of a product (Brito et al, 2002). For end-of-use returns, the challenge is determining the match between demand for used products and supply, impacted by time, quantity and quality, which could be managed by information (Sarkis et al, 2004).

- *Environmental Protection and Legal Regulations*

Environmental protection, competitive advantage and regaining value are also reasons for a growing interest in reverse logistics (Brito et al, 2002). Legal regulations in returns management are becoming strict in countries like Germany, Netherlands, Japan, USA and in Europe due to negative effects of wasting energy and resources, pollution of air, water and environment (Fernandez, 2003). These governments have imposed disposal tariffs and bans, placed restrictions on waste transportation, waste prevention and emission which have increased producers' responsibility to manage returns effectively (Krikke, Pappis, Tsoulfas and Bloemhof-Ruwaard, 2001).

One of the future challenges of solid waste management is managing the disposal of upcoming out-dated computers. The number of personal computer returns is predicted to be 500 million units by 2007 in the U.S.A. As electronic products are the source of many toxic substances such as lead, mercury, cadmium, chromium and bromine they can have an impact on he environment (Kokkinaki et al, 2001; Europa, 2000). Accordingly, disposing of returned personal computers has become a good example of a product for which reverse logistics is required for environmental protection as well as recapturing value or assets.

- *Economic Gains*

Economic gains from reuse of products, parts or recycling material are a growing interest in managing of reverse logistics in recent years (Brito et al 2002). For example over 70,000 re-manufacturing firms in the U.S.A. typically for jet and car engines, auto parts and copiers are relying on reverse logistics for reuse of products, modules, parts, components or recyclable raw materials (Lund 1998; Roy 2003). As a result, the manufacturers are able to provide quicker, inexpensive, better customer service using the recovered material, parts, components and modules (Roy 2003). Similarly it is important to have effective reverse logistics system, because of increased emphasis on new products and product "freshness". On the other hand reverse logistics helps to clear the distribution channel. It is an efficient means of bringing back obsolete, outdated and clearance items. For example Xerox replace or upgrade hundreds of office printing machines every month, because it is able to provide state of the art technology while increasing customer retention rates (Roy 2003).

- *Customer Loyalty and Retention Rate*

Jiang et al (2005) and Sharma et al (2005) identified a number of potential benefits of strategic reverse logistics management such as increased customer retention rates and increased customer loyalty. Retention of online customers is highly beneficial compared to other channels. As cited in Jiang et al (2005), Boston Consulting Group estimate that acquiring a new customer through an online channel costs US$ 82 compared to US$ 38 for store based sale and US$ 11 for catalogue-based sales. Further, they emphasised that excellence in delivery of services in forward logistics and reverse logistics in an electronic business could create a powerful competitive differentiation.

BARRIERS OF REVERSE LOGISTICS

Although reverse logistics is an important business process, it is faced with many challenges and barriers. Important barriers identified from literature are discussed in the following section of this paper.

- Lack of Attention

Reverse logistics has not been given the same importance as forward supply chain due to a lack of understanding about return management benefits by organisations (Lee, 2002). Therefore companies have not allocated the necessary resources for the management of reverse logistics. Tibben-Lembke (2002) is of the opinion that many highly touted electronic businesses have suffered due to an under estimation of difficulties and physical and psychological values of reverse logistics. This view is supported by Spiegel (2000) describing that most electronic businesses lose 70% of their time due to a lack of technologically advanced and integrated logistics operations including reverse logistics. In addition Rogers et al (1998) identified the unimportance of reverse logistics compared relative to other business operations.

- *Lack of Dedicated Technology for Reverse Logistics Management*

According to Caldwell (1999) commercial software specifically designed for supporting reverse logistics is not available. Nagel and Meyer (1999) are of the opinion that a lack of information management causes bottlenecks, which brings difficulties for the management of recycling systems leading to substantial loss to companies. Further, due to a lack of technology-supported systems incorporated in reverse logistic processes, longer times and higher costs of operations, invisibility of retuned products handled by third parties, repeat operations leading to additional cost to manufacturer and unawareness and inattention of process partners result (Zhao, 2001). As a result of unavailability of sophisticated systems to identify validity of returned products at the point of entering leads to additional cost due to illegal returns, for example, "non-defective returns" can actually account for 55% or more of the total returns (Lee, 2002).

- *Poor Relationship Management with Intermediaries*

It is not always easy to predict the exact amount of material in a certain collection centre (Carella et al, 2002). Therefore companies sometimes use third parties for reverse logistics operations. However Krumwiede and Sheu (2002) emphasise that a lack of guidelines and collaborative information sharing also affects the relationship with intermediaries for reverse logistics management. Accordingly, returns are invisible, not recorded in the organisations information systems, and not considered for next business operations until they physically arrive in store without warning. This makes it difficult for operations managers to schedule re-work on returned goods. Manufacturers also have to bear costs of holding goods and experience lost opportunity from these goods (Zhao 2001).

- *Meeting Recovery Deadline*

Reverse logistics management is essential for companies that produce make-to-order contracts or repair large industrial machinery, such as conditioners, heating plants, trains, ships, airplanes or components for their clients. Without a better reverse logistics management model the manufacturers are unable to respond to their customers in a timely manner and help them avoid shortages or delays. Failing to do so leads to huge penalty costs to the company (Carella et al, 2002) as well as lost business opportunities.

CONCEPTUAL MODEL FOR REVERSE LOGISTICS MANAGEMENT

Due to the importance of reverse logistic in capturing business efficiencies and value, it is important to understand the preliminary and recovery processes of reverse logistics. It was noted that although different aspects of reverse logistics have been considered by other researchers discussed above, a logical and comprehensive presentation of the processes showing the logistics of effectively managing returned goods has not been developed.

An analysis of reverse logistics processes identified from the works of (Kokkinaki, Dekker, Nunen and Pappis (1999:2000), Thierry et al (1995), Fleischmann et al (2002), Lee (2002), Krikke et al (2001), Roger et al (1998), Brito et al (2003), De Koster, de Brito and van de Vendel (2002), Lonn and Stuart (2003), Meade and Sarkis (2002), Kokkinaki, Dekker, Koster and Pappis (2001) and Kokkinaki, Dekker, Lee and Pappis (2001)) is presented in Figure 1. This conceptual model of reverse logistics provides a comprehensive illustration of reverse logistics management process and operations.

Figure 1 was derived from a thorough analysis of reverse logistics issues from the works of authors presented in column one in tables 1 and 2. Table 1 presents product recovery processes, and Table 2 includes all other recovery processes that can be performed on returned goods.

Table 1: Product Recoveries of the Models

Reverse Logistics Model	Redistribution	Resale	Reuse	Donate	Upgrade	Repair
Kokkinaki, Dekker, Nunen and Pappis (1999:2000)	X		X			
Thierry et al (1995)		X	X			X
Fleischmann et al (2002)	X		X			X
Lee (2002)		X	X			
Krikke et al (2001)	X	X	X		X	
Roger et al (1998)	X	X	X		X	
Brito et al (2003)	X		X			X
De Koster, de Brito and van de Vendel (2002)						X
Lonn and Stuart (2003)	X	X				X
Meade and Sarkis (2002)			X			
Kokkinaki, Dekker, Koster and Pappis (2001)	X	X	X			
Kokkinaki, Dekker, Lee and Pappis (2001)	X	X	X			

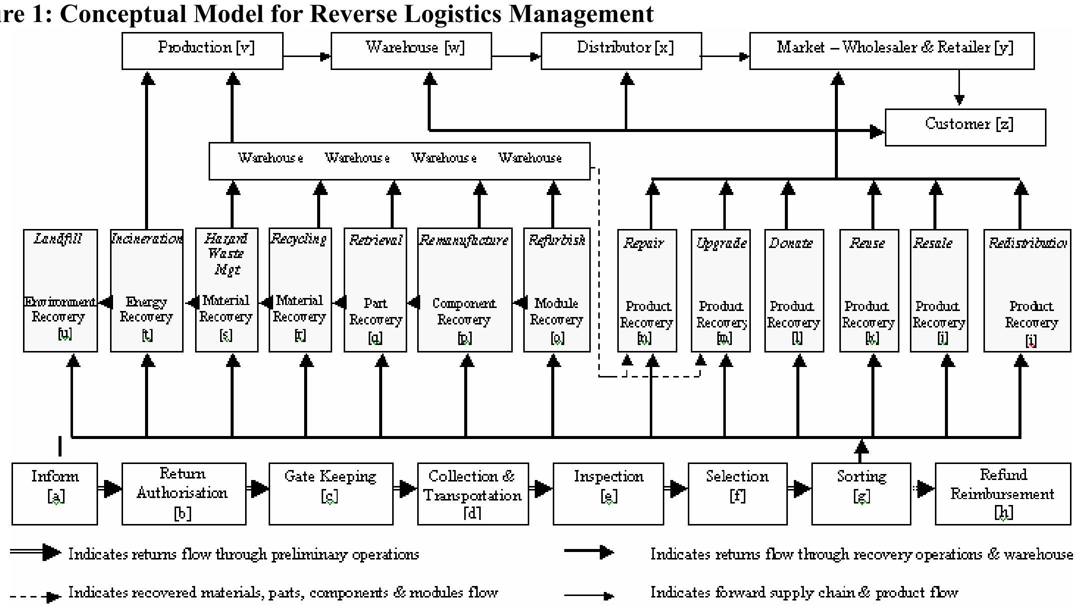
Table 2: Module, Component, Part, Material, Energy and Environment Recoveries of the Models

Reverse Logistics Model	Refurbish	Remanufacture	Retrieval	Recycling	Hazard Waste Mgt	Incinerate	Landfill
Kokkinaki, Dekker, Nunen and		X		X			

Pappis (1999:2000)							
Thierry et al (1995)	X	X	X	X		X	X
Fleischmann et al (2002)	X			X			X
Lee (2002)	X		X				
Krikke et al (2001)	X	X		X			X
Roger et al (1998)	X	X		X			X
Brito et al (2003)	X	X	X	X			
De Koster, de Brito and van de Vendel (2002)	X	X	X	X			X
Lonn and Stuart (2003)							X
Meade and Sarkis (2002)		X		X			X
Kokkinaki, Dekker, Koster and Pappis (2001)		X		X			X
Kokkinaki, Dekker, Lee and Pappis (2001)		X		X			X

Based on analysis of the above models, important reverse logistics issues were identified, that have been included in Figure 1. Other processes and flow of goods included in Figure 1 were identified from literature (Roy 2003 and Amini and Retzlaff-Roberts 2003).

Figure 1: Conceptual Model for Reverse Logistics Management



Recovery processes included in figure 1 are redistribution, resale, reuse, upgrade, repair, refurbish, remanufacture, retrieval, recycling, incineration and landfill processes. Other processes are donation and hazard waste management recovery.

This conceptual model (Figure 1) includes three main sections:

- *preliminary reverse logistics operations*
- *reverse logistics recovery operations and,*
- *forward logistics operations.*

- *Preliminary reverse logistics operations;*

This represents all the preliminary operations, which take place from the time a customer or a dealer decides to return a product or products till the same product(s) are to be reprocessed through a recovery processes. Eight preliminary recovery processes were identified from the literature.

- inform [a], return authorisation [b], gate keeping [c], collection and transportation [d], inspection [e], selection [f], sorting [g] and, refund/reimbursement [h].

- *Reverse logistics recovery operations;*

Methods of returned products reprocessed are considered as recovery processes. Thirteen recovery processes were identified from the literature review based on the level of changes made on and methods of handling of the returned product(s).

- redistribution [i], resale [j], reuse [k], donate [l], upgrade [m], repair [n], refurbishing (module level recovery) [o], remanufacturing (component level recovery) [p], retrieval (selective part level recovery) [q], recycling (material level recovery) [r], hazard waste management (material level recovery) [s], incineration (energy level recovery) [t] and, landfill (environment recovery) [u].

- *Forward logistics section*

This is the traditional supply chain through which finished products are passed from manufacturer to final customer.

- production [v], warehouse [w], distributor [x], wholesaler and retailer [y] and, customer [z].

Arrows in Figure 1 show the flow of returned goods. Returns commence at point [a] and flow to [g]. From [g] the returned goods flow to points [i], [j], [k], [l], [m], [n], [o], [p], [q], [r], [s], [t] and, [u] according to recovery decisions made in the first eight preliminary operations [a, b, c, d, e, f, g and h]. From product recovery operations [i], [j], [k] [l], [m] and, [n] goods flow to [w] [x] [y] and [z]. Outcomes of module, component, part, material and hazard waste recovery operations flow to [v], [n] and [m] via the warehouse. This indicates that any part recovered from reverse logistics is either used for the production of other goods or used to repair or upgrade products. Some goods may be selected for [t] and [u].

The above ideas have been put together in figure 1 to show the processes their sequences in reverse logistics management. As shown in the proposed model (figure 1), reverse logistics begins at point [a]. Any partner of the forward supply chain [v, w, x, y and z] can take the decision to return goods due to commercial and warranty issues, end life and end of use issues, environmental issues, sustainable development issues and strategic marketing and retailing policies. The partner or the customer then informs the manufacturer or its agent of the decision.

All of the above discussed recovery operations depend on product life and the level of depreciation. Brito and Dekker (2003) identified three kinds of product depreciation; intrinsic, homogeneity and economic. Intrinsic deterioration is age of the product. Homogeneity deterioration is age of a product's parts and economic deterioration is value of depreciation.

As suggested by Croxton et al, (2001) managing all the above recoveries is a complex task that consists of a number of sub processes. These sub processes can be divided into two groups such as strategic processes and operational processes. Strategic processes address environment and legal, gate keeping, disposition guidelines, mange return flow, credit handling and secondary markets. Operational sub processes manage returns on ground such as preparing to accept returns, receiving returns, recovery decisions, customer and supplier credit management and analyzing returns. Further according to them these two groups of processes require an interface to cover customer service and relationship management, demand management, order fulfillment, manufacturing management, supplier relationship management and product development management. All this processes and their relationships emphasize the importunacy of having very effective and efficient information management system.

To investigate the reverse logistics processes in the Australian original machine and equipment manufacturing organisations a postal questionnaire survey was undertaken with 310 companies.

(6) RESEARCH APPROACH AND METHODOLOGY

The research on Australian original machine and equipment manufacturing organisations was accomplished via a postal questionnaire survey. The survey consisted of carefully structured questions to elicit reliable responses. Survey research is a positivistic methodology whereby a statistical sample is drawn from the population of subjects to make deductive inferences about the population as described by Collis et al (2003). The survey study was planned to identify and investigate the type of recovery processes used by the Australian original machine and equipment manufacturing organisations. According to Cavana et al (2001) use of a postal questionnaire survey is more appropriate to gather large amounts of data statistically sampled from a survey population, which represent a large geographical area. Therefore the postal questionnaire survey was the most suitable method for this research, which investigated organisations in all cities and states of Australia. In addition some other considerations such as budget constraints, time constraints, geographical limitations, and width of the sample frame, were also taken into account when selecting this method.

The postal questionnaire was pre-tested with thirty-eight researchers at RMIT University. Based on suggestions of nineteen responses, the questionnaire was modified to avoid ambiguities and to ensure understanding.

Australian original machine and equipment manufacturing organisations were selected as the sample population for this research based on the findings of the previous postal survey study conducted among Australian manufacturing organisations. Rogers and Tibben-Lembke (1998) are also of the opinion that electrical and electronic products manufacturers receive returns at a rate of 8 to 20% of sales. The questionnaire survey collated data from a randomly selected sample of 310 Australian original machine and equipment manufacturers with over AUS$ 1 million annual income. The manufacturers were randomly chosen from two national business directories (The Business Who's Who of Australia, May 2005, http://bww.dnb.com.au/ and IBIS World of Australia, May 2005, http://www.ibisworld.com.au) for the survey. A sample of 310 Australian original machine and equipment manufacturing organisations was determined from the views of Cavana et al (2001), and Sekeran (2003) suggesting that sample sizes larger than 30 and less than 500 are appropriate for most research.

A total of 51 usable responses were received. A response rate of 16.45% was contemplated to be a reasonable accomplishment as described by Collis and Hussey (2003) that 10% or less response rate of postal questionnaires was common.

(7) DATA ANALYSIS

The data elicited through the survey was collated from Likert Scales. Data analysis included in this paper was tabulated to show some preliminary findings. Tabular methods of presentation of the data gathered from the postal questionnaire were summarised and analysed using SPSS, a statistical analysis software package.

Findings and Discussions

Findings of this research highlight commonly practiced preliminary and recovery processes of reverse logistics management, how these processes are managed, amount of returned products reprocessed under each recovery method and technologies used for managing reverse logistics information.

In Table 3, below we have presented product categories and level of reverse logistics currently performed in these organisations.

Table 3 Product Categories and Reverse Logistics Performance

Product Categories	Responses (Percentage)	Organisations Performing Reverse Logistics	
		Yes (Percentage)	No (Percentage)
Electronic and Electrical Equipment and Parts	43.14	77.27	22.73
Automotives and Parts	25.49	76.92	23.08
Mining and Irrigation Equipment	11.76	83.33	16.67
Construction Equipment	11.76	50.00	50.00
Medical and Scientific Equipment	7.84	75.00	25.00
Overall		**74.51**	**25.49**

In Table 3 above, column 2 indicates the number of respondents from the industries listed in column 1, column 3 of Table 3 indicates that in Australia reverse logistics is performed in electrical and electronic equipment and parts, automotive parts, mining and irrigation equipment, and medical and scientific equipment manufacturing industries.

Recovery Processes

Preliminary and recovery processes of reverse logistics, and how theses processes are performed are presented in Table 4.

Table 4 Preliminary Recovery Processes of Reverse Logistics

Preliminary Recovery Processes	Percentage	In house Performed (Percentage)	Third Party Performed (Percentage)
Receive return information	100.00	97.37	2.63
Inspection of returned goods	100.00	86.84	13.15
Authorisation to return goods	97.37	92.11	5.26
Collect and Transport of returned goods	94.74	47.37	47.38
Sorting of returned goods	94.74	84.21	10.53
Refund / Reimbursement	92.11	84.21	7.69
Selecting a recovery method	84.21	81.58	2.63
Gate keeping for return goods	76.32	73.68	0.00

Preliminary reverse logistics processes identified from the data are presented in columns 1 of Table 4. The processes listed in column 1 of the Table 4 have been ranked according to the percentages indicated in column 2. From column 2 it is seen that processing return information and inspection of goods are important preliminary processes in reverse logistics management. These processes are mostly managed in-house. Once the goods are inspected authorisation of return is granted, before it is transported back to the seller. At the premises of the seller sorting, refund or reimbursement and selection of returned goods take place. From the column 2 in the Table 4 it is apparent that gate keeping is a preliminary process not as important as the other preliminary processes for the sample of manufacturing organisation investigated. Column 3 and 4 indicates that most of the processes are performed in-house except collection and transportation for which companies depend on intermediaries.

From the findings listed in Table 4 it is seen that in Australia the preliminary reverse logistics processes performed are similar to those listed in the conceptual model (Figure 1). The only difference is that gate keeping is not as important as indicated in literature discussed above on which the conceptual model is developed. Literature review indicates that gate keeping enables organisations to check in the goods returned from customers and sellers, however according to this research it seems that this process might be combined with inspection.

Table 5 Recovery Processes of Reverse Logistics and How the Processes are Performed

Recovery Processes	Percentage of Organisations Perform the Process	Percentage of Returned Goods in each Recovery Process	Percentage of Organisations Performed the Process In House	Percentage of Organisations Performed the Process by Third Party
Repair	71.05	24.82	63.16	7.89
Resale	47.37	17.31	44.74	2.63
Recycling	39.47	11.76	31.58	7.89
Landfill	31.58	8.49	23.68	7.89
Remanufacture	26.32	7.60	26.32	0.00
Redistribution	23.68	9.17	23.68	0.00
Upgrade	23.68	8.11	23.68	0.00
Refurbish	23.68	4.16	21.05	2.63
Reuse	21.05	3.84	21.05	0.00
Others	10.53	1.76	10.53	0.00
Incineration	5.26	2.11	5.26	0.00
Hazardous Waste Mgt	5.26	0.73	2.63	2.63
Donation	2.63	0.14	2.63	0.00
Retrieval	0.00	0.00	0.00	0.00
Total		**100.00**		

Table 5 represents the reverse logistics recovery processes, percentage of goods managed through each of this and how the processes are performed. From the data in Table 5 it is apparent that repair of goods is the most important recovery process and the highest number of goods are put through this process. The least important recovery processes are incineration, hazard waste management, donation and retrieval. Data in Table 5 indicates that repair of returned goods is generally managed in-house.

Table 6 Technologies used to manage Reverse Logistics Information

Technologies	Electronic & Electrical Equipments & Parts	Automotives & Parts	Mining & Irrigation Equipments	Constructio n Equipment	Medical & Scientific Equipment
Land Telephone/Fax	100.0	100.0	100.0	100.0	66.7
Computers-Laptop/Desktop, Servers	94.1	100.0	100.0	66.7	66.7
Mobile phone/PDA/Other Palm device	64.7	70.0	80.0	66.7	33.3
Intranet	64.7	70.0	40.0	33.3	0.0
Tracking & Tracing System	52.9	20.0	60.0	100.0	33.3
Enterprise Resource Planning	41.2	50.0	20.0	33.3	33.3
Data Mining & Data Logging System	41.2	30.0	20.0	33.3	33.3
Barcode & Scanning System	41.2	20.0	0.0	66.7	66.7
Extranet	41.2	20.0	20.0	33.3	0.0
Internet	29.4	30.0	20.0	66.7	0.0
Electronic Data Interchange	17.6	50.0	0.0	33.3	66.7
Decision Support System	29.4	0.0	20.0	33.3	0.0
Global Positioning & GIS	11.8	0.0	0.0	33.3	0.0
Radio Frequency Identification	0.0	0.0	0.0	33.3	0.0

Table 6 lists the types of technologies used for information management in reverse logistics by the responding industries. It shows that land telephones, computers, mobile phones and intranets are the important technologies used in the management of reverse logistics information. It also indicates that mobile technologies are increasingly becoming important for reverse logistics information management. All of the technologies listed in column 1 are used to some extent. However some are more popular in one industry than other.

DISCUSSION

Form research findings discussed above it is informed that reverse logistics management is relevant to electronic and electrical equipment and parts, automotive and parts, mining and irrigation equipment and medical and scientific equipment manufacturing industries in Australia. Although most of the preliminary recovery processes presented in the conceptual model are performed by Australian organisations, gate keeping is one that stood out be not as important as indicated in literature review. This could be due to it being combined with inspection which, is clearly an important preliminary recovery processes.

Recovery process that stood out in this research was repair indicating the returned goods are mostly repaired for improvement. Other recovery processes for returned goods are performed, however they are very insignificant, except incineration, hazard waste management, donation and retrieval recoveries which are as close to not being performed at all. This is interesting because in Australian law mainly enforce product returns option as a consumer protection aspect rather than forcing organisations for resources recovery, where as other parts of the world such as Germany, Netherlands, Japan, USA, and Europe imposed laws mandating accept of returned goods for the purpose of recovering parts, materials and environment (Brito et al, 2003 and Fernandez, 2003).

Although organisations are aware of new information and communication technologies discussed above the finding indicate that telephone and fax are still important technology supporting reverse logistics information management.

This shows that reverse logistics is not a widely adopted in Australia. Finding discussed above confirm that a conceptual model for reverse logistics management such as Figure 1 is needed to supported efficient reverse logistics management.

Although it is apparent from literature review and the resulting conceptual model that effective reverse logistics management adds value to business, the findings discussed in this paper highlight that all recovery processes shown in Figure 1 are not equally important business processes for value creation among responded organisations. This indicates that the opportunities and benefits of reverse logistics has not been identified many Australian organisations, as a result many organisation in Australia miss out of substantial value that could be recovered from returned goods.

CONCLUSION

This paper is research in progress, which has discussed some preliminary findings for reverse logistics management. Further analysis of data will reveal issues for further investigations and analysis. Accordingly a modified model to guide reverse logistics management will be developed.

REFERENCES

Amini, M, M., Retzlaff-Roberts, D., 2003, Reverse Logistics Process Reengineering: Improving Customer Service Quality http://www.people.memphis.edu/~cscm/ctr5/ReverseLogistics.pdf

Andel T, 1997, 'Reverse logistics:a second chance to profit', Transportation & Distribution, vol.38, no.7, pp. 61

Bayles DL, 2001, 'E-commerce logistics and Fulfillment, Delivering the goods', Send It Back! The Role Of *Reverse Logistics*, Prentice Hall PTR, California.

Boyer KK, Hallowell R and Roth A, 2002, 'E-services : operating strategy, a case study and a method for analyzing operational benefits', *Journal Operations Management*, vol.20 no 2, pp.177-190.

Brito MP, Flapper SDP, Dekker R, 2002, 'Reverse Logistics: a review of case studies', *Econometric Institute Report* EI 2002-21, May.

Brito MP, Flapper SDP, Dekker R, 2003, 'Reverse Logistics: a review of case studies', *ERIM Report Series Research in Management*, ERS 2003-12-LIS, February.

Brito, MP, Dekker, R, 2003, 'A Framework for Reverse Logistics', *Erim Report Series Research In Management*, ERS-2003-045-LIS, April.

Caldwell B, 1999, 'Reverse Logistics', *InformationWeek*, April, No. 729, pp48-56.

Carella, G, Murino, T, Santillo, LC, 2002, 'Optimization of collection and transportation in a recovery process for a make-to-order firm', *Proceeding of Symposium on logistics*, pp 595-599.

Cavana, RY, Sekaran U, Delahaye BL, Hanney P, 2001, *Applied business research: qualitative and quantitative methods*, Milton, Qld.: John Wiley & Sons Australia, 2001.

Collis J, and Hussey R, 2003, *Business research, a practical guide for undergraduate and postgraduate students*, 2nd edition, Palgrave Macmillan.

Croxton, KL, García-Dastugue, SJ, Lambert, DM, Rogers, DS, 2001, 'The Supply Chain Management Processes', *The International Journal of Logistics Management*, Vol 12, No.2, p 14.

De Koster, RBM, de Brito, MP, van de Vendel, MA, 2002, 'Return handling: an exploratory study with nine retailer

warehouses', *International Journal of Retail & Distribution Management* Vol. 30, No. 8, pp. 407-421
Dekker, R, de Brito, MP, 2002, 'A framework for Reverse Logistics' in R. Dekker, K. Inderfurth, van Wassenhove, L, Fleischmann, M. (eds.), *Quantitative Approaches to Reverse Logistics*.
Delaney, B, 2001, 'Source for U.S. total costs', *12th annual State of Logistics Report*, National Press Club, Washington, D. C. June.
Dorp KJV, 2002, 'Tracking and tracing: a structure for development and contemporary practices', *Logistics Information Systems* Vol.15, No.1, pp24-33.
Europa, 2000, 'Commision tackles growing problem of electrical and electronic waste', Brussels, 13 June, http://europa.eu.int/comm/environment/docum/00347_en.htm
Fernández, I, 2003, 'The Concept Of Reverse Logistics. A Review Of Literature', http://www.uwasa.fi/~ifq/NOFOMApaper.pdf
Fleischmann, M, Nunen, J, Grave, B, 2002, 'Integrating Closed-loop Supply Chains and Spare Parts Management', *IBM ERIM Report Series Research In Management*, ERS-2002-107-LIS, November.
Harris, R, Goodman, J, 2001, 'B2B: back to basics – driving shareholder value', *Strategy and Leadership*, MCB University press, pp27-33.
Jiang, P, Rosenbloom, B, 2005, 'Customer intention to return online: price perception, attribute-level performance, and satisfaction unfolding over time', *European Journal of Marketing*, vol.39 no.1/2, pp150-174.
Kokkinaki AI, Dekker, R, Lee, R, Pappis, C, 2001, 'Integrating a Web-based System with Business Processes in Closed Loop Supply Chains', *Econometric Institute* EI 2001-31,
Kokkinaki AI, Dekker, R, Nunen, JV, Pappis, C, 1999, 'An Exploratory Study on Electronic Commerce for Reverse Logistics', Suppy chain forum, www.supplychain-forum.com, pp10-17
Kokkinaki AI, Dekker, R, Nunen, JV, Pappis, C, 2000, 'An Exploratory Study on Electronic Commerce for Reverse Logistics', *Econometric Institute Report* EI-9950/A,
Kokkinaki, AI, Dekker, R, Koster, MBMD, and Pappis, C, 2001, 'From e-trash to e-treasure: how value can be created by the new e-business models for reverse logistics', *EIReport*, Erasmus University Rotterdam.
Krikke, H.R., Pappis, C.P., Tsoulfas, G.T., Bloemhof-Ruwaard J., 2001, 'Design Principles For Closed Loop Supply Chains', http://ideas.repec.org/p/dgr/eureri/2001124.html
Krumwiede, DW, Sheu, C, 2002, 'Reverse Logistics Strategic Decision-Making Model:Management Of Returned Goods Through Third Party Providers', *Decision Sciences Institute, Annual Proceeding*, pp.2184-2189.
Lee, J, 2002, 'Critical Issues in Establishing a Viable Supply Chain/Reverse Logistic Management Program', *Proceeding of IEEE Conference*, May.
Lonn, S, Stuart, JN, 2003, 'Increasing service through aggressive dealer inventory return policies', *International Journal of Physical Distribution & Logistics Management*, Vol. 33, No. 6, pp. 35-47.
Lund, R, 1998, 'Remanufacturing: An American Resource', *Proceedings of the Fifth International Congress for Environmentally Consius Design and Manufacturing*, Rochester Institute of Technology, June 16 & 17.
Mason, S, 2002, 'Backward progress: turning the negative perception of reverse logistics into happy returns', IIE Solutions, vol.34 no.8, pp42-46.
Meade, L, Sarkis, J, 2002, 'A conceptual model for selecting and evaluating third party reverse logistics providers', *Supply Chain Management: An International Journal*, vol.7, no.5, pp.283-295.
Morrell, AL, 2001, 'the forgotten child of the Supply Chain', *Modern Materials Handling*, May 15,
Nagel C, Meyer, P, 1999, 'Caught between ecology and economy: end-of-life aspects of environmentally conscious manufacturing', *Computers & Industrial Engineering*, Vol.36, No.4, pp 781-792.
Nairn G. 2003, 'Reverse logistics causes headaches & eats into already thin margin', *FinancialTimes*, 5th February, pp5.
Pan, X, Ratchford, BT, and Shankar, V, 2001, 'Why aren't the prices of the same item the same at Me.com & You.com? Drivers of price dispersion among e-tailers', working paper, University of Maryland, MD.
Rogers, DS, and Tibben-Lembke, RS, 1999, 'Going backwards: Reverse Logistics trends and practices'. *Reverse Logistics Executive Council*, Pittsburgh, P.A.
Rogers, D.S., Tibben-Lembke, R.S. 1998, 'Going backwards: reverse logistics trends and practices'. The University of Nevada, Reno, Center for Logistics Management, Reverse Logistics Council.
Roy. A, 2003, 'How Efficient is Your Reverse Supply Chain?, Supply chain management', special issue, ICFAI Press, January. http://www.genco.com/news/how_efficient.pdf
Sarkis, J, Meade, LM, Talluri, S, 2004, 'elogistics and natural environment', *Supply Chain Management: International Journal*, vol. 9, no.4, pp.303-312, ISSN 1359-8546
Sekaran U, 2003*Research methods for business:a skill-building approach*, 4th ed. NewYork: John Wiley & Sons
Sharma SK, Wickramasinghe N, Singh M, 2005, 'Building systems around reverse value chain:a new system development approach', *International Journal of Management & Enterprise Development*, vol2, No1, pp93-105.
Smith, M, Bailey, J, and Brynjolfsson, E, 2000, 'Understanding digital markets: review and assessment', in Brynjolfsson, E., and Kahin, B., (Eds), *Understanding the digital economy*, MIT Press, Cambridge, MA.

Spiegel R, 2000, 'Report: 70 percent of retailers lack of e-commerce strategy', *E-commerce times*, 26 January,
Thierry, M, Salomon, M, Van Nunen, J, Wassenhove, LV, 1995, 'Strategic Issues in Product Recovery Management', California Management Review,
Tibben-Lembke, RS, 2002, 'Life after death: reverse logistics and the product life cycle', *International Journal of Physical Distribution and Logistics Management*, vol.32, no. 3, pp. 223-244.
Trebilcock, B, 2002, 'Return to Sender', *Modern Material Handling*, May, http://www.manufacturing.net/mmh/index.asp?layout=articlePrint&articleID=CA215481
Yeldham P, 2003, 'Gone to dogs: what to do with those old computers', *The Australian Financial Review*, www.atr.com, 11 September 2003.
Zhao, H, 2001, 'Simulation and analysis of dealers' returns distribution strategy', *Proceedings of the Winter simulation conference*, pp.1109-1116.

Assessing Business & IT Alignment maturity Applying Luftman's model in practice (Research in progress)

Drs. A.J. Gilbert Silvius MBA
Utrecht University of Professional Education
Utrecht, the Netherlands
Email: gilbert.silvius@hu.nl

Abstract

A key success factor for a successful company in such a dynamic environment is an effective and efficient information technology (IT) supporting business strategies and processes. In a recent survey by Synstar however, 78% of European IT managers indicate that their IT is not aligned with business strategy. The alignment between business needs and IT capabilities is therefore still a prominent area of concern. This paper builds upon Luftman's business & IT alignment maturity model and reports the application of the model to 15 Dutch firms. The survey is still in progress, but the results will be reported on the IMB 2006 conference.

Keywords:

Business & IT Alignment, IT maturity, IT management, IT planning

INTRODUCTION

In almost all industries developments like new technologies, mergers and acquisitions, entrepreneurial initiatives, regulatory changes anstrategic alliances create a dynamic business environment. A key success factor for a successful company in such a dynamic environment is an effective and efficient information technology (IT) supporting business strategies and processes. The alignment between business needs and IT capabilities is therefore a prominent area of concern. A study by the Society for Information Management (SIM), identified IT-business alignment as the No. 1 management concern among all groups surveyed, which included 300 senior IT managers (Illustration 1, Luftman 2003).

Illustration 1. *Top 10 concerns.*

In a recent survey by Synstar, 78% of European IT managers indicate that their IT is not aligned with business strategy (Synstar, 2004). Another recent survey shows similar results (Winmark, 2004). Given the 'buzz' around 'Business & IT alignment' (BIA) in recent years, these results should be surprising. BIA doesn't seem to live up to it's promise (Bloem & van Doorn, 2004).

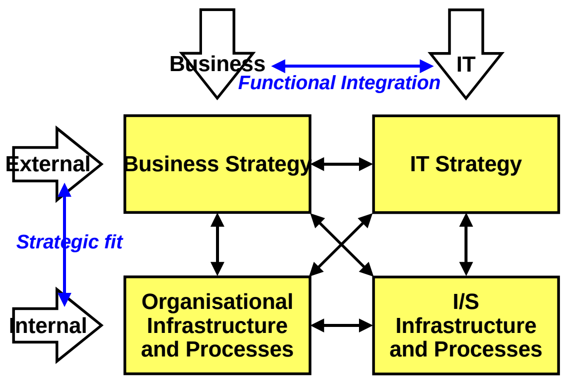


Illustration 2. *The 'Strategic Alignment Model'.*

BUSINESS & IT ALIGNMENT

Despite of the apparent importance of aligning IT and business, the majority of publications are rather vague in terms of how to define or practice alignment (Maes, et al. 2000). In well over a million Google his, a definition is seldom found. Especially consultants and IT companies tend to use the term in unclear and probably different ways. One of the reasons for this is that the theoretical foundation of BIA is still young. The most widespread and accepted framework of alignment is that of Henderson and

Venkatraman (Henderson & Venkatraman, 1993). This model, also known as the Strategic Alignment Model, describes BIA along two dimensions (Illustration 2). The dimension of strategic fit differentiates between external focus, directed towards the business environment, and internal focus, directed towards administrative structures. The other dimension of functional integration separates business and IT. Altogether, the model defines four domains that have been harmonized in order to achieve alignment. Henderson and Venkatraman pay extensive attention to the different approaches of achieving this alignment. In the model this can be visualized by starting the process of alignment from any one of the four domains.

Based on the scarce definitions in literature (A.o. Tallon, et al. 1999), Silvius defines BIA (Silvius, 2005) as:

Business & IT Alignment is the amount to which the IT applications, infrastructure and organization, the business strategy and processes enables and shapes, as well as the process to realize this.

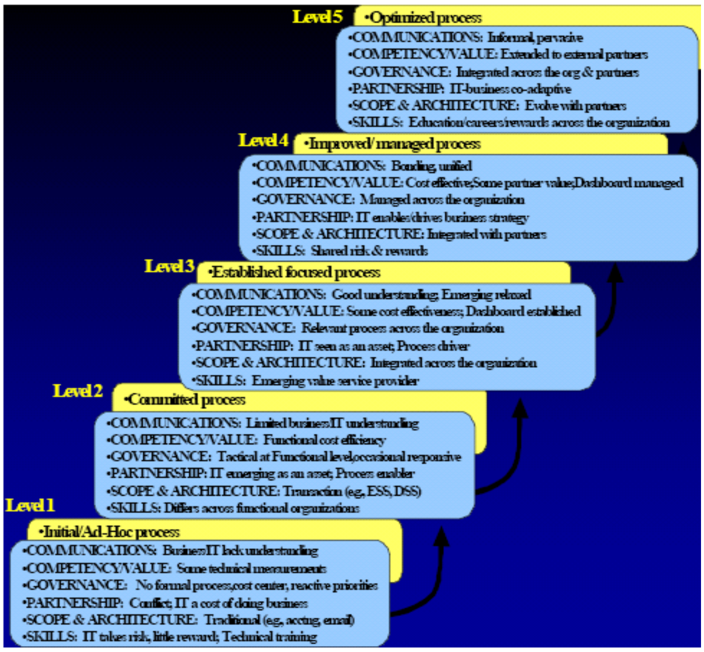


Illustration 3. *The five maturity levels*

In this definition, Business & IT Alignment can express both a 'state', the amount of alignment, as a 'process', the activities to reach a certain state of alignment. The concept of BIA as a 'state' is developed by Luftman in his assessments of the maturity of BIA (Luftman, 2000). This paper builds upon this 'BIA maturity assessment' and reports a survey to the maturity of BIA in Dutch firms.

MATURITY OF BUSINESS & IT ALIGNMENT

Luftman's BIA Assessment Model involves five levels of strategic alignment maturity:

1. Initial / Ad Hoc Process
2. Committed Process
3. Established Focused Process
4. Improved / Managed Process
5. Optimized Process

These five levels are illustrated in illustration 3. Each of the five levels of focuses, in turn, on a set of six alignment criteria. These six criteria are:

1. Communications Maturity

   How well do the technical and business staff understand each other? Do the

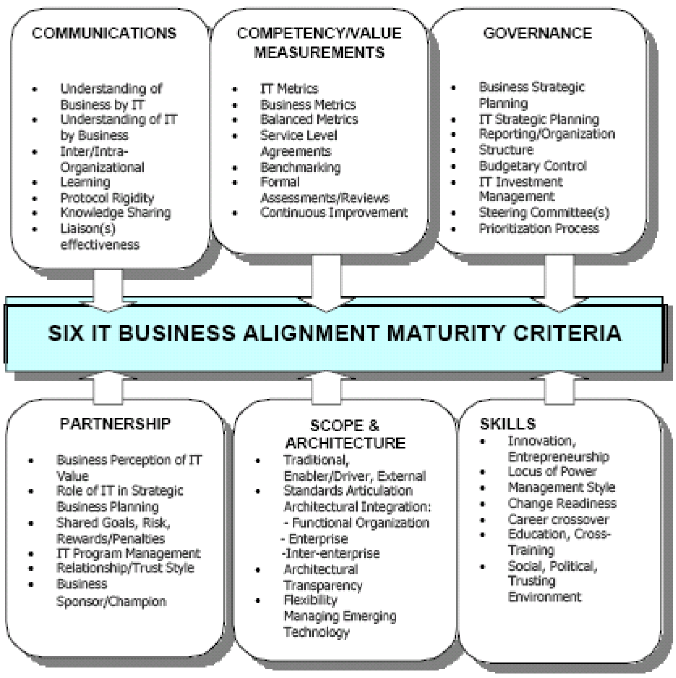


Illustration 4. *The criteria of BIA maturity.*

connect easily and frequently? Does the company communicate effectively with consultants, vendors and partners? Does it disseminate organizational learning internally?

2. Competence / Value Measurement Maturity

   How well does the company measure its own performance and the value of its projects? After projects are completed, do they evaluate what went right and what went wrong? Do they improve the internal processes so that the next project will be better?

3. Governance Maturity

   Do the projects that are undertaken flow from an understanding of the business strategy? Do they support that strategy?

4. Partnership Maturity

   To what extend have business and IT departments forged true partnerships based on mutual trust and sharing risks and rewards?

5. Scope & Architecture Maturity

   To what extend has technology evolved to become more than just business support? How has it helped the business to grow, compete and profit?

6. Skills Maturity

   Does the staff have the skills needed to be effective? How well does the technical staff understand business drivers and speak the language of the business? How well does the business staff understand relevant technology concepts?

An extensive description of the assessment model and its validation can be found in Luftman, 2000.

PILOT RESEARCH

This paper reports the assessment of BIA Maturity in 15 Dutch firms. These assessments are the first application of Luftman's assessment in the Netherlands and are a pilot for a research program aimed at the relationship between maturity of BIA and company performance.

The goal of this pilot study is to:

1. translate the BIA Maturity model to Dutch;
2. develop Dutch questionnaires;
3. get acquainted with the methodology of assessment;
4. get a first impression of the maturity of BIA of the participating firms;
5. test whether the results are recognizable.

The participating companies in the pilot study were selected from a group of participants of a series of Masterclasses organized by bITa center, an independent 'knowledge hub' on BIA (see http://www.bita-center.com/masterclasses). In this way we assured that the participants had an interest in BIA. The 15 participating companies are listed in table 1.

	Name	*Industry*	*# Employees*
1	UWV	Social security	> 5000
2	Rabobank	Finance	> 5000
3	SNS Reaal	Insurance	> 5000
4	Interpolis	Insurance	> 5000
5	Getronics	IT service provider	> 5000
6	Essent	Utilities	> 5000
7	Nuon	Utilities	> 5000
8	BMC	IT	250 - 500
9	Inter Access	IT service provider	500 - 1000
10	Quint	Business Consultancy	100 - 250
11	Deloitte	Auditing	500 - 1000
12	v Lanschot	Finance	500 - 1000
13	IBM Nederland	IT	1000 - 5000
14	DTO	Pubic	1000 - 5000
15	CFI	Pubic	500 - 1000

Table 1. *The companies participating in the pilot research.*

CONDUCTING ASSESSMENTS

Luftman proposes that the assessments should be conducted in workshops with business executives and IT exectives. The maturity 'score' should then be a consensus between all participants. Ekstedt point out the benefits, e.g. a better understanding of the business by IT staff, and drawbacks, e.g. biased opinions because of personal goals, (Ekstedt, et al, 2005). He therefore proposes the 'Organization-wide Assessment Method', a mix of individual interviews with participants to gather data and a workshop to create consensus and acceptation. In the pilot

From each company two to four persons were interviewed in positions differing from IT management, information management, general management, financial management and commercial management. In an individual structured interview an assessment of maturity was made on each of the aspects of the six criteria shown in illustration 3. The assessments were scored on a five point scale, corresponding with the five levels of maturity. The interviews were conducted in the period November 2005 – January 2006.

RESEARCH FINDINGS

Since the pilot study is still in progress at this moment (November 2005), the findings of the 15 BIA maturity assessments cannot yet be reported. It is planned however to report the experience and results of the assessments on the IMB2006 conference in February 2006.

However, the impression that remains after the first interviews seems to support the idea that the final score of alignment is not the most important output of the assessments (Ekstedt, et al. 2005). Especially if the assessment is performed for the first time in an organization, its main value is an enhanced understanding of the different aspects of BIA Maturity. It is expected that the maturity as a 'score' becomes more valuable when the assessments are a periodical process.

CONCLUSIONS

Aligning IT to business needs is still an important challenges for many organizations. This BIA requires fuzzy conditions like 'being on speaking terms with each other' and 'partnership between business and IT'. These fuzzy conditions are indications for an organization's maturity of the relationship between business and IT. Luftman introduced an assessment model to assess the maturity of BIA. This paper reported a pilot study of the BIA maturity in 15 larger Dutch companies. Since the study is still in progress at this moment, the results can only be reported at the IMB2006 conference.

However, regardless the results of the pilot study, the concept of 'maturity', in any field, raises the question: 'Does maturity matter?'. Further research to the BIA Maturity should therefore try to establish a relationship between an organization's performance and its' BIA maturity.

REFERENCES

BLOEM, J. & DOORN, M. VAN. (2004). 'Realisten aan het Roer, Naar een prestatiegerichte governance van IT' (in Dutch,), Sogeti VINT

EKSTEDT, M., JONSSON, N,. PLAZAOLA, L., MOLINA, E.S. & VARGAS, N. (2005). 'An Organization-Wide Approach for Assessing Strategic Business and IT ALignment', proceedings of the PICMET 2005 conference.

HENDERSON, J.C. & VENKATRAMAN, N. (1993). 'Strategic alignment: Leveraging information technology for transforming organizations', *IBM Systems Journal*, Vol. 32, no. 1.

LUFTMAN, J.N., (2000). 'Assessing Business-IT Alignment Maturity', *Communications of the Association for Information Systems*, Vol 4, Article 14.

LUFTMAN, J.N., (2003). 'Measure your Business-IT Alignment. The longstanding business-IT gap can be bridged with an assessment tool to rate your efforts', *Optimize Magazine*, Issue 26.

MAES, R., RIJSENBRIJ, D., TRUIJENS, O. & GOEDVOLK, H. (2000), 'Redefining Business-IT Alignment through a unified framework', white paper, http://imwww.fee.uva.nl/~maestro/PDF/2000-19.pdf.

SILVIUS (2005). 'Business & IT Alignment in theory and practice', proceedings of the IMB 2005 conference.

SYNSTAR (2004). 'The Pressure Point Index: V', Synstar.

TALLON, P.P. & KRAEMER, K.L. (1999). 'A Process-oriented Assessment of the Alignment of Information Systems and Business Strategy: Implications for IT Business Value', *Proceedings of the Fourth Americas Conference on Information Systems (AIS)*.

WINMARK & BMC SOFTWARE (2004). 'The Communication Gap: The Barrier to Aligning Business and IT'.

Developing a collaborative decision model in an ERP system

Dr. Liang-Tu Chen
Chin Min Institute of Technology
Toufen, Taiwan
Email: ltchen@ms.chinmin.edu.tw

Lecturer Mei-Chen Chien
Vannung University
Jhongli, Taiwan
Email: chien@msa.vnu.edu.tw

Dr. Jen-Ming Chen
National Central University
Jhongli, Taiwan
Email: jmchen@mgt.ncu.edu.tw

Abstract

Although the lately evolved manufacturing technologies such as manufacturing resource planning (MRP II) and enterprise resource planning (ERP) provide a unified platform for managing and integrating core business processes within a firm, the decision-making between marketing and production planning still remains rather disjoint. It is due in large part to the inherent weaknesses of ERP such as the fixed and static parameter settings and uncapacitated assumption. To remedy these drawbacks, we develop a tactical-level decision model that solves the production scheduling problem taking into account the dynamic nature of customer's demand which is partially controllable through pricing schemes as well as the restriction of finite capacity in a plant. The models can be used as an add-on optimizer like an advanced planning system (APS) in an ERP system that coordinates distinct functions with the objective of maximizing total profit.

Keywords:

enterprise resource planning (ERP), advanced planning system (APS), dynamic programming, integrating planning, lot-size and scheduling, pricing.

INTRODUCTION

The basic architecture of an MRP II/ERP system builds upon one database and a unified interface across the entire enterprise providing integrated business solutions for the core processes and the main administrative functions of an enterprise. In addition to increasing operational efficiencies, the highly integrated system can strengthen strategic advantages and generate financial and non-financial benefits that have been well documented in the literature (Langenwalter 2000, Stefanou 2001, Mandal and Gunasekaran 2002, Gattiker and Goodhue 2004, and Robinson and Dilts 1999). However, ERP benefits cannot be fully realized without a finely tuned alignment and reconciliation between system configurations, organizational imperatives, and core business processes (Al-Mashari et al. 2003). Further, the fundamental basis of planning and scheduling in an MRP II/ERP system is based on the fixed and static parameter settings (e.g., lead time, lot size, safety stock, and costs) with infinite capacity (Hsiang 2001 and Petty et al. 2000). As a result, the system generates suboptimal solutions to the production lot-size/scheduling problem.

To remedy these drawbacks, we develop a decision support model that solves the production lot-size/scheduling problem taking into account the dynamic nature of customer's demand which is partially controllable through pricing schemes. As analogous to the sales and operations planning proposed by Van Landeghem and Vanmaele (2002) or the advanced planning and scheduling systems given by Gayialis and Tatsiopoulos (2004), the proposed scheme can be served as a coordination center of the APS system within a generic ERP framework, which integrates and coordinates distinct functions within a firm (Langenwalter 2000). APS uses operations research and management science techniques to jointly optimize sales and production planning at varying levels within an enterprise. In this paper, we restrict our research scope within a single capacitated plant which produces and sells a deteriorating item in the market with the objective of maximizing the total profit as well as minimizing the inventory investment over finite planning horizon. Typical solution techniques in currently available APS systems include mixed integer programming formulation, solved by branch and bound (Simchi-Levi et al. 2000) and finite capacity heuristics such as simulated annealing and tabu search (Cohen and Huchzermeier 1999). In this paper, we use calculus-based formulation coupled with dynamic programming and iterative search techniques to solve the cross functional decision problem.

There is substantial literature on issues involving the coordination of pricing and production planning decisions. A comprehensive survey of this literature was given by Eliashberg and Steinberg (1993). To remain focus, we only review those works dealing with deteriorating items. Deterioration or decay defined by Raafat (1991) is the process that prevents an item from being used for its intended original use such as spoilage of foodstuffs, physical depletion due to pilferage, evaporation of liquids, decay of radioactive substances, degradation of high-tech products, and loss of potency of photographic films and pharmaceutical drugs. These phenomena of deterioration are prevalent and should not be disregarded in production lot-size/scheduling.

This paper considers the effect of deterioration as a function of the on-hand level of inventory. Nahmias (1982), Raafat (1991), and Goyal and Giri (2001) provided comprehensive reviews of the related work dealing with perishable or deteriorating items. However, the aforementioned literature and most of the references therein do not incorporate pricing or any other marketing related decision. Some exceptions include the models proposed by Cohen (1977), Kang and Kim (1983), Rajan et al. (1992), Abad (1996), Wee (1997), and Wee and Law (1999). The models of Cohen and Rajan et al. jointly determine the optimal replenishment cycle and price for inventory that is subject to continuous decay along product lifetime under standard EOQ cost assumptions. Abad extended the model of Rajan et al. by allowing shortages that can be partially backlogged at the end of the cycle. Wee and Law developed a model that determines the optimal number of production cycles as well as the static price over a finite horizon so as to maximize the net present worth profit. Kang and Kim, Wee, and the subsequent models developed by Abad (2000, 2001, and 2003) are variants of Abad (1996) that deal with single period setting. Instead, we present a dynamic version of the pricing-production decision model under multi-period setting, in which the deterioration rate of the on-hand inventory follows a continuous function of product's lifetime, the demand is a function of price and time, and the selling price, production cost, and production rate are allowed to vary along time.

In the following sections, we describe the problem context including assumptions and notations, present the mathematical model for the decentralized and the coordinated policies, and develop the four-layer decision support system for marketing and production planning. We also conduct a numerical study showing that the difference between the solutions generated by the two policies, and performs sensitivity analysis with respect to major parameters. Concluding remarks, applicability discussions and future research directions are given in the last section.

THE PROBLEM CONTEXT

In this section, we describe the underlying problem and settings, including assumptions and necessary notations. We consider a manufacture firm who produces and sells a single product that is subject to continuous decay over lifetime, faces a price-dependent and time-varying demand function $D(p,t)$, and has the objective of determining price and production lot-size/scheduling so as to maximize the total profit stream over multi-period planning horizon. The reason of using time-varying demand under multi-period setting is twofold: to reflect sales fluctuation over time and to reflect sales trend in different phases of product life cycle in the market. We assume the production rate is finite and shortages are allowed and completely backordered. To remain focus, we assume no inventory is held at the beginning and at the end of the time horizon. If the initial inventory level is positive in the system, no action will be taken until the depletion of inventory. In addition, we assume the deterioration of units occurs only when the item is effectively in stock, and there is no repair or replacement of deteriorated units during the planning period.

A typical behavior of the production schedule coupled with time-decreasing demand is illustrated in figure 1. Each cycle, say [z_{i-1} , z_i], of the production/inventory system starts with the production run lasting over period [z_{i-1} , T_1], followed by the consumption period over [T_1, T_2] due to demand requirements and deterioration loss, after which the system starts with the shortage period over [T_2 , T_3], at which the production resumes over [T_3 , z_i] to meet the backlogged and demand requirements.

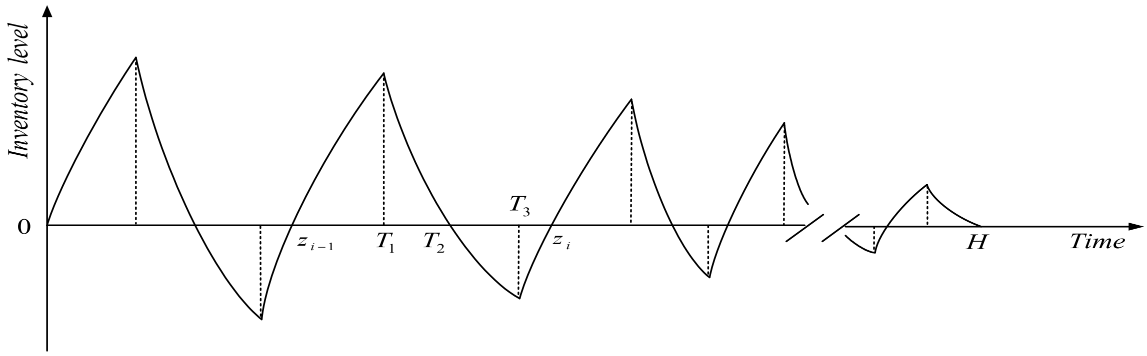


Figure 1. The production schedule for a deteriorating item with time-decreasing demand.

To incorporate the multi-period setting in a typical MRP II/ERP environment, we assume the inventory and selling price are reviewed periodically at time t, t = 0, 1, 2,…, H, where H is the planning horizon. At the beginning of each period, a joint decision is made regarding the lot-size and schedule of a new production run (if any) and its associated selling price p. The problem is equivalent to determining the optimal sequence of times z_{i-1}, i = 1, 2,…, n, at which a new cycle starts, the selling price is reset, and the production schedule (i.e., T_1, T_2, and T_3) and lot-size Q are specified simultaneously so that the total profit stream over [0, H] is maximized. It is worth noting that $n \le H$, $z_0 = 0$, $z_n = H$, z_{i-1} is integer and $z_{i-1} \in [0, H)$, and n is the total number of productions to be scheduled over the horizon.

THE MODEL

In the proposed system, a production/inventory cycle can be divided into four periods (see figure 1): the demand, deterioration, and production simultaneously incurred over period $[z_{i-1}, T_1]$, the demand and deterioration incurred over $[T_1, T_2]$, the demand only over shortage period $[T_2, T_3]$, and the demand and production over period $[T_3, z_i]$.

We first derive the optimal lot-size and price joint decisions over an arbitrary production period $[z_{i-1}, z_i]$, then prove the uniqueness of the solutions, and finally determine the optimal production schedule z^*_{i-1} and associated price p^*_{i-1} and lot-size $Q^*_{z_{i-1}}$ for i = 1, 2,…, n, using dynamic programming technique. The developed profit model (the detail is omitted) is a function of the selling price p which can be determined either by the decentralized policy or by the centralized (coordinated) policy. In the decentralized decision process, the marketing department sets the price by maximizing its gross profit function disregarding the production cost, the market responds with a specific demand, and the production department makes the lot-sizing and scheduling decision with the objective of minimizing the total production cost while satisfying the demand. As contrast to the sequential process, the coordinated policy makes the pricing and production decisions at a time.

The optimization model has emerged as a computer-based decision support system that assists decision-makers in dealing with production, marketing, and cross-functional decision problems, such as the advanced planning and scheduling (Gayialis and Tatsiopoulos 2004), sales and operations planning (Van Landeghem and Vanmaele 2002), and operations and marketing integration (Pullman, and Moore 1999 and Klassen and Rohleder 2001). The main purpose of this study is to propose a conceptual framework of a robust decision support system that can be served as an add-in optimizer like an advanced planning system in an ERP system by accessing and using data and analytic models. The four-layered framework given in figure 2 represents one of the emerging extensions of ERP systems, focusing on operations research methods embedded in commercial software applications with an aim at cross-functional cooperation and tactical-level planning. The first layer of the framework is the data needed by the decision system, including the characteristics of the product such as deterioration property, the demand type and sales trend, cost structure, capacity of production line, and inventory level. The proposed decision support system or the optimizer is outlined in the second layer which consists of three stages under the production system: problem formulation, algorithmic development, and software development. The decision process in the decision support system can be either decentralized (sequential) or coordinated (simultaneous). The third layer reports the output of the decision system including the joint decisions on production scheduling and pricing. Performance evaluation is given in the fourth layer, based on which the decision-maker can assess both operational and financial indicators such as the sales volume, service level inventory level, cost, revenue, and net profit.

IMPLEMENTATION AND SENSITIVITY ANALYSIS

We implemented the dynamic solution procedure for both policies on a personal computer with Pentium CPU at 2.0 GHz under Windows XP operating system using Mathematica version 5.0. In the experiments, we first verified numerically the concavity of the models, and then determined the optimal price and associated lot-size, and the production schedule. For the iterative procedure, the process continued until the absolute value of relative error between consecutive iterates was less than or equal to 10^{-5}. The process took about 3-5 iterations to converge in all experiments being studied. The computer time required for the dynamic programming was, on the average, less than 5 seconds. Several experiments were conducted to attend qualitative insights into the structures of the proposed policies and their sensitivity with respect to major parameters. We particularly focused on investigating the solution property as well as the benefit of the coordinated policy compared to the decentralized policy in settings with time-decreasing demand. Further, we explored the impacts of price-sensitivity coefficient of demand, coefficient of production rate, and production unit cost on the profits generated by the two policies.

Comparing the solutions generated by the two policies, the coordinated policy generates higher price and more profit. One additional performance measure is introduced in the study: the coordinated policy produces smaller value of inventory which implies less storage capacity required and, as a result, less investment needed in warehousing and material handling system. Using the settings in the example, we compare the net profits generated by the two policies with respect to major parameters, such as the production unit cost, coefficient of production rate, and price-sensitivity coefficient of demand. The coordinated policy outperforms the decentralized significantly in the total net profit as price-sensitivity coefficient and production unit cost increase, while it is insignificant to the larger value of coefficient of production rate. Next, we investigate the impacts of several factors such as deteriorating rate, set-up cost, holding cost, and shortages cost on the total profit generated from the coordinated policy. It suggests that the total profit is more sensitive to the setup cost and the deteriorating rate, but is less sensitive to the holding and shortage costs.

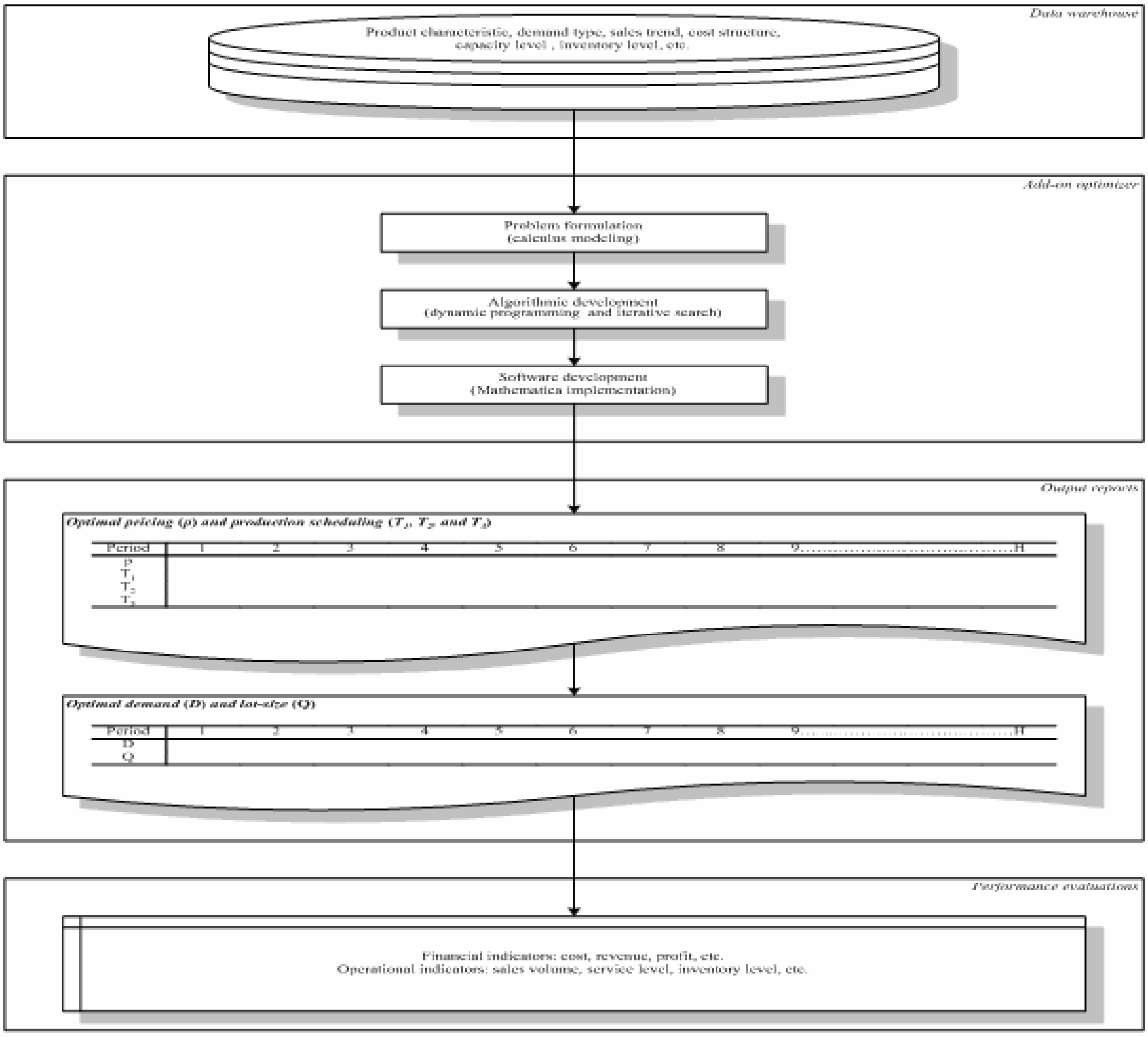


Figure 2. Four-layered decision support system for marketing and production planning.

CONCLUSION

In this paper, we has developed the four-layer decision support system for marketing and production planning and has proposed two decision-making policies, coordinated and decentralized, that determine the optimal price and production lot-size/scheduling for a deteriorating item over finite planning horizon. We have presented the necessary and sufficient conditions to the maximization problem, formulated the problem as a dynamic programming model, and provided the solution procedure. An extensive numerical study has been conducted to attend qualitative insights into the structures of the proposed policies and their sensitivity with respect to major parameters such as price-sensitivity coefficient, production coefficient, and unit production cost. The numerical results have shown that the solution generated by the coordinated policy outperforms that by the decentralized in maximizing the net profit and other quantifiable measures such as minimizing inventory investment and storage capacity. Further, we have shown that the percentage of profit difference between the two policies increases significantly in price-elasticity coefficient of the demand function as well as the production unit cost. The proposed model is based on a periodic review policy which makes it applicable in many manufacturing planning and control practices. It can be used as an add-on optimizer like the advanced planning system in an ERP system. A nature extension of this research is to consider more complicated and practical demand and deterioration functions in the model, such as the stochastic demand and the fuzzy-modeling deterioration. Another direction of this research is to develop a prototype of an advanced planning system with an ERP system that integrates the management science techniques into commercial software for collaborative and robust planning.

REFERENCES

Abad, P.L., 1996. Optimal pricing and lot-sizing under conditions of perishability and partial backordering. Management Science 42 (8), 1093-1104.

Abad, P.L., 2000. Optimal lot size for a perishable good under conditions of finite production and partial backordering and lost sale. Computers & Industrial Engineering 38, 457-465.

Abad, P.L., 2001. Optimal price and order size for a reseller under partial backordering. Computers & Operations Research 28, 53-65.

Abad, P.L., 2003. Optimal pricing and lot-sizing under conditions of perishability, finite production and partial backordering and lost sale. European Journal of Operational Research 144, 677-685.

Al-Mashari, M., Al-Mudimigh, A., Zairi, M., 2003. Enterprise resource planning: a taxonomy of critical factors. European Journal of Operational Research 146, 352-364.

Cohen, M., Huchzermeier, A., 1999. Global supply chain management: a survey of research and applications. In: Tayur, S., Ganeshan, R., Magazine, M. (Eds.), Quantitative Models for Supply Chain Management, Kluwer Academic Publishers, Dordrecht, pp.671–702.

Cohen, M.A., 1977. Joint pricing and ordering policy for exponentially decaying inventory with known demand. Naval Research Logistics Quarterly 24, 257-268.

Gattiker, T.F., Goodhue, D.L., 2004. Understanding the local-level costs and benefits of ERP through organizational information processing theory. Information & Management 41, 431-443.

Gayialis, S.P., Tatsiopoulos, I.P., 2004. Design of an IT-driven decision support system for vehicle routing and scheduling. European Journal of Operational Research 152, 382-398.

Gilbert, S.M., 2000. Coordination of pricing and multiple-period production across multiple constant priced goods. Management Science 46 (12), 1602-1616.

Goyal, S.K., Giri, B.C., 2001. Recent trends in modeling of deteriorating inventory: an invited review. European Journal of Operational Research 134 (1), 1-16.

Hsiang, T., 2001. The illusion of power. OR/MS Today February.

Kang, S., Kim, I.T., 1983. A study on the price and production level of the deteriorating inventory system. International Journal of Production Research 21, 899-908.

Langenwalter, G.A., 2000. Enterprise Resources Planning and Beyond Integrating Your Entire Organization. Boca Raton FL: CRC Press, LCC.

Mandal, P., Gunasekaran, A., 2002. Application of SAP R/3 in on line inventory control. International Journal of Production Economics 75, 47-55.

Nahmias, S., 1982. Perishable inventory theory: a review. Operations Research 30 (4), 680-708.

Petty, D.J., Stirling, M.D., Travis, L.C., Bennett, R., 2000. Conditions for the successful implementation of finite capacity/MRP II hybrid control systems. Proceedings of the institution of Mechanical Engineers 214, 847-851.

Raafat, F., 1991. Survey of literature on continuously deteriorating inventory models. Journal of the Operational Research Society 42 (1), 27-37.

Rajan, A., Rakesh, Steinberg, R., 1992. Dynamic pricing and ordering decisions by a monopolist. Management Science 38 (2), 240-262.

Robinson, A.G., Dilts, D.M., 1999. OR & ERP: a match for the new millennium? OR/MS Today June.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E., 2000. Designing and Managing the Supply Chain. Irwin, Homewood/McGraw-Hill, Boston.

Stefanou, C.J., 2001. A framework for the ex-ante evaluation of ERP software. European Journal of Information Systems 10, 204-215.

Van Landeghem, H., Vanmaele, H., 2002. Robust planning: a new paradigm for demand chain planning. Journal of Operations Management 20, 769–783.

Wee, H.M., 1997. A replenishment policy for items with a price-dependent demand and a varying rate of deterioration. Production Planning & Control 8 (5), 494-499.

Wee, H.M., Law, S.T., 1999. Economic production lot size for deteriorating items taking account of the time-value of money. Computers & Operations Research 26, 545-558.

Using Enterprise Resource Planning Tools in Real-time Supply Chain Coordination

Mohammad J. Tarokh, Javad Soroor
Postgraduate Information Technology Engineering Group, Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran

Abstract

The transaction based integrated Enterprise Resource Planning (ERP) software provides different tools that can support real-time supply chain coordination and integration but at the same time it has several features that obstruct the integration with business partners. We concentrate on the inventory management aspects of real-time supply chain coordination reviewing the recent quantitative modelling and organizational results available in literature. We summarize the results of a detailed numerical and sensitivity analysis based on our previously published models for supply chain coordination and joint optimal ordering and shipment policies for the buyer and supplier. These results can be used in enterprise software to measure the potential monetary value of policy coordination, to facilitate real-time coordination, and minimize the total supply chain system cost. Our further goal is to combine quantitative tools with organizational and management factors, and to integrate them in a multi-level framework of policy coordination.

Keywords:

Supply chain, Real-time coordination, ERP, Inventory management, Organizational factors

1. INTRODUCTION

Effectively integrating the information and material flows within the demand and supply process is what Supply Chain Management (SCM) is all about (Soroor, 2005a). The potential for improved productivity, cost reduction, and customer service are enormous (Tarokh & Soroor, 2005a). Enterprise Resource Planning (ERP) software systems have focused on internal process coordination and integration of traditional functions, such as sales, production, and inventory management. The transaction-based integrated processing provides different tools that can support real-time supply chain coordination but at the same time it has several aspects that obstruct the integration with business partners. By gaining access to the suppliers' production and delivery schedules, buyers can improve their own production plans and delivery schedules. Correspondingly, suppliers can use the buyer's real-time store level data to plan their inventory levels, and production schedules. Sharing order status information among the supply chain partners improves customer service quality, speeds up the payment cycle, and provides cost savings. Sharing data regarding to performance metrics such as lead times, quality specifications, return status, etc., helps supply chain partners to identify and overcome the bottlenecks in the supply chain and enables real-time supply chain coordination.

In this paper first we summarize the most important tools and concepts of ERP systems that help in supply chain information sharing, real-time coordination, and cost optimization. On the other hand, the obstacles of coordination are characterized and suggestions for improvement are listed. In the third section, we provide the quantitative analysis and support for the three different development phases: first we discuss the traditional, adversary relationship between buyer and supplier, and then the case of partnership and the joint optimal policies are discussed, followed by the consideration of the effects of supply chain network. In the fourth section we deal with the organizational analysis and summarize the barriers, bridges of inter-organizational coordination and cooperation. Next, we outline a multi-level framework of real-time coordination for supply chain partners. In the last section conclusions and further research plans are summarized.

2. ERP TOOLS: OPPORTUNITIES AND OBSTACLES FOR REAL-TIME SUPPLY CHAIN COORDINATION

The implementation of ERP enables the companies to move towards an extended enterprise business model that enhances value across the total supply chain. In order to gain real-time supply chain coordination, companies need to exchange large amount of planning and operational data, ranging from information for annual contracts and periodic progress reporting to real-time delivery and invoicing data. The advantages and obstacles of ERP tools have been discussed in several research papers. Next we quote some relevant statements from the recent articles that underline our research problems and approaches:

- Although ERP packages strive to integrate all the major processes of a firm, customers discover that some essential functionality is lacking (Scott & Kaindl, 2000).
- Traditional ERP infrastructures failed to support an extended business model across the supply chain (Edwards et al., 2001).
- The challenge is to figure out what, how, where, who, when, and why manufacturing operations can feed the ERP beast (Harrold, 2001).
- Since ERP philosophy is process based, rather that function based, it necessitates disruptive organizational changes (Hong & Kim, 2002).
- ERP systems mostly adopt a myopic view of planning, based on pure deterministic planning methods (Landeghem & Vanmaele, 2002).

ERP provides several tools; the two most important for real-time supply chain coordination are the real-time transaction tracking and the internal process coordination. Next we outline the four main opportunities offered and the obstacles of using them. In Table 1 we formulate four research questions and indicate the directions how we try to answer these questions in the subsequent sections using quantitative and organizational analysis.

The traditional vertically integrated business model requires re-evaluation. The ERP software vendors saw the above problems and started providing advanced decision support tools that are the new ERP software extensions. Among them the most important directions are: the Advanced Planning and Scheduling (APS), Demand Planning and Revenue Management (DPRM), Customer Relationship Management (CRM), Sales Force Automation (SFA), and Supply Chain Management (SCM).

In this paper, we concentrate on the inventory management aspects of real-time supply chain coordination. With our quantitative analysis we try to present results in information systems area by

- providing guidelines for transaction tracking,
- promoting visibility of information for the supply chain, and by
- supporting real-time coordination among business partners.

In the decision support area we try to

- improve the quality of ordering and transportation decisions (for SCM and DPRM),
- provide quantitative tools to find the joint optimal policy and help in contract negotiations (for SCM, APS, and CRM),
- estimate the fair amount of compensation necessary (for CRM),
- facilitate real-time coordination between the buyer and the supplier (for CRM).

Different theories are available to support real-time supply chain coordination. Management theory states that efficient supply chain requires real-time coordination. Quantitative modeling proves that the total system cost can be reduced by coordinated policy. At the same time, practice, and theory shows several barriers in organizational factors (organizational theory) and subjective factors (behavioral theory). Few results integrate the results and combine the quantitative and organizational effects. Our goal is to

- provide quantitative modeling, numerical and sensitivity analysis to measure the potential monetary value of real-time coordination,
- combine it with organizational and management factors, and
- integrate in a multi-level framework of policy coordination.

Table 1: Research questions

Real-time transaction tracking

Opportunity 1: tool for information sharing

Major obstacle: unwillingness to share information.

How to motivate companies for information exchange?

We show the potential monetary benefits and savings achievable by information sharing.

Opportunity 2: availability of large amount of planning and operational data

Major obstacle: too much, too detailed information.

How to select and aggregate data to share with partners?

We provide data analysis and the prospective of activity based costing (ABC).

Internal process coordination

Opportunity 3: improves operational decision within the company

Major obstacle: fails to support operational decisions with supply chain partners.

How to extend the traditional vertically integrated business model?

We provide coordinative models for buyer and supplier.

Opportunity 4: tool for policy coordination

Major obstacle: fails to support operation policy for real-time coordination with supply chain partners.

What is the best operation policy of real-time coordination?

We provide joint policy optimization for real-time coordination.

3. QUANTITATIVE SUPPORT FOR SUPPLY CHAIN OPTIMIZATION

We start this section with a brief summary and classification of recent literature followed by the summary of our quantitative results in the three phases of the supply chain development: in adversary relationship, in partnership, and in extended supply chain network. In the quantitative analysis, our objective is to quantify the supply chain relations and evaluate the costs and benefits of relations that can be used in enterprise software to facilitate real-time coordination and minimize the total supply chain system cost.

We consider three phases in the subsequent three subsections, in Section 3.1 the traditional, adversary relationship is discussed, where either the supplier's dominance has the effect of large production lot sizes or the buyer's dominance affects small, frequent shipments. In both cases there is a considerable loss, the high total system cost, for which we provide quantitative evaluation. In Section 3.2 the partnership and JIT case is analyzed. We provide joint optimization and its quantitative evaluation. These quantitative results can also serve as motivating and negotiating tool for sharing information and providing joint operating policies. Section 3.3 extends the results to supply chain network and in particular summarizes the quantitative evaluation of bullwhip effect and its remedies.

3.1. Traditional, adversary relationship

The traditional way of supply is to have a large number of suppliers, where relationship is based on price. Buyer gives detailed technical specifications; still a limited communication and information exchange is typical. This relationship has some advantages for buyer like easy switching between suppliers, lower costs, having the supplier as the shock absorber, and not sharing the confidential information. The major advantage for suppliers is that confidential information are not shared. These advantages are dominated by the disadvantages, like for the buyers the large amount of inventory, low quality, and delivery problems; for the suppliers the short-term contracts and arms-length transactions, limited input, large amount of inventory, and increased pressure for price reduction.

The optimal policy for the buyer is to have frequent small-size shipments that result in a large loss for the supplier. Similarly, the large lot sizes and shipment sizes, preferred by the suppliers, result in large loss for the buyers. Quantitative models to evaluate these losses have been provided by (Kelle & Miller, 2000). Further extensive numerical analysis, recently executed, has provided the results, summarized briefly and illustrated next.

The buyer's demand and cost parameters are:

D	buyer's demand with constant rate
A_B	buyer's ordering cost
C_B	selling price
r_B	buyer's annual inventory carrying cost rate
Z_B	receiving cost for the buyer
L_B	cost rate of losing flexibility for the buyer—similar to carrying cost rate but the flexibility loss rate is related to the value of contract commitment, not the value on hand.

The supplier's production and cost parameters are

p	production rate
A_S	supplier's fixed setup cost
C_S	production cost
r_S	supplier's annual inventory carrying cost rate
Z_S	supplier's fixed shipment cost

We assume that the buyer's order, Q_B, is delivered in n shipments of size $q=Q_B/n$. If the buyer is strong and forces its optimal shipment size and frequency on the supplier, the supplier can choose its economic lot size as an integer multiple of the shipment quantity, $Q_S = mq_B$.

During the execution of different full and partial experimental designs, the extensive numerical and sensitivity results show that the loss in setup, inventory holding, and shipment related cost of the supplier is typically in the 50-90% range compared to the supplier's optimal policy. The supplier's loss is higher if the buyer's carrying cost, or selling price, or the production rate increase. Moreover, the loss is decreasing with the increase in setup cost, production cost, or demand rate.

The optimal policy for the supplier is to have a large lot size and shipment size. This policy results in a considerable loss for the buyer, typically in the 80-180% range compared to the buyer's optimal policy. Our numerical results show that the buyer's loss is higher for large setup cost, or carrying cost rate, or selling price. Higher ordering cost or production cost or demand rate decreases the buyer's loss.

Generally, if the r_B/r_S, or C_B/C_S, or p/D rates are high, the weaker party's loss will be larger compared to its optimal policy. The increase in the A_B/A_S rate has an increasing effect on the supplier's loss but a decreasing effect on the buyer's loss. The selling price vs. production cost rate and the carrying cost for buyer versus supplier rates have the largest effect on the loss for the buyer. The shipment cost for the supplier vs. the buyer and the carrying cost for buyer vs. supplier have the largest effect on the loss for the supplier.

These numerical and sensitivity results, together with further numerical results illustrated next, can be applied in practice to measure the potential benefits of changing policies, estimate the value of coordinated policy, and help in negotiation between buyer and supplier and to facilitate real-time coordination.

3.2. Partnership, JIT, joint optimal policies

The buyer-supplier partnership reduces the number of suppliers and results in a fair price structure for both parties. It has the potential of joint investments in product and process development and an extensive communication and information sharing. The main advantages for the buyers include reduced manufacturing and labor costs, improved quality, supply assurance, reduced inventories, and frequent deliveries in small lots. For the suppliers, contract predictability, buyer assistance, increased R&D, influence on buyer's future decision making are among the main advantages. The partnership has potential disadvantages. For the buyers, the disadvantages are increased dependence on suppliers, new negotiating styles, communication costs, loss of direct contract with 2nd tier suppliers. For the suppliers, the disadvantages are the risk of loosing confidential information, increased communication and coordination costs, and increased pressure among others.

The typical partnership is in the form of a JIT agreement. JIT supply calls for frequent deliveries in small quantities that require the supplier's flexibility, which is a burden on supplier. On the other hand, long-term contracts and exchange of information are burdens on buyer. Our goal in this paper is to provide tools for negotiation and real-time coordination and show the potential large savings in total supply chain cost provided by the partnership can be used to overcome these burdens of JIT partnership from both sides.

The three decision variables are the shipment size, q, the buyer's order quantity: $Q_B = nq$ and the supplier's production lot size: $Q_S = mq$. Our objective is to find the joint optimum of (q, m, and n) and compare its cost with the other policies. In negotiation it will always benefit the weaker party to entice the stronger party to agree to the joint optimal shipment policy. The weaker party can offer compensation to the stronger party and still improve their cost position. Extensive numerical and sensitivity analysis results supported the following tendencies: The savings incurred by the weaker party in negotiating the joint optimal policy increases as the buyer's carrying cost, the selling price, and the production rate increase. Moreover, these savings will decrease with the increase in the setup cost, the cost rate of the buyer for losing flexibility, the production cost, and the demand rate.

If the buyer is the weaker party, the savings will be larger for higher setup cost or supplier's carrying cost. However, if the supplier is the weaker party the opposite will occur. The selling price vs. production cost and the carrying cost for the buyer vs. the supplier have the largest effect on the savings for the buyer. The shipment cost for the supplier vs. the buyer and the carrying cost for the buyer vs. the supplier rates have the largest effect on the savings for the supplier.

Optimal joint policies under uncertainty effects have been analyzed in details previously (Kelle & Miller, 1998). The results showed the typical inventory cost saving of the optimal joint policies with safety stock coordination over the policies:

	Random Lead Time (%)	Random Yield (%)	Random Demand (%)
Optimal joint policy with no safety stock coordination	10-38	15-40	15-35
Purchaser's optimal policy	13-60	16-70	15-60
Buyer's optimal policy	15-300	20-400	15-300

The main results of this quantitative analysis can be summarized as:

- The joint optimal policy will always result in savings in the total system cost for the supply chain.
- The savings achieved by the weaker party by negotiating for the joint optimal policy rather than the other party's optimal policy will be in the range of:
  - 1-30% for the supplier, and
  - 25-60% for the buyer.
- The joint optimal policy is closer to the buyer's optimal JIT policy.

- The savings in total system cost of the joint optimal policy can be split between the two parties providing a potential benefit for both buyer and supplier (win-to-win).
- Coordinating the safety stock policy can result in a large cost saving (typically 15-400%).

There is a major practical question: How can we motivate the companies for information exchange? Several recent papers deal with the buyer-supplier cooperation, negotiation, and contracts (Kelle & Miller, 1998; Kelle et al., 2001). Other recent publications include: Hoffmann (2000) about channel coordination through quantity discounts with multiple price brakes; Corbett and de Groote (2000) providing the supplier's optimal quantity discounts under asymmetric information; Chen et al. (2001) dealing with discount based on sales volume, order quantity and frequency; Ganeshan et al. (2001) analyzing the effect of forecast error, communication and planning frequency; Dong et al. (2001) showing that JIT purchasing directly reduces costs for buyers, suppliers too can benefit, at least indirectly, from JIT purchasing.

Next question is to specify the content and aggregation of information sharing. The most important groups of data to share include:

- Operations information (Production schedules, Order tracking, Return status, Volume of operations, Inventory levels)
- Planning information (Forecasting, Sales, Production plans)
- Customer requirement information
- Financial information

The ideal accounting tool available in the enterprise software products is the Activity Based Costing (ABC). The ABC approach has been used for inventory holding cost, ordering and setup cost evaluation among others. The general way is to find the cost drivers, resources, activities, and make the appropriate cost assignment to products and materials. Player and Kramer (1995) report on a successful implementation of ABC system based automatic replenishment system; Baxendale and Gupta (1998) deal with the implementation of ABC for inventory and capacity management; Damito et al. (2000) outline the design and implementation of ABC; Gupta and Galloway (2001) discuss an ABC-based information system to support effective operations decision making.

Several methodologies are available for measuring information sharing. Frohlich and Westbrook (2001) introduce the concept of "arc of integration", which measures the direction and degree of integration; Barut et al. (2002) introduce a generic measure to what degree a firm is integrated with the members of its supply chain. Several vendors and consulting companies provide tools for determining information requirements including Distribution Sciences, Inc., Ross Systems, Sabre Decision Technologies, etc. Some of the most powerful methods are:

- Cross-organizational Joint Application Design (JID) Session
- Business System Planning (BPS) Technique—IBM
- Critical Success Factor (CSF)
- Ends/Means (E/M) Analysis

3.3. Supply chain network

Extending the relations from simple buyer-supplier real-time coordination to a whole supply network, from raw material suppliers to final customers provides several advantages that include: for the buyers, a large pool of suppliers, reduced transaction costs, market transparency, purchase transparency, lower prices, dynamic pricing models, control of maverick, buying, and lower inventory costs (Soroor, 2005b). For the suppliers, the benefits among others are large pools of buyers, real-time information, time to market, aggregation of small orders, efficient fund transfer.

On the other hand, real-time supply chain coordination has some potential disadvantages like: for the buyers, unqualified suppliers, miscommunication, failed promises, hidden switching costs, missed value-creating opportunities; for the suppliers, the disclosure of confidential information, pressure for price reductions, easy supplier switching, loss of established relationships, high initial investment.

One critical problem in an extended supply chain is the bullwhip effect. Empirical observations show that the variability of orders may considerably increase relative to the variability of demand (Tarokh & Soroor, 2005b). The effect is an accumulating uncertainty in the supply chain, which causes high cost of safety stock or poor service. Several papers in integrated supply and logistics are dealing with different aspects of the bullwhip effect (Anupindi & Akella, 1993; Lau & Zhao, 1993; Ramasesh et al., 1993; Nahmias & Smith, 1994; Schneider et al., 1995; Lee et al., 1997; Cahon, 1999; Xu et al., 2001).

Some recent papers such as Kelle and Milne (1999), Cachon and Fisher (2000), Chen et al. (2000), and Li and O'Brien (2001) deal with quantifying the magnitude of the bullwhip effect. Others such as Bourland et al. (1996), Chen et al. (1998), Gavirneri et al. (1999), Cetinkaya and Lee (2000), Lee et al. (2000), Lee and Whang (2000), and Yao et al. (2001) try to measure the effect of counteractions (like VMI, CPFR).

Next the results are summarized, where simple approximations are provided to show the effect of demand, cost, and policy parameters (measure and dependence of bullwhip effect): Variability increase at a retailer due to an (*s*,*S*) ordering policy can be approximated by a simple analytic formula:

$$\mathbf{(2y^2)/[x^2 + (2y + x^2 + 1)],}$$

where x is the demand variability, and y the average number of periods between orders.

Result 1. Variance amplifying effect is very large for large y and small x values.

- Variability of aggregate orders of N retailers:

Result 2. Variance amplifying effect does not depend on correlation between markets.

- Effect of autocorrelation in orders on lead time variability:

Result 3. Variance reducing effect is usually small relative to the amplifying effects. Remedies of bullwhip effect include:

- Improving forecast
  - Sharing point-of-sale data
  - Collaborative Forecasting and Planning
- Coordinating orders
  - Building strategic partnerships
  - Vendor Managed Inventory (VMI)
- Continuous Replenishment Programs (CRP)
  - Role of Enterprise Systems
  - Improved connectivity, Customer Relationship Management
  - Computer-assisted ordering, analytical applications, APS
  - Reducing lead time, lot sizes, SCM systems

4. ORGANIZATIONAL FACTORS

While the numerical modeling in the previous section shows the potential advantages of buyer-supplier policy coordination in the supply chain, there are a number of organizational factors that may influence whether firms will participate in this type of inter-organizational real-time coordination. In Section 4.1 findings from seven recent empirical studies on participation in real-time supply chain coordination are summarized listing the ten most important barriers and bridges. The purpose of Section 4.2 is to present a conceptual framework for theoretical development in this area that is cross-disciplinary and cross-level, in order to provide more comprehensive views and facilitate theoretical synthesis. Frameworks of this type, for example, could provide conceptual bridges between an analytical operations research perspective and an organization/management theory perspective, leading to new theories and important new insights.

4.1. Barriers and bridges

Based on earlier literature (Akintoye et al., 2000; Fawcett & Magnan, 2002; Gebauer & Buxmann, 2000; Gregor, 2000; Shore, 2001; Wilson & Nielson, 2001; Wong, 1999), we summarize the top 10 barriers to buyer-supplier real-time coordination:

- Inadequate information systems
- Poor/conflicting measurement
- Inconsistent operating goals
- Organizational culture and structure
- Resistance to change—lack of trust
- Poor alliance management practices
- Lack of supply chain vision/understanding
- Lack of managerial commitment
- Constrained resources
- No employee passion/empowerment

According to those references, the top10 bridges to real-time coordination are (Fawcett & Magnan, 2002):

- Senior and functional managerial support
- Open and honest information sharing
- Accurate and comprehensive measures
- Trust-based, synergistic alliances
- Supply chain alignment and rationalization
- Cross-experienced managers
- Process documentation and ownership
- Supply chain education and training
- Use of supply chain advisory councils
- Effective use of pilot projects

4.2. Multi-level framework of real-time coordination

As described below, five perspectives represented in the framework provides a valuable viewpoint from which to understand the factors that may influence participation in buyer-supplier policy coordination.

- *Individual:* At the level of the individual, for example, we can draw from theory in managerial and organizational psychology to develop an understanding of how managers gather information, identify opportunities, and make decisions regarding policy coordination relationships (see, e.g. Walsh, 1988).
- *Two or more individuals:* The second theoretical perspective offered by the framework can be used to examine relationships between/among individuals in the buyer and supplier organizations. For example, a potential cause for the inability to form inter-organizational alliances is a clash of executive personalities (Moss-Kanter, 1994). Alternatively, social bonding—the degree of mutual personal friendship and liking shared by the individuals representing the buyer and the seller—may facilitate the formation of these real-time coordination relationships (Crotts, 1998). Social network theory provides another theoretical lens at this level of analysis, e.g., Uzzi's (Uzzi, 1997) constructs and measures to examine the behavioral and attitudinal orientations of exchange parties.
- *Single organization:* Using the single organization perspective, we can investigate the factors that may influence a particular organization's (buyer or supplier) decision to engage in policy coordination arrangements. For example, economic theories of the firm (see Seth & Thomas, 1994), or theories that predict organizational actions in response to threats and opportunities (Chattopadhyay, 2001) could provide valuable insights at this level of analysis.
- *Buyer-supplier dyad:* Pairs of buyer and supplier organization are the fourth perspective, enabling the examination of elements of the inter-organizational relationship such as power, conflict, and motivations. Organizational differences, such as incompatibility of organizational culture and values could also affect the formation of policy coordination arrangements.
- *Multiple organizations/Organizational environment:* The last perspective provides a level of analysis that can incorporate multiple buyers and sellers (e.g., industry groups) as well as other elements of the organizational environment. Organization theory, for example, offers several theories that can be applied at this level of analysis. Resource dependence theory (Pfeffer & Salancik, 1978) incorporates the concepts of exchange relations, exchange networks, resource dependence, and power to describe and explain how external environments affect and constrain organizations and how organizations respond to those constraints. Institutional theory (Scott, 1995) and population ecology (Hannan & Freeman, 1977) can be used to examine how institutions (e.g., legal and regulatory systems, industry structures and norms, IT standards) affect and are affected by organizations. This level of analysis, for example, could be used to investigate participation in consortia arrangements for buyer-supplier real-time coordination.

While each of these levels of analysis offers a distinct and valuable perspective, combining multiple levels of analysis (e.g., industry group and individual organization) may also yield valuable insights.

5. SUMMARY AND FUTURE RESEARCH DIRECTIONS

In this section we summarize our major contribution and future plans in the sequence of the four basic research questions listed in the first section of this paper.

- *How to motivate companies for information exchange?*

We quantified potential benefits and margins for negotiation. Results of our modeling and sensitivity analysis show that if one of the supply chain partners forces its optimal policy on the other partner, the total operating cost of the system can be much higher than with a real-time coordinated ordering/setup and shipment policy. The large cost benefit potential in policy coordination motivates the sharing of confidential information. Further organizational analysis of barriers and motivation based on company interviews and surveys has to be planned.

- *How to select and aggregate data to share with partners?*

We provided data analysis, and guidelines using activity based costing principles. We plan to provide a data aggregation prototype using activity based costing.

- *How to extend the traditional vertically integrated business model?*

We provided quantitative models for optimization and real-time coordination. We plan to extend the integration model including more organizational factors and using results of company interviews and surveys.

- *What is the best operation policy of real-time cooperation?*

We provided joint policy based on optimization models for operation and negotiation. Policy coordination has larger saving effect than information sharing alone. We show how cost and demand parameters influence the value of policy coordination for the buyer and for the supplier, and which factors are the most important in gaining more savings in operations cost. We plan to extend the optimization model portfolio.

While the numerical modeling shows potential advantages, several organizational factors may pose barriers to real-time coordination. We extended the results of prior research by drawing from organizational and management theory to identify the key factors most likely to prevent policy coordination, in particular, and organizational strategies to address them. The next steps will be to extend the framework in two ways:

(1) Incorporating the results of prior research on buyer-supplier policy coordination and other types of buyer-supplier real-time coordination, and

(2) Incorporating relevant organizational theories (including factors and relationships predicted by theory).

Extending the framework in these ways will enhance its value in guiding future empirical research by offering an integrated set of perspectives with which to identify and examine key research questions related to supply chain real-time coordination. In the future, an integrated framework could also be useful guiding management practice, for example, by identifying the key barriers to policy coordination and management strategies to reduce those barriers.

REFERENCES

Akintoye, A., McIntosh, G., & Fitzgerald, E. (2000). A survey of supply chain collaboration and management in the UK construction industry, *European Journal of Purchasing & Supply Management*, 6, 159–168.

Anupindi, R., & Akella, R. (1993). Diversification under supply uncertainty, *Management Science*, 39, 944–963.

Barut, M., Faisst, W., & Kanet, J.J. (2002). Measuring supply chain coupling: An information system perspective, *European Journal of Purchasing and Supply Management*, 8(3), 161–171.

Baxendale, S., & Gupta, M. (1998). Aligning TOC and ABC for silkscreen printing, *Management Accounting*, 39–44.

Bourland, K.E., Powell, S.G., & Pyke, D.F. (1996). Exploiting time demand information to induce inventory, *European Journal of Operational Research*, (92), 239–253.

Cachon, G.P., & Fisher, M. (2000). Supply chain inventory management and the value of shared information, *Management Science*, 46(8), 1032–1048.

Cahon, G. (1999). Managing supply chain demand variability with scheduled ordering policies, *Management Science*, 45(6), 843–856.

Cetinkaya, S., & Lee, C.Y. (2000). Stock replenishment and shipment scheduling or vendor managed inventory systems, *Management Science*, 46, 217–232.

Chattopadhyay, P., Glick, W.H., & Huber, G.P. (2001). Organizational actions in response to threats and opportunities, *Academy of Management Journal*, 44(5), 937–955.

Chen, F., Drezner, Z., Ryan, J.K., & Simchi-Levi, D. (1998). The bullwhip effect: Managerial insights on the impact of forecasting and information on the variability in a supply chain, Chapter 14 in Quantitative Models for Supply Chain Management, *International Series In OR and Management Science*, 17, Kluwer Publications, Boston, MA, USA.

Chen, F., Drezner, Z., Ryan, J.K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times and information, *Management Science*, 46(3), 436–443.

Chen, F., Federgruen, A., & Zheng, Y-S. (2001). Coordination mechanisms for a distribution system with one supplier and multiple retailers, *Management Science*, 47, 693–708.

Corbett, C. J., & de Groote, X. (2000). A supplier's optimal quantity discount policy under asymmetric information, *Management Science*, 46, 440–450.

Crotts, J., Aziz, A., & Raschid, A. (1998). Antecedents of supplier's commitment to wholesale buyers in the international travel trade, *Tourism Management*, 19(2), 127–134.

Damito, J., Hayes, G., & Kintele, P. (2000). Integrating ABC and ABM at Dow Chemical, *Management Accounting Quarterly*, 22–26.

Dong, Y., Carter, C.R., & Dresner, M.E. (2001). JIT purchasing and performance: An exploratory analysis of buyer and supplier perspectives, *Journal of Operations Management*, 19(4), 471–483.

Edwards, P., Peters, M., & Sarman, G. (2001). The effectiveness of information systems in supporting the extended supply chain, *Journal of Business Logistics*, 1–28.

Fawcett, S.E., & Magnan, G.M. (2002). Achieving world-class supply chain alignment: Benefits, barriers, and bridges, from http://www.capsresearch.org/completed.htm.

Frohlich, M.T., & Westbrook, R. (2001). Arcs of integration: An international study of supply chain strategies, *Journal of Operations Management*, 19, 185–200.

Ganeshan, R., Boone, T., & Stenger, A.J. (2001). The impact of inventory and flow planning parameters on supply chain performance: An exploratory study, *International Journal of Production Economics*, 71, 111–118.

Gavirneri, S., Kapuscinski, R., & Shindhar, T. (1999). The value of information in capacitated supply chains, *Management Science*, 45, 16–24.

Gebauer, J., & Buxmann, P. (2000). Assessing the value of interorganizational systems to support business transactions, *International Journal of Electronic Commerce*, 4(4), 61–82.

Gregor, S., & Johnston, R.J. (2000). Developing an understanding of interorganizational systems: Arguments for multi-level analysis and structuration theory, *Proceedings of the Eighth European Conference on Information Systems*, Vienna, Austria, 575–582.

Gupta, M., & Galloway, K. (2001). Activity-based costing management and its implications for operations management, *Technovation*, 1–8.

Hannan, M.T., & Freeman, J. (1977). The population ecology of organizations, *American Journal of Sociology*, 82, 929–964.

Harrold, D. (2001). How manufacturing benefits by understanding ERP and IT, *Control Engineering*, 48, 26–36.

Hoffmann, C. (2000). Supplier's pricing policy in a just-in-time environment, *Computers and Operations Research*, 27, 1357–1373.

Hong, K.K., & Kim, Y.G. (2002). The critical factors of ERP implementation: An organizational fit perspective, *Information and Management*, 40, 25–40.

Kelle, P., & Miller, P.A. (1998). Transition to just-in-time purchasing—handling uncertain deliveries with vendor—purchaser co-operation, *International Journal of Operations and Production Management*, 18(1), 53–65.

Kelle, P., & Milne, A. (1999). The effect of (s, S) ordering policy on the supply chain, *International Journal of Production Economics*, 59, 113–122.

Kelle, P., & Miller, P.A. (2000). Partnership and negotiation support by joint optimal ordering/setup policies for JIT, Presentation, *International Symposium on Inventories*, Budapest, Hungary.

Kelle, P., Pawlowski, S., & Akbulut, A. (2001). The influence of quantitative and organizational factors on the cooperation in supply chain, *Proceedings of the Conference on Production Economics*, Igls, Austria.

Landeghem, H.V., & Vanmaele, H. (2002). Robust planning: A new paradigm for demand chain planning, *Journal of Operations Management*, 319, 1–15.

Lau, H., & Zhao, L. (1993). Optimal ordering policies with two suppliers when lead times and demands are all stochastic, *European Journal of Operational Research*, 68, 120–133.

Lee, H., Padmanabhan, P., & Whang, S. (1997). The paralyzing curse of the bullwhip effect in a supply chain, *Sloan Management Review*, 93–102.

Lee, H.L., So, K.C., & Tang, C.S. (2000). The value of information sharing in a two-level supply chain, *Management Science* 46, 626–643.

Lee, H., & Whang, S. (2000). Information sharing in a supply chain, *International Journal of Manufacturing Technology and Management*, 79–93.

Li, D., & O'Brien, C. (2001). A quantitative analysis of relationships between product types and supply chain strategies, *International Journal of Production Economics*, 73, 29–39.

Moss-Kanter, R. (1994). Collaborative advantage: The art of alliances, *Harvard Business Review*, 72(4), 96–108.

Nahmias, S., & Smith, S.A. (1994). Optimizing inventory levels in a two-echelon system with partial lost sales, *Management Science*, 40(5), 582–596.

Pfeffer, J., & Salancik, G.R. (1978). *The external control of organizations: A resource dependency perspective*, Harper & Row, New York, USA.

Player, S., & Kramer, M. (1995). Using ABC to increase revenues, In: Player, S., & Keys, D.E., (Eds.), Activity Based Management: Arthur Andersen's Lessons from the ABM Battle-field, *Mastermedia Limited*, New York.

Ramasesh, R.V., Ord, J.K., & Hayya, J.C. (1993). Dual sourcing with nonidentical suppliers, *Naval Research Logistics*, 40, 279–288.

Schneider, H., Rinks, D. B., & Kelle, P. (1995). Power approximations for a two-echelon inventory system using service levels, *Production and Operations Management 4(4)*, pp. 381-400.

Scott, W.R. (1995). *Institutions and Organizations*, Sage Publications, Thousand Oaks, CA, USA.

Scott, J.E., & Kaindl, L. (2000). Enhancing functionality in an enterprise software package, *Information and Management*, 37, 111–122.

Seth, A., & Thomas, H. (1994). Theories of the firm: Implications for strategy research, *Journal of Management Studies*, 31(2), 165–191.

Shore, B. (2001). Information sharing in global supply chain systems, *Journal of Global Information Technology Management*, 4(3), 27–50.

Soroor, J. (2005a). Application of Intelligent WW Services to Mobile Real-time Coordination in Supply Chains, *4th IASTED International Conference on Communications, Internet and Information Technology (CIIT 2005)*, Cambridge, USA.

Soroor, J. (2005b). Coordination Mechanisms in Multi Agent Supply Chains for Multi-Agent Knowledge Management, *IEEE International Conference on Computational Intelligence for Modeling, Control and Automation (CIMCA'2005)*, Vienna, Austria.

Tarokh, M.J., & Soroor, J. (2005a). Supply Chain Management Information Systems Critical Failure Factors, *European Management & Technology Conference of Technology Research Institute of Florida*, Rome, Italy.

Tarokh, M.J., & Soroor, J. (2005b). Supply Chain Management Information Systems Critical Failure Factors, *the European Management and Technology Conference*, Rome, Italy.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of imbeddedness, *Administrative Science Quarterly*, 42, 35–67.

Walsh, J.P. (1988). Selectivity and selective perception: An investigation of managers' belief structures and information processing, *Academy of Management Journal*, 31, 873–896.

Wilson, E.J., & Nielson, C.C. (2001). Cooperation and continuity in strategic business relationships, *Journal of Business-to-Business Marketing*, 8(1), 1–24.

Wong, A. (1999). Partnering through cooperative goals in supply chain relationships, *Total Quality Management*, 10(4/5), 786–792.

Xu, K., Dong, Y., & Evers, P.T. (2001). Towards Better Coordination of the Supply Chain, *Transportation Research*, Part E 37, 35–54.

Yao, Y., Evers, P., & Dresner, M. (2001). Value of information sharing in vendor-managed inventory, *Proceedings of the Seventh America Conference on Information Systems*, 670–676.

Developing ERP performance indicators on the ERP implementation objective view

Chun-Chin Wei, Kuo-Liang Lee
Department of Industrial engineering and management,
Ching-Yun University, Chung Li City, 320, Taiwan

Abstract

Enterprise Resource Planning (ERP) system is a long-term investment for a company and affects significantly its competitiveness. This study presents a performance indicator development framework for assessing the performance of the adopted ERP system on the ERP implementation objective view. The framework can systematically identify the appropriate performance indicators based on the knowledge of ERP implementation objectives and set up consistent measurement standards for facilitating the complex ERP performance evaluation process. The evaluating results can represent the achievement of the objectives and the directions for improving the adopted ERP system. A real case in Taiwan is illustrated to demonstrate the feasibility of the proposed framework.

Keywords:

Enterprise Resource Planning (ERP), performance indicators, performance measurement

1. INTRODUCTION

Owing to the highly severe market competition and the immense impact of advances in information technology (IT) progress, a number of companies have widely implemented the Enterprise Resource Planning (ERP) systems. A comprehensive ERP implementation project should involve selecting an appropriate ERP software system and a cooperative vendor, implementing the selected ERP system, managing business processes change (BPC), and examining the practicality of the adopted ERP system (Wei and Wang, 2004). However, most of companies never select their ERP systems carefully, nor evaluate the performance of their adopted ERP systems systematically. Unfortunately, information managers are often swamped by the diverse requirements of users, instead of evaluating the advantages and drawbacks of the ERP system and further improving its performance.

"How to assess the success of the adopted ERP system" is still an important and difficult issue. There are many academic and practical researches to introduce the ERP or other IT performance measurement methods. Traditionally, the financial analysis is a conventional means of evaluating the success of IT systems, like return on assets (ROA), return on investment (ROI), and Cost-Benefit analysis (Hunton et al., 2003; Murphy and Simon, 2001). However, the financial analysis may seldom suffice, ignoring other critical qualitative performance indicator aspects, like system quality, and the impact on the organization and the individual. Instead, some subjective and surrogate performance indicators need to be developed.

The system and data quality assessment of IT systems have been widely studied (Lee et al., 2002; Palvia et al., 2001). The quality measurement reflects the engineering-oriented performance characteristics of the system itself and the quality of information and data. Some critical performance indicators were proposed, like data accuracy, timeliness, reliability, and completeness and IT system flexibility, integration, response time, and ease of use and learning.

Additionally, a more popular term, "User satisfaction", is used to measure the quality of IT or data. The discrepancy between a user's expectations before and after using the IT system consists of the user's satisfaction evaluation of such utility. Related studies evaluated the IT performance using the experience and perspective of various users, like employees, middle managers, top managers and system engineers (McHaney et al., 2002; Doll et al., 1994). However, in real life it is difficult for users to constitute the so-called pre-usage expectation due to lack of information or historical experience as a reference framework.

DeLone and McLean (1992) surveyed 180 articles from seven publications to categorize IT performance indicators into six major dimensions, including system quality, information quality, use, user satisfaction, individual impact, and organization impact. Their model indicated that the system and data quality would affect the system use and user satisfaction. Then, user satisfaction could cause to the impact of individual and organization. This study offers many critical directions to the later IT performance research efforts (Heo and Han, 2003; Choe, 1996).

However, the most frequently adopted performance indicators are to refer to the common indices without developing tailor-made measures that echo the objectives of ERP implementation. Success is often defined as a favourable or satisfactory result or outcome. The objectives of ERP implementation describe exactly the states of an ERP system which the managers wish to realize. Then, the explicit definition of “the success of an ERP system” can be defined that the adopted ERP system really achieves the objectives that the managers strive. That is, the methodologies for evaluating the practicality of the adopted ERP system should be based on the knowledge of these objectives. Hence, a systematic ERP performance indicator development methodology is necessary to align the system’s performance indicators with the ERP implementation objectives and to highlight the effectiveness of the system.

This study presents an ERP performance indicator development framework to extract suitable performance indicators on the ERP implementation objective view. This paper will focus on the ERP performance indicator development process, but not on the ERP performance assessment skills. Then, the performance indicators will be aligned with the ERP implementation goals and business strategies.

2. METHODOLOGY

The ERP implementation project team should identify some critical ERP implementation objectives at the ERP implementation stage, please refer to (Wei et al., 2005). First, the team members should discuss the strategic objectives of the ERP implementation project according to their ERP implementation reasons, business environment, and industrial characteristics. Then, decompose these strategic objectives into two kinds of detailed objectives, including fundamental-objectives and means-objectives. Fundamental-objectives are those that are important because they reflect what the decision makers really want to accomplish and are organized into a hierarchy. Meanwhile, means-objectives are those which help the fulfilment of fundamental-objectives and are organized into networks. The fundamental-objectives can extract to the proper attributes for selecting an ERP system. Additionally, the means-objectives can develop the guidelines and standards of group decision-making. After implementing the selected ERP system, we should understand which objectives have been achieved and how the ERP system performs.

A stepwise procedure is first described to present the proposed ERP system performance indicator development framework clearly:

Step 1. Form an ERP performance indicator development team.

Step 2. Identify the critical directions of ERP performance evaluation.

Step 3. Choose some suitable performance indicators from the means-objectives.

Step 4. Add other crucial performance indicators.

Step 5. Construct the proper performance indicator structure.

Step 6. Develop the detailed assessment content of each performance indicator.

The procedure is illustrated in Figure 1 and elaborated as follows.

2.1 Form an ERP performance indicator development team

First, an ERP performance indicator development team must be formed. This team should involve some crucial members, including senior managers, information managers and experts, representatives of user departments, even external ERP consultants. They are responsible to formulate an ERP performance indicator development plan, identify appropriate performance indicators, and develop consistent evaluation guidance of each performance indicator.

2.2 Identify the critical directions of ERP performance evaluation

A company formulates the objectives of ERP implementation to select an appropriate ERP system and vendor at the ERP implementation phase. Therefore, these objectives generally indicate the direction in which the managers should strive to do better. In the ERP implementation objective structure, the fundamental-objectives are those that are important to specify the goal of the ERP implementation project. The objectives highlight why the managers are concerned with the selection and what criteria the managers should be reviewing [10]. On the other hand, the means-objectives specify how to accomplish the desired fundamental-objectives. Having sorted them, the members can evaluate various alternatives whose conditions are consistent with the company’s concerns (Wei et al., 2005).

According the definitions of the two kinds of objectives, this study find that the fundamental-objectives describe some critical directions of ERP performance evaluation. Based on the fundamental-objective classification in Wei et al. (2005), they can be classified into three dimensions, including project management factors, software system factors, and vendor factors. Additionally, the fundamental-objectives would be constructed into a hierarchy. Then, this fundamental-objective hierarchy can be utilized to express the crucial direction of ERP performance evaluation. The three dimensions are introduced as below.

(1) Project management factors: attributes involved in project management, such as total cost and time of implementation;

(2) Software system factors: features of the software and system, including functionality, reliability, user friendliness, and flexibility of the ERP system;

(3) Vendor factors: indicators for assessing the performance of the ERP vendor, such as technology capability and service.

2.3 Choose some suitable performance indicators from the means-objectives

After identifying the direction of ERP performance evaluation, the team members must select proper ERP performance indicators. A good performance indicator can clearly express whether the corresponding implementation objective has been attained and involve sufficient information about this objective. Practically, too detailed and many performance indicators would cause the performance measurement process too complex and inefficient. According to the definition of means-objective, some means-objectives themselves are suitable to be performance indicators to evaluate whether the fundamental-objectives have been accomplished as promised. Especially, some first tier means-objectives in the means-objective network are very adequate performance indicators.

This study proposes a systematic discussion method to select suitable ERP performance indicators. The team members can start with a nearest means-objective of a bottom level fundamental-objective in the objective structure to discuss, "Whether the means-objective can be used as a suitable ERP performance indicator?". After such as discussion, if the means-objective is a suitable ERP performance indicator, then it is added to the performance indicator set. If it is not suitable, the members can further discuss, "How to evaluate whether this means-objective has been achieved?". The answers can reveal more detailed and new ERP performance indicators, which can be incorporated into the set. The process is repeated with all means-objectives in the ERP objective structure. Figure 2 displays the systematic discussion. Put the performance indicators into the original ERP performance indicator set.

2.4 Add other crucial performance indicators

The performance indicators in the original ERP performance indicator set are extracted from the objective structure. However, it is not sufficient for assessing the ERP performance only on an input view. The ERP performance assessment should involve many important outputs of the adopted ERP system. The team members should survey other proper performance indicators based on the output aspects of ERP system execution from the relevant literatures or their concernment, like the impact of organization and individuals. Then, add these performance indicators into the ERP performance indicator set.

By the same reasons, some performance indicators extracted from the implementation objectives should be modified on an ERP output view.

2.5 Construct the proper performance indicator structure

As mentioned above, the fundamental-objectives point out the crucial direction of ERP system performance assessment. Some means-objectives can be selected as suitable ERP performance indicators. Then, the team members should construct the performance indicators into a systematic structure. Organizing the performance indicators means organizing them so that they describe in detail what the managers want to achieve and can be incorporated in a proper method into the performance evaluation model. According the knowledge of the fundamental-objective hierarchy, this study categorizes the performance indicators to be compatible with the ERP objective structure (Wei et al., 2005), including project management factors, software system factors, and vendor factors. Additionally, we should add an important aspect, "impact factors". Then, these four main dimensions construct the ERP performance indicator hierarchy.

The team members can review the performance indicators in the performance indicator set and place them into the four dimensions. After specifying the ERP performance indicator hierarchy, the members may find themselves refining the content and establishing the performance evaluation directions. The performance indicators at the lowest level are specific and operational in practice and are perceived as the means to achieve those higher objectives.

2.6 Develop the detailed assessment content of each performance indicator

After organizing the ERP performance indicator hierarchy, the project team must discuss the evaluation content and method of each performance indicator. The managers should investigate what types of data they need to collect and how to collect the data for evaluating each indicator. Especially, the team members should pay attention to the accessibility and reliability of data which need to be collected. A standard form can help them to collect the data and conduct the performance assessment. Knowledge of the objective structure can not only help in identifying the performance indicators, but also beyond that the objectives indicate how outcomes must be measured and what kinds of uncertainties should be considered in the evaluation.

Some performance indicators are quantitative and are easy to assess with exact numbers. However, many performance indicators are not. The units and assessment models of performance indicators need to be identified first. The evaluation results of quantitative and qualitative performance indicators should be aggregated into a score to show the ERP performance. The analytic hierarchy process (AHP), fuzzy set theory, or other scoring and ranking methods can be applied.

3. A REAL CASE

The proposed ERP performance indicator development framework was applied to measure the performance of the ERP system at an electronics company in Science-Based Industrial Park in Taiwan. After adopting the ERP system, the information managers hoped to know how the ERP system is currently performing and how it should be performing at a future point in time. Additionally, they want to justify the success and the value-added contribution of the ERP system to accomplish the objectives set at the ERP system implementation stage. The stepwise procedure is presented in the following.

Step 1. An ERP performance indicator development team with 6 team members was formed, including 3 major managers, the information manager, and 2 user representatives. The major managers were the Marketing, Manufacturing, and Financial manger. All team members participated in the ERP system implementation project and selected their ERP system.

Step 2 and 3. The objectives of ERP implementation have been established in the ERP selection phase (Wei et al., 2005). According to the fundamental-objective hierarchy, the team members chose four main direction of ERP performance assessment, that is, the project management, software system, vendor, and impact factors.

The means-objectives in the objective structure were discussed whether they were suitable ERP performance indicators. For example, the means-objectives of a fundamental-objective “system stability” were “system maturity”, “few bugs”, and “minimum of system break down”. Following the systematic discussion method as mentioned above, the members discussed, “Whether system maturity, few bugs, and minimum of system break down can be used as suitable performance indicators of system stability?” The result of this examination derived core contents of the original performance indicator set. Initially, the original set of 39 performance indicators was extracted from relevant means-objective network.

Step 4. Some additional performance indicators surveyed from prior IT performance literatures were recommended to compensate the original performance indicator set. After examining the necessity of these indicators with the panel members, 23 performance indicators were incorporated into the original performance indicator set, increasing the number of performance indicators to 62.

Step 5. The performance indicator set must be refined. After discussing, some performance indicators were cancelled or combined. Finally, total 39 performance indicators were selected.

A hierarchy was constructed from the remaining performance indicators based on the four main performance indicator categories. The ERP performance indicator hierarchy comprised five levels, as illustrated in Figure 3. Level 1 reveals the ultimate mission for assessing the performance of the adopted ERP system. Level 2 consists of four main categories: project management, system, vendor, and impact factors. Level 3 contains the major fundamental-objectives that the managers hoped to achieve. Level 4 describes the basic fundamental-objectives. The bottom level comprises the associated PIs that were used to measure the performance of the adopted ERP system.

Step 6. Then, the team members discussed how to evaluate the performance indicators and how to collect the data of the ERP system performed. Using the knowledge of ERP implementation objectives, we assisted the team members to establish the specific data requirements of each performance indicator. The detailed evaluation guidance and an assessment questionnaire or form for each performance indicator was also developed. This process can ensure that everyone who undertakes the ERP performance measurement can follows the same criteria in the performance measurement process consistently.

After the detailed contents of performance indicators identifying, the team members can undertake the performance measurement according to the current status of their ERP system.

4. CONCLUSION

ERP systems are seen to be effective in tying the business functional units with the various organizational information systems and their associated databases, which in the end can support the strategic aims of modern organizations. Many companies are rushing to implement ERP systems under the competitive pressure unleashed by the processes of globalization. All of the modern technology in the world is worthless if the adopted ERP system can't attain the expected objectives which managers want to strive. However, if we do not have a standard measure of ERP implementation success, it is difficult to expect practitioners to undertake such evaluations. This study proposed an ERP performance indicator development framework based on the ERP implementation objectives. The managers can easily follow the steps of the proposed framework to develop the suitable performance indicators for assessing their ERP system.

The ERP performance indicator development framework can establish a feedback mechanism between the desired objectives of ERP adoption and the substantial effects of ERP execution. The proposed procedure expands the structure of fundamental-objectives and means-objectives, which had been constructed in the ERP system selection methodology to identify the hierarchy of ERP performance indicators and formulate the detailed performance evaluation guidance. Then, the evaluation results can truly reflect the current situation of the adopted ERP system and the accomplishment of the expected objectives.

REFERENCES

[1] Wei C. C., Wang M. J. J., A comprehensive framework for selecting an ERP system. International Journal of Project Management 2004; 22: 161-169.

[2] Hunton J. E., Lippincott B., Reck J. L.,Enterprise resource planning systems: comparing firm performance of adopters and nonadopters. International Journal of Accounting Information Systems 2003; 4: 165–184.

[3] Murphy K. E., Simon S. J., Using cost benefit analysis for Enterprise Resource Planning project evaluation: A case for including intangibles. Proceedings of the 34th Hawaii International Conference on System Sciences, 2001; 1-11.

[4] Lee Y. W., Strong D. M. , Kahn B. K., Wang R. Y., AIMQ: a methodology for information quality assessment. Information & Management 2002; 40: 133–146.

[5] Palvia S. C., Sharma R. S., Conrath D. W., A socio-technical framework for quality assessment of computer information systems. Industrial Management & Data Systems 2001; 101 (5): 237-251.

[6] McHaney R., Hightower R., Pearson J., A validation of the end-user computing satisfaction instrument in Taiwan. Information & Management 2002; 39: 503-511.

[7] Doll W. J., Xia W., Torkzadeh G., A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Quarterly 1994; 18(4): 453-461.

[8] Delone W. H., McLean E. R., Information systems success: the quest for the dependent variable. Information Systems Research 1992; 3: 60-95.

[9] Heo J., Han I., Performance measure of information systems (IS) in evolving computing environments: an empirical investigation. Information & Management 2003; 40: 243-256.

[10] Wei C. C., Chien C. F., Wang M. J. J., An AHP-based approach to ERP system selection. International Journal of Production Economics 2005; 96: 47-62.

[11] Choe J. M., The relationships among performance of accounting information systems, influence factors, and evolution level of information systems. Journal of Management Information Systems 1996; 12(4): 215-239.

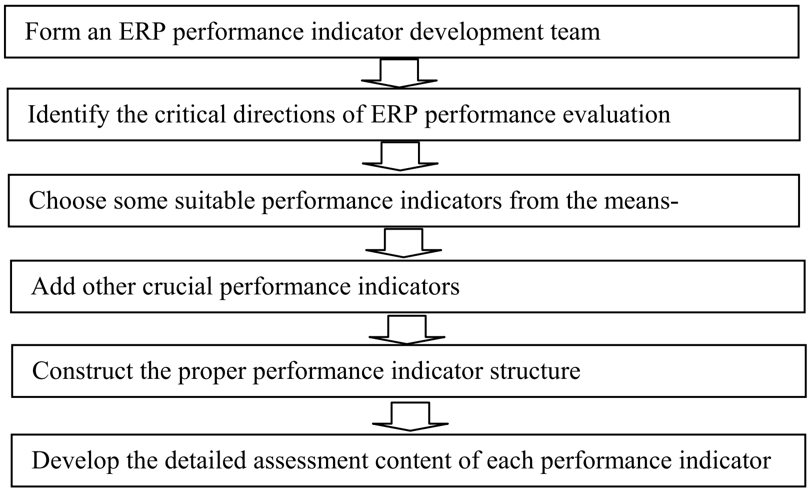

Form an ERP performance indicator development team
Identify the critical directions of ERP performance evaluation
Choose some suitable performance indicators from the means-
Add other crucial performance indicators
Construct the proper performance indicator structure
Develop the detailed assessment content of each performance indicator

Figure 1. ERP performance indicator development framework

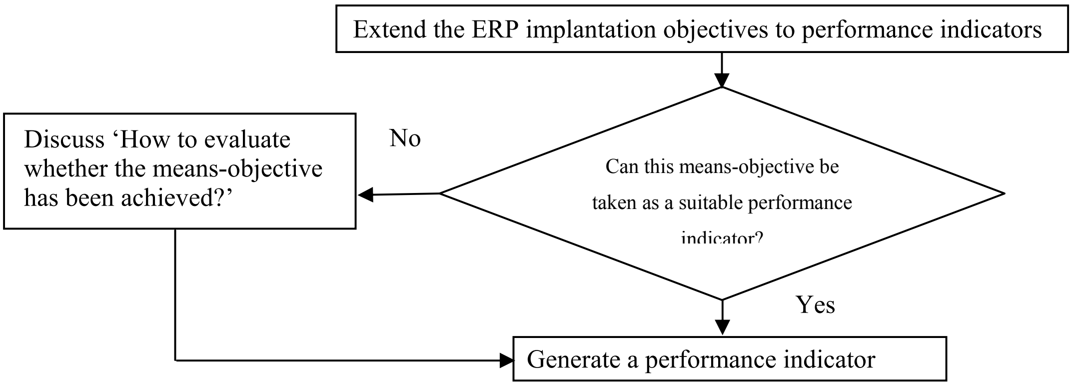
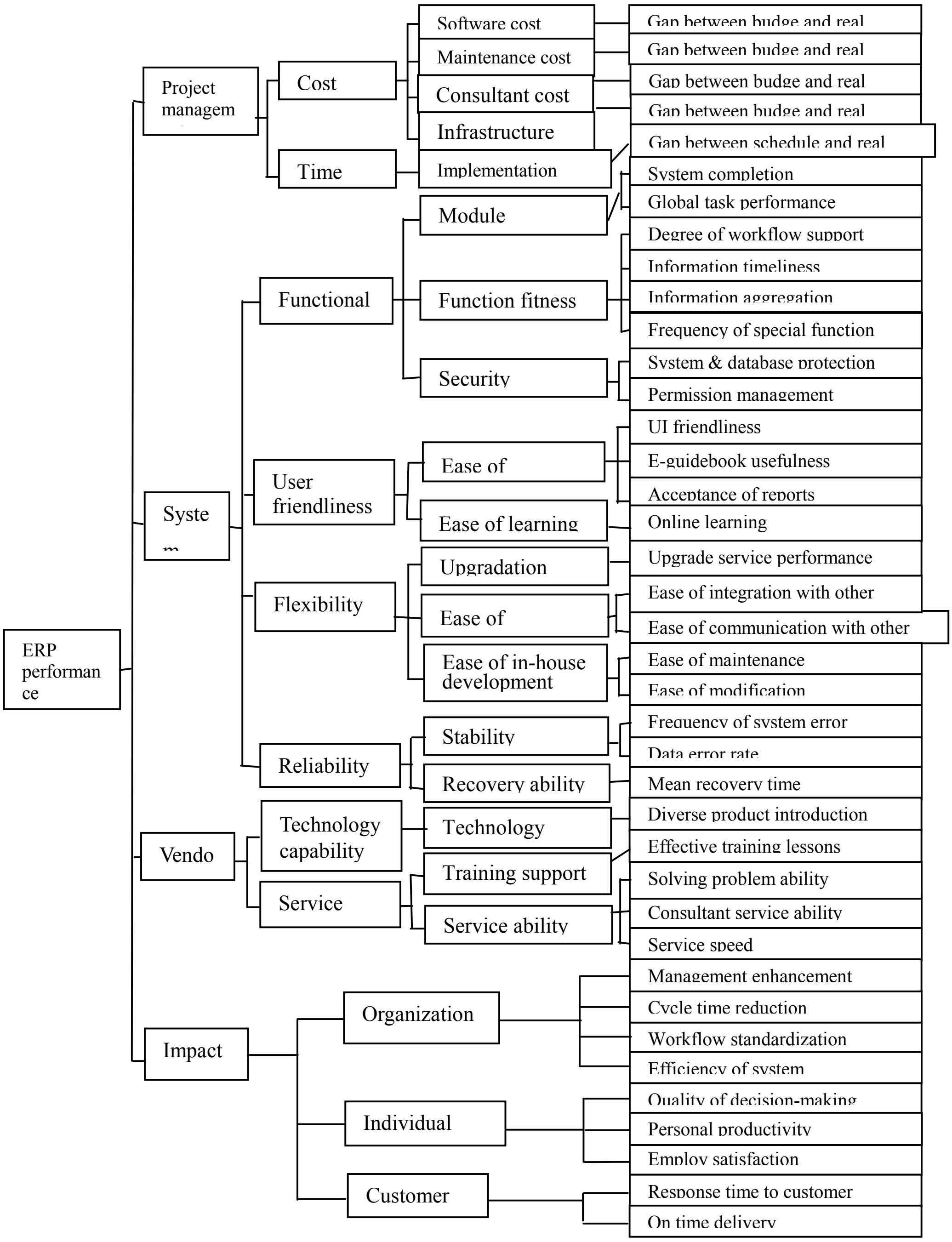


Figure 2. Systematic performance indicator discussion



Rethinking User Acceptance: An Examination of User Resistance in Mandatory Adoption of Enterprise Systems

Tim Klaus, ABD
Texas A&M-Corpus Christi
J. Ellis Blanton, PhD
University of South Florida

Abstract
User resistance has been found to be an important issue in the implementation of an Enterprise System (ES). However, despite the prevalence of user acceptance literature, user resistance literature is scarce. Although some studies have conceptualized user resistance as the opposite of user acceptance, a mandatory, role-transforming system such as an ES can clearly demonstrate that users may adopt or accept a system while resisting it. Although this relatively unstudied area is highly relevant, it is theoretically underdeveloped. This study examines user resistance at the individual level of analysis to determine the underlying reasons for user resistance and the types of resistant behaviors.

Keywords:
User Resistance, ERP Systems, Enterprise Systems, Mandatory Adoption, Resistant Behaviors

INTRODUCTION

Enterprise Systems (ESs) are software packages used for integrating and managing business processes across organizational activities and are widely deployed in organizations from numerous industries. ESs refer to commercial software packages that enable the integration of business processes and transaction-oriented data throughout an organization (Markus et al. 2003). They include organizational-wide software such as Enterprise Resource Planning (ERP) systems, scheduling, customer relationship management, product configuration, and sales force automation (Markus et al. 2000). Not only are an increasing number of organizations installing full ESs or ES modules, but also organizations currently using ESs are expanding their use. AMR Research estimates that the ERP market alone will grow to 31.4 billion by 2006 at a rate of 10 percent annually (Surmacz 2002).

ESs have gained credibility as their widespread implementations have led to the creation of more stable and adaptable systems and improved management tools. Through removing inefficiencies in business processes, ESs have led many organizations to greater profitability. For example, Hitt, Wu, and Zhou (2002) finds that financial markets have consistently rewarded adopters through an increased market valuation. ES implementations are important as they are pervasive, ongoing and require a fair amount of compliance as well as job transformation. These large-scale projects require changes that upset the status quo of individuals in the organization. Successful implementations remain a daunting issue as numerous articles report on implementation catastrophes as well as implementations that have failed to provide projected benefits (i.e., Bingi et al. 1999; Robey et al. 2002). One important reason for this is user resistance (Jiang et al. 2000).

User resistance is an important issue in ES implementations and has been said to be "at the root of many enterprise software project failures" (Hill 2002). For example, Callahan (2002) finds a significant amount of user resistance even after nine months of ERP integration testing, partly due to the many interfaces with existing systems. Maurer (2002) states that the reason for low ES return on investments is because of resistance. Hines (2002) notes that since end user resistance often is cited as an important cause of organizations failing to achieve projected benefits, PeopleSoft, an ES vendor, purposely made user-related improvements in version 8.8. Furthermore, Cooke et al., 1998, a report on 186 companies that implemented the SAP ES finds that resistance was the second most important contributor to time and budget overruns and was the fourth most important barrier to SAP implementation. Other studies have revealed how users' resistance can cause ES implementation failures (Barker et al. 2003; Krasner 2000; Robey et al. 2002; Umble et al. 2002; Wah 2000).

Although user resistance is an important issue, especially in ES implementations, Marakas and Hornik (1996) points out that "few theoretical foundations currently exist in the literature for explaining user resistance" (p. 209). Although there are some studies (i.e., Jiang et al. 2000; Shang et al. 2004) which describe user resistance, the concept of user resistance still lacks a theoretical underpinning as to its cause. Research literature has focused much more on user acceptance rather than user resistance. This is understandable as many types of systems or technologies have voluntary acceptance and thus user resistance is not an issue. Unfortunately for ES research, user acceptance literature fails to account for the mandatory and job transforming nature of ES implementations.

Although there is a lack of theoretical foundations, user resistance remains an important and relevant issue faced by numerous organizations. User resistance must be reduced in order to reap efficiency benefits, particularly for systems that transform business processes such as ESs. As an ES is used to transform an organization by fundamentally changing business processes, user resistance can greatly affect an ES implementation. Understanding the causes of user resistance could lead to improved implementation strategies and results.

NATURE OF ES IMPLEMENTATIONS

One of the major benefits of ES implementation, as well as a catalyst for user resistance, is efficiency achievements through process reengineering. An ES is only a tool, yet as a master craftsman uses a chisel to carve a piece of wood, management can use an ES to chip off the inefficiencies from the organizational processes. Effective ES use not only changes organizational technologies, but through redesign, fundamental business processes are transformed. Cooper (2000) found that IT can be used as an effective reengineering tool, although the appropriate organizational climate is required that fosters creativity.

There are several key differences between an ES implementation and other types of system implementations. First, ESs require mandatory usage throughout all affected levels of the organization. Mandatory usage is necessary for the system to integrate the data and produce organizational snapshot and trend analysis reports. Second, an ES implementation generally results in the reengineering of jobs, often requiring changes in job tasks and reward structures. A clear benefit of and reason for ES implementation is the efficiency gains through process reengineering and thus these changes are made during the system implementation. Third, in order to minimize cost and time of future upgrades, standardized modules are only partially customized for employees as opposed to a full customization that may be performed for software produced in-house. Customization is only minimally performed since every upgrade that an ES software vendor delivers will also need to be customized, thus increasing both initial and future costs. Due to these increased costs, managers are discouraged from making modifications unless they are absolutely necessary.

Because of the three contextual differences noted above, the end-user's perceived usefulness and perceived ease of use is not a priority; rather, the goal of implementation is to achieve efficiencies through reengineered processes and provide better organizational reports to managers for improved decision-making. This inherent nature of ES that revolves around the business processes, not the user, and can both breed and proliferate resistance. A business process has been defined as "a set of logically related tasks performed to achieve a defined business outcome" (Wu 2003, p. 2). Business process reengineering received much attention around 1990 (Hammer 1990; Hammer et al. 1993) by both systems and business people and has been defined as "The fundamental rethinking and radical design of business processes to achieve dramatic improvement in critical, contemporary measures of performance such as cost, quality, service, and speed" (Maurer 2002, p. 2). Reengineering can entail eliminating or transforming organizational processes and change the way transactions are performed with suppliers and customers. ESs are not needed for reengineering to occur, but one main benefit of an ES is the process reengineering that occurs as the technology is implemented.

There are several contributions of this study. First, the ES implementation is examined from a user resistance perspective; as user resistance is a reason why a technology is not adopted, this research may modify the current understanding of the user acceptance literature. A second contribution is to understand what motivates users to resist an ES. In spite of the recent increases in the number of ES publications, there is not a compelling theoretical explanation in describing the phenomenon user resistance and its underlying causes. A third contribution is the understanding of how user resistance manifests itself through attitudes and behaviors. Although some studies have suggested ways users may resist a system, this study looks specifically at ES implementations, and the types of attitudes and behaviors that are exhibited by users.

RESEARCH QUESTIONS AND CONCEPTUAL MODEL

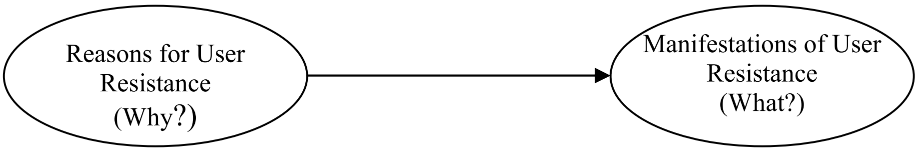
Because of the reports that address the failure of ES implementations and the importance of minimizing user resistance (i.e., Callahan 2002; Hill 2002; Krasner 2000; Maurer 2002), it seems important to better understand the nature of user resistance. In previous paragraphs, the context of ES implementations has been described. As user resistance is an issue affecting most ES implementations, it is important to examine this area. All resistance does not hurt an organization and there are valid reasons why users both passively and actively resist an implementation (Keen 1981); in fact, Fiorelli and Margolis (1993) argues that some level of resistance can be beneficial to the organization as it may draw attention to problems in the change and issues that need to be addressed. There is a difference between resisting a system that an employee believes will make the organization worse off and resisting due to selfish ambitions; however, in either case, the employee is resisting and thus hindering the implementation of the system. Whether or not resistance is beneficial in specific cases, it is still an important phenomenon that must be addressed so that proposed changes can be either effectively implemented or modified. Understanding the issues of why users resist can help in identifying important underlying issues that will ultimately help to bring a greater degree of long-term success for the company. Following are the first two research questions:

1) What motivates users to resist an ES implementation?

2) How does user resistance manifest itself in an ES implementation?

The two research questions all revolve around user resistance in an ES implementation and are identified in the study's conceptual model illustrated in Figure 1:

Figure 1. Conceptual Model



METHODOLOGY

Qualitative research methods enable researchers to examine social and cultural issues through the use of interviews, observation, questionnaires, manuscripts, and researcher's impressions (Lee 2002). Case methodology is useful when a natural setting is required and researchers can use multiple data collection methods that converge to support research findings (Benbasat et al. 2002). For this study, interviews, archival records, and documents, such as emails and training materials, were collected and analyzed.

Data was collected in three steps from three sources: an expert panel, an in-depth case study of a large University, and two case studies in an Asian airline and a cellular company. The goal of the first step of the expert panel was to understand the major issues that arise in an ES implementation. For the expert panel, a focus group of seven IT professionals that have been involved with ESs was used to extract their perspectives on the reasons for resistance, the resistant behaviors and the management strategies to minimize resistance. This focus group was recorded and lasted over an hour. Also, an expert was interviewed who was a CIO of a Fortune 500 organization that has led the rollout of several ESs. This interview lasted approximately one hour. Both the focus group and the expert interview recordings were transcribed and analyzed to extract the major principles and concepts.

The second step was an in-depth case analysis. A large public university was selected that is located in the southeastern United States. With approximately 40,000 students and 10,000 employees, the rollout of the ES was a major change, affecting many people and departments. The selected system contained nine modules: Purchasing/Procure to Pay/Order to Cash, Grants, Accounts Payable, Asset Management, Accounts Receivable/Billing, Budgets, General Ledger, Project Costing, and Record to Report. The primary data collection method was through semi-structured interviews. Walsham (2002) warns that data may lose its richness if the interviewing style of the interviewer is over-directing the interview through tight controls. Therefore, when questions were asked during the interviews, the researcher tried to listen well while conveying a non-judgmental attitude.

The data collection at the primary organization continued until a point of theoretical saturation, which is when the value of an additional interview was considered to be negligible (Eisenhardt 1989). There were 22 people interviewed from all levels of the organization, including clerical staff, IT professionals, trainers, managers, and superusers. The average interview length was 47 minutes, ranging from 25 to 77 minutes. The interviews were recorded and transcribed in order to acquire a full description of the interviewee's comments, yielding 242 pages of single-spaced transcripts (135,200 words).

Techniques were used from grounded theory (Glaser et al. 1967) in an attempt to derive basic concepts and structures among the concepts. All the sentences from the transcripts were first all marked whether or not they had any direct relevance to the areas under investigation (there were a number of paragraphs that provided extra information such as backgrounds on the individual or system, but did not relate directly to any of the research questions). Each relevant sentence (and context, if useful in understanding the sentence) was then extracted to a separate document, from which they were all analyzed to identify themes. As the researcher progressed through the transcripts, there was a need to refine the emerging themes.

Each interview transcript was analyzed in depth. The major concepts identified in each transcript were highlighted and put into separate documents. Statements related to reasons for user resistance, resistance behaviors, and management strategies to minimize resistance were extracted for further analysis. After extracting the statements, the statements were analyzed and classified into themes that emerged among the reasons, the behaviors, and the management strategies. After the initial identification of themes, there were three iterative rounds of analyzing the themes, then reclassifying statements according to what emerging themes improved the classification. This essentially followed the hermeneutic process laid out by Klein and Myers (1999) which suggests an iterative process of reflecting on the interdependent meanings of the parts (individual statements) and the whole (evolving themes or conceptual framework).

As this research is exploratory in understanding the reasons for resistance, the behaviors that are manifested, and the management strategies to minimize resistance, an a priori list was not available. Thus the themes that emerged through the iterative process of analyzing the data and restructuring the themes each were assigned a code. In order to both check these codes/themes for reliability and definitional clarity (Miles et al. 1994), two graduate research assistants, taking part only in the coding and unfamiliar with the research, were used to read and code the transcripts. Both research assistants were provided with a one-page coding scheme that identified each code/theme and its operational definition. Each paragraph in the transcripts could be assigned zero, one, or multiple codes. The research assistants first examined one interview transcript, discussed discrepancies, and then continued to code the remaining transcripts. The Cohen's Kappa statistic was used to analyze the level of correspondence between the coders and was well above the 61% level that is suggested to have "substantial strength of agreement" (Landis et al. 1977, p. 165).

It has been suggested that "every organizational situation is likely to be filled with multiple and frequently conflicting interpretations and meanings" (Prasad 1993, p. 1404). Thus in a case study, it is important to establish construct validity. Construct validity is supported through the use of multiple sources and multiple data collection methods (Benbasat et al. 1987; Yin 2003). In regards to the multiple sources, statements made from one interviewee were compared and contrasted with statements made with other interviewees in order to triangulate the ideas suggested by the interviewees. Multiple data collection methods were also used, since the interviewer was given access to training manuals, emails, memos, and other written documentation concerning the project. Also, the employees were selected based on three criteria: 1) represent different departments; 2) represent different positions (i.e., manager, super-user, end-user); 3) they use (or used) the system regularly.

The third step was the use of two case studies to validate the findings of the first two steps. One case study was conducted with an airline located in Asia and the second case study was conducted with a cellular phone company located in the U.S. Since the second step encompassed users from all levels of the organization, this step focused only on IT professionals heavily involved in the rollout of an ES. Because of their widespread experiences, they were useful in validating the findings based on the first two steps. Phone interviews were conducted with 7 IT professionals for the airline company and with 4 IT professionals in the cellular phone company. All of these interviews were recorded, lasting an average of 40 minutes, ranging from 25 to 51 minutes. The recordings were all transcribed, yielding 106 single-spaced pages (47872 words).

The research assistants coded these transcripts as well, to establish validity among the identified constructs. The Cohen's Kappa statistic for the coding of the third step also was well above the 61% threshold. External validity is in regards to the generalizability of the study's findings (Yin 2003). This was established through the focus group of IT professionals that have been involved with ESs within different industries as well as using multiple case studies from organizations representing three different industries.

RESULTS AND IMPLICATIONS

Based on the iterative construct formation process, reasons for user resistance and resistant behaviors emerged. Tables 1 and 2 identify the emergent constructs. As noted in these tables, the constructs have also been placed into categories: for example, Table 1 categorizes the constructs into individual, system, organizational, and process issues. The use of categories were added to provide a better understanding of the types of reasons and behaviors. The classification of resistant behaviors found in ES implementations that is shown in Table 2 is based on a classification of behaviors proposed by Bovey et al. (2001a, p. 375) and Bovey et al. (2001b, p. 534).

Table 1. Reasons for User Resistance

	Construct	Definition	Examples
Individual Issue	Uncertainty	User is unclear of the future	Unknown future, potential threat, lack of clarity
	Lack of Input	User's opinions are not considered	The thoughts and opinions of users were not sought out
	Loss of Control/Power	User loses control or loss of recognition as the expert	Leveled playing field, not the expert anymore
	Self-Efficacy	Perceived lack of capability	Lack of confidence, lack of computer skills/abilities
System Issue	Technical Problems	Problems with the system	Bugs in system, features that don't work right
	Complexity	System is complicated to use	Difficult to access, Poor user interface that lacks logic or is not intuitive
Organizational Issue	Lack of Facilitating Environment	Organizational culture is not conducive to the change	Lack of technology usage in organization, bureaucracy that is slow to change
	Poor Communication	Communication to users is problematic	Lack of communication, users not hearing benefits of system, lack of coordination, users not understanding why
	Poor Training	Training does not meet organizational needs	Lack of training, training seems to be a waste of time, incompetent trainers, timing of training, sufficiency of training
Process Issue	Changed Job/Job Skills	User's job or job skill requirements changes	Revised job description, different job tasks, new skills, new way of thinking
	Additional Workload	User is required to put forth additional effort	Extra work, more work to get same info, extra time
	Lack of Fit	Process problem between the system and organizational structure	Problematic changes to processes, new processes not working as planned

Table 2. Manifestations of User Resistance

	Overt (clearly expressed)	**Covert** (minimally expressed)
Active (takes action)	• Refusal to use system • Challenge system/plan • Hack at system • Don't follow process • Quit job/job change	• Use shadow system • Try to use old system • Avoid system use • Enter in info inappropriately
Passive (inert)	• Complaints • Lower morale • Defensive • Turnover Intention	• Not Motivated • Less Productive • Impatient • Not paying attention • Procrastinate • Don't want to learn

Table 1 identifies the reasons for user resistance that were identified through this research. From the perspective of management, if the decision is made to rollout a system, it is very important to minimize the amount of user resistance. Through understanding the reasons for why users resist, strategies can be employed to minimize the resistance. For example, the first reason that is identified is Uncertainty. In order to minimize the user resistance that may stem from this reason, management can clearly convey to users the type of changes that will occur and communicate to them the change management plan. To minimize the user resistance stemming from the second reason, Lack of Input, management can set up user group meetings to obtain input from users who would like to contribute their ideas to the changes that potentially may occur. Likewise, with the rest of the reasons, management can utilize strategies that address these reasons for user resistance.

Markus (1983) is a well-cited publication on resistance, and identifies user attributes, technical attributes, power issues, and organizational attributes that affect the level of resistance. In particular, Markus (1983) focuses on power issues, discussing how the user resistance remained until the users felt compensated for the lost power. The results shown in Table 1 identifies and describes reasons for user resistance in all four of the areas identified by Markus (1983). Furthermore, it adds constructs that fit under process issues that Markus (1983) did not discuss. The changed job/job skills, the lack of fit between new processes and the system, and additional workload were all reasons for user resistance that were not discussed by Markus (1983).

Table 2 identifies potential manifestations of user resistance. There are several different categories of resistant behaviors that were found. Passive resistance tends to decrease productivity although passive forms of resistance still leads to system utilization as planned. Active resistance, tends to affect the organization to a greater degree, as it generally leads both reduced productivity as well as problems with the system. For example, users who do not follow the processes or enter in information inappropriately can cause problems within the system that negatively affect users who are not resisting. Managers who understand the types of resistant behaviors will most likely be better prepared to minimize these types of behaviors.

The implementation of ESs in organizations has forced many employees to adopt a system that changes their job duties and reward structure. These mandatory, role-transforming systems have faced considerable resistance, though often it is covert. There are obviously contributing factors that affect employee responses to a system, as mentioned earlier in this paper, such as politics (Markus 1983), top management support, and project team competence (Akkermans et al. 2002). However, even when the appropriate planning, analysis, and design has been performed, there are still many times implementations have failed or faced unwarranted difficulties because of user resistance. Furthermore, ES success does not just occur from a one-time implementation, but rather through on-going improvements (Kraemmergaard et al. 2002). Although these are interesting areas for future studies, this paper has provided a foundation on the examination of user resistance upon which other studies can build.

There are several contributions of this study. One contribution is an understanding of what motivates users to resist an ES. In spite of the recent increases in the number of ES publications, there is not a compelling explanation in describing the phenomenon of user resistance and its underlying causes. A second contribution is the understanding of how user resistance manifests itself through behaviors. Although some studies have suggested ways users may resist a system, this study looks specifically at ES implementations, and the types of behaviors that are exhibited by users. Through a qualitative analysis, specific resistant behaviors are identified, described, and set into a framework. Third, the ES implementation is examined from a user resistance perspective; as user resistance is a reason why a technology is not adopted, this research modifies the current understanding of the user acceptance literature. Although many studies have examined user acceptance, user resistance sometimes is considered the opposite of user acceptance. This study argues that user resistance is not the opposite of user acceptance and differentiates the two concepts, since user resistance can still occur, even when acceptance appears to have occurred.

REFERENCES

Akkermans, H., and Van Helden, K. "Vicious and Virtuous Cycles in ERP Implementation: A Case Study of Interrelations between Critical Success Factors," *European Journal of Information Systems* (11) 2002, pp 35-46.

Barker, T., and Frolick, M.N. "ERP Implementation Failure: A Case Study," *Information Systems Management* (20:4), Fall 2003, pp 43-49.

Benbasat, I., Goldstein, D.K., and Mead, M. "The Case Research Strategy in Studies of Information Systems," *MIS Quarterly* (11:3), September 1987, pp 369-386.

Benbasat, I., Goldstein, D.K., and Mead, M. "The Case Research Strategy in Studies of Information Systems," in: *Qualitative Research in Information Systems,* M.D. Myers and D. Avison (eds.), Sage Publications, London, 2002, pp. 79-99.

Bingi, P., Sharma, M.K., and Godla, J.K. "Critical Issues Affecting an ERP Implementation," *Information Systems Management* (16:3), Summer 1999, pp 7-14.

Bovey, W.H., and Hede, A. "Resistance to Organizational Change: The Role of Cognitive and Affective Processes," *Leadership & Organization Development Journal* (22:8) 2001a, pp 372-382.

Bovey, W.H., and Hede, A. "Resistance to Organizational Change: The Role of Defense Mechanisms," *Journal of Managerial Psychology* (16:7) 2001b, pp 534-548.

Callahan, T. "Credit Research Foundation Examines ERP Technology," 2002.

Cooper, R.B. "Information Technology Development Creativity: A Case Study of Attempted Radical Change," *MIS Quarterly* (24:2), June 2000, pp 245-276.

Eisenhardt, K.M. "Building Theories from Case Study Research," *Academy of Management Review* (14:4), October 1989, pp 532-550.

Fiorelli, J.S., and Margolis, H. "Managing and Understanding Large Systems Change: Guidelines for Executives and Change Agents," *Organization Development Journal* (11:3), Fall 1993, pp 1-13.

Glaser, B.G., and Strauss, A.L. *The Discovery of Grounded Theory: Strategies for Qualitative Research* Aldine, Chicago, 1967.

Hammer, M. "Reengineering Work: Don't Automate, Obliterate," *Harvard Business Review* (68:4), July-August 1990, pp 104-112.

Hammer, M., and Champy, J. *Reengineering the Corporation: A Manifesto for Business Revolution* HarperCollins, New York, 1993.

Hill, K. "System Designs That Start at the End (User)," 2002.

Hines, M. "PeopleSoft 8.8 to be More Usable, Easier to Integrate," 2002.

Hitt, L.M., Wu, D.J., and Zhou, X. "Investment in Enterprise Resource Planning: Business Impact and Productivity Measures," *Journal of Management Information Systems* (19:1), Summer 2002, pp 71-98.

Jiang, J.J., Muhanna, W.A., and Klein, G. "User Resistance and Strategies for Promoting Acceptance across System Types," *Information & Management* (37) 2000, pp 25-36.

Keen, P.G.W. "Information Systems and Organizational Change," *Communications of the ACM* (24:1), January 1981, pp 24-33.

Klein, H.K., and Myers, M.D. "A Set of Principles for Conducting an Evaluating Interpretive Field Studies in Information Systems," *MIS Quarterly* (23:1), March 1999, pp 67-94.

Kraemmergaard, P., and Rose, J. "Managerial Competences for ERP Journeys," *Information System Frontiers* (4:2) 2002, pp 199-211.

Krasner, H. "Ensuring E-Business Success by Learning from ERP Failures," *IT Professional* (2:1), January/February 2000, pp 22-27.

Landis, J.R., and Koch, G.G. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* (33), March 1977, pp 159-174.

Lee, A.S. "A Scientific Methodology for MIS Case Studies," in: *Qualitative Research in Information Systems,* M.D. Myers and D. Avison (eds.), Sage Publications, London, 2002, pp. 147-167.

Marakas, G., and Hornik, S. "Passive Resistance Misuse: Overt Support and Covert Recalcitrance in IS Implementation," *European Journal of Information Systems* (5) 1996, pp 208-219.

Markus, M.L. "Power, Politics, and MIS Implementation," *Communications of the ACM* (26:6) 1983, pp 430-444.

Markus, M.L., Axline, S., Petrie, D., and Tanis, C. "Learning from Adopters' Experiences with ERP: Problems Encountered and Success Achieved," in: *Second-Wave Enterprise Resource Planning Systems,* G. Shanks, P.B. Seddon and L.P. Willcocks (eds.), Cambridge University Press, Cambridge, 2003.

Markus, M.L., and Tanis, C. "The Enterprise System Experience - From Adoption to Success," in: *Framing the Domains of IT Management,* R.W. Zmud (ed.), Pinnaflex Educational Resources, Cincinnati, Ohio, 2000, pp. 173-208.

Maurer, R. "Plan for the Human Part of ERP," in: *Workforce Online*, 2002.

Miles, M.B., and Huberman, A.M. *Qualitative Data Analysis: An Expanded Sourcebook* Sage Publications, Thousand Oaks, CA, 1994.

Prasad, P. "Symbolic Processes in the Implementation of Technological Change: A Symbolic Interactionist Study of Work Computerization," *Academy of Management Journal* (36:6) 1993, pp 1400-1429.

Robey, D., Ross, J.W., and Boudreau, M.-C. "Learning to Implement Enterprise Systems: An Exploratory Study of the Dialectics of Change," *Journal of Management Information Systems* (19:1), Summer 2002, pp 17-46.

Shang, S.S.C., and Su, T.C.C. "Managing User Resistance in Enterprise Systems Implementation," Americas Conference on Information Systems, New York, 2004, pp. 149-153.

Surmacz, J. "Mix-and-Match ERP," *CIO* (http://www2.cio.com/metrics/2002/metric381.html:Printed 10/22/04), June 5 2002.

Umble, E.J., and Umble, M.M. "Avoiding ERP Implementation Failure," *Industrial Management* (44:1), January/February 2002, pp 25-33.

Wah, L. "Give ERP a Chance," *Management Review* (89:3), March 2000, pp 20-24.

Walsham, G. "Interpretive Case Studies in IS Research: Nature and Method," in: *Qualitative Research in Information Systems,* M.D. Myers and D. Avison (eds.), Sage Publications, London, 2002, pp. 101-113.

Wu, I.-L. "Understanding Senior Management's Behavior in Promoting the Strategic Role of IT in Process Reengineering: Use of the Theory of Reasoned Action," *Information & Management* (41) 2003, pp 1-11.

Yin, R.K. *Case Study Research: Design and Methods*, (3rd ed.) Sage, Newbury Park, CA, 2003.

Environment Uncertainty, Cooperation Relationship and Performance in 3rd Party Logistics---Transaction Cost, Agent and Resource Dependent Theories Approach

Dr. Jao-Hong Cheng
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: jhcheng@yuntech.edu.tw

Chih-Sheng Chen
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: torshell@seed.net.tw

Jr-Huei Tang
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: ellis_tang@hotmail.com

Abstract

At the end of the 20th century and the beginning of the 21th, the third party logistics issues (3rd Party Logistics; The assistance that hereafter referred to as 3PL) have emerged as an important and increasing interest research area in a variety of supply chain academic setting. In practice, enterprises through the third party logistics were in an attempt to arise the promotion that couldn't merely reach one's own logistics and dealt in the performance, but also supply chain reached the synergy of the integrated logistics, and made the supply chain of the industry have competition advantages. Therefore, it is one of the important subjects for enterprises to cooperate with the new intermediation well now.

Until comparatively recently researches, they had not only been guided by the single theory concept but been proved by constructing the integrality of theory framework. In this paper, we constructed model by the multi-theory completely based on transaction costs theory, agency theory, resource-dependence theory to analyze the relationships of enterprises and the 3PL. In addition as to enterprise, final purpose of all operation, investment, and the behavior for improving action was to improve performance of enterprise.

The linear structural relationship (LISREL) model was used to analyze survey data collected from 147 companies in Taiwan. These results provided the influence of exhibiter significant relationships with enterprises and 3PL, utilizing uncertainty factor of transaction costs theory, 3PL's opportunity cost of logistic function with outsourcing of agency theory, and complex logistic function of resource-dependence theory. These factors were crucial to integrate partnership of 3PL with outsourcing of enterprises. Throughout this paper, our goal has been to provide the integral partnership and performance of enterprises and 3PL with significant. It will take a concerted effort by researchers of a supply chain to shape a comprehensive understanding of multi-theory construct with logistic. Such an effort could make an important contribution, not only to researchers, but to practicality as well.

Keywords:

Transaction Costs Theory, Agency Theory, Resource-Dependence Theory, 3rd Party Logistics, Cooperative Relationship and Performance

1. INTRODUCTION

For integration of supply chain, enterprise organizations are intend to seek for alienation and cooperation with other enterprises, and try to make the pea profit in temporary, complicated networks. The third party professional logistic plays an important role which is through professional distribution of labors and gets its economic in the supply chain network, it was responsible to act as the cooperation and distributive role to combine from levels to levels. Thus, the whole efficiency can be largely enhanced. For this reason, the third party professional logistic has to develop and maintain the relationship of the positive and good distributed with enterprises in every level, hence the supply chain can be improved.

Way back the researches in the past decade, they had indicated that a grant for out of organization enterprise for supply service, mainly transition, storage and many kinds of transportation (Miller, 2000). A view on the transportation cost analysis, to compare their customer's partnership between the so called " 3rd " party professional logistic and traditional goods sending service, came a clean definition: the 3rd party professional goods supply service needs a broader vision and a more integrated perspective to manage goods supply chain. (Peter & Magnccs, 2000).

The goods flows had gradually become a corporation's potentially competitive advantage (Stank & Daughterty, 1997) service comprises transportation. The third party goods had attracted the global attention, cit serves more than international trade, service such as emergent supply. The patterns of competition had been changed: trends such as the rapid introduction into markets and the decrease of risk self – owned facilities in investments. The global cooperation between producers and retailers led to technical innovation. Examples are: information technology application and complex storage management.

2. LITERATURE REVIEW

2.1 Content of cooperative Relationship

The third middle-organization energy is from the supply chain system. They aim at offering better service for this entire network organization, it may include the substantial goods distribution and the collection of information, in the processes of offering service to both side in a trade, the relationship of cooperation is hence established. Cooperation means that in different points, different individuals combine with each other for the shared goods. The ultimate purpose is to create mutual and multiple – aspects. Kuman (1995) had pointed out that when different other and the barricade or obstacles stood in between as the entrance appease to be essential, the motivation became strong enough to maintain their relationship of cooperation, it was when they felt that they were in the same side, the mutual understanding of the state of "mutual win " and cooperation was thin feet choice.

In this study research, the definitional of "cooperation" means the close relationship between enterprise's cooperation. According to relevant research, the level of close relation differ of the two sides on the scale, one side of extreme is very close which is called strategic partner the other side is pretty loosely which is called the relationship of general market trade.

2.2 The meaning and Content of the Transaction Costs Theory

Transaction cost theory (Williamson, 1975, 1985) offered another perspective to help us to understand better the forces shaping structure of supply chain. The basic premise of transaction cost theory (TCT) was that the firm will internalize activities that it is able to perform at lower cost and will rely on the market for activities in which other providers have an advantage(Coase, 1937). Williamson (1975) & Kogut (1998) argued that the nature of cooperation was an exchange behavior. When " cost of exchange " plus "cost" was smaller than " internalized cost ", and the relative were willing to search the external resource to support the internal operation, thus the cooperative relationship established. Beside, the most important goal of cooperative relationship were to eliminate the cost and risk if operating and, to avoid any unnecessary investment. From the view of trading cost, the align cooperation focused on the reduction of the trading cost. The substantial motivation includes: reducing cost, sharing risk, and avoiding unnecessary cloudlet investment and to get sizable economic gain.

In contemplating the relationship, reviewing the literature of the theories were between the cost exchange and cooperative relations. Borrowing from the scholars' analysis and assumptions, the author conducted and found the outcome in researches that the founding perspectives of exchanging cost: the character of assets, the uncertainty, and the frequency of exchanging. The above three variables are substantial for explaining how the cooperative relation happened.

The characteristic unique of the assets are importance that the investment incurred the cooperative relations wouldn't stop until the sustainable function of the assets being officially expired. If, only if, the contract had being invalidated on destroyed, the asset provider would first face the difficulty for return to the investment. Besides, the other party of the contract would suffer loss because the ensuing situation that no appropriate target person to collaborate with. Therefore, the more unique the asset is, the more willing (full) the two sides of the investment project are to keep the mutual cooperative relationship.

2.3 The meaning and Content of the Agency Theory

Jensen and Meckling (1976) provoked that the deal with people mainly for preventing his agents' aberration behavior, would refer to invoke the contract design to supervise on motivate his agents. The expense which cost the bore extra money was the expenditure paid for supervision, in the agent's side, for convincing his lose that the utmost benefit was always the priority and the agents did their jobs considering on behalf of their clients' benefit. The agents would also self, regulate (with costs) themselves in order not to hurt the clients. This was the agents' cost to control them-selves. As to the opportunity cost, it meant that under the supervision and the context of organization construction, the agents couldn't take actions early enough and missed the chance of gaining benefits for his clients, it was also named " remains loss " in other words, the clients, though supervise and monitor, could keep the agents from bias and wrong decision was made and the loss was caused. The affirmed points that in the " Agent Theory ", three variables could be taken effectively to ascertain the above: the cost of supervision; the cost of self – control, and the opportunity cost.

As soon as the clients cooperate with the agents, the issue of substitute (agent) raised. When the organization developed the cooperative and managerial systems, people focused on eliminating or reducing the pseudo – cost (pre & post). Any sub – structure in a supply chain can get the relevant information by way of the monitor system. The purpose is to reduce the managerial cost. Or by motivation the clients can reach some kind of cooperative relation to make achieve a consensus and reduce the gamble behavior. However, the non – symmetrical information system existed in reality and the IMM (International Monetary Market) of yield, or the transformation of the nation on the industry. The products would inevitably be uncertain. All these would reduce the monitoring mechanism and motivation in function.

2.4 The meaning and Content of the Resource-Dependence Theory

In essence, from viewpoint of the "resource dependence" perspective, organization was an open systems. As organizations were operation under uncertain contexts, they weren't sufficient in sustaining. For survival, the organizations depend upon the resource. And for obtaining resource, the organizations had interactions with the external dominants of resource. They referred to the environment in which they existed. The "resource dependence theory" had its main emphasis. Enterprises depended on the resource owned by other corporations. Enterprises manage inter-organization relation in order to dominate and reduce such as a kind of dependence. Another option was to enhance the intra-organizational dependence.

Pheffer & Gerald (1978) issued the environmental mission concept that the organizational mission is comprised of three facets: the concentration, the openness and the coherence of resource. According to the disposition of the natural resource in the environment and facet analysis of the mission context, they separated individual the construction: resource concentration, resource munificence, and the interconnectedness of resource. The activity and varied value in the environment were hence interpreted (Aldrich, 1979).

Pheffer and Gerald (1978) Cheon et al. (1995) urged that the strategies relied on the environments accordingly and have relationship with other units on enterprises. In our research, the resource dependence theory specializes how enterprises can obtain the resources without risk and further and what they are capable of building the resources with other organization in the environment. To its utmost, the organizations survive forever in competition. It leads to the appearance of agents on weaker dominant agencies. As to the operational procedure of the resource based on RDT, the mission context of an organization (concentration, munificence, interconnectedness) would definitely determine the structure (the significance, contemplation, and options). And, the resource structure and the policies together, would jointly have an impact on agent decision-making function (Cheon et al., 1995). In other words, in the RDT perspective, the mission context of the designed organization would influence choices of cooperative relations.

2.5 The meaning and Content of the supply the performance of the chain and define it with intension

Beamon (1996) reviewed the existed literature and came to a conclusion that the moot common evaluative approaches could be divided into qualitative and quantitative. The qualitative indicators were used to assess the characters which couldn't be numerically expressed. And the others could stand for accurate amount, extends, degrees, in digital way. The quantitative approach which represents the quantitysuch as: the satisfaction of consumers, flexibility, the integration of information and goods. The quantitative indicative refer to figures and mumbles which represent the descriptive measurement.

After reflecting literature analysis, we know that there isn't definite standard on criteria. Beamon's (1996, 1998) qualitative and quantitative surveys were the most widely applied tools. It was the reason the author used Beamon's (1996) scales (of qualitative and quantitative) and integrated the both and transformed the individual efficiency measurement into an efficiency system. The multiple indicators developed by the scholars mentioned above are used as the variables of the supply chain. The purpose is to figure out the correlation among " the uncertainty of the cooperative relationship in the supply chain network and the efficacy".

3 RESEARCH DESIGN

From the problem solving viewpoint, the researcher conducted with the literature to develop a logical framework model and the hypothesis. Then, the variables were defined and data was collected, the analytical approaches were under examination at last it was all about the validity and efficiency.

3.1 Framework and Hypothesis of this Research

A Theoretical Perspective is applied for depicting the framework of this study, as shown in Figure 1. The null hypothesis in it showed H1, H2, H3 … H10 of a reflective viewpoint in “The Third Party’s Professional Goods Distribution, its cooperative Relationship, it was apparent that the general was like Young’s (2002) point. As far as the entrepreneur enterprises owned relatively limited resource, being faced with the highly uncertain market, competition and technological environment, they tended to seek for a collaborative relationship and partnership with other enterprises’ in order to reduce the threats of uncertainty. Besides, Hallijas *et al.* (2002) thought that the environmental uncertainty played a key sole to force organizations push and hunger for inter organizational cooperation’s.

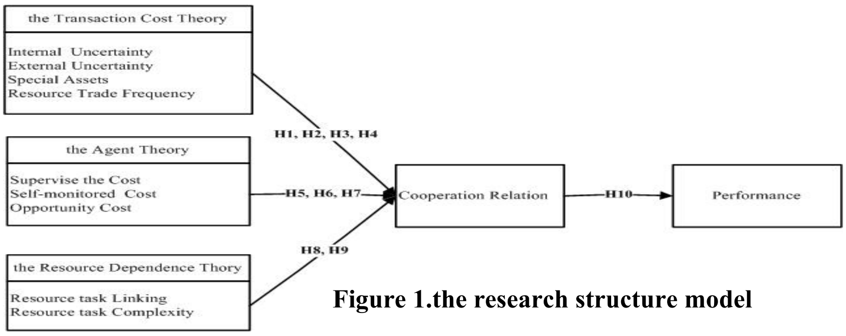


Figure 1.the research structure model

3.2 The hypothesis in this research

From this dynamism it follows that some sort of uncertainty always exists inside the transactional network environment. Hallkas *et al.*(2002) deemed uncertainty had an influence on the decision-making process and particularly the individual behavioral attitude. And, as long as the future profits of the company depend on the future potential of partners, it’s essential to be able to evaluate transaction specific uncertainty in terms of the transaction costs and the option value of the partner. So, here come the two hypotheses:

H1: When an enterprise’s awareness of the uncertainty is stronger, it intends to build construct a closer cooperative relation with other enterprise on organization.

H2: Where an enterprise is aware of its own feeling that the uncertainty of its internal goods distribution capacity is at a higher level, it tends to build a chooser cooperative relationship with other enterprise on organization.

The “monopoly investment” means that a certain asset which is not transferable and the extend in which the restriction to keep it from other usage. The restraint is in degree, the stricter the monopoly on what unique belonging is. The value of that asset is hence determined by the endurance of such a partial trading relation. Once the deal relationship expires, it is impossible to duplicate its former value in any other forms of trading. Next, we think the monopoly investment would inevitably restrict the two parties of the deal in a specified. Stank & Daugherty (1997) had discovered that in the goods distribution function area, if an asset is unique belonging to in character, then its cooperative relation with the agent system is a positive correlation, Here comes the third hypothesis:

H3: When an enterprise is aware of the highly monopoly nature in its goods distribution, it’s inclined to find a closely cooperative relationship with other enterprise on organization.

Young (2002) found that as middle and small enterprises trade frequently with thin partners. The formation of this kind of societal web net can not only enhance the transmission of knowledge and technology but also help to reduce the cost in mutual adaptation and delaying. The result, then, is that the frequent trading between enterprises can lead to the construction and application of organizational and individualistic in building the social insertion. However, Stank and Daugherty (1997) discovered in thin findings that the frequency of goods distribution (transportation, purchases, and consumption), but they didn’t show a significant correlation ship with “ agent service “ organizations. Based on that conclusion we have the fourth hypothesis H4:

H4: When an enterprise needs higher frequency of goods distribution, it tends to construct a closely integrated cooperative relationship.

These supervisions want to know, if the Third Party Goods Distributions were doing what they were supposed to do. Sbcctt's (1995) finding was approximately 38% of the consumption enterprises monitored constantly on thin retailers. In the past literature and papers, we don't know if the monitoring or supervision would have an impact on the agent's willingness and compliance. In concept, the context in monitoring is quite different with that of motivation, for example, motivation and attraction are based on the creation of the shared profits which could clearly motivate the retailers in cooperation. The more often an enterprise used the monitoring strategies.

On the contrary, monitoring is a way invoking the equilibrium of information to reduce the standoff between the client enterprise and the agent company, the more likely the client company to exploit the self – behavior of the agent company. It leads to the punishing act towards to the agent company (of goods distribution). In other words, the disclosure of the agent good's distribution's act of breach of the contract leaves a space for the enterprise to hold on its power is a coercive mechanism. (Blau, 1995 ; Blau & Scott, 1962). We develop the fifth hypothesis: H5:

H5: The cost of monitoring is more risen, a trend of the construction of a loosely integrated cooperative relationship is more likely to happen.

Eisenhardt (1985) drew out that the client and agent might belong to different, separate organizations accordingly. They might have diversified goals and human resource in place. The client plays as either the consumer or the user of an agent who assigns a job to the agent organization. So efficacy is a concerned. There are inevitably explicit or tact potential of conflict in between the two parties and also the contradictory risk preference.

Fama and Michael (1983) discovered in their research that the ownership and control of an organization are separated, the agents and clients become cooperative in a certain way. This relationship develops as a result of the tangling combination of profits and risk sharing between each party. The client one party in change has to control the agent's behavior. Or otherwise the agent has to control himself. The purpose is set to reduce the loss of the client's. Thus, there are the hypotheses H6-H7:

H6: When an enterprise is aware of the increasing cost of self – restriction of the third party professional goods distributions, it tends to build a chooser and integrated, cooperative relationship with other enterprises.

H7: When an enterprises is aware that the opportunity cost immerged from a third party goods distributor becomes lover (in cost), it tends to build a closely integrated, cooperative, relationship with its partner.

Pheffer and Gerald (1978) interpreted that the "commission linkage " meant " a state of combination, coherence or linkage between the resource and an organization. " Knemeyer el. at (2003) & Bowersox el at. (2000) referred " the consequential, ensuing combination concept " to " the combination of resource comes from either inter – personal or organizational relationship. it might include emotional factors. The combination would have a lasting effect on the ensuing organizational relational lets assume that the mission connection could influence the third parties goods distributors' relationship with both the upper Stream enterprises and the consumers (customers). According to this assumption, there appears Hypothesis H8:

H8: The higher the mission connection is, between an enterprises goods distribution practice with professional Third party goods distributor, the more likely they will tend to build up a closely coordinate (integrated) cooperative relationship.

The mutual inter dependence on resource become increasingly higher, mutual organizational interaction is more likely to occur the complex mission will be accomplished earlier on in time. Young (2002) when the entrepreneurs needed to pour in more kinds and amounts of resource to produce on sell their products but because of thin owned limited resource, they tended to make use of their social web net to pull together whatever resource available to continue the operation of their enterprises. However, in Rao & Young's resesnck (1994), they found the real reason why this agent's system worked. it was because of the Logistics Complexity of the goods Distribution. And the contextual factors came from the related geographical, product character and communication and monitor of the goods distribution. Complicacies in the mission would have an impact on the policies of the agenting decision. Here, then came the 9th hypothesis. The complicativeness of goods distribution emerges where as the communication, and control among sale goods, its geographical, good's characteristic and the chain of supply. Level of the complicativeness in an enterprise's goods distribution will inevitably have an impact on the enterprise's on agent policy. Therefore, there's hypothesis H9:

H9: As the level of complicativeness arises goods distribution, the enterprise tends to build a more concrete and close relationship of cooperation.

Dyer & Singh (1998) depicted that as the willingness to integrate between different enterprises, it was more likely that the value chain will become more productive. Also, with the more effective integration undergoing with thin resources, there will be the more advantage of competition in place. Broadly speaking, the linkage between the specified organizations will perform in the key role of the source of productivity and competitive advantage.Holm, Eriksson & Johansom (1996) asserted that the coordinative interaction was heightened on the willingness to mutual adaptation for further cooperation under way, the coordination was easier to go on. Their unified productivity will increase. Asanuma (1989) and Dyen (1996) both indicated that when the partners agreed to invest under this kind of particular relationship and combined unique resources, then their profit gained in the chain of value would be enhanced. The cooperative benefits in the chain supply were also called the " shared benefits " (Newton el at., 1997). In some case, the potential effect emphasizes included the time in cycle, the decrease of operational cost. The main purpose of enterprise's agent service is to enhance the effect of the organization. Here comes the tenth hypothesis H10:

H10: The higher the integration of the cooperative relationship is, the higher effect emerged from the cooperative relation is.

4. RESULT

With SEM used for analyzing, the researchers proceeded with Testing Factors Analysis and followed by exploring the potential variables' cause and effect relationship.

4.1 Testing Factors Analysis:

Under SEM analysis, in order to assess the entire pattern's fitness and elaborating the Letter degree and validity, we needed to proceed with the Testing Factors'Analysis. Based on the basis SEM hypothesis, the following tests were worked out :

(1) Cronbach's αcoefficient to analyze the Letter degree of each scale analysis on the Letter degree included trade frequency, opportunity cost. In which the conceptual constructure belonged to unique measuring variable other variables such as the external uncertainties, the internal uncertainties, the exclusiveness of the asset, the internal uncertainties, the exclusiveness of the asset, the supervision cost self – restrictive cost, the resource mission combination, the complicativeness of the resource mission to what extended the cooperation integrated and the efficiency. The total was Cronbach's coefficients were 0.76, 0.75, 0.74, 0.76, 0.74, 0.71, 0.84, 0.78 and 0.78. Nunnally (1978) suggested that if the Letter degree level were above 0.7 then it would be sufficient

On testing the convergent validity, there was a more common approach appeared in literature, suggested by Formall & Lanker (1981). The condition in which each constructure and its correspond question where their average extracted variance achieved above 0.50, it was called sufficiently convergent validity. According to the external and internal uncertainties, the exclusiveness of its resourceful assets, the supervision cost, 3 PL – self – restrictive cost, the combination of resourceful mission and its complicativeness, the integrity level of cooperation and the efficiency. Testing the above items and the respective AVE values were 0.674, 0.667, 0.795, 0.593, 0.553, 0.762, 0.863, 0.786, 0.818, 0.795, 0.792 and 0.825. There figures were all above that of what Fornell and Larker (1981) suggested. Therefore, the questionnaires of this research were convergent valid.

Besides, according to Fornell, Larker (1981) and Espinoza (1999), AVE values could also be used to test the "differential validity ": When and as the AVEs in each constructure are more than the square value of the coefficients between each c constructure, it's defined that it achieves the differential validity standard. To this research, the minimal AVE value of each constructure was 0.553 and higher than the square of the ultimate coefficients between each constructure (0.549). Thus, there existed the differential validity.

A fitness test of the total models in this research was 114.64 (P=0.000) in Fitness clndicator X2. Given a =0.05 in significant criteria, the test result was the rejection of Ho (hypothesis). It indicated that " the observed co – variant matrix equals to the assumption co – variant " was rejected. This result didn't necessarily indicate that the theoretical models couldn't match up the observed practices. The reason is that the X2 value would too easily swing as the sample figures Changes. (Joreskog & Sorbom, 1989). Once the amount of samples is too many, it is inclined to reject the null hypothesis. Therefore, in this research referred to other indicators as the tools to evaluate the fitness in full range model, to see if it was ideal. Besides, GFI indicators mean the extend level to which the theoretical model can explain the variants and co – variance. The higher the GFI value is, the better fit the mode can match up the full range. In this research, the GFI figure was 0.93 and was higher than Subhark's (1996) suggested – 0.90. It meant the acceptability. Another AGFI value was another indicator which is adjusted from GFI through freedom range. In this study, the value was 0.83 and higher than Subhark's suggested value (1996) of 0.80. It also meant the model fitness was acceptable.

The RMR value was to mix incomplete and bad variation and count the square root of the average of the parameter altogether, it reflected incomplete and bad size, the more so the one that exceed Subhash (1996) and propose in its value was close to 0, the better the ones that show the way was suitable for degree, RMR value of this research was 0.053, very close to 0. There were two indexes NFI and IFI value besides, it was value by way of theory or the degree of freedom that and came compared with 2 value of ¦Ö of the datum line way or the degree of freedom, because it will be the worst way that datum line way was suitable for degree, so what these two indexes reflect was both theory ways ' appreciation was suitable for degree ' (Incremental Fit), value of NFI lie between 0-1, Bentler and Bonett (1980) think if index this greater than 0.9 more than showed to suitable for degree extremely good way all, the value of this research NFI was 0.91. And the minimum of IFI index was 0, but the maximum will exceed 1, Bollen (1989) thinks that it was good that IFI index must be close to 1 or greater than 1, IFI value of this research was 0.94. To comment ancient bronze mirror result show to mix appropriate to all accord with ideal value model the above-mentioned all index.

4.2 Study the verification of the structure hypothesis relation

Under theory view, logistics by the other person special business factor and professional logistics cooperation degree of company, logistics of company committee enterprise, and the analysis of cooperation and performance, assay and suppose H1 to H10 through the γ value of the way, judge with t-value whether the relation between the two influences the existence showing the relation at the same time, the analysis result is as form table 1.

Uncertain impact on the other person's professional logistics cooperation degree of outside that enterprises face, can be from proving that the result learns under $\alpha = 0.05$, unable to refuse to suppose H1, namely enterprises are when being the higher in the face of its external environment condition uncertainty, cooperation degree outside higher than to adopt and walk logistics homework committee of enterprises. The effect of how the internal organizational uncertainty has influence on the cooperative level, with the Third party Professional Goods distribution agent, can be discovered that the testing outcome, when a = 0.01, we couldn't reject H2. That is, facing the higher uncertainty inside the organization itself, the more tendencies the organization is to adopt and proceed with a higher cooperation with professional goods distribution agent. When the enterprise recognizes that price of the service by the third party, the goods' distribution is lower in its opportunity cost, the level of cooperative relation is closer which is likely to have. From the result of the test we find that when a = 0.05, the H7 hypothesis is rejected, it means that the lower the opportunity cost of enterprise 's post understanding on professional Goods' Distribution was, the tendency is more likely that the enterprise would adopt a closer cooperation with the agent company. The more complicated the resource to pour in for professional goods' distribution; the adoption of the agent company is more intense in their cooperative relationship. Result of the testing shows that under the condition a = 0.01. H9 is impossible to be rejected. That is the more in amount and the more complicative of the resource an enterprise has to invest in the past, the more tight relation it seeks to pursue in its relationship with its cooperation with the professional Goods' distribution agent company.

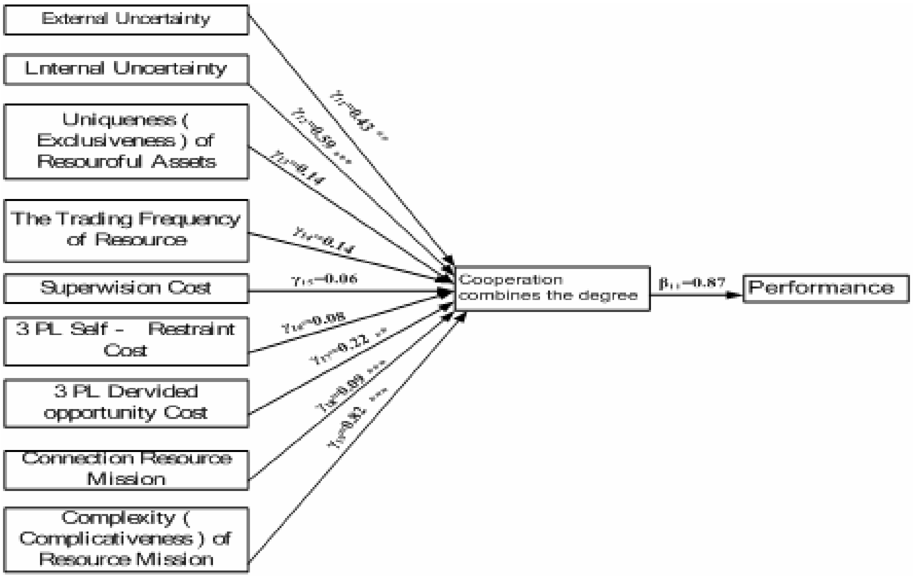
Table 1.The coefficient and hypothesis of the route are assayed

H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}
γ_{11}= 0.43	γ_{12} =0.59	γ_{13}= 0.14	γ_{14} =0.14	γ_{15} =0.06	γ_{16} =0.08	γ_{17} =0.22	γ_{18} =0.09	γ_{19} =0.82	β_{11} =0.87
** t=2.35	*** t=2.74	t=1.47	t=1.62	t=0.59	t=0.79	** t=2.16	t=0.75	*** t=4.57	*** t=2.90
▯	▯	○	○	○	○	▯ ▯3	○	▯	▯

"*" ▯ $\alpha < 0.1$▯ "**" ▯ $\alpha < 0.05$▯ "***" ▯ $\alpha < 0.01$▯ "▯ " ▯ There is positive correlation of showing "▯ "▯ Shoulder apparently relevantly "○"▯ Not apparent

In order to shoulder relevantly, and weigh and ask that the verification of one is to the design in hypothesis of H7, so, although coefficient of γ17 is, it should be in order to shoulder apparently relevantly.

The analysis towards whereas the goods. Flow of an enterprise by professional third party Agent, to figure out the relation between the level of cooperation and the efficacy, the researchers apply the direct effect – β value test H10 and t – value, to determine whether there is the significant correlationship. The result comes in Table 1. It describes that the better the relationship exists, the more effective the cooperation is. In other words, under the condition as a = 0.01, the H10 is not to be rejected. In all, as in the past, an enterprise has the closer cooperative relationship the performance of that enterprise is more effective. The pattern is shown figure 2.



Figur.2 The antithetical couplet of the cooperative factor outside theory structure and the other person's professional logistics committee sets up the structure and concerns way pictures.

1..“ * “$\rightarrow \alpha = 0.1$ ” ** “ $\rightarrow \alpha = 0.05$ ” *** “ $\rightarrow \alpha = 0.01$.

2..The thick line is a hypothesis that is originally researched and proposed.

3..It is relevant that the full line shows, the dotted line is not apparent.

5. CONCLUSION AND SUGGESTIONS

5.1 Conclusion

This paper takes a different view with the multiple theory perspectives, which is based on the transaction cost theory, agent theory, and resource dependency theory, to construct the models and argue that the 3PL maximum utility. Consequently, the call for the practice and academic to have the fresh eyes necessary to select and utilize the 3rd logistic partners become loud and clear. Pattern and models were developed. Followed by analysis on how enterprises do cooperate with the professional good’s flow agents, and how the enterprises worked through such a kind of cooperation to improve the integration among related industrial chains and lead to their corporate effectiveness. The construction on a “ cause and effect “ model could help to coordinate and explain how enterprises could cope with the professional goods flow agents, in theory as well as in practice. The following is our result:

1. After our pragmatic (empirical) research, we discovered that by gaining the evidence in testing the relevant hypothesis on trading cost theory that as the enterprise faced the external uncertainties, they teuded to be influenced and responded naturally as to what have caused impacts upon for the ultimate purpose of survival. As to what extension or what level they adapted to cooperate with the Third Party Professional Goods & Service Distribution companies. First to mention is, when α = 0.05, it was impossible to reject H 1. In other words, the higher the level the level of external untainty is, the more possible an enterprise is out of“ the system “ willing to adopt the policy of agent’s service and thus to construct a closer cooperation ship with the latter. When an enterprise is facing a more unstable.

Internal organizational situation, to what extension does it have an impact on agents’ service strategy ? We gain from the data and analysis and come to a conclusion, that is when the α is 0.01, H2 is impossible to be rejected. In other words, the higher the level an enterprise is facing its internal uncertainty, the more possible it would adopt a higher cooperative, agent’s service relationship with the Third Party Professional Good’s Service.

2. In this study, with an empirical overview on the Agent’s Theory, we appear that enterprise recognized that the lover the chance cost of the Third Party Professional Goods Service was, the more willing it would cooperate with the agent’s service company. The result shows under α = 0.05, H7 is impossible to be rejected. In a short word, when an enterprise

knowingly recognizes that the agent's service is the bower – priced, it would be more inclined to adopt the strategy to have a tighter cooperative relationship with the Third Party Professional Goods' Distribution Service.

3. As to empirical perspective on the Resource Dependence Theory, we discover that the more complicated the organizations become, the more resource necessary to pour into Good's Distribution is, the more likely it would seek an agent's service system to under the Good's Flows. We seem in the result that under the situation α = 0.01, H9 is impossible to be rejected. Broadly speaking, the level of complicates rises in investing resource or the needed amount cooperative relationship, it would naturally incline to adopt a tight cooperative relative relationship with the Third Party Professional Good's Distribution company.

4. Enterprise learns from his experience. As they maintain the letter relationship with the Professional Good's Distribution system's, they seem that the effectiveness is better, too. In our testing, under the situation when α = 0.01, the H10 is not reject. Surely, the letter the cooperative relationship is, the more efficient the enterprise performances.

5.2 Suggestion

In our study, we appear that " uncertainty " has an important role in deciding to what extension an enterprise pursue the cooperative relationship with the Third Party Agent's Service company in good's distribution. In practice, there are the different levels of uncertain factors. For further study, researcher might add to more level and elements of uncertain factors and undergo more analysis, test, and develop patterns to enterprise the theories and practice, and their correlation. Otherwise, we construct the framework should be test with many example of the operational ideology, like the cooperative management of the 3G's, e-health model.

REFERENCES

Aldrich, H.E.（1979）, Organizations and Environment, NY: Prentice-Hall.

Aldrich and Pfeffer（1976）, Environments of Organizations," Annual Review of Siciology, pp79-105.

Arrow, K. J.（1985）, "The Economics of Agency, inPrincipals and Agents: The Structure of Business", J.W. Pratt and Zeckhauser, R. J. eds. Boston, MA: Harvard University Press, pp37-51.

Beamon, B. M.（1996）, Measuring Supply Chain Performance," International Journal of Operations and Production Management, Vol.3, p275-292.

Beamon, B. M.（1998）, Supply Chain Design and Analysis: Models and Methods," International Journal of Production Economics.

Bergen, M., Dutta, S. and Walker, O.（1992）, "Agency Relationships in Marketing: A Review of the Implications and Applications of Agency and Related Theories," Journal of Marketing, Vol.3, pp 1-24.

Bentler, P. M.（1990）, "Comparative fit indexes in structural models,"Psychological Bulletin, Vol.107, No.2, pp238-246.

Blau, P. M.（1995）, The Dynamics of Bureaucracy. Chicago:University of Chicago Press.

Bourgeois, L. J.（1985）, "Strategic Goals, Perceived Uncertainty and Economic Performance in Volatile Environments," Academy of Management Journal, Vol.3, pp548-573.

Bowersox, D. J., David, J. C., and Theodore, P. S.（2000）, "Ten Mega-Trends That Will Revolutionize Supply Chain Logistic," Journal of Business Logistics, Vol.21, No.2, pp1-16.

Cannon, J. P., William, D. C., and Perreault, J. R.（1999）, "Buyer-Seller Relationships in Business Markets," Journal of Marketing Research, Vol.31, November, pp439-460.

Cheon, M. J., Grover, V. and Teng, T. C. J.（1995）, "Theoretical Perspectives on the Outsourcing of Information Systems," Journal of Information Technology, Vol.10, pp 209-219.

Child, J.（1972）, "Organizational Structure, Environment and Performance: The Role of Strategic Choice," Sociology, Vol.6, pp1-22.

Coase, R.（1937）, "The Nature of the Firm, Economica, 4, November, 386-405. Reprinted in Oliver E.

Dess, G. G., and Beard, D. W.（1984）, "Slocum, Dimensions of Organization Task Environments, Administrative Science Quarterly, Vol.29, pp52-73.

Dyer, J. H., and Singh, H.（1998）, “The Relational View: Cooperative and Sources of Interorganizational Competitive Advantage,” Academy of Management Review, Vol. 23, pp 660-679.

Eisenhardt, K. M.（1985）, “Control: Organizational andEconomic Approaches,”Management Science, Vol.31, pp134-49.

Emery, F. E. and Tris, R.（1965）, “The Causal Texture of Organizational Environments,” Human Relation, pp21-31.

Espinoza, M. M.（1999）, “Assessing the Cross-cultural Applicability of Service Quality Measure: A Comparative Study Between Quebec and Peru,”International Journal of Service Industry Management, Vol.10 No.5, pp449-468.

Fama, E. and Michael, J.（1983）, "Separation of Ownership and Control," Journal of Law & Economics, No. 26, pp301-325.

Fornell, C. and Larcker, D.（1981）, “Evaluating Structure Equations Models WithUnobservable Variables and Measurement Error,” Journal of Marketing Research, Vol.18, pp39-50.

Ganesan, S.（1994）, “Determinants of Long-Term Orientation in Buyer-Seller Relationships,” Journal

of Marketing, Vol.58, pp1-19.

Gooely, T. B.（2000）,” Growth Spurt,” Logistics Management & Distribution Report, Vol.39, No.11, pp77-84.

Hallikas, J., Virolainen, V. M. and Tuominen, M.(2002), “Understanding Risk and Uncertainty in Supplier Networks--a Transaction Cost Approach,” International Journal of Production Research, Vol. 40 Issue 15, pp3519.

Heide, J. B. and John, G.（1990）, “Alliances in Industrial Purchasing: The Determinants of Joint Action in Buyer-Supplier Relationships,” Journal of Marketing Research, Vol.27, P24-36.

Heide, J. B. and John, G.（1992）, “Do Norms Matter in Marketing Relationships?” Journal of Marketing, Vol.56, pp32-44.

Holm, D. B., Eriksson, K. and Johanson, J.（1996）, “Business Networks and Cooperation in International Business Relationships,” Journal of International Business Studies, Special Issue, pp1033-1053.

Jap, S. D.（1999）, “Pie-Expansion Efforts: Collabratuon Process in Relationships,” Journal of Marketing Research, Vol.36, pp461-475.

Jensen, M. C. and Meckling, W. H.（1976）, “Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure,” Journal of Financial Economics, Vol.3, No.2, pp 305-360.

J. Hallikas, V.-M. Virolainen and M. Tuominen(2002)’ Understanding risk and uncertainty in supplier networks-a transaction cost approach.’ International Journal of Production Research Accounting and Economics, Volume: 40, No.15, pp. 3519-3531

Joreskog, K. G. and Sorbom, D.（1989）, LISREL 7: A Guide to The Program and Application, Chicago, Ill: SPSS Inc.

Joskow, P. L.（1987）, “Contract Duration and Relationship-Specific Investments: Empirical Evidence From Coal Markets, ”American Economic Review, Vol.77, pp168-185.

Knemeyer, A. M., Corsi, T. M., Murphy, P. R.（2003）, “Logistics Outsourcing Relationships: Customer Perspectives, “Journal of Business Logistics, Vol. 24 Issue 1.

Kogut, B.（1988）, “Joint Ventures: Theoretical and Empirical Perspectives,” Strategic Management Journal, Vol.9, pp319-332.

Kumar, N., Lisa, K. S. and Steenkamp, Jan-Benedict E. M.（1995）, “The Effect of Perceived Interdependence on Dealer Attitudes, Journal of Marketing Research, Vol. 32, Issue 3, pp 348-356.

Lee, H. L. and Billington, C.（1993）, “Material Management in Decentralized Supply Chain,” Operation Research, Vol.41, No.5, pp835-847.

Logan, M. S.（2000）, “Using Agency theory to Design Successful Outsourcing Relationships.International,” Journal of Logistic*s* Management, Vol.11 Issue 2, p2.

Michael, G. J.（1988）, "Selection of Partners of International Joint Venture," Business Quarterly, pp31-36.

Newton, B. F., John, C. L., and Gary, R. A.（1997）, “Third-Party Logisrics Study,” Detroit, MI: Cap Gemini Ernst and Young.

Pfeffer, J.and Gerald, S.（1978）, The External Control of Organizations: A Resource Dependence Perspective. New York: Harper and Row.

Rindfleisch, A., and Heide, J. B.（1997）, Transaction Cost Analysis Past, Present, and Future Applications, Journal of Marketing, Vol.61, pp30-54.

Robbins, S. P.（1990）, Organization Theory, Structure, Design and Applications. N. Y.: Prentice Hall Inc. 3rd ed.

Scott, W. R.（1989）, Organization: Rational, Nature, and Open Systems, 2th ed Englwood Cliffs, NJ: Prentice-Hall, Inc.

Shutt, C. A.（1995）, "1996 P -O-P Times Trends Surveys," P-O-P Times, Vol.8, No.10, pp54-80.

Stank, T. P., Daugherty, P. J.（1997）,"The impact of operating environment on the formation of cooperative logistics relationships," Transportation Research Part E: Logistics and Transportation Review Volume: 33, Issue: 1, pp53-65

Subhash, S.（1996）, "Applied Mutivariate Techniques,"New York: John Wiley and Sons, Inc..

Teigen, R.（1997）, Information Flow in a Supply Chain Management System, Available at http://www.eil.utoronto.ca/profiles/rune/dip-thesis.html.

Thompson, J. D.（1967）, Organizations in Action (McGrawHill, New York). Ulrich, D. and Barney, J.B.（1984）Perspectives in organizations: resource dependence, efficiency, and population. Academy of Management Review, Vol.9, pp471-481.

Williamson, O. E.（1975）, Markets and Hierarchies: Analysis and Antitrust Implication, New York: The Free Press.

Williamson（1990）, Industrial Organization, New York: Edward Elgar Publishing Company, pp3-22.

Williamson, O. E.（1991）, Comparative Economic Organization: The Analysis of Discrete Structural Alternatives, Administrative Science Quarterly, Vol.36, No.2, June, 269-296. Reprinted in Oliver E.

Williamson and Scott E. Masten（1999）, The Economics of Transaction Costs, New York: Edward Elgar Publishing Company, pp101-128.

Williamson, O. E.（1996）, The Mechanisms of Governance, New York: : The Free Press.

Process Driven Business Value Assessment of ERP Solutions: An Overview of the Extended SAP Method

Prof. Dr. Joerg Becker
ERCIS - European Research Center for Information Systems at the University of Muenster
Muenster, Germany
Email: becker@ercis.de

Dr. Victor Taratoukhine
SAP AG
Moscow, Russia
Email: victor.taratoukhine@sap.com

Lev Vilkov
ERCIS - European Research Center for Information Systems at the University of Muenster
Muenster, Germany
Email: lev.vilkov@ercis.de

Tobias Rieke
ERCIS - European Research Center for Information Systems at the University of Muenster
Muenster, Germany
Email: tobias.rieke@ercis.de

Abstract

Profitability analysis in preliminary stages of ERP implementation projects is a critical factor for the success of ERP investments. This paper summarizes the results of an international research project with SAP, aiming at methodological and software-technical extension of the existing SAP business value assessment methodology. The extended SAP method is characterized by: (1) a process oriented analysis of ERP driven benefits, (2) a unified knowledge base structure for storing reference data related to ERP profitability analysis, (3) using reference data for effective business value assessment of ERP solutions and (4) reduced cost in preliminary stages of ERP projects. The SAP Value Analyzer is a software-technical implementation of the extended method.

Keywords:

IT value, business value assessment, profitability analysis, process orientation, ERP, SAP.

INTRODUCTION

Enterprise Resource Planning (ERP) systems are known as effective in supporting business process optimization and data integration throughout an enterprise (Martin et al., 2002; Wallace, 2001; Scheer, 2000; Laudon, 1998). At the same time, reorganization of business processes and the complexity of an ERP implementation projects require high IT investments (Bancroft et al., 1998; Buxmann, 1996; v. Arb, 1997). Therefore, a profitability analysis in preliminary stages of ERP implementation projects is a critical factor for the success of ERP investments (Eschinger et al., 2004; Holland, 1999; Parr, 1999; Muschter, 1999).

The SAP business value assessment (BVA) methodology aims at forecasting the profitability of investments in Enterprise Application Software (EAS) of SAP, including the ERP system as an integrated part of the SAP EAS (SAP, 2004). The SAP is aiming at the continuous improvement of the methodology and initiated a joint international research project at the beginning of 2005. The project is funded by the SAP University Alliance Program and SAP Research. The aim of the project is to provide methodological and software-technical extensions to the existing SAP BVA methodology.

The aim of the paper is to summarize the developed extensions. We start with the review of the used research methods pointing out key findings from our common practice analysis. The key findings represent the aspects we followed to address the major theoretical paradigms employed in prior IS research. Based on this awareness we present the procedure model and point out the major extensions to the existing SAP BVA method. The extended SAP BVA method is the conceptual basis of the developed software prototype “SAP Value Analyzer”. We conclude by summarizing the limitations and the identified gaps in knowledge, showing the propositions to guide future research.

RESEARCH METHOD

Using structured interviews we gained insights into the common ERP BVA practice of 5 major car manufacturers, 6 large automotive suppliers and 5 consulting companies with a focus on IS/ERP implementations. The focus on the automotive industry was selected for two reasons: first, according to the SAP business consultancy, ERP systems are widely recognized in the sector due to high process automation level and the necessity for deep data integration; second, it was necessary to ensure comparable interview results. In the context of further ERP investments, companies point out the necessity of:

- stronger orientation on specific business processes,
- transparent and detailed profitability analysis of ERP solutions,
- assessment of the ERP business value based on reference experience from other projects.

Focusing on these key findings we analyzed the tools for ERP BVA used by the SAP business consultancy. Following the prototyping paradigm (Gordon, Bieman, 1993) we developed a concept of the extended method and a new software tool. To evaluate our research we conducted a field study with GM-AvtoVAZ, a joint venture established by the General Motors, AvtoVAZ (the largest car manufacturer in Russia) and European Bank for Reconstruction and Development. For the Parts & Accessories Department we developed a detailed profitability analysis of the mobile barcode-scanning infrastructure in combination with the SAP Warehouse Management.

RELATED RESEARCH

Orientation on specific business processes

Even from as far back as the 1930's, Nordsieck highlighted the necessity of a process-oriented corporate design and continued to do so in the 1970's (Nordsieck, 1934, 1972). In spite of the early discussion of this topic in academic literature, process orientation only began to be practiced in the 1980's, after Gaitaniedes (1983), Porter (1985), Scheer (1990), Davenport (1993) as well as Hammer and Champy (1993) published their approaches. While the organizational structure traditionally divides the company into sub-systems with their assigned tasks, a business process deals with the execution of these tasks as well as their coordination in time and sequence (Esswein, 1993). ERP systems aim at the integration and automation of the business processes and, therefore, are able to increase its performance (Becker et al., 2003).

Business value of IT

The measurement of the economic value of IT/ERP is the basis for the well founded IT investment decisions (Wegen, Hoog, 1996). In this paper we use the term 'business value of IT' with regard to a profitability analysis of IS, particularly as a decision support instrument for ongoing investment in IS (Schumann, 1993). It becomes obvious that not IT itself but the impacts which are closely related to the application context are the drivers of the IT cost and benefits (Remenyi, 2000). Following the concept, which considers IT value as the positive impact of IT on organizational performance (Mukhopadhyay et al., 1995) it is necessary to compare the cost and the benefit of the selected IS (Stahlknecht, 2005). Due to the unstructured decision situations (Farbey et al., 1999) the identification of the IT impacts becomes a critical issue for transparent and reliable IS profitability analysis, particularly in the context of such complex systems like ERP (Renkema et al., 1997).

Porter (1985) defined the value chain placing processes in the center of attention and paved the way for further process-oriented IT business value research. Numerous research approaches studied the cause-effect relationships between IT and its impact on the level of business processes (Mende et al., 1994; Mooney et al., 1996; Barua et al., 1995). Also an empirical research of ERP system implementations showed that the biggest benefits have been achieved at the level of business processes (Martin et al., 2002). Melville et al. (2004) presented one of the latest IT business value models with the main focus on specific business processes. According to this model, IS support the execution of different activities, increasing business process performance, and generate beneficial effects in this way. Walter et al. (2004) suggested to divide the ex-ante evaluation approaches of IS investments into effect-assessing and effect-locating approaches. Combining this point of view with the research of Andresen (2002), which categorized 79 IT evaluation methods into financial, quantitative and qualitative approaches, we address a pool of potential methods in the context of process driven BVA of ERP solutions.

Case-based reasoning

Addressing the last key finding from our practice analysis we noticed that the companies' requirements for an ERP BVA based on reference experience from other projects indicate a number of similarities with the paradigm of Case-based Reasoning (CBR). The CBR approach was developed in the research field of artificial intelligence, particularly at the field of dynamic/reminding memory and situation patterns (Schank, 1982), and analogical reasoning (Gentern, 1983). CBR uses the specific knowledge of previously experienced concrete problem situations for dealing with new but similar situations (Aamodt, Plaza, 1994). The methodological framework of Aamodt and Plaza (1994) describes a CBR cycle. Starting from the appearance of a new case the framework retrieves in a first step the most similar cases from a knowledge database containing a number of previous cases. The reuse of this information leads to a suggested solution by combining the existing case with the new case. The suggested solution will be revised by applying it to the real world or evaluated by an expert. The evaluated or adapted case is regarded as a confirmed solution. This confirmed solution retains as a new knowledge case in the repository, extending the existing knowledge base.

PROCESS DRIVEN BUSINESS VALUE ASSESSMENT METHOD

Although the key issues pointed out in the interviews with ERP customers and vendors/providers have been repeatedly regarded in the prior IS research, we identified a large gap between the numerous theoretical paradigms and methodological approaches and its use in the real business environment. Motivated by the project objective, we developed methodological extensions to the existing body of knowledge aiming at an appropriate software-technical support of the process driven BVA of ERP solutions in the future. Here, we focused on the operational requirements of the SAP business consultancy.

Procedure model

In order to accommodate the above-mentioned key findings and identified ERP specifics, we developed a procedure model for a process driven ERP BVA method (cp. Figure 1). The procedure model follows the pragmatic, straightforward approach of the phase models. It describes a break down of the ERP BVA process into single, logical and chronological sequential or partially overlapping tasks (Schwarze, 1995).

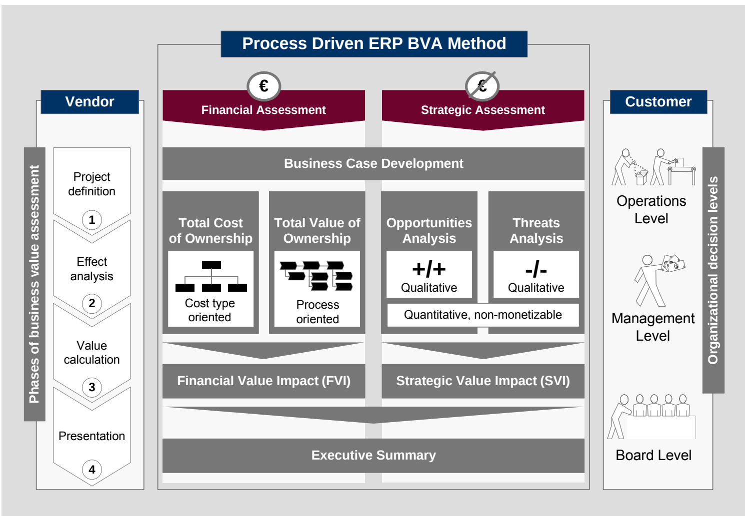


Figure 1. Process driven ERP BVA method: procedure model

The procedure model consists of four phases (cp. Figure 1) and represents a link between ERP vendor/provider and the ERP customer. During the 'project definition phase' the project scope, aim, and resources are determined. In the 'effect analysis phase' the ERP driven effects on the company's business processes are located and assessed, examining costs and benefits with and without conducting the ERP project. Based on the analyzed effects the business value of the ERP project is calculated in the 'value calculation phase'. Here, we differentiate between financial value impacts (FVI) and strategic value impacts (SVI). Financial assessment operates with monetary costs and benefits. Strategic assessment addresses qualitative, positive (opportunities) and negative (threats) aspects, as well as quantitative, non-monetizable aspects of an ERP implementation. Finally, these different value types are communicated to the customer's decision makers in the 'presentation phase'.

Due to the limited size of this paper we will focus the 'effect analysis' as the most important and critical phase of our method. The major methodological innovations also took place in this phase. The core of the developed process driven ERP BVA method is a unified reference database.

Unified reference database structure

The core of the process driven ERP BVA method is a unified reference database. At the present time, SAP stores the customer specific profitability analyses in different Excel files with variant structure. Although SAP, for example, conducted a large number of similar ERP projects, the structured collection and assessment of this data is difficult. Taking into consideration companies' specific situations, ERP value assessment is often time and resource consuming. The developed unified structure is capable of storing essential reference data for future ERP projects. This idea goes back to the CBR concept described in the previous section. The experiences and figures of previous projects and other reference data are used as a starting point for the customer specific IT profitability analysis. The data related to a BVA process is continuously stored and updated within the unified data structure.

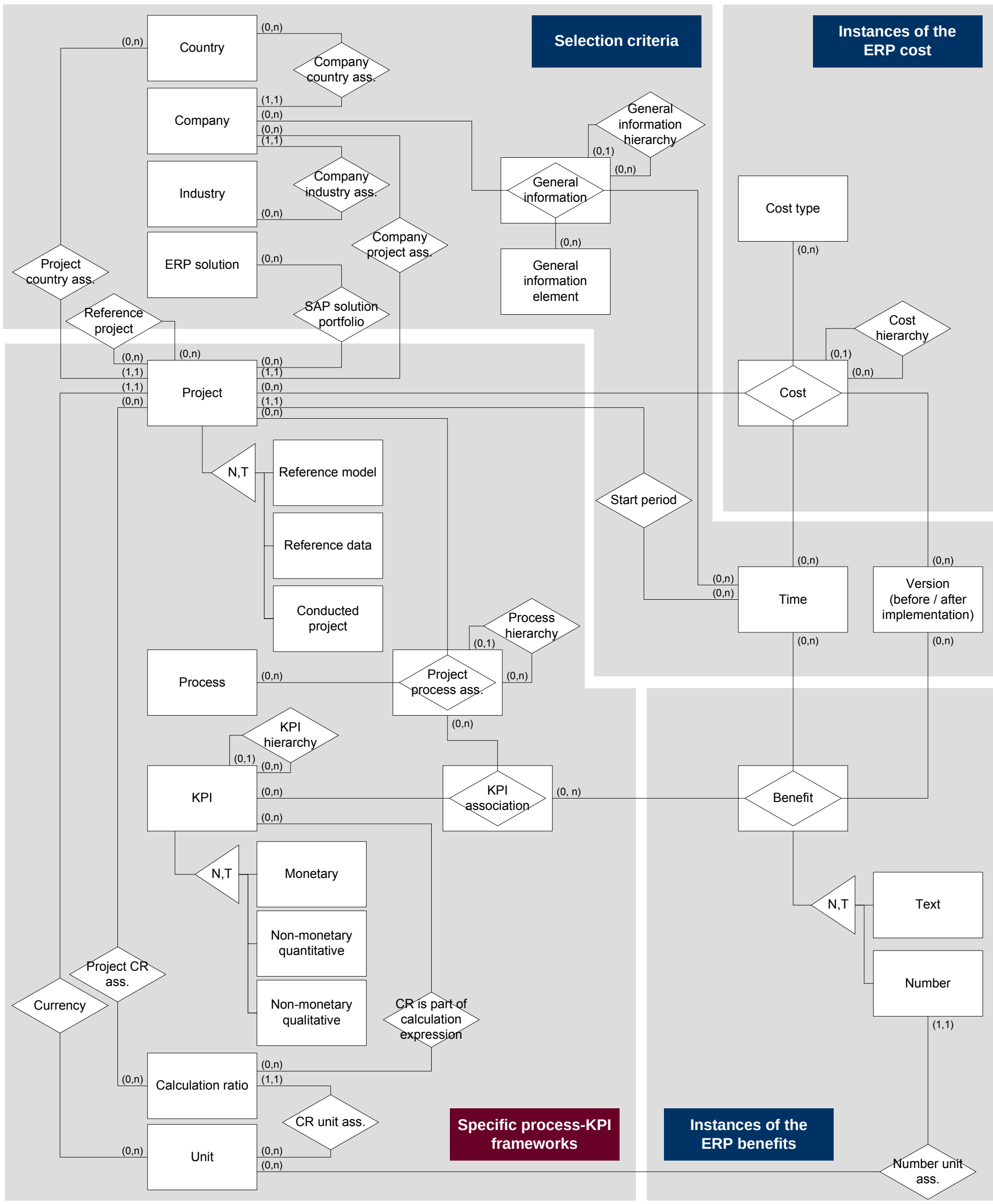


Figure 2. ERM of the reference database (selection)

The conceptual model of the developed reference database is displayed in Figure 2. For formal representation we chose the Entity Relationship Model (ERM) (Chen, 1976; Becker, Schütte, 2004). For better readability Figure 2 is divided into four clusters. The cluster 'specific process-KPI frameworks' can store 3 data types of the so-called reference projects (cp. Table 1). Processes with associated process performance metrics or key performance indicators (KPI) and used calculation ratios are assigned to a particular reference project. The environmental parameters in the cluster 'selection criteria' are important for further case(s) selection. The single instances of the ERP costs and benefits of a particular reference project are stored in two other clusters.

	Data type	Level of detail	Information source
1	Conducted project	Very detailed monetary and non-monetary figures for each period	Projects conducted by SAP BC or other SAP partners
2	Reference model	Detailed documentation of process and/or KPI systems	Handels-H, SCOR, SAP solution maps, IDS reference models, Y-CIM
3	Reference data	Selected information about relative process performance improvements	Public information, success stories, company reports etc.

Table 1. Three data types of the reference projects

The reference database is used for two different purposes (in the procedure model phase 'effect analysis'):

process driven location of the ERP implementation effects and

process driven assessment of the ERP implementation effects.

Process driven location of the ERP implementation effects

As assumed in the previous section, an ERP implementation influences performance of certain processes. Therefore, we capture the affected business processes and define appropriate metrics for further assessment of the ERP driven changes. The formalization of the identified ERP implementation effects is supported by the so-called 'customer specific process-KPI framework' (cp. Figure 3). The processes have to be stored in a hierarchical order to allow the later aggregation of the assessed effects. To each process/sub-process one or more KPI or KPI system is assigned. The process performance can be measured by monetary and non-monetary KPIs. The value of non-monetary KPIs can be measured quantitatively (duration, weight etc.) or expressed by means of qualitative aspects. The imaginary example in Figure 3 illustrates the structure of a customer specific process-KPI framework. Process hierarchies with assigned KPI structures are used to asses the benefits of planned SAP solution implementations. Here, the customer - a car manufacturer - considers the implementation of the mobile barcode infrastructure in combination with SAP Warehouse Management.

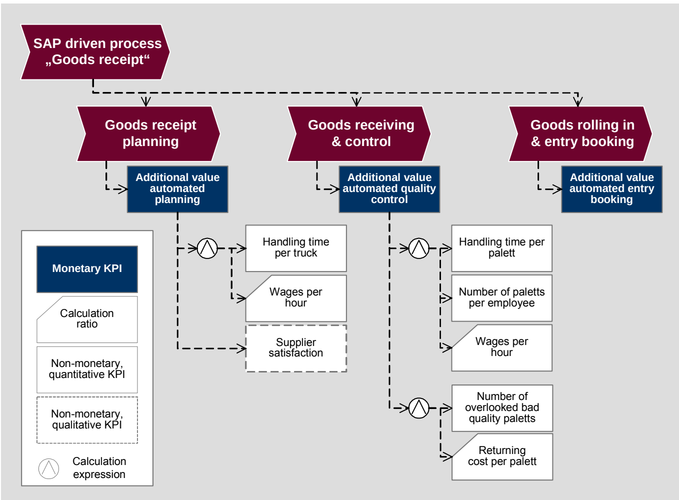


Figure 3. Specific process-KPI framework (selection)

Without the reference database, the development of such a process-KPI framework with a necessary level of detail can be very cost intensive. The reference database (cp. Figure 2) supports the specification of a process-KPI framework in two ways. On the one hand, the use of the existing reference process models (cp. Table 1) can significantly accelerate and simplify the domain-specific know-how transfer (Becker, 2001; Becker, Schütte, 2004; Hay, 1996; Scheer, 1994). On the other hand, a comparable reference project can be used as a template for a new process-KPI framework.

Process driven assessment of the ERP implementation effects

In the next step, the project specific process-KPI structure is assessed in terms of particular benefits. For this purpose the data about the relative improvements of the selected KPIs can be retrieved from the reference database. In the database these values are captured in percent as well as the absolute values (in case of the reference project type 'conducted projects') and stored as the instances of the ERP benefits of the previous reference projects (cp. Figure 2). Figure 4 illustrates the use of these values for reference based process improvement analysis. The figure depicts a non-monetary quantitative KPI 'handling time per pallet' from the specific process-KPI framework (cp. Figure 2). The selection criteria (cp. cluster 'selection criteria' in Figure 2), in particular process-KPI combinations, define the filtering of the reference database. Figure 4 exemplarily shows 2 references, which are used to calculate average relative improvements of the selected KPI. Three scenarios (conservative, realistic and optimistic) are hidden due to better readability. Entering the as-is KPI value for period 0 (before ERP implementation) initializes the calculation of the expected KPI improvement for the periods after the ERP implementation (periods 1 and 2). Further, the KPI is monetizable using customer specific calculation ratios.

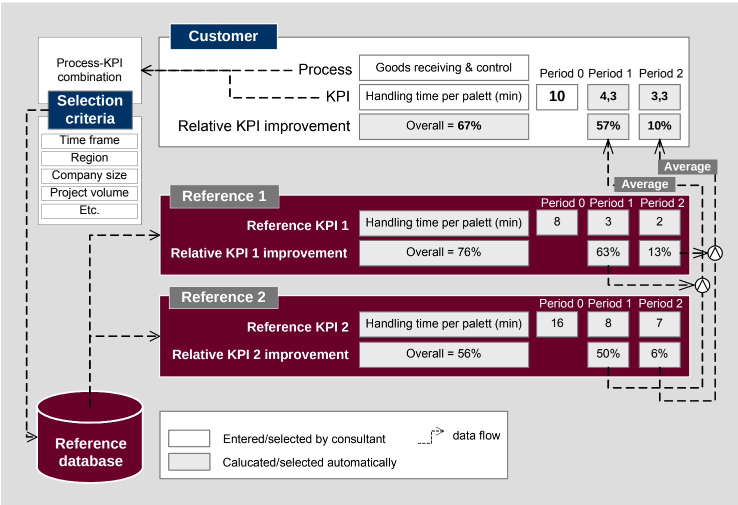


Figure 4. Reference based process improvement analysis

Total value of ownership (TVO) and total cost of ownership (TCO)

The process driven location and assessment of the ERP implementation is a part of the TVO analysis. The TVO concept goes back to McKinsey (Dempsey et al., 1998). In our method, every IT impact on the process level we understand as a positive (process performance improvement resp. process cost reductions) or a negative benefit (process performance deterioration resp. process cost increase). These IT impacts can be represented by monetary and non-monetary quantitative/qualitative (opportunities and threats) values.

The assessment of different cost types over the whole life cycle of the ERP system is based on the TCO model of SAP (Ullerich, 2004). The model has seven cost types on its first level: hardware/software investment, implementation, hardware/software ongoing costs, operations, continuous improvement projects, upgrade projects, and end-user usage. The SAP TCO model has up to 4 cost type levels. The cost type structure is also stored in the reference database (cp. Figure 2). Originally, the TCO model was developed by the Gartner Group in 1987 (Wild, Heges, 2000). The core idea of the TCO concept is the consideration of IT costs over the entire lifetime of an IS. Similar approaches are available from the MetaGroup, Forrester Research, Fraunhofer-IAO, Giga Information Group, GSM Software Management or International Data Corporation (Wild, Heges, 2000).

The calculated TVO und TCO are the basis for the next procedure model phases 'value calculation' and 'presentation'.

SAP VALUE ANALYZER

SAP Value Analyzer is a software-technical implementation of the process driven ERP BVA method with a focus on the ERP solution of SAP. The software prototype is a three-tier desktop application, which is built upon common used Microsoft software and therefore does not require additional licensing:

The data tier has been developed with Microsoft Access. It stores the complete data related to the BVA process.

The application tier has been developed with Microsoft's C# and runs in a Microsoft.NET environment. The application tier (cp. Figure 5) is the graphical user interface for entering and editing BVA related data:

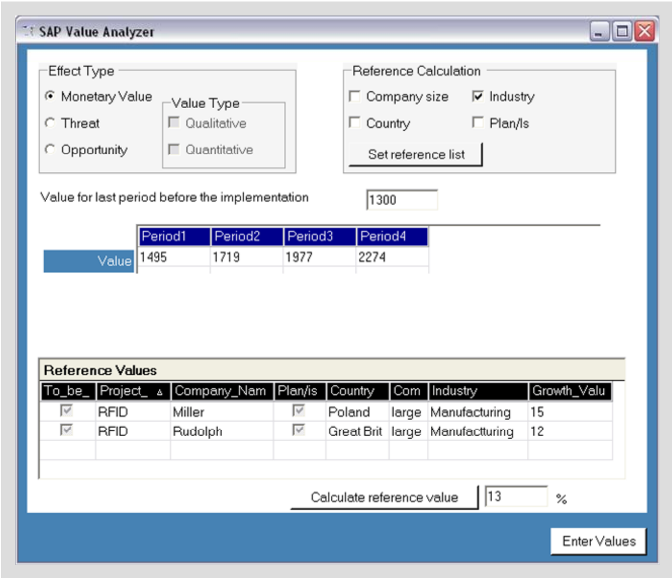


Figure 5. Application tier of the SAP Value Analyzer

The presentation tier has been developed in MS Excel. Processed data is passed to dynamically configured Excel templates. The customer specific analysis can be stored as a separate Excel file and handed out to the customer. Confidential data of other customers and internal calculation rates remain in SAP's internal reference database (cp. Figure 6).

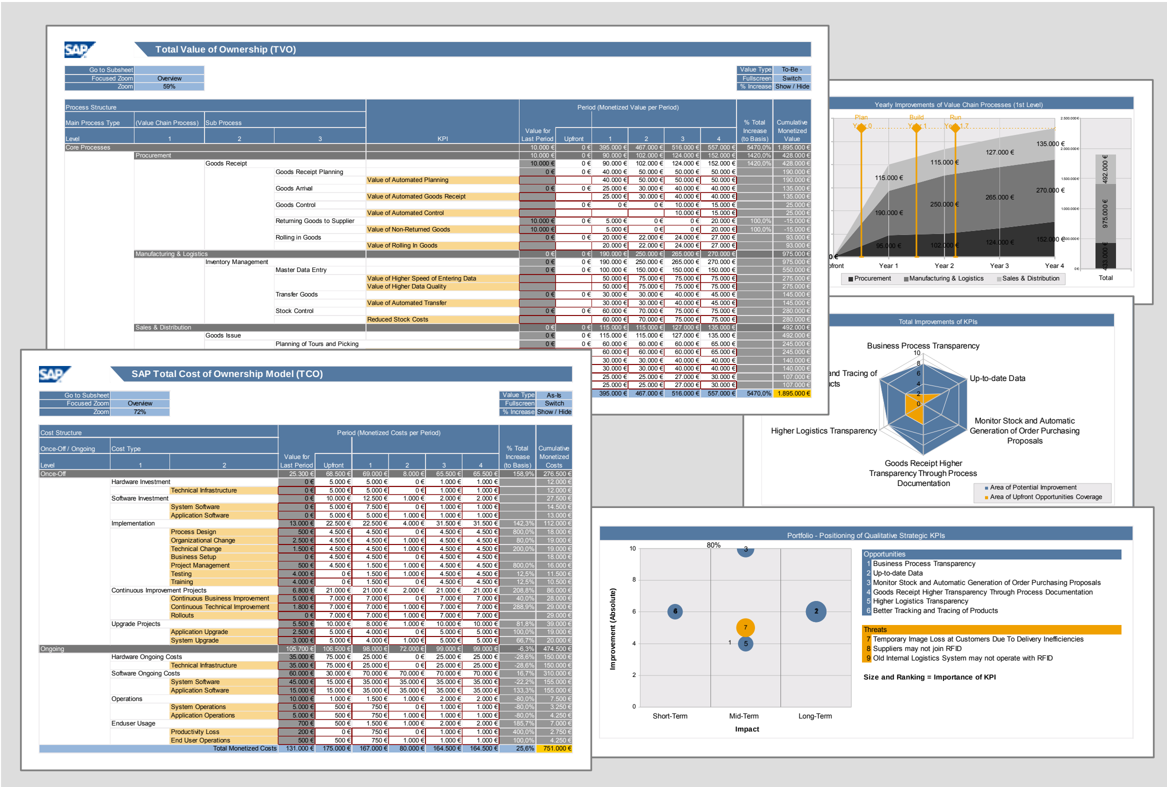


Figure 6. Presentation tier of the SAP Value Analyzer

Usage scenarios of the SAP Value Analyzer

The primary user groups of the SAP Value Analyzer are:

SAP consultants (BVA of SAP solutions in preliminary stages of ERP projects),

SAP customers (decision making regarding new ERP investments).

During process analysis on the operational level, SAP consultants create a process model and corresponding KPI structures with the necessary level of detail (cp. Figure 3). This step is supported by using the existing reference process models and reference KPI systems (cp. Table 1), which are stored in the database of the SAP Value Analyzer (cp. Figure 2). The relative improvement of KPIs can be based on reference values from the database or has to be estimated anew (cp. Figure 4 and 5). The process based aggregation of particular improvements leads to the TVO calculation. The TCO calculation is based on the existing SAP TCO models. Furthermore, the structured SVI analysis is supported (Cp. Figure 1). Finally, the results of the BVA process on a highly aggregated level are presented in an executive summary (cp. Figure 6).

Furthermore, the SAP Value Analyzer can support project managers and IT controlling during the long-term monitoring of ERP investments. After an SAP implementation the relevant data can be stored in the database and compared to data before the SAP implementation. The findings of such an analysis continuously improve the quality of the reference database.

The usage of the SAP Value Analyzer is not restricted to a particular industry, e. g. automotive. Reference data of the SAP Value Analyzer can be extended for usage across other industries. Moreover, the SAP Value Analyzer allows the SAP consultancy to profit from its cross-industry experience. Due to process oriented capturing of SAP benefits, the reference values can be obtained not only within one particular industry. For example, the information about achieved improvements in the retail industry can be useful for supply chain management concepts of the automotive industry. Finally, multilingual functionality of the SAP Value Analyzer enables the adoption of the SAP Value Analyzer in other SAP offices worldwide.

LIMITATIONS AND FURTHER RESEARCH

Although we have endeavored to develop a unified method for process driven BVA of ERP solutions with the high level of validity, the appropriate application of the method and of the supporting software requires following conditions:

an ERP implementation is not exclusively motivated by long-term strategic and/or political factors,

an ERP solution portfolio/configuration proposed by a customer is approximately defined,

a customer is supervised by a competent partner with a longtime ERP implementation experience, which invested necessary personal and financial resources in continuous development of the reference data base,

a customer is willing to develop a business case before the proposed ERP implementation,

general conditions of the market are varying in an expected range.

The research work presented in this paper addresses a number of general conceptual limitations of existing ERP BVA methods. However, we still require additional research regarding following conceptual defiancies:

interrelations between correlating process hierarchies and KPI hierarchies,

multiple capturing of ERP benefits,

definition of the best suitable selection criteria for the reference values,

semi-automated definition of the process metrics to measure ERP impacts.

Particularly, the semi-automated definition of the ERP impact metrics can be an important contribution to the existing body of normative IS literature. Our future research will concentrate on improving our method, including the application of this new approach in practice, assessing its applicability and efficiency in real world scenarios. For this purpose a number of additional evaluation projects with SAP customers in Germany and Russia are planned within the next year. Furthermore, we intend to prove adaptability of our method for usage across other industries.

REFERENCES

Andresen, J. L. (2002) How to select an IT evaluation method - in the context of construction. Retrieved 2005-05-24 from http://itc.sicx.net/data/works/att/w78-2002-16.content.pdf.

Bancroft, N., Seip, H., Sprengel, A. (1998) *Implementing SAP R/3 - How to introduce a large system into a large organization.* Greenwich CT.

Becker, J. (2001) Referenzmodel. In Mertens, P., Back, A., Becker, J., König, W., Krallmann, H., Rieger, B., Scheer, A.-W., Seibt, D., Stahlknecht, P., Strunz, H., Thome, R., Wedekind, E. (Eds.), *Lexikon der Wirtschaftsinformatik*. Berlin, 399-400.

Becker, J., Schütte, R. (2004*) Handelsinformationssysteme* (2nd ed.). Frankfurt/Main: Redline Wirtschaft.

Becker, J., Kugeler, M., Rosemann, M. (Eds.). (2003) *Process Management: A Guide for the Design of Business Processes*. Berlin, Heidelberg, New York: Springer.

Buxmann, P., König, W. (1996) *SAP R/3-Einführungsstrategien – Eine empirische Untersuchung*. Institut für Wirtschaftsinformatik, Johann Wolfgang Goethe-Universität Frankfurt.

Chen, P. P.-S. (1976) The Entity Relationship Model - Toward a Unified View of Data. *ACM Transactions on Database Systems, 1* (1), 9-36.

Davenport, T. H. (1993) *Process Innovation: Reengineering Work Through Information Technology*. Boston, MA, USA: Harvard Business School Press.

Dempsey, J., Dvorak, R.E., Holen, E., Mark, D., Meehan, W. F. (1998) A hard and soft look at IT investments. *The McKinsey Quarterly, 1*.

Eschinger, C., Pang, C. (2004) *Market Trends: Enterprise Resource Planning, Worldwide, 2003-2004* (Executive Summary), Gartner Group.

Esswein, W. (1993) Das Rollenmodell der Organisation. *Wirtschaftsinformatik, 35* (6), 551-561.

Farbey, B., Land, F., Targett, D. (1993) *How to Evaluate Your IT Investment: A study of methods and practice*. Oxford: Butterworth Heinemann.

Farbey, B., Land, F., Targett, D. (1999) Moving IS evaluation forward: learning themes and research issues. *Journal of Strategic Information Systems, 8* (2), 189-207.

Gaitanides, M. (1983) *Prozessorganisation. Entwicklung, Ansätze und Programme prozessorientierter Organisationsgestaltung.* München, Vahlen.

Gentner, D. (1983). Structure mapping - a theoretical framework for analogy. *Cognitive Science, 7* (2), 155-170.

Gordon, V. S., Bieman, J. M. (1993) Reported Effects of Rapid Prototyping on Industrial Software Quality, *Software Quality Journal, 2*, 93-110.

Hammer, M., Champy, J. (1993) *Reengineering the Corporation. A Manifesto for Business Revolution* New York, NY, USA: Harper Business.

Hay, D. C. (1996) *Data Model Patterns - Conventions of Thought*. New York.

Holland, C. P., Light, B. (1999) A Critical Success Factors Model for ERP Implementation. *IEEE Software 16* (3), 30–36.

Laudon, K. C., Laudon, J. P. (1998) *Management Information Systems – New Approaches to Organization & Technology*. (5th ed.) London: Prentice Hall.

Martin, R., Mauterer, H., Gemünden, H.-G. (2002) Systematisierung des Nutzens von ERP-Systemen in der Fertigungsindustrie. *Wirtschaftsinformatik, 44* (2), 109-116.

Melville, N., Kraemer, K., Gurbaxani, V. (2004) Review: Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly, 28* (2), 283-322.

Mende, M. W., Brecht, L., Österle, H. (1994) *Evaluating existing information systems from a business process perspective.* Paper presented at the SIGCPR '94: Computer Personnel Research Conference on Reinventing IS: managing information technology in changing organizations. Alexandria, Virginia, USA, 289-296.

Mooney, J.G., Gurbaxani, V., Kraemer, K.L. (1996) A process oriented framework for assessing the business value of information technology (Reprinted from Proceedings of the sixteenth annual International Conference on Information Systems, pg 17-27, 1995). *Data Base for Advances in Information Systems 27* (2), 68-81.

Mukhopadhyay, T., Kekre, S., Kalathur, S. (1995) Business Value of Information Technology - a Study of Electronic Data Interchange. *MIS Quarterly, 19* (2), 137-156.

Muschter, S., Österle (1999) Investitionen in Standardsoftware: Ein geschäftsorientierter Ansatz zur Nutzenmessung & -bewertung. In Scheer, A.W., Nüttgens, M. (Eds.), *Electronic business engineering*, 443-468.

Nordsieck, F. (1934) *Grundlagen der Organisationslehre.* Stuttgart: Poeschel.

Nordsieck, F. (1972) *Betriebsorganisation. Lehre und Technik.* Stuttgart, Germany: Schäffer Verlag.

Parr, A. N., Shanks, G., Darke, P. (1999). *Identification of necessary factors for successful implementation of ERP systems*. Paper presented at the New Information Technologies in Organizational Processes: Field Studies and Theoretical Reflections on the Future of Work, IFIP, St. Louis, Missouri, USA.

Porter, M. E. (1985) *Competitive Advantage.* New York, NY, USA: The Free Press.

Remenyi, D., Money, A. H., Sherwood-Smith, M., Irani, Z. (2000) *The effective measurement and management of IT costs and benefits*. (2nd ed.) Oxford: Butterworth-Heinemann.

Renkema, T. J. W., Berghout, E. W. (1997) Methodologies for information systems investment evaluation at the proposal stage: a comparative review. *Information and Software Technology, 39* (1), 1-13.

SAP (2004) *Value based methods: Business value assessment, value assessment and business case, TCO - analysis*. SAP Business Consulting

Schank, R. (1982) *Dynamic memory; a theory of reminding and learning in computers and people*. Cambridge: Cambridge University Press.

Scheer, A.-W. (1990) EDV-orientierte *Betriebswirtschaftslehre. Grundlagen für ein effizientes Informationsmanagement.* Berlin: Springer.

Scheer, A.-W. (1994) *Business Process Engineering. Reference Models for Industrial Enterprises* (2 ed.). Berlin: Springer.

Scheer, A.-W., Habermann, F. (2000) Making ERP a success. *Communications of the ACM 43* (4), 57-61.

Schumann, M. (1993) Wirtschaftlichkeitsbeurteilung für IV-Systeme. *Wirtschaftsinformatik, 35* (2), 167-178.

Schwarze, J. (1995) *Systementwicklung: Grundzüge der wirtschaftlichen Planung, Entwicklung und Einführung von Informationssystemen.* Berlin: Herne.

Stahlknecht, P., Hasenkamp, U. (2005) *Einführung in die Wirtschaftsinformatik.* (11th ed.). Berlin: Springer.

Ullerich, T. (2004) *TCO-Modell für SAP-Systeme am Beispiel mySAP CRM mit SAP Enterprise Portal*. Bonn: Galileo Press.

v. Arb, R. C. (1997) *Vorgehensweisen und Erfahrungen bei der Einführung von Enterprise-Management-Systemen dargestellt am Beispiel von SAP R/3*. Unpublished Thesis Dissertation, Bern.

Wallace, T.F., Krezmar, M.H. (2001) *ERP: making it happen: the implementer's guide to success with enter-prise resource planning*, (3rd ed.) New York, Chichester: Wiley.

Walter, S. G., Spitta, T. (2004) Approaches to the ex-ante evaluation of investments into information systems. *Wirtschaftsinformatik, 46* (3), 171-180.

Wegen, B. V., Hoog, R. d. (1996) Measuring the economic value of information systems. *Journal of Information Technology, 11* (3), 247-260.

Wild, M., Herges, S. (2000) *Total Cost of Ownership (TCO): Ein Überblick*. Giessen.

Transaction Cost, Cooperation Relationship and Boundary Spanning Activities for Selecting Third Party Logistics

Dr. Jao-Hong Cheng
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: jhcheng@yuntech.edu.tw

Shu-Nu Chan
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: g9323205@yuntech.edu.tw

Yu-Wei Chuang
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: g9323727@yuntech.edu.tw

Sheng-Wen Hung
National Yunlin University of Science & Technology
Yunlin, Taiwan.
Email: g9223705@yuntech.edu.tw

Abstract

By looking over the available literature about the manufactory and 3PLs in supply chain management, we found that there is an abundance of literature concerning the cooperation relationship and boundary spanning activities. The concept of taking the cooperation relationship as a medium that constructed boundary spanning activities was extracted from the previous research; hence the cause and effect framework was analyzed. On the other hand, transaction cost factors that cause companies' selfish benefits and risks dispersedly behaviors, such as environmental uncertainty, the frequency of contract noncompliance and companies' cooperation relationship with 3PLs, will change into boundary spanning activities. In other words, companies change the cooperation relationship with a specific 3PLs owing to take account of reducing their transaction cost and effective production.

The linear structural relationship (LISREL) model was used to analyze survey data collected from 102 who had adapted 3PLs services in Taiwan. Our results showed that the more companies notice the uncertainty, the more cooperation relationship and boundary spanning activities. The more transaction frequency happened, the less cooperation relationship. If the cooperation relationship between companies and 3PLs is strong, the fewer boundary spanning activities will be implemented.

Keywords:
Transaction Cost Theory (TCT), Boundary Spanning, Third Party Logistics (3PLs), Cooperation Relationship, Structural Equation Modeling (SEM)

1. INTRODUCTION

By the assistance of third party logistics, enterprises can strengthen their core business without spending too much time on transportation or warehousing of products. It not only can elevate the business performance, but also attain the synergy of integrated logistics of the whole supply chain; hence the whole industry possesses a competitive advantage. How the enterprises utilize the third party logistics effectively will then be the key issue in the phase.

Previous studies had discussed the issue of organization performance, and most of them separated cooperation relationship from boundary spanning activities. Better performances between organizations will ensue by launching the cooperation relationship. However, when self-interest and risk-aversion behaviors occur, organizations are more likely to change their partners, and then be engaged in boundary spanning activities.

On the basis of this reason, our study would like to analyze the impact of boundary spanning activities when choosing third party logistics through the intensity of the cooperation relationship (viewed as the intermediary dimension) and the transaction cost theory.

This paper adopts 1000 fronts from the manufacture industry with 3PLs in Taiwan, and is based on the construct of boundary spanning activities, cooperation relationship, and boundary spanning activities' implement level regarding as a think over of the main shaft. There are 6 significant hypotheses are in our empirical result.

1.1 Transaction Cost Theory

Transaction cost theory (Williamson 1975, 1985) offers another perspective to help us to better understand the forces shaping structure of supply chain. The basic premise of transaction cost theory (TCT) is that the firm will internalize activities that are able to perform at lower cost and will rely on the market for activities in which other providers have an advantage(Coase, 1937). TCT is built on a micro-analytic framework with strong behavioral reality. The members are assumed to be subject to bounded rationality. Furthermore, at least some actors are assumed to be opportunistic (i.e., having a tendency to cheat other parties) if given the chance. Imperfect or asymmetric, information may give such actors an exploitable advantage in their dealings with other parties.

For saving the cost, enterprises tend to expand thin size. This doesn't stop the internal expansion of organizations until the operational cost. Coase (1937) invoked that after production process being distributed according to separate skills, if it's achieved just through market exchange, then the cost would increase. The integration should be transferred into intra-organization management. This also means the range of an organization's production will be influenced by the cost of trading.

Williamson (1975) and Kogut (1988) argued that cooperation itself was an exchange behavior. When "cost of exchange" plus "cost" are smaller than "internalized cost", and the relative are willing to true the external resource into internal resource; hence the cooperation relationship established. Besides, the most important goal of cooperative relationship is to eliminate the cost and risk to avoid any unnecessary investment. From the view of trading cost, the align cooperation focused on the reduction of the trading cost. The substantial motivation includes: reducing costs, sharing risks, avoiding unnecessary cloudlet investments, and getting sizable economic gains.

In contemplating the relationship, between the literature of the cost exchange and the cooperation relationship was reviewed. Borrowing from the scholars' analysis and assumptions, the perspectives of exchanging cost are as follows: the character of assets, the uncertainty, and the frequency of exchanging. The variables mentioned above are substantial for explaining how the cooperation relationship ensues.

The unique characteristics for the assets are important, and the investment that carries out the cooperation relationship won't stop until the sustainable function of the assets is officially expired. If, and only if, the contract being invalidated or destroyed, the asset provider would first face difficulties when returning back the investment. Besides, the other party of the contract would suffer the loss because the ensuing situation that there is no appropriate target person to collaborate with. Therefore, the more unique the asset is, the more willingness for (full) both sides of the investment project are to keep the mutual cooperation relationship.

1.2 Boundary Spanning

Organizations are open systems that need to interact with and adjust to different kinds of external environments. Boundary spanning is the activity by which individuals within organizations bridge external needs and provide information for internal users (Scott, 1995). Boundary spanning activity is broadly defined as the amount of cross-boundary information that managers exchange. Those managers who cross the boundary between the organization and the environment are called boundary spanners. These boundary spanners, because of their multi-dimensional activities, have been called information gatekeepers, external representatives of the firm, conduits for resource acquisition, and influence agents (Aldrich & Herker, 1977; Ancona & Caldwell, 1992).In the microbusiness, the entrepreneur occupies the boundary spanning role. Therefore, the strategic activities in which the owner/operators engage manifest themselves as boundary spanning activities.(Dollinger, 1984). Adams(1976)listed five types of boundary spanning activity: filtering, transacting, buffering, representing, and protecting. Evan(1966) believed that by studying the boundary spanning activity of an organization, one could deduce its goal structure, strategy, and implementation plan. Organ(1971) viewed the boundary spanner as the "linking pin" between the organization and its environment.

1.3 Third Party Logistics

Susanne and Monica (2003) indicated that a third party logistics provider is an external provider who manages, controls, and delivers logistics activities on behalf of a shipper. This relationship can be formal or informal. The intention is that it should be a mutually beneficial and continuous relationship. The activities performed can include all or a part of the logistics activities but at least management and execution of transport and warehousing should be included (Berglund,2000).

A strategic alliance between the third party logistics provider and the client is often necessary to guarantee the quality of the performance (Andersson, 1995).

Typical services outsourced to third party logistics providers are transport, warehousing, inventory, value-added services, information services and design, and reengineering of the chain. (Andersson,1997;Berglund,1997).

1.4 Cooperation Relationship

The third middle-organization energy is from the supply chain system. Its aim is to offer better service for this entire network organization. It may include the substantial goods distribution and the collection of information, and processes of offering service to both sides in the trade; the relationship of cooperation is hence established. Cooperation means that different individuals work with each other for the shared goods. The ultimate purpose is to create mutual and multiple aspects. Kumar et. al (1995) has pointed out that when different other and the barricade or obstacles stand in between as the entrance appease to be essential, the motivation become strong enough to maintain their relationship of cooperation, it is when they feel that they are in the same side, the mutual understanding of the state of "mutual win" and cooperation is thin feet choice.

In this study, the definition of "cooperation" means the intimate working relationship between enterprises'. According to relevant research, the degree of close relationship differs from both sides on the scale. One side of extreme is very close, which is called strategic partner, while the other side is pretty loose, and therefore it is called the relationship of general market trade.

2. RESEARCH METHOD

This study used 1000 fronts from the manufacture industry in Taiwan as the research setting. The subjects are boundary spanners, product procurement managers or managers who have experiences of the 3PLs. 1000 questionnaires were sent to the investigation target. After two weeks, we adopted various means, such as, telephone calling, mail and direct visit to raise our response rate. Therefore, 193 questionnaires were sent back and the response rate was 19.3%. After eliminating 4 invalid questionnaires, 189 valid questionnaires were collected.

102 companies that had adopted 3PLs services took 53.96% in the 189 received questionnaires, and 87 companies that didn't implement 3PLs services took 46.04% of the valid questionnaires. Figure. 1 presented the conceptual model of our study. To our knowledge, few empirical studies ever attempted to examine cooperation relationship and boundary spanning activity relationship. Therefore, we adopted quantitative research approach. A quantitative survey was carried out to verify the proposed relationships among the variables.

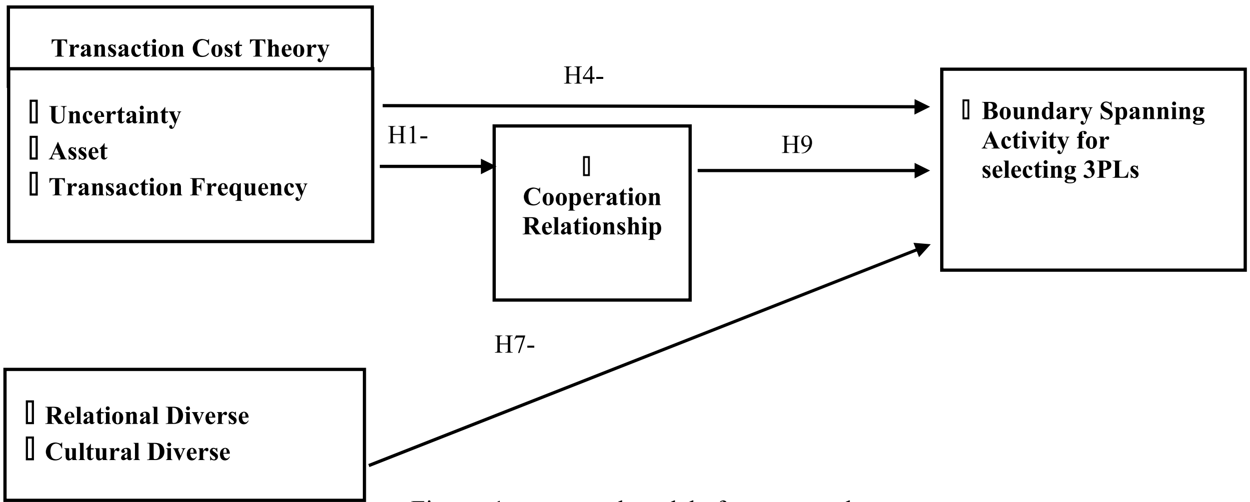


Figure. 1 conceptual model of our research

2.1 Hypotheses formulation

• The influence factors of transaction costs theory regarding cooperation relationship and boundary spanning activities

From this dynamism it reveals that some sort of uncertainty always exists inside the transactional network environment. Hallikas et al. (2002) deemed the uncertainty has an influence on the decision-making process, especially on the individual behavioral attitude. In addition to this, it is essential to be capable of evaluating transaction specific uncertainty in terms of the transaction costs and the option value of the partner as long as future profits of companies that depend on the future potential of partners. Specifically, we hypothesize:

H1: The higher the uncertainty, the stronger the cooperation relationship with third party logistics (3PLs).

Ancona and Caldwell (1988) referred that if the boundary spanner has stronger interactions and contacts with other groups, or even, owns more experiences on choosing partners, high frequencies of implementing boundary spanning activities will then occur. Hence negative effects on current cooperation relationship ensue. As a result, too many boundary spanning activity experiences will bring negative effects to cooperation relationship. Specifically, we hypothesize:

H2: The higher the asset of boundary spanners, the weaker the cooperation relationship with third party logistics (3PLs).

Stank and Daugherty (1997) indicated that logistics function use frequency does not have any significant relationship with outsourcing organization.

Concerns about opportunistic behavior on the part of even a few of a firm's potential alliance partners has ramifications for judgments regarding the potential behavior of all of the firm's alliance partners.(Weaver & Dickson, 1998). Specifically, we hypothesize:

H3: The higher the transaction frequency, the weaker the cooperation relationship with third party logistics (3PLs).

Environments differ in their degree of uncertainty, and different degrees of uncertainty require appropriate degrees of intelligence collection and information processing (Dollinger, 1984)

When the boundary spanners discover some environment uncertainties, they tend to adopt more boundary spanning activities.(Au and Fukuda, 2002) Specifically, we hypothesize:

H4: The higher the uncertainty, the more likely for companies to construct more boundary spanning activities.

People with more experiences on choosing other functional groups are more likely to participate in boundary spanning activities (Ancona & Caldwell, 1988)

Mendenhall, Kullmann, Stahl, and Osland (2002) gave some examples from the representatives. When they experience more cross boundaries tasks, they will contact with more culture groups and management experiences; thus, the more cross boundaries experiences will enhance the degree of boundary spanning activities. Specifically, we hypothesize:

H5: The higher the asset of the boundary spanner, the more likely for companies to construct more boundary spanning activities.

Transaction cost theory provides a direct link between a failure to live up to the behavioral expectations provided by an alliance contract and low quality outcomes (Williamson, 1991).Concerns about opportunistic behavior on the part of even a few of a firm's potential alliance partners has ramifications for judgments regarding the potential behavior of all of the firm's alliance partners.(Weaver & Dickson, 1998). Specifically, we hypothesize:

H6: The higher the transaction frequency, the more likely for companies to construct more boundary spanning activities.

• Other factors influencing boundary spanning activities.

The networking activities of expatriates can contribute to boundary spanning. This occurs when they build and maintain links with different groups of insiders and outsiders (Krackhardt, 1994). Social capital is a feature that is constituted by certain types of relationships among people, and its possession can facilitate their actions (Coleman, 1988). Because information is costly to collect, knowing the right contacts can help individuals to acquire information. Oftentimes, the contacts can also provide the gist of the information and point to trustworthy sources (Coleman, 1988). Boundary spanners must relate to a variety of groups that serve different function for the organization (Manev & Stevenson, 1996). In technical terms, having a variety of contacts means that an individual's social network contains diverse relationships (Blau, Ruan, & Ardelt, 1991). Specifically, we hypothesize:

H7: The higher the relationally diverse of boundary spanner for third party logistics (3PLs), the more likely for companies to construct more boundary spanning activities.

A culturally diverse network is conducive to boundary spanning for expatriates because different kinds of information are embedded within different cultural groups (Cohen, 1977; Smith, 1999). Yun (1988) found that immigrants in Hawaii were more able to locate crucial information when they had ethnically diverse networks. Specifically, we hypothesize:

H8: The higher the culturally diverse of boundary spanner for third party logistics (3PLs), the more likely for companies to construct more boundary spanning activities.

Van Weele (2002) referred that enterprises will purchase products from the supplier with a better relationship, or sound out other competitors frequently. Cooperation relationship means that the enterprise is willing to share confidential information and construct contracts with the supplier. When enterprises adopt cooperation relationship, they may as well be engaged in boundary spanning activities with the current or new supplier. Akacum and Dale(1995) indicated that four miniature companies in their study were all trying to build cooperation relationship with their major customers. In these cases, they were worried that the major customers will change the supplier. This shows that if the boundary spanner adopts more boundary spanning activities, the cooperation relationship will thus be endangered. Specifically, we hypothesize:

H9: The higher the cooperation relationship with specific third party logistics (3PLs), the poorer the performance of other third party logistics (3PLs).

3. RESULTS

Structural equation modeling was used to test the hypothesized relationships in our model. We performed the two-step approach as suggested by Anderson and Gerbing (1988) using LISREL 8.52. First, we reported the assessment of the measurement properties of the scales used in our model. Second, we focused on the estimation of the structural model and the testing of the proposed hypotheses.

3.1 Measurement Model Evaluation

The items were first factor analyzed to assess their dimensionality and measurement properties. An assessment of the eigenvalues and scree plot suggested a seven-factor solution (uncertainty, asset, transaction frequency, relationally diverse, culturally diverse, cooperation relationship, and boundary spanning activity) and all items loaded highly on the respective factor. In this study, cooperation relationship was measured with three global indicators of coordination (Y1) , flow (Y2) and common view (Y3), which were developed by summing across the three individual items, respectively. Boundary spanning activity was measured with three global indicators of spend time (Z1) , contact people number (Z2) and spend cost (Z3), which were developed by summing across the three individual items, respectively. Uncertainty was composed of (1); (2); and (3). Asset was measured with (1); (2). All measures were then subjected to a confirmatory factor analysis using LISREL 8.52. Specifically, a measurement model was estimated in which every item was restricted to its a priori specified factors, and the factors themselves were allowed to correlate (Gerbing and Anderson, 1988). All items performed well and were retained in the model. The chi-square of the measurement model was significant ($\chi 2$ =145.445, df=133). The large chi-square value was not a surprising result since the chi-square statistic in LISREL has been shown to be directly related to sample size (Doney and Cannon, 1997). Other goodness-of-fit indices less sensitive to sample size indicated that the model achieved an acceptable fit (GFI= 0.867; AGFI= 0.810; CFI= 0.976; RMSR= 0.038; RMSEA=0.042).

Table1. structural model fit

Fit benchmark	χ^2 / df	GFI	AGFI	RMSR	RMSEA	NFI	NNFI	CFI	IFI
Ideal	< 3	> 0.9	> 0.8	< 0.05	< 0.05	> 0.9	> 0.9	> 0.9	> 0.9
result	1.180	0.867	0.810	0.038	0.042	0.867	0.969	0.976	0.977

We used composite reliability to assess the reliability of the constructs. All of the composite reliability values, ranging from a low of 0.73 to a high of 0.93, exceeded the recommended cut-off value of 0.60 (Fornell and Larcker, 1981). Besides, all items loaded significantly and substantially on the latent variables (factors) were intended to provide the evidence of convergent validity. The average variance extracted was greater than 0.50 in all cases, which meant that the variance accounted for by each of the constructs was greater than the variance accounted for by measurement error (Fornell and Larcker, 1981). Discriminant validity was assessed by determining the confidence intervals of plus or minus two standard errors around the correlation between any two factors never included 1.0 (Anderson and Gerbing, 1988). Results provided strong evidence of discriminant validity among the constructs. In sum, the evidence suggested that all scales had adequate measurement properties.

3.2 Structural Model Evaluation

The relationships among the variables were assessed simultaneously via covariance analysis. Maximum likelihood (ML) estimation was used to estimate model parameters with the covariance matrix as data input. The ML estimation method has been described as being well suited to theory testing and development (Anderson and Gerbing, 1988). Results of the model indicated the model fit was acceptable (Table 1). The model showed significant chi-square value ($\chi 2$ =156.889, p = 0.077), which is not unexpected given the sample size. The model performed favorably on other fit diagnostics (GFI= 0.867; AGFI= 0.810; CFI= 0.976). The RMSR of 0.038 was also below the suggested value of 0.08 or lower (Hu and Bentler, 1999). Additionally, the RMSEA value of 0.042 indicated a close fit between the model and the data (Jöreskog, 1993).

With the exception of information technology, all other hypothesized relationships were supported. Uncertainty (H1:γ= 0.225, t = 2.012), asset (H2:γ= -0.193, t= -1.448), transaction frequency (H3:γ= -0.381, t= -2.850), Uncertainty (H4:γ= 0.466, t = 4.102), asset (H5:γ= 0.148, t= 1.220), transaction frequency (H6:γ= 0.045, t= 0.364), relationally diverse (H7:γ= 0.250, t= 2.164) were significantly associated with boundary spanning activity. culturally diverse (H8:γ= 0.230, t= 2.094) were significantly associated with boundary spanning activity. And the relationship between cooperation relationship and boundary spanning activity was a little significantly supported (H9:γ= -0.188, t= -1.654). Thus, the results supported H1, H3, H4, H7, H8, and H9. H2, H5 and H6 was not supported. Parameter estimates and t values for the hypothesized relationships are shown in Table 2. Overall, the model explains 26.7% of the variance in cooperation relationship and 34.8% of the variance in boundary spanning activity.

Table 2. hypotheses test result

	Hypotheses	Parameter estimates and *t* values
H1	The higher the <u>uncertainty</u>, the stronger the cooperation relationship with third party logistics (3PLs).	0.225(2.012)**
H2	The higher the <u>asset</u> of boundary spanners, the weaker the cooperation relationship with third party logistics (3PLs).	-0.193(-1.448)
H3	The higher the <u>transaction frequency</u>, the weaker the cooperation relationship with third party logistics (3PLs).	-0.381(-2.850)***
H4	The higher the <u>uncertainty</u>, the more likely for companies to construct more boundary spanning activities.	0.466(4.102)***
H5	The higher the <u>asset</u> of the boundary spanner, the more likely for companies to construct more boundary spanning activities.	0.148(1.220)
H6	The higher the <u>transaction frequency</u>, the more likely for companies to construct more boundary spanning activities.	0.045(0.364)
H7	The higher the <u>relationally diverse</u> of boundary spanner for third party logistics (3PLs), the more likely for companies to construct more boundary spanning activities.	0.250(2.164)**
H8	The higher the <u>culturally diverse</u> of boundary spanner for third party logistics (3PLs), the more likely for companies to construct more boundary spanning activities.	0.230(2.094)**
H9	The higher the <u>cooperation relationship</u> with specific third party logistics (3PLs), the poorer the performance of other third party logistics (3PLs).	-0.188(-1.654)*

note: in the parentheses is the t-value

*Represent $p<0.1$
** Represent $p<0.05$
*** Represent $p<0.01$

4. Discussions and Limitations

4.1 Discussions

Concerning hypothesis1 and hypothesis3, the uncertainty will affect companies' adopting of outsourcing services directly. This explains that the higher the uncertainty, the stronger the cooperation relationship with 3PLs; the higher the transaction frequency, the weaker the cooperation relationship with 3PLs. Therefore, the uncertainty and transaction frequency are important factors regarding the cooperation relationship with 3PLs.

In hypothesis4, the higher the uncertainty, the more likely for companies to construct more boundary spanning activities. As a result, uncertainty is a crucial factor for constructing more boundary spanning activities with 3PLs.

As for hypothesis7 and hypothesis8, the higher the relationally and culturally diverse of boundary spanner for 3PLs, the more likely for companies to construct more boundary spanning activities. Transparently, the interpersonal relationship and social contact ability of boundary spanner are of high importance.

In hypothesis 9, the higher the cooperation relationship with specific 3PLs, the poorer the performance of other 3PLs. This indicates that when companies with specific 3PLs with close cooperation relationship, the opportunity of choosing other 3PLs will decrease.

4.2 Limitations

The main foci of this study are boundary spanning activities and the analysis of choosing 3PLs. Therefore, we can only infer the factors of outsourcing. Hence, further research can follow our model to examine other selection activities.

5. CONCLUSIONS AND FURTHER RESEARCH

5.1 Conclusions

This study contributes to our understanding of how the factors affecting the quality of relationships between the cooperation relationship and boundary spanning activity in Taiwan. The research results indicate that cooperation relationship perceives, in order of relative importance, uncertainty, asset, and transaction frequency, as the major factors contributing to cooperation relationship. However, the research results also indicate that boundary spanning activity perceives, in order of relative importance, uncertainty, asset, transaction frequency, relationally diverse, culturally diverse and cooperation relationship as the major factors contributing to boundary spanning activity.

5.2 Further Research

Our study is based on the transaction cost theory framework, thus further research can study the outsourcing boundary spanning activities through other theories such as, social exchange theory, network theory, resource-based theory and more. On the other hand, our study focuses on the 1000 fronts of the manufacture industry in Taiwan. As a result, further research can verify in other industries or investigate cross industry differences.

REFERENCES

Adams, J. S. , (1976). *The structure and dynamics of behavior in organizational boundary roles*, In M. D. Dunnette(Ed.), Handbook of industrial and organizational psychology. New York: Rand McNally.

Akacum, A. and Dale, B.G.,(1995).Supplier Partnering: Case Study Experiences. *Journal of Supply Chain Management*, 31(1), 38-44.

Aldrich, H., & Herker, D.,(1977). Boundary spanning roles and organization structure. *Academy of Management Review*, 2:217-230.

Ancona, D. G., & Caldwell, D. F. ,(1988). Beyond task and maintenance: Defining external functions in groups. *Group and Organization Studies*, 13: 468-494.

Ancona, D. G.., & Caldwell, D. F. , (1992). Bridging the boundary: External activity and performance in organizational teams. *Administrative Science Quarterly*, 37: 634-665.

Anderson, J. C., & Gerbing, D. W.(1988). Structural Equation Modeling in Practice: A Review and Recommended two-step Approach. *Psychological Bulletin*, 103, 411-423

Andersson D.(1995). *Logistical alliances and structural change*. Tech. Lic. Thesis No. 470, Linköping University.

Andersson D. (1997).*Third party logistics—outsourcing logistics in partnership*. Linköping Studies in Management and Economics, Linköping University. Dissertation No. 34 (Doctoral dissertation).

Au, D.Y., and Fukuda, J.(2002). Boundary Spanning Behaviors of Expatriates. *Journal of World Business*,37,285-296.

Berglund M.(2000) *Strategic positioning of the emerging third party logistics providers*. Dissertation no. 45 (Doctoral dissertation), Linköping Studies in Management and Economics.

Berglund M. Third party logistics providers—towards a conceptual strategic model. No. 642 (Licenciate Thesis), Linköping University, Linköping Studies in Management and Economics, 1997.

Blau, P. M., Ruan, D., & Ardelt, M. ,(1991). Interpersonal choice and networks in China. *Social Forces*, 69: 1037-1062.

Coase, R.(1937), The Nature of the Firm, *Economica*, 4, 386-405. Reprinted in Oliver E.

Cohen, E. (1977). Expatriate communities. *Current Sociology* 24:5-133.

James S. Coleman, (1988). Social Capital in the Creation of Human Capital, *American Journal of Sociology Supplement* ,94,95-120.

Dollinger, M. J. ,(1984). Environmental boundary spanning and information processing effects on organizational performance. *Academy of Management Journal*, 27:351-368.

Doney, P. and J. P. Cannon (1997). An Examination of The Nature of Trust in Buyer-Seller Relationships, *Journal of Marketing*, 61(April),35-51.

Evan, W. M. ,(1966). The organization-set. In J. D. Thompson(Ed.), *Approaches to organization design. Pittsburgh*, Pa: University of Pittsburgh Press.

Fornell, C. and Larcker, D.(1981). Evaluating Structure Equations Models With Unobservable Variables and Measurement Error, *Journal of Marketing Research*, 18:39-50.

Gerbing, D.W., and Anderson, J.C.(1988). An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment, *Journal of Marketing Research*, 25, 186-192.

Hallikas, J., Virolainen, V. M. and Tuominen, M.(2002).Understanding Risk and Uncertainty in Supplier Networks--a Transaction Cost Approach, *International Journal of Production Research*, 40(15), 35-43.

Hu, L. and P. M. Bentler, (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives, *Structural Equation Modeling*, 6(1): 1-55

Jöreskog, K. G., and Sörbom, D.(1993) .*LISREL8.14: Structural Equation Modeling with the Simplis Command Language*, Chicago: Scientific Software International.

Kogut, B., (1988). Joint Ventures: Theoretical and Empirical Perspective, *Strategic Management Journal* ,9,319-332

Krackhardt, D. ,(1994). *Constraints on the interactive organization as an ideal type*. In c. Heckscher & A. Donnellon (Eds.), The post-bureaucratic organization: New perspectives on organizational change. Thousand Oaks, CA: Sage.

Kumar, N., Lisa K. Scheer, & Jan-Benedict E. M. Steenkamp(1995). The effects of perceived interdependence on dealer attitudes, *Journal of Marketing Research*, 32(August),356-384

Manev, I. M., & Stevenson, W. B. ,(1996). *Balancing ties: Internal and external contacts in the organization's extended network of communication*. Paper presented at the Academy of Management Meeting, Cincinnati, OH.

Mark K., and Dickson H.(1998) . Outcome Quality of Small To Medium-Sized enterprise-Based Alliances: The Role of Perceived Partner Behaviors, *Journal of Business Venturing*, 13, 505-522.

Mendenhall, M. E., Kuhlmann, T., Stahl, G. K., and Osland, J. S.(2002) . *A Review of the Expatriate Adjustment Theory Literature: Implications for Future Research and Practice*. In M. J. Gannon and K. L. Newman(Eds.) , Handbook of Cross-Cultural Management, 155-183. Oxford, UK: Blackwell.

Organ, D. W. ,(1971). *Linking pins between organizations and environment*. Business Horizons.

Scott, W. R., (1995). *Institutions and organizations*. Thousand Oaks, CA: Sage.

Smith, L. R. ,(1999). Intercultural network theory: a cross-paradigmatic approach to acculturation, *International Journal of Intercultural Relations*, 23,(4),629-658.

Stank, T. P.; Daugherty, P. J.(1997) .The impact of operating environment on the formation of cooperative logistics relationships, *Transportation Research Part E: Logistics and Transportation Review* ,33(1), 53-65

Susanne Hertz, Monica Alfredsson. (2003). Strategic development of third party logistics providers. *Industrial Marketing Management* 32:139-149

Van Weele, A. J.(2002) . *Purchasing And Supply Chain Management: Analysis, Planning and Practice*. United Kingdom: Thomson Learning.

Weaver K. M. & Dickson P. H. ,(1998). Outcome quality of small-to medium-sized enterprise-based alliances: the role of perceived partner behaviors, *Journal of Business Venturing*, 13, 505-522.

Weick, K. (1969). *The social psychology of organizing. Reading*, Mass.: Addison-Wesley.

Williamson, O. E.(1975) , *Markets and Hierarchies: Analysis and Antitrust Implication* , New York: The Free Press.

Williamson, O. E.(1985) , *The Economic Institutions of Capitalism*, New York: The Free Press.

Williamson, O. E.(1991) .Comparative Economic Organization: The Analysis of Discrete Structural Alternatives, *Administrative Science Quarterly*, 36(2),June, 269-296. Reprinted in Oliver E.

Yum, J.O.(1988) . *Social Network Patterns of Five Ethnic Groups in Hawaii*. In Y. Y. Kim and W. B. Gudykunst(Eds.) , Theories in Intercultural Communication, 239-258. Newbury Part, CA: Sage.

Approach for Enterprise Systems Implementation: Organizational fit Perspective

Dr. Vineet Kansal
Arab Open University, Kuwait
Kuwait

Abstract

The potential of the integrated information systems is widely accepted. Though, Enterprise Systems (ES) implementation is complex undertakings, many of the implementations turned out to be less successful than originally intended. The difficulties of ES implementations have been widely cited in literature. The objective of the study has been to address the key issues need to be focused for ES implementation. The questionnaire based responses from 92 organization of ES implementation were collected and result has been analyzed by the different stakeholders of ES project implementation. The study outlines the structured approach while having identified the important factors influencing ES implementation covering organizational context, business support to changed processes and risk management.

Keywords:
Enterprise Systems, ERP, Information System, Survey, Case Study

INTRODUCTION

As the pace of change accelerates in the twenty-first century as a result of technological opportunities, liberalization of world markets, demands for innovation, and continually decreasing life cycles, organizations are finding that they have to continuously re-adjust and re-align their operations to meet all these challenges (Motwani, et al., 2002). This pace of change has increasingly forced organizations to be more outward looking, market-oriented, and knowledge driven. A useful tool that businesses are turning to, in order to build strong capabilities, improve performance, undertake better decision-making, and achieve a competitive advantage is Enterprise Systems (ES) software.

However, it has been estimated that about half of ES implementations fail to meet expectations (Appleton, 1997). Other reported figures show that more than 70 percent of ES implementation fails to achieve their estimated benefits. According to a survey (Themistocleous et al. 2001), organizations adopting ES certainly acquire benefits such as an increase suppliers' and customers' satisfaction and an increase in productivity but the level of return on investment (ROI) is rather low. Other findings of the same survey suggest that many organizations adopting ES have serious conflicts with their business strategies and the majorities of ES projects are often characterized by delays and cost overruns.

The important issue concerning managers is the appropriateness of ES to meet organization's needs. The study discusses approach for ES implementation process, which can assist managers considering their ES projects and stimulate further research in the domain of ES acquisition issues.

METHODOLOGY

This paper is based on three sources of data. The first source is a questionnaire based responses from 92 organizations experienced ES implementation in India, the second is articles drawn from the web and the respected practitioners' magazines reporting ES implementation cases and the third is personal semi-structured and structured interviews conducted through e-mail and personal meetings with twenty one ES consultants and project implementation leaders. The size of the sample is rather small but not unusual for this kind of qualitative research. The interviews served as a means to test the validity of the theoretical framework proposed in this paper. The textual data of the transcribed interviews with the ES consultants were combined into single file and analyzed by calculating the frequency with which words or small phrases appeared in the text. It should be noted that the required validity of the findings of this analysis is mainly dependent upon the coherence of the interpretation (Weber, 1990). It should be also noted that Shang and Seddon (2000) have been argued, convincingly, that ES case studies reported in the trade press and the web can provide reliable data to conduct academic research and can be used as a starting point for understanding the benefits and the costs involved in ES.

DISCUSSION OF ISSUES IDENTIFIED

Given the organizational, technological and behavioral impact of ES, a broad perspective of ES adoption process is needed. Technological, business and organizational contexts should be studied in a unified way, which encourages the examination of interrelated key success factors. The certain issues specific to these systems have to be taken into account, such as unsuitability, most of the time, of ES modifications to meet institutionalized business operations and the extent of business processes re-engineering required prior to the implementation of the software. In traditional IS development theory, the software has to fit in to certain business processes, probably adopting and reproducing organizational inefficiencies. In ES implementation the reverse course is usually effective. Due to complexity of the system, enterprise prefers to adapt their business processes to software's in-built best business practices. Modification of the standard ES configuration options to fit business processes is costly, risky, time consuming and difficult (Devenport, 1996). In order to develop approach of ES implementation following issues should be considered.

ES Characteristics

ES is both strategic and operational. Strategic systems aim at making the organization more flexible and responsive to customer needs. The goal of rendering the enterprise readily adaptable to changing competitive conditions makes strategic systems fundamentally different from back office applications. The evaluation of strategic systems has to be based on cost (Clemons, 1991). Moreover, as has been argued by Kaplan, cited by (Clemons, 1991), financial techniques such as discounted cash flow, are constantly misused when applied to evaluation of strategic IT decisions, due to difficulty of quantification of the value of strategic systems. ES software's operational costs and benefits are more easily identifiable and quantifiable than the strategic ones.

As far as the strategic aspect is concerned, a key factor is the identification of the degree to which the implementation of an ES contributes to business strategy of the organization (Fitzgerald, 1998). While this degree is difficult to assess in quantitative terms, a qualitative assessment is nevertheless possible by interviewing, for example, senior managers or by using other qualitative techniques such as Likert type scales and cognitive mapping. Various methods can also be employed for assessing the relative importance of alternative options. It has been argued (Clemons, 1991) that, sometimes, when alternative outcomes can be ranked in a structured manner, decisions that cannot be based on numerical data can be made rationally and analytically without having precise estimates of the individual courses of actions. Simulation, probability and sensitivity analysis can be extremely helpful under these circumstances. Although, some overlapping between strategic and operational factors is inevitable. For example, business processes integration is a strategic activity as long as (successful) integration has an impact on the competitive position of the organization, while at the same time is also operational as long as it results in cost reductions in the daily activities of the organization.

Stakeholders Analysis

A number of ES's stakeholders operate outside the organization's boundaries (customers, suppliers, business partners in the value chain). For achieving the full potential of ES, especially, under collaborative business structures such as Supply Chain Management (SCM), the co-operation between business partners is essential. Other contemporary business paradigms, such as Customer Relationship Management (CRM), require the cooperation with the organization's customers. In that respect, the evaluators should consider the impact of Enterprise Systems on external stakeholders, specifically the customers, suppliers and business partners. In fact, alliance is one of the strategic benefits incurring from ES implementations (Shang & Seddon, 2000). Successful ES cases demonstrate the importance of estimating ES's effects on external stakeholders.

Cost and Benefit Analysis

There is a high percentage of intangible costs and benefits. According to Brynjolfsson and Yang (1997), there is empirical evidence to suggest that up to nine-tenths of the costs and benefits of computer capital are embodied in intangible assets. Intangible assets are created by investments in software, training and organizational transformations induced by IT. These assets, although, not measured financially, have the potential of increasing the value of IT investments. The estimation of the value of these assets in monetary terms is clearly a very difficult endeavor. However, it is important that both tangible and intangible assets and hidden costs should be taken into account from the outset when considering ES projects. For example, reductions in transaction systems and technical support personnel, cost savings resulting from better inventory management or value chain optimization, and savings from not upgrading legacy systems can be calculated. Other benefits, such as perceived customer satisfaction and benefits arising from rapid decision making are more difficult to be calculated, but nevertheless existent.

Intangible or hidden and underestimated costs are also a major concern among ES specialists (Slater, 1998). Underestimation of the time it takes to implement an ES is very common in ES projects. Consultants' fees, personnel training, data conversion, software's integration testing and self developed software (Slater, 1998; Rosemann & Wiese, 1999) to name but a few, can be a very heavy burden on the budget for supportive activities. In fact, according to some estimates, services by ES support industry can exceed the initial software cost by a factor of seven to ten (Martin, 1998; Hecht, 1999). Other costs, characterized by a behavioral aspect, are difficult to be identified and estimated. Such costs, for example, include the lack of commitment to change, which can lead to a dysfunctional operating environment and user resistance resulting in increased operational costs.

Organizational Impact

ES implementation results in a major organizational change. A major implication of ES deployment is that it involves drastic changes in the organizational structure, business processes and the people of an organization. These changes are the source of both costs and benefits, tangible and intangible. The re-engineering exercise is undertaken with the aim of achieving the optimization and integration of business processes according to the software's inbuilt best practices. Thus, in so far change management is effective, competitive advantage and financial returns on investment are expected. However, the estimation of the effectiveness of change management is not straightforward as it is dependent on the analysis of many uncontrollable factors related to human resources and the psychological climate of the organization (Stafyla, 2000).

Product, vendor and support services evaluation

A careful attention while considering the selection of ES modules that support critical business functions and of any other needed additional application, such as for example SCM is must. Certain weighted criteria for the selection of vendor, product and implementation partner should be set and evaluated at this phase (Travis, 1999). According to a recent International Data Corp (IDC) survey (Moss, 2000), users, who implemented ES systems, rate the ability of the vendor to deliver the promised system on time and on budget as the most important issues involved in the ES process. Other important buying criteria are the scalability and flexibility of the ES solution and the confidence in both the solution and the provider.

Although every one of the established ES packages offers a broad functionality, they certainly exhibit individual strengths and weaknesses compared to individual business requirements. Certain packages are regarded as having an exceptional functionality in some of their modules, as is the case, for example, with PeopleSoft's Human Resources module. Other vendors are regarded as specializing in certain industries, supporting industry specific best practices, as for example SAP in Chemicals and Pharmaceuticals, Oracle in Energy and Telecommunications and Baan in Aerospace and Defense industries. The availability and functionality of additional applications to support current and future business needs such as SCM or CRM is an important factor in ES software selection. It should be also examined if the packages under consideration support a certain business practice or operation, which is considered critical, such as make-to-order or make-to-stock manufacturing. Certain characteristics, such as multi-language and multicurrency capabilities can be the key drivers for selection of an ES (Bancroft et al., 1998). Among other factors considered in selecting an ES is the availability of experts in the system, the partnering company that will assist in the implementation, the training courses available by the vendor or third parties as well as vendor's financial position and pricing models.

All-in-one vs. best-of-breed ES software

Enterprises searching for competitive advantage have the option of acquiring an all-in-one or a best-of-breed ES. Additional applications can be acquired from the vendor of the ES, from another vendor closely collaborating with the first, from a third party vendor, built-in-house or outsourced. An example of an all-in-one approach is provided by the multinational Colgate-Palmolive's SAP R/3 solution, which integrated the processes of the company and connected ten thousand users worldwide after a 5-year implementation effort. Worries about the risk of relying on only one vendor were put aside as the company was convinced that this integrated environment offers systems robustness and the additional required functionality to support the company's operations. On the other hand, in an attempt at achieving increased functionality, the Boeing Commercial Airplane Group has adopted the best-of-breed approach by implementing demand-forecasting software by i2 Technologies, ES by Baan and product data management by Structural Dynamics Research. It has been reported that industry watchers agree that about 80% of companies will adopt the all-in-one solution but the remaining 20% will demand best-of-breed applications from multiple software vendors (Stein, 1999). Obviously, substantial empirical work is needed in order to identify the merits of these two approaches.

APPROACH FOR ES IMPLEMENTATION

On the basis of the above discussion an approach for ES implementation has been proposed as follows (Figure 1).

Business Practices

Business practices cover the 'clarity of business vision' and 'requirement analysis' in order to understand the organization support for changed business processes.

Clarification of the business vision

The first phase of the proposed approach, the clarification of the business vision, is a starting point for Enterprise Systems acquisition. Investment in ES is a strategic action, which can have significant consequences for the competitive position of the organization. It has been argued that effective IT project implementation requires a clear business vision, which clarifies the organization's direction, the goals, and the business model behind the implementation of the project (Holland & Light, 1999). It was explained above that ES requires substantial business process re-engineering and as Davenport and Short (1990) have pointed out, the first step in IT enabled process re-engineering is to develop the business vision and process objectives.

Business requirements

Both current and future business needs, arising mainly from external competitive pressures, have to be balanced against various technological, work and organizational constraints. Companies engaging in e-commerce or supply chains operate in a sophisticated business and technological environment. In such cases, the effectiveness of ES, which span beyond traditional organizational boundaries, require collaboration between partners, coordination of decisions, as well as accurate and real-time information flow in a network of enterprises. There is a great likelihood that the examination of needs and constraints will reveal that for a successful ES implementation, a radical change in business processes, towards simplification and efficiency, must take place. Such is the case, for example, when developing systems with a customer perspective or adopting best practices from industry (Avison & Fitzgerald, 1995). It is also likely that ES acquisition will have to be postponed or rejected in view of the high risks involved (Stefanou, 2000). Therefore, a critical factor that should be considered at this stage is the desire and the commitment to continuous change not only by top management but also by the steering committee, the systems' users and by all members of the project's implementation team.

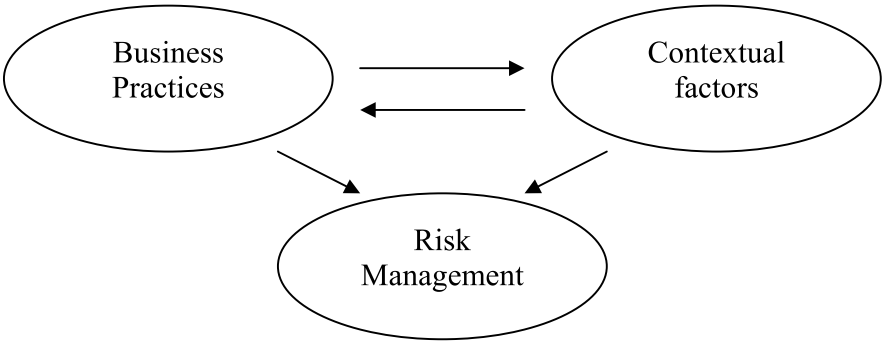


Figure 1. Approach for Enterprise Systems Implementation

Contextual factors

The decision concerning the adoption of an ES has to be made according to both the current and the future status of the enterprise, which is constrained by various technological, organizational and financial inefficiencies. Therefore, at this stage, a detailed critical ES functionality and enhancements requirements matrix, followed by a list regarding the organizational and technological changes required for the successful implementation of the ES system should be developed and evaluated according to certain criteria. This is, for example, exactly what Gillette India had to do with its re-engineering exercise. For this company, re-engineering for ES involves extracting the best finance, logistics and sales practices from multiple divisions and then standardizing them into an integrated ES system that could provide business consistency, such as credit terms, across all company's divisions. Martin, 1998 and Moss, 2000 after describing a Baan software implementation case, concludes that to achieve the goal of business support by implementing an ES system, companies should avoid the design of a system that the ES software is capable of providing but which is beyond the capabilities of the company to absorb as a daily routine. For companies wishing to achieve this goal, a well defined set of objectives and an on-going commitment to meeting them are essential from the outset of an ES project.

Risk Management

The risk management has been classified in five categories viz. technical, organizational, human, financial and time aspects as these constructs may proved to be substantial risk for ES implementation projects.

Technical aspects

Costs incurring from using multiple hardware and software platforms could be significantly reduced if there were a common IT architecture, including software and hardware platform, networking and communications, and applications development. Scalability and flexibility of the IT infrastructure is critical in order to support additional applications and systems and it should be assured before proceeding to the ES procurement process. Changes in the IT infrastructure may be necessary in order to support the ES system and any other additional applications. This poses another major evaluation problem, because the IT infrastructure is a supportive IT investment with no immediate measurable benefits by its own, but it still needs to be evaluated as far as alternative solutions or vendors are concerned (Fitzgerald, 1998).

Organizational aspects

These include, among others, the degree of the decentralization, the management structure, the style of leadership, the rigidity of business processes, and the company's culture. Resistance to change, prestige, job security feelings and departmental politics are also involved (Bancroft et al., 1998). It should be noted that organizational and cultural factors seem to be very important for successful implementation of ES and Supply Chain Management (SCM) systems (Stefanou, 1999).

Human resources aspects

A cross functional implementation team consisting of both business and IT people and of internal personnel and external consultants can be very effective in implementing ES software. However, the lack of experienced external consultants and trained and educated employees in ES philosophy represents a serious constraint that could jeopardize the implementation project.

Financial and time aspects

Any project of the scale of ES systems implementation should have adequate financial resources. A lot of hidden costs, such as the period of training required and unanticipated fees of external consultants, may prove to be a barrier to successful implementation. One final constraint is the time allowed for the selection and implementation process. Unrealistic time frames and deadlines may add unnecessary pressure and lead to project failure.

CONCLUSIONS

ES is seen as a technological artifact bundling material and symbolic properties in a certain recognizable form. The literature on ES implementation found unstructured and fragmented. A questionnaire based survey was conducted followed by interviews with the executives involved in ES project implementation. Five case studies were also prepared to understand the qualitative issues of ES implementation. The synthesis of quantitative and qualitative analysis leads to proposing the approach for ES implementation assuming ES project a socio-technical system. It describes that ES implementation process is heavily influenced by uniqueness of business processes, risk and contextual factors.

REFERENCES

Appleton, E. (1997). How to survive ERP, Datamation, 43, 50-53.

Avison, D.E., & Fitzgerald, G. (1995). Information System Development: Methodologies, Techniques and Tools, McGraw Hill, London.

Bancroft, N.H., Seip, H., & Sprengel, A. (1998). Implementing SAP R/3, Manning Publications Co., Greenwich, CT.

Brynjolfsson, E., & Yang, S. (1997). The intangible benefits and costs of computer investments: evidence from the financial markets, In proceedings of the international conference on Information systems, Atlanta, GA.

Clemons, E. K. (1991). Evaluation of strategic investments in information Technology, Communication of ACM, 34, 22-36.

Davenport, T. H., & Short, J. E. (1990). The new industrial engineering: Information technology and business process redesign, Sloan management Review, 31, 11-27.

Devenport, T. H. (1996). Holistic Management of megapackage change: the case of SAP, In proceedings of the second Americas conference on information systems, 51a-51c, Phoenix, Arizona.

Fitzgerald, G. (1998). Evaluating information systems projects: a multidimensional approach, Journal of Information Technology, 13, 15-27.

Holland C. P., & Light B. (1999, May / June). A critical success factors model for ERP implementation, IEEE

Software,30-35.
Hecht, B. (1999). Choose the right ERP software: a neglected technology?. Journal of Global Information Management, 6, 3-4.
Martin, M. H. (1998). An electronics firm will save big money by replacing six people with one and lose all this paperwork using ERP software, Fortune, 132, 149-153.
Moss, P. (2000). ERP implementation success in an SME company. ITtool box, (http://www.techrepublic.com).
Motwani, J., Mirchandani D., Madan, M., & Gunasekaran, A. (2002). Succesful implementation of ERP projects: Evidence from two case studies, International Journal of Production Economics, 75, 83-96.
Rosemann, M., & Wiese, J. (1999). Measuring the performance of ERP software – a balanced scorecard approach, In proceedings of the 10th Australasian Conference on information systems, Wellington, 773-784.
Shang, S., & Seddon, P. B. (2000). A comprehensive framework for classifying the benefits of ERP systems, In proceedings of the 2000 Americas conference on information systems, Long Beach, CA, 1005-1014.
Slater, D. (1998). The hidden costs of enterprise software, CIO Magazine. http://www.cio.com
Slater, D. (1999, February 15). An ERP package for you and you and even you, CIO Magazine.
Stafyla, A. (2000). Change climate, work team and leadership style: an empirical study on information systems users in Greece. In proceedings of the sixth Americas conference of Information Systems, Long Beach, CA, 1401-1406.
Stein, T. (1999, February 8). The great ERP debate- whether all-in-one or best-of-breed systems are best depends on your company's needs, Information Week, Issue 720.
Themistocleous, M., Irani, Z., & O'keefe, R.M. (2001). ERP and applications integration Exploratory Survey, Business Process Management Journal, 7(3), 195 – 204.
Travis, D. M. (1999). ERP Selection: the foundation for enterprise, APICS 9, 36-39.
Weber, R. P. (1990). Basic Content Analysis, 2nd edition, Sage Publications, Beverly Hills, CA.

Interception and Wiretapping Technology for Internet Telephony – Case Study on SIP-Based VoIP System

Roy Chaoming Hsu, Cheng-Ting Liu, Wen Ping Huang, Jen-Bon Tsai, and Jun-Long Tseng
*Department of Computer Science and Information Engineering
National Chia-Yi University, Chia-Yi City, Taiwan
rchsu@mail.ncyu.edu.tw
Department of Management Information Systems
National Chengchi University, Taipei, Taiwan

Abstract

One of the advantages of using Internet Telephony is the convenience and cost-effectiveness it provides. Besides, new generation of Internet Telephony can also provide advanced applications such as unified message service, video conference and multimedia on demand. However, Internet Telephony also brings some unnecessary evils due to its private characteristics of virtual and anonymous alias names or addresses. If a malicious user tries to make use of Internet Telephony for crime behaviours, such as black-mail, harassment, commercial spy, etc…, it will be hard to catch the users. Thus, similar technique for interception and wiretapping technique in PSTN system should be developed for Internet Telephony as well such that the information technology will not be abused. In the paper, we suggest an interception and wiretapping technique for SIP based VoIP telecommunication system. The mechanism of interception and wiretapping will be embedded into the SIP VoIP server system such that the SIP signalling and RTP voice packet can be analysed and forward to a monitoring server or handheld device (Notebook or PDA) for real-time inspection or a data storage for later forensics evidence via an embedded web server.

Keywords:

Internet Telephony, Cyber Forensics, Session Initiation Protocol, Interception and Wiretapping

INTRODUCTION

Owing to the advance of information and communication technology (ICT) and the popularity of Internet, Internet Telephony has become the most prevailing Internet service in the ICT era. Besides, utilization of Internet and the explosive growth of Internet services push the research of Internet Telephony and Voice over IP (VoIP) to leap forward. Different to the traditional Public Switched Telephone Network (PSTN) telecom system, users of VoIP communicate with each other and exchange voice signal by employing the Internet Protocol (IP), while the Internet Telephony is the physical realization of the VoIP system. Contrast to the traditional PSTN of separating voice and data network, Internet Telephony not only offer integrated voice and data communication services over the same IP network but also provide brand new and cost-effective services and lower the development and maintenance cost to Internet users. Usage of Internet Telephony only occupies very small bandwidth (about 5 to 15 kbps) of the information network and, once the Internet infrastructure and service are established, the cost for using the Internet Telephony is near free (computer call to computer call). The Internet Telephony not only provides basic voice communication, it also can offer value-added service such as integrated voice and video conference, short message, voice mailbox, file transfer and fax function etc… Being offering so many attractive services, the Internet Telephony has become the most cost-effective ICT services the normal user, small firms, enterprise and corporation would like to try.

The technique and system of Internet Telephony can provide more convenient and cost-effective voice, video and multimedia services to its users, however the call initiators often use anonymous alias names, IP addresses, or telephony numbers randomly selected by the Internet Telephony carriers such that it is not easy to monitor or track the users in real-time, not even to mention the wiretapping of the voice communication due to the characteristics of packet switching. It hence becomes the abuse of ICT invention if the users use the Internet Telephony to perform crime behaviours, such as black-mail, harassment, commercial spy, under-table transactions, traitor etc…. Thus, it is very important to apply the similar technique of PSTN's interception and wiretapping to the Internet Telephony such that the employment of new ICT to crime can be prevented in advance. Lawful Interception (LI) are regarded as the exercise of interceptions and wiretappings under the confinement of law. The function of interception and wiretapping provided compulsorily by the telecom carriers or service providers through government legislation such that the lawful interception can be exercised by the law enforcement agency have been on-going for many years. To perform normal operation in the US, telecommunication service providers are regulated by the Communication Assistance for Law Enforcement Act (CALEA, 1994) to offer interception and wiretapping method. Under similar circumstance of providing telecommunication services, the CALEA shall equally apply to the Internet Telephony service providers as well.

In the PSTN telecom system, the signalling and voice are transmitted within a fixed network and the bi-directional signalling and voice are transmitted over a dedicated wire of the fixed network during the whole conversation session such that the interception and wiretapping are easier. Interception can be done by crossing a second copper wire over the dedicated wire for the conversation in the local loop, while wiretapping can also be performed by adding recording equipments. The interception and wiretapping work can be carried out by the law enforcement agency by applied a LI permission from the court. Because the interception and wiretapping are done in the local loop of the traditional PSTN telecom carrier site such that the users of the conversion normally will not notice about the situation. Contrast to the traditional PSTN, based on packet switching IP network, the signalling and voice data packet of Internet Telephony are transmitted through different yet unpredicted routes. In some IP network which provides quality of service, the data packet can be guaranteed to transmit through a dedicated route using protocol such as Resource Reservation Protocol (RSVP) from one direction; however on the returning end the signalling and voice data packet might suffer from network teardown and re-establishment such that the routing paths could be varied. Besides, the convenient and important function of user mobility the Internet Telephony provides could increase the difficulties for interception and wiretapping. Location service of Internet Telephony enables the users to use the same alias name from different end-point with different IP address to communicate such that the method and technique of interception and wiretapping shall be able to monitor and detect the users registered from different Registration Server. In PSTN, the users hardly notice he/she is under interception and wiretapping while he/she is talking, yet users of the Internet Telephony are mostly skilful computer users and he/she might use sniffer tool to monitor the operation of packet flow, which could disclose the situation of interception and wiretapping. In this study, the proposed interception and wiretapping methods are designed and embedded in the proxy server of SIP-based Internet Telephony system. The arrangement of the paper is divided into 4 parts. Section I explains the importance and difficulties of interception and wiretapping method in Internet Telephony. The related work on the topic is described in section II. Section III illustrates the method of interception and wiretapping for SIP-based Internet Telephony system. Section IV exhibits the implementation results. And section V is the conclusion and further work.

RELATED WORK

The most acceptable protocols for VoIP service are H.323 (ITU, 2000) and SIP (Session Initiation Protocol) (Rosenberg, 2002). H.323, recommended by ITU for multimedia communication, provides an umbrella structure and method for Internet Telephony. Even though the first version of H.323 based Internet Telephony system was defined by the ITU since 1996, based on PSTN and LAN, H.323 needs to accommodate other ITU defined signalling and control protocols such as H.225 and H.245. Besides, lack of interoperability make H.323 difficult to communicate with other VoIP system. H.323 once was the most widely adopted VoIP protocol, but it lost its popularity since SIP was introduced by IETF as a Request For Comment (RFC). SIP is a signalling protocol which defines the initiation, modification and teardown of user interaction and multimedia sessions. SIP is an ASCII based protocol which is similar to HTTP and SMTP which provides the following 3 major functions

- The initiation, establishment, and teardown of conversion session.
- establishment of the flexible distributed system, and
- interoperation between other web-based applications

Because the simplicity and advantages of SIP, H.323-based Internet Telephony has been replaced by the SIP VoIP telecommunication system. Initiations of SIP protocol design and SIP VoIP research was proposed by Schulzrinne and Rosenberg (1998). In the next year, the RFC for SIP was formally proposed by IETF and published in IEEE Network (Schulzrinne and Rosenberg, 1999) since then numerous research works has been focus on the SIP based Internet Telephony Architecture and system. Bellcore has used Java Telephony API and SIP to design telecommunication system called ChaiTime (Anjum et. al, 1999), which was based on PBX (Public Branch exchange) system and can be used to integrate with other telecommunication service. Gbaguidi of the Swiss Federal Institute of Technology established a VoIP telecommunication system (Anjum et. al, 1999), which integrated different Java Bean package for the VoIP applications. The VoIP system developed by Cornell University, called ITX (Gbaguidi et at., 1999) is similar to the Gbaguidi's, which also use the mix-and-match strategy to design the VoIP telecommunication service. Different with Gbaguidi's is that the ITX adopt the "Template" model to develop a series of high level Java API such that the developers of the VoIP system can re-use the software component to integrate a VoIP system which fit the individual's requirement. Integration of VoIP service with wireless network is only recent. Rao, Lin, and. Cho (2000) have explored the feasibility of integrated VoIP service into the existing GSM system, which is called iGSM. In their study, the authors have investigated how to use H.323 terminals or the GSM handsets to access VoIP service and to realize the registration, voice data transmission, and teardown without changing the current GSM system. Recently, SIP-based VoIP systems with new and innovative services are proposed. A SIP-based VoIP system with basic function of multi-point call, multiple audio coding formats and complex functionalities such as call muting, call-hold and personal mobility support is proposed and developed by Zeadally, and Siddiqui (2004). Prasad, Hurni, and Jamadagni (2003) considered the limitations of existing SIP conferencing methods and proposed a distributed architecture using Controller (SIP Proxy Servers) and Conference Servers which facilitates the control and media handling of a VoIP conference. Another novel software architecture providing the connected multimedia services through networked embedded systems using SIP protocol are proposed by Kwak, *et al*. (2003)

Computer Forensics or Cyber Forensics is a technical term of the information technology era, which refers to the collection, inspection, and forensics of crime evidences committed by suspect(s) employing information technology and the Internet. The collected and inspected digital evidence will later be used as the crime evidences for appearing to the court. Research and discussion of Computer Forensics are only recent due to the increasing number of crime of information and communication technology. In [Cassey, 2000], the authors discuss in details about the digital evidences from the computer forensics which includes the theories of the networking, the acquisition of the digital evidence and how to apply the behaviour evidence in analysing the digital evidence. Several books on the topics of computer crime and forensics have been published in 2001 (Middleton, *et, al*. 2001). In (Middleton, 2001), the procedures and the software and hardware tools for collecting digital evidences are introduced. Another topic covered is on how to collecting digital evidence as legal evidence and, at the same time, not to violate privacy. In (Albert J., *et, al*. 2001), collection, inspection, and preservation methods, the confinement of federal and international law, and tools of computer forensics for cyber crimes are illustrated. General discussions on how to acquire, preserve, and restore the cyber crime evidence, and re-establish the crime event can be found in (Vacca, 2002). Not only the collections and analysis of digital evidence were discussed in the book, how to actively exercise information warfare was also proposed in this book.

Utilization of H.323 and SIP to create innovative telecommunication services have been proposed numerously, yet it is becoming more important to prevent the cyber crime committed by abusing the privacy and anonym characteristics the Internet Telephony provides. However, the existing research and professional book on computer forensics only provide general discussion of the procedures and tools for collecting digital evince, techniques of interception and wiretapping for Internet Telephony were not even mentioned or surveyed. In this study, an interception and wiretapping technique is proposed for the SIP-based Internet Telephony, which could be used by the law enforcement agency to collect the digital evidence and to prevent illegal behaviour committed by abusing the information and communication techniques.

METHOD OF INTERCEPTION AND WIRETAPPING

The scenario of interception and wiretapping of Internet Telephony can be classified into two types, which are online and off-line operations. Online operation refers to the situation when the law enforcement agency performs the interception and wiretapping while the users of Internet Telephony are still in the conversation state. The offline operation is to locate the users of the Internet Telephony first and when the users start the conversation, the system will perform interception and wiretapping automatically and record and store the audio data to remote storage devices, such as hard disk, for later analysis by the law enforcement agency. Major advantage of online operation is the real-time characteristics in obtaining the control signalling and voice data streams, however, the real-time online operation will increase the VoIP server's loading and reduce the efficiency such that the voice stream from the users end might suffer delay. The situation will become even more obvious when the law enforcement agency perform the interception and wiretapping from a remote area. Besides, to implement the real-time characteristics, the software in the server must be modified a lot, which might cause difficulties and increase the cost in the interoperability between different VoIP system and the integration between heterogeneous platforms. Offline operation is to make a complete copy of the control signalling and voice data stream to a recording and storage device such as hard disk or flash memory for later analysis. Advantage of using offline operation is that it can decrease the unnecessary loading of the online operation causes by consistently requesting voice data stream from the servers. The drawback of the offline operation is that the hard disk or memory space reserved for the recording is hard to predict due to the inherent characteristics of conversation. Besides, the offline operation will lost the real-time opportunity to prevent the crime in advance. Both scenario and structure of interception and wiretapping methods have its own advantages and drawbacks, and if the advantages can be integrated while the drawback is expelled, the interception and wiretapping system can possibly provide added value to the Internet Telephony system such that the computer crime can be prevented.

Realization of the Interception and Wiretapping for SIP-Based Internet Telephony

In this study, an integrated offline and online operation of interception and wiretapping method will be proposed to SIP based Internet Telephony. In SIP based VoIP system, the function of *proxy server* is to forward the SIP packet to the User Agent Client (UAC) of the callee and the RADIUS protocol usually can be added to provide communication to a *billing server* for authorization and conversation time record. *Registrar server* is to provide the registration and location query service for users or proxy server, whose function is similar to the DNS (Domain Name Service), however, the records in

DNS is permanent while the registration records in the *Registrar server* can be modified through registration mechanism from time to time. *Redirect server* is to support and assist the *proxy server* in redirecting the signalling packets. The function of *location server* is to act in-between the *redirect server* and user information database (UIDB), which stores the IP address, ID, Password and authorization information, etc…, to obtain the users information. To simply the structure, a SIP-based Internet Telephony within a local area network (LAN) is illustrated in Figure 1, where the various functions of servers are integrated into a single proxy SIP server. However, in reality the various functions of SIP servers may be distributed in the network to provide signalling service.

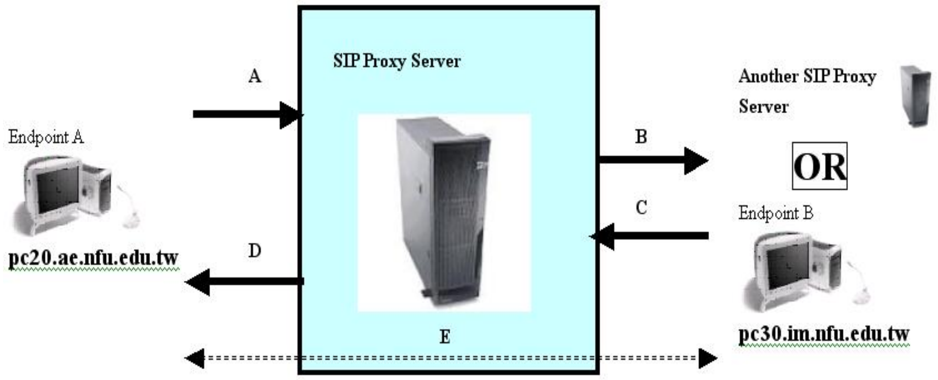


Figure 1. SIP-based Internet Telephony within a LAN, where various functions of SIP servers are integrated into a single proxy server.

The simplified SIP server can be described as the following:

Path A: User A (end point A with IP address pc20.ae.nfu.edu.tw) would like to dial to user B (end point B with IP address pc30.im.nfu.edu.tw) and initiate an INVITE to the proxy server

Path B: The proxy server will forward the INVITE to endpoint B after checking the UIDB (via the Registrar) and find out the address of user B. (Assuming the end point B is the user B, however, the endpoint can be another proxy server which will check and find the address of user B)

Path C: User B will reply the proxy server with a 200 OK.

Path D: The proxy server will forward the 200 OK along with the address of user B to the user A, and the signalling control is complete at this moment.

Path E: User A will establish another voice packet stream route with user B and the voice data will be exchanged along this data stream until one of the user hang on the IP telephone.

User the operation of SIP Internet Telephony, the proposed online and offline interception and wiretapping method are discussed.

Online Real-Time Operation

Under the above-mentioned SIP Internet Telephony structure, the online operation of interception and wiretapping method are proposed. The scenario can be illustrated as in Figure 2, while each step is described as below.

Step 1-8. Step 1-8 is standard SIP signalling control, the description is similar to the description from Path A-E in Figure 1.

Step 9. When notified by special signalling events within a proxy server, the inspector can plug in his handheld device (such as PDA) or access via wireless LAN to browse into the build-in web server of the proxy server 1 if he/she would like to perform interception and wiretapping. The inspector can monitor the on-going and to-be established signalling sessions and perform real-time analysis for the IP addresses flow to decide which suspicious users to be intercepted next.

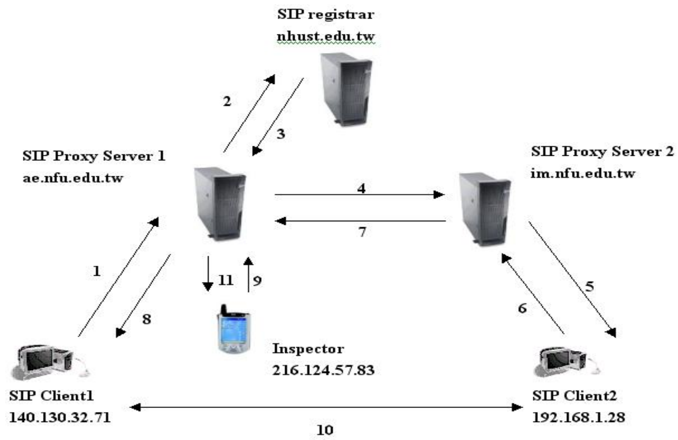


Figure 2. Online operation of interception and wiretapping method

Step 10. Path 10 is the RTP streaming channel for voice packet. Assuming the users of SIP based Internet Telephony rely on NAT (Network Address Translation) of the proxy server to access the Internet such that path 10 is physically via the SIP proxy server 1 (with domain name ae.nfu.edu.tw) to access the Internet. Similarly, the SIP proxy server 2 (domain name im.nfu.edu.tw) also provide user B to access the Internet using NAT. Here, the RTP streaming channel have established between proxy server 1 and 2 for caller (user A) and callee (user B), respectively and the RTP session can be start.

Step 11. The inspector can now use a handheld device or a notebook computer to locally connect the proxy server 1 to intercept the signalling and voice data stream while Client 1 and 2 are setting up call session or during a conversation such that the possible crime can be prevented.

Offline Operation

The most important steps in the offline operation are to store the specific control signalling and voice packet to a storage media, and the method of retrieval and presentation of the stored voice data. How the specific voice and signalling will be stored to a storage media under the offline operation is introduced as in Figure 3.

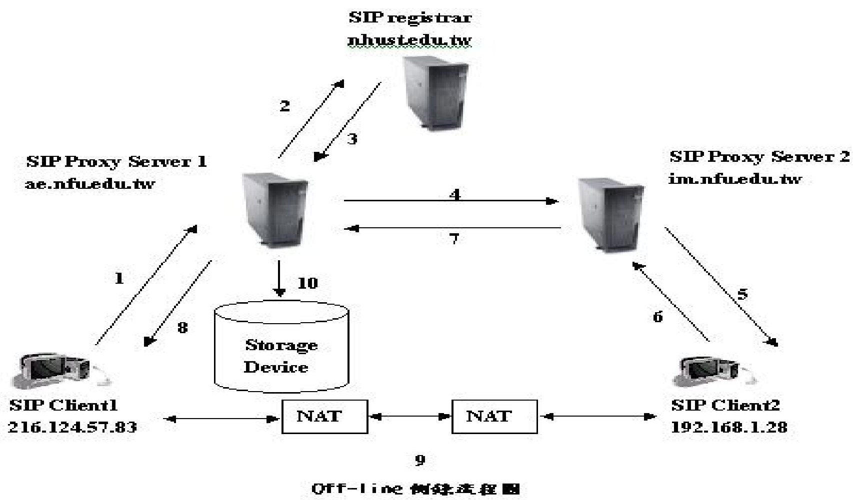


Figure 3. Data flow for offline operation of interception and wiretapping.

Step 1-8. Step 1-8 is standard SIP signalling control, the description is similar to the description from Path A-E in Figure 1.

Step 9. Step 9 is the RTP streaming channel for voice packet. Assumption of the proxy server with the NAT function is same as the Step 10 of online operation such that the voice stream is physically connected between SIP client 1, and 2 via proxy server 1 and 2, respectively.

Step 10. Whenever the suspicious users are required for interception and wiretapping and the IP addresses are located by the law enforcement agency, the system will automatically make a copy of the signalling and voice data flowing in and out of the located IP address and store to storage media. The storage media can be attached to the proxy server, located within the LAN or in a remote network, depending on the inspecting agency's requirement.

IMPLEMENTATION RESULTS

In the proposed interception and wiretapping method, function of the Registrar Server, Redirect server, Location server and Proxy server are integrated as a single SIP server which is shown in Figure 4.

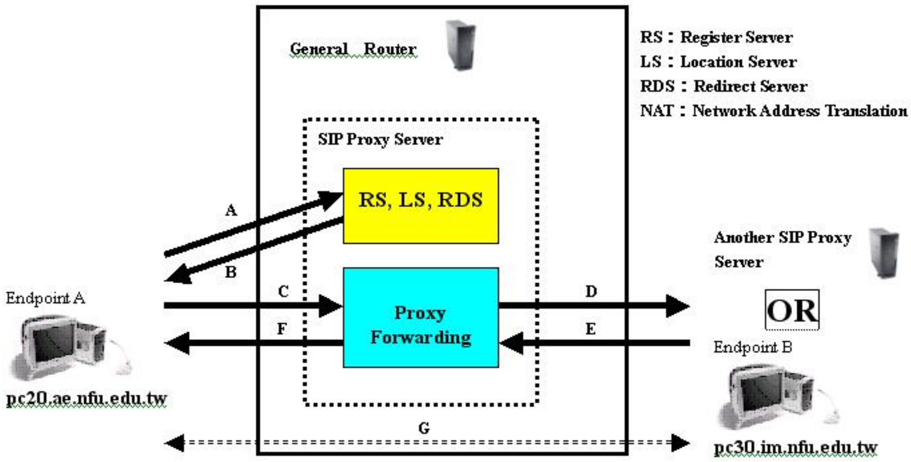


Figure 4. An integrated SIP server, as shown in the block diagram.

In the integrated SIP server, as shown in Figure 4, the function NAT server has been configured in the General Router to enable the function of DNAT (Destination NAT) and SNAT (Source NAT) such that the users can communicate with each other point-to-point by exchanging the RTP packet (as indicated in the path G) through the SIP server in the LAN. Without changing the hardware of the SIP server, the interception and wiretapping can be done by breaking the point-to-point communication channel G established for endpoint A and B and redirect to a third party. The NAT function has the benefit of preventing user from Internet to access the user in the LAN and it can also prevent user in the LAN to directly access to the Internet such that the path G in the Figure 4 can be physically break using the iptable command of the NAT, yet both users can still be virtually connect through the Gateway function of the router. Figure 5 shows the interception and wiretapping method and a web server embedded in the integrated SIP server. The web server is established to interactively provide user information of the SIP server. The original RTP path G is break into path G2 and G3 by the function of RTP interception and wiretapping, while the RTP content is redirected and copied to a recording device attached to the SIP server or located in the LAN through the function of RTP proxy. As indicated in Figure 5, the interception and wiretapping scenario can be described as below,

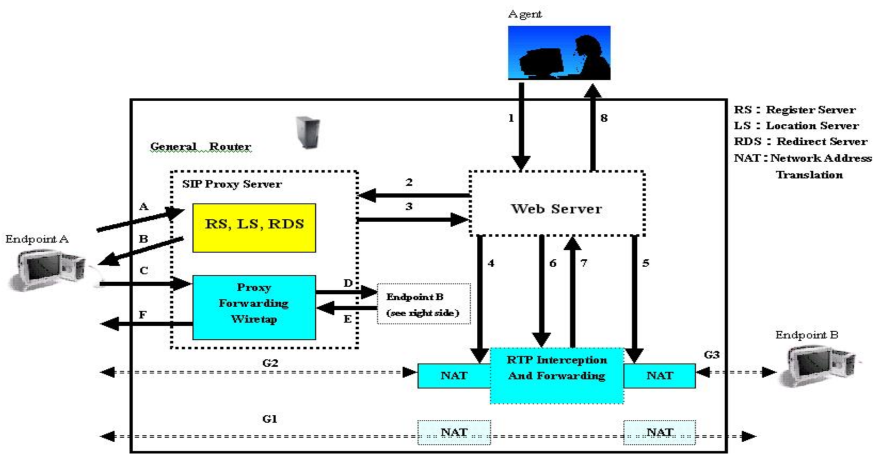


Figure 5. The implemented interception and wiretapping method and a web server as embedded in the integrated SIP server.

Path 1: The law enforcement agent orders a command to perform interception and wiretapping for a suspicious user through the web server after the authorization process as shown in Figure 6.

Path 2: Web server requests the SIP server for interception and wiretapping.

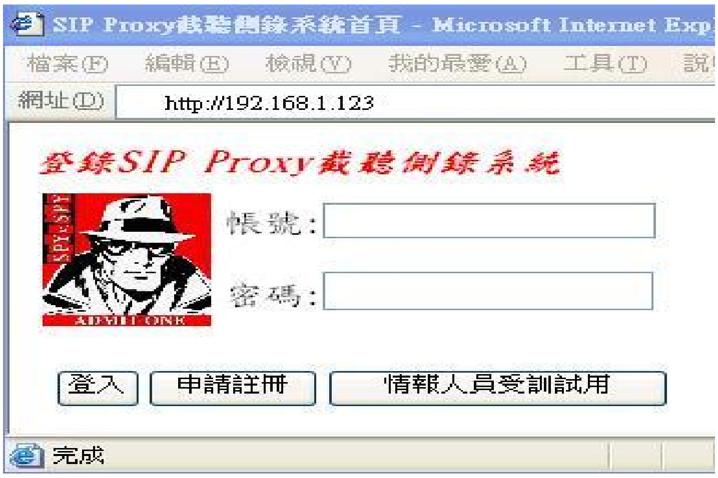


Figure 6. The authorization page before the agent can initiate an interception and wiretapping order.

Path 3: SIP server sends out requested users' information to the web server for displaying to the agent's browser as shown in Figure 7.

Figure 7. The requested user information provided by the SIP server through the web server.

Path 4 and path 5: The agent gives iptable command and the associated parameters to the Linux operating system through the webs server to close the DNAT function such that both users is virtually connected with each other via the Gateway.

Path 6: The agent initiates the voice data acquisition program through the web server, and the voice data will be redirected and send to agent via the web server afterwards.

Path 7: The recorded voice data is stored in the database of the SIP server and the information will be provided to the web server as shown in Figure 8.

Path 8: The received user and voice data information is provided to the agent through the interactive web server as shown in Figure 8, the agent then can decide to perform analysis on the voice data.

Figure 8. The interception and wiretapping data after retrieval from the SIP server and display to the agent through the interactive web server.

In the proposed online interception and wiretapping method, the SIP control signalling header and content for the call setup can be acquired by a handheld device or a notebook computer. The call setup information of the caller and calee is obtained using the interactive web server embedded in the SIP server and display at the agent's browser. The agent can also issue commands through the web server to request the voice data going through the SIP server

CONCLUSION

In this paper, an interception and wiretapping method for SIP-based Internet Telephony is proposed and implemented in a LAN environment. The proposed method can be performed in two types which are offline operation and online real-time operation. The signalling and voice data communication of the caller and callee can be obtained by a law enforcement agent by issuing commands from a web server embedded in an integrated SIP server. With some modification and extension, the proposed method can be applied to the real Internet as well. The proposed method can be utilized to collect, analyse, and preserve the digital evidence from users of the SIP based Internet Telephony, hence the users would be afraid of risking his legal status and would not use the Internet Telephony to commit illegal behaviours, such as black-mail, harassment, commercial spy, under-table transactions, traitor etc….. It can thus prevent the users from abusing the information and communication technology. .

REFERENCES

CALEA (1994). "Communications Assistance for Law Enforcement Act of 1994" (CALEA), Pub. L. No. 103-414, 108 Stat. 4279, Congress of the United State of America.

ITU-T (2000). Packet-based multimedia communication systems." ITU-T, Recommendation H.323 version 4, November 2000.

Rosenberg, J. et al. (2002). "SIP: Session Initiation Protocol," RFC 3261, June 2002.

Schulzrinne, H. and Rosenberg, J. (1998). "The Session Initiation Protocol: Providing Advanced Telephony Services Across the Internet," Bell Labs Tech. J., vol. 3, Oct.-Dec. 1998, pp. 144-160.

Schulzrinne, H. and Rosenberg, J. (1999). "The IETF Internet Telephony Architecture and Protocols," IEEE Network, May/June 1999, pp. 18-23; http://computer.org/internet/telephony/index.htm

Rao, H. C. H., Lin, Y-B, and Cho, S-L. (2000). "iGSM: VoIP Service for Mobile Networks," *Communication Magazine*, vol. 38 No. 4, April 2000, pp. 62-69

Anjum, F. et. al., (1999). "ChaiTime: A System for Rapid Creation of Portable Next-Generation Telephony Services Using Third-Party Software Components," OPENARCH '99, 1999.

Gbaguidi, C. et at., (1999). " An Architecture for the Integration of Internet and Telecommunication Services, " OPENARCH '99, 1999.

Bergmark, D. and Keshav, S. (2000). "Building Blocks for IP Telephony," *IEEE Communications Magazine,* pp. 88-94, vol. 38 No. 4, April 2000

Zeadally, S., and Siddiqui, F. (2004). "Design and Implementation of a SIP-based VoIP Architecture," *18th International Conference on Advanced Information Networking and Applications (AINA2004)*, Vol. 2, 29-31 March 2004 pp.187 – 190

Prasad, R. V., Hurni, R. and. Jamadagni, H. S. (2003). "A Scalable Distributed VoIP Conferencing Using SIP," *Proceedings of Eighth IEEE International Symposium on Computers and Communication (ISCC 2003)*, vol.1, 30 June-3 July 2003, pp. 608 – 613.

Kwak, J. Y., Sul, D. M., Ahn, S. H., Kim, D. H. (2003). "An embedded software architecture for connected multimedia services in ubiquitous network environment," *2003 IEEE Workshop on Software Technologies for Future Embedded Systems*, 15-16 May 2003, pp. 61 – 64.

Cassey, Eoghan (2000). "Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet," Academic Press, London, UK, 2000.

Middleton, Bruce (2001). "Cyber Crime Investigator's Field Guide," Auerbach Publications, Washington D.C. USA, 2001.

Marcella, Albert J., Greenfield, Robert S. (2001). "Cyber Forensics: A field Manual for Collecting, Examining, and Preserving Evidence of Computer Crimes," Auerbach Publications, Washington D.C. USA, 2001.
Vacca, John R. (2002) "Computer Forensics: Computer Crime Scene Investigation," Charles River Media, Inc., Massachusetts, USA, 2002.

The Design of Load-Balancing Computer-Assisted Instruction System with Embedded 3D Virtual Instruments

Dr. Fu-Chien Kao
Da Yeh University
Changhua, Taiwan
Email:fuchien@mail.dyu.edu.tw

Kun-Yi Chiang, Chia-Liang Kuo
Da Yeh University
Changhua, Taiwan
Email:R9306022@mail.dyu.edu.tw

Abstract

To promote the goal of web-based education, this paper proposed a multi-tier Computer-Assisted Instruction (CAI) system with load-balancing function, focusing on the basic electricity course of electronics and electrical engineering. The proposed CAI system's functions contain multimedia teaching, animation simulation, random simulated examination and weakness analyzing, 3D virtual instruments, and on-line maintenance. It aims at guiding students to build up theoretical concepts. The proposed teaching units contain serial-parallel circuit unit, simple circuit unit, general circuit unit, voltage division and current division unit, and network theory unit. To improve the learning interest and efficiency on the practice course, the proposed interactive 3D virtual instruments allow learners to operate the 3D virtual instruments in the front-end computers by embedding technology. In addition, the proposed embedded teaching modules and load-balancing module can obviously reduce and distribute the loads on the middle-tier servers and transfer the loads to the clients' computers when the users connect to the Web Server. It makes Internet learning more effective and efficient.

Keywords:
load-balancing, ActiveX, virtual instruments

INTRODUCTION

There is no doubt that technology has gradually been incorporated into our school systems. Computers are used not only as a means of helping schools analyze data; computers have become a pervasive tool toward optimizing student learning. Personalizing information allows CAI to increase learning interest in the given tasks (Padma,A.G., 1987) and increase the internal logic and organization of the material (Anderson, R.C., 1984). Current research shows the mechanisms by which computer programs facilitate this learning: (1) animating objects on the screen, (2) providing practice activities that incorporate challenges and curiosity, (3) providing a learner with choice over his/her own learning. Providing students with choice over their own learning provides learner-controlled instruction, which contributes to motivation. Increased motivation in turn increases student learning(Traynor, 2003). While the factors contributing to learning using CAI have been well established in the literature, there is a lack of data based research showing the impact of CAI on various types of students. Therefore, the researchers have integrated 3D virtual instruments into the CAI system to improve learning interest and efficiency.

As a result of the rapid growth in Internet web traffic, more and more web sites replicate information across independent or coordinated servers. Among various solutions, the distributed web server system is the most promising one(T.H.Chien and F.C.Kao, 2005). Due to enhanced capability, distributed processing and information replication, the cluster of web servers could significantly increase the processing throughput of the system and reduce the response delay. Two-tier client/server application architecture is implemented when a client talks directly to a server, with no intervening server. It is typically used in small environments of less than 50 users. Its disadvantage is that it provides limited flexibility in moving program functionality from one server to another without manually regenerating procedural code. Three-tier client/server application architecture separates the user interface, the data logic and the database server (data services) onto three separate tiers. Generally this means that there are three computers involved, though it is possible to program the application so that the data logic part runs as an independent process on either the client or server computer. It has been shown to improve performance for groups with a large number of users (in the thousands) and improves flexibility when compared to the two-tier approach.

The proposed load-balancing CAI system with 3D virtual instruments mainly contains the theoretical teaching courses and the practical interactive 3D virtual instruments. The proposed system is integrated by the multi-tier network architecture, the embedded teaching modules, embedded load-balancing module, embedded 3D virtual instruments, and database. By TCP/IP connection with the users' computer (client), Middle-Tier Server, and back-end (SQL Server) database. The proposed 3D virtual instruments are packaged into embedded ActiveX objects, and embedded into learners' computers to reduce the load on the Web Server. The researchers integrate the load-balancing module into the system to distribute the loads on the middle-tier servers, and promote the operating performance of the network.

THE ARCHITECTURE AND FUNCTIONS OF THE CAI SYSTEM

The proposed system functions are described as the following (shown in Figure 1).

(1). Multimedia teaching: There are five teaching units using this function, as shown in Figure 2. All teaching units are made into embedded learning objects and embedded into the learners' computers when learners connect to the Web Server. The theoretical teaching units comprise serial-parallel circuit section, simple circuit section, general circuit section, voltage division, current division section, and network theory section.

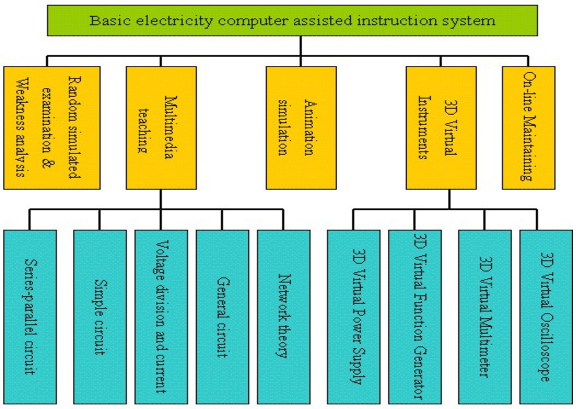


Figure 1 The architecture of CAI system functions

Each section contains several examples. Learners can select these examples to practice. Finally, learners can check whether the calculated results are correct by the function of answer window. After example explanation,

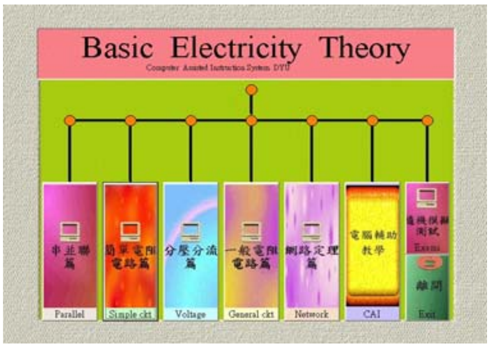


Figure 2 The theoretical teaching units of Basic Electricity

learners will understand how to analyze circuits. The proposed CAI system also allows learners to set element's values by themselves and measure the values of voltage and current by the virtual multimeter. Each section is divided into four units:

a. Example explanation unit: This unit provides various kinds of circuits and their analysis methods, as shown in Figure 3.

b. Circuit measuring unit (virtual multimeter): Learners can set elements' values by themselves and then obtain the values of the element's voltage and current from the function of circuit measuring.

c. Formula teaching unit: The relevant electrical formulas are incorporated in the animated presentations by the software tools of Director, VRT, Photoshop and VB. These tools provide the technologies for designing multimedia content. The proposed animated formulas help learners to understand abstract electrical theories through these animated multimedia presentations.

d. Exercise unit: Each section offers many exercises. Learners can build up correct electrical concepts by doing these exercises.

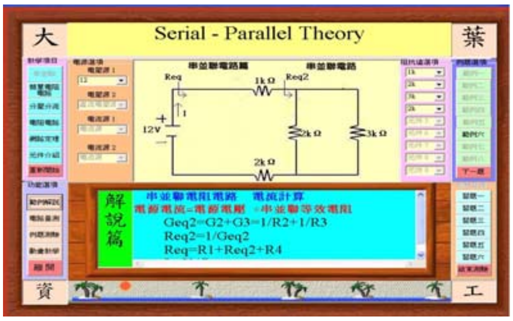


Figure 3 Example explanation function

(2). Random simulated examination and Weakness analysis function: Learners can decide the number of questions, and randomly select the examination questions from database server. After simulated examination, the system automatically lists the test results and the teaching units which need to be improved by learner.

(3). Animation simulation function: To improve learners' interest and efficiency, this system allows learners to learn the relevant electrical theories by the function of animation simulation.

(4). 3D virtual instruments: To reduce training time in the operation of electronic instruments and improve the learning interest of practice courses, this function allows learners to operate 3D virtual instruments with their computers. The proposed interactive 3D virtual instruments include Function Generator, Oscilloscope, Multimeter and Power Supply, as shown in Figure 4 ~ 6. All 3D virtual instruments are packaged into embedded practice modules by ActiveX technology. It is supported by VB Web browsing object and plug-in Viscape to demonstrate the 3D virtual instruments, as shown in figure 7.

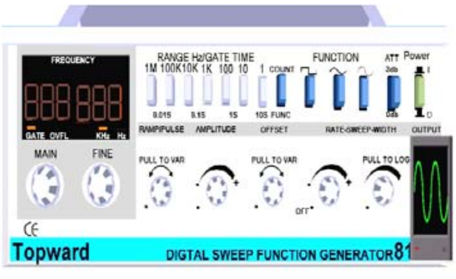


Figure 4 3D virtual Function Generator

Figure 5 3D virtual Power Supply

Figure 6 3D virtual Multimeter

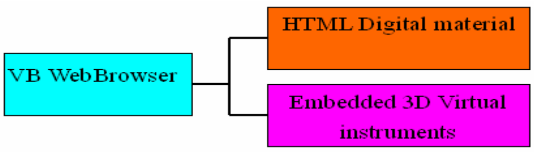


Figure 7 System integrated interface

THE DESIGN OF MULTIMEDIA CONTENT

The proposed teaching materials are designed by multimedia technology. By the application of Cool 3D, Director, and PhotoShop, the teaching materials are therefore produced, as shown in Figure 8.

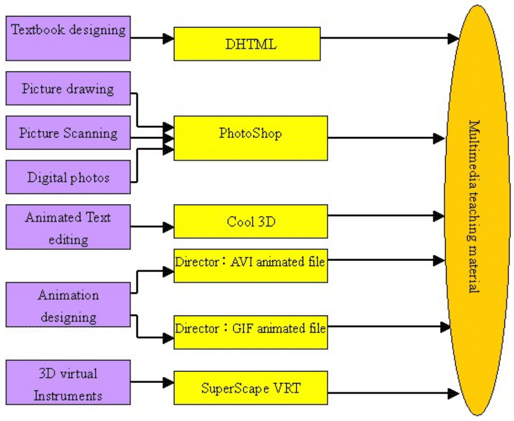


Figure 8 Design of multimedia teaching material

THE PROPOSED LOAD-BALANCING NETWORK ARCHITECTURE

The proposed multi-tier network architecture with load-balancing and embedded function is shown in figure 9. The multi-tier network architecture has three tiers. The client end is the user end, the embedded practice modules are in the middle-tier server, and the back end is the database server. The relevant operating steps of the system are listed below.

(1) Users connect to the Web Server by Web browser.
(2) Check if the client has installed the embedded objects. If the client has not or the version is old, then the embedded objects are automatically downloaded for the client.
(3) The load-balancing module in the front end checks the load on the middle-tier servers by polling method and chooses one to connect to, and then allows learners to proceed with the random simulated examination.
(4) The middle-tier server with the communication function of TCP/IP crates the connection.
(5) The front-end learners are able to access to the back-end SQL server through the middle-tier server with the communication function of TCP/IP and ADO object.

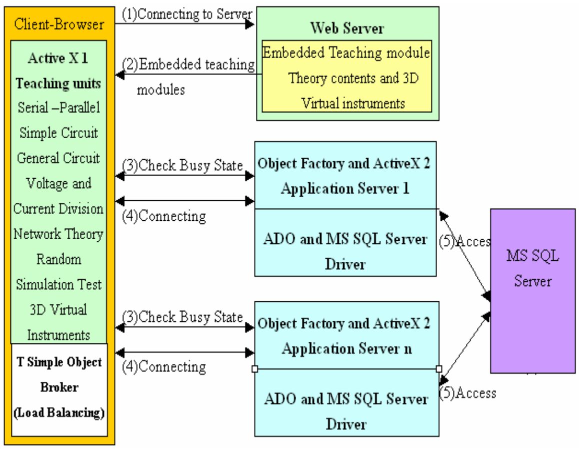


Figure 9 Load-balancing network architecture

THE ADVANTAGES OF THE PROPOSED EMBEDDED MODULES

For reaching the interactive effect between client and server and reducing the load on the middle-tier servers, the practice modules are packaged into embedded ActiveX objects. The proposed embedded ActiveX modules contain load-balancing modules, random examination, score-figuring modules, on-line maintenance modules and 3D virtual instruments. The advantages of the proposed embedded modules are described as below:

(1) Offers front-end users an interactive learning interface and a strong function for dealing with the back-end database.
(2) Allows front-end users to practice with 3D virtual instruments through the performance of front-end computers.
(3) The web speed connection is fast. Clients only need to connect once; the embedded ActiveX modules are automatically embedded into the users' computers.
(4) The embedded load-balancing module is able to check the loads of the middle-tier servers by polling mechanism and distribute the loads on Servers, therefore improving the performance of network operation.

(5) Clients don't need to install any plug-in software to browse and operate the embedded 3D virtual instruments.
(6) By the application of embedded TCP/IP module and the ADO object on middleware server, the system allows learners at the front end to take part in the random simulated exam.
(7) Allows database keeper to maintain and update the database through the Internet.

The proposed embedded load-balancing module can obviously reduce the respond time when the users connect the servers to request the teaching materials, the results as shown from Figure 10 to Figure 13 and the comparison results are shown in Figure 14. The maximum respond time in the proposed system with load-balancing module is 10.46ms(as shown in Figures 12 and 13), but the maximum respond time in the system without load-balancing module is 12.4ms. The proposed embedded load-balancing module therefore can effectively reduce the respond time, and also the curve of respond time almost maintain as a horizontal line. It means that the system with embedded load-balancing module would not change the respond time when the users continuously increase.

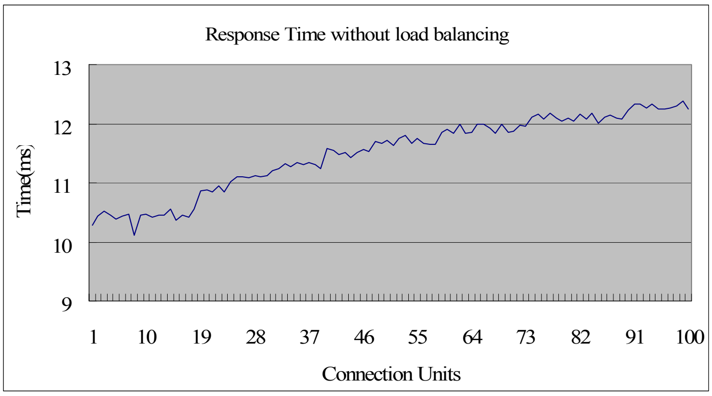


Figure 10 Respond time without Load-balancing function

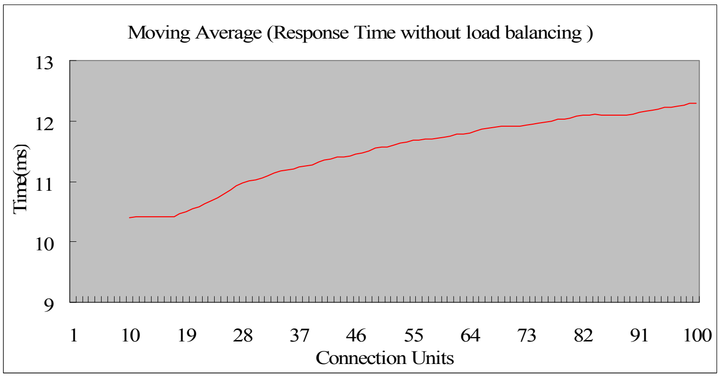


Figure 11 Moving average time without Load-balancing function

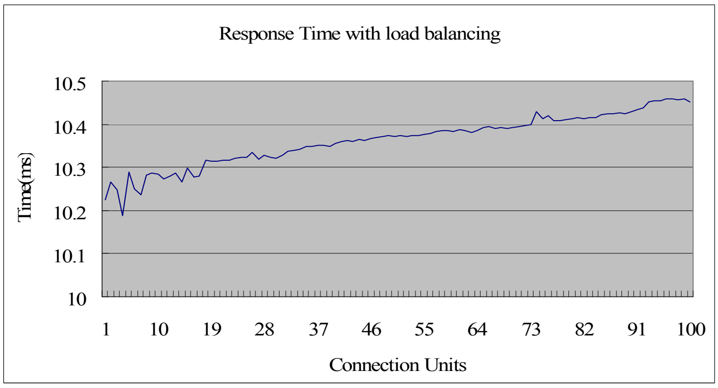


Figure 12 Respond time with Load-balancing function

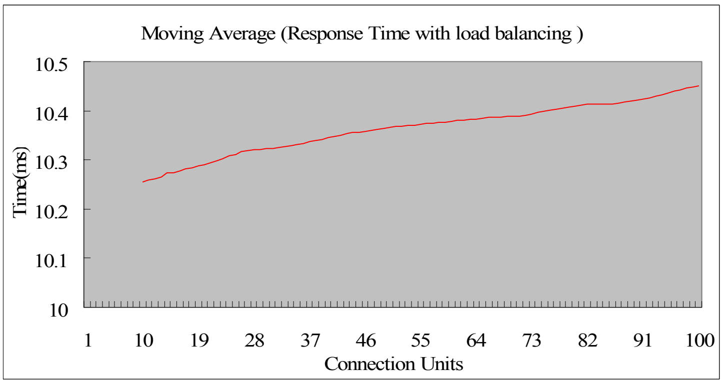


Figure 13 Moving average time with Load-balancing function

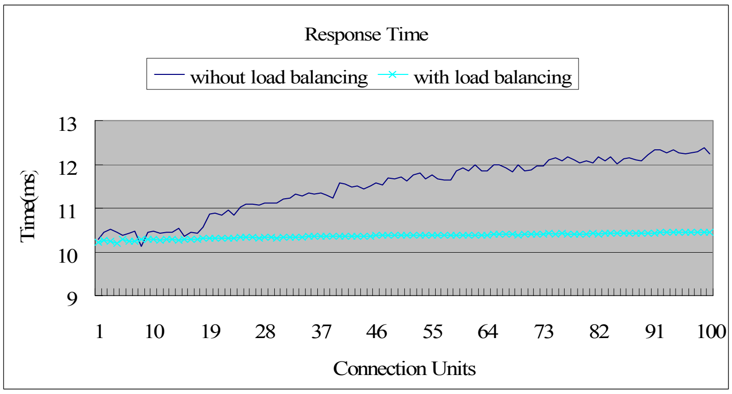


Figure 14 The comparison of respond times

Table 1 Results of examination

Examination of Basic Electricity Theory			
Without CAI System		With CAI System	
Grades	Numbers	Grades	Number
41~50	1	41~50	0
51~60	5	51~60	0
61~70	17	61~70	8
71~80	5	71~80	17
81~100	2	81~100	5
Average score: 66.5		Average score: 77.6	

Table 2 Improvement of learning efficiency and interest

The CAI System of Basic Electricity Theory and Practice			
The Improvement of Connecting State		3D virtual instruments' operation	
Without Load-Balancing module	With Load-Balancing module	Improve Learning Efficiency	Improve Learning Interest
53.33% (16/30)	100.0% (30/30)	100.0% (30/30)	100.0% (30/30)

CONCLUSIONS

This study proposed a load-balancing CAI system to improve the learning environment on the Internet. The multi-tier network architecture, embedded 3D virtual instruments, embedded load-balancing module and database mainly integrate the proposed load-balancing network. The proposed multi-tier network structure provides the users in the front end to download the embedded 3D virtual instruments from the middle-tier servers, and therefore transfer the loads on Servers to the client. The proposed embedded load-balancing module is able to distribute the load on servers and improve the operating performance of networks. By the application of embedded practice modules and embedded load-balancing modules, the operating performance of the system network is obviously improved, and it is also hoped that this improved network environment can promote the learners' interest and efficiency.

The proposed CAI system with load-balancing function improves the operating performance of networks, and improves learners' interest and efficiency. The average exam scores improved from 66.5 to 77.6 after the assistance of CAI system, as shown in Table 1. Among 30 learners, there were only 53.33% (16/30) of learners satisfied with the network connection quality for the system without load-balancing module installed. In comparison, for the proposed CAI system with load-balancing function, all learners (30/30) were satisfied with the network quality, as shown in Table 2. The results prove that the function of load-balancing obviously promotes the efficiency of the network operation. For practical teaching, the training with embedded 3D virtual instruments is really able to improve the learning interest and efficiency, as show in Table 2. This paper proposed a load-balancing CAI system with embedded 3D virtual instruments and hopes it can improve the learning environment on the Internet and promote learning efficiency and interest.

ACKNOWLEDGES

This paper has been supported in part by the grant NSC93-2520-S212-002 from the National Science Council of Taiwan.

REFERENCES

[1] Padma, A. G., Ross, S. M., "Using computer-assisted-instruction to personalize arithmetic materials for elementary school children", Journal of Educational Psychology, 79, 1987, pp. 72- 78.

[2] Anderson, R. C., "Some reflections on the acquisition of knowledge", Educational Researcher, 13, 1984, pp. 5- 10.

[3] Ausubel, D. P, "Educational psychology: A cognitive approach". New York: Holt Rinehart and Winston, 1968.

[4] Mayer, R. E., "Information processing variables in learning to solve problems", Review of Educational Research,1975.

[5] Rumelhart, D. E., Ortony, A. "The representation of knowledge in memory". In R.C. Anderson, S. J. Sprio, W.E. Montague (Eds.), Schooling and the acquisition of knowledge, Hillsdale, NJ: Earlbaum., 1977, pp. 99-135.

[6] Kinzie, M. B., Sullivan, H. J., Berdel, R. L., "Learner control and achievement in science computer-assisted-instruction", Journal of Educational Psychology*, 80, 1988, pp. 299- 303.*

[7] Dr. Patrick L. Traynor, "Effects of Computer-Assisted-Instruction On Different Learners", Journal of Instructional Psychology*, 825 - 30 (2), 2003, pp. 137 – 143.*

[8] Tzu-Hua Chien, Fu-Chien Kao, "The Design of Load-Balancing LMS Based on Decomposition Structure", 2005, 5th IEEE International Conference on Advanced Learning Technologies, pp. 783 - 787.

[9] Tao Cai, Shi-Guang Ju, Zhen-Jin Qian, Dian-Chuen Guo, Si-Chuen Cai, Hai-Hua Qian, "Implement of a Web-based and serial port agent remote control demonstrating systems [Implement read Implementation]", Computer Supported Cooperative Work in Design, 2004. Proceedings. *The 8th International Conference on, Volume: 2, 26-28 May 2004, Pages: 567 - 572 Vol.2.*

[10] Frank, M., "Characteristics of engineering systems thinking - a 3D approach for curriculum content", Systems, Man and Cybernetics, Part C, IEEE Transactions on*, Volume: 32, Issue: 3 , Aug. 2002 , pp. 203 – 214.*

[11] M. Colajanni, P. S. Yu, and D. M. Dias. Analysis of task assignment policies in scalable distributed web-server systems. IEEE Trans. on Parallel and Distributed Systems*, 9(6), June 1998.*

[12]Bilski, P., Winiecki, W., "Time optimisation of soft real-time virtual instrument design", Instrumentation and Measurement Technology Conference, 2004. IMTC 04.Proceedings of the 21st IEEE*, Volume: 3, 18-20 May 2004, pp. 2223 - 2228 Vol.3.*

[13]Morales, C.R., Meador, S.W., "Trial results from a custom distance learning system based on hybridization of CD-ROM, Web, and Active-X component technology", Information Visualization, 2000. Proceedings. IEEE International Conference on*, 19-21 July 2000, pp. 39 – 43.*

[14] Garcia, R.C., Heck, B.S., "Enhancing engineering education on the Web: the use of ActiveX controls and automation server technology", Southeastcon 2000. Proceedings of the IEEE, 7-9 April 2000, pp. 515 – 518.

[15] Zhong Xu, Rong Huang, Bhuyan, L.N., "Load balancing of DNS-based distributed Web server systems with page caching", Parallel and Distributed Systems, 2004. ICPADS 2004. Proceedings. Tenth International Conference on*, 7-9 July 2004, pp. 587 – 594.*

[16] Grosu, D., Chronopoulos, A.T., "A truthful mechanism for fair load balancing in distributed systems", Network Computing and Applications, 2003. NCA 2003. Second IEEE International Symposium on, *16-18 April 2003, pp. 289 – 296.*

A Conceptual Framework and Propositions for the Acceptance of Mobile Services

Dr Indrit Troshani, Dr Sally Rao
School of Commerce
University of Adelaide
Adelaide, SA 5005, Australia
indrit.troshani@adelaide.edu.au
sally.rao@adelaide.edu.au

Abstract

Mobile services are heralded to create a tremendous spectrum of business opportunities. User acceptance of these services is of paramount importance. Consequently, a deeper insight into theory-based research is required to better understand the underlying motivations that lead users to adopting mobile services. Most of the theoretical acceptance models available in the literature originate from organisational contexts. As mobile services bring additional functional dimensions, including hedonic and experiential aspects, using extant models for predicting mobile services acceptance by individuals may be inadequate. The aim of this paper is to explore, analyse and critically assess the use of existing acceptance theories in the light of evolving mobile services and their underlying technologies. Constructs affecting adoption behaviour are discussed and relevant propositions are made. Managerial implications are explored and future research directions are also identified.

Keywords:
Mobile services, technology adoption, innovation adoption, mobile services adoption, mCommerce

INTRODUCTION

Mobile technologies and services are heralded to create a tremendous spectrum of business opportunities. According to the International Telecommunications Union (ITU) mobile phone penetration rates have increased significantly in many countries in Northern Europe (e.g. Sweden 98.05%, Denmark 88.72%, Norway 90.89%) (Knutsen et al., 2005). Similarly, Japan and Korea have consistently experienced very high diffusion rates of mobile devices and services (Carlsson et al., 2005; Funk, 2005). Nevertheless, the current penetration rate in many countries in the Western hemisphere and Asia-Pacific, including the US and Australia lags behind the forerunners (Funk, 2005; Ishii, 2004; Massey et al., 2005). Given the difference between rapid growth rates in the adoption of mobile technologies and associated services in some countries and the relatively slow growth rates in others (Bina & Giaglis, 2005; Knutsen et al., 2005) it is important to identify the factors and predictors of further adoption.

Mobile technology is enabled by the collective use of various communication infrastructure technologies and portable battery-powered devices. Mobile devices are powered by applications which deliver various services while enhancing flexibility, mobility, and efficiency for users within business and life domains. Despite the availability of technologically advanced mobile devices there is evidence that advanced mobile services which run on these have not been widely adopted (Carlsson et al., 2005; Khalifa & Cheng, 2002). The recognition of this fact is important for the mobile telecommunications industry because mobile services associated with technologically advanced devices constitute a massive source of potential revenue growth (Alahuhta et al., 2005; Massey et al., 2005).

The adoption of advanced mobile technologies and services requires special consideration. Most of the current acceptance models are based on research conducted in organisational contexts (Carlsson et al. 2005). There has been only limited research from the consumers' perspective in the area of adoption of mobile technologies and services (Lee et al., 2003). Mobile services, such as Short Message Service (SMS), Multimedia Messaging Service (MMS), email, map and location services, allow for single wireless devices, such as mobile phones, to be used seamlessly and pervasively across traditionally distinct spheres of life, such as work, home, or leisure, and with various levels of time commitment (e.g. busyness) and self-ascribed roles (e.g. professional and on duty or private and off duty) (Dholakia & Dholakia, 2004). In addition, as mobile technologies and services add other functional dimensions, including hedonic and/or experiential aspects (Kleijen et al., 2004; Mathwick et al., 2001), applying extant theories outright to determine the acceptance and adoption by individual users may be inadequate (Knutsen et al., 2005).

Moreover, more research is called for in the adoption of mobile services because of the levels of complexity and diversity that may be encountered during their adoption. A number of factors contribute to this. First, there is a strong relationship between the mobile devices and their users because the former always carries the identity of the latter (Chae & Kim, 2003). As a result, spatial positioning and identification of users is easier in the mobile context than in the traditional innovation adoption (Figge, 2004). Second, most mobile devices have limited available resources including memory, processing power, and user interface, which have the potential to offset ubiquity benefits (Chae & Kim, 2003; Figge, 2004). Third, the lifecycle of mobile technologies is usually short, which increases adoption risks because new technologies become rapidly obsolete, and may, therefore, need to be replaced by newer ones. During this process, a certain amount of recurring learning might be required before adopters can be confident and satisfied in using the mobile devices and services (Saaksjarvi, 2003).

This paper focuses on mobile services. Examples of mobile services include mobile e-mail, commercial SMS and MMS services, downloads to portable devices, access to news through a mobile phone, mobile ticket reservations, mobile stock trading, as well other customised services which may be made available by mobile phone operators (Bina & Giaglis, 2005). The aim of this paper is to extend the existing models and to propose an integrated conceptual and parsimonious framework which attempts to explain adoption behaviour of users of mobile services. To accomplish this, we firstly start with a critical assessment of existing acceptance models. Next, acceptance constructs and their relevance to mobile services is discussed. Theoretical propositions are derived from this discussion to facilitate future research. The constructs are then integrated into a new adoption framework suitable for mobile services. Finally, the implications of this model and future research directions are also discussed.

A CRITICAL REVIEW OF THEORETICAL MODELS OF TECHNOLOGY ACCEPTANCE

A review of technology acceptance literature revealed many competing theoretical models, each with different focus and tested in different contexts. A significant amount of research effort has been put into building theories to examine how and why individuals adopt new information technologies and predict their level of adoption and acceptance. While one stream of research focuses on individual acceptance of technology (Compeau & Higgins, 1995; Davis et al., 1989a), other streams have focused on implementation success at the organizational level (Leonard-Barton & Deschamps, 1988).

Many of the previously empirically researched models have been drawn from social psychology, for example, Theory of Reasoned Action (TRA), Motivational Model, Theory of Planned Behaviour (TPB) and sociology, for example, Social Cognitive Theory (SCT) and Innovation Diffusion Theory (IDT). Others specifically apply to technology adoption, for example, Technology Acceptance Model (TAM). While each of these models made unique contributions to the literature on technology acceptance and adoption, most of these theoretical models theorise behaviour intention and/or usage as the key dependent variable in explaining acceptance of information technology because behavioural intentions are motivational factors that capture how hard people are willing to try to perform a behaviour (Ajzen, 1991). For example, TPB suggests that behavioural intention is the most influential predictor of behaviour; after all, a person does what s/he intends to do. In a meta-analysis of 87 studies, an average correlation of 0.53 was reported between intentions and behaviour (Sheppard et al., 1988). As mobile services and underlying technologies are emerging information technologies, it is appropriate to consider this as the point of departure and use it to form the basis of a theoretical framework in mobile services and technology acceptance and adoption. The models that have been most frequently quoted in the technology acceptance and adoption literature are discussed next.

Theory of reasoned action (*TRA*). Theory of Reasoned Action models are considered to be the most systematic and extensively applied approaches to attitude and behaviour research. According to TRA, the proximal determinant of a behaviour is a behavioural intention, which, in turn, is determined by attitude. These models propose that an individual's actual behaviour is determined by the person's intention to perform the behaviour, and this intention is influenced jointly by the individual's attitude and subjective norm. Attitude is defined as "a learned predisposition to respond in a consistently favourable or unfavourable manner with respect to a given object" (Fishbein & Ajzen, 1975, p. 6). A person's attitude towards a behaviour is largely determined by salient beliefs about the consequences of that behaviour and the evaluation of the desirability of the consequences (Fishbein & Ajzen, 1975). Subjective norm is defined as "the person's perception that most people who are important to him think he should or should not perform the behaviour in question (Dillon & Morris, 1996). In brief, TRA asserts that attitude and subjective norm and their relative weights directly influence behavioural intention.

Theory of planned behaviour (TPB) and Decomposed theory of planned behaviour. TPB, which generalizes TRA by adding a third construct – perceived behavioural control (Ajzen, 1991) – has been one of the most influential theories in explaining and predicting behaviour, and it has been shown to predict a wide range of behaviours (Sheppard et al., 1988). TPB asserts that the actual behaviour is determined directly both by behavioural intention and perceived behavioural control. Behavioural intention is formed by one's attitude, subjective norm and perceived behavioural control (Ajzen, 1991). Further, a decomposed TPB includes constructs such as relative advantage, compatibility, influence of significant others, and risk from the innovation diffusion literature, and decomposing the three perceptions in TPB into a variety of specific belief dimensions. This model offers several advantages over TPB and is considered more complete and management-relevant by focusing on specific factors that may influence adoption and usage (Teo & Pok, 2003).

Technology acceptance model (TAM). TAM can be seen as an adaptation of the theory of reasoned action (TRA) and was developed to predict and explain individual system use in the workplace (Davis, 1989). This model further suggests that two beliefs - perceived usefulness and perceived ease of use - are instrumental in explaining the user's intentions of using a system. Perceived usefulness refers to the degree to which "a person believes that use of the system will enhance his or her performance" whereas perceived ease of use is the degree to which "a person believes that using the system will be free of effort". Simply put, a technology that is easy to use and is useful will lead to a positive attitude and intention towards using the technology.

The main advantage of this model over others is that the two related beliefs can generalize across different settings. Thus, some argue that it is the most robust, parsimonious and influential model in explaining information technology adoption behaviour (S. Elliot & Loebbecke, 2000; Teo & Pok, 2003; Venkatesh et al., 2003). Indeed, since its development, it has received extensive empirical support through validations, applications and replications for its prediction power (Taylor & Todd, 1995, 1995a; Venkatesh & Morris, 2000a). A number of modified TAM models were proposed to suit new technologies including Internet and intranet (Agarwal & Prasad, 1998; Chau, 1996; Chau & Hu, 2001; Horton et al., 2001).

A major theoretical limitation of TAM is the "exclusion of the possibility of influence from institutional, social, and personal control factors" (Elliot & Loebbecke, 2000, p. 49). Thus the suitability of the model for predicting general individual acceptance needs to be re-assessed as the main TAM constructs do not fully reflect the specific influences of technological and usage-context factors that may alter user acceptance (King et al., 1994; Taylor & Todd, 1995). In response to this, a number modifications and changes to the original TAM models have been made. The most prominent of these is the Unified Theory of Acceptance and Use of Technology (UTAUT), a unified model that integrates constructs across eight models (Venkatesh et al. 2003). UTAUT provides a refined view of how the determinants of intention and behaviour evolve over time and assumes that there are three direct determinants of intention to use (performance expectancy, effort expectancy and social influence) and two direct determinants of usage behaviour (intention and facilitating conditions). However, both TAM and UTAUT have received criticisms with the fundamental one being about the problems in applying these beyond the workplace and/or organisation for which originally created (Carlsson et al., 2005).

Motivational theories. Motivation theories are rooted in psychological research to understand individuals' acceptance of information technology (Davis et al., 1992; Igbaria et al., 1996). These theories often distinguished extrinsic and intrinsic motivation. While extrinsic motivation refers to the performance of an activity in helping achieve valued outcomes, intrinsic motivation puts emphasis on the process of performing an activity (Calder & Staw, 1975; Deci & Ryan, 1985). For example, perceived usefulness is an extrinsic source of motivation (Davis et al., 1992) while perceived enjoyment (Davis et al., 1992), perceived fun (Igbaria et al., 1996) and perceived playfulness (Moon & Kim, 2001) can be considered intrinsic sources of motivation. Both sources of motivation affect usage intention and actual usage. Therefore, in addition to ease of use and usefulness, intrinsic motivators, such as playfulness, will also play an important role in increasing usability in a usage environment in which information technology applications are both used for work and play (Moon & Kim, 2001).

Innovation diffusion theory. The innovation diffusion theory is concerned with how innovations spread and consists of two closely related processes: the diffusion process and adoption process (Rogers, 1995). Diffusion is a macro process concerned with the spread of an innovation from its source to the public whereas the adoption process is a micro process that is focused on the stages individuals go through when deciding to accept or reject an innovation. Key elements in the entire process are the innovation's perceived characteristics, the individual's attitude and beliefs, and the communication received by individuals from their social environment. In relation to the factors pertaining the innovation, factors such as, relative advantage, complexity, trialability, observability and compatibility were considered important in influencing individual's acceptance of the innovation (Rogers, 1995).

A PROPOSED ACCEPTANCE MODEL FOR MOBILE SERVICES

This section develops an acceptance model for mobile technology and services that may be empirically tested. This development begins with identifying the latent constructs in extant technology adoption literature. However, mobile services differ from traditional systems in that mobile services are ubiquitous, portable and can be used to receive and disseminate personalised and localised information (Siau et al., 2001; Teo & Pok, 2003). It follows that the models examined in the previous section and the underlying constructs are not readily applicable to mobile services adoption. In particular, we discuss the various antecedents of attitude towards mobile services and develop a new framework based on the widely used TAM model to predict adoption of new mobile services.

User predisposition

User predisposition refers to the internal factors of an individual user of mobile services. Personal differences strongly influence adoption. There is evidence that successful acceptance of innovations depends as much on individual adopter differences as on the innovation itself. Recognising individual differences that impact technology adoption is important because it helps identify segments of adopters who are more likely to adopt technology innovations than others, which in turn, helps providers address adopter needs more closely (Massey et al., 2005). Further, these individual adopters can then act as opinion leaders or change agents to facilitate the diffusion of the technology further (Rogers, 1995). Diffusion resources can also be used more effectively and efficiently (Agarwal & Prasad, 1998). In this paper, we define user predisposition as the collection of a number of factors including the individual's prior knowledge and experience of existing mobile services, compatibility, behavioural control, personal innovativeness, perceived enjoyment, and price sensitivity which are examined below.

First, *prior knowledge* is essential for the comprehension of the technology and related services. According to Rogers (1995) knowledge occurs when a potential adopter learns about the existence of an innovation and gains some understanding concerning its functionality. Knowledge consists of two components, namely, familiarity and expertise. For instance, the former constitutes the number of mobile services-related experiences accumulated by consumers over time, which includes exposure to advertising, information search, interaction with salespersons, etc. The latter represents the ability to use the mobile services, and it includes beliefs about service attributes (i.e. cognitive structures) as well as decision rules for acting on those beliefs (i.e. cognitive processes) (Alba & Hutchinson, 1987). However, while knowledge and its components is is important, by itself, it has limited usefulness. As a matter of fact, "knowledge alone cannot determine the basis for adoption" (Rogers, 1995, p. 167) of a technology or service. Adopters' previous positive or negative *experiences* with a technology or service can have a significant impact on their perceptions and attitudes towards that technology or service (Lee et al., 2003; Taylor & Todd, 1995a). Because of their greater clarity and certainty, direct prior experiences are likely to have a stronger impact on perceptions and attitudes towards usage than indirect or incomplete evidence (i.e. pre-trial) (Knutsen et al., 2005; Lee et al., 2003).

The second variable within the user predisposition construct is *compatibility*. Rogers (1995) defines compatibility as the degree to which an innovation is perceived to be consistent with existing values of potential adopters. In general, high incompatibility will adversely affect potential adopters of an innovation, which decreases the likelihood of adoption (Saaksjarvi, 2003). In the context of wireless devices, lifestyle compatibility is the extent to which adopters believe mobile devices and services can be integrated into their daily lives. For example, adopters' lifestyle in terms of degree of mobility is likely to have a strong impact on their decision to adopt the technology (Pagani, 2004; Teo & Pok, 2003). For example, a person who leads a busy lifestyle, and is employed in an information-intensive job, and is always on the move is more likely to adopt a wireless device and its associated services compared to a person who leads a sedentary lifestyle.

Third, perceived *behavioural control*, a dynamic and socio-cognitive concept, has attracted a lot attention in adoption literature. More recent empirical findings suggest that perceived behavioural control has two distinct components: self-efficacy which is an individual's judgement of their capability to perform a behaviour and controllability which constitutes an individual's beliefs if they have the necessary resources and opportunities to adopt the innovation. It denotes a subjective judgment of the degree of control over the performance of a behaviour, not the perceived likelihood that performing the behaviour will produce a given outcome (Ajzen, 1991). In the context of mobile service adoption, perceived behavioural control refers to the individual perception of how easy or difficult it is to getting mobile services.

Fourth, *personal innovativeness* is the willingness of an individual to try out and embrace new technologies and their related services for accomplishing specific goals. Also known as technology readiness, personal innovativeness embodies the risk-taking propensity which exists in certain individuals and not in others (Agarwal & Prasad, 1998; Massey et al., 2005; Parasuraman, 2000). This definition helps segment potential adopters into what Rogers (1995) characterises as innovators, early adopters, early and late majority adopters and laggards. Personal innovativeness represents a confluence of technology-related beliefs which jointly contribute to determining an individual's predisposition to adopt mobile devices and related services. Therefore, given the same level of beliefs and perceptions about an innovation, individuals with higher personal innovativeness are more likely to develop positive attitudes towards adopting it than less innovative individuals (Agarwal & Prasad, 1998).

Fifth, *perceived enjoyment* refers to the degree to which using an innovation is perceived to be enjoyable in its own right and is considered to be an intrinsic source of motivation (Al-Gahtani & King, 1999). Because the market for mobile innovations and services is comprised of both corporate users and consumers, factors focusing on perceived enjoyment constitute an important consideration (Carlsson et al., 2005; Pagani, 2004). That is, adopters use an innovation for the pleasure or enjoyment its adoption might bring and, therefore, serve as an end unto itself. Further, intrinsic enjoyment operates outside valued outcomes or immediate material needs (i.e. extrinsic motivations), such as enhanced job performance, increased pay etc. (Mathwick et al., 2001; Moon & Kim, 2001). Upon adoption, individuals are more likely to use the mobile services that offer enjoyment more extensively than those which do not (Kaufaris, 2002; Novak et al., 2000).

The final variable is *price sensitivity*. In the original technology acceptance models, from an individual point of view, the costs of adopting an innovation was not considered to be a relevant construct because the actual users did not have to pay for the technology. In an organisational setting, the cost would be incurred by the organisation. However, in the context of individual adoption, cost becomes a relevant factor. There is evidence showing that perceived financial resources required to adopt mobile technologies and services affect behavioural intention (Kleijen et al., 2004; Lin & Wang, 2005). This is particularly relevant in the adoption of mobile services, which might be 'nice-to-have', but not an absolute necessity. Further, evidence also shows that adopters of mobile devices and services also attempt to assess the value of adoption by comparing perceived costs against the benefits (Crawford, 2002; Lu et al., 2003; Pagani, 2004).

In the context of the above discussion concerning user predisposition, we have the following propositions:

P1a: Prior knowledge and experience will have a positive influence on the attitude towards adopting mobile services.
P1b: Life-style compatibility will have a positive influence on the attitude towards adopting mobile services.
P1c: Perceived behavioural control will have a positive influence on the attitude towards adopting mobile services.
P1d: Personal innovativeness will have a positive influence on the attitude towards adopting mobile services.
P1e: Perceived enjoyment will have a positive influence on the attitude towards adopting mobile services.
P1f: Price sensitivity will have a positive influence on the attitude towards adopting mobile services.

Perceived Usefulness

Perceived usefulness is "the degree to which a person believes that using a particular system would enhance his or her job performance." (Davis, 1989, p. 320). Perceived usefulness is also known as performance expectancy (Venkatesh et al., 2003). An innovation is believed to be of high usefulness when a potential adopter believes that there is a direct relationship between use on the one hand and productivity, performance, effectiveness, or satisfaction on the other (Lu et al., 2003).

Usefulness recognition is important because it has been found to have a strong direct effect on the intention of adopters to use the innovation (Adams et al., 1992; Davis, 1989). In addition, potential adopters assess the consequences of their adoption behaviour and innovation usage in terms of the ongoing desirability of usefulness (Chau, 1996; Venkatesh & Davis, 2000). Perceived usefulness can be split into two parts. Near-term usefulness is perceived to have an impact on the near-term job fit, such as job performance or satisfaction (Thompson et al., 1994). Long-term usefulness is perceived to enhance the future consequences of adoption including career prospects, opportunity for preferred job assignments or social status of adopters (Chau, 1996; Thompson et al., 1994). Although perceived near-term usefulness has the most significant impact on the behavioural intention to adopt an innovation, perceived long terms usefulness also exerts a positive, yet lesser impact (Chau, 1996; Jiang et al., 2000). In the case of mobile technology and services, perceived usefulness is defined as the degree to which the mobile technology and services provide benefits to individuals in every day situations (Knutsen et al., 2005).

Considering the arguments above we define the following propositions:

P2a: Perceived near-term usefulness will have a positive influence on the attitude towards adopting mobile services.

P2b: Perceived long-term usefulness will have a positive influence on the attitude towards adopting mobile services.

Perceived Ease of Use

Perceived ease of use is the "degree to which a person believes that using a particular system would be free of effort." (Davis, 1989, p. 320). Other constructs that capture the notion of perceived ease of use, are complexity and effort expectancy (Rogers, 1995; Venkatesh et al., 2003). Perceived ease of use may contribute towards performance, and therefore, near-term perceived usefulness. In addition, lack of it can cause frustration, and therefore, impair adoption of innovations.

In the mobile setting, perceived ease of use represents the degree to which individuals associate freedom of difficulty with the use of mobile technology and services in everyday usage (Knutsen et al., 2005). For example, there is evidence in the media that using certain services on a mobile device can be quite tedious, especially when browsing Internet-like interfaces on mobile devices is required (Teo & Pok, 2003). Together with relatively small screen sizes and associated miniaturized keypads, the overall usage experience may be adversely affected. In addition, user-friendly and usable intuitive man-machine interfaces, including clear and visible steps, suitable content and graphical layouts, help functions, clear commands, symbols and meaningful error messages are likely to influence adoption as well (Condos et al., 2002). As a consequence, we propose the following propositions:

P3a: Perceived ease of use and usability will have a positive influence on the attitude towards adopting mobile services.
P3b: Perceived ease of use will have a positive influence on perceived usefulness.

Social Influences

Social influence constitutes the degree to which individuals perceive that important or significant others believe they should use an innovation (Venkatesh et al., 2003). Social influence seems to be significant in the early phases of adoption and its effect decreases with sustained usage (Thompson et al., 1994; Venkatesh & Davis, 2000). Also, in voluntary settings, social influence appears to have an impact on perceptions about the innovation (Venkatesh et al., 2003). In Taylor & Todd's (1995) study, social influences include the influence of other people's opinions otherwise known as referent groups. These include peers, friends, superiors, computer and technology experts. Research shows that pressure from referent groups to adopt an innovation is effective because it contributes to reducing perceived risk associated with adoption (Ishii, 2004; Lu et al., 2003; Teo & Pok, 2003). As a result we make the following proposition:

P4: Perceived social influences will have a positive influence on the attitudes towards adopting mobile services.

Facilitating Conditions

Facilitating conditions refer to external controls and catalysts in the adoption environment which aim at facilitating adoption and diffusion of new technologies (Terry, 1993). Facilitating conditions are important because they are likely to make adoption behaviour less difficult by removing any obstacles to adoption and sustained usage (Thompson et al., 1994; Venkatesh et al., 2003). These conditions can be provided by both governments and mobile operators. For example, governments or the representative agencies can act as facilitators by bringing together the telecommunication industry, academia and research community. Government agencies can also set up protocol standardization policies and regulations favouring the future growth of mobile communication systems (Lu et al., 2003). Likewise, mobile operators can encourage adoption by mass advertising campaigns and active promotion aimed at increasing awareness about mobile services (Teo & Pok, 2003).

Facilitating conditions also capture the existence of a trusting environment that is external to the mobile operator's control and determine the user's expectations from the relationship with their service providers, and increase their perceived certainty concerning the provider's expected behaviour. Generally, trust is essential in all economic activities where undesirable opportunistic behaviour is likely to occur (Gefen et al., 2003). However, trust becomes vital in a mobile environment, where situational factors such as uncertainty or risk and information asymmetry are present (Ba & Pavlou, 2002). On the one hand, adopters of mobile technology are unable to judge the trustworthiness of service providers, and on the other, the latter can also easily take advantage of the former by engaging in harmful opportunistic behaviours. For example, service providers can sell or share the transactional information of its users or their personal information. In mobile adoption research the trust environment has been encapsulated in a construct called perceived credibility (Lin & Wang, 2005; Wang et al., 2003). As a result, we propose the propositions that follow:

P5a: Facilitating conditions will have a positive influence on the attitude towards the adoption of mobile services.

P5b: Perceived trust environment will have a positive influence on the attitude towards the adoption of mobile services.

Moderating variables

Evidence shows that gender and age might influence the adoption of technology and related services due to their moderating effects on other constructs (Venkatesh et al., 2003). In general, men tend to exhibit task-oriented attitudes suggesting that usefulness expectations might be more accentuated in men than women (Minton & Scheneider, 1980). This is particularly the case for younger men (Venkatesh & Morris, 2000a). On the other hand, ease of use expectations are more salient for women and older adopters (Bozionelos, 1996). Further, women are predisposed to be more sensitive to the opinions of members of their social structure. As a result women are more likely to be affected by social influence factors when deciding to adopt new mobile technologies and services (Venkatesh & Morris, 2000a). Similarly, because affiliation needs increase with age (Rhodes, 1983), older adopters are more likely to be affected by social influence.

In this context we define the following proposition:

P6: The collective effect of moderators, particularly gender and age, will have an influence on the attitude towards adopting mobile services.

To summarise the constructs discussed in this section, Figure 2 portrays a proposed model of acceptance of mobile services. The managerial implications that the constructs of this model have are discussed next.

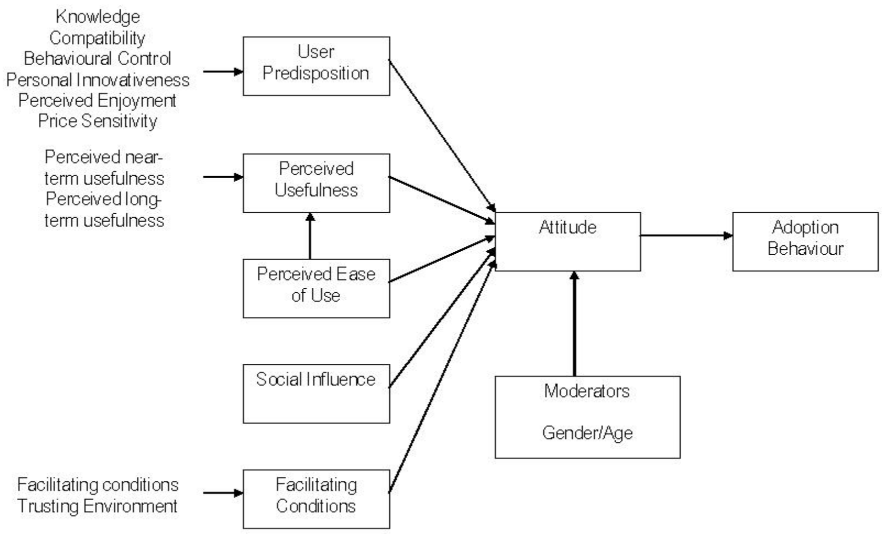


Figure 1. Proposed model of acceptance of mobile services

MANAGERIAL IMPLICATIONS AND CONCLUSIONS

Mobile technologies and the associated services integrate both the business and social domains of the user's life (G. Elliot & Phillips, 2004; Knutsen et al., 2005). This creates the opportunity for developing accurate adopter profiles both in their work- and leisure-related domains. While such information can help address the needs of adopters better, it can also be misused by businesses for unethical direct business-to-consumer marketing (Casal et al., 2004), raising privacy concerns, overcontrol and overwork of individual adopters (Yen & Chou, 2000). As a result, existing privacy protection, and industrial relations policies and regulations about employees and consumers should reflect these new conditions.

Designing content suitable for mobile phones constitutes an important issue that affect the adoption and diffusion of mobile technology and associated services. This has implications for content providers, developers, policymakers and academics. Content providers must design content "for value-contexts specific for mobile use which provide users freedom from complicated configuration procedures, and ubiquitously serve and support current day-to-day individual social practices" (Knutsen, 2005a, p. 7). Further, the content design of mobile applications should encompass both intrinsic and extrinsic motivation dimensions (Moon & Kim, 2001). Developers of mobile applications need to recognise that mobile applications are quite different from PC applications (Funk, 2005). Further, because 'made-for-the-medium' content type and design may be required (Massey et al., 2005), the available technologies which determine screen size, display quality and processing speeds should be taken into consideration as well (Funk, 2005). The combined effect of these factors on navigation patterns, adopters' cognitive overload and subjective perceptions about the usability and ease of use of mobile applications can have a critical impact on uptake (Chae & Kim, 2004).

Segmentation of mobile service adopters must not only be based on adopter type (e.g. pioneers, early adopters, majority adopters, and laggards) but also on individual differences. The basis of segmentation should constitute the foundation in developing marketing strategies. According to Rogers (1995) there are different categories of adopters, such as innovators, early and early majority and late majority adopters, and laggards. Marketing mobile applications for adopters in one category is likely to frustrate adopters in the other. Therefore, developers and marketers should be prudent in recognising that the confluence of various individual characteristics with varying levels of prior experience, perceptions and learning predispositions are all likely to influence adoption and retention patterns (Card et al., 1983; Hung et al., 2003; Massey et al., 2005).

User acceptance of mobile technology and related services is of paramount importance. Consequently, a deeper insight into theory-based research is required to better understand the underlying motivations and barriers that will lead users to or inhibit them from adopting these technologies and services. In this paper we have explored and critically reviewed existing technology acceptance theories. Relevant constructs of extant models were discussed in the light of evolving mobile technologies and services and then incorporated into a synthesised acceptance model of mobile services. The proposed model attempts to view acceptance of mobile services beyond traditional organisational borders. The proposed model which can be tested empirically provides the foundation to guide further validation and future research in the area of mobile services adoption.

REFERENCES

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: a replication. *MIS Quarterly, 16*(2), 227-247.

Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research, 9*(2), 204-215.

Ajzen, I. (1991). The theory of planned behavior. *Organisational Behavior and Human Decision Process, 52*(2), 179-211.

Alahuhta, P., Ahola, J., & Hakala, H. (2005). *Mobilising business applications: a survey about the opportunities and challenges of mobile business applications and services in Finland* (Technology Review No. 167/2005). Helsinki: Tekes.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research, 13*(3), 411-454.

Al-Gahtani, S. S., & King, M. (1999). Attitudes, satisfaction and usage: factors contributing to each in the acceptance of information technology. *Behaviour & Information Technology, 18*(4), 277-297.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. *MIS Quarterly, 26*(3), 243-268.

Bina, M., & Giaglis, G. M. (2005). *Exploring early usage patterns of mobile data services.* Paper presented at the International Conference on Mobile Business, July 11-13, 2005, Sydney, Australia.

Bozionelos, N. (1996). Psychology of computer use: prevalence of computer anxiety in British managers and professionals. *Psychological Reports, 78*(3), 995-1002.

Calder, B. J., & Staw, B. M. (1975). Self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology, 31*(4), 599-605.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Carlsson, C., Hyvonen, K., Repo, P., & Walden, P. (2005). *Adoption of mobile services across different platforms.* Paper presented at the 18th Bled eCommerce Conference, June 6-8, Bled, Slovenia.

Casal, C. R., Burgelman, J. C., & Bohlin, E. (2004). Propects beyond 3G. *Info, 6*(6), 359-362.

Chae, M., & Kim, J. (2003). What's so different about the mobile Internet? *Communications of the ACM, 46*(12), 240-247.
Chae, M., & Kim, J. (2004). Do size and structure matter to mobile users? an empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour & Information Technology, 23*(3), 165-181.
Chau, P. Y. K. (1996). An empirical assessment of a modified technology acceptance model. *Journal of Management Information Systems, 13*(2), 185-204.
Chau, P. Y. K., & Hu, P. J.-H. (2001). Information technology acceptance by individual professionals: a model comparison approach. *Decision Science, 32*(4), 699-719.
Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: development of a measure and initial test. *MIS Quarterly, 23*(2), 189-211.
Condos, C., James, A., Every, P., & Simpson, T. (2002). Ten usability principles for the development of effective WAP and m-commerce services. *Aslib Proceedings, 54*(6), 345-355.
Crawford, A. M. (2002). International media habits on the rise. *Ad Age Global, 2*(11).
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance in information technology. *MIS Quarterly, 13*(3), 319-340.
Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989a). User acceptance of computer technology: a comparison of two theoretical models. *Management Science, 35*(8), 982-1002.
Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology, 22*, 1111-1132.
Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
Dholakia, R. R., & Dholakia, N. (2004). Mobility and markets: emerging outlines for m-commerce. *Journal of Business Research, 57*(12), 1391-1396.
Dillon, A., & Morris, M. (1996). User acceptance of information technology: theories and models. *Journal of American Society for Information Science, 31*, 3-32.
Elliot, G., & Phillips, N. (2004). *Mobile Commerce and Wireless Computing Systems*. Harlow: Pearson Education Limited.
Elliot, S., & Loebbecke, C. (2000). Interactive, inter-organizational innovations in electronic commerce. *Information Technology & People, 13*(1), 46-66.
Figge, S. (2004). Situation-dependent services - a challenge for mobile operators. *Journal of Business Research, 57*(12), 1416-1422.
Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: an introduction to theory and research*. Reading, MA: Addison-Wesley.
Funk, J. L. (2005). The future of the mobile phone Internet: an analysis of technological trajectories and lead users in the Japanese market. *Technology in Society, 27*(1), 69-83.
Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: an integrated model. *MIS Quarterly, 27*(1), 51-90.
Horton, R. P., Buck, T., Waterson, P. E., & Clegg, C. W. (2001). Explaining intranet use with the technology acceptance model. *Journal of Information Technology, 16*, 237-249.
Hung, S.-Y., Ku, C.-Y., & Chang, C.-M. (2003). Critical factors of WAP services adoption: an empirical study. *Electronic Commerce Research and Applications, 2*(1), 42-60.
Igbaria, M., Parasuraman, S., & Baroudi, J. J. (1996). A motivational model of microcomputer usage. *Journal of Management Information Systems, 13*(1), 127-143.
Ishii, K. (2004). Internet use via mobile phone in Japan. *Telecommunications Policy, 28*(1), 43-58.
Jiang, J. J., Hsu, M. K., Klein, G., & Lin, B. (2000). E-commerce user behaviour model: an empirical study. *Human Systems Management, 19*(4), 265-276.
Kaufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behaviour. *Information Systems Research, 13*(2), 205-223.
Khalifa, M., & Cheng, S. K. N. (2002). *Adoption of mobile commerce: role of exposure.* Paper presented at the 35th Hawaii International Conference on System Sciences, Hawaii.
King, J. L., Gurbaxani, V., Kraemer, K. L., McFarlan, F. W., Raman, K. S., & Yap, C. S. (1994). Institutional factors in information technology innovation. *Information Systems Research, 5*(2), 139-169.
Kleijen, M., Wetzels, M., & de Ruyter, K. (2004). Consumer acceptance of wireless finance. *Journal of Financial Services Marketing, 8*(3), 206-217.
Knutsen, L., Constantiou, I. D., & Damsgaard, J. (2005). *Acceptance and perceptions of advanced mobile services: alterations during a field study.* Paper presented at the International Conference on Mobile Business, July 11-13, 2005, Sydney, Australia.

Lee, M. S. Y., McGoldrick, P. J., Keeling, K. A., & Doherty, J. (2003). Using ZMET to explore barriers to the adoption of 3G mobile banking services. *International Journal of Retail & Distribution Management, 31*(6), 340-348.

Leonard-Barton, D., & Deschamps, I. (1988). Managerial influence in the implementation of new technology. *Management Science, 34*(10), 1252-1265.

Lin, H., & Wang, Y. (2005). *Predicting consumer intention to use mobile commerce in Taiwan.* Paper presented at the International Conference on Mobile Business, July 11-13, 2005, Sydney, Australia.

Lu, J., Yu, C., Liu, C., & Yao, J. E. (2003). Technology acceptance model for wireless internet. *Internet Research: Electronic Networking Applications and Policy, 13*(3), 206-222.

Massey, A. P., Khatri, V., & Ramesh, V. (2005). *From the web to the wireless web: technology readiness and usability.* Paper presented at the 38th Hawaii International Conference on System Sciences, Hawaii.

Mathwick, C., Malhotra, N., & Rigdon, E. (2001). Experiental value: conceptualization, measurement and application in the catalog and Internet shopping environment. *Journal of Retailing, 77*(1), 39-56.

Minton, G. C., & Scheneider, F. W. (1980). *Differential Psychology*. Prospect Heights, IL: Waveland Press.

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management, 38*(4), 217-230.

Novak, T. P., Hoffman, D. L., & Yung, Y. (2000). Measuring the customer experience in online environments: a structural modeling approach. *Marketing Science, 19*(1), 22-42.

Pagani, M. (2004). Determinants of adoption of third generation mobile multimedia services. *Journal of Interactive Marketing, 18*(3), 46-59.

Parasuraman, A. (2000). Technology readiness index: a multiple item scale to measure readiness to embrace new technologies. *Journal of Service Research, 2*(4), 307-320.

Rhodes, S. R. (1983). Age-related differences in work attitudes and behavior: a review of conceptual analysis. *Psychological Bulletin, 93*(2), 328-367.

Rogers, E. M. (1995). *Diffusion of Innovatoins*. New York: Free Press.

Saaksjarvi, M. (2003). Consumer adoption of technological innovations. *European Journal of Innovation Management, 6*(2), 90-100.

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theoryof reasoned action: a meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*(3), 325-343.

Siau, K., Lim, E. P., & Shen, Z. (2001). Mobile commerce: promises, challenges, and research agenda. *Journal of Databases Management, 12*(2), 4-13.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: a test of competing models. *Information Systems Research, 6*(2), 144-176.

Taylor, S., & Todd, P. A. (1995a). Assessing IT usage: the role of prior experience. *MIS Quarterly, 19*(4), 561-570.

Teo, T. S. H., & Pok, S. H. (2003). Adoption of WAP-enabled mobile phones among Internet users. *Omega: The International Journal of Management Science, 31*(6), 483-498.

Terry, D. J. (1993). Self-efficacy expectancies and the theory of reasoned action. In D. C. Terry, C. Gallois & M. McCamish (Eds.), *The Theory of Reasoned Action: Its Application to AIDS-preventive Behaviour*. Oxford: Pergamon.

Thompson, R., Higgins, C., & Howell, J. (1994). Influence of experience on personal computer utilization: testing a conceptual model. *Journal of Management Information Systems, 11*(1), 167-187.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science, 46*(2), 186-204.

Venkatesh, V., & Morris, M. G. (2000a). Why don't men ever stop to ask for directions? gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115-139.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly, 27*(3), 425-478.

Wang, Y.-S., Wang, Y.-M., Lin, H.-H., & Tang, T.-I. (2003). Determinants of user acceptance of Internet banking: an empirical study. *International Journal of Service Industry Management, 14*(5), 501-519.

Yen, D. C., & Chou, D. C. (2000). Wireless communications: applications and managerial issues. *Industrial Management & Data Systems, 100*(9), 436-443.

E-Commerce Adoption Barriers in Small Businesses: A Comparative Study of Three Regional Areas in New South Wales (Australia)

Dr Rob MacGregor
School of Information Technology and Computer Science
University of Wollongong, Wollongong, Australia
rmacgreg@uow.edu.au

Dr Lejla Vrazalic
Information Systems
University of Wollongong
lejla@uow.edu.au

Dr Deborah Bunker
School of Information Systems, Technology and Management
University of New South Wales
d.bunker@unsw.edu.au

Abstract

The adoption of e-commerce in small businesses located in Australian regional areas remains alarmingly low. The study presented in this paper indicates an overall adoption rate of less than 15% in three major regional areas in the state of New South Wales – the Illawarra, the Hunter and the Central West. The purpose of the study was to examine and compare e-commerce adoption barriers across the three regions in order to determine whether the same barriers apply in all locations. Furthermore, the study also investigated the effects of organisational factors such as business size and the gender of the primary owner/manager on e-commerce adoption barriers. The results show that there e-commerce barriers across the three regions are not standard and that specific organisational factors are associated with small businesses located in the Illawarra region.

Keywords:

E-commerce, barriers, SMEs, regional, Australia

INTRODUCTION

It has been well documented that small businesses, through the advent of e-commerce, are increasingly turning to global markets. E-commerce, defined as "the buying and selling of information, products, and services via computer networks" (Kalakota & Whinston, 1997, p.3), has been shown to provide a cost-effective way of accessing customers by being 'wired to the global marketplace'. Yet despite the exponential growth of e-commerce, recent studies (MacGregor, 2004; MacGregor et al, 2004) of the Australian small business sector, particularly the regional small business sector, have shown that it is lagging behind many similar economies in Europe and North America.

The sluggish pace of e-commerce diffusion in the small business sector has been attributed to various adoption barriers that are faced by small businesses. These barriers have been well documented in previous research studies (Riquelme, 2002; Van Akkeren & Cavaye, 1999; Quayle, 2002; Purao & Campbell, 1998; Lawrence, 1997). Studies have also shown that unlike the large business sector, many of the adoption barriers are associated with organisational factors such as the number of employees, market focus, business structure, and alliances with other businesses. However, relatively little research has been undertaken into e-commerce adoption barriers across different locations, particularly regional locations. Our understanding of how geograohically universal e-commerce adoption barriers are remains limited. This paper attempts to add to the existing literature on e-commerce adoption in small businesses by presenting the results of three studies carried out in regional small businesses located in the most populous state in Australia – New South Wales.

The study presented in this paper represents some of the findings of a larger, ongoing study of e-commerce adoption in small businesses located in regional areas of developed OECD countries (including Sweden, Australia and the USA). The research presented here is part of an extended study of e-commerce adoption in regional small businesses located in Australia only. The results from other locations are reported elsewhere (see: MacGregor et al, 2004). Regional small businesses were chosen due to the disadvantages inherent to their location (Larsson et al, 2003; Keniry et al, 2003). While it is beyond the scope of this paper to describe the entire Australian study (which examined a number of issues associated with e-commerce), one of the primary aims of the research was to determine whether any differences existed between e-commerce adoption in small businesses located in three different regional areas.

The paper begins by briefly examining the nature of small businesses. A consolidated discussion of barriers to e-commerce adoption based on previous research is then presented, followed by the methodology. A series of chi-square tests were carried out to determine whether there are significant differences in the perception of importance of barriers across the three locations. A second series of chi-square tests were applied to the perception of the importance of the barriers in each location to determine whether the association of the organisational factors (number of employees, market focus, business structure, alliances with other businesses and gender of the owner/manager) differs from location to location. The results of these statistical analyses are then discussed. Finally, the limitations of the study are presented and conclusions drawn.

SMALL BUSINESSES

There are a number of definitions of what constitutes a small business. Some of these definitions are based on quantitative measures such as staffing levels, turnover or assets, while others employ a qualitative approach. Meredith (1994) suggests that any description or definition must include a quantitative component that takes into account staff levels, turnover, assets together with financial and non-financial measurements, but that the description must also include a qualitative component that reflects how the business is organised and how it operates. In this paper, small businesses are defined as organizations employing less than 50 people, while medium enterprises have more than 50 but less than 250 employees. While these figures may vary from previous studies, an examination of recent literature shows that more often than not, these figures are becoming the international norm.

Qualitatively, any description of an small business must be premised on the notion that they are not simply scaled down large businesses (Wynarczyk et al 1993) and although size is a major distinguishing factor, small businesses have a number of unique features that set them apart from larger businesses. Small businesses tend to be more prone to risk than their larger counterparts (Delone, 1988). Cochran (1981) found that small businesses tended to be subject to higher failure rates while Welsh and White (1981) found that small businesses suffered from a lack of trained staff and had a short-range management perspective. They termed these traits 'resource poverty' and suggested that their net effect was to magnify the effects of environmental impacts, particularly where information systems were involved. These early studies have been supported by more recent research that has found most small businesses lack technical expertise (Barry & Milner, 2002), and adequate capital to undertake technical enhancements (Raymond, 2001).

These characteristics that are unique to the small business sector set the context for this study. However they also provide an insight into some of the reasons why small businesses have found it more difficult to take-up e-commerce technologies. Furthermore, small businesses located in regional areas have faced more problems than their counterparts in major cities and metropolitan areas.

SMALL BUSINESSES IN REGIONAL AREAS

Small businesses located in regional areas are affected by circumstances inherent to their location. The Australian Bureau of Statistics defines regional areas as geographical areas located outside metropolitan centres and major cities and classifies these areas as inner and outer regions, remote and very remote areas. The study presented in this paper only examined urbanised regional areas (which include inner and outer regions), rather than rural areas (which include remote and very remote areas). Regional areas are of particular interest to governments because they are characterised by high unemployment rates (Larsson et al, 2003), a shortage of skilled people, limited access to resources and a lack of infrastructure (Keniry et al, 2003). Yet, at the same time, businesses located in regional areas have the potential to play a major role in developing those areas. This potential has not gone unnoticed by government organisations (Keniry et al, 2003).

To encourage growth and development in regional areas, government organisations have been heavily promoting the adoption of information and communication technology (ICT) by small businesses. The most well known of these initiatives was the “Networking the Nation” program, which was launched by the Commonwealth Government in 1997 to promote the economic development of regional and rural areas through sustainable use of ICT and the provision of ICT infrastructure. Yet, despite such programs and initiatives, the rate of e-commerce adoption in small businesses has been reported as being low. The reasons for this are diverse, however they are generally categorised as barriers or inhibitors to e-commerce adoption. The following section will examine some of the barriers to e-commerce adoption that are faced by small businesses.

BARRIERS TO E-COMMERCE ADOPTION IN SMALL BUSINESSES

E-commerce has been widely touted as providing small businesses with an opportunity for instant access to global markets and customers (Coviello & McAuley, 1999). However, research shows that it is mostly larger businesses that have benefited from e-commerce adoption (Riquelme, 2002) with small businesses showing a much slower pace of adoption. The reasons for this are diverse and have been examined in various studies as inhibitors or barriers that prevent small businesses from adopting and, subsequently fully reaping the benefits of e-commerce.

Hadjimanolis (1999), in a study of e-commerce adoption by small businesses in Cyprus, classified these barriers into two types: internal and external. External barriers could be further categorised into supply barriers (difficulties obtaining finance and technical information), demand barriers (e-commerce not fitting with the products/services or not fitting with the way clients did business) and environmental barriers (security concerns). Internal barriers were further subdivided into resource barriers (lack of management and technical expertise) and system barriers (e-commerce not fitting with the current business practices). A summary of different e-commerce adoption barriers in small businesses based on an extensive literature review is presented in Table 1.

Table 1: Summary of Research - E-commerce adoption barriers in small businesses

Barriers to E-Commerce Adoption	Related Literature
High cost of implementation; Internet technologies too expensive to implement	Riquelme (2002), Van Akkeren & Cavaye (1999), Purao & Campbell (1998), Lawrence (1997), Iacovou et al (1995)
E-commerce is too complex to implement	Fielding (1996), Quayle (2002)
Small businesses require short-term ROI and e-commerce is long-term	Lawrence (1997), McGowan & Madey (1998)
Resistance to change because of the fear of new technology amongst employees	Van Akkeren & Cavaye (1999), Lawrence (1997)
Preference for and satisfaction with traditional manual methods (phone, fax, etc)	Lawrence (1997) Venkatesan & Fink (2002)
Lack of technical skills and IT knowledge amongst employees; Lack of computer literate/specialised staff	Riquelme (2002), Van Akkeren & Cavaye (1999), Lawrence (1997), Iacovou et al (1995), Quayle (2002), Damsgaard & Lyytinen (1998)
Lack of time to implement e-commerce	Van Akkeren & Cavaye (1999), Lawrence (1997) Walczuch et al (2000)
E-commerce not deemed to be suited to the way the organisation does business, or the way our clients do business	Poon & Swatman (1997), Hadjimonolis (1999), Iacovou et al (1995)
E-commerce not deemed to be suited to the products/services	Poon & Swatman (1997), Hadjimonolis (1999)
E-commerce perceived as a technology lacking direction	Lawrence (1997)
Lack of awareness about business advantages/opportunities e-commerce can provide	Iacovou et al (1995), Quayle (2002)
Lack of available information about e-commerce	Lawrence (1997)
Concern about security of e-commerce	Riquelme (2002), Van Akkeren & Cavaye (1999) Purao & Campbell (1998), Hadjimonolis (1999), Quayle (2002)
Lack of critical mass among customers, suppliers and business partners	Hadjimonolis (1999)
Heavy reliance on external consultants to provide necessary expertise	Van Akkeren & Cavaye (1999), Lawrence (1997)
Lack of e-commerce standards	Tuunainen (1998), Robertson & Gatignon (1986)

A number of previous studies (MacGregor, 2004; Vrazalic et al, 2002) have shown that while the size of the small business (in terms of employee numbers) appears to have a bearing on the importance of barriers to e-commerce adoption, e-commerce barriers would seem to be uniform across different business types. Other studies (Yap et al, 1992; Thong et al, 1996; MacGregor & Bunker, 1996) have shown that while vendor support, as a surrogate for IT skill, can reduce the impact of technical barriers, there is no equivalent reduction in organisational barriers.

There are limited studies (MacGregor et al, 2004) available about the extent of the e-commerce barriers in regional areas. However, they have shown that many of the barriers listed in Table 1, particularly those related to technical infrastructure, are more pronounced in businesses located in regional areas than those in metropolitan areas.

Despite the relatively large number of studies that have examined e-commerce adoption barriers in small businesses, few studies have concentrated their efforts on regional areas, and even fewer on comparing e-commerce barriers across different locations. The study described in the following section aims to address these issues as a starting point for filling the gap in our current knowledge.

METHODOLOGY

A number of research frameworks, including Davis' (1989) Technical Acceptance Model (TAM), Rogers' (1995) Diffusion of Innovation theory, and the use of Grounded Theory (Glaser & Strauss, 1967) were considered and examined. While TAM is concerned with a system's perceived usefulness and ease of use, these factors are normally considered from an individual's viewpoint and in a static business structure. Several studies (Lee, 2001; Damanpour, 2001) have suggested that for e-commerce adoption to be valid, organisational re-structuring must occur. Furthermore, this study was concerned with e-commerce barriers from an organisational decision-making, rather than a, personal perspective. As such, TAM was considered to be unsuitable. Rogers' (1995) theory of innovation diffusion was also considered. While it may be argued that e-commerce adoption in regional small businesses is a diffusion of innovation, Rogers' theory requires the 'trialability' of technology. As this study is concerned with non-adopters, Rogers' framework was found to be unsuitable for the research objectives. Furthermore, diffusion of innovation defines how an innovation is spread through a social system. Finally, as there were no overriding hypotheses, Grounded Theory was also examined. However, since the study did not employ ethnographic research methods, this framework was also rejected.

A number of previous studies (Teo & Tan, 1998; Shiels et al, 2003; Daniel, 2003), examining the role of business characteristics in the adoption and non-adoption of e-commerce by small businesses were examined. In all cases, the authors expressed the view that no framework adequately suited their approach and that they were ultimately attempting to develop a model from quantitative analysis. As this study (and others that have been derived from the data presented in this paper) are ultimately attempting to develop a model of non-adoption of e-commerce by small businesses, a similar viewpoint is taken.

The study presented here is part of a larger ongoing study into IT and e-commerce adoption in small to medium enterprises which began in 2002. The larger study was primarily concerned with small businesses located in regional areas, especially since no other research has investigated e-commerce adoption in these areas specifically. As a result the study was conceived primarily as exploratory in nature. The larger study has been undertaken in Sweden and Australia, and more recently in the USA, with all three countries having similar economies according to the Organisation for Economic Co-operation and Development (OECD) classification, thus allowing for comparisons. The portion of the study presented in this paper is concerned only with e-commerce adoption barriers in three regional areas in the state of New South Wales (Australia). Other issues such as e-commerce adoption drivers, and advantages and disadvantages of e-commerce, and the results from the Swedish and USA studies, are reported elsewhere (see: MacGregor et al 2004).

An examination of the literature concerned with e-commerce adoption in small businesses identified ten of the most frequently cited e-commerce barriers, all of which were uniquely different, and common to the studies examined. In order to both determine the applicability of these barriers and to ensure that no other barriers have been excluded, a series of six in-depth semi-structured interviews were conducted in Australia, where the study was conceived. All interviews were with owners/managers of small businesses (three of those had adopted e-commerce, while the other there had not). All of the identified barriers were found to applicable and no additional barriers were forthcoming by the majority of the interviewees. Based on the six in-depth interviews, a survey instrument was developed to collect data about e-commerce adoption barriers (amongst other things). A series of pilot tests of the instrument were conducted in regional small businesses in Australia (see MacGregor et al 2004). The survey was administered to owners of small businesses as the ultimate decision-makers for e-commerce adoption or non-adoption (Reynolds et al, 1994; Welsh & White, 1981). Respondents who had not adopted e-commerce were asked to rate the importance of each barrier to their decision not to adopt e-commerce using a standard 5 point Likert scale (as shown in Figure 1 below). The Likert scale responses were assumed to posses the characteristics of an interval measurement scale for data analysis purposes.

The survey also contained questions used to build a profile of the respondents, including the size of the small business (measured in terms of number of full-time equivalent employees), the type of business, the market focus, alliances with other organisations and the gender of the primary owner/manager. These factors have been termed "organisational factors" and Table 2 shows the scales that were used to capture them.

23. This question relates to the reasons why your organisation is not using e-commerce. Below is a list of statements indicating possible reasons. Please rate each statement on a scale of 1 to 5 to indicate how important each reason was to your decision NOT to use e-commerce, as follows:

1 = the reason was very unimportant to your decision not to use e-commerce
2 = the reason was unimportant to your decision not to use e-commerce
3 = the reason had little importance to your decision not to use e-commerce
4 = the reason was important to your decision not to use e-commerce
5 = the reason was very important to your decision not to use e-commerce

Our organisation does not use e-commerce because:	**Rating**				
E-commerce is not suited to our products/ services.	1	2	3	4	5
E-commerce is not suited to our way of doing business.	1	2	3	4	5
E-commerce is not suited to the ways our clients (customers and/or suppliers) do business.	1	2	3	4	5
E-commerce does not offer any advantages to our organisation.	1	2	3	4	5
We do not have the technical knowledge in the organisation to implement e-commerce.	1	2	3	4	5
E-commerce is too complicated to implement.	1	2	3	4	5
E-commerce is not secure.	1	2	3	4	5
The financial investment required to implement e-commerce is too high for us.	1	2	3	4	5
We do not have time to implement e-commerce.	1	2	3	4	5
It is difficult to choose the most suitable e-commerce standard with so many different options available.	1	2	3	4	5

Figure 1: Question about barriers to e-commerce adoption used in survey

Table 2: Scales used for capturing Organisational Factors

Number of employees (FTE)	**Business type**	**Market focus**	**Private Alliance Membership**	**Government Alliance Membership**	**Gender of Primary Owner/Manager**
0 1-5 6-10 11-20 21-5- 51-200	Sole owner Partnership Incorporated	Local Regional National International	Yes No	Yes No	Male Female

The following criteria were developed and applied to several areas in New South Wales (Australia) in determining the areas for investigation:

- The location must be an urban regional area, and not a major/capital city or rural area;
- A viable government initiated Chamber of Commerce must exist and be well patronised by the small business community;
- The location should have access to a full range of educational facilities (including a University/College);
- The business community must represent a cross-section of business ages, sizes and market sectors.

As a result of applying the above criteria, three regional areas in New South Wales were selected: the Central West region, the Hunter region and the Illawarra region. A total of 289 surveys were administered by phone to randomly selected small businesses in each region – 55 in the Central West, 74 in the Hunter and 160 in the Illawarra. The mode of the data collection was selected based on previous research by de Heer (1999) that indicated that Australia had low mail survey response rates.

RESULTS

Of the total number of respondents, 49 in the Central West region, 62 in the Hunter region and 135 in the Illawarra region indicated that they had examined and rejected the adoption of e-commerce. The overall non-adoption rate for all three regions combined was 85.4%. The e-commerce adoption rates for the three individual regions are shown in Table 3.

All responses were examined to determine completeness. The responses formed the basis for the statistical analysis carried out using SPSS. An inspection of the frequencies indicated, in all three locations, that the full range of the scale was utilised by the respondents (i.e. every barrier had at least one instance of each rating from 1 to 5).

Table 3: E-Commerce Adoption Rates in 3 New South Wales Regions

Region	Adoption Rate
Central West	10.9%
Hunter	16.2%
Illawarra	15.2%

A series of Chi-Square tests was applied to the data to determine whether there were any statistically significant differences in the perception of importance of the barriers to e-commerce adoption between the locations. Six of the ten barriers showed differences, as shown in Table 4.

Table 4: Differences in Perception of Barriers to E-commerce Adoption across the 3 Regions

EC doesn't offer any advantages		**Very Unimportant**	**Unimportant**	**Little Importance**	**Important**	**Very Important**	**Missing**
	Cent.West	3 (6.1%)	19 (38.8%)	10 (20.4%)	7 (14.3%)	5 (10.2%)	5(10.2%)
	Hunter	8 (12.9%)	16 (25.8%)	20 (32.3%)	0 (0%)	6 (19.4%)	12(19.4%)
	Illawarra	11 (8.1%)	22 (16.3%)	33 (24.4%)	20(14.8%)	42(31.1%)	7(5.2%)
p<.001							
Too complicated		**Very Unimportant**	**Unimportant**	**Little Importance**	**Important**	**Very Important**	**Missing**
	Cent.West	4 (8.2%)	12 (24.5%)	10 (20.4%)	10(20.4%)	9 (18.4%)	4(8.2%)
	Hunter	4 (6.5%)	8 (12.9%)	14 (22.6%)	18(29%)	12(19.4%)	6(9.7%)
	Illawarra	17 (12.6%)	31 (23%)	24 (17.8%)	12 (8.9%)	42(31.1%)	9(6.7%)
p<.05							
Not secure		**Very Unimportant**	**Unimportant**	**Little Importance**	**Important**	**Very Important**	**Missing**
	Cent.West	3 (6.1%)	7 (14.3%)	12 (24.5%)	16(32.7%)	7 (14.3%)	4(8.2%)
	Hunter	6 (9.7%)	10 (16.1%)	18 (29%)	18((29%)	4 (6.5%)	6(9.7%)
	Illawarra	21 (15.6%)	25 (18.5%)	40 (29.6%)	13 (9.6%)	28(20.7%)	8(5.9%)
p< .05							
Too costly		**Very Unimportant**	**Unimportant**	**Litle Importance**	**Important**	**Very Important**	**Missing**
	Cent.West	2 (4.1%)	10 (20.4%)	8 (16.3%)	18(36.7%)	7 (14.3%)	4(8.2%)
	Hunter	6 (9.7%)	6 (9.7%)	10 (16.1%)	28(45.2%)	6 (9.7%)	6(9.7%)
	Illawarra	22 (16.3%)	31 (23%)	18 (13.3%)	22(16.3%)	33(24.4%)	8(5.9%)
p< .001							
Limited time		**Very Unimportant**	**Unimportant**	**Little Importance**	**Important**	**Very Important**	**Missing**
	Cent.West	4 (8.2%)	8 (16.3%)	17 (34.7%)	11(22.4%)	5 (10.2%)	4(8.2%)
	Hunter	8 (12.9%)	12 (19.4%)	22 (35.5%)	10(16.1%)	4 (6.5%)	6(9.7%)
	Illawarra	20 (14.8%)	25 (18.5%)	21 (15.6%)	24(17.8%)	36(26.7%)	9(6.7%)

p< .01							
Too many options		**Very Unimportant**	**Unimportant**	**Little Importance**	**Important**	**Very Important**	**Missing**
	Cent.West	4 (8.2%)	7 (14.3%)	20 (40.8%)	9 (18.4%)	5 (10.2%)	4(8.2%)
	Hunter	6 (9.7%)	16 (25.8%)	24 (38.7%)	6 (9.7%)	4 (6.5%)	6(9.7%)
	Illawarra	24 (17.8%)	29 (21.5%)	27 (20%)	16(11.9%)	27 (20%)	12(8.9%)
p< .05							

A second series of chi-square tests was undertaken in each of the three locations to determine whether any of the factors (number of employees, market focus, business structure, alliances with other businesses and gender of the owner/manager) was associated with the perception of importance of any of the barriers to e-commerce adoption. The data showed that there were no associations of any of the factors with barriers to e-commerce adoption in either the Central West or the Hunter regions. Two barriers were each associated with one of the factors in the Illawarra region. These are shown in Tables 5 and 6 below.

Table 5: E-Commerce Barrier: Cost too high (Illawarra region)

Factor – No. of Employees	Very Unimportant	Unimportant	Little Importance	Important	Very Important
0	4	0	3	9	10
1-9	14	25	12	12	21
10-19	2	5	2	0	0
20-50	3	1	1	0	2

p<.05

Table 6: E-Commerce Barrier: Too many options (Illawarra region)

Factor – Gender	V.Unimportant	Unimportant	Little Importance	Important	V.Important
Female	3	12	7	10	18
Male	21	20	20	9	12

p<.05

DISCUSSION

Before examining the data in detail, a number of comments are appropriate. Firstly, while the rating of importance of barriers differs from one location to another, the results suggest that each of the ten barriers is applicable to regional small businesses in Australia. Secondly, the results show that there is a very low uptake of e-commerce in regional small businesses in New South Wales. The overall adoption rate for all three regions combined is under 15% which is significantly lower than the adoption rate in similar regions in other OECD countries. This is a major source of concern and also highlights that efforts to promote ICT, and e-commerce, adoption in regional small businesses have not had the desired effect. One of the major outcomes of the study presented in this paper is the necessity of reviewing current initiatives aimed at promoting e-commerce adoption and developing more effective strategies with a systemic focus that will assist small businesses to get “e-commerce ready”.

An examination of Table 4 shows that despite the three locations being similar, and indeed subject to the same government incentives and regulations, the importance of barriers is substantially different. While 27% of the Central West and 12% of the Hunter consider that e-commerce offers few advantages to their business, 48.4% of the Illawarra respondents gave a similar response. This is a somewhat surprising result since the Illawarra region is considered to be the most economically developed of the three regions due to its proximity to Sydney. In all cases, a very high percentage of respondents felt that e-commerce was too complicated, however, the Hunter respondents showed a 10% higher response than the other two regions.

Security and cost have always been issues precluding many small businesses from adopting e-commerce (Riquelme, 2002; Quayle, 2002; Lawrence, 1997; Iacovou et al, 1995). An examination of the data in Table 4 shows that there appears less concern with these barriers in the Illawarra than the other two regions. The Illawarra respondents are, at best, equivocal on both these concerns. Indeed, it would appear that small businesses in the Illawarra were making e-commerce adoption decisions based on the advantages that would flow from e-commerce, rather than any financial or technical (security) considerations. It is also interesting to note that the Illawarra respondents are more concerned with limited time to consider or undertake adoption of e-commerce than the other two regions. The implication is that small businesses in the Illawarra are "too busy" with day-to-day business operations to find the time to implement an e-commerce system.

A number of studies (Hyland & Matlay, 1997; Moran, 2000; Blackburn & Athayde, 2000; Mazzarol et al, 1999) have shown that factors such as the number of employees, market focus, business structure, alliances with other businesses and the gender of the owner/manager appear to affect the decision-making processes (either to adopt e-commerce or reject it). An examination of Tables 5 and 6 shows that for two locations the findings of these authors is not supported. Indeed, only two barriers (cost of e-commerce and large number of choices) were associated in any way with any of the organisational factors and then only in the Illawarra region. As would be expected, small businesses are far more conscious of costs than medium-sized enterprises. This result is in accord with the findings of Quayle (2002). However, the result begs the question as to why similar findings were not in evidence in either the Central West or the Hunter regions. An examination of Table 6 shows that female respondents are far more concerned with the plethora of options in the uptake of e-commerce than their male counterparts. This would appear to support previous research studies (Gilroy & Desai, 1986; Meier & Lambert, 1991; Perez et al, 2002) which suggest that males were less anxious about computing technology than females, leading to a greater concern amongst women about the different e-commerce systems available.

LIMITATIONS

It should be noted that the study presented here has several limitations. The data for the study was collected from various industry sectors so it is not possible to make sector specific conclusions. Also, the choice of variables selected for the study is somewhat problematic because of the complex nature of adoption barriers which change over time. Furthermore, according to Sohal and Ng (1998), the views expressed in the surveys are of a single individual from the responding organisation, and only those interested in the study are likely to complete and return the survey. Finally, this is a quantitative study, and further qualitative research is required to gain a better understanding of the key issues.

CONCLUSION

The main aim of this paper was to determine whether barriers to e-commerce adoption were standard and thus perceived in the same way by small businesses in different, but inherently similar locations. A study of small businesses in three regional areas of New South Wales (Australia) was undertaken. The results showed not only that the perception of importance varied across the three locations, but that the association of factors (the number of employees, market focus, business structure, alliances with other businesses and the gender of the owner/manager) differed between the locations.

The implications of the study are significant for government organisations engaged in promoting e-commerce adoption, especially in regional small businesses. This research suggests that small business variations do exist depending on the location of the small business. This would suggest that any strategies developed to promote e-commerce in small businesses located in different areas should address barriers differently. Further research needs to take place, both to address the reasons behind the differences and possible solutions to alleviate the concerns of small businesses.

REFERENCES

Australian Bureau of Statistics (2001) www.abs.gov.au, 10.12.03.

Barry, H. & Milner, B. (2002) SMEs and Electronic Commerce: A Departure from the Traditional Prioritisation of Training?, *Journal of European Industrial Training*, 25(7), pp 316-326.

Blackburn R. & Athayde R. (2000) Making the Connection: The Effectiveness of Internet Training in Small Businesses *Education and Training* vol 42, no. 4/5

Blili, S. & Raymond, L. (1993) 'Threats and Opportunities for Small and Medium-Sized Enterprises', *International Journal of Information Management*, Vol 13, pp. 439-448.

Bunker, D.J. & MacGregor, R.C. (2000) 'Successful Generation of Information Technology (IT) Requirements for Small/Medium Enterprises (SME's) – Cases from Regional Australia', *Proceedings of SMEs in a Global Economy*, Wollongong, Australia, pp. 72-84.

Bunker, D.J. & MacGregor, R.C. (2002) The Contextof Information Technology and Electronic Commerce Adoption in Small/Medium Enterprises *Proceedings of the Americas Conference on Information Systems*

Cochran, A.B. (1981) Small Business Mortality Rates: A review of the Literature, *Journal of Small Business Management*, 19(4), pp 50-59.

Coviello, N. & McAuley, A. (1999) ‘Internationalisation and the Smaller Firm: A Review of Contemporary Empirical Research’, *Management International Review*, Vol 39, No 3, pp. 223-240.
Cragg, P.B. & King, M. (1993) ‘Small-firm computing: Motivators and inhibitors’, *MIS Quarterly*, Vol 17, No 1, pp. 47-60.
Damanpour F. (2001) E-Business E-Commerce Evolution: Perspective and Strategy *Managerial Finance* vol 27, no. 7, pp 16 - 33
Damsgaard, J. & Lyytinen, K. (1998) ‘Contours of diffusion of electronic data interchange in Finland: Overcoming technological barriers and collaborating to make it happen’, *Journal of Strategic Information Systems*, Vol 7, pp. 275-297.
Daniel E. (2003) An Exploration of the Inside-Out Model: E-commerce Integration in UK SMEs *Journal of Small Business Enterprise Development*vol. 10, no. 3, pp 233 - 249
DeLone, W.H. (1988) Determinants for Success for Computer Usage in Small Business, *MIS Quarterly*, 51-61.
Dennis C. (2000) ‘Networking for Marketing Advantage’, *Management Decision*, Vol 38, No 4, pp. 287-292.
Europa (2003) *SME Definition*, europa.eu.int/comm/enterprise/enterprise_policy/ sme_definition, 15.12.03.
Fallon M. & Moran P (2000) Information Communications Technology (ICT) and manufacturing SMEs *2000 Small Business and Enterprise Development Conference* University of Manchester, pp 100 - 109
Fielding, J. (1996) ‘Getting Out in Front with EDI’, *Inform*, Vol 10, No 9, pp. 12-18.
Frazer, L. & Lawley, M. (2000) *Questionnaire Design and Administration*, Wiley.
Gilbert D., Lee-Kelley L. & Barton M. (2003) Technophobia, Gender Influences and Consumer Decision-Making for Technology-Related Products *European Journal of Innovation Management* vol. 6, no. 4, pp 253 - 263
Hadjimonolis, A. (1999) ‘Barriers to Innovation for SMEs in a Small Less Developed Country (Cyprus)’, *Technovation*, Vol 19, No 9, pp. 561-570.
Hill, R. & Stewart, J. (2000) ‘Human Resource Development in Small Organisations’, *Journal of European Industrial Training*, Vol 24, No 2/3/4, pp. 105-117.
Hyland T. & Matlay H. (1997) Small Businesses, Training needs and VET Provisions *Journal of Education and Work* vol 10, no. 2
Iacovou, C.L., Benbasat, I. & Dexter, A.S. (1995) ‘Electronic Data Interchange and Small Organisations: Adoption and Impact of Technology’, *MIS Quarterly*, Vol 19, No 4, pp. 465-485.
Johansson, U. (2003) *Regional Development in Sweden: October 2003*, Svenska Kommunförbundet, www.lf.svekom.se/tru/RSO/Regional_development_in_ Sweden.pdf, 14.12.03.
Kalakota, R. & Whinston, A. (1997) *Electronic Commerce:A Manager’s Guide*, Addison-Wesley.
Keniry, J., Blums, A., Notter, E., Radford, E. & Thomson, S. (2003) *Regional Business – A Plan for Action*, Department of Transport and Regional Services, www.rbda.gov.au/ action_plan, 13.12.03
Larsson, E., Hedelin, L. & Gärling, T. (2003) Influence of Expert Advice on Expansion Goals of Small Businesses in Rural Sweden, *Journal of Small Business Management*, 41(2), pp 205-212.
Lawrence, K.L. (1997) ‘Factors Inhibiting the Utilisation of Electronic Commerce Facilities in Tasmanian Small- to Medium- Sized Enterprises’, *Proceedings of the 8th Australasian Conference on Information Systems*, Adelaide, pp. 587-597.
Lee C.S. (2001) An analytical Framework for Evaluating E-commerce Business Models and Strategies *Internet Research: Electronic Network Applications and Policy* vol 11, no. 4, pp 349 - 359
MacGregor R.C., Vrazalic, L., Bunker, D.J., Carlsson, S. & Magnusson, M. (2004) Comparison of Factors Pertaining to the Adoption and Non-Adoption of Electronic Commerce in Formally Networked and Non-Networked Regional SMEs: A Study of Swedish Small Businesses. In Corbitt B.J. & Al-Qirim N.A.Y (eds) *E-Business, E-Government & Small and Medium-Sized Enterprises: Opportunities and Challenges*, Idea Publishing Group, pp 206-243.
MacGregor, R.C. (2004) Factors Associated with Formal Networking in Regional Small Business: Some Findings from a Study of Swedish SMEs, *Journal of Small Business and Enterprise Development*, 11(1), pp 60-74.
MacGregor, R.C. & Bunker, D.J. (1996) ‘The Effect of Priorities Introduced During Computer Acquisition on Continuing Success with It in Small Business Environments’, *Proceedings of the Information Resource Management Association International Conference*, Washington, pp. 271-277.
MacGregor R.C. (2003) Are the Perceptions of Barriers to Electronic Commerce Adoption Different for Formally Networked and Non-Networked SMEs? *Journal of Systems and Information Technology* vol. 7, no. 1, pp 27 - 48
McGowan, M.K. & Madey, G.R. (1998) ‘The Influence Of Organization Structure And Organizational Learning Factors On the Extent Of EDI Implementation In U. S. Firms’, *Information Resources Management Journal*, Vol 11, No 3, pp. 17-27.
Martin, L.M. & Matlay, H. (2001) ‘“Blanket” Approaches to Promoting ICT in Small Firms: Some Lessons from the DTI Ladder Adoption Model in the UK’, *Internet Research: Electronic Networking Applications and Policy*, Vol 11, No 5, pp. 399-410.

Mazzarol T., Volery T., Doss N & Thein V. (1999) Factors Influencing Small Business Start-ups *International Journal of Entrepreneurial Behaviour and Research* vol 5, no. 2, pp 48 - 63
Miller, N.L. & Besser, T.L. (2000) 'The Importance of Community Values in Small Business Strategy Formation: Evidence from Rural Iowa', *Journal of Small Business Management*, Vol 38, No 1, pp. 68-85.
Poon, S. & Swatman, P. (1997) 'The Internet for Small Businesses: An Enabling Infrastructure', *Fifth Internet Society Conference*, pp. 221-231.
Purao, S. & Campbell, B. (1998) 'Critical concerns for small business electronic commerce: Some reflections based on interviews of small business owners', *Proceedings of the Association for Information Systems Americas Conference*, Baltimore, 14-16 August, pp. 325-327.
Quayle, M. (2002) 'E-commerce: the challenge for UK SMEs in the twenty-first century', *International Journal of Operations and Production Management*, Vol 22, No 10, pp. 1148-1161.
Raymond, L. (2001) Determinants of Web Site Implementation in Small Business *Internet Research: Electronic Network Applications and Policy*, 11(5), pp 411-422
Reynolds, W., Savage, W. & Williams, A. (1994) *Your own business: A Practical guide to success*, ITP.
Riquelme, H. (2002) 'Commercial Internet adoption in China: comparing the experience of small, medium and large businesses', *Internet Research: Electronic Networking Applications and Policy*, Vol 12, No 3, pp. 276-286.
Robertson, T. & Gatignon, H. (1986) 'Competitive Effects on Technology Diffusion', *Journal of Marketing*, Vol 50.
Rogers E.M. (1995) *Diffusion of Innovations* 4th Edition, New York, Free Press
Shiels H., McIvor R. & O'Reilly (2003) Understanding the Implications of ICT Adoption: Insights from SMEs *Logistics Information Management* vol. 16, no. 5, pp 312 - 326
Sohal, A. S. & Ng, L. 1998, The Role And Impact Of Information Technology In Australian Business, *Journal of Information Technology*, 13(3), pp 201-217.
Teo T.S.H. & Tan M. (1998) An Empirical Study of Adopters and Non-adopters of the Internet in Singapore *Information & Management* vol. 34 pp 339 - 345
Tetteh, E. and Burn, J. (2001) 'Global Strategies for SME-business: Applying the SMALL Framework' *Logistics Information Management*, Vol 14, No 1/2, pp. 171-180.
Tuunainen, V.K. (1998) 'Opportunities of Effective Integration of EDI for Small Businesses in the Automotive Industry', *Information & Management*, Vol 36, No 6, pp. 361-375.
Van Akkeren, J. & Cavaye, A.L.M. (1999) 'Factors Affecting Entry-Level Internet Technology Adoption by Small Business in Australia: An Empirical Study', *Proceedings of the 10th Australasian Conference on Information Systems*, Wellington, New Zealand, 1-3 December.
Venkatesan, V.S. & Fink, D. (2002) 'Adoption of Internet Technologies and E-Commerce by Small and Medium Enterprises (SMEs) in Western Australia', *Proceedings of the Information Resource Management Association International Conference*, pp. 1136-1137.
Walczuch, R., Van Braven, G. & Lundgren, H. (2000) 'Internet Adoption Barriers for Small Firms in The Netherlands', *European Management Journal*, Vol 18, No 5, pp. 561-572.
Welsh, J.A. & White, J.F. (1981) A Small Business is not a little Big Business, *Harvard Business Review*, 59(4), pp 46-58.
Westhead, P. & Storey, D.J. (1996) 'Management Training and Small Firm Performance: Why is the Link so Weak?', *International Small Business Journal*, Vol 14, No 4, pp. 13-24.
Wynarczyk, P., Watson, R., Storey, D.J., Short, H. & Keasey, K. (1993) *The Managerial Labour Market in Small and Medium Sized Enterprises*, Routledge.

Usage Discontinuance in the Context of Mobile Service

Siu-Man LUI
The Hong Kong University of Science and Technology
Clearwater Bay, Kowloon, Hong Kong
Email: imcarrie@ust.hk

Abstract

The mobile industry is probably the most dynamic and turbulent, characterized by the rapid technological change and evolving competition. This study aims at investigating the post-adoption behaviour of users in the mobile services industry. This paper details the development and empirical validation of a model of service provider switching, a post-adoption behaviour, by individuals. A research model is proposed based on the Theory of Planned Behaviour (TPB) as a framework. An online survey was used to test the proposed model. The results indicate that perceived switching cost and perceived alternative attractiveness are significant precursors to the attitude towards switching mobile service provider. In addition, perceived switching cost, perceived homogeneity and perceived competitiveness of the mobile market have a positives effect on the individual's perceived behavioural control toward switching service provider.

Keywords:

Mobile data services, switching intention, theory of planned behavior

INTRODUCTION

During the last decade, the mobile industry has experienced an explosive growth; mobile services are no longer a luxury but a popular consumer product. According to a recent report from the International Telecommunication Union (ITU), the number of mobile subscribers reached 1,155 million, as compared to 23 million in 1992 (ITU 2002).The phenomenal growth was partially attributable to the intense market competition and the ensuing price wars in which service providers shell out high subsidies to give new subscribers cheap or even free handsets. However, little evidence of brand loyalty has been found in this industry, which witnesses an average churn rate of 24% per year now, a rate that is expected to double to 50% when local number portability system is initiated (Meisner 2004). The high churn rate not only means that service providers will not be able to recoup their initial subsides, it also has negative effects on market share and profitability (Anderson, et al. 1994, Oliver 1997, Parthasarathy and Bhattacherjee 1998). To understand the high churn rate and possibly take actions to curb it, we need to first investigate the switching behaviour of the mobile service subscribers.

Besides the practical importance, a good understanding of service provider switching -- the behavioural consequence of discontinuance decision -- in the context of mobile service industry adds to the innovation adoption research on the differences between the post-adoption behaviour of mobile subscribers and that of other IT services (e.g., system development project) and products (e.g., spreadsheets, personal computers), especially at the later stage of the innovation diffusion cycle (Rogers 1995). Individual decision about an innovation is a process consisting of a series of actions and decisions. While prior research mainly has focused on the decision stage of the innovation-decision process, which occurs when an individual engages in activities that lead to his/her decision to adopt or reject an innovation, the current research concentrates on the confirmation stage, when individuals seek reinforcement of the innovation-decision already made or reverses the pervious adoption decision.

Compared with previous studies focusing on how and why individuals adopt an innovation (Venkatesh, et al. 2003), research in post-adoption behaviour is scant. Among the latter, user satisfaction (Devaraj, et al. 2002, Rai, et al. 2002, Van Dyke, et al. 1997) and usage continuance (Bhattacherjee 2001, Karahanna, et al. 1999, Venkatesh and Morris 2000) are the major issues under study. Only limited amount of research has been done on individuals' innovation discontinuance. For example, (Parthasarathy and Bhattacherjee 1998) distinguished potential discontinuers from continuing adopters by examining individual characteristics and perceptual beliefs related to online services. (Chen and Hitt 2002a) used Web site traffic data to measure switching cost for online service providers and identified individual and firm-specific characteristics affecting switching and attrition. (Lee, et al. 2003) utilized observations of individual web browsing activities panel dataset, which contained individual characteristics and web usage pattern, to explore the dynamics of consumer switching behaviour across major Internet portals. In summary, prior literature on innovation discontinuance mainly involved descriptive surveys and data mining. There has been no process study using behavioural model to understanding individuals' discontinuance.

The current work differs from previous innovation adoption research in that it shifts the focus onto a later stage of the innovation diffusion by investigating the individual's discontinuance decision. To examine the individual's post-adoption decision in perspective, we look into service provider switching with consideration of the need recognition stage in the consumer decision-making process, in particular, the factors that affect the user's intention to switch (i.e., recognition of the need for switch). In this study, a mobile service provider switching model based on the Theory of Planned Behaviour (TPB)(Ajzen 1991) is proposed for studying determinants of switching intention. The paper is organized as follows. First to be discussed are the characteristics of mobile services, which will help highlight the importance and relevance of the current work. This is followed by the development of a theoretical model, as well as the research methodology and empirical analysis for this study. Then, discussions and implications of the findings will be presented. Finally, the paper concludes with a look at the limitations of this study.

DISCONTINUANCE IN MOBILE SERVICES

In this section, we highlight the characteristics of mobile services, which help illustrate the relevance and importance of studying discontinuance in the context of mobile services.

Subscription Service Provision

The cost structure of mobile services posits the importance of investigating consumers' post-adoption behaviour and makes the differences between the post-adoption behaviour of mobile users and that of other IT products users. Mobile services are subscription-based IT services: the setup cost for subscribing mobile services is minimal but adopters are charged a flat fee or usage-based fee monthly for using the services. In contrast, the adoption of other IT products (e.g., personal computer) means substantial acquisition cost but minimal operating cost afterwards. Compared with the adoption of other IT products, that of mobile services is a less prominent decision; adopter can discontinue the services any time without significant sunk cost. While initial adoption decision is very important in the case of all IT products, post-adoption behaviour assumes greater importance for subscription-based IT services. Innovation diffusion literature (Rogers 1995) suggests that consumers' discontinuance can be classify into replacement discontinuance (i.e., replacing an adopted innovation with another innovation perceived as being better) and disenchantment (i.e., terminating the use of an innovation because of dissatisfaction). In the case of mobile services, however, the behavioural consequences of both types of discontinuance are most likely to be the same -- switching to a new mobile service provider. This is because disenchanted users usually turn to a competitor's services instead of terminating the use mobile service. This compels the importance of studying service provider switching in the context of mobile services, which makes this study different from existing studies on adoption or usage continuance.

Gateway Technology and Number Portability

Standardization activities for mobile services can be categorized into three generations. Despite the complicated technology composition of the market, the availability of gateway technology sheltered the compatibility and network interconnection issue to the end users. An individual subscribing to any one of the technology standards from any one service provider can easily make calls to, and receive calls from, subscribers to any one of the other standards and any one of the other service providers. This also makes the switching of mobile services provider different from the switching of other IT products (e.g., your switch from PC to Mac may affect the file-sharing compatibility between you and others who use PC) and compels the importance of studying switching

Market Structure

Compared with the majority of IT products and services, mobile services are extremely competitive, deregulated, and have low entry barriers. Individuals face numerous and highly substitutable service providers choices. Unlike the markets for other IT products (e.g., operating system) where one or two firms dominate, the mobile services market has over 600 firms worldwide. More than one third of the countries in the world have three or more competing mobile operators each (ITU 2002). That the availability of alternative providers is much greater on mobile services market than on any other IT products markets posits the relevance of studying switching in the context of mobile services.

THEORETICAL DEVELOPMENT

In this study, a mobile service provider switching model linking factors related to service provider efforts with market environment is developed. The model based on the Theory of Planned Behaviour (TPB) attempts to associate the characteristics of mobile services industry with beliefs relevant to and intentions towards service provider switching. Figure 2 depicts the research model.

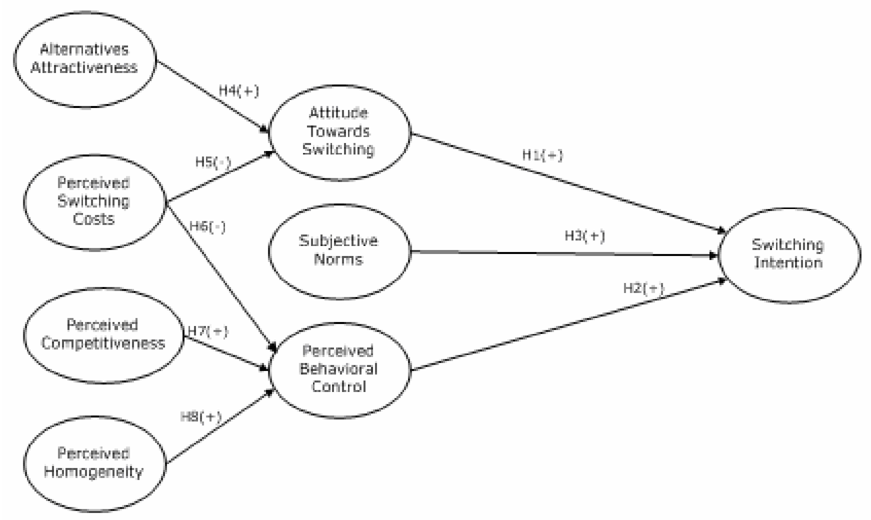


Figure 1. Research Model

Theory of Planned Behaviour

The Theory of Planned Behaviour (TPB) has been utilized in various domains and significant supportive results were found (Armitage and Conner 2001). Moreover, TPB has been widely use in IS research for innovation adoption and continuance in prior studies (Karahanna and Straub 1999, Taylor and Todd 1995, Venkatesh and Morris 2000). Therefore, using TPB in this study provides the basic for comparing prior innovation adoption researches with the current work. TPB identifies three independent determinants of intention (Ajzen 1991). The first of these is the attitude toward the behaviour and refers to the degree to which a person has a favourable or unfavourable evaluation or appraisal of the behaviour studied. The second is a social factor called subjective norm. It refers to the perceived social pressure to perform or not to perform the behaviour. The third determinant of intention is the degree of perceived behavioural control, which refers to the presence or absence of requisite resources and opportunities reflected by past experience as well as anticipated barriers. Generally, the more favourable the attitude and subjective norm with respect to a behaviour and the greater the perceived behavioural control, the stronger an individual's intention to perform the behaviour under consideration. The relative importance of attitude, subjective norm, and perceived behavioural control in predicting intention is expected to vary across behaviours and situations (Ajzen 1991). The attitude and beliefs about adopting an IS are different from the attitude and beliefs about continuing to use an IS. (Karahanna, et al. 1999) found that adoption intention is solely determined by normative consideration from the social environment, while continuance decision are determined by attitude and perceived behavioural control (Venkatesh and Morris 2000). In line with existing literature, we agree that attitude and perceived behavioural control would be the determinants of behavioural intention for switching mobile services providers. However, we suggest a significant effect of subjective norms on behavioural intention in the context of mobile service provider switching, because of the network externality of using the same mobile service provider with other users. In other words, adopters could benefit more by using the same service provider which their important others are using. Therefore, we hypothesize that

H1: A more positive attitude towards switching mobile service providers will lead to greater intention to switch mobile service providers.

H2: A higher level of perceived behavioural control will lead to greater intention to switch mobile service providers.

H3: A higher level of subjective norms supportive of switching mobile service providers will lead to greater intention to switch mobile service providers.

Social exchange theory

Thibaut's and Kelly's Social exchange theory (Thibaut and Kelley 1996) provides well-developed insights into how a relationship is maintained and terminated. This theory first proposed that there is a difference between a person's attitude with a relationship (know as attraction or affect) and a person's commitment to the relationship (the tendency to maintain the relations). In short, it is possible for an individual to have positive attitude but not committed, or vice versa. Social exchange theory is based on two important concepts, which are the standards for evaluation the relationship -- the comparison level (CL) -- and the comparison level of alternative relationship (CL-alt). Comparison level is defined as the standard against which the participant evaluates the attractiveness of a relationship or how satisfactory a current relationship is. For example, if a person's mobile service experience is above his/her CL, he/she will be satisfied; if it is below the CL, he/she will be dissatisfied. Thus, a person's satisfaction level is related to the extent that the service experience is above the CL. By contrast, the comparison level of alternative (CL-alt) is defined as the lowest level of outcomes a person will accept in the light of available, and most attractive, alternative opportunities in other relationships. If the outcome from a relationship drops below the CL-alt, the person will leave the current relationship. As a result, as the outcome of a current relationship exceeds the CL-alt by a large margin, the person should be more dependent on this relationship. It follows that consumers would have less favourable attitude towards switching if they perceived potential alternatives to be no better than their current service provider(s). In marketing literature, the quality of the outcome available from the best alternative supplier is one of the key factors in maintaining or breaking a buyer-seller relationship. The availability of high quality alternative suppliers is found to weaken the dependence on the current supplier (Anderson and Narus 1990).

In this study, perceived alternative attractiveness refers individuals' perception of the relative superiority of the potential alternative mobile service provider(s) over the current one. The concept is similar to the notion of relative advantage in innovation literature (Rogers 1995), which refers to the degree to which an innovation provider benefits from being superior to its precursor. According to the expectancy-value model of attitude (Fishbein and Ajzen 1975), attitude is developed from beliefs people hold about the object of the attitudes. In case of attitudes toward a behaviour, each belief link the behaviour to a certain outcome, or to some other attribute such as the cost incurred from performing the behaviour. Perceived alternative attractiveness can be incorporated into the TPB framework by predicting the attitudes towards switching. It is posited that the more attractive of the alternative service providers, the more positive the attitude towards switching, as a greater benefit will be obtained for switching. Consequently, the following hypothesis is proposed:

> *H4: A higher level of the perceived attractiveness of alternative mobile service providers will lead to a more positive attitude towards switching.*

Perceived Switching Cost

Perceived switching cost is defined as individual's perception of one-time cost associated with the process of switching from one mobile service provider to another. (Porter 1980) suggests that switching cost is a "one-time" cost (p.10), as opposed to the on-going cost associated with using a product or provider once a repeat purchase relationship is established. Empirical findings from IS research suggest that, customers will persist with one IT service provider because the cost of switching to another is too high (Chen and Hitt 2002b, Reichheld and Schefter 2000). In the context of mobile services, switching cost includes the cost and time for terminating the subscription from the current service provider, the cost and time for registering with a new service provider, the lost of benefits accumulated in the current subscription, and so on. Since switching cost is directly related to the expected outcome of switching mobile service provider, the perceived switching cost can be included in the TPB framework as an antecedent of attitude. As the perceived switching cost increases, the individual's attitude toward switching will become less positive. This leads to the following hypothesis:

> *H5: A higher level of perceived switching cost will lead to a more negative attitude towards switching mobile service providers.*

According to the TPB, among the beliefs that ultimately determine intention, control beliefs are related to the presence or absence of requisite resources and opportunities. These beliefs exert influences through factors that affect the perceived difficulty of performing an act. Ajzen (1991) refers to both the internal notion of individual "self-efficacy" (Compeau, et al. 1999) and the external resource constraints. The internal "self-efficacy" control belief is related to one's perceived ability and knowledge for performing the act (Ajzen 1991, Taylor and Todd 1995, Taylor 1995). With respect to mobile service provider switching, user-perceived switching cost reflect their self-efficacy on handling the switching process. In line with this, we suggest that perceived switching cost is the factor that affects the perceived difficulty of mobile service provider switching.

> *H6: A higher level of perceived switching cost will lead to lower level of behavioural control for switching mobile service provider.*

Market Competition and Product Homogeneity

With respect to IT usage, the facilitating conditions construct consists of a two-dimensional control belief: one relating to resources availability and the other relating to technology compatibility issues that may restrict usage (Taylor and Todd 1995). These two dimensions are the objective factors in the environment affecting the ease of performing an act (Thompson, et al. 1991). The more resources and opportunities individuals believe they possess, and the fewer obstacles or impediment they anticipate, the greater their perceived control over the behaviour should be (Venkatesh 2000). A highly competitive market provides more options for users. And that facilitates service provider switching.

> *H7: A higher level of perceived market competitiveness will lead to a lower level of perceived behavioural control for switching mobile service provider.*

Perceived homogeneity is defined as the extent to which the providers in a market are seen as the same or as substitutable. Similar to the case of IT usage, individuals would be constrained by the compatibility of the services provided by different service providers in their switching behaviour and would perceive higher behavioural control (Taylor and Todd 1995). Therefore, perceived homogeneity is incorporated as an antecedent of the control beliefs in the TPB framework. Higher homogeneity of the mobile service providers means that users can switch among any service providers without worrying about the compatibility of the new mobile services with their existing services or mobile devices.

> *H8: A perception of mobile service providers as being more homogeneous will lead to a higher level of behavioural control for switching mobile service providers.*

RESEARCH METHODOLOGY

Eight constructs were measured in this study, as described below. Scales for all constructs were adopted from prior research and measured in Likert scale (1~7). The measurement of the concepts under study involved the adaptation of existing scales to the current research setting. To measure perceived behavioural control, items reflecting an individual's belief regarding access to resources and opportunities for engaging in behaviour of switching were required. Consistent with that notion, the scale for perceived behavioural control was developed by adapting items from the perceived behavioural control scale used by Taylor (1995) .The scale for attitude toward switching was adapted from the scale for measuring attitude toward act used by Ajzen (1991). The scale for subjective norms was adapted from the Taylor (1995) scale of the same name. For measuring switching intentions, items were derived from the scale of behavioural intention (Oliver and Swan 1989). Measurement for perceived switching cost and perceived alternative attractiveness is adopted from Jones . Perceived homogeneity is measured by two items adopted from Burnham (2003) which are modified reflect the substitutability of different mobile services providers. Perceived competitiveness contains two items developed by the researcher to reflect the degree of the availability of choice and options for switching.

Sample

Empirical data for testing the hypotheses were collected via an online questionnaire. Email invitations to participate in the online survey were sent to members of an official governmental portal in Hong Kong. The portal was selected because it was the exclusive portal for Government-to-Citizen activities/transactions, providing a reliable source of subjects, which overcome some of the limitations for online survey, such as multiple responses from one individual. As Pitkow and Recker (1995) points out, online survey provides several advantages including the use of adaptive questionnaires, reliability, and cost saving. The survey was conducted over a period of one month in later 2003. A lucky draw was arranged to provide an incentive to participate. A pilot test of the online questionnaire was conducted among staff and students of a university in Hong Kong before the formal survey was launched. Feedback from the pilot test was taken into consideration when finalizing the online questionnaire. All together, 4,159 valid responses were collected; and all respondents are mobile phone users. Over 80% of the respondents had experience in switching mobile service provider.

Results

A confirmatory factor analysis using Lisrel 8 was conducted to test the measurement model. The fit of the overall measurement model was estimated by various indices provided by Lisrel (see Table 1). The goodness-of-fit (GFI) and adjusted goodness-of-fit (AGFI) were 0.96 and 0.94, respectively. The normalized fit index (NFI), non-normalized fit index (NNFI) and comparative fit index (CFI) were also used. Values typically range from 0 to 1, with values greater than 0.9 representing reasonable model fit. For the measurement model, we observed values of 0.97, 0.97, and 0.98 for NFI, NNFI, and CFI respectively, all indicating good model fit. Finally, root mean square residual (RMSR) provides an indication of the proportion of the variance not explained by the model, whereas root mean square error of approximation (RMSEA) describes the discrepancy between the proposed model and the population covariance matrix. Observed values were 0.041 and 0.046, which were within the recommend cut-off values of 0.10 for RMSR and 0.08 fro RMSEA for goodness of fit. Then, we proceed to evaluate the psychometric properties of instrument in terms of reliability, convergent validity, and discriminant validity.

Fit Indices	Recommended Value	Measurement Model	Structural Model
X2	N/A	2187.61	2583.41
Degree of freedom	N/A	224	234
Goodness-of-fit (GFI)	≥ 0.90	0.96	0.95
Adjusted goodness-of-fit (AGFI)	≥ 0.80	0.94	0.94
Normalized fit index (NFI)	≥ 0.90	0.97	0.97
Non-normalized fit index (NNFI)	≥ 0.90	0.97	0.96
Comparative fit index (CFI)	≥ 0.90	0.98	0.97
Root mean square residual (RMSR)	≤ 0.10	0.041	0.095
Root mean square error of approximation (RMSEA)	≤ 0.08	0.046	0.049

Table 1. Fit Indices

Item	Mean	Standard Deviation	Standardized Loading	Item Reliability	Cronbach's Alpha	Average Variance Extracted
INTEN1	4.25	1.718	0.90	0.82	0.94	0.80
INTEN2	4.22	1.641	0.97	0.94		
INTEN3	4.11	1.525	0.89	0.80		
ATT1	4.32	1.317	0.88	0.77	0.92	0.79
ATT2	4.41	1.295	0.91	0.83		
ATT3	4.26	1.284	0.89	0.79		
PBC1	6.48	1.121	0.70	0.50	0.73	0.59
PBC2	5.72	1.308	0.83	0.68		
SN1	2.94	1.655	0.87	0.76	0.94	0.79
SN2	2.91	1.572	0.87	0.76		
SN3	2.79	1.552	0.90	0.80		
SN4	2.84	1.630	0.92	0.84		
COMP1	5.91	1.248	0.77	0.59	0.71	0.56
COMP2	5.75	1.307	0.73	0.53		
ALTER1	3.68	1.347	0.78	0.61	0.93	0.78
ALTER2	3.64	1.349	0.92	0.85		
ALTER3	3.63	1.323	0.95	0.89		
ALTER4	3.54	1.365	0.89	0.78		
COST1	3.38	1.600	0.79	0.62		
COST2	3.22	1.559	0.87	0.76		
COST3	2.91	1.414	0.89	0.80		

COST4	3.02	1.477	0.87	0.75		
HOMO1	4.55	1.415	0.78	0.61	0.73	0.58
HOMO2	4.41	1.496	0.74	0.55		

Table 2. Reliability and Convergent Validity

The psychometric properties of the constructs and items are summarized in Table 2. Covariance matrixes are listed in Table 3. Cronbach's alphas for all constructs were over 0.70 threshold for field research and most of the scales showed good reliability with alphas greater than 0.70 (Nunnally and Bernstein 1994). Convergent validity can also be accessed by factor loadings and squared multiple correlations from the confirmatory factor analysis (see Table 3). Hair et al. (1998) recommend that factor loading greater than 0.50 be considered highly significant, while Fornell (1981) proposes a more stringent criterion whereby loading is considered significant only if it is greater than 0.70. All factor loadings of items in the current research model were above 0.70, with each item loaded significantly ($p < 0.01$) in all cases, on its underlying constructs. Therefore all constructs in the model have adequate reliability and convergent validity.

Discriminant validity is the degree to which items differentiate between constructs. We compared the shared variances between the constructs with the average variance extracted from individual constructs according to Fornell & Larcker (1981). The results show the shared variance between constructs, confirming the discriminant validity (see Table 3). To conclude, the measurement model and all items of the instrument have demonstrated adequate reliability, convergent validity, and discriminant validity.

	1	2	3	4	5	6	7	8
1. Switching intention	0.80							
2. Attitude	0.45	0.79						
3. Perceived behavioral control	0.14	0.04	0.59					
4. Subjective norm	0.22	0.16	0.06	0.79				
5. Alternative attractiveness	0.24	0.42	0.02	0.45	0.56			
6. Perceived switching cost	0.08	0.17	0.25	0.16	0.08	0.78		
7. Perceived homogeneity	0.01	0.01	0.11	0.14	0.09	0.14	0.73	
8. Perceived competitiveness	0.06	0.04	0.38	0.06	0.01	0.17	0.00	0.58

Notes: Diagonals represent the average variance extracted. Other entries represent the shared variance.

Structure Model

Table 3. Discriminant Validity

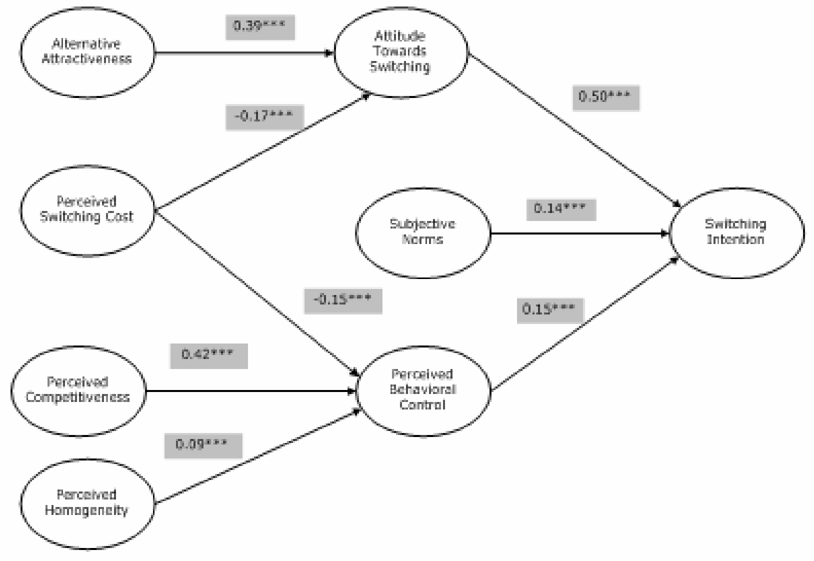


Figure 2. Path Coefficients

A similar set of fit indices was used to examine the structural model (see Table 1). When all the fit indices were compared with their corresponding recommended values, the structural model showed a good fit (GFI =0.95, AGFI=0.94, NFI=0.97, NNFI=0.96, CFI=0.97, RMSR=0.095, RMSEA=0.049). We then examined the estimation results of the structural model over the entire sample. Properties of casual paths, including path coefficients, t-values, and variances explained for each equation in the structural model are presented in Figure 2, which also shows that all path coefficients in the model were significant with a p-value of <0.001.

DISCUSSIONS AND IMPLICATIONS

That all the hypotheses were supported indicates the usefulness of the proposed model for analysing mobile service provider switching decision, which is a special type of post-adoption discontinuance. It is interesting to note that attitude toward switching is by far the strongest predictor of mobile service provider switching. The importance of the beliefs prediction intention can be expected to vary across situation, so it is useful to determine which factors are the most important in the specific case of mobile service provider switching. The impacts of perceived behavioural and subjective norm on switching intention are similar as shown in the model. Unlike pervious research on innovation continuance or usage (Karahanna, et al. 1999), which found that the impact of subjective norm on intention, an important determinant for adoption intention, had no impact on post-adoption usage, the results of the current study suggest that subjective norm has a significant impact on the post-adoption discontinuance decision in the case of mobile service provider switching and therefore indicate that individual decision on post-adoption discontinuance of mobile service provider switching is different from post-adoption continuance or usage of other IT products. A possible explanation of this is the existence of network externality of using the same service provider. This calls for further research in the innovation post-adoption behavioural, especially, for innovation at the later stage in the innovation diffusion cycle.

The results also provide implications for practitioners. The results validate the importance of relative service attractiveness on the attitudes of subscribers' switching intention. The relative importance of perceived alternative attractiveness and perceived switching cost on determining switching attitudes suggest that subscribers pay more emphasis on forthcoming benefits that they will received from the potential “switch to” service provider than the cost from the “switch from” service provider. This paper has several limitations that should be taken into consideration when interpreting the results. First, using online survey in the data collection may bias sample towards Internet users, there may be subjects who are mobile users but do not user Internet services. However, given the large sample size and high penetration rate of mobile and internet users in Hong Kong, this sample bias is keep at an insignificant level.

CONCLUSION

This study builds on pervious innovation adoption research to develop a model of mobile service provider switching. The proposed model relates the characteristics of the mobile service industry for identifying the beliefs that determine the main TPB constructs. On the theoretical side, the current work proposed and validated a theoretical model for studying the switching intention, a post-adoption decision, pushing the innovation adoption research to the later stage of the diffusion cycle. On the managerial side, the current work provides insights to service providers and corporate IT managers on formulating retention strategy.

REFERENCES

Ajzen, Icek, "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes*, 50, (1991), 179-211.

Anderson, Eugene W., Claes Fornelland Donald R. Lehmann, "Customer Satisfaction, Market Share, and Profitability: Findings From Sweden," *Journal of Marketing*, 58, 3, (1994), 53-66.

Anderson, James C. and James A. Narus, "A model of distributor firm and manufacturer firm and manufacturer firm working partnerships," *Journal of Marketing*, 54, 1, (1990), 42.

Armitage, Christopher J. and Mark Conner, "Efficacy of the Theory of Planned Behaviour: A meta-analytic review.," 40, 4, (2001), 471.

Bhattacherjee, Anol, "Understanding Information System Continuance: An Expectation-Confirmation Model," *MIS Quaterly*, 25, 3, (2001), 351-370.

Burnham, Thomas A., Judy K. Frelsand Vijay Mahajan, "Consumer Switching Costs: A Typology, Antecedents, and Consequences," *Journal of the Academy of Marketing Science*, 31, 2, (2003), 109-126.

Chen, Pei-Yu (Sharon) and Lorin M. Hitt, "Measuring Switching Cost and the Determinants of Customer Retention in Internet-Enabled businesses: A Study of Online Brokergae Industry," *Information System Research*, 13, 3, (2002a), 255-274.

Chen, Pei-Yu (Sharon) and Lorin M. Hitt, "Measuring Switching Costs and the Determinants of Customer Retention in Internet-Enabled Businesses: A Study of the Online Brokerage Industry.," 13, 3, (2002b), 255.

Compeau, D., C. A. Higginsand S. Huff, "Social cognitive theory and individual reactions to computing technology: A longitudinal study," *MIS Quarterly*, 23, 2, (1999), 145-158.

Devaraj, S., M. Fanand R. Kohli, "Antecedents of B2C channel satisfaction and preference: Validating e-commerce metrics," *Information Systems Research*, 13, 3, (2002), 316-333.

Fishbein, M. and I Ajzen, *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*, Addison-Wesley Publishing Company, Reading, MA, 1975.

Fornell, Claes and David F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18, 1, (1981), 39-50.

Hair, Joseph F., Ronald L. Tatham, Rolph E. Andersonand William Black, *Multivariate Data Analysis*, Prentice Hall, Upper Saddle River, N.J, 1998.

ITU, "ITU Internet Reports 2002: Internet for a Mobile Generation," *International Telecommunication Union*, (2002).

Jones, Michael A., David L. Mothersbaughand Sharon E. Beatty, "Why Customers Stay: Measuring the Underlying Dimensions of Services Switching Costs and Managing their Differential Strategic Outcomes," *Journal of Business Research*, 55, 6, (2002), 441-450.

Karahanna, Elena and Detmar W. Straub, "Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post Adoption Beliefs," *MIS Quarterly*, 23, 2, (1999), 183-213.

Karahanna, Elena, Detmar W. Strauband Norman L. Chervany, "Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs," *MIS Quarterly*, 23, 2, (1999), 183-213.

Lee, Sukekyu, Fred Zufrydenand Xavier DrÃ¨ze, "A Study of Consumer Switching Behavior Across Internet Portal Web Sites.," *International Journal of Electronic Commerce*, 7, 3, (2003), 39.

Meisner, Jeff, "Telecom companies expect bumpy '04 ride," *The Business Journal*, 20, 48, (2004), 36.

Nunnally, Jum C. and Ira H. Bernstein, *Psychometric theory*, McGraw-Hill, New York, 1994.

Oliver, Richard L., *Satisfaction: A Behavioral Perspective on the Consumer*, McGraw-Hill, New York, 1997.

Oliver, Richard L. and John E. Swan, "Equity and Disconfirmation Perceptions as Influences on Merchant and Product Satisfaction," *Journal of Consumer Research*, 16, 3, (1989), 372-383.

Parthasarathy, Madhavan and Anol Bhattacherjee, "Understanding Post-Adoption Behavior in the Context of Online Services," *Information System Research*, 9, 4, (1998), 362-379.

Pitkow, James E. and Margaret M. Recker, "Using the Web as a Survey Tool: Results from the Second WWW User Survey," *Journal of Computer Networks and ISDN systems*, 27, 6, (1995), 809-822.

Porter, Micheal E., *Competitive Advantage*, Free Press, New York, 1980.

Rai, A., S. S. Langand R. B. Welker, "Assessing the validity of IS success models: An empirical test and theoretical analysis," *Information Systems Research*, 13, 1, (2002), 50-69.

Reichheld, Frederick F. and Phil Schefter, "E-Loyalty.," 78, 4, (2000), 105.

Rogers, Everett M., *Diffusion of Innovations*, The Free Press, New York, 1995.

Taylor, S. and P. A. Todd, "Understanding Information Technology Usage - a Test of Competing Models," *Information Systems Research*, 6, 2, (1995), 144-176.

Taylor, Shirley, "Assessing IT Usage: The Role of Prior Experience," *MIS Quarterly*, 19, 4, (1995), 561-570.

Thibaut, John W. and Harold H. Kelley, *The Social Psychology of Groups*, Wiley, New York, 1996.

Thompson, R. L., C. A. Higginsand J. M. Howell, "Personal Computing: Toward a Conceptual Model of Utilization," *MIS Quarterly*, 15, 1 (March), (1991), 125-142.

Van Dyke, Thomas P., Leon A. Kappelmanand Victor R. Prybutok, "Measuring Information Systems Service Quality: Concerns on the Use of the SERVQUAL Questionnaire," *MIS Quarterly*, 21, 2 (June), (1997), 195-208.

Venkatesh, V. and M. G. Morris, "Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior," *MIS Quarterly*, 24, 1, (2000), 115-139.

Venkatesh, Viswanatch, Micheal G. Morris, Gordon B. Davisand Fred D. Davis, "User Acceptance of Information Technology: Toward a Unified View," *MIS Quaterly*, 27, 3, (2003), 425-478.

Venkatesh, Viswanath, "Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model.," 11, 4, (2000), 342.

Information Management in the Mobile Hospital Work Domain

Dr Julie Fisher
Monash University
Melbourne, Australia
Email: Julie.fisher@infotech.monash.edu.au

Dr Linda Dawson
Monash University
Email: Linda.Dawson@infotech.monash.edu.au

Dr Liza Hesslop
Monash University
Email: liza.hesslop@med.monash.edu.au

Mr Stephen Weeding
Monash University
Email: stephen.weeding@med.monash.edu.au

Abstract

Mobile computing can potentially change the way medical, health and information services are delivered. Much of the focus has been on mobile telephone use particularly from a commercial and marketing perspective. It is often assumed that 'mobile' means using a mobile telephone. The growth in mobile devices usage in hospitals, is contributing to better health and medical services delivery. This paper describes two hospital based case studies involving different mobile devices. We argue that understanding the mobile domain in hospitals requires an understanding of both the context and the user, we propose a definition of the mobile hospital work domain.

Keywords:

Mobile, mobile devices, health information, health, hospitals

INTRODUCTION

When the term 'mobile' is used, particularly in an Australian context, it is often assumed to refer to a mobile telephone. Much of the research reported to date on mobile devices focuses on the use and applications for mobile phones (Carroll et al, 2003; Carlsson et al, 2004). It is not surprising that much of the research has focused on mobile telephones. The mobile phone is small, portable and wireless. In many contexts, mobile devices that are used extend well beyond the mobile phone. In a hospital setting, for example, numerous mobile devices can be found (Heslop et al, 2003). In this context, connectivity of devices is of critical concern rather then the size of the device (Fernando, 2003).

There is yet to be articulated a definition of 'mobile' environments in the hospital context that is clear or accepted. This lack of a contextual definition makes it difficult to base mobile-based research on a solid foundation. This paper explores, from the perspective of the current literature and research, the range of devices, users and processes that can be defined as mobile. We present two case studies involving the use and implementation of mobile devices in a hospital environment. Finally, the paper proposes a definition of a 'mobile' environment in a hospital context and presents a model of a mobile hospital work domain.

Analysis of the literature 2005

Over the last five years there has been a significant increase in the number of papers published reporting research in the area of mobile devices. Here, we examine the papers published in 2005 from a range of journals and conferences which include: ECIS, Bled, ICMB, MISQ, JAIS, CAIS. The focus was on journals and conferences in the information systems field as the researchers were interested in the use of mobile devices in practice rather than discussions and research on technical issues. Specifically we looked for papers describing case studies involving the implementation and use of mobile devices in practice or the work environment. Hence papers with a theoretical approach and papers on device applications have not been included. Table 1 presents a list of papers found. A brief description of the mobile device is given and a definition provided.

Area of application	Device/s	Definition /Authors
International Conference on Mobile Business Sydney 2005		
Mobile workforce - Field Force or mobile employees	Laptop touch screen Smart telephone	"Wireless FFA is usually used to describe mobile employees, remote from their base of operations, utilising wireless technologies to perform their specific business tasks." (Innes et al. 2005)(49)
Use in a restaurant	PDA	"..Mobile electronic commerce can occur anywhere and anytime via handheld devices such as cellular phones and personal digital assistants." (Prasad et al. 2005) (69)
Railway workers workforce deployed to facilitate more efficient communication between railway staff.	Specifically designed PDAs and SMS	(Wang et al. 2005) No definition provided.
Mobile monitoring system for people with Parkinsons disease	Patient monitoring mobile unit included camera, PC, fax, speakers	The authors describe this as " Mobile monitoring" or a "portable or mobile medical monitoring system".(Biemer and Hampe 2005) (464)
Use of wireless solutions in hospitals	Variety of devices deployed including hand held devices, web phone, laptop computers.	"Wireless information technologies is used as an umbrella term that includes mobile and handheld computers, personal digital assistant, mobile and wireless communications. They offer greater flexibility and mobility" (Gururajan and Murugesan 2005) (473)
European Conference on Information Systems 2005		
IT consultants use for work	Customised device tablet pc/pda, mobile phone	"Mobile working solutions typically require a combination of IT networks and devices, such as WLANs, Internet enabled mobile phones, and portable computers." (Breu et al. 2005)
Tourism – implementation of a hotel guide for consumers	Wap enabled phone, pda and IMode	(Carlsson et al. 2005) No definition provided

Table 1 Research reported in the literature on mobile device use

This list is not intended to be a comprehensive analysis of papers published in the area but provides an analysis illustrating that what is defined as a mobile device can be very broad, from a mobile phone to a full-sized computer based in a patient's home. What is consistent across these papers and reported research is that critical to the concept of 'mobile', is wireless capability. This distinguishes these devices from standard personal computers or laptops which do not have a wireless component. It is significant that in the forum we chose for this analysis (a range of journals and conferences from 2005: ECIS, Bled, ICMB, MISQ, JAIS, CAIS) that for CAIS, MISQ, JAIS and Bled there were no publications which focussed on the use of mobile technology in hospital settings based on a definition of 'mobile' in that context.

USE OF MOBILE AND WIRELESS TECHNOLOGIES IN HOSPITAL SETTINGS

Poor access to health and patient information such as diagnostic results can delay clinical decision making. Frequently delays are compounded by clinicians' poor access to computers and/or their lack of confidence in the ability of information technology (IT) to positively affect clinical outcomes. The deployment of mobile devices in hospitals goes some way to addressing these issues. Further, wireless networks, hand held and portable devices and associated applications are technologies that are ideal for nomadic workers such as clinicians in ward settings. They can potentially enhance clinicians' use of patient management and clinical systems by providing decision support and clinical information exchange at the bedside or point of care (McLeod et al. 2003; Al-Ubaydli 2004; McAlearney et al. 2004). Such technologies however, need to be critically assessed in a hospital environment for their wider potential and application for the delivery of health information at the point of care.

Increasingly handheld devices, particularly PDAs are used in hospitals and health care (Gururajan and Murugesan 2005). PDAs are by no means the only form of mobile device used in health care. Other devices, also defined as 'mobile' are becoming more prevalent such as tablet PCs, and laptops. In one case reported in the literature the mobile system included an entire monitoring system with a standard personal computer and video camera mounted on a tripod located in the patient's home (Biemer and Hampe 2005).

Gururajan and Murugesan (2005) suggest that PDAs or digital tablets that are wireless are the preferred device choices in hospitals. Their study concluded that "while these developments achieved their software development objectives, failed to address the acute issues of healthcare management in Australia." (477). Another study by McAlearney et al. (2004) investigated the use of PDAs by doctors in the United States. The study found that the doctors expect such devices to become more useful though issues still remain that relate to reliability and security. Benefits perceived by the doctors included enhanced productivity and quality of patient care and services and improved access to information. Use and usefulness of the devices related strongly to the devices being able to connect to wireless services thereby providing access to patient and other data. The complexity of health care organisations and in particular hospitals requires organisational changes for the implementation of such technologies to realise real benefits (Istepanian et al. 2004).

The entering of data and data integrity is another concern. The use of mobile devices by volunteers in a clinical trial reported by Koop and Mosges (2002), found that the users had problems with data entry that resulted in missing values. The problem was further exacerbated when the system did not notify the users of the missing values. Also, the handheld devices had a high error rate because the application was not able to handle the speed with which users were entering data. A range of technology issues are also impacting on uptake. Screen size and the limitations this places on information presentation is a problem (Gururajan and Murugesan 2005). Other issues include data transmission problems, a lack of integration with other hospital systems and concerns relating to security (Gururajan and Murugesan 2005; Istepanian et al. 2004).

With such a rise in usage and the use of such a variety of technology it is important to understand the context and the issues around the implementation of mobile technology as it is with any technology. There are however few reported case studies describing the deployment and integration into hospital settings. The following two case studies describe how mobile devices were implemented in hospitals, how they are used particularly in terms of information management, how the decisions were made and discusses the issues implementation raised.

RESEARCH

The aim of this research project is to identify the key issues in the implementation and use of mobile devices in a hospital setting. The specific research questions relevant to this paper are:

- What do clinicians regard as 'mobile' and how does this help us understand a mobile hospital environment?
- How are mobile devices used in hospital settings and what determines device usefulness and appropriateness in this environment?

This paper presents two case studies which describe the use of different mobile devices in two different hospital settings.

Data collection

The first case study involved anaesthetists using a PDA application. Data for this project was collected by face-to-face and email interviews. Interview questions addressed issues of purpose, usability, implementation and use. Interviews were conducted with the key players involved which included the developer of the application, senior hospital anaesthetist who was championing the use of the application and three anaesthetists. E-mail was used for follow-up questions.

The second case study involves the implementation of three different mobile devices in a hospital acute ward. Interviews and observations on the ward have been conducted with key personnel regarding the current clinical ICT environment and the wireless environment. In addition observations and notes taken at project planning meetings have also contributed to our data. These interviews and observations have provided the preliminary findings discussed in this paper. Key personnel include the Nurse Unit Manager (NUM) of the ward and the Stroke Liaison Manager (SLM).

RESULTS

The next section describes the settings in which the mobile devices were implemented, the purpose and choice of device, value to the users, use and problems encountered.

Case study 1: Implementation and use of PDA based application for anaesthetists

The application is a simple logging application implemented on a Palm PDA based on CUSUM Charts (Shewhart 1931; Page 1954) which allows anaesthetists to monitor their performance of various procedures. A report of another project where this application was implemented has been previously published by others (Bent et al, 2002).

The developers of the PDA based application work for a company experienced in the development of technology products for the health-care industry. The application has been recently deployed as a pilot project with six registrars in three hospitals in Australia and New Zealand. The application is designed to record anaesthetic procedures and incidents occurring during the procedure. Most of the data is recorded by selecting options from drop-down menus or check boxes, however, users may also record information using the free text facility available. The application is very quick and easy-to-use, one of the anaesthetists demonstrated recording a recently completed procedure in under 40 seconds.

- Purpose

A prime reason for this project in the eyes of the Senior Anaesthetist was the long term high level advantages for using IT and mobile solutions for hospital data integration and data management. *"We have simplified greatly the amount of data we actually collect in our audit [a related IT application] because you just don't collect information for information's sake and the more information you want to collect the less the quality of it is that you will get. Our goal really is automated anaesthesia recordkeeping whereby it collects the lot, single point and you can do all your pre-anaesthesia assessment and all your postoperative assessment and all that."* The PDA based application used by the anaesthetists recorded at the conclusion of a procedure the outcome of that procedure. The PDA is later synchronised with a desktop PC. The purpose is to track the anaesthetist's skills and monitor whether additional training is required.

- Choice of mobile device

A non-wireless PDA was the only mobile device used in this environment. The Senior Anaesthetist had been using a handheld device for many years and was therefore enthusiastic about its usefulness in hospital particularly as it offered users a range of functionality and was small and as he said "*fitted in a shirt pocket*". He had championed the choice of using a PDA and assisted with the design of the application and implementation. The Senior Anaesthetist's choice of mobile device was also influenced by what was happening in other hospitals and because it was supported by the College of Anaesthetists who had funded a pilot project.

Commenting on the choice of technology The Senior Anaesthetist said: "*We have had implementations of systems where people have put in desktop computers in the operating room you couldn't find a more hostile environment for a desktop computer, or somewhere where you have a premium on real estate*." His key point was that decisions had been made in the past by people who really did not understand the environment in which they were working, that is an operating theatre. Employing a device that did not take up space in an operating theatre and did not require the anaesthetist to look away from their work was critical.

- Use and value to users

For the Senior Anaesthetist the principal reason for use was to improve the skills of the anaesthetists "*This thing is good because I think what I would like to do with our registrars here is part of their training is so they do leave here and see as a valuable clinical professional activity to do is to monitor performance throughout their life because not only is good for acquisition of skills but deterioration of skills and performance be it due to a physical illness that sort of thing.*" In his mind the Palm pilot and application was a good way to achieve this. The Senior Anaesthetist, because he had been using a mobile device for some time saw other possibilities for use. He explained "*we have tried to put value in them [PDAs] like putting anaesthesia database in and some sort of books and text which are really good decision support texts or 'aide memoirs' things like that and departmental policies. So you have got to do that*." It is clear that he recognised that the additional functionality such as ability to access information in medical databases was important to encourage extensive use of the device.

The users generally were enthusiastic about using a PDA, there was no reluctance on their part. One anaesthetist commented: "*For myself I use it as logbook so on top of all the monitoring and data other people are using it for I am using it myself as a database of the cases I have done*." However there was limited use among the three anaesthetists interviewed who were trialling the system and the device. Generally they did not find the application or the device to be of great value in the hospital environment. A comment from a female anaesthetist was that she was not using it because she had her own logbook in which procedures were recorded and she was keeping data on her own computer at home and did not really want to enter data for a third time. However there were some downloadable texts she was using and found that useful to have on the PDA. Two of the anaesthetists did not carry their PDA with them but left them in their lockers. The critical issue was not the usability of the device or application but the lack of additional functionality as this comment illustrates "*I am waiting for the technology to improve to the point where you have a Palm, somewhere where someone can send me my theatre lists for the next day and I can just look at it and say that is what I am supposed to be doing instead of go hunting around the department trying to look for the most up-to-date theatre list.*" Another responded to this comment "*If we had that one thing we would all be impossible to separate from our Palm's*." There was unanimous agreement on this.

- Problems

Implementation issues were one of the more serious problems that the hospital faced. The Senior Anaesthetist explained that there was an initial enthusiasm but then there were delays in the delivery of the device is. He said "*I think people's initial enthusiasm was high, yes it's coming yes its coming. Then it was just a 'oh well took so long to get here I'm not going to be interested'. I think it didn't deliver what they were expecting, some people were very good users and took to it and got a lot out of it. Some people didn't bother. I don't know why that is I think some people are probably just Luddites or something like that.*" In his mind the two key problems of why uptake has been slow were delays in issuing the devices and reluctance on the part of the users. This however was not reflected in the comments from the users.

The critical issue for the users was functionality particularly as at this point the device had no wireless functionality. The anaesthetists agreed that the size was appropriate but they would not be prepared to carry anything larger. Whilst they could appreciate the potential usefulness of the device at this stage the functionality was not there. With all the other equipment they were expected to carry and given its limited functionality, the PDA was dispensable. The following comments illustrate these anaesthetists' frustrations: "*It probably isn't an issue for a lot of people but you put it in your pants pocket and your pants fall down*.", "*I started out using it for everything and taking it everywhere and that is when I started using the Emins and things like that but overtime it stays in my locker now before I go home I log my cases on it. It falls out of your pocket and skids across the operating theatre it is not necessarily practical*." *"Sometimes if I do take it with me, I leave it in the operating theatre on the anaesthetic trolley. But I sometimes worry it is going to get swiped." "I think that is a real issue having to carry another thing on the body because we've got our pagers. Sometimes we have two pagers and by the time you have your stethoscope and all the rest of it, quite a few things."*

Further problems the hospital faced were with regard to funding such initiatives. The Senior Anaesthetist described the problem "*We work in organisations that are very strapped for cash and our IT department is grossly under resourced still cottage industry stuff. We don't measure what we do and until we are adequately resourced in our IT department then they start exploring these sorts of issues and they have personnel and resources to say let's go ahead let's do this*."

For users in this environment in this case study the uptake and use of a PDA will be difficult. Although the device is small and mobile without the additional functionality that comes with a wireless infrastructure a device that serves only one purpose in a hospital environment, is of limited value. As this case study has illustrated the users are not prepared to carry the device therefore the fact that it is mobile is irrelevant.

Case study 2: Implementation and use of mobile devices in an acute hospital ward

The Neuroscience ward has 30 beds and for the last 18 months has had wireless base stations installed. There are notionally three units containing 16 neurosurgical beds, 4 neurology beds and 10 stroke beds but these numbers are flexible depending on patient needs. The neurology and stroke units tend to be occupied by patients who are predominantly emergency admissions and the neurosurgical patients are a mix of elective surgery and emergency admissions. There are 42 effective full time staff (EFTS) staff of nurses and registrars. The resident medical officers turn over every 2-3 months. In addition there are approximately 30 allied health professionals – physiotherapists, occupational therapists, speech pathologists, neuropsychologists, social workers, dieticians etc. all rotating through the unit at any one time. This results in a large number of staff who need to access the same records and the same data to track the patients.

- Purpose

The two key champions for the implementation of mobile devices in this ward are the Nurse Unit Manager (NUM) of the ward and the Stroke Liaison Manager (SLM). Both are enthusiastic supporters of technology and have invested significant time in investigating and establishing a wireless infrastructure.

The NUM and the SLM began two years ago investigating the possible use of mobile devices on this ward to improve the delivery of patient care. They have had discussions with IT professionals and researchers about a potential wireless infrastructure. Currently the ward has two desktop PCs, pagers, fax and deskbound phones. Some clinicians use mobile phones. The wireless infrastructure recently installed includes, 3 wireless base stations, 2 wireless laptops, 1 wireless tablet PC (for displaying x-rays in particular), 1 ruggedised PDA (including bar code reader), 2 customised trolleys for moving laptops around the ward, 3 wireless IP phones.

The SLM is specifically interested in using the devices to collect data and manage information flow, as he said " .. *how patients travel through the service, what the time frames are for that service, what kind of clinical assessments have been done for these patients, what are the risk factors, treatment plans, what investigations they have had done as well as looking at the projected outcomes of these patients based on the data*." It is also expected that the pharmacy information will be available through the mobile devices.

There is also a need for better management and coordination of information of patients at the time of discharge. As explained by the NUM "*We need to know what services are available at home. So if someone is going home do they have counsel services, visiting nursing service, physiotherapist, case worker in the community. We need to know what that information is. None of those services communicate with each other. So it can be a huge issue for us*." Currently much of this information can be collected four or five times by different people.

- Choice of mobile devices

The selection of the different mobile devices implemented has been very much based on the experience of the NUM and SLM and their perception of the technology needs within the ward. The different devices that have been deemed to be the most useful for this ward include:

Laptops on trolleys: The NUM worked with an engineering firm to design an appropriate trolley. The laptops are placed on top of individual trolleys so staff can wheel them to each patient's bed and around the ward for administrative use. The use of the trolley has been very important in improving the usefulness of the laptop. For example the trolley must also hold a range of documents such as pathology request forms, drug charts and equipment, discharge summaries, in patient care progress notes and patient files kept in an A4 folder. One of the key features of the trolley are the wings on the side of the trolley which can be raised or lowered offering the clinicians a table at the bed side. Figure 1 is a photograph of one laptop on the trolley in use.

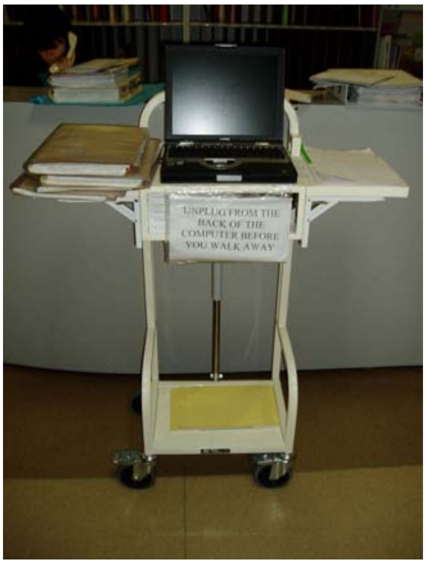


Figure 1 Laptop on a trolley.

The laptop and trolley are extensively used during multi-disciplinary team meetings. The NUM explained: "*We have formal sit down meetings where we discuss everything and we have our multi disciplinary handover each morning. Now D. (SLM), while having the laptop, has established a reporting program, so that minutes can be taken as these meetings are being held, which can then be printed onto a progress page which is then inserted into the medical record. So instead of leaving the meeting and making your own notes about what we have heard, we are reporting a 'set of minutes' for want of a better word, an 'Action Plan' of what has been discussed in that meeting. You then put the pages into the printer, print it up and put that page into the medical records*". The wireless infrastructure facilitates the printing of these notes the NUM explained "*we print to a wireless system printer so no matter what room we are in while we are having that meeting, we can move lap top around and print to a central printer.*"

Physiotherapists are using the laptops to check the CT scans of patients who have chest problems. Speech Therapists use the laptops to check the Videofluoroscopy of swallowing for patients. Also Stroke and Neurology rounds are conducted using the laptops.

PDA or Tablet PC enable nursing staff to look up MIMS Online to help in the administration of drugs. The SLM thinks it would be desirable for this to be located in the drug room. The SLM described how admitted patients have a barcode on their wrist label "*One of the tools we have been trialling has the capacity to barcode the patient and get that data up on the screen*." The use of bar codes was also identified as important for ordering of medical supplies and ordering pathology services. It was noted by the SLM that the drug room is too small for a laptop and given that the PDA has a bar code scanning facility it can be used to scan the barcodes of the drugs.

Tablet PC: X-rays in this hospital are no longer printed in hardcopy but are available electronically. Using a tablet PC enables the X-rays to be viewed at the patient's bedside. Use is also made of barcode technology.

- Use and value to users

The current infrastructure has only recently been successfully configured and the feedback is positive but anecdotal. There is a high level of enthusiasm and a high demand for the technology as the SLM describes "*We've got more computers than any other ward in the hospital with the exception of Intensive Care. We're quite privileged in that regard but we have so many people that want to use computers that it doesn't make it any easier. We almost need one for every single person on the floor*." There is also interest from staff from other wards. Given the demand the ward is currently investigating increasing the number of lap tops on trolleys.

- Problems

Currently this is the only ward with a wireless infrastructure there are therefore limitations on the use of the devices. One major problem has been the time it has taken to get the infrastructure working, over two years. This has created a great deal of frustration amongst the staff. There are also security matters that need to be addressed at the hospital IT administration level. The SLM described some of the problems"*No matter how many we have, people are still wanting a computer they can use and they cannot get on to them when they want them; there are issues of logging on and off; having your own passwords; so you can have things open where anyone can access; so we need security there. In the bigger picture the hospital needs to look at how people can be identified onto system, so that what's happening is happening in the right name. You know, so it's traceable, and that's a huge can of worms. And at the moment our system is very slow and just doesn't meet our needs. That's something that I can't fix.*"

Other technical problems also exist. The battery life of the laptop is not adequate. Breakages and technical problems need more support. There have also been technical problems with the wireless infrastructure.

DISCUSSION

One of the outcomes of this research was to provide a definition for 'mobile' in a hospital work environment. Mobile systems for this research are defined at the device level and include infrastructure to support wireless and mobile devices such as tablet PCs, PDAs and laptops. Mobile systems allow users to physically move around while using a device. In a hospital this means being able to access and deliver data at the point of care – at the bedside. Wireless networks and mobile infrastructure allow real-time access to network, communication and computer resources while moving between physical locations. The key element in the usefulness of mobile devices in a hospital setting is clearly the availability of a wireless infrastructure. The wireless infrastructure provides for a more extensive and even creative use of the technology as illustrated in the two case studies.

In addition to mobile systems there is also a need to understand the context of use and the users. Dawson et al, (2002) suggested a conceptual framework for understanding mobile systems in work places called the Mobile Work Domain. In this framework,

“Device-level communities are defined as environments where various types of mobile and wireless devices are continually connected to one another by wired and wireless networks.

User-level communities are defined as environments where users are able to communicate using device level communities (and wired and wireless networks). These communities are often communities within organizations such as work teams, departments, business units, etc, but can also represent supply chain communities of interrelated suppliers and customers.

Nomadicity is defined as that characteristic of a worker who, as a member of a user level community, conducts his/her day-to-day work activities by moving from place to place in a virtual office supported by device-level communities. Nomadic workers require access to data and services independent of location. [p 5]”

Based on the findings of our case studies we suggest an extension of that framework from device and actor centred definitions to include a process centred definition called a **user-activity-level community**. User-activity-level communities are defined as environments where processes are set up which allow users to communicate using devices at a process level, that is, a user defines a process and carries out a process using device-level community. These communities are workflow communities which exist to support a process for the length of time it takes to accomplish that process. These workflow communities would form and reform as necessary to carry out a specific task or process. Figure 2 provides a model describing a Mobile Hospital Work Domain. The figure identifies the relationship between the device level communities (characteristics of mobile device required and examples of suitable devices) and user-activity-level communities (processes performed by users using the device level communities) in the hospital settings discussed in the case studies. All the devices listed can be used without a wireless infrastructure but what makes them valuable in a hospital setting is the facility to access the wireless infrastructure.

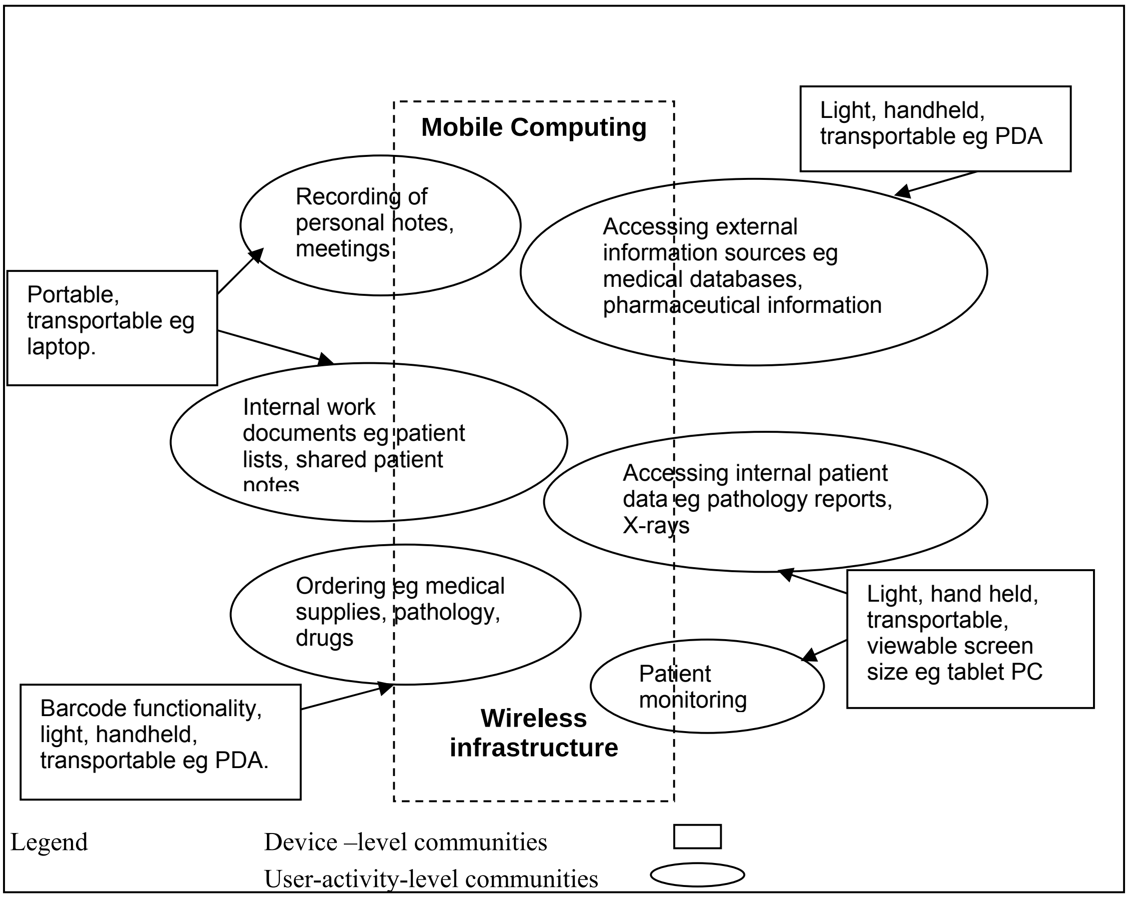


Figure 2: Mobile Hospital Work Domain

Istepanian et al (2004) define mobile health as “mobile computing, medical centre, and communication technologies for healthcare.” (405). This definition focuses only on technical infrastructure and does not really capture the essence of a mobile environment in a hospital setting. The definition we believe needs to be extended to include wireless infrastructure which is essential for a truly mobile work domain, and the business or work processes which are built on that infrastructure.

We therefore propose that a mobile environment in a hospital context is defined as *wireless infrastructure supporting mobile systems which allow nomadic hospital workers real-time access to network, communication and computer resources independent of location so that they can carry out their work.*

CONCLUSION

This paper has presented a working definition of a ‘mobile system’ in a hospital ward setting. This definition is necessary in order to provide a foundation and context for research and discussion in mobile computing systems and work environments in hospitals. To date no such definition which incorporates both the social and technical dimensions of mobile computing systems in hospital settings has been proposed. This paper also presents a model of the Mobile Hospital Work Domain for ward based settings which extends an existing Mobile Work Domain model to include process modelling in user-level and device-level communities.

Future research will extend this model into broader healthcare environments including the “road warrior” section of the clinical healthcare community who attend to patients outside hospital ward settings often in the home where modem-based, satellite, microwave or global wireless networks will provide the infrastructure for a Mobile Computing Healthcare Domain.

REFERENCES

Al-Ubaydli, M. (2004). "Handheld computers." *British Medical Journal*. 328: 1181-4.

Bent, P., Bolsin, S., Creati, B., Patrick, A., and Colson, M. (2002) "Professional monitoring and critical incident reporting using personal digital assistants", *Medical Journal of Australia*, 177, 496-499.

Biemer, M. and F. Hampe (2005). "A mobile medical monitoring system: concept, design and deployment". *International Conference on Mobile Business, Sydney, IEEE.* 464-471

Breu, K., C. Hemingway and C. Ashurst (2005). "The impact of mobile and wireless technology on knowledge workers: an exploratory study". *13th ECIS*, Regensburg, University of Regensburg. CD

Carlsson, C., J. Carlsson and W. Pirkko (2005). "Mobile services for the hospitality industry". *13th ECIS*, Regensburg, University of Regensburg. CD

Carlsson, C., Hyvonen, K., Repo, & Walden, P. (2004). "It's all about my phone! Use of mobile services in two Finnish consumer samples." *17th Bled eCommerce Conference*, Bled, Slovenia, June 21-23, 2004.

Carroll, J., Howard, S., Peck., J. and Murphy, J. (2003). "From adoption to use: the process of appropriating a mobile phone", *Australian Journal of Information Systems*, 10:2, 38-38.

Fernando, J. (2004) "Factors That Have Contributed to a Lack of Integration in Health Information System Security.", *The Journal of Information Technology in Healthcare,* 2, 5, 313-328.

Gururajan, R. and S. Murugesan (2005). "Wireless solutions developed for the Australian health care: A review". International Conference on Mobile Business, Sydney, IEEE. 472-478

Heslop L Howard A Fernando J Rothfield A Wallace L (2003) "Wireless communications in acute health care.", *Journal of Telemedicine and Telecare 9 (4) 187-193*

Innes, D., S. Barnes and E. Scornavacca (2005). "The impact of wireless Field Force automation of New Zealand trade services organisations". *International Conference on Mobile Business,* Sydney, IEEE. 49-55

Istepanian, R., E. Jovanov and Y. Zhang (2004). "Guest Editorial. Introduction to the special section on M health: beyond seamless mobility and global wireless healthcare connectivity." *IEEE Transactions on Information Technology and Biomedicine* 8 (4): 405 -- 414.

Koop, A. and R. Mosges (2002). "The use of handheld computers in clinical trials." *Controlled Clinical Trials* 23(4): 469-480.

McAlearney, A. S., S. B. Schweikhart and M. A. Medow (2004). "Doctors' experience with handheld computers in clinical practice: qualitative study." *British Medical Journal.* 328: 1162-6.

McLeod, T. G., J. O. Ebbert and J. F. Lymp (2003). "Survey Assessment of Personal Digital Assistant Use among Trainees and Attending Physicians." *Journal of the American Medical Informatics Association*. 10: 605-607.

Page, E. S. (1954). "Continuous Inspection Schemes." *Biometrika* 41.

Prasad, M., E. Scornavacca and H. Lehmann (2005). "Using Wireless Personal Digital Assistants in a Restaurant: impact and perceived benefits". *International Conference on Mobile Business*, Sydney, IEEE. 69-81

Shewhart, W. A. (1931). *Economic Control of Quality Manufactured Product.* New York, Nostrand-Reinhold.

Wang, Y., E. van de Kar, G. Meijer and M. Hunteler (2005). "Improving business processes with mobile workforce solutions". *International Conference on Mobile Business,* Sydney, IEEE. 94-102

Integrating RFID Technology In To Mobile Business. About the Possibilities to Distribute Free and Paid Content on Mobile Devices Through RFID Technology in Retail Business

Florian Stahl

Center for Business Metrics - University of St. Gallen

Marion Freudenschuss

Vienna University of Economics and Business Administration

Abstract

The attention on Radio Frequency Identification (RFID) and Mobile Business and Mobile Content has recently attracted in various industry sectors, the media and in academic research. This article focuses on the convergence of RFID technologies and mobile devices and presents applications in the retail business. Based on related literature and existing models and technologies, a business framework and service architecture is derived which describes new ways of distributing product information through the convergence of RFID technology and mobile devices in the retail business, as well as identifying the key players involved in this. The article outlines applications of the business framework and service architecture in product advertisment, Mobile Business and Paid Mobile Content which involve connecting physical products with virtual product information made available through integrating RFID technologies and mobile devices.

Keywords:

Mobile Business, Mobile Content, RFID Technology, Retail Business

INTRODUCTION

For several years there has been debate on Radio Frequency Identification (RFID) technologies from a technological point of view. Several case studies with a business perspective also exist for the retail segment, describing and predicting a fundamental transformation of the supply chain due to RFID technologies. Shopping has become more individualised and more comfortable (Loebbecke 2005a). RFID technologies make it easier for consumers to compare products and to utilise product information for their purchasing decisions. According to Loebbecke (2005a) customers have become sophisticated and have gained a clear understanding of value, generated from vast information resources. Delivering this value is an enormous challenge for retailers (Loebbecke 2005a).

The basic idea of RFID is storing and remotely retrieving data using devices called RFID tags or transponders. According Wikipedia.org a RFID tag is a small object that can be attached to or incorporated into a product, animal, or person. RFID tags contain antennas to enable them to receive and respond to radio-frequency queries from an RFID transceiver. Passive tags require no internal power source, whereas active tags require a power source (Wikipedia.org 2005). RFID tags are often envisioned as a replacement for barcodes in retail business, having a number of important advantages over the older barcode technology. An RFID tag contains an incredible amount of information. The data contained in the RFID tag can be read by so-called RFID readers. RFID tags ‚identify' themselves when they detect a signal from a compatible device (Loebbecke 2005b). As a tag passes through a radio frequency field generated by a compatible reader, it transmits its stored data to the reader, communicating information about the object to which the tag is attached and further content.

Whereas most business case studies on RFID technologies in retail business discuss (Strüker et al. 2004, Loebbecke 2005b, METRO Group 2005a, METRO Group 2005b) RFID readers and content provision devices like personal shopping assistants or information terminals, the customer's personal mobile devices such as cell phones are not considered as content provision devices. Certain cell phone manufacturers such as Nokia, however, have long taken an interest in RFID. At the CTIA Wireless IT & Entertainment Trade Show 2004 Nokia presented an early prototype built in collaboration with VeriSign, who propose a central repository for RFID data that companies can use to relay information about inventory and deliveries to customers and suppliers. The prototype was based on a Nokia cell phone, with an RFID reader contained in a shell attached to the phone. Nokia's prototype has fundamental implications for the delivering of product information to a mobile device using RFID and can extend the technology "beyond the supply chain and into customer service, merchandizing, marketing and brand management" (ZDNet 2004).

<u>Research Question:</u>

Given the subject's relevance the following research questions will be considered in this paper. What are the consequences of the convergence of RFID technology with mobile devices – and the association of physical products with corresponding virtual information and content which thus becomes possible – on the way retail products are sold, on the conception of advertising and on the emergence of new business models for cell phone operators and producers of paid content?

The paper is organised as follows. The following section will give an overview over the literature on consumer choice, mobile business, mobile content and RFID technologies in retail business. Based on a model of object hyperlinking a theoretical business framework and services architecture will be derived. In the last section, business applications derived from the theoretical business framework and services architecture and the implications of these business applications are discussed.

RELATED LITERATURE

Consumer Choice

Many goods and services traded are experience goods, meaning that the buyer only recognises the good's or service's true properties after the purchase, when consuming the good or service. Often these goods are refereed to as credence goods, as the vendor's reputation determines whether the buyer trusts the information on the good or service offered by the vendor. The asymmetric distribution of information for goods with experience and credence properties leads to a classical principle-agent relationship between the buyer and the vendor at the time of purchase – the vendor of the product is better informed about the product's true properties than the consumer (Milgram, Roberts 1992).

Given incomplete information consumers buy goods and services based on different attributes such as price or quality. The information about a good or service can be imperfect for a number of reasons, such as the proliferation of competing brands, the difficulties of exhaustive search or sampling, biases in product evaluation, constant product innovation or consumer mobility (Newmann 1977, Thorelli and Thorelli 1977). Thus, price and quality of a good or service are the dominant attributes for the choice of a brand, a provider or a vendor. Information concerning the quality of a good or a service is more problematic, as it is more difficult to communicate before or at the time of purchase (Tellis, Gaeth 1990). Tellis und Gaeth (1990) state that the effect of incomplete information on consumer choices is therefore important for analysing consumer behaviour and firm strategy.

As discussed in the introduction, RFID technologies have the potential to provide consumers with information about goods and services at the time of purchase, which alleviates the principle-agent problem and has direct influence on consumer choice, as discussed by Newmann (1977), Thorelli and Thorelli (1977) or Tellis and Gaeth (1990). The direct influence of RFID technologies on purchasing and consumption behaviour is explained with a model for RFID technologies in mobile devices which also takes account of the integration of paid content from third party providers. The implications of the integration of paid content from third party providers at the time of purchase using RFID technologies is dealt with in the last section of the article.

Mobile Business and Mobile Content

In recent years, computing has been moving toward creating pervasive and ubiquitous environments. Countless small mobile devices such as cell phones, palm pilots, and personal digital assistants (PDAs) have been developed to provide users with ubiquitous access to information. This section provides a general overview of the value chain in mobile industries and of mobile content, the major driving force for the telecommunications and media industries, as the basis of a general business and service framework of RFID technologies in mobile devices. The value chain concept is used according to the definition given by Timmers (Timmers 1998), which he developed for the analysis and design of business models for electronic markets. Timmers (1998) bases his concept of a value chain on Porter's management approach (Porter 1985, 1998). In analogy to Porter (1985, 1998), the value chain determines the architecture of the product, the service and the information flows in Timmer's (1998) concept. Due to this relationship the analysis of a business model can only be carried out by taking account of the value chain. „A systematic approach to identifying architectures for business models can be based on value chain de-construction and re-construction, that is identifying value chain elements, and identifying possible ways of integrating information along the chain" (Timmers 1998).

The value chain that defines the mobile content environment depends upon a wide range of factors. According to the OECD Mobile Content Report (OECD 2005) the underlying value chain for mobile content is long, complex, varies for different types of content and is still developing. The handset or device frequently dictates or constrains a user's access to mobile content (OECD 2005).

As many players in the mobile content value chain originate from mature industry sectors such as wireless telecommunications, media and entertainment and the device industries, the OECD claims that the mobile content industry is experiencing the integration of the very different value chains of these three industries. The differing strength of the individual industries and the integration of different value chains has resulted in the vertical and horizontal integration of value chains in the mobile content industry in recent years. Major industries such as movie studios, record labels and game developers are moving into mobile content, and each is vying for a central position in the value chain. Mobile operators are at the centre of a complex web of value creation for mobile content because of their primary interface with mobile consumers (OECD 2005).

As content becomes increasingly digital, more and more of it will be delivered over mobile platforms. The capabilities and bandwidth of the mobile networks are continually being expanded by mobile operators. Greater bandwidth enables the download and sharing of larger, more sophisticated files in an acceptable timeframe for a particular content application via WAP or Web browsers. Next to entertainment, video, music and game content, contents like mobile tourist guides or shopping guides for consumers are further examples of new innovative mobile content business models.

Not only paid mobile content, however, is gaining importance in the mobile business, but free mobile content is too. The term mobile marketing has entered the discussion designating services and business models where users are informed about products and services by means of free content or where free on-demand information and content is made available to users of mobile devices. Mobile marketing can be defined as the use of interactive wireless media to provide customers with time and location sensitive, personalised information that promotes goods, services and ideas, thereby generating value for all stakeholders (Dickinger et al. 2004, Bragge et al. 2005). “As mobile devices are very personalized and they allow the interaction with the users virtually any time and anywhere, mobile advertisements must be custom-made and also adaptable to different types of terminals” (Bragge et al. 2005).

The integration of RFID technologies in mobile devices can not only influence the value chain of mobile business models of paid content but also make new kinds of mobile marketing feasible by means of personalised advertising in the sales room and at the moment of decision making. The next section will elaborate the ways in which the value chain is influenced and which possibilies for personalised advertising emerge due to the integration of RFID technologies in mobile devices with the help of a business framework and a service architecture.

RFID Technologies in Retail Business

In retail business a RFID tag is best compared to a conventional barcode. A numeric, article-specific code stored in the RFID chip is assigned to each product, also refereed to as an „Electronic Product Code“ (EPC). The EPC is a 96-bit tag which contains a number called the Global Trade Identification Number (GTIN). Unlike a UPC number, which only provides information specific to a group of products, the GTIN gives each product its own specific identifying number, giving greater accuracy in tracking (Wikipedia 2005b). As soon as the chip with the ”Electronic Product Code“ is within range of a RFID reader (>1m) it sends its numeric code to the reader. The reader then matches this code to further information and content about the product or service such as its price, size or weight which the reader either has stored or which it obtains from other data sources.

In keeping with Porter’s (Porter 1998:46) value chain concept, RFID technology can be employed all along the value chain of retail stores and offers the retail business efficiency and cost advantages at every step of the value chain as has been shown by a number of studies (Loebbecke 2005b, Kaerkkaeinen 2002, METRO Group 2005c).

- The impact of RFID on the *inbound logistics* of the value chain of the retail business is related to the possibility to monitor incoming and outgoing goods quickly and precisely. RFID allows several processes concerning incoming goods to be automated simultaneously (“merchandise picked for an order is matched with the shipment and approved for transport, while the merchandise planning and control system registers the outgoing goods” METRO Group 2005c) as well as allowing for the automatic monitoring of stocks by retailing and industry partners, making time-consuming, manual scanning obsolete (METRO Group 2005c).
- The *operations* of the retail value chain can be optimised with RFID technology with respect to warehouse management, storage processes, automating goods handling and administrative processes (e.g. correct sorting of merchandise, constant and accurate overview of the location of goods). As there is a growing tendency toward smaller shipments and mixed loads being delivered to stores, requirements for handling efficiency in the supply chain will increase (Fitzek 2003).
- RFID influences *outbound logistics* in retail business in particular with regard to the payment procedure in the store (automatic check-out at the cash desk by reading several articles simultaneously without each item having to be

scanned). Automating the check-out process attains cost savings (Hawkes 1994) and improves theft prevention with shelves that report if a large amount of goods is removed from them as well as increasing authenticity control (Ashton 2000).

- RFID has an impact on *marketing* and *sales*, as retailers manage their stocks more efficiently and reliably. Because consumers always find the desired goods on the store shelves, they are not led to purchase alternative products or patronise competitors. Brand loyalty is supported and industry partners do not lose sales (METRO Group 2005c). Personalised advertisment displays and self-check-out increases customer loyalty and promotes sales. Electronic price labelling in stores makes price tags easier to understand and more up-to-date (Loebbecke 2005b).
- The *services* of the retail business value chain are influenced by optimised traceability. Each transponder-equipped object is positively identifiable. Should a recall programme be necessary, RFID enables retailers and manufacturers to conduct it in a targeted, safe and efficient manner (METRO Group 2005c).

In these scenarios, the benefits described are often significant, and a total re-invention of the supply chain way of working is promised by enabling consumer packages to efficiently communicate with their environment. However, an exact description of how the benefits are attainable in practice has often remained vague, and the process and costs of implementing the solution have not been explained (Kaerkkaeinen 2002).

This paper aims to integrate these different research areas about consumer choice, mobile business, mobile content and RFID Technologies with the aim of developing a business framework and service architecture which describes how new forms of product information, personalised advertising and promotion are being created and which new fields of business are emerging for mobile operators thanks to the integration of RFID technology in mobile devices. The business framework and the service architecture focus primarily on the influence of RFID technology on marketing and sales within the value chain of retail business.

THEORETICAL BUSINESS FRAMEWORK AND SERVICE ARCHITECTURE

In this section a generic business framework and service architecture for developing business models and mobile applications is presented. As described in the previous section, RFID technology is at present primarily being discussed with regard to the logistics between producer, supplier and retail store or in connection with in-store product information and is being tested by individual firms such as METRO, Tesco or Wal-Mart. Figure 1 presents the applications of RFID technology for retail business addressed in the current discussion..

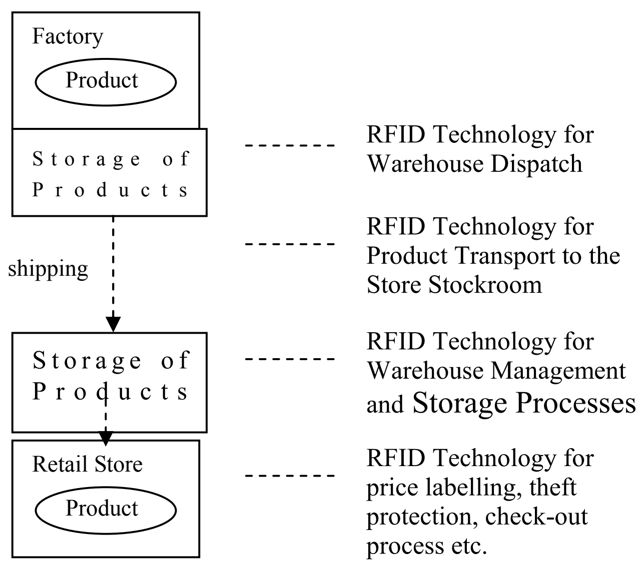


Figure 1: Main applications of RFID technology in retail business logistic processes and value chain

As mentioned in the introduction prototypes of RFID readers integrated into mobile devices like cell phones already exist. The scientific literature discusses this convergence of mobile technologies and RFID technologies and the resulting application possibilities using the terms “websigns” and “object hyperlinking” (Pradhan et al. 2001, Sphorer 1999). According to Pradhan et al. (2001) the key idea of „websigns“ is that users would benefit from having access to devices that combine the advantages of wireless technology and ubiquitous computing to provide a transparent link between the physical world around them and the resources available on the Web. Websigns are an alternative way to map e-services using a simple form of augmented reality (Pradhan et al. 2001). To create a hyperlink between a physical entity and a Web resource, Pradhan et al. attach RFID tags to people, places, and things to associate them with an appropriate universal resource identifier, resolving the URI in the network if it is not already a URL. “These hyperlinks rely on commonly available wireless mobile devices to help users automatically access services associated with physical objects. [...] To receive a URL, users point their mobile device in the direction of a RFID tag and receive periodic broadcasts“ (Pradhan et al. 2001). Figure 2 illustrates the way „websigns“ work.

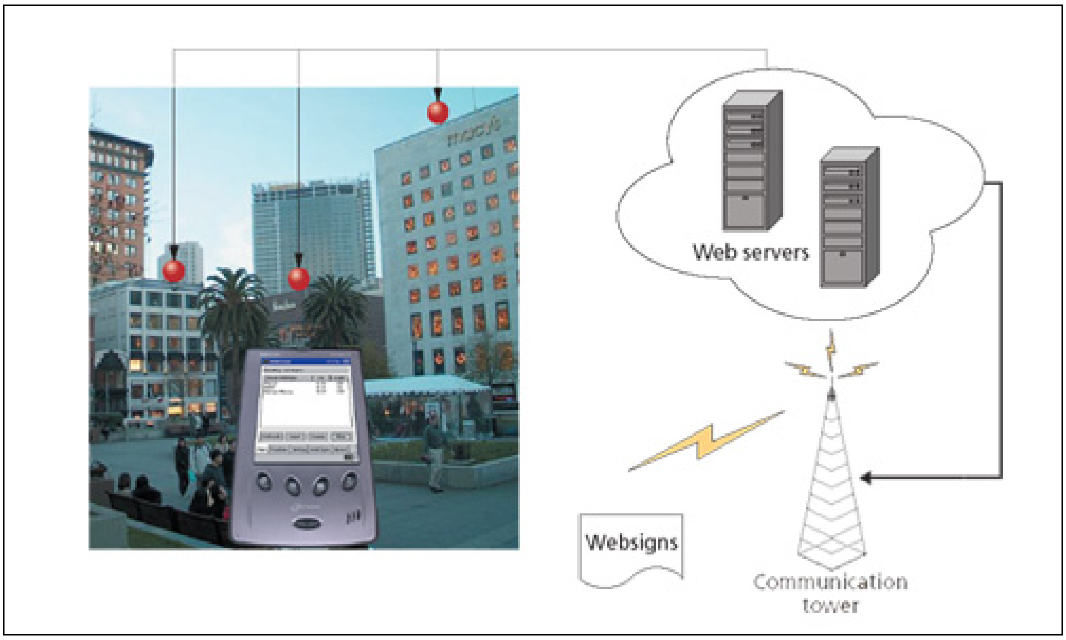


Figure 2: Example of a “websign”: Web servers detecting websigns with a wireless PDA in an urban environment (Source: Pradhan et al. 2001)

Within the CoolTown research programs at Hewlett Packard Laboratories Pradhan et al. (2001) have developed a ubiquitous computing “websign” system which can be personalised for individuals, group-targeted for users with specific interests or needs, or made universally available to anyone who wants to use them. To facilitate class-specific filtering, it is possible to categorise websigns by type – for example, restaurants, theatres, or historic sites.

Based on the ubiquitous computing “websign” system a business framework and service architecture is developed, specifying how RFID technology can be employed in retail stores to communicate and inform customers at the moment of decision making. The business framework and service architecture does not focus on the RFID applications for logistics as in other scientific articles (cf. figure 1) but rather first and foremost on the processes involving marketing and sales in the value chain of retail business. Figure 3 presents the business framework und service architecture..

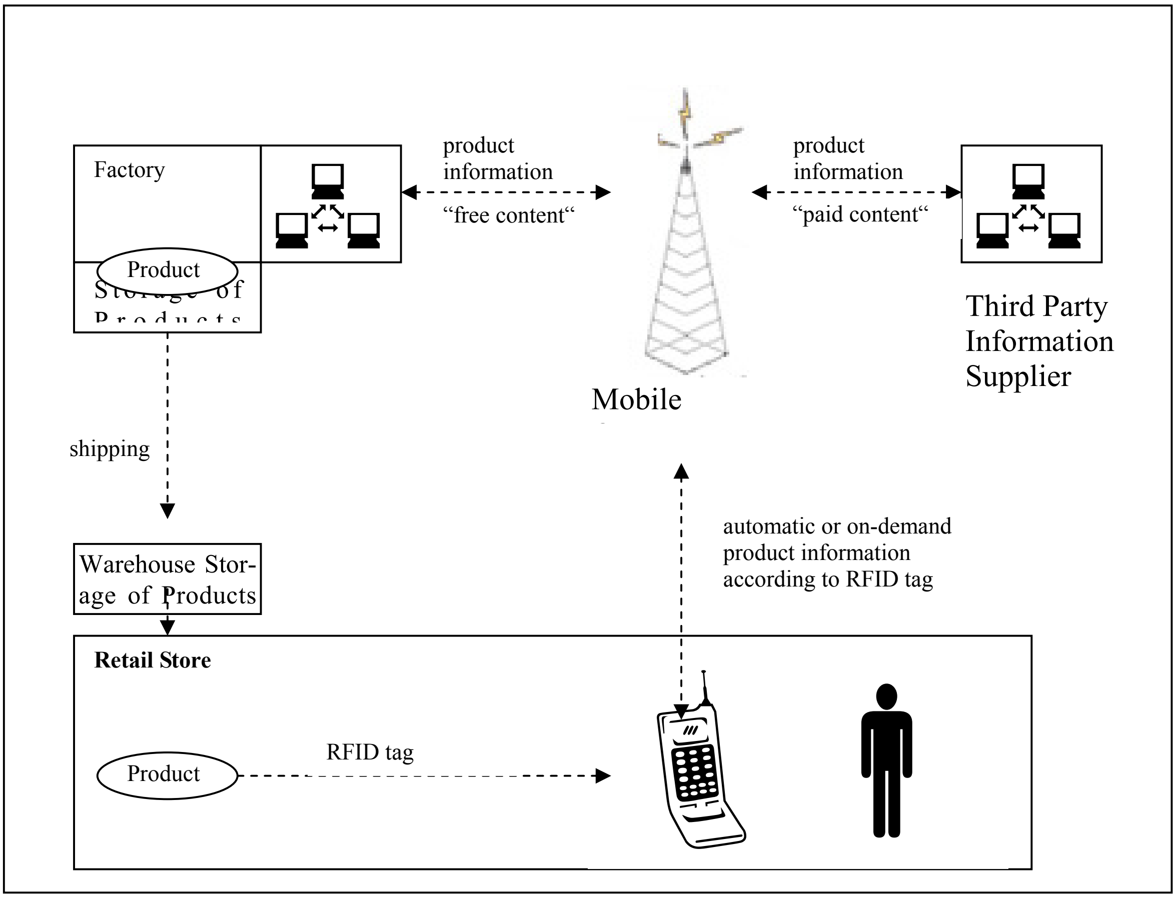


Figure 3: Business Framework and Service Architecture to distribute free and paid content about products on mobile devices through RFID technology in retail business

Within this business framework and service architecture the customer can be informed about a product at the time of purchase in a retail store (e.g. advertisements) or she can access product information from the product's manufacturer or a third party information supplier (e.g. comparative product test) through a RFID reader in her mobile device (e.g. cell phone). In this business framework the key players in the value chain are not only producers, suppliers und retailers of a product as in other RFID frameworks in the retail sector. Rather, mobile operators and third party information suppliers also become key players in the framework.

The key essence of the framework is not logistics of products but information for customers about a product at the moment of decision making. This framework introduces transformations into two aspects of the existing value chain in the retail sector. One concerns product advertising, the other the degree of information a customer has and consumer choice. Furthermore, the technological implementation of this business framework has the potential to change the existing balance of market power. The implications of the business framework for product advertising, customer choice and market power are discussed in the next section.

DISCUSSION AND IMPLICATIONS

The business framework and service architecture which was presented in the last section is a further step in the investigation of the growing complexity and connectivity between three formerly separate interactive spaces – the real world of retail stores, the mobile world and the virtual world.

As described in the context of consumer choice, goods and services are often experience and credence goods, because the true properties of a good or a service do not become apparent until it is consumed. Thus a good or a service is more than a physical object or service, as the product has properties which cannot be recognised without consuming it. For this reason information about the properties and attributes of a product are necessary for the purchasing decision to minimise the buyer's risk. Nowadays such information – e.g. product information provided by the manufacturer, product ratings by other buyers or product tests by independent test institutes – is available on the Internet. The physical object or service which represents the product has a virtual counterpart (Nicolai 2005). With RFID technologies, the product can be digitally augmented if the digital data is attached to it, for example, by sticking a RFID tag or integrating it into an embedded system. "Attaching a digital tag enables the connection of the product to the virtual world by either storing information on the chip itself or storing a link to the Internet which can be used by mobile devices as the premier source of information" (Nicolai 2005).

The connectivity between three formerly separate interactive spaces – the real world of retail stores, the mobile world and the virtual world – which is presented in the business framework and service architecture influences product advertisement as well as the degree of information of the customers at the moment of decision making.

Due to the connection of physical products in real retail stores to information about customers and their purchasing behaviour by means of RFID technologies, the possibilities for product advertising are expanded and the efficiency of advertising measures increases. It has never before been so easy for retailers and producers to place strongly personalised advertising at the place and at (or near) the time of the purchasing decision, taking into account the attributes of the customers and consumer behaviour. „Sellers now have far more information about their customers through personalization technologies. This make it possible for retailers to know the identity of their consumers in "real-time", to access data on past shopping patterns, and to change prices and offerings where appropriate" (Vulkan 2003). Due to this technological convergence mobile marketing gains further importance. By means of the connection of the real products to offers of digital virtual information about the products with RFID technologies and mobile devices, new forms of customer relationship management become possible for providers, as not only a customer's purchasing history is known but also her purchasing intentions as the access of information can be observed. In principle this opens up the opportunity to engage in the strategies of "push" marketing (tailoring marketing messages to individual customers) and mass customisation (low cost tailoring of the same basic product for individual customers) (Vulkan 2003). Whereas up until now it was only possible for the retailer to advertise a product at the place of purchase (usually depending on the products margins), the producer now can also target the customer at the time and place of purchase. The producer of a product will now gain a part of the hitherto exclusive power of the retailer regarding promotion and advertising at the place of purchase.

The connectivity of a physical product with virtually provided digital information all about the product by means of RFID technologies in mobile devices such as cell phones means the customer can easily and conveniently access information about the product at the time and place of purchase. Products can thus be compared with other products regarding price and other properties at the moment of decision. This possibility to access information at the time and place of purchase results in a much higher degree of information for the customer for her purchasing decision. A possible hypothesis is that the convergence of RFID technologies with mobile devices will change the demand elasticity of products. From the perspective of consumer choice the possibilities of on-demand information at the time and place of purchase are comparable to the "Shop Bots" offered on the Internet, which visit many Web sites, interrogate them to find the price and return with the lowest price. It thus becomes unnecessary to compare product prices in different retail stores. „Shop Bots have the potential to reduce consumer search costs to zero" (Vulkan 2003). This leads to the question of whether the convergence of RFID technologies and mobile devices and the easy and convenient access to information on the product and on product comparisons that this makes possible results in a reduction of the prices. As the search costs for price information on electronic markets decrease, the price differences for comparable goods between different vendors should be small (Luo & Chung 2002). However, several empirical studies concerning Shop Bots on the Internet (Bailey 1998, Clemons, Hann, & Hitt 1998, Brynjolfsson & Smith 2000, Lee & Gosain 2000), show that price dispersion (the difference between the lowest and the highest price for same goods on a market) still exists on markets with easily accessible information and easy product comparisons and has not decreased.

Apart from the effect on product advertising and the consumer's degree of information the business framework and service architecture presented in the previous section has an effect on the value chain of mobile operators and paid content producers. In this business framework and service architecture mobile operators connect consumers to the providers of information. Due to the convergence of RFID technologies and mobile services mobile operators not only have access to data about the communication behaviour of their customers but also to data concerning their consumption behaviour and purchasing intentions, seeing as all this can be tracked when information from the producer of a product or third party information is accessed thanks to an RFID tag. For the providers of information mobile operators are the central access channel to a customer, as they are in possession of all the relevant customer information. This opens up the possibility of new business fields for mobile operators such as offering producers or vendors of goods and services personalised advertising and marketing strategies (in the moment of decision making). Through integration of RFID Technologies in mobile devices mobile operators get key competences in market and marketing research.

However, not only mobile operators but also producers of paid content concerning products such as product tests or products comparisons can boost their sales volume thanks to the convergence of RFID technologies and mobile devices. In addition to offering their paid content over the Internet they can now make it easily and conveniently available as mobile content by means of RFID technology. It also makes it possible to increase the mobile operator's sales volume through mobile content, as these new content services, which meet a need of consumers, are billed via the mobile operators. RFID technologies in mobile services thus indirectly represent new applications for the mobile operators' UMTS (3G) technologies.

The applications for product advertising, product producers, mobile operators or paid content producers considered in connection with the business framework and service architecture, however, are subject to a restriction which will severely limit the realisation of the applications: the privacy issue. The use of RFID technology has engendered considerable controversy and even product boycotts. The aspect which concerns consumer advocacy groups is the globally unambiguous labelling of everyday objects, which thus can be identified anywhere in the world without the owner being aware of this. The convergence of RFID technology and mobile devices would be a big step towards the transparent consumer. To avoid the privacy problems of the described applications of the business framework and service architecture the proposed privacy related guidelines for the use of RFID-based EPC by EPCglobal would have to be adhered to, which require giving consumers clear notice of the presence of EPC and the possibility to turn off the RFID reader functionality in their mobile devices. In addition and thanks to the sales generated by new business fields, mobile operators could leave the choice to the customer, offering the customer lower rates or free mobile services in return for her revealing her consumer behaviour through RFID technology.

A further aspect of the applications of the business framework and service architecture of RFID in mobile devices and retail business that could be criticised could be that RIFD is too expensive a technology, and that it is unlikely that investments in RFID will pay off (Riso 2001, Burnell 1999). The sales revenues from new business fields for mobile operators and the advertising possibilities for producers of products, however, seem to answer to this criticism.

CONCLUSIONS AND FUTURE WORK

This paper discusses the applications emerging from the convergence of RFID technology and mobile devices which are: personalised advertising at the moment of decision for retailers, a higher degree of information for consumers and new business models for mobile operators and content producers for the distribution of free and paid content. The framework presented leads to connectivity between three formerly separate interactive spaces – the real world of retail stores, the mobile world and the virtual world. Physical products are linked with virtual information – with their digital counterparts. At the same time the previously virtually-limited interactive frame of the Internet is connected to the real world. "Enhancing a physical product with digital information objects can lead to completely new creations and process chains. [...] Real and virtual worlds emerge into a new expanded interactive space" (Nicolai et al. 2005).

However, the framework presented poses questions which should be the object of further research. These include the question of which types of personalised advertising can be used and how, further how a record of consumer behaviour can be stored without violating privacy issues, thus being available for the development of new products and services or finally how the technology can be implemented without RFID resulting in an information overload for the customer in a retail store.

In the future, companies will be able to utilise the numerous possibilities offered by RFID technologies. The future standards are, however, still being developed and it is difficult to foresee exactly what the future of RFID will bring us. The integration of RFID technologies into mobile telephones promises to make the much discussed concept of "augmented reality" a feasible possibility. The real and virtual worlds will become connected thanks to information and new services with the aim of supporting customers in whatever environments they may be in.

REFERENCES

Adipat, B. Zhang, D. (2005): Developing Adaptive and Personalized Mobile Applications: A Framework and Design Issues, Proceedings of the Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA

Alba, J., Lynch, J. G., Weitz, B., et al (1997): "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces." Journal of Marketing, Vol. 61, 3, p. 38-53.

Bailey, J. P. (1998). Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce. Massachusetts Institute of Technology, Cambridge.

Bragge, J., den Hengst, M., Tuunanen, T., Virtanen, V. (2005): A Repeatable Collaboration Process for Developing a Road Map for Emerging New Technology Business: Case Mobile Marketing, Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA

Brynjolfsson, E., Smith, M. (2000): Frictionless Commerce? A Comparison of Internet and Conventional Retailers. Management Science, 46(4), 563-585.

Clemons, E., Hann, I., Hitt, L. (1998). The Nature of Competition in Electronic Markets: An Empirical Investigation of Online Travel Agent Offerings (Working Paper): The Wharton School, University of Pennsylvania.

Dickinger, A., Haghirian, P., Murphy, J., Scharl, A. (2004): An Investigation and Conceptual Model of SMS Marketing, Proceedings of the 37th Hawaii International Conference on System Sciences, January 5-8, Big Island, Hawaii, USA, IEEE.

Fitzek, D. (2003): Application of RFID in the Grocery Supply Chain, http://www.unisg.ch/org/item/scmweb.nsf/ download on August 09, 2005;

Gerard J. Tellis, Gary J. Gaeth (1990): Best Value, Price-Seeking, and Price Aversion: The Impact of Information and Learning on Consumer Choices. Journal of Marketing, April 1990, Vol. 54 (April 1990), 34-35,

Hawkes P. (1994): Supertag—stock counting off its trolley, Sensor Review, Volume 14(3), pp. 2-2(1)

Kaerkkaeinen, M. (2002): RFID in the grocery supply chain – a remedy for logistics problems or mere hype?, http://www.ecr-academics.org/partnership/award/index.html download on August 21, 2005;

Lee, Z., Gosain, S. (2000): Price Comparison for Music CDs in Electronic and Brick-and-Mortar Markets: Implications for Emergent Electronic Commerce. Paper presented at the Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, Hawaii.

Loebbecke, C. (2005a): Emerging Information Systems Applications in Brick-and-mortar Supermarkets: A Case Study of Content Provision Devices and RFID-Based Implementations, Proceedings of Pacific Asia Conference on Information Systems (PACIS), Bangkok, Thailand, July.

Loebbecke, C. (2005b): RFID Technology and Applications in the Retail Supply Chain: The Early Metro Group Pilot, International Electronic Commerce Conference (Bled), Bled, Slovenia, June.

Luo, W., Chung, Q. B. (2002): An Empirical Investigation of Reputation and Price Dispersion in Electronic Commerce. Paper presented at the AMCIS 2002, Dallas, Texas.

METRO Group (2005a): "Personal Shopping Assistant". http://www.future-store.org/servlet/PB/menu/1000376_l2/1074195494455.html, download on August 15, 2005.

METRO Group (2005b): "Background Information on METRO Group Future Store Initiative: Platform for the Future of Retailing". http://www.future- store.org/servlet/PB/show/1000674/presse- hintergrundinfofsi_e.pdf, download on August 17, 2005;

METRO Group (2005c): "RFID in practice: A pilot project of Kaufhof and Gerry Weber". http://www.future-store.org/servlet/PB/show/1000672/presse- factsheet_e.pdf, download on August 17, 2005;

Nicolai, T., Resatsch, F., Michelis, D. (2005): The Web of Augmented Physical Objects, Proceedings of International Conference on Mobile Business (ICMB'05), Sydney, Australia

Pradhan, S., Brignone, C., Cui, J-H., McReynolds, A., Smith, M.(2001): Websigns: Hyperlinking Physical Locations to the Web, . IEEE Computer, 34(8):42-48

Strüker, Jens, Sackmann, Stefan, Müller, Günter (2004): Case Study on Retail Customer Communication Applying Ubiquitous Computing in: Proceedings of the IEEE Conference on E-Commerce Technology 2004(CEC'04), 6.-9. Juli 2004, San Diego, California, USA;

OECD - Organisation for Economic Co-Operation and Development (2005): „Digital Broadband Content: Mobile Content New Content for New Platforms", http://www.oecd.org/dataoecd/19/7/34884388.pdf download on June 22, 2005;

Sphorer, J. (1999): Putting Information in Places, IBM System Journal, Sept. 1999, pp. 602-628

Vulkan, N. (2003): *The Economics of E-Commerce*. Princeton and Oxford: Princeton University Press.

Wikipedia.org (2005a): RFID, downloaded from http://en.wikipedia.org/wiki/RFID at August 23, 2005

Wikipedia.org (2005b): Electronic Product Code, downloaded from http://en.wikipedia.org/wiki/Electronic_Product_Code at August 25, 2005

ZDNet (2004): RFID cell phones take shape at Nokia, http://news.zdnet.com/2100-1035_22- 5424528.html download on August 21, 2005

A Case Study of Mobile Device Use: Motivations And Trends

Linda Dawson, Susan Foster, Andrew Barnden
Monash University
Email: Linda.Dawson@infotech.monash.edu.au

Abstract

Mobile devices are increasingly being considered a key component of the connected enterprise reflecting the increasingly mobile and geographically dispersed workforce. This paper describes a pilot study whose aim was to gain an understanding of the use and motivations for use of mobile devices by individuals for business and/or personal activities. Major findings from this study indicate that motivations for the ownership of mobile devices are heavily influenced by social integration and lifestyle choice rather than work place practices while PDAs are mainly used as business tools replacing business diaries. Other findings suggest that usability problems are still evident and need to be addressed for take-up of the devices to increase beyond simple communication or calendar devices and that the notion of choice for the user in an ubiquitous environment needs further exploration. In the long term this project will focus on research in the emerging and evolving area of mobility-based business and associated mobile business models.

Keywords*:*

Mobile device, mobile commerce, survey

1. INTRODUCTION

The mobile revolution based on new wireless technologies is changing the way we do business and creating explicit mobile information environments (Carroll *et al.*, 2003; Marchegiani, 2004; Urbaczewski *et al.*, 2003). IDC (2000) a leading computing industry watchdog, estimates that the U.S. mobile work force - defined as employees spending at least 20% of their time away from the office - will be at 55 million in 2004, up from 39 million in 2000. With global sales of mobile phones in 2005 predicted to reach 740 million (GMS-Group, 2002) and an estimated 650 million 3G phone users by 2010 (2002) it is little wonder that organisations are seeking the means to extend corporate data to an increasingly mobile workforce.

Most businesses already understand the value of wireless connectivity for their employees. Primary reasons cited for deploying a wireless solution in the enterprise are to: Service mobile workers, provide access to mobile email, provide access to intranet applications, deploy specific corporate applications, improve scheduling and enable mCommerce (Guizani & Aghvami, 2002). Another factor cited for mobile computing and wireless enabling is that employees are blending company and personal time. Attention to personal telephone calls, voicemail, and even email is common during the daily commute.

The purpose of this paper is to describe a pilot study whose aim was to gain an understanding of the use of mobile devices (mobile phones, Personal Data Assistants and hand held computers) by individuals for business and/or personal activities.

The next section of this paper describes some related work in the field of mobile business, specifically in the area of user acceptance. The paper goes onto describe the research approach and the survey instrument. The findings from this study are then presented and are followed by an analysis and discussion of implications of these findings for understanding and eventually improving mobility based business activity.

2. BACKGROUND AND RELATED WORK

Kalakota and Robinson (2002) suggest that " *... the most successful appliances will be easy-to-use, task-oriented devices that leverage the benefits of Internet access to enhance their core functionality. p[24]*." Keen and Mackintosh (Keen & Mackintosh, 2001) suggest that "*There is no doubt whatsoever that, over time, M-commerce will dominate every one of the ecommerce tools and their descendants. The issue is not if, but when.* p[5]"

Although there is a growing body of literature discussing technical issues in the uptake of mobile technologies (Carlsson *et al.*, 2004; Chan *et al.*, 2002) comparatively little research has been published which explores the usage patterns, motivations or business processes being changed by the use of mobile devices. Katz and Aakhus (2002) suggest that mobile phones and other wireless devices have " *…transformed social practices and changed the way we do business, yet surprisingly little serious academic work has been done on them.*" Much of the research published to date has focused on the technology, such as networking and communications, bandwidth and location management. Some research on usability and mobile devices has focused primarily on the limitations of the technology and the size of the display (Chan et al., 2002). Palen at al (2000) suggest in their study on mobile phone use that "*During this time of rapid adoption [of mobile telephony] we are experiencing societal growing pains. Some people are highly enthusiastic about mobile telephony's possibilities, while others still question its discretionary and social benefits*." [p 210]

However, a study of user acceptance of the mobile Internet in Korea (Lee et al., 2002) reported that although 80% of mobile phone users in Korea subscribe to mobile internet services only 20% of those subscribers actually use the mobile Internet.

Recent research by Carlsson et al (2004) addressing the motivation to change mobile technology suggests in a study of Finnish consumers that technology is outpacing consumers' willingness to upgrade to new devices. That is, *"... [o]ur data indicates that the majority of Finnish mobile phone users still do not have the appropriate phones to use advanced mobile services ... [such as] ... WAP, GPRS and MMS ..." and that " ...consumers apparently are not willing to part with their old phones ... and value-added in terms of technology is not necessarily value-added in terms of service. [p 12]"*

Jarvenpaa et al (2003) suggest that consumers need strong incentives to adopt new mobile services *"... consumers may lack a compelling motivation to adopt new for-pay service offerings unless they create new choices where mobility really matters [p 44]"*. These authors further suggest that users are attracted to mobile technologies that provide "freedom of choice" (Marchegiani, 2004).

Grantham and Tsekouras (2002) analyse the claims that mobile technology is different from other information and communication technologies (ICTs) in that it is transformative and has become integral to the lives of users. They suggest that political and societal questions such as *" ... involuntary exclusion, surveillance, privacy, accountability and the need for openness [p 1]*" need to be considered.

To successfully evaluate mobile commerce, Dempsey and Donnelly (2002) suggest that the advantages (always available, availability of aggregated and location-aware services) must be measured against the disadvantages (small screens and high costs) leading to a deduction that "*M-Commerce will be at least as successful as e-commerce if the underlying technology and standards accentuate the advantages and minimise the disadvantages*. [p 11]" while (van der Heijden & Valiente, 2002) suggest *"...the benefits of mobile technology are hard to quantify in isolation, and that the unit of analysis to identify value should be the business process [p 1145]."*

Research addressing these issues is needed to understand motivations and trends in these new business environments for mobile technologies.

3. RESEARCH APPROACH

The research approach used for this project was a case study method based on a survey. This study was an "exploratory" study as defined by Yin (1994) *"... in which the topic is the subject of an "exploration" ... [and] ... should state a purpose*". The purpose of this study was to gain a better understanding of the main uses of mobile devices in day-to-day business activities and in particular the motivations, usage patterns and perceived advantages and disadvantages of owning and using mobile devices.

The data collection method was based on a questionnaire which was used to collect data which could be analysed in the form of "descriptive statistics" (Cavana *et al.*, 2001, p 405) which " *... describe a set of factors in a situation ... [and] ... are provided by frequencies, measures of central tendencies and dispersion*." The questionnaire also collected qualitative data. This initial data will provide preliminary contextual information for a larger study of the use of mobile devices in the workplace.

The pilot study used a population of 52 MBA and final year undergraduate students studying at a major university. It was considered that a higher proportion of these students would be in employment than average undergraduates. There was an explanatory section on the questionnaire itself and two of the researchers explained the reasons for the questionnaire and the voluntary and anonymous nature of the questionnaire.

A major advantage of face-to-face invitations to complete questionnaires is the high response rate (in this case almost 100%) although the sample can only be representative of MBA and final year undergraduate students or possibly mature age University students. The external validity in terms of the general population is relatively low but for a pilot study the findings can be argued to be useful.

The questionnaire collected both quantitative and qualitative data. It was divided into three major sections and was timed to take 20-25 minutes to complete. Question types ranged through yes/no, tick a box, Lickert scales, and comments.

4. FINDINGS

Demographics and Context Information

Findings regarding age, gender, industry, and job title data were as expected for the population. Seventy-nine percent of respondents were between the age of 20 and 40 years. Forty-six percent were female and sixty-three percent were employed part or full-time. Twenty-four percent of respondents' job titles contained the word "Manager". The most common industries in which respondents worked were IT, retail and manufacturing which together accounted for thirty-eight percent. Perceived expertise with mobile devices was high with sixty-five percent of respondents saying they were experienced or very experienced.

Motivation for using or owning a mobile device

Seventy-five percent of respondents owned or used a mobile phone. Seventeen percent owned a PDA. Fourteen percent of mobile phones and PDAs were company supplied. Fourteen percent of respondents reported their reason for owning or using a mobile phone was the low price. No respondent reported buying a PDA because of its low price. The main reason given for owning a mobile device was "I felt it was necessary for my lifestyle" (50% – phone; 57% – PDA). For 16% of mobile phone users and 21% of PDA users "I was interested in mobile technology" was given as the reason. See Figure 1.

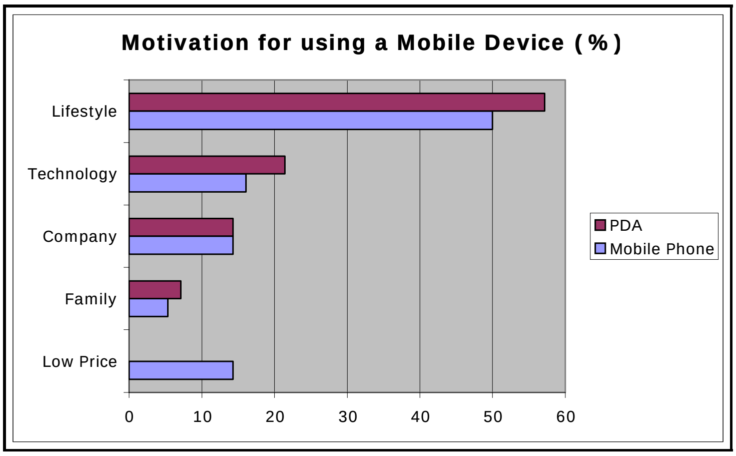


Figure 1 - Motivation for owning or using a mobile device

Usage Patterns

Mobile phones were used nearly twice as much for personal communications (48%) and tasks than for business tasks (28%) while PDAs were used as much for business as personal tasks. There was an open-ended question on specific uses posed as "*List 3 main uses for each device. (eg talk to friends, email friends, look up address book, monitor financial services, monitor sports results, surf the web, do tutorial exercises, play games, keep appointments etc)*". This resulted in 100 nominations across twelve categories for mobile phones and 20 nominations across 5 categories for PDAs (see Figure 2). The specific uses cited for PDAs were organising appointments (10 nominations), looking up addresses (5 nominations), email or internet browsing (3 nominations), playing games (1 nomination) and alarm clock (1 nomination). The main uses of mobile phones were contacting friends (38 nominations), looking up phone numbers or addresses (19 nominations), and messaging (9 nominations).

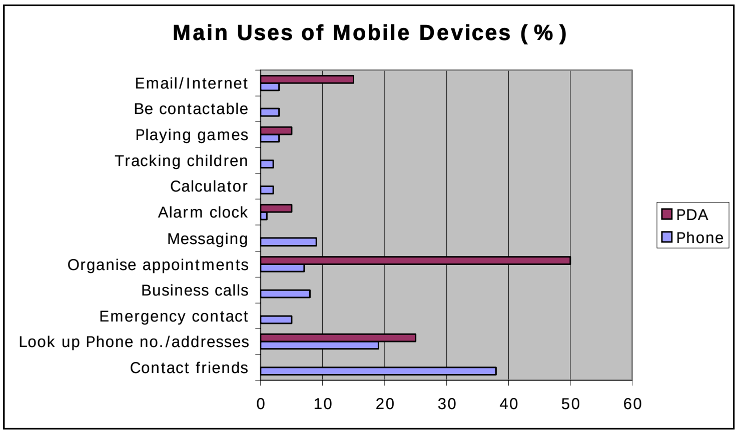


Figure 2 Main Uses of Mobile Devices

Main Problems with Mobile Devices

A quarter of respondents (22% – phone; 25% – PDA) explicitly reported no problems as distinct from making no response at all. Common problems with both phones and PDAs were either technology related: limited functionality, short battery life, and lack of memory; or usability related: small keypad, small screen, poor backlighting, and difficulty in getting to know the capabilities of the device. Phone specific problems emphasised network issues: bad connections, line drop out and busy networks. (see Figure 3). Table 1 presents the findings together with some illustrative comments given by respondents.

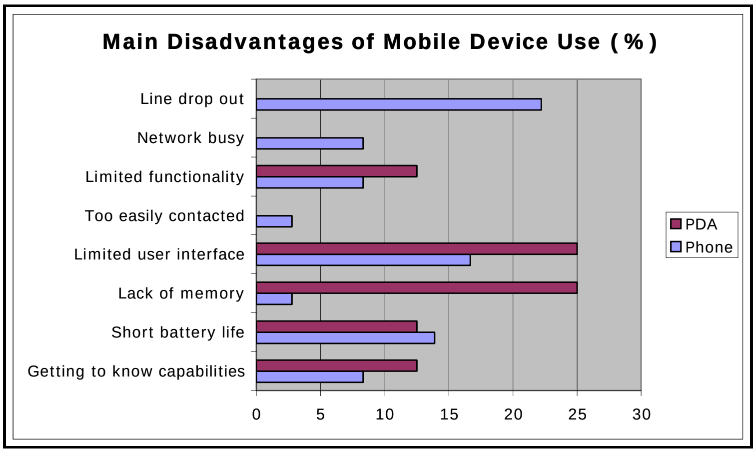


Figure 3 Main Disadvantages of Mobile Device Use

Issue Type	Specific Issue	Phon	PDA (‘)	Sample Comments
Network-related	Network busy	8	0	*“I get too many “network busy” messages.”*
	Line drop out	22	0	*“I don’t like it when the line drops out all the time and the screen goes blank.”*
Process-related	Too easily contacted	3	0	*(see general comments below)*
Technology-related	Limited functionality	8	13	*“My PDA is too basic, it needs more multimedia functionality”* *“My phone is too big and has only limited functionality”* *“My PDA is easy to use and quite effective but it needs to have greater functionality – the addition of more advanced audio and video features.”* *“I don’t have colour and only a limited text display.”*
	Short battery life	14	13	*“Main problem is battery time – if a heavy use day, by 5 pm it is going flat”*
	Lack of memory	3	25	*“My phone does not have enough memory”*
Usability-related	Getting to know capabilities of the device	8	13	*“I am still learning to use the technology and trying to find the best way to do things.”*
	Limited user interface (including poor backlighting, small keypad, small screen and input methods)	11	13	*“I don’t like typing using the stylus (PDA)”* *“I don’t like the small keypad on my phone”* *“SMS is quite tedious. It is very hard to input text especially entering URLs” (WAP phone user).* *“The main problem I see is the input of text and the information into the device as well as the amount of information that can be displayed at any one time on the screen. There needs to be a move towards a more effective means of inputting text on the device. For mobile phones perhaps voice control and use of voice for entering information”*

Table 1 Network, usability and technology issues

Perceived Advantages of Mobile Devices

The terms most commonly used by respondents to describe the advantages of using mobile devices were “portable”, “convenient” and “handy” rather than specific tasks such as looking up phone numbers or using the Internet. Not surprisingly, the main perceived advantage of mobile phones was being “contactable anywhere, any time”, although this contrasts with some of the comments made by some respondents elsewhere in the survey where this was seen as a disadvantage. The other main perceived advantage of PDAs was storing information (see Figure 4).

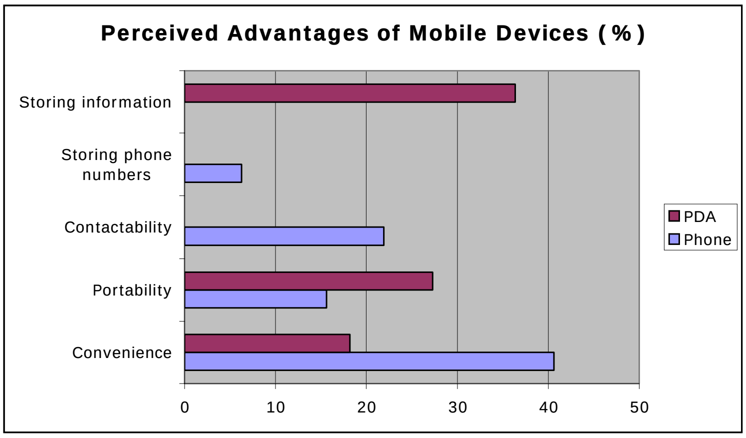


Figure 4 Perceived Advantages of Mobile Devices

Table 2 presents the findings together with some illustrative comments given by respondents.

Issue Type	Phone (%)	PDA (%)	Sample Comments
Contactability	22	0	*"I like the freedom. I can be in contact all the time."* *"It raises my level of contactability."* *"I can be contacted whilst out of the office or at home."* *"Convenient. Contactable – stay in touch."* *"I can be contacted anytime, anywhere – especially in an emergency."*
Portability	16	25	*"All my data is portable. It never leaves me." (PDA)* *"I like the ease of use and portability."(PDA)* *"I can use it anywhere."* *"I am able to call or access people at any time" (phone)* *"The main advantage is the portability" (phone)* *"I like the mobility" (phone).*
Convenience	28	17	*"Quicker, more convenient, can do it on the run (in taxis for example)" (phone).* *"Using the device for keeping appointments and contacts is very convenient." (PDA)* *"Alarms, reminders, notes all help with remembering tasks." (PDA)* *"Convenience!"* *I like my WAP phone because I can check my email.*
Storing information	0	33	*"Keeping my calendar with me allows quick access without logging into my laptop. Easy access (quickly) when I am off site or interstate." (PDA)* *"I can back up my data and store large amounts of information."*
Storing phone numbers	6	0	*"I can store 500 names/contacts including name, address, email, fax, mobiles etc."*

Table 2 Convenience, contactability and portability aspects that users find advantageous.

General Comments

Respondents were given a large blank space at the end of the questionnaire to provide general comments on using mobile devices. Those respondents who chose to comment were either very enthusiastic or very unenthusiastic.

Enthusiastic response included:

"Useful tool"

"Love it! Makes my life easy!"

"Way to go!"

"Very convenient. Allows you to do things in times/spaces which would otherwise be wasted time (eg at airport waiting for flights, taxis)"

"A great way for communicating and allowing access to data."

Unenthusiastic responses included:

"The costs are too expensive. SMS should be free. I am also concerned about health related matters for long hours of use."

"Security concerns are an issue for me. I think personalised software is important – instead of the set standard menu. I should be able to choose the options when I buy the phone."

"I don't have any mobile device (personal or work) and I have no interest in having one. I don't feel the need to be constantly in touch, playing with technology (enough of that at work)"

"I do not own or use a mobile phone"

"I don't like them and I refuse to use them as people at work can contact me. Interestingly work people don't feel comfortable to ring my home number as it is a bit 'intrusive'."

"Although security is meant to be of a suitable standard, I still believe and have heard of ways to break security."

A major issue that emerged for mobile phones was that of contactability. Several negative comments in the general comments section of the questionnaire were specifically about the intrusive nature of being contactable all the time especially in a business or work context. When comparing these comments with the specific comments about contactability being an advantage the difference seems to be along the lines of being contactable as a private person is desirable whereas being contactable by employers or work colleagues is not. This issue will be explored in future research.

5. DISCUSSION & IMPLICATIONS

This study provided three main findings which will be explored further in subsequent research. The main findings within the identified limitations are:

A major motivation for mobile device ownership/use is lifestyle choice

The findings indicate that ownership is a lifestyle choice and not imposed by work environments. This confirms the concept that mobile technology has become integral part of people's lives as suggested by Grantham and Tsekouras (2002). Private mobile phone ownership was high although a significant proportion of mobile phone users have been supplied with company-owned phones. PDAs were not perceived as low cost items but were mainly used as business tools replacing standard business diaries. Comments from respondents emphasised personal social advantages which were described using words like "freedom" and "convenience". This reinforces the notion of social integration and ubiquitousness of mobile devices (Grantham & Tsekouras, 2002; Katz & Aakhus, 2002; Keen & Mackintosh, 2001; Palen et al., 2000).

The concept of "ease-of-use" within technological limitations needs to be strongly addressed

Respondents were generally happy with the freedom mobile devices offered them and the ease-of-use of the devices though many respondents would like more functionality. Major complaints were in the usability area and in particular in the interface - small screens and poor backlighting, small keypads and difficulties in entering text. These findings suggest the need to address the importance of ease-of-use issues in the uptake, motivations and usage of mobile devices (Dahlberg & Mallat, 2002; Kalakota & Robinson, 2002) and that benefits for business processes and universal acceptance require solutions to technology-related limitations (Dahlberg & Mallat, 2002; van der Heijden & Valiente, 2002).

The concept of "contactability" needs to be qualified by the concept of "by choice"

Most general comments related to mobile phones, contactability and lifestyle. Respondents were either vehemently against or enthusiastically for the use of the devices and the impact the devices had on their lives. This reinforces Palen at al's (2000) findings described in section 2 above. Being "always connected" or "always available" in the technical sense is part of the ubiquitous and pervasive computing philosophy (Weiser et al., 1999) where technology is supposed to recede into the background of people's lives and activities, where computing technology is a part of life and is embedded in the environments in which human beings live and work. According to the ubiquitous philosophy people operating in ubiquitous environments need "invisible", unobtrusive, computing and computing devices which are constantly connected, continually available, portable and intuitive. The responses from participants in this study indicate that most people like to be "always contactable" as long as they have the choice in whether to respond to a request for attention. The social implications of these "always available" environments still needs to be researched and understood especially in a context of usability.

6. CONCLUSION

This paper described the findings from a study of mobile devices focussed on usage patterns and motivations for using mobile devices. Findings indicated that motivations are heavily influenced by social integration and lifestyle choice rather than imposition from the workplace. Findings also indicated that usability issues need to be addressed for mobile devices to be integrated into business as more than simple communications devices.

Mobile devices are now perceived as an essential component of the next generation of automated enterprise workflows and business processes. Adding mobility to a business process environment calls for a re-examination of the business system itself. The pilot study reported in this paper is part of a larger project whose goals include understanding and improving mobile work environments. The data gathered by this pilot study therefore, by describing usage patterns assists in gaining a preliminary understanding of changes in the ways people do business. Future research will focus on the emerging and evolving area of mobility-based business and associated mobile business models.

In the longer term the data collected in this pilot study and subsequent surveys will be used in the design of larger case study projects in specific industries (current candidates are manufacturing, wholesale distribution and health care) thus enabling a taxonomy of mobile commerce business models to be formulated which can be used to assist organizations in improving business processes associated with mobile commerce.

7. REFERENCES

Carlsson, C., Hyvonen, K., Repo, & Walden, P. (2004). *It's all about my phone! Use of mobile services in two finnish consumer samples.* Paper presented at the Seventeenth Bled eCommerce Conference, Bled, Slovenia, June 21-23, 2004.

Carroll, J., Howard, S., Peck, J., & Murphy, J. (2003). From adoption to use: The process of appropriating a mobile phone. *Australian Journal of Information Systems, 10*(2), 38-48.

Cavana, R. Y., Delahaye, B. L., & Sekaran, U. (2001). *Applied business research: Qualitative and quantitative methods*. Milton, Queensland: John Wiley & Sons.

Chan, S., Fang, X., Brzezinski, J., Zhou, Y., Xu, S., & Lam, J. (2002). Usability for mobile commerce across multiple form factors. *Journal of Electronic Commerce Research, 3*(3), 187-199.

Dahlberg, T., & Mallat, N. (2002). *Mobile payment service development - managerial implications of consumer value perceptions.* Paper presented at the Tenth European Conference on Information Systems, Gdansk, Poland.

Dempsey, S., & Donnelly, W. (2002). *Identifying the building blocks of mobile commerce.* Paper presented at the First International Conference on Mobile Business, Athens, Greece.

GMS-Group. (2002). The world mobile phone market. Retrieved October, 2002, from http://www.gmsgroup.net/statistics/9.htm

Grantham, A., & Tsekouras, G. (2002). *The global information age: Wireless technology's transformative potential in knowledge-based society.* Paper presented at the First International Conference on Mobile Business, Athens, Greece.

Guizani, M., & Aghvami, H. (2002). Profile of mobile and wireless computing. *Wireless Communications and Mobile Computing Journal, 2*(3), 319-338.

IDC. (2000). Internet and ebusiness has impacted australian businesses. Retrieved May 2003, 2003, from http://www.idcresearch.co.nz/Newsletters/eBusiness_news/ebt20011024.htm

Jarvenpaa, S. L., Lang, K. R., Takeda, Y., & Tuunainen, V. K. (2003). Mobile commerce at the crossroads. *Communications of the ACM, 46*(12), 41-44.

Kalakota, R., & Robinson, M. (2002). *M-business: The race to mobility*. New York: McGraw-Hill.

Katz, J. E., & Aakhus, M. (Eds.). (2002). *Perpetual contact: Mobile communication, private talk, public performance*: Cambridge University Press, Cambridge.

Keen, P. G. W., & Mackintosh, R. (2001). *The freedom economy: Gaining the m-commerce edge in the era of the wireless internet*. New York: Osborne/McGraw-Hill.

Lee, W. J., Kim, T. U., & Chung, J.-Y. (2002). *User acceptance of the mobile internet.* Paper presented at the First International Conference on Mobile Business, Athens, Greece.

Marchegiani, L. (2004). Mobile telecommunications at the dawn of a new era. How new mobile technologies drive innovative business models, *Social Science Research Network - Working Paper Series*: Social Science Electronic Publishing.

Palen, L., Salzman, M., & Youngs, E. (2000). *Going wireless: Behaviour and practice of new mobile phone users.* Paper presented at the ACM Conference on Computer Supported Cooperative Work (CSCW 2000), Philadelphia PA.

Urbaczewski, A. J., Valacich, J. S., & Jessup, L. M. (2003). Mobile commerce opportunities and challenges. *Communications of the ACM, 46*(12), 30-32.

van der Heijden, H., & Valiente, P. (2002). *The value of mobility for business process performance: Evidence from sweden and the netherlands.* Paper presented at the Tenth European Conference on Information Systems, Gdansk, Poland.

Weiser, M., Gold, R., & Brown, J. S. (1999). The origins of ubiquitous computing research at parc in the late 1980s. *IBM Systems Journal, 38*(4), 693-696.

Yin, R. K. (1994). *Case study research: Design and methods* (2 ed. Vol. 5). Thousand Oaks, CA: Sage Publications Inc.

An Overview Of Mobile Banking Adoption Among Urban Community

Ainin Sulaiman , Noor Ismawati Jaafar, Suhana Mohezar
University of Malaya, Kuala Lumpur.
ainins@um.edu.my , isma_jaafar@um.edu.my , suhanamohezar@yahoo.com

Abstract

Technological advances have changed how financial services are transacted, with mobile banking being the latest development in this domain. Today, many bank customers can carry out transactions such as these via their mobile devices: checking their account balances and transaction histories, transferring funds, paying bills, trading stocks and obtaining stock quotes, managing their portfolio, and ordering insurance. This paper focuses on the adoption of mobile banking services by consumers, by viewing this development as a service delivery innovation. The study employed Rogers' diffusion of innovations model to understand the consumers' behavior and motivation towards this innovation. The personal characteristics of mobile banking users, including their personal characteristics and level of innovativeness, were found to be important determinants of their adoption decisions. This finding provides financial services industry with a better understanding of customer perceptions of mobile banking services, and helps them plan their marketing strategies and promotion approaches for mobile banking services in the future.

Keywords

Mobile banking, financial services, consumer behavior, diffusion innovation, mobile commerce

INTRODUCTION

The development of new technologies has profoundly changed the way in which customers interact with service providers. Traditionally, the distribution of services in the retail banking industry largely meant customers having to visit a physical branch ("bricks and mortar") to access any financial services (Stewart, 2002). Today, the situation has changed. Technology has become an increasingly vital element in the competitive landscape of the financial service industry. It has changed the very nature of selling and buying financial services (Suoranta and Matilla, 2003).

Mobile banking, where customers access bank services remotely by using mobile devices with wireless connectivity, is one of the latest service delivery modes available from banks. Customers can conveniently check their account balances and transaction histories, transfer funds, pay bills, trade stocks and obtain stock quotes, manage their investment portfolio, and order insurance by using mobile devices (Durkin, Howcroft, O' Donnel, Quinn, 2003). However, even though the technology has been available for some time, financial institutions in most developed countries have only recently begun to offer mobile banking to their customers, and where it has been available, it's usage has been fairly limited.

This paper is structured around Rogers' research on the diffusion of innovations (1995). Rogers (1995) recognized five categories of consumers who differed in terms of innovativeness, which influenced the adoption of technological innovations. These categories were: innovators, early adopters, the early majority, the late majority and laggards. This paper will applies Rogers' work into the domain of mobile banking, by seeking to identify the typical characteristics of mobile banking users, including their personal innovativeness, compared to non-mobile banking users. The objective of this research is to identify the characteristics of typical mobile banking users and to determine the level of innovativeness of the mobile banking adopters compared to non-mobile banking adopters. In addition, this study also attempts to determine the factors that influence the adoption of mobile banking among the urban community using diffusion innovation theory. The paper begins by reviewing the relevant literature to provide the theoretical background of the study. Thereafter, the methodology is discussed, followed by a description of the findings. The paper concludes with a discussion of the implication of the findings and identifies areas for future research.

LITERATURE REVIEW

The theoretical framework for this study is the research on innovation diffusion by Rogers (1995). Rogers (1995) defines diffusion as the adoption of an innovation 'over time' by the given social system as a consequences of diffusion processes, which result in the acceptance or penetration of a new idea, behavior, or physical innovation. Prior studies have traditionally analyzed consumers' adoption of innovation using Rogers' (1995) five categories of adopters; innovators, early adopters, early majority, late majority and laggards.

Personal Innovativeness and Innovation Diffusion

Personal innovativeness is found to be best predictors of technology adoption which is often used in previous research (Agarwal and Prasad, 1998; Locket and Litler, 1997). Personal innovativeness embodies the risk-taking propensity that is higher in certain individuals. In innovation diffusion research, it has long been recognized that highly innovative individuals are active information-seekers of new ideas. They are able to cope with high levels of uncertainty, and develop more positive intentions towards acceptance (Rogers, 1995). Agarwal and Prasad (1998) defined personal innovativeness in the domain of information technology as "the willingness of an individual to try out any new information technology". They postulated that individuals with a higher level of innovativeness with respect to information technology would be expected to develop more positive perceptions about an innovation in terms of its advantages, ease of use, and compatibility. Hence, they would have a greater intention to adopt new technology. Different categories of adopters may differ in terms of their preferences for varying information sources, as well as their propensity to rely on information provided by marketing departments. In this instance, individuals will vary in the amount of trust they place on information provided by marketing departments, independent third parties and personal sources.
The distinction between Rogers' five categories of adopters is based on innovativeness, and suggests that new products and services should be targeted at innovators, who start the diffusion process by communicating to other adopter segments (Black, Lockett, Winklofer, Ennew, 2001).

''Innovators' are the first adopters (Saaksjarvi, 2003), and they tend to be interested in the technology itself, possessing positive attitudes towards technology (Mohr, 2001). Innovators tend to be heavier users of professional communication sources such as sales people and governments. These consumers are seen as leaders and technology pioneers. They recognize the benefits of new technology earlier than others, adopt it, and communicate these benefits to other adopter segments. They are willing to test and reduce errors in innovative products just to get access to the latest technologies. According to Parasuraman and Colby (2001), access to the new technologies provides innovators with mental stimulation. Rogers (1995) stated that the prerequisites of innovators include having substantial financial resources to absorb possible losses from an unprofitable innovation, the ability to understand and apply complex technical knowledge, and the ability to cope with a high degree of uncertainty. An innovator's position makes them more exposed to technological innovations than other segments of adopters. Exposure is important to innovators who are interested in technology for its own sake and enjoy examining technological innovations. According to Saaksjarvi (2003), innovators possess a greater degree of innovativeness, followed by early adopters, the early majority and laggards. Hence, innovativeness is closely related to consumers' willingness to learn about new product, and have positive attitudes towards new products, which serves as a basis for adopting the technology.

Demographic Profile and Innovation Diffusion

In terms of the effects of demographics on innovation diffusion, Venkatesh and Morris (2000) recently investigated gender differences in the context of individual adoption and sustained usage of. They found gender to be an important determinant of technology adoption and usage. According to the Target Group Index (Target Group Index Europe Survey, 2000), there were markedly more men in Germany's market, which had the lowest level of mobile phone ownership in Europe (60% men vs. 40% women). In the same vein, a recent survey conducted in the UK found that men spend more time on mobile phones than women (NST, 2002). In contrast, Wan et al. (2005) found that gender has least useful dimension for market segmentations as there was only light differences between males and females in internet banking adoption.

Besides gender, Venkatesh and Morris (2000) also suggested that gaining a better understanding of age differences is important, particularly as it relates to user acceptance and usage of new information technologies. Early adopters of new products are commonly thought to be young in most technology markets. According to Polatoglu and Ekin (2001), demographic factors that describe electronic banking services adopters include young, affluent and highly educated. Similarly, a Finnish study (Matilla et al., 2003) reported that the Internet banking user is middle aged, relatively wealthy and highly educated.

RESEARCH METHODOLOGY

A research model is developed to show the relationship between the independent variables and the dependent variable. With reference to the literature review (Rogers, 1995; Venkatesh and Morris, 2000; Polatoglu and Ekin, 2001), the following research model was developed (Figure 1).

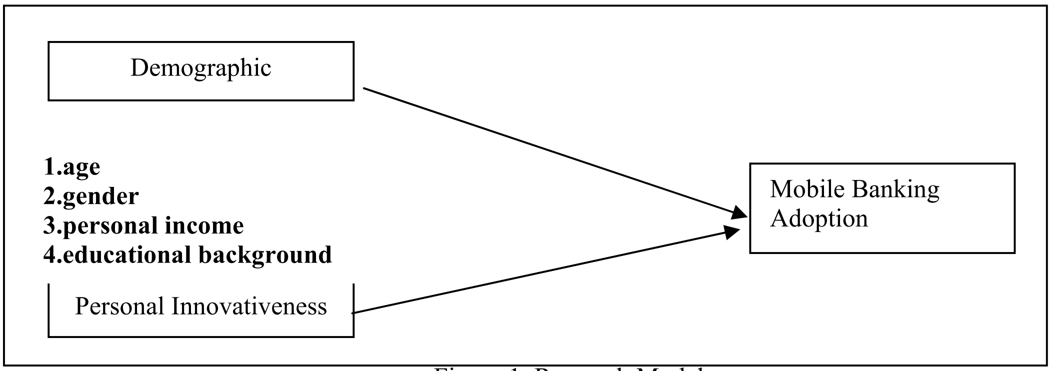


Figure 1. Research Model

Primary data was collected by using a questionnaire with two screening question; "do they have a bank account" and "do they use the Internet". The questionnaire was divided into 2 sections. Section A focused on collecting the respondent's demographic details, such as gender, age, and personal income. Section B was used to determine the respondents' characteristics and to determine the level of innovativeness. The items used to determine the level of innovativeness of respondents were adapted from Rogers (1995); Locket and Littler (1997). A 5-point Likert scale was used for all items. Respondents indicated their level of agreement with carefully constructed statements that ranged from positive to very negative toward the attitudinal object.

In the interests of expediency and practicality, a convenience sampling method was used to obtain respondents. The limitation of this method is that that we are unable to ascertain whether the information collected is representative of the population as a whole. We will bear this limitation in mind when we report the results. The target respondents were working adults, aged between 18 to 55 years old, who were both bank customers as well as Internet users. The questionnaire was self-administered and was distributed to 300 randomly selected respondents located around the Klang Valley, an area within the capital city of Malaysia, Kuala Lumpur. This area is chosen as the area has the highest Internet penetration as well as Internet Users in Malaysia (Eight Malaysia Plan 2001-2005). The drop-off method was used, that is, the questionnaires were distributed and collected after half an hour. Two hundred seventy nine complete responses were received.

DATA ANALYSIS AND RESULTS

Profile of Respondents

Approximately fifty four percent of the respondents were male, while 45.9% of the respondents are females. Previous research has shown that males outnumbered females in Internet usage. Hence the results of this study can be used to make generalization. The respondents can be classified into 3 different groups based on their educational background. 57% of the respondents currently possess a degree, 18.6 % possess a master degree, while 24.4% have a secondary school education. Out of the 279 respondents, 1.1% earn below RM1,500, 46.2% earn between RM1,500 to RM3,000, 42.3% earn between 3000 to RM5,000, and 10.4% earn between RM5,000 to RM7,000. The respondents can be classified into 3 segments based on their age. Those between 21 and 30 years old formed the highest percentage (50.9%), while respondents between 41 and 50 years old were the fewest (6.5%). 42.7% of the respondents were between 31 and 40 years old.

Characteristics of Mobile Banking Users

The characteristics of mobile banking users were identified by examining the relationship between the respondents' demographic profiles and whether they adopted mobile banking. The results of the study are elaborated in the following paragraph.. Out of 279 respondents, 149 (53.4%) have adopted mobile banking.

The relationship between the mobile banking adoption and gender is shown in Table 3. While 69.5% of males use mobile banking, in comparison, only 34.4% of the females did so. This is consistent with the previous research, which found males are more likely to adopt technological innovation than females (Venkatesh, Morris, 2000).

Table 1. Relationship between Mobile Banking Adoption and Gender

Gender	Adopt Mobile Banking		Do not Adopt Mobile Banking		Total	
	Frequency	%	Frequency	%	Frequency	%
Male	105	69.5	46	30.5	151	100
Female	44	34.4	84	65.6	128	100

Mobile banking is most popular among respondents from the 21 to 30 years old group (Table 2). In contrast, only 5.6% of the respondents between 41 and 50 years old adopt mobile banking. This shows that mobile banking is more popular among younger consumers.

Education Level	Adopt mobile banking		Do not adopt mobile banking		Total	
	Frequency	%	Frequency	%	Frequency	%
Secondary and Below	2	2.9	66	97.1	68	100
First Degree	108	67.9	51	32.1	159	100
Master Degree	39	75.0	13	25	52	100

Table 2: Relationship between Mobile Banking Adoption and Age

Age Category (years)	Adopt mobile banking		Do not adopt mobile banking		Total	
	Frequency	%	Frequency	%	Frequency	%
21-30	88	62	54	38	142	100
31-40	60	50.4	59	49.6	119	100
41-50	1	5.6	17	94.4	18	100

Table 3 represents the relationship between mobile banking adoption and education level. Only 2.9% of the respondents, who have secondary education and below, adopted mobile banking. In comparison, 75% of the respondents who possessed master's degree used mobile banking. The result shows that mobile banking adopters are typically customers who have high educational background. This is consistent with the previous research by Polatoglu and Ekin (2001).

Table 3: Relationship between Mobile Banking Adoption and Education Level

Table 4 displays the relationship between income level and mobile banking adoption. None of the respondents who earned below RM1,500 adopted mobile banking, while 89.7% of the respondents who earned between RM5,000 to RM7,000 did so. Hence, mobile banking adoption is prevalent among high-income earners, consistent with Matilla (2001).

Table 4: Relationship between Mobile Banking Adoption and Income Level

Income (RM)	Adopt mobile banking		Do not adopt mobile banking		Total	
	Frequency	%	Frequency	%	Frequency	%
Below 1,500	0	0	3	100	3	100
1500-3000	45	34.9	84	65.1	129	100
3000-5000	78	66.1	40	33.9	118	100
5000-7000	26	89.7	3	10.3	29	100

The relationship (Table 5) between the adoption of mobile banking and demographic profile (gender, age, education level and income level) was tested using the Pearson's Chi-square test. All four variables were found to be significantly associated with mobile banking. This implies that there are higher levels of adoption among male, younger people, those with higher income and those who have a higher level of education.

Table 5: Relationship between Mobile Banking Adopters and Demographic Variables

Relationship	Value	Asymp sig.	Significant
Mobile banking and gender	34.419	.000	Yes
Mobile banking and age	21.176	.000	Yes
Mobile banking and education level	92.806	.000	Yes
Mobile banking and income level	44.180	.000	Yes

Personal Innovativeness and Mobile Banking Adoption

Sixteen items were used to measure the level of personal innovativeness among mobile banking adopters. It was observed that mobile banking adopters were more eager to try new ideas, compared to non-mobile banking adopters (Table 6). Mobile banking users tend to be the majority (M=3.99) of the group who firstly try new products in the community compared to non-users (M=2.65) (item 6). This implies that the mobile banking adopters tend to be people who have pioneering characteristics. Table 8 also illustrates that mobile banking adopters have a favorable attitudes towards change, as shown by the higher mean scores (M=4.12) of mobile banking users compared to non-mobile banking users (M=2.85) on the statement "I like to keep up with technological advances". Conversely, non-mobile banking adopters are unwilling to adopt new ideas, unless compelled to do so. This is illustrated by their higher scores on items 12, 15 and 16. Mobile banking adopters have a greater exposure to the mass media and often seek out information about innovations. They use commercial media and professional sources more extensively (item 10), compared to non-mobile banking users. Interestingly, mobile banking adopters (M=4.15) had higher scores on social participation (item 11) than non-mobile banking adopters (M=2.9). In addition, the average total score of personal innovativeness for mobile banking users were slightly higher (M=66.1) than that for non-mobile banking users (M=49.5) (Table 9). This implies that people with high level of innovativeness tend to adopt mobile banking services. This finding is consistent with Agarwal and Prasad (1998).

Table 6: Relationship between Mobile Banking Adoption and Personal Innovativeness

No.	Items	Mean Scores		Mean Differences
		Adopt mobile banking	Do not adopt mobile banking	
1	I like to be considered as a leader	3.88	3.11	0.77
2	My friends and neighbors often come to me for advice about new products and innovation	3.85	2.71	1.14
3	I often seek out the advice from friends regarding new product or innovation	3.77	3.29	0.48
4	I like to buy new and different things	4.07	2.85	1.22
5	I am eager to try new ideas.	4.1	2.91	1.19
6	I am usually among the first to try new products.	3.99	2.65	1.34
7	I have more self-confidence than others	4.01	2.88	1.13
8	I want to look a little different than others	3.95	2.83	0.12
9	I like to keep up with technological advances.	4.12	2.85	1.22
10	I often make extensive use of commercial media and professional sources in learning of new products.	4.21	2.9	1.31
11	I am a socially active person.	4.15	2.94	1.21
12	I have an old-fashioned taste and habits.	3.77	4.64	0.87
13	My social status is important part of my life.	4.11	2.75	1.36
14	It's very important to me to feel I am a part of a group.	4.15	2.98	1.17
15	I am only accepting and use new products	4.0	4.49	0.49

	pressures.			
16	I am a person who is “skeptical” about new ideas.	4.04	4.48	0.44

Table 9: Total Scores of Personal Innovativeness

	Mobile banking adoption	N	Mean	Std. Deviation
Total scores of personal innovativeness	Adopt	149	66.0537	7.81785
	Do not adopt	130	49.5077	11.41823

(8) CONCLUSION AND RECOMMENDATION

Discussion and Conclusion

This paper provides new, interesting insights into the diffusion pattern of mobile banking services. The fundamental question we are asking is “why do people adopt new products?” The literature review described how demographic and psychographic variables affect the adoption of innovations, and our findings are in line with this argument. Specifically, this study found that demographic factors, such as age, gender, personal income, and educational background, do affect the adoption of mobile banking services.

Inline with previous study done by Black et al. (2001); Lockett and Littler (1997); Gilbert et al. (2003), this study also found that males are more willing to adopt new technology than females. The results could be attributed to the male’ masculine personality, which is often willing to take risks and anxious to try out new technological products. Besides this, men also read more technological magazines and have greater involvement with technological products.

Similar to prior studies, which showed that younger consumers have a greater tendency to adopt innovations, such as Internet banking, this study found that mobile banking is most commonly adopted by consumers aged between 21 to 30 years old. In this context, this is understandable since mobile banking requires the use of mobile devices, such as personal digital assistant (PDA) or 3G mobile phones, which are usually purchased by young consumers with more disposable income. The high educational background of mobile banking users is explained by the need to master the complexity of conducting financial transactions via mobile channels.

Finally, this study also suggests that personality traits do have an effect on the likelihood of the decision to adopt mobile banking. The dominant idea within the field of innovation research is that individuals can be categorized according to their innovativeness. These categories, innovators, early adopters, early majority, late majority and laggards, have different level of innovativeness, and this affects their likelihood of adopting mobile banking. Consumer innovativeness can be measured through their attitudes to change, peer influence, communication behaviors and pioneering characteristics. The more innovative the consumer is, the more likely they will adopt mobile banking.

In conclusion, this paper has identified several factors, which influence the adoption of mobile banking. The results are consistent with Rogers’ diffusion model (1995).

Suggestion for Additional Research

This study only focuses on individual differences and how individuals perceive mobile banking. Other factors such as the type of innovation decision, communication channels, and the nature of the social system are also important in determining the factors that influence the adoption of mobile banking. For example, the type of innovation decision is related to an innovation’s rate of adoption. We suggest that additional research should be conducted to examine these other factors.

Diffusion studies have yet to determine the relative contribution of each of the variables. When such enquiry is accomplished, we shall have a much more adequate basis for planning and allocating the inputs going into campaigns designed to speed up adoption.

Implications of the Research

This paper's results can help the financial services industry identify the typical users of mobile banking based on their demographics, and thus point out the right market segment to target, This knowledge affects a range of decisions, such as the amount of resources to allocate to train salespeople and the type of advertising campaigns to run.

In addition, financial services providers can also use the findings to improve the way information about mobile banking is disseminated. Different target market needs different types of advertising campaigns, and a bank's communication style should be compatible with the information processing styles of potential adopters. Disseminating information through the right mode of communication for different consumer segments is likely to increase each segment's probability of adopting mobile banking.

REFERENCES

Agarwal,R., and Prasad, J., (1998)" A conceptual and operational definition of personal innovativeness in the domain of information technology"*, information System Research,* Vol. 9, no.2, pp. 204-215

Black, N.J, Lockett, A., Winklofer, and H., Ennew, C.(1997) "The adoption of internet financial services: a qualitative study", *International Journal of retail and Distribution Management*, Vol. 29, No. 8, pp.390-398.

Bumiputra-Commerce Bank Bhd (2002), "Highlights of The Malaysian Banking and Financial Services Summit 2002, *Economic Research Services*, June.

Bandura, A., (1982)"Self efficacy mechanism in human agency*", American Psychologist*, vol 37,pp.122-147.

Culnan, M.J,"Georgetown Internet privacy policy survey: report to the Federal Trade Commission, available at www.msb.edu,

Datamonitor, (2002)"Mbanking in Europe 2000: the killer application of m-commerce", May 2002, available at www.datamonitor.com

Mitchell,V. M. (1998)"Consumer-perceived risk: conceptualizations and models", *European Journal of marketing*, Vol.33, pp163-195.

Choon, S.L, Hyung, S .S., and Dae,S.K, (2004)"A classification of mobile business models and its applications", Industrial Management and Data System, vol.104, pp.78-87

Morna, L., Mcgoldrick,P., Keeling,K., and Doherty,J,(2003) "Using ZMET to explore barriers to the adoption of 3G mobile banking services,*International Journal of Retail and Distribution Management* , vol. 31,pp. 340-348.

Cellular-news, 2000, available at www.cellular-news.com

CIT Publications ,"3G mobile in Europe: future markets- preparing for launch", CIT Publication, October,2002

Dalbhokar, P.A, (1992)"Role of affect and need for interaction in on-site service encounters",in *Advances Consumer Research*,pp.536-539,1992

Dickerson, M.D. and Gentry,J.W, (1983)"Characteristics of adopters and non adopters of home computers", *Journal of Consumer Research*, Vol. 10, No.9, pp.225-235,1983

Durkin, M., Howcroft, B., O'donnell, A., and Quinn,D.,(2003)"Retail bank customer preferences:personal and remote interactions", *International Journal of Retail and Distribution Management*, Vol 31, No. 4, pp. 177-189.

Ennew, C.(1992)"Consumer attitudes to independent financial advice", *International Journal of bank Marketing*, Vol. 10 No. 5, pp. 13-18.

Fitchard, K, “The two m’s of commerce”, Primedia Business Magazine, April 19,2004

Gatignor, H. and Robertson, T.S,(1985) ,”A prepositional inventory for new diffusion research”, *Journal of Consumer Research*, Vol. 11, No.4,pp. 849-867,1985

Haque, A. (2004),” Mobile commerce:customer perception and its prospect on business operation in Malaysia, Vol. 4, No.8,pp 257

Hamman, H.,(2001) “wireless woes”, IDG News Service, Available http:// www.computerworld.com,.

Harrison, A.W, and Rainer, P.K,(1997)” The influence of individual differences on skill in end-user computing*”, Journal of Management Information System*, Vol.9, no.1, pp 93-111.

June,L., Chun,Y., Chang,L., and James, Y (2003),”Technology acceptance model for wireless internet”, *Electronic Networking Applications and Policy*”, Vol.13, no. 3, pp.206-222

Kohli, K, (2004)”SMS in banking:mitigating the risk”,Paladion Knowledge series.

Kwon, T.H, and Zmud, R.W, (1997)“Unifying the fragmented models of information systems implementation”, *Information System Research*, Vol 14, No.2, pp317-319

Labay, D.G. and Kinnear, T. C.(1981)”Exploring the consumer decision process in the adoption of solar energy system”, *Journal of Marketing Management*, Vol.13, pp. 791-811.

Lee, D.S, (2003)”Wireless internet security “, Information System Security, Vol 11, No.3, pp 34-50

Lockett, A. and Littler, D (1997),”The adoption of direct banking services”, *Journal of Marketing Management*, Vol. 13, pp. 791-811.

Mohr, J. (2001)“*Marketing of high-technology products and innovations*”, Prentice Hall, Upper Saddle River, USA.

Matilla M, (2001)” *Essay on Customers in the Dawn of Interactive Banking*”, doctoral dissertation, Jyvaskyla Studies in Business and Economics, University of Jyvaskyla.

Matilla, M., Karjaluoto,H., Pento,T., (2003),”Internet banking adoption among mature customers:early majority or laggards?”, *Journal of Services Marketing,* Vol 17 No. 5, pp.514-528.

Nelson, D.L (2001).,”Individual adjustment to information-driven technologies: a critical review”, *MIS Quarterly*, Vol.14, No.1,pp. 79-98.

New Straits Times, (2002)”*Tech and gender factor*”, New Straits Times, November 18.

O’connell, B., (2004)” Wireless banking: news of its demise is greatly exaggerated; sure full-fledged wireless banking flopped when wells tried it in 2001.but other banks have had success with elements like text-message alerts. Next up: bill payment. *Bank Technology News*, Vol 17, No. 8, pp 43

Olshavsky, R.W. (1996),”an exploratory study of the innovation evaluation process”,*Journal of Product Innovation Management*, Vol.13, pp23-29.

Parasuraman, A., and Colby, C.L (2001)”*Techno- Ready Marketing*, The Free Press, New York.

Polatoglu, V.N., and Ekin, S. (2001)” An empirical investigation of the Turkish consumers’ acceptance of internet banking services”, *International Journal of Bank Marketing*, Vol.19, No.4, pp. 156-165.

Robertson, T.S.(1971), *Innovative Behavior and Communication*, Holt Rinehart and Winston, New York.
Rogers, E.M., *The Diffusion of Innovations*, Free Press, New York

Rugimbana, R. and Iversen(1997), ,"Perceived attributes of ATMs and their marketing implications", *International Journal of Bank Marketing*, Vol.12, No.2,pp.30-50.

Suoranta, M., and Matilla, M,(2003)"Mobile banking and consumer behavior : new insights into the diffusion pattern", *Journal of Financial Services Marketing*, Vol.8, No.4, pp354-356.

SaaksjarvI, M,(2003)" Consumer adoption of technological innovations", *European Journal of Innovation Management*, Vol. 6, No. 2, pp. 90-100

Turban, E, (2004)" *Electronic Commerce: A managerial Perspectives*", Prentice Hall, Upper Saddle River, New Jersey.

Venkatesh, V., and Morris, M.G.(2000),"Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and use behavior', *MIS Quarterly*, Vol. 24, no.1, pp115-139.

Wan, W.N., Luk, C.L, Chow,W.C., (2005) "Customers adoption of banking channels in Hong Kong,"*International Journal of Bank Marketing* , Vol. 23 No. 5, pp 255-272.

Zmud, R.W(1979)"Individual differences and MIS success: A review of the empirical literature", *Management Science*, vol.25, no. 10, pp966-979

APPENDIX

Questionnaire

Section A: Demographic Information

Check (✓) the suitable box

1. Age

[ ] 20 and below
[ ] 21-30
[ ] 31-40
[ ] 41-50
[ ] Above 50

2. Gender

[ ] Male [ ] Female

3. Marital Status

[ ] Single
[ ] Married
[ ] Divorced
[ ] Widowed

4. Education Level

[ ] Secondary School and below
[ ] Degree
[ ] Master degree
[ ] Doctoral Degree

5. Personal Income
[ ] Below RM 1,500
[ ] RM1,500 - RM 3,000
[ ] RM3,000-RM5,000
[ ] RM5,000-RM7,000
[ ] Above RM7,000

Section B: Personal Innovativeness

Check (✓) the suitable box

1= strongly disagree, 2= disagree, 3= neutral, 4= agree, 5 strongly agree.

		1	2	3	4
1	I like to be considered as a leader.	[ ]	[ ]	[ ]	[ ]
2	My friends and neighbors often come to me for advice about new products and innovation.	[ ]	[ ]	[ ]	[ ]
3	I often seek out the advice from friends regarding new product or innovation.	[ ]	[ ]	[ ]	[ ]
4	I like to buy new and different things.	[ ]	[ ]	[ ]	[ ]
5	I am eager to try new ideas.	[ ]	[ ]	[ ]	[ ]
6	I am usually among the first to try new products.	[ ]	[ ]	[ ]	[ ]
7	I have more self-confidence than others.	[ ]	[ ]	[ ]	[ ]
8	I want to look a little different than others.	[ ]	[ ]	[ ]	[ ]
9	I like to keep up with the technological advances.	[ ]	[ ]	[ ]	[ ]
10	I often make extensive use of commercial media and professional sources in learning of new products	[ ]	[ ]	[ ]	[ ]
11	I am a socially active person.	[ ]	[ ]	[ ]	[ ]
12	I have an old-fashion taste and habits.	[ ]	[ ]	[ ]	[ ]
13	My social status is important part of my life.	[ ]	[ ]	[ ]	[ ]
14	It's very important for me to feel I am a part of a group.	[ ]	[ ]	[ ]	[ ]
15	I am only accepting and new products because of economic necessity and social pressures.	[ ]	[ ]	[ ]	[ ]
16	I am a person who is "skeptical" about new ideas.	[ ]	[ ]	[ ]	[ ]

Investigating the Factors Impacting Personal Use of Mobile Technologies

Abbass Ghanbary, Dinesh Arunatileka
PhD Research Student, University of Western Sydney
ghanbary@cit.uws.edu.au
darunati@cit.uws.edu.au

Abstract

This paper investigates the factors influencing the use of mobile technologies by personal users. This investigation is different to most contemporary literature on mobile technologies which predominantly focuses on mobile applications from a business viewpoint. An outline of the research framework and a corresponding survey is discussed for individual usage of the mobile devices in various applications such as mobile shopping and mobile banking as well as its application for individuals in the work environment. This research will also focus on the user's experience of limitations of mobile technologies especially as people have increasingly turned to rely on this technology.

Keywords:

Mobile technologies, personal mobile phone use, mobile banking, mobile shopping, mobile at work.

INTRODUCTION

This paper outlines a research framework for investigation of the factors that impact the use of mobile technologies by individual users. As a part of this research, a questionnaire is designed to investigate the usage of mobile devices by individuals in activities related to shopping, banking and their regular work. The foundation of this study is based on the behaviour of the people who live and work in the modern environment that is inundated with wireless communication devices.

The aim of this paper is to describe the study, distinguishing the limitations and the applications of mobile technologies – particularly in the context of their impact on such areas like social implications and its impact on the mobile users, change management in a way that the user adapt to the new applications, customer relationship management from the customer's viewpoint and how individual users will benefit from the improved communication ability, knowledge management and the distribution of knowledge to mobile user, legal issues and cultural issues that affect all parties involved.

Technology affects individuals in the way they operate and manage their daily activities. The biggest paradigm shift in the technology area came with the introduction of personal computers. This was followed by the Internet, which raised the usage of technology to a new height. The current wave of technology transformation is in the form of mobile technologies giving anytime anywhere accessibility to people. Toffler (1980) was right when he said that not only are we undergoing change, but the rate of change is also high. In fact, we seem to have transcended Toffler's three waves (agriculture, industrialisation and information) and find ourselves in the fourth wave – the age of communication. An important part of this 'fourth wave' is the impact of mobile technologies in our day to day lives. While the mobile technology domain has been increasingly trying to address its initial technical challenges such as the reliability of connections, screen size of mobile/wireless units, limited processing resources of mobile/wireless units and the limited wireless bandwidth (Samaras, 2002), what is even more important is to recognise and address these and related challenges from an individual user's viewpoint.

As per Lacohee, et al, (2003) huge advances in technology have undoubtedly played a role in the rapid and unprecedented take up and wide spread availability of mobile communications. Perry et. al (2001) state that the rapid and accelerated move towards the mobile technologies has provided the individuals, the ability to work away from the office and on the move. Thus we can surmise that the primary areas of impact of mobility include: general mobile usage, mobile banking, mobile shopping and the use of mobile devices at work.

The users get interested in technology only when it provides them with satisfaction in terms of facilitating them to achieve their product/services from the relevant supplier. As per Sun (2003) the user context is the user sensible information related to consumer behaviour, including their locations, surroundings and timings of the available information. Therefore, the data gathered during this research will be statistically analysed based on the given demographic and the personal experiences of the users to construct a theory on how mobile technology could advance its capabilities to create better value to individuals and their personal processes. Arunatileka & Unhelkar (2003) described how an m-enabled Customer Relationship Management (CRM) would elevate the CRM into real time interactions with the customer and enhance the Business to Customer(B2C) and Business to Business (B2B) relationship into an Individual to Individual (I2I) relationship. Such individual - I2I - relationships are also the focus of this research.

To achieve our aim of research, we have classified this paper in the following sections: a) abstract b) keywords c) introduction d) the literature review on mobile technology e) discussing the proposed hypothesis f) questionnaire development g) methodology in gathering the data h) proposed analyses of the data and i) conclusions and future directions.

CURRENT MOBILE TECHNOLOGY LITERATURE

The concepts of mobile communication, digital networks and service providers have advanced very rapidly in the last few years. This has lead to consideration of mobile communications and corresponding network providers way beyond the well known data and voice transmissions and into the realm of mobile web services. Lacohee (2003) describes how the increase in the number of service providers to the market in the early 1990s served to further increase the number of subscriber and consumer popularity. With increasing users, the mobile devices such as personal digital assistants (PDAs) can also be used to access the Internet, schedule work, identify locations and organise the entire personal daily activities. In fact the statistics of last few years show that the number mobile devices have more than doubled the number of Personal computers (ITU-T estimation, 2003). Thus the usage of mobile technologies has caused the mobile devices to be extremely popular as these devices are not only used for business activities but also for socialising, personal leisure and pleasure in the technological fabric of today's society.

As stated by Unhelkar, et.al, (2005), the world is undergoing a transformation with the communication technologies and the flow of data and information had become a crucial resource for individuals for their personal activities as much as for their business activities. Furthermore, this mobile connectivity has also improved the ability of individuals to connect to business processes to exchange data and conduct transactions. This transformation of individuals and the way they make use of business processes that allow mobile connectivity has been evolutionary rather than revolutionary. For example, at the beginning of the Internet age, with the aid of its communications capabilities, businesses were transferred to e-business and individuals had the opportunity to do their daily business activities from home. Ghanbary (2003) has described the Internet as the most powerful tool that bring information to our home through communication lines like water and electricity that come by power lines and pipes. Powerful search engines and capability of sharing information are the great advantages of the Internet. This has also lead to the businesses to form clusters, collaborate and communicate with each other in order to satisfy increasingly sophisticated customer requirements. Lientz and Rea (2001) describe the business activities and technologies that help support individual processes as follows: *" The Internet, telephone systems, computer networks, electronic mail, voice mail and facsimile machines are examples of infrastructure that support business operations. You cannot ignore such factors as facilities, location, and office layout."*

Vaghijiani, et al, (2005), in discussing the phenomena of mobile technologies, explain the effect of mobile transitions on the cultural aspect of the workforce within the organization. Adapting and influencing employee attitudes toward working in the new mobile environments requires carefully thought out change management strategies. The area of influence of the Internet communication abilities (including the ability of mobile devices) include our ability to communicate, gather information, use email facilities, make payment, carry out most of our banking transactions (referred to as mobile banking) and satisfying our shopping needs (referred to as mobile shopping). Furthermore the current mobile devices could also check on government bulletins, undertake leisure activities such as games and reading material, provide telemetric services and also provide mobile based advertising.

Herzberg (2003) states that the use of mobile devices enable secure, convenient authorizations of e-banking, retail payment, brokerage and related transactions. Therefore individuals could conduct most of their banking needs from wherever they are irrespective of the time and place. Stafford (2003) states that the convergence of mobile technologies with Internet connectivity will result in GPS-linked PDA phones that will automatically accommodate our shopping needs with location dependent vendors offers and promotions. Arunatileka & Unhelkar (2003) have also discussed in detail how the application of mobility to Customer Relationship Management (CRM) can be based on the four major features of mobile connectivity: Real-time Interaction, Spot-based, Dynamic Organizational Processes and Dynamic Customer Group.

Real time interaction meant that the users of mobile devices could interact with each other in a real time fashion resulting in discussions and clarifications of issues in real time – as compared with a Personal Computer (PC) based system wherein the interactions could only be communicated via emails. Spot based implies that the customer's physical location is used to undertake promotions such as offering a discount in the next hour for customers in a certain shopping centre. Dynamic organisation processes allow individuals of an organisation to schedule their work more dynamically such as road service providers using mobile technology to strategically locating service providers in order to respond to customers quicker. Dynamic customer groups mean that the grouping of the customers is done according to their demographics at a particular point in time and location. For instance, members of a frequent flier programme could be notified that a flight is delayed on their mobile devices.

Geer (2004) state that many people now purchase books, CDs, travel and other products and services over the Internet using e-micro payments. Tatar et. al (2003) describes the usage of mobile hand held devices in order to do online learning featuring leaning anywhere, any time. Most mobile service providers facilitate such micro payments and allow the amounts to be charged to mobile bills. However this has constraints to develop further since trust relationships between service providers and individuals have to develop to extend such facility to large amounts of money. Rao & Minakakis (2003) state that the knowledge of the end user's location is used to deliver relevant, timely and engaging content and information.

Therefore, it is realised that the application of this technology is unlimited and the impact of this technology is on individual as well as the organization. Unhelkar (2005) describes this impact as a unique dimension to the globalization process and mentions that the mobility makes connectivity totally independent of space.

On the other hand, Spatz and Wieners declare that People, processes and technology—these three, in that order, are the keys to managing any successful it organization. To gain the best results in each of these key areas, a CIO should emphasize skills and competencies, services management and governance, and, as the organization matures, implement an adequate project investment portfolio. (www.cioinsight.com)

People, process and technology are contributing to the success of the mobile organisation. People play a major role in this equation as they are the fundamental dimension of the transformation evolution. Process and technology are in use when people are involved. The process and technology are supporting the mobile organisation as long as make the life easier for the people. Figure 1 represent the importance of people by being presented on the top, however the importance of the technology and process including the crucial role of the organisation and their infrastructure has also been illustrated.

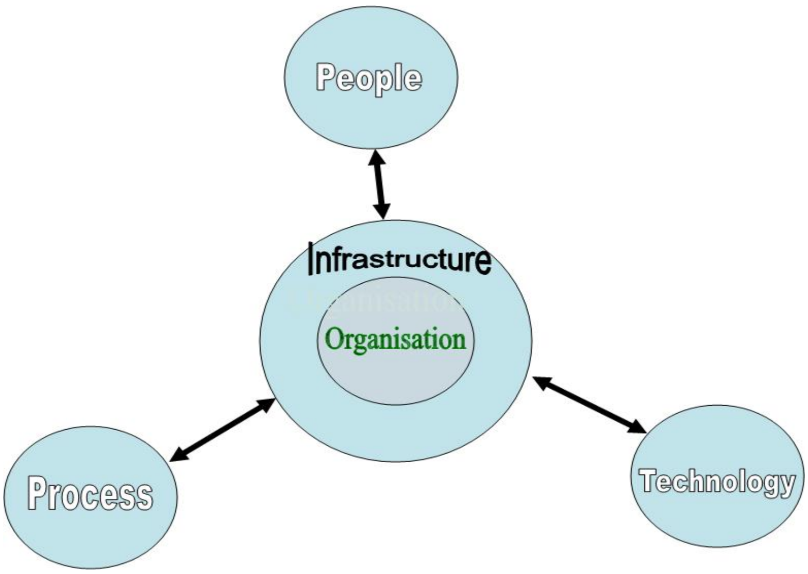


Figure 1: Relationship of Organisation, People, Processes and Technology

There is no doubt that mobility has brought critical issues that are unresolved in the technological world. Mobile technology and its related devices have raised the concern of the user under certain circumstances. Alahuhta, et al, (2005) identify the real challenge for mobile services as mostly non-technical issues such as understanding the needs of customer and whether a mobile solution responds to these requirements. In the design of our survey, we ask the respondents to identify these critical non-technical issues and provide us with their personal experience. The descriptive answers will be transferred into numeric values to gather results for the given hypothesis in the next section.

PROPOSING THE HYPOTHESIS

This study is based on the following hypothesis and questions:

1. Is the continuity of transactions on the move a major concern of the mobile users?

2. Do the users of mobile technology decide on the methods to transit already e-transformed businesses into collaborative m-enablement in order to fully capture the mobile opportunities?

3. Do the limitations of mobile technology such as security, privacy and integrity concerns, the major drawbacks for the mobile users to fully rely on this technology for their daily activities such as banking, shopping and working?

We strongly believe there are still other questions existing which we propose to explore after the statistical analyses of the data arising from the questionnaire. The development of the questionnaire is described in the following section.

Arising from the above questions, the main hypothesis of this study has been identified as following:

H1: Mobile technology has influenced the way people adapt to their banking, shopping and work related activities

QUESTIONNAIRE DEVELOPMENT

The survey will be validated and tested by collecting the sample data and placing the result in the spread sheet already designed in Microsoft Excel to test and verify our final result. The descriptive collected data will be transferred into the numeric values. The descriptive data will also be statistically analysed and they will be tested against the hypothesis of this research. The purpose of the validation is to observe the changes that can be attributed to the research in identifying the full potential of the mobile applications and the shortcomings of this technology (Internal Validity). In this study, the researcher is not targeting any specific group of people, however the people who are using a mobile device are preferred because their experience could be used (External Validity). This research would be followed by further publications once the data is collected and analysed at which point we could show construct and conclusion validity. (www.socialresearchmethods.net)

The questionnaire has been categorised into five different sections:

- Section A: General mobile usage
- Section B: Mobile Banking
- Section C: Mobile Shopping
- Section D: Mobile at work
- Section E: Demographics

Section A General mobile usage: This section classifies the participants who use mobile devices and whether they are connected to Internet using the personal device. The reasons for their purchase of a mobile device are also queried together with their perceived advantage and disadvantage of the mobile Internet.

Section B Mobile Banking: The questions asked in this section of the questionnaire give us the opportunity to identify the usage of mobile devices by the individuals in the mobile banking domain. The participants are asked to identify the advantages and disadvantages of using mobile devices in their banking activities. This will probe in to the sort of transactions that people conduct on their mobile devices. We also query on the transactions that people would want to conduct through the mobile devices. Any problems/difficulties/complaints in the use of mobile banking will also be queried. Finally, the participants will be asked to list improvements they would like to see in their mobile banking experience.

Section C Mobile Shopping: In this section the participants are asked about the usage of their mobile devices to shop on line. The participants are requested to list the perceived advantages and disadvantages of mobile shopping. They are also queried on the sort of products/services that is bought using their mobile devices and the improvements they would require to make their mobile shopping experience more satisfactory. The problems/difficulties/complaints they have in mobile shopping and the sort of products/services they would not buy using their mobile devices will also be probed. The participants will also be queried on the setup applications and facilities they would like to see, which would comfort and attract them into the mobile shopping environment.

Section D Mobile at Work: The questions on this section are concentrated on how mobile devices could improve the daily processes of persons at work. The participants will be queried, why they need mobile devices to do their given tasks and the advantages and disadvantages, using their mobile devices at work environment. Problems/difficulties/complaints experienced by using mobile devices at work will also be queried. The preference of the participants, whether to take given tasks in the office or anytime, anywhere using mobile technology will also be probed.

Section E Demographics: Participants are asked to give details about their age, level of education, level of income, occupation and their gender. The reason for this section is to identify the type of people and how the usage of the mobile device is influenced by their demographics. The sample of the questions is presented in the appendix.

Figure 3 represent the core for the submission of this paper since it is clearly identifying the selected design of the survey box presented in figure 2, and distinguish the task constructed in developing the survey.

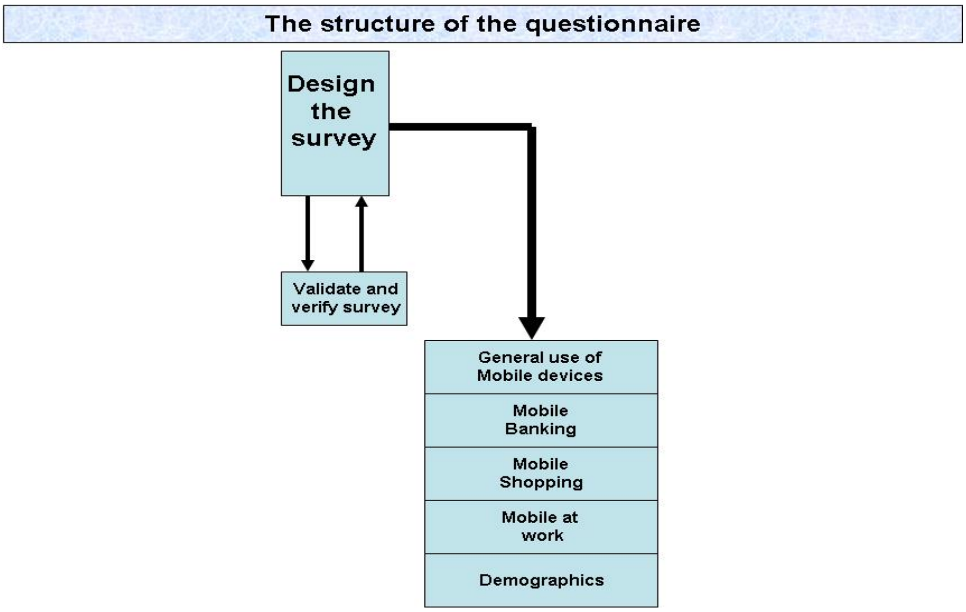


Figure 3: The Structure of the Questionnaire

The sample of the participants would be urban based users. The demographics would be varied due to the sensitivity of the data which will differ as per the demographic factors of the samples. Especially in age category and the level of education would be very important factors when investigating the usage of mobiles. The gender of the user also could have a bearing to certain usage. All these areas are tested in the questionnaire which would be analysed to see whether such demographic factors have any impact on the mobile usage.

No children will be included in the sample. The questionnaire is anonymous with straightforward questions that any adult could easily answer while all ethical and privacy issues were carefully observed. The methodology in gathering data for the questionnaire is discussed in the following section.

METHODOLOGY IN GATHERING THE DATA

The research was firstly initiated by literature review to understand the nature of the existing problems in the mobile environment. All the data will be collected by a questionnaire while the participants remain anonymous. This is a quantitative research and as per Leedy (1993) Quantitative methodologies manipulate variables and control nature phenomena. They construct hypothesis and "test" them against the hard facts and reality.

The primary reason for selecting a survey-based quantitative approach was considered the best research methodology since the study can approach the participants and gather the required variable to test them against the constructed hypothesis. The data will be analysed and processed by statistical methods for deduction and testing of the hypothesis. Figure 2 represent the path that will take place for this study to establish a robust ground into test of the hypothesis.

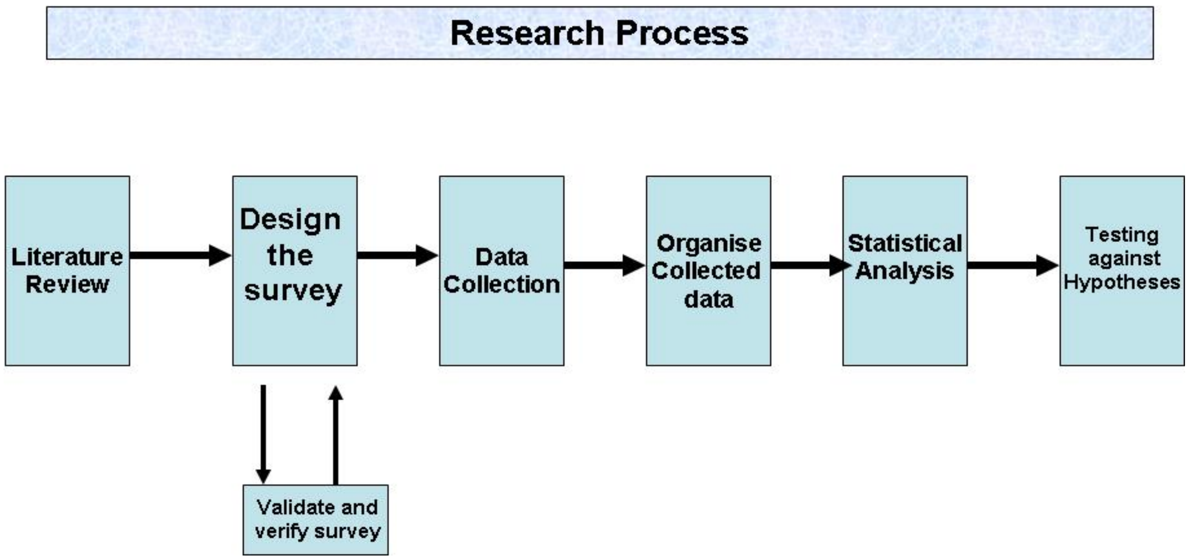


Figure 2 – Research Process

For the creation of the sample frame, first the research must identify its population that are to be studied in the survey. As mentioned earlier the research is not targeting any specific group however the mobile users are preferred for the purpose of their previous experiences. The research is also not targeting at any children.

Sample is a subset of the target population chosen so as to be representative of that population. The general aim of all sampling methods is to obtain a sample that is representative of the target population (www.deakin.edu.au). Since there is no classification of participants (Except children), the study must identify a sample size for the progress of the study. Consequently, the participants will be randomly selected from the general public.

There are different methods available in collecting and filling the questionnaire. Later when the method of data collection has to be decided, that decision will be based on the following options. However it appears that the best option for this study is the personal interview since some of the answers are descriptive and interactive which would have to be transferred to numerical values in order to analyse them statistically.

The available options in collecting the data are as follows:

1. **Postal Survey:** The Ethics Committee of UWS rejected this method, as they wanted to know where we get the participants address from and whether it would be ethical and legal. The cost could also be unaffordable for students.

2. **Telephone Interview:** After examining this method, it was realised that the same problems would apply as in a postal survey.

3. **E-mail:** This method was rejected on the assumption that respondents are already computer users, excluding those who do not use it.

4. **Questionnaire and Personal Interview:** This method was considered the most suitable for the project and was adopted.

5. **On line:** Provide an online document for the participants to answer and email it back. The advantages of this method are the low cost as well as the collected data could automatically go to the designated destination ready to be processed.

The process for data collection in this study would be very time consuming. It will take a few months to complete all the questionnaires and gather the required data. The descriptive data items will be transferred into quantified numeric values. The proposed analyses of the data are explained in the next section.

PROPOSED ANALYSES OF DATA

For the statistical analyses of the data, a spread sheet is already created using Microsoft Excel. The results will be exported to Microsoft Excel spreadsheets where detailed graphical representations would be constructed and statistical values would be calculated. Also, it has been decided to change all the descriptive answers to the numeric values and place them on Microsoft Excel. For the purpose of this research, we graphically present the result and use different statistical methods to test against the hypothesis. We also propose to test whether there are any relationship between mobile banking, mobile shopping and work related activities for the personal users. The demographics of users against their particular usage will also be tested and tabulated. We will also test whether there is any relationship between male or females of certain age, education or level of income use their mobile more for particular personal activities rather than others.

CONCLUSION & FUTURE DIRECTION

In this paper, we have outlined a research framework for identifying how mobile technology influences the use of mobile devices for individual users. A design of the survey and the way we try to analyse the gathered data is discussed. The hypothesis is identified and the numeric result will be statically related to address the proposed hypothesis.

This research will reveal some issues that contribute to the increasing usage of mobile devices in the day to day activities such as in banking, shopping and work related activities identifying the limitations and applications based on the user experience. The limitations and the opinions of the users will be very valuable to service providers who could improve the user experiences by implementing some of the requirements stated by the users.

The dominating information and mobile technology has also created some physical, social and psychological factors for the users which could be identified through this research enabling the research community to work on them in order to improve the user experience of mobile technology, a more pleasurable one.

REFERENCES

Alahuhta, P., Ahola, J., Hakala, H., (2005). *"Mobilizing Business Application: A Survey About the opportunities and Challenges of Mobile Business Applications and Services in Finland".* (Technology Review 167/2005) . Finland: Helsinki. www.tekes.fi/julkaisut/mobilizing.pdf . TEKES Downloaded 28/11/05

Arunatileka, D., Unhelkar, B., (2003), *"Mobile Technologies, Providing New Possibilities in Customer Relationship Management"*, *Proceedings of 5th International Information Technology Conference*, December 2003, Colombo, Sri Lanka.

Geer, D. (2004)*." E-Micropayments Sweat the Small Stuff". Computer, 37*(8), 19-22.

Ghanbary, A. (2003). *" Effects of Computers on Family and Leisure Time"*. B. Applied Science, Honour Thesis. Australia, Sydney: University of Western Sydney.

Herzberg, A., (2003),*"Payments and Banking with Mobile Personal Devices"*, Communications of the ACM, 46(5), pp 53-58

Lacohee, H., Wakeford, N., Pearson, I. (2003). *"A social history of the mobile telephone with a view of its future",* BT Technology Journal, Vol. 21(3),July 2003..

Leedy, D. P. (1993). *"Practical Research"*. USA. New Jersey: Prentice Hall.

Lientz, P. B., Rea, P. K. (2001). *"Start Right in E-Business. A Step by Step Guide to Successful E- Business Implementation "*. Academy Press. ISBN: 0-12-449977-5(pb)

Perry,M.,O'Hara,K.,Sellen,A.,Brown,B. and Harper R.(2001), *"Dealing with Mobility: Understanding Access Anytime, Anywhere"*, ACM Transactions on Computer Human Interaction, Vol. 8(4),pgs 323-347, December 2001

Rao, B., Minakakis, L. (2003). Evolution of Mobile Location-based Services. *Communications of the ACM, 46*(12), 61-65.

Samaras, G. (2002), *"Mobile Commerce: Vision and Challenges (Location and its Management)"*, proceedings of the 2002 Symposium on Applications and the Internet, 2002.

Stafford, T. F., Gillenson, M. L. (2003). *"Mobile Commerce: What it is and what it could be". Communications of the ACM, 46*(12), 33-34.

Sun, J. (2003). *"Information Requirement Elicitation in Mobile commerce",* Communications of the ACM, Vol. 46(12), December 2003..

Tatar, D., Roschelle, J., Vahey, P., & Penuel, W. R. (2003). "*Handhelds Go to School: Lessons Learned. Computer, 36*(9), 30-37

Toffler, A. (1980). "*The Third Wave".* William Morrow and Company Inc. ISBN: 0-688-03597-3

Unhelkar, B. (2005a). *"Web Services and their Impact in creating a paradigm Shift in the Process of Globalization"* book chapter in Global *Information* Society: Operating information systems in a Dynamic Global Business Environment, Ed. Lan, Y., IDEAS Group Publication, USA, ISBN: 1-59140-306-5

Unhelkar, B. (2005b). *"Web Services Extending BPR to Industrial Process Reengineering"* Proceedings of Information Resource Management Association, International Conference. San Diego, California. May 2005.

Vaghjiani, K., Teoh, J, (2005). "*Comprehensive Impact of Mobile Technology on Business*" book chapter in Global *Information* Society: Operating information systems in a Dynamic Global Business Environment, Ed. Lan, Y., IDEAS Group Publication, USA, ISBN: 1-59140-306-5

www.cioinsight.com/article2/0,1540,1632985,00.asp Downloaded: 13/09/2005
www.socialreserachmethods.net/kb/intval.htm Downloaded 29/11/2005
www.socialreserachmethods.net/kb/external.htm Downloaded 29/11/2005
www.deakin.edu.au/~agoodman/sci101/chap7.html Downloaded 29/11/2005

APPENDIX

A sample of the questions is presented:

➢ **What do you consider to be the advantages of using your mobile device?**

a) Great communication tool [ ]
b) Use of mobile devices in education [ ]

c) Leisure (Music, video clips,…..) [ ]
d) Availability of information [ ]
e) Telemetric (Global positioning services, location based services) [ ]
f) Office information [ ]
g) Others (Specify)

➢ **What do you consider to be the disadvantages of using your mobile device?**

a) Security [ ]
b) Loss of work if disconnected [ ]
c) Coverage [ ]
d) Cost [ ]
e) Ease of navigation [ ]
f) Anxiety of using the new technology [ ]
g) Perpetual contact [ ]
h) Data transmission speed [ ]
i) Health hazards [ ]
j) Others (Specify)

➢ **What do you consider to be the main reasons for mobile banking?**
➢ **What sorts of transactions are taking place via mobile banking?**
➢ **For what type of transactions / enquiries do you prefer physically going to a bank?**
➢ **What are the main reasons for mobile shopping?**
➢ **What (problems/difficulties/complaints) are you experiencing by mobile shopping?**

➢ **What kind of set-up would you like mobile shopping to take?**
a) Bigger screen [ ]
b) Ease the navigation [ ]
c) Providing the sufficient information [ ]
c) Link to other web sites [ ]
d) Others (Specify)

➢ **Do you work out of office by using any mobile device?**
a) Yes [ ]
What type of mobile device? How often?
How many hours last week?
What type of work?
If sale, please mention the product
If service, please mention the service
Please mention the name of the device
b) No [ ]
c) Don't Know [ ]

➢ **If you had the choice to either entirely work in the office, entirely work out of the office using your mobile device, which one would you choose? Why?**
➢ **What improvements would you like to see in the conditions of your mobile device?**

➢ **Sex**
a) Male [ ]
a) Female [ ]

➢ **Age Group**
b) 18 - 21 years [ ]
c) 22 - 26 years [ ]
d) 27 - 31 years [ ]
e) 32 - 36 years [ ]
f) 37 - 41 years [ ]
g) 42 - 46 years [ ]

h) 47 or Older [ ]

➢ **Income Group**

a) $ 0 - $ 10 000 [ ]
b) $ 10 001 - $ 20 000 [ ]
c) $ 20 001 - $ 30 000 [ ]
d) $ 30 001 - $ 40 000 [ ]
e) $ 40 001 - $ 50 000 [ ]
f) More than $ 50 000 [ ]

The Architecture and Development of a Secure On-line Transaction and Monitoring System

Instructor RongJou Yang
WuFeng Institute of Technology
MingHsiung, ChiaYi, Taiwan
Email: rjyang@mail.wfc.edu.tw

Dr. Hann-Jang Ho
WuFeng Institute of Technology
MingHsiung, ChiaYi, Taiwan
Email: hhj@mail.wfc.edu.tw

Dr. I-Ming Shieh
WuFeng Institute of Technology
MingHsiung, ChiaYi, Taiwan
Email: pimshieh@mail.wfc.edu.tw

Abstract

With the increasing popularity of Internet, the e-commerce market has grown rapidly in recent years. The secure electronic transaction system (SETS) to process securely electronic transaction is currently the main e-commerce application. Due to the increasing malicious activity (e.g. computer viruses, Trojan horses, Internet worm, Easter Eggs, logic bomb, and intrusion attacks), the SETS developers frequently provide the security updates for their products. In this paper, we improve the existing SETS by adding an extremely important feature of the defense-in-depth strategy. It provides an alternative for businesses to protect their SETS servers by integrating the intrusion detection with secure on-line transaction system. We study the characteristics of network traffic in the insecure phases that are threatened by the malicious code attacks and Internet worm propagation and compare it with the phases in which such activities will be secure in the secure on-line transaction system. Moreover, an experimental platform is proposed for testing the architecture of security that is established.

Keywords:

Secure Electronic Transaction System, Intrusion Detection, Security Monitoring, Defense-in-depth, Firewall

INTRODUCTION AND LITERATURE REVIEW

Recently, with the rapid development and progress of IT and Internet, the flow of information and the amount of transactions are drastically increasing. Many traditional brick-and-mortar stores have been gradually shifted to on-line virtual stores as an additional marketing channel, which affect many facets of a variety of areas including logistic, cash flow, merchandise flow, and information flow. Meanwhile, the traditional way of payments such as cash, credit cards, and checks have also been gradually shifted to the on-line virtual mechanism for order fulfillment in order to accomplish the transaction processing. Although the on-line transaction can reduce the cost and increase the efficiency the issue of information security is still the spotlight of concerns. We can even say that no matter how convenient the transactional platform is it just can not allure the interested parties without providing an environment of security and certain degree of assurance in the commercial circumstance.

Currently, SSL (Secure Socket layer) connection technology is most often used as the secure on-line transaction mechanism even though it does not verify that the transaction is from both the right clients and merchants so that it gives the network hackers chances to grab the transaction information or simulate transactional subjects in Freier, Karlton, and Kocher (1996). Although the verification technology in the secure electronic transaction mechanism such as SET (Secure Electronic Transaction) and 3D-Secure has been largely improved it is complicated in use comparing with SSL (MasterCard Inc., 1997 and Visa International, 2002) and unbearable of the abnormal attacks from the intended hacker.

No matter which kind of mechanisms for the secure electronic transaction is adopted on the on-line transactional host it just can not completely avoid the possible attacks or intrusions. Moreover, the processing of transaction encryption on the transactional host with a secure on-line transaction mechanism deployed can not be detected by the computer virus scanner and monitored by IDS (Intrusion Detection System) so that the leak of security often becomes the convenient way for hackers' intrusion. The security concern of a server with security mechanism installed is often misled due to such a blind spot of security. In fact, we are confronted by the ineffective shield against the intrusion and by much more security threat in Karve (2001). Therefore, it is still an important issue how to enhance the security of transactional host by employing the monitoring technology and by constructing defensive system in the area of information security today.

In the domain of information security, the firewall has been the most extensively used network technology in Guster and Hall (2001). It has been usually used with anti-virus software tools so as to protect the significant resources within Intranet and block the improper and illegitimate visits and the insecure data transmission.

Generally, although the firewall can filter the illegitimate visits and the flow of abnormal packets it can not assure the security and reliability of the network. Thus, the IDS have been usually employed to supplement the inadequacy of firewall in Yasin and Awan (2004) and Sailer, Perez, Dyer, and Doorn (2001). It can prompt the alerts and record the system log after detecting the illegitimate packets in the network timely and prevent and respond to the unauthorized activities within Intranet by the effective detection of intrusion. Also, although the firewall can resist the externally insecure packets in order to protect Intranet from hackers' attacks it still could be attacked due to the undetectable vulnerability of legitimate flow of the packets. Therefore, the legitimate flow of the packets recognized by the firewall may result in circumstance of the leak of potential threat within Intranet. We can draw a parallel the EC (Electronic Commerce) Web site as a bank in that the firewall is like the entrance of the bank and IDS is like the alarm system. The relationship between them is complementary in stead of substitution Sailer, Perez, Dyer, and Doorn (2001).

Currently, there are plenty of researches and applications in IDS as in Yasin and Awan (2004), Roesch (1999), and Sailer, Perez, Dyer, and Doorn (2001) and IPS (Intrusion Protection System) in Desai (2003). Although the implementation of those systems are not particularly aimed at the secure mechanism of electronic transaction we can focus on the improvement of them in order to be able to detect the common protocol packets of secure electronic transaction and record the system log for the reference resources of promotion for the secure on-line electronic transaction on server.

RESEARCH AND ANALYSIS METHOD

In terms of the security concern of on-line transaction, this research integrated the functionality of NIDS (Network IDS) with the transactional host to develop a monitoring and detecting system for SSL security mechanism of the on-line transaction which is most often used currently. An open source GPL (General Public Licenses) IDS Snort in Roesch (1999) is employed to detect and avoid the intrusion through the analysis of the packets filtering and to block the attacks from the abnormal packets through the protection mechanism provided by the firewall in order to discover the intrusion of the system such as port and weak scanning so as to assure the security of on-line transaction.

This paper is focused on the rules chain detection, system log monitoring, and accesses filtering in the constructed monitoring platform, described in the following subsection.

Rules Chain

The flexible rules chain setting for the system is aimed at the packets filtering functions of the IDS Snort and firewall so as to integrate intrusion detection with protection monitoring and generate the related alerts prompts and packets log in order to track the suspicious attacks and abnormal activities.

Database Accesses Logs Monitoring

MySQL is employed to store and import Snort-related files into the tables of the databases. Besides, the accesses logs of the system are placed in the directory /var/log/ and MySQL database to provide the important information of user accesses to the major resources for the system and application in order to track the hackers' attack. The syslog is used as log monitoring and rules setting to reveal the information of real-time packets filtering.

Accesses Filtering

The connection control of bandwidth and packets filtering mechanism are aimed at the circumstance of transactional host in Intranet as follows:

- Configure xinetd and TCP wrappers to analyze, diagnose, filter, and intercept the packets so as to create a network security environment based on the management of xinetd.
- Implement the functions of firewall such as redirecting, IP masquerade (i.e., network IP address translation) and the reassembling of the packets.

- Exercise the QoS (Quality of service) management of bandwidth and connection accesses in order to avoid or alleviate the buffer overrun attacks by refusal to the service such as DoS (Denial of Service) and DDoS (Distributed DoS).

SYSTEM ARCHITECTURE

In the environment of the on-line transaction monitoring system deployed, not only the passive blocking of hacker's attacks but also active intrusion avoidance is achieved by employing IDS. The on-line monitoring system proposed in this paper is a three-layered structure of security indicated as in Fig. 1 and described as follows.

Layer 1 – IDS at the Gateway

In layer 1, the main purpose of deploying IDS at the gateway connected to Internet is to detect and avoid a variety of malicious code attacks and conceal scanning mode such as FIN and Null in order to complement the single functionality of SYN-related flooding vulnerability scanning provided by the firewall so as to prompt an alert for the transactional host in case of intrusion.

Layer 2 – NAT Server

In Layer 2, the packet filtering functionality of firewall is integrated with IDS to block the abnormal packets. The transactional host is protected in this way and the shield of system security is enhanced through the functionality of IP address translation provided by the deployment of NAT (Network Address Translation) server.

Layer 3 – xinetd Configuration for Access control

In layer 3, the access control mechanism of IP address, host, domain, and duration of connection is done through the configuration of xinetd and integrated with the functionality of QoS network transmission and system logs monitoring in order to assure that the regular services and secure connections are provided by the transactional host.

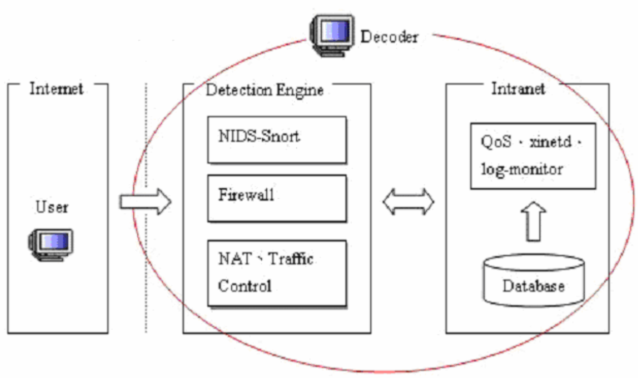


Figure 1. The Architecture of Monitoring System

EXPERIMENTAL SYSTEM DESIGN

The experimental system design is based on Linux operating system. A monitoring system is deployed on the transactional host and there are six steps to proceed as described in the following subsections:

Constructing Experimental Platform

First, the IDS Snort and firewall (including NAT server) are deployed on the transactional host and the functions of detecting and filtering the attacking packets are started. Second, the file Kiwi_Syslogd_Service is downloaded from the remote host which acts as a log server and installed and run on the local host as indicated in Fig. 2 and Fig. 3.

```
database: password is set
database: database name = snort
database:          host = localhost
database:   sensor name = 140.130.127.151
database:     sensor id = 8
database: inconsistent cid information for sid=8
          Recovering by rolling forward the cid=84
database: schema version = 106
database: using the "alert" facility
1864 Snort rules read...
1864 Option Chains linked into 188 Chain Headers
0 Dynamic rules
+++++++++++++++++++++++++++++++++++++++++++++++++++

Warning: flowbits key 'realplayer.playlist' is checked but not ever set.

+------------------------[thresholding-config]----------------------------------
| memory-cap : 1048576 bytes
+------------------------[thresholding-global]----------------------------------
| none
+------------------------[thresholding-local]-----------------------------------
| gen-id=1      sig-id=2523      type=Both        tracking=dst count=10   seconds=10
| gen-id=1      sig-id=2494      type=Both        tracking=dst count=20   seconds=60
| gen-id=1      sig-id=2496      type=Both        tracking=dst count=20   seconds=60
| gen-id=1      sig-id=2495      type=Both        tracking=dst count=20   seconds=60
| gen-id=1      sig-id=2275       type=Threshold tracking=dst count=5    seconds=60
+------------------------[suppression]------------------------------------------
--------------------------------------------------------------------------------
Rule application order: ->activation->dynamic->alert->pass->log

        --== Initialization Complete ==--

-*> Snort! <*-
Version 2.2.0 (Build 30)
By Martin Roesch (roesch@sourcefire.com, www.snort.org)

You have new mail in /var/spool/mail/root
[root@localhost root]# █
```

Figure 2. The Configuration of IDS Snort

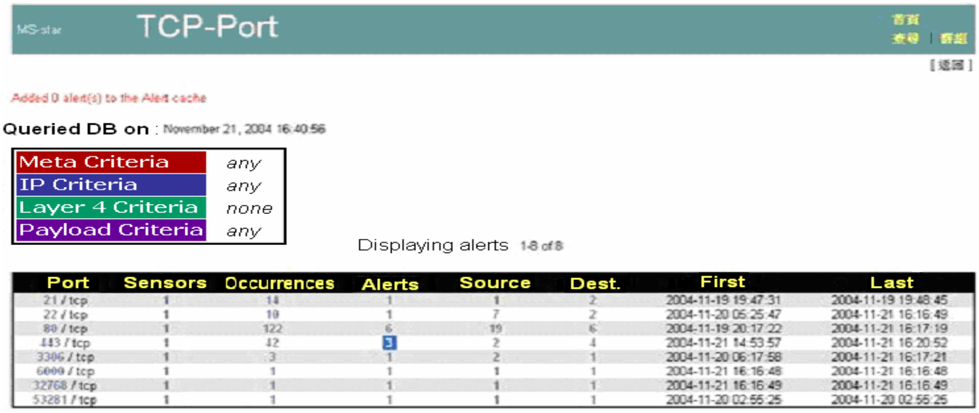
```
[root@localhost root]# iptables -L
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN,RST,PSH,ACK,URG/FIN,PSH,URG
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN,RST,PSH,ACK,URG/FIN,PSH,URG
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN/FIN,SYN
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN/FIN,SYN
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN/FIN,SYN
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN/FIN,SYN
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN/FIN,SYN
DROP       tcp  --  anywhere             anywhere           tcp flags:FIN,SYN/FIN,SYN

Chain FORWARD (policy DROP)
target     prot opt source               destination

Chain OUTPUT (policy ACCEPT)
target     prot opt source               destination

Chain ping (0 references)
target     prot opt source               destination
RETURN     icmp --  anywhere             anywhere           icmp echo-request limit: avg 1/sec burst 5
REJECT     icmp --  anywhere             anywhere           reject-with icmp-port-unreachable
[root@localhost root]# █
```

Figure 3. The Configuration of Firewall

Transactional Packets and Protocols Analysis

The IDS Snort acts as a detecting engine for the transactional packets and is responsible for the loading of the detecting rules after startup and the matching analysis of the characteristic of packets. The detecting results from the IDS Snort are then presented via ACID (Analysis Console for Intrusion Databases). This paper focuses on the matching and testing of secure protocols in the on-line transaction as indicated in Fig. 4 and Fig. 5.

Figure 4. The Display of Detected Packet Information



Port	Sensors	Occurrences	Alerts	Source	Dest.	First	Last
21 / tcp	1	14	1	1	2	2004-11-19 19:47:31	2004-11-19 19:48:45
22 / tcp	1	10	1	7	2	2004-11-20 05:25:47	2004-11-21 16:16:49
80 / tcp	1	122	6	19	6	2004-11-19 20:17:22	2004-11-21 16:17:19
443 / tcp	1	42	3	2	4	2004-11-21 14:53:57	2004-11-21 16:20:52
3306 / tcp	1	3	1	2	1	2004-11-20 06:17:58	2004-11-21 16:17:21
6000 / tcp	1	1	1	1	1	2004-11-21 16:16:48	2004-11-21 16:16:48
32768 / tcp	1	1	1	1	1	2004-11-21 16:16:49	2004-11-21 16:16:49
53281 / tcp	1	1	1	1	1	2004-11-20 02:55:25	2004-11-20 02:55:25

Figure 5. The Analysis (TCP-Port Flow Summary) of On-line Transaction Protocols

Detecting Rules of Transactional Host

The flexible detecting rules are configured based on the demands of system environment and the detecting actions for the transactional host. The unrelated IP information appearing in the system log is decreased through the use of IDS Snort. This paper focuses on SSL connection as indicated in Fig. 6.

```
alert tcp $EXTERNAL_NET any -> 140.130.109.99/24 443 (msg:"SSL(HTTPS)
text alert";content:"/selector/showcode.asp";nocase;)
alert tcp $EXTERNAL_NET any -> 140.130.109.99/24 443 (msg:"SSL(HTTPS)
SYN/FIN packet";flags:SF,12;)
alert tcp $EXTERNAL_NET any -> 140.130.109.99/24 443 (msg:"SSL(HTTPS)
preprocessor alert";sameip;flow:stateless;)
alert tcp any any -> 140.130.109.99/24 443 (msg:"SSL(HTTPS) SYN
alert";flags:SF,12;)

alert tcp $EXTERNAL_NET any -> 140.130.109.99/24 80 (msg:"HTTP text
alert";content:"/selector/showcode.asp";nocase;)
alert tcp $EXTERNAL_NET any -> 140.130.109.99/24 80 (msg:"HTTP SYN
alert";flags:SF,12;)
alert tcp $EXTERNAL_NET any -> 140.130.109.99/24 80 (msg:"HTTP
preprocessor alert";sameip;flow:stateless;)
```

Figure 6. The Setting of Detection Rules

Transaction Logs and Monitoring

The main purpose of the system is to block the malicious attacks before it is breached. The detection of network hackers and the tracking of the attacking IP are done by employing the IDS Snort as indicated in Fig. 7 and Fig. 8. The event logs in the log file are integrated with the SNMP-supported adapter to filter the packets. The detected message of events is passed to the transactional host by employing two separate subsystems: Message Alerting System and System Logging System. In the latter, the packets are decoded as the readable format TCP DUMP and are recorded while in the former, the alert messages are prompted and recorded in the syslogd, flat files of documents, Linux concatenated string, or database in order that the related alerts can be prompted to the on-line transaction system.

```
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/stylesheet.css HTTP/1.1" 200 6154
140.130.109.99 - - [29/Oct/2004:16:43:14 +0800] "GET /shop/login.php?
osCsid=590b1b4d884f74c04d784b71401a45b2 HTTP/1.1" 200 26195
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/oscommerce.gif HTTP/1.1" 200 3656
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/header_account.gif HTTP/1.1" 200
439
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/header_cart.gif HTTP/1.1" 200 535
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/header_checkout.gif HTTP/1.1" 200
505
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/infobox/corner_left.gif HTTP/1.1"
200 123
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/pixel_trans.gif HTTP/1.1" 200 43
140.130.109.99 - - [29/Oct/2004:16:43:15 +0800] "GET /shop/images/infobox/corner_right_left.gif
HTTP/1.1" 200 52
140.130.109.99 - - [29/Oct/2004:16:43:16 +0800] "GET /shop/images/infobox/arrow_right.gif HTTP/1.1"
200 69
140.130.109.99 - - [29/Oct/2004:16:43:16 +0800] "GET /shop/images/gt_interactive/wheel_of_time.gif
HTTP/1.1" 200 3828
140.130.109.99 - - [29/Oct/2004:16:43:16 +0800] "GET /shop/includes/languages/tchinese/images/
buttons/button_quick_find.gif?osCsid=590b1b4d884f74c04d784b71401a45b2 HTTP/1.1" 200 554
140.130.109.99 - - [29/Oct/2004:16:43:16 +0800] "GET /shop/includes/languages/tchinese/images/
buttons/button_continue.gif HTTP/1.1" 200 1315
140.130.109.99 - - [29/Oct/2004:16:43:16 +0800] "GET /shop/images/table_background_login.gif
HTTP/1.1" 200 1381
140.130.109.99 - - [29/Oct/2004:16:43:16 +0800] "GET /shop/images/infobox/corner_right.gif
HTTP/1.1" 200 123
```

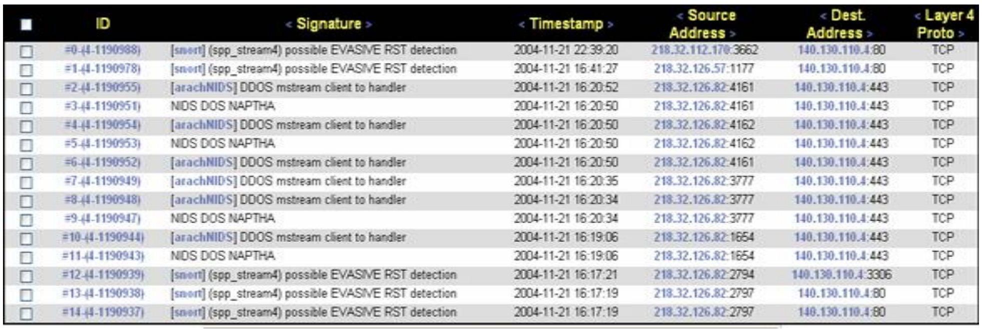
Figure 7. The Log File of Connections

```
[Tue Mar 08 09:38:51 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Tue Mar 08 09:38:51 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Tue Mar 08 14:33:11 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Tue Mar 08 14:33:11 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Tue Mar 08 14:33:16 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Tue Mar 08 14:33:16 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Tue Mar 08 17:56:01 2005] [error] [client 140.130.109.82] Directory index forbidden by rule: /var/www/html/
[Tue Mar 08 17:56:09 2005] [error] SSL handshake interrupted by system [Hint: Stop button pressed in browser?!]
[Tue Mar 08 17:56:58 2005] [error] [client 140.130.109.82] Directory index forbidden by rule: /var/www/html/
[Tue Mar 08 17:57:06 2005] [error] SSL handshake interrupted by system [Hint: Stop button pressed in browser?!]
[Wed Mar 09 03:18:25 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 03:18:25 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 03:18:30 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 03:18:30 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 04:02:36 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 04:02:36 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 07:11:42 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 07:11:42 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 07:11:46 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 07:11:46 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 08:46:55 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 08:46:55 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 08:47:00 2005] [warn] RSA server certificate is a CA certificate (BasicConstraints: CA == TRUE !?)
[Wed Mar 09 08:47:00 2005] [warn] RSA server certificate CommonName (CN) `localhost.localdomain' does NOT match server name!?
[Wed Mar 09 09:02:58 2005] [error] [client 140.130.127.152] Directory index forbidden by rule: /var/www/html/
[Wed Mar 09 09:03:06 2005] [error] SSL handshake interrupted by system [Hint: Stop button pressed in browser?!]
```

Figure 8. The Access Records of On-line Transactions

Intrusion Alerts

The alerts are prompted in response to the detected abnormal packets by employing the IDS Snort in accordance with the functions of the OS built-in firewall that are configured in terms of the filtering rules, IP translations, and packets reassembling rules. The firewall is implemented through the use of ipchains and iptables to provide the functions of gateway or router in order to set up a flawless security mechanism as indicated in Fig. 9 and Fig.10.



	ID	Signature	Timestamp	Source Address	Dest. Address	Layer 4 Proto
	#0-(4-1190988)	[snort] (spp_stream4) possible EVASIVE RST detection	2004-11-21 22:39:20	218.32.112.170:3662	140.130.110.4:80	TCP
	#1-(4-1190978)	[snort] (spp_stream4) possible EVASIVE RST detection	2004-11-21 16:41:27	218.32.126.57:1177	140.130.110.4:80	TCP
	#2-(4-1190955)	[arachNIDS] DDOS mstream client to handler	2004-11-21 16:20:52	218.32.126.82:4161	140.130.110.4:443	TCP
	#3-(4-1190951)	NIDS DOS NAPTHA	2004-11-21 16:20:50	218.32.126.82:4161	140.130.110.4:443	TCP
	#4-(4-1190954)	[arachNIDS] DDOS mstream client to handler	2004-11-21 16:20:50	218.32.126.82:4162	140.130.110.4:443	TCP
	#5-(4-1190953)	NIDS DOS NAPTHA	2004-11-21 16:20:50	218.32.126.82:4162	140.130.110.4:443	TCP
	#6-(4-1190952)	[arachNIDS] DDOS mstream client to handler	2004-11-21 16:20:50	218.32.126.82:4161	140.130.110.4:443	TCP
	#7-(4-1190949)	[arachNIDS] DDOS mstream client to handler	2004-11-21 16:20:35	218.32.126.82:3777	140.130.110.4:443	TCP
	#8-(4-1190948)	[arachNIDS] DDOS mstream client to handler	2004-11-21 16:20:34	218.32.126.82:3777	140.130.110.4:443	TCP
	#9-(4-1190947)	NIDS DOS NAPTHA	2004-11-21 16:20:34	218.32.126.82:3777	140.130.110.4:443	TCP
	#10-(4-1190944)	[arachNIDS] DDOS mstream client to handler	2004-11-21 16:19:06	218.32.126.82:1654	140.130.110.4:443	TCP
	#11-(4-1190943)	NIDS DOS NAPTHA	2004-11-21 16:19:06	218.32.126.82:1654	140.130.110.4:443	TCP
	#12-(4-1190939)	[snort] (spp_stream4) possible EVASIVE RST detection	2004-11-21 16:17:21	218.32.126.82:2794	140.130.110.4:3306	TCP
	#13-(4-1190938)	[snort] (spp_stream4) possible EVASIVE RST detection	2004-11-21 16:17:19	218.32.126.82:2797	140.130.110.4:80	TCP
	#14-(4-1190937)	[snort] (spp_stream4) possible EVASIVE RST detection	2004-11-21 16:17:19	218.32.126.82:2797	140.130.110.4:80	TCP

Figure 9. The Latest Alert Listing

Figure 10. The Statistical List of Alerts

Flow Information on Host

The flow monitoring on the transactional host is based on the MRTG (Multi Router Traffic Grapher) in order to monitor the current states of the flow and performance of services on the host as indicated in Fig. 11 and Fig. 12.

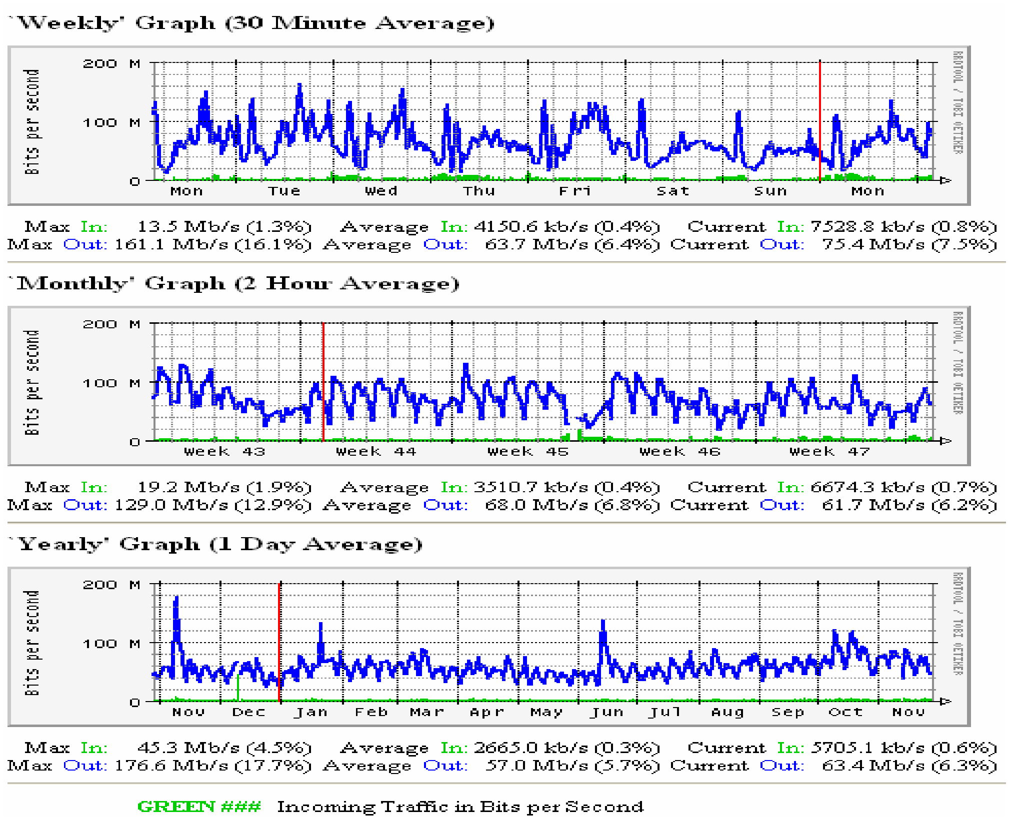


Figure 11. The Traffic Flow Analysis on Host

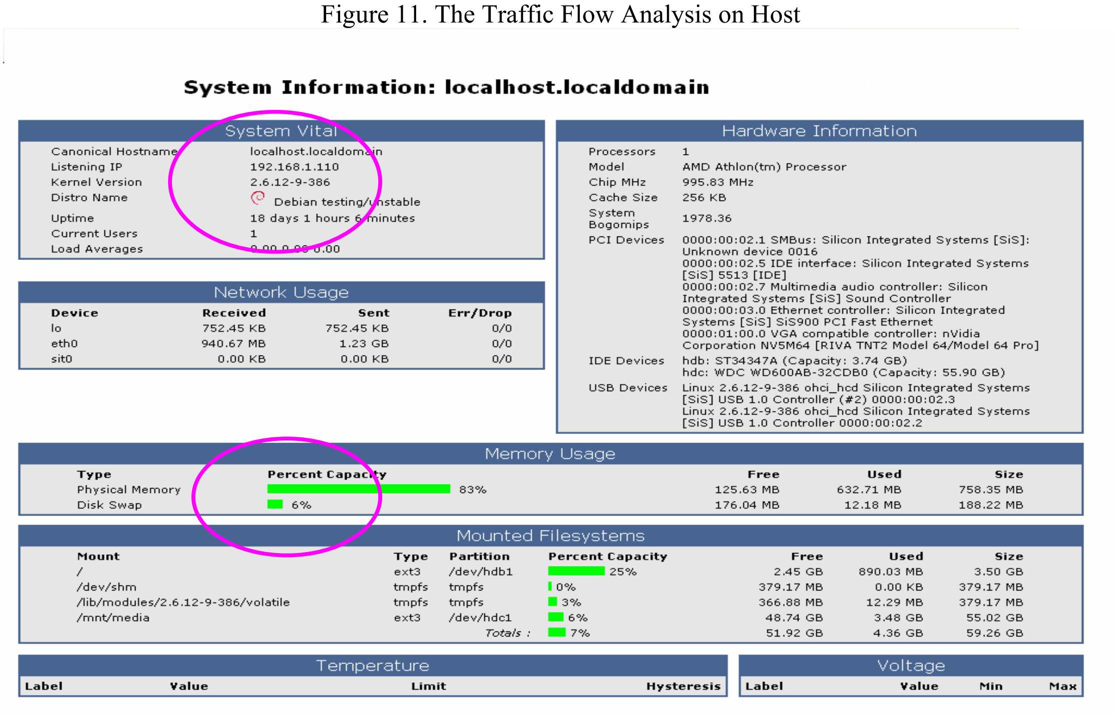


Figure 12. The System Information on Host

EXPERIMENTAL RESULTS AND ANALYSIS

The user interface is implemented by using the monitoring system. All the main connection packets such as TCP, UDP, and ICMP flowing through the transactional host are analyzed, recorded, and filtered. The connection time, the source and destination port, the source and destination IP, and the length of the data field in the PDU (Protocol Data Unit - packet) are checked according to the configuration of the filtering rules set in the IDS Snort and firewall so as to diagnose the current states of normal and abnormal packets. The packets are regarded as in the abnormal state if the communication protocols and ports included in the packets remain the same within a period of time. The abnormal packets are stored and recorded and the alert prompts are issued at the same time. Meanwhile, the abnormal packets are blocked by the monitoring system based on the detecting rules.

The host within Intranet can be started up standalone to listen to the requests from the clients and to restrain the frequency and number of connections. The sub-processes can be prawn to provide more kinds of services in case of large amount of requests. In addition, the xinetd daemon and the rules of authentication within the firewall are configured to filter a variety of packets so as to block the illegitimate accesses and unusual attacks from the network hackers.

In the course of packets accessing and transactions processing, the monitoring functions of the monitoring system can be brought into full play in the effect of early detection of intrusion and prevention from attacks. In the aspect of intrusion detection on the transactional host, the system performance might be degraded due to the SYN flooding records (logs) and the detecting rules must be updated frequently.

Undetectable of Some Particular Attacks

Although the network connections and the switching of packets can be monitored through the IDS Snort some peculiar attacks from the hackers due to the leaks of buffer overflow in the application layer might not be detected and blocked effectively. In the case that the attacks is unable to be blocked by the NIDS (network IDS) and firewall a vertical security mechanism HIDS (Host IDS) is employed to complement the insufficiency of the network IDS. The applications are integrated with the OS on the host to realize the undetectable attacks so as to provide the transactional host with more robust protection measurement.

The Packets Loss Rate Increases while the Flow Increases

Although the network connections and the switching of packets can be monitored through the IDS Snort some peculiar attacks from the hackers due to the leaks of buffer overflow in the application layer might not be detected and blocked effectively. In the case that the attacks is unable to be blocked by the NIDS (network IDS) and firewall a vertical security mechanism HIDS (Host IDS) is employed to complement the insufficiency of the network IDS. The applications are integrated with the OS on the host to realize the undetectable attacks so as to provide the transactional host with more robust protection measurement.

The Limit of IDS

In accordance with the increasing and encrypted network flows, the difficulty of dealing the leak of security, and the increasing dissemination of IPv6 technology, the data encrypting process of SSL is not examined by ordinary security scanners or IDS. Thus, the difficulty in the interception of IP and the characteristics of attacks is increased. Therefore, the detecting technology of IDS and the performance of hardware must be also constantly researched and developed and the functionality be further promoted. And it relies on the development of IT and the security facilities of hardware and software.

CONCLUSIONS AND DISCUSSION

This paper is aimed at the on-line transaction mechanism. An intrusion detection system software tool, Snort, is employed to integrate the hardware firewall with the protection functionality of the system in order to monitor the on-line secure transaction and the behavioral characteristic of the hackers in which the detected data can be stored and managed and the transaction status of, and the network flow from/onto the host can be monitored effectively. Through the retrieval of the detected data from Snort by ACID a variety of information about the transmission packets can be obtained fast and the firewall and internal mechanism can be notified in case of the attacks to response with the enhancement of the security of the server in time in order to reduce the damage from the intrusion.

As a mater of fact, we can promote the security of electronic transaction through the employment of hardware and software in response to the emergence of digital era in that most of time diverse techniques are mixed and employed by the attackers in order to achieve the expected effect of destruction in most perpetration of electronic transaction and the intrusion of hackers can be succeeded only when the loopholes of hardware and software exist at the same time.

REFERENCES

[1] A. O. Freier, P. Karlton, and P. C. Kocher, The SSL Protocol Version 3, *Netscape Communications Corp*., 1996, http://home.netscape.com/eng/ssl3.

[2] MasterCard Inc., SET Secure Electronic Transaction Specification, Book 1: Business Description, *MasterCard Inc.,* 1997.

[3] Visa International, 3-D Secure Version 1.0.2, *Visa International,* 2002.

[4] A. Karve, Can Intrusion Detection Keep an Eye on Your Network's Security? *Network Magazine*, 2001.

[5] D. Guster, C. Hall, A firewall configuration strategy for the protection of computer networked labs in a college setting, *Journal of Computing in Small Colleges*, vol.17 no.1, pp.187-193, 2001.

[6] M. M. Yasin, A. A. Awan, A study of host-based IDS using system calls, *INCC*, pp. 36- 41, 2004.

[7] M. Roesch, Snort: Lightweight Intrusion Detection for Networks, *13th USENIX conference on System Administration*, pp.229-238, 1999.

[8] R. Sailer, R Perez, J.G. Dyer, L.van Doorn, Personal Firewalls and Intrusion Detection System, *Proceedings of the 2nd Australian Information Warfare and Security Conference (IWAR)*, 2001.

[9] N. Desai, Intrusion Prevention Systems: the Next Step in the evolution of IDS, *Security Focus*, 2003. http://www.securityfocus.com/infocus/1670.

Creation of a framework for transitioning to mobile business processes

Dinesh Arunatileka, Abbass Ghanbary
PhD Research Student, University of Western Sydney
Email: darunati@cit.uws.edu.au
Email: ghanbary@cit.uws.edu.au

Abstract
This paper outlines a framework for transitioning of business processes in organisations enabling them to take advantage of mobile technologies .A design of a survey, as part of the research methodology to gather data, is also discussed.. Data collected from the responses to the survey will be analysed in order to verify the validity of the research questions. Outline of a resultant systematic methodology for mobile transformation (m-transformation) of business processes for an organisation has also been proposed..

Keywords:
M-transition, m-transformation of business processes

INTRODUCTION

The objective of this study is the creation of a framework to enable systematic transformation of business processes of organisations to enable mobile technology. Mobile technology has become a key area of interest in domestic markets as well as in the corporate world. This is so because mobile technology has created a substantial impact on how consumers and organisations carry out their business activities. Mobile technology is no longer restricted to voice transmission but has advanced to include mobile data, location based services, mobile payments and specific target marketing to name but a few. Nicopolitidis (2004) states that mobile wireless markets is one of the fastest growing segments in the telecommunication industry. The business integration of Mobile technology is popularly known as m-business, which has aggregated Internet, wireless and e-business together (Kalakota & Robinson, 2002). The major drivers for m-business are: convenience due to ubiquity provided, better conceived security, revenue drives by operators pushing more services, declining costs and opportunities for other service providers (www.acma.gov.au). Rao & Minakakis (2003) describe location based services being one of the key applications in m-commerce. Schilit et. al(2003) state that location based services would become the next killer application in personal wireless applications.

Organisations, particularly service-based organizations, have seriously taken up mobile technology to facilitate their business processes. For example, in the banking industry, mobile technology has provided secure and convenient use of e-banking, payments, brokerage and other such transactions which are integral to the banking sector (Herzberg, 2003). Mallat et. al (2004) state that mobile banking services are valued by the users due to the time and place independence. The freedom from space and time for the carriers of mobile gadgets needs to be identified, studied and applied within two main contexts in an organisational environment: mobility in delivery of service to customers and the mobility at work place for its employees. As per Paul Stroud, Managing Director of Mnetics, a leading provider for supply chain and retail solutions in Europe (www.ecominfo.net), mobility is a suite of mobile applications that link host computer systems and mobile workforces. Mobility combines some of the following features:

- Intuitive mobile applications that compliment and enhance working practices, while verifying and validating actions in real time.
- A selection of mobile devices from leading manufacturers which include bar code scanners, digital cameras and GPR and GPRS radios - Mobility can even include card swipe devices for cash transactions on the move or receipt validation.
- A secure, efficient and cost-effective wide-area mobile data network.
- Inputs from external devices, such as GPS for vehicle location and tracking or more far reaching Telematics equipment if comprehensive or specific vehicle monitoring is a prerequisite.
- Internet interfaces for easy access and data sharing.
- Direct integration into host systems and applications.

These features are further verified by Murugesan (2005) in the description of four key factors which makes mobile applications successful, namely :

- Technology(devices, networks, standards and software)
- Applications(including the value of the application to the users, the business model, the mobiles carrier and process refinement)
- Regulatory framework(legal, privacy and ethical requirements to be compiled with)

- User adoption and usage

It is also important to focus our attention to the regulatory body, the Australian Communications and Media Authority (www.acma.gov.au) which indicates the communications policy of Australian government up to year 2020. This report indicates the importance of developing and implementing a comprehensive and “wholesome” business approach to regulating the following in the areas of communications technology, especially, mobile technology. These areas include:

- Monitoring industry performance of network security and network integrity risk management
- Managing issues from voice, data and multimedia convergence
- Market based approaches to voice over IP and public switched telephone network interconnection
- Updating consumer information to reflect trends in services convergence
- Risks to legacy network reliability and accessibility
- Taking broader measures of consumer benefit and consumer satisfaction and
- Working more closely with other regulating bodies

This approach very clearly indicates that the policy making bodies in Australia are focused and committed to facilitating the development of the mobile communications technologies. However, in order to achieve the aforementioned use of mobility, there is a need to undertake a systematic study of the processes for mobile transformation. These change processes can be referred to as the “m-transformations” (Unhelkar, 2005). M-transformation may be defined as the process of subjecting the business processes of an organisation to undergo systematic changes in order to incorporate mobile technologies in them to gain time and cost savings and increase the productivity of those processes.

As a part of this investigation into how organisations adopt mobile technologies, a survey has been designed. The survey is aimed at the corporate management in order to elicit their views, and the approach they would take on adopting mobile technologies. Prior to the survey, there was a ‘case study’ that was conducted at a leading bank in Australia, wherein certain business processes were investigated in terms of how they are currently conducted. This was followed by a suggested methodology to m-transform the processes by incorporation of mobile gadgets in those processes. The authors of this research paper are proposing to extend the case study and augment it with a survey aimed at organisations at large. The survey, discussed in this paper, will look at the current business processes in order to investigate whether an M-transformation model could be applied to these business processes in order to transform them to mobile enabled business processes.

The structure of the paper is as follows: a)Introduction; b)Proposed Research Approach; c)Scale of four stages in m-transformation in an organization; d) Mobile Transition Road Map; e)Research Questions and Hypotheses; f)Structure of the Questionnaire; g)Analysis of the Survey; h)Conclusion and Future Work.

PROPOSED RESEARCH APPROACH

The proposed approach is a questionnaire aimed at business organisations surveying the current status with regard to the use of mobile technology. Nardi (2003) describe questionnaires as ideally suited for measuring people’s attitudes and opinions. Thus the questionnaire would give us a good insight into the corporate management in terms of mobile technology. The questionnaire will be sent to organisations, especially in Australia, India and Sri Lanka in order to verify the Mobile transition road map and m-transition looked at, in three different perspectives namely: technology, methodology and sociology. The different country environment would provide an international flavour in terms of different market environments as well. The road map and the three perspectives are explained later in the paper. The majority of the sample would be from Wester Sydney area business organisations which is also explained later.

The questionnaire is designed with different sections probing in to different areas of information with regard to use of mobile technology by the organisation. Before analysing the questionnaire itself, let’s look at the fundamental questions that are to be addressed in order to verify the use of mobile technologies in the business processes of an organisation.

Initially, depending on the current situation of the organisation as assessed by the survey results, the organisation could be placed on a scale of one to four stages with regards to where it stands in terms of using mobile technology. This is shown in Figure 1. These proposed four stage scale of m-transformation in organisations has been arrived by the authors after careful consideration of earlier research in this area, notably by Kalakota and Robinson (2002). Kalakota and Robinson (2002) have depicted the market evolution of m-business in five stages. They are: System Integration, Business

Integration, Web enabled commerce, Web enabled transformation and Multi-device multi web transformation (M-business). The Kalakota and Robinson model depicts a roadmap towards M-business. However, that model misses out on the rigours of identifying and analysing where the organisation is currently standing in terms of m-transformation. This is partly due to the timing of the research since this road map was introduced in 2002 when m-business was at very early stages. Lan & Unhelkar (2005) have also considered a Globalization Transition road map (GET) for organizational transitions. Their roadmap has four steps namely: Arguments and approach, Goals and vision, Process Framework and Enactment. This model is concentrating more on how organisations react globally towards the globalisation. We have considered the aforementioned models in arriving at our proposed four stages in m-transformation. Since the focus of both these models is different from our main objective of m-transformation, it is difficult to compare and contrast these models with the proposed model. However they were used as background knowledge due to lack of direct material available depicting the stages in m-transformation. Thus comparing the proposed model directly with any of the existing models is not possible at this point of time. The following section describes the new model to identify the stage of an organization in terms of m-transformation.

SCALE OF FOUR STAGES IN M-TRANSFORMATION IN AN ORGANIZATION

Figure 1 depicts the four stages where a current organisation could be placed at in terms of m-transformation. This identification is very important in deciding how, further m-transformation of business processes should take place in the organisation. In Kalakota and Robinson (2002), the adaptation of mobile technology is an evolutionary step from Web enabled transformation. In the Scale of four stages, Web enabled transformation is already happened and is a given. We are probing into the Multi-device multi web transformation stage of Kalakota and Robinson(2002) and further analyse the organisation to identify level of maturity of m-transformation of the organisation. The proposed four stages of m-transformation of an organisation is as follows:

Stage 1(preliminary stage) – where the organisation although has some idea that mobile technology exists, there is no effort to improve the work methods and business processes, using this technology. This is the stage where the organization has started looking at ways and means of considering mobility in its business processes. A study of the existing business processes with and without mobility may be undertaken at this stage. A collation of existing software applications, databases, security and other software assets may also be undertaken here.

Stage 2(ad-hoc stage) – Mobile technology is used as a tool in an ad-hoc way. There is no calculated and planned usage of mobile technology yet. The organisation is at the initial stage in adopting some business processes m-transformed. There is average use of mobile technology but there is also very large room for improvement in the usage of technology and improving business processes. Individual players (such as employees and customers) may be using hand held gadgets to accomplish business processes.

Stage 3(Advanced stage) – At this stage, mobile technology is used in a substantial way. Most departments and business units use mobile technology in their business processes. Intra business process communications are transformed to use mobile technology. However further improvement could be done by transforming the inter business process communications to use mobile technology as well. For instance, selling of a product through mobile technology is possible but the payment for the product still requires some non mobile process. Thus systematic transformation of the organisation could still be improved in order to achieve the best results through the new technology.

Stage 4(M-transformed stage) - Mobile technology is used as a true technology tool in the day to day running of the business. All the significant processes are fully m-transformed so that most combined business processes are tackled fully by mobile technology. For example, a purchase of a product, payment for it and the receipting could all be done through mobile technology. Time savings and cost savings are achieved through the extensive and systematic use of mobile technology while offering the customers a new and fully integrated technology to do business transactions. Forward and backward integrations with suppliers and value added service providers also could be facilitated through mobile technology.

Figure 1 depicts the Scale of four stages in M-Transformation of an Organisation. It is a four step stair case model where the stages are achieved in a linear scale with m-transformed stage being the highest and most desirable stage in terms of the usage of mobile technologies. This stage could also be described as a perfectly m-transformed organisation gaining full advantage of the new technology.

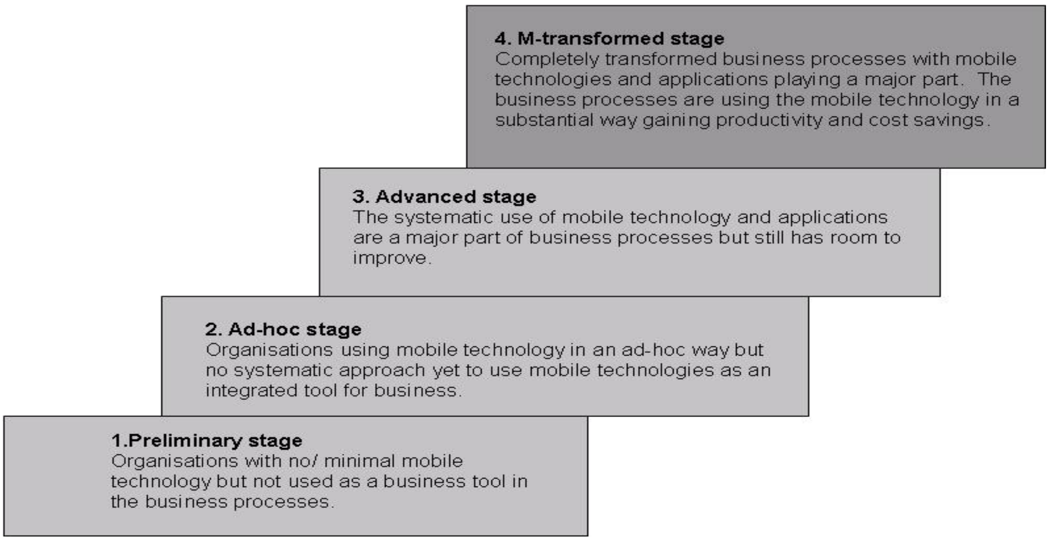


Figure 1 – Scale of four stages in an organisation in m-transformation

Once the stage where the organisation stands with regard to m-transformation is identified, the mobile transition road map could be applied to the business processes. This could be done either organisation wide or department or business unit wise as per the requirements of the particular organisation under investigation.

The application of the mobile transition road map will vary depending on the stage of m-transformation of the organisation. If the Organisation is in the Preliminary stage, the business processes needing to be transformed has to be identified. These business processes have to be prioritized while a survey is done on the existing applications. Similar applications by the competitors and the requests and aspirations of customers and employees also have to be considered in the m-transformation process.

If the organisation is in the ad-hoc stage of m-transformation, the business processes should be further investigated to identify the processes that could be transformed while looking at other processes which are to be introduced as new. The processes which will be made obsolete with new technology also need to be identified.

If the organisation is in the advanced stage in m-transformation, refinement of the m-transformed business processes will be considered. Fine tuning of the business processes while enhancing inter-department, inter-business unit communication using new technology will be looked at in this stage.

At the m-transformed stage, the organisation is fully m-transformed in carrying out its business processes. However, the organisation should still be sensitive to new technology breakthroughs, to further improve the m-transformed processes in terms of better mobile devices, higher bandwidths and increased cost savings. The following section explains the mobile transition road map which will be used as the model to m-enable business processes of an organisation.

MOBILE TRANSITION ROAD MAP

Figure 2 depicts the mobile transition roadmap, a generic model to transform a business organisation into a mobile enabled organisation. This roadmap idea has been adopted from the electronic transformation roadmap (Ginige et. al, 2001). The survey is also to gather data in order to verify this model as per the practical situations in organisations.

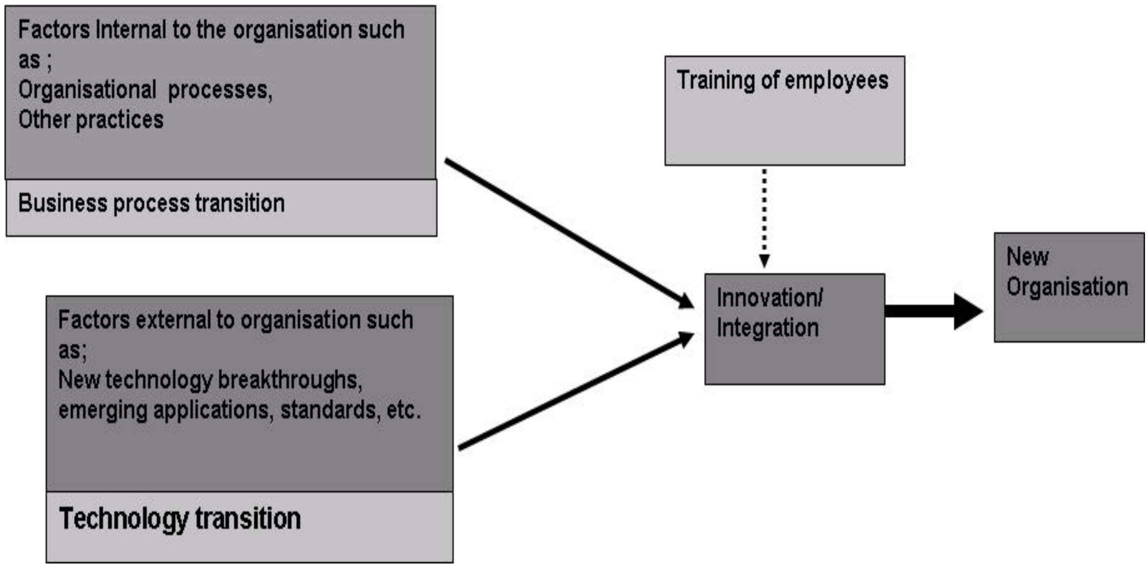


Figure 2: Mobile transition roadmap (Adapted from Ginige et. al, 2001)

The Business process transformation and the Technology transition could be further expanded into different steps. In the business process transition area, factors internal to the organisation are considered. The current operations of the organisation would be looked at in the wake of customers and employees demanding new technology and easier work methods leading to new thinking, in terms of new technology and new work processes.

The technology transition also complements the business process transitions with emerging new technology, new tools and new applications. The establishment of such new technology in the same business sector would add additional pressure for the organisation to effect change in its technology usage. This would lead to new thinking and new business processes to get the full effect of emerging technology. .

The training of employees and the forming of the new look organisation will create a more customer oriented work ethic. The transitioned organisation would be able to follow the well-known objective of being customer-centric by making judicious use of mobile technology. As a result, the service standards, delivery periods, response times and similar attributes would change positively in the organisation leading to a new business culture. Thus, three perspectives namely: technology with regard to emerging technologies, methodology with regard to business processes and sociology with regard to new business culture towards customer orientation are identified in the transitioned organisation.

Figure 3 describes how the three perspectives of technology, methodology and sociology (Unhelkar, 2005), are considered in transforming the organisation in to a more technology savvy and flexible organisation. It further looks at the business processes and ensures that the transformable processes are transitioned to engulf new technology while looking at new business processes to be introduced when the new technology is introduced. This model will also be used when looking at organisations in order to decide how the business processes should be transformed. It would be very important to identify the processes that have to be scrapped all together and the ones that could be transformed and then to identify the new processes which should be a part of the m-transformed organisation.

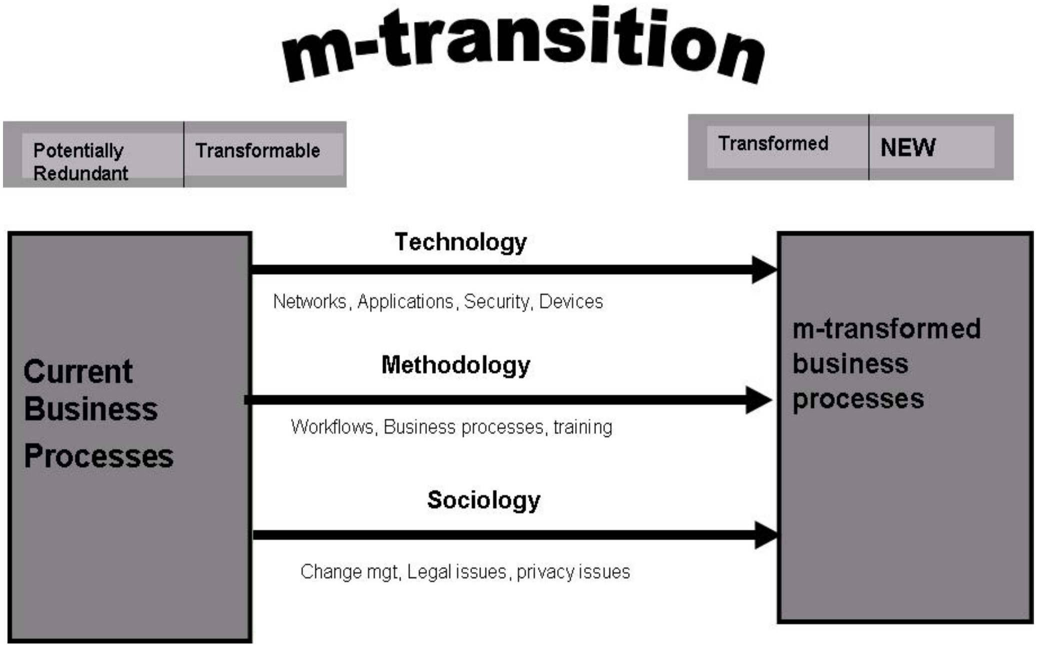


Figure 3 : Mobile transformation of business processes in Technology, Methodology and Sociology perspectives

Finally the ultimate state of the m-transformed organisation would suffice with strongly founded pillars and the business processes which are closely bound together and working together. The perfect m-transformed organization should have a strong foundation on the three perspectives of technology, methodology and sociology. It is pertinent to look at the hypotheses and the research questions that need to be answered in order to verify the models herein discussed. The following section, define the hypotheses and the research questions under consideration.

RESEARCH QUESTIONS AND HYPOTHESES

The main research question under investigation is:

"Does m-transformation of business processes enhance the competitive delivery of service in business?"

In order to probe into this research question we use the hypothesis;

H1 : M-transformation of business processes enhance the delivery of service in business.

To establish the above hypotheses, there are several leading research questions to be answered. The following questions were considered when a banking case study was done in order to identify the m-transition process (Arunatileka, 2006).

- Which stage does the organisation stand in terms of m-transformation?
- How could we apply the m-transition road map at this stage?
- What is the best possible method to apply m-transition road map? Is it organisation wide, Department wise or business unit wise?
- What are business processes that could be transformed in the unit already identified?
- What are the changes happening in the industry where the organisation belongs to in terms of this unit?
- What is the expected impact after mobile technology is introduced to selected business units in the organisation?
- What would be the direct impact on the customers once the new technology is fully implemented?
- Would there be any anticipated problems during the changeover?

After considering these questions the mobile transition road map is applied by identifying the internal business processes and the external enablers in order to systematically transform the organisation. Before looking at the Analysis of the survey lets consider the structure of the questionnaire.

STRUCTURE OF THE QUESTIONNAIRE

The questionnaire is divided into five areas, namely;

- Section 1 - Organisation (Questions 1-3)
- Section 2 - Mobile technology information (Questions 4-11)
- Section 3 - Mobile technology management (Questions 12-13)
- Section 4 - Mobile technology process issues (Question 14)
- Section 5 - Mobile web services (Question 15)

The set of questions on section one is aimed to get an idea of the organisation for classification purposes in the final analysis. These questions are looking at the size of the organisation and the industry in which the organisation belongs to and also the position and the area of work of the person who answers the questionnaire.

Section two is gathering data on mobile technology on the view of the organisation. The questions are probing to see what mobile gadgets/applications are used by the organisation in its day to day operations. This section also aims to establish the perceived advantages for the organisation in using mobile technology. This also probes into the current applications of the organisation using mobile technology and the possible future uses they could have. Further, it investigates how mobile technology could improve the day to day business activities of the organisation. Finally the section also queries whether the use of existing mobile gadgets imposes any problems or difficulties. Data gathered in this section would be used to classify the position of the organisation with regard to m-transformation stage, before applying the roadmap.

The third section is probing into the area of Mobile technology management. In this section, the questions probe into the areas of existing mobile networks and infrastructure and also probing into any general issues such as data integrity, reliability and quality assurance. These set of questions also probe into data processing, office automation and open system standards and whether the current technology poses any issues towards these areas.

Section four of the questionnaire covers the area of mobile technology process issues. This is a very important section since the authors are investigating into the necessary changes that would be required into current business processes of the organisations in order to gain the best competitive advantage of mobile technologies. This section investigates the change in business processes under three perspectives namely technology, methodology and sociology. The technology perspective is looking at the technological changes taking place with the introduction of mobile technology. The methodology perspective investigates the change in the business processes and training and how the overall change management is done in the organisation. The sociology perspective looks at the relationships, such as between the employees, with the customers and the competitors and how they are handled in the m-enabled environment.

The requested sample for this project would be to target organizations which have shown an interest in mobile technologies. Ginige, et al, (2000) grouped their industries for the study on the use of Information Technology on western Sydney area based on the organization's population size and the industry category. In particular we would be aiming at IT managers and CEO's of businesses using mobile technology. The sample size for this project is about 250 questionnaires which would be around 5% of the population of western Sydney businesses when non mobile using organisations are eliminated. This estimate was derived from the Western Sydney IT survey of 2003. In addition there would about another 100 questionnaires from India and Sri Lanka in order to validate data gathered from the main sample area. We believe, this sample size is appropriate for the quantitative answers to make decision based on the statistical methods for deduction and testing the hypotheses. A brief analysis of the survey follows.

ANALYSIS OF THE SURVEY

The purpose of this survey is to establish the hypothesis stated above. The survey will first look at current mobile technologies/applications used in the organisation. The survey will establish if there are issues when mobile applications are used as a tool in the business activities of the organisations.

The participants are also queried on possible shortcomings of the mobile technologies as they perceive in order to establish the general thinking on the technology at present. These shortcomings would be looked at in order to establish the general areas where organisations have issues in using mobile technologies. Then the opinions of the participants are also sought to establish why they think mobile technology is required in their organisations in the context of improving business processes.

The final outcome of the exercise would be to become a perfectly m-transformed organisation where the mobile technologies are used in a determined way to enhance the performance of the organisation and improve the productivity while achieving substantial time and cost savings. The case study carried out by the authors in a leading Australian bank proposing m-transformation of some business processes actually proved this fact. The table 1 and table 2 depict a comparison with the current business processes and the proposed m-transformed business processes in the banking context (Arunatileka, 2006). The two tables are based on two processes, namely the travel process and the customer meeting process of the Finance Manager/International Business Manager (FM/IBM) of the bank, attached to the financial markets division.

Wider issues such as the management of the entire change process along with any legal implications and privacy issues in the new organization have to be worked out. Providing training to employees in these areas is also of utmost importance. In the travel example, the training of FM/IBM about the new processes and any legal issues falling therein in giving the Managers the authority to provide preliminary proposals etc. will fall within the Sociology perspective.

Table 1: Comparison of the current processes with the proposed processes with regard to travel of FM/IBM

Activity/Attribute	Current Process	M-transformed Process
1. Downloading info. Before travel	**Stock/currency indicator updates at office**	**Stock/currency indicator updates at office**
2.Dial up for updates while travel	**Too slow and difficult for updates**	**Mobile dial up for updates** **As frequently as necessary**
3. Liaison with office	**In frequent/ almost nil due to bad lines etc.**	**Always in touch with frequent periodic updates**
4. E mail contacts	**Almost nil**	**Periodic updates every evening with office**
5.Feedback to the customers visited	**Several days since the visit**	**The proposal could be delivered within a day**
6. Timeliness of information on hand	**Could be several days old**	**Only few hours old since last update**

The table 1 shows that considerable time savings could be achieved after m-transformation. There are other measures as well which are significant to consider. The customers who are mostly corporate clients would like to get up to date information.

Table 2: Comparison of the current processes with the proposed processes with regard to Customer Meeting of FM/IBM

Activity/Attribute	Current Process	M-transformed Process
1. Downloading info. Before visiting customer	**Relevant forms are downloaded at office**	**Relevant form downloaded at office is updated with current info just before visit**
2. Verification of Customer Info.	**Makes notes to do corrections**	**Updates and verifies data at**

	later while entering	**current time before providing a draft to customers at his site**
3. Re-Verification of Customer info.	**May be needed to be done via email due to manual entering**	**Not required since it is already done.**
4. Liaison with office	**In frequent/ almost nil due to bad lines etc.**	**Always in touch with frequent periodic updates**
5. E mail contacts	**Almost nil**	**Periodic updates every evening with office**
6.Feedback to the customers visited	**Several days since the visit to enter the info and verifications**	**The proposal could be delivered within a day**
7. Timeliness of information on hand	**Could be several days old**	**Only few hours old since last update**

Table 2 shows the comparison of the current and the proposed processes for Customer Meetings activity. Considerable time savings appear on delivery of proposals etc. in this activity. This also leads to better accuracies as the data is entered at the customer site and verified then and there. Thus considerable monetary savings due to speeded up process and also additional revenue due to customers being signed up earlier than the current system.

The survey data will be analysed to investigate the validity of the hypotheses already presented which is, "M-transformation of business processes enhance the delivery of service in business". This would be used to further verify this proposed m-transformation methodology and the possibility of generalizing the hypotheses into service organisations at large.

CONCLUSION AND FUTURE WORK

Once the survey is sent out and the responses are received, we will be analysing the data as outlined in this paper. A statistical analysis would be conducted including classification of the organisations, how each organisation perceives the existing technology and the areas that need improvement in order to achieve the best results for an m-transformed organisation.

After identifying the stage at which the organization is with regard to m-transformation, further investigations can be carried out to create a practical m-transformation path for the business. Firstly, the business units have to be identified where the m-transformation could take place. Then m-transition road map could be applied in order to see the best possible way to m-transform identifying the external forces such as the technology, competition and business applications. Internal processes that has to be dealt would be the business processes. The identification of business processes that has to be transformed should also be identified to effect necessary changes to them. There would be other business processes which would be made redundant by m-transformation. These redundant processes will be removed while some new processes are identified to fine tune the m-enabled organisation.

If any of the surveyed organisations are willing to take steps in m-transforming their business processes using our methodology, such action could be undertaken as a separate project by the authors. Identification of business processes and their transformation needs, could be proposed after the identification of the stage of m-transformation of the organisation by the use of the models explained in this research work.

References

Arunatileka, D. (2006). "Applying Mobile Technologies to Banking Business Processes." In B. Unhelkar (Ed.), *Mobile Business : Technological, Methodological and Social Perspectives*. Idea Group Publishing, USA.

www.acma.gov.au/ACMAINTER.4849984:STANDARD:1004384283:pc=PC_7126 last accessed 07.12.2005 ; Australian Communications and Media Authority (2005),"Vision 20/20: Future scenarios for the communications industry – Implications for regulation", *Final Report*, April 2005

www.acma.gov.au/ACMAINTER.4849984:STANDARD:1004384283:pc=PC_7126 last accessed 07.12.2005 ; Australian Communications and Media Authority(2003),"Mobile Commerce: Regulatory and Policy Outlook Discussion

Paper", August 2003.

Ginige A., Murugesan S., Kazanis P. (2001), A Roadmap for Successfully Transforming SMEs in to E-Businesses, Cutter IT Journal, May 2001, Vol 14.

Ginige, A., et al, (2000). /"Information Technology in Western Sydney. Status & Potential). /SCIT. University of Western Sydney. ISBN: 1-86341-899-7

Herzberg, A., (2003),"Payments and banking with mobile personal devices", Communications of the ACM, 46(5), pp 53-58

Kalakota, R., and Robinson, M.(2002), "M-Business : The Race to Mobility", McGraw-Hill Professional, New York, USA.

Lan, Y. and Unhelkar, B. (2005), "Global Enterprise Transitions : Managing the Process" Idea Group Publishing, USA.

Mallat, N., Rossi,M. and Tuunainen,V.K. (2004). Mobile Banking Services. Communications of the ACM, 47(5), 42-46.

Murugesan, S..(2005), "Opening Statement", Cutter IT Journal, August 2005, Vol. 18(8).

Nardi, P.M., (2003), "Doing Survey Research : A guide to quantitative methods", Pearson Education Inc. USA ,

Nicopolitidis, P., Papadimitriou,G.,Obaidat,M.S. & Pomportsis,A.S. (2004). The economics of wireless networks. Communications of the ACM, 47(4), 83-86.

Rao, B. and Minakakis, L. (2003). Evolution of Mobile Location-based Services. Communications of the ACM, 46(12), 61-65.

Schilit, B., Hong, J., & Gruteser, M. (2003). Wireless l

ocation privacy protection. Computer, 36(12), 135-137.

Unhelkar, B.(2005), "Transitioning to a Mobile Enterprise : A Three-Dimensional Framework", Cutter IT Journal, August 2005, Vol. 18(8).

www.ecominfo.net/arts/884_mnetics.htm last accessed 08/09/2005

An Exploratory Study of SME Barriers for Adoption of ICT and e-commerce in the Developing Countries - an empirical pilot study of Sri Lanka

Mahesha Kapurubandara
University of Western Sydney
mahesha@cit.uws.edu.au

Abstract

Embracing ICT and e-commerce for stability in international markets and competitive advantage are becoming imperative for Small and Medium Enterprises (SMEs,) to survive in a global economy. Yet, SMEs in developing countries, forming the backbone of the economy, are relatively slow in adopting ICT and ecommerce. Literature reveals many significant reasons contributing towards this reluctance.This paper looks into more in-depth information about the reasons why SMEs in Sri Lanka – a developing country in Asia, are reluctant to adopt ICT and e-commerce technologies. . The barriers were identified through a pilot studyof 17 SMEs carried out in Sri Lanka. It identifies the similarities and differences between the SMEs in developing countries and the developed. The author hopes to develop a methodology to effectively help e-transform SMEs in developing countries.

Keywords:

E-commerce, SMEs, ICT, adoption, developing countries, factors, barriers

INTRODUCTION

The emergence of the Internet has allowed Small- and Medium-sized Enterprises (SMEs) to compete effectively and efficiently in both domestic and international markets (Poon and Swatman, 1999). It is a well-known fact that e-commerce and Internet technologies can benefit an organisation (Akkeren and Cavaye, 1999). Developing countries have the potential to achieve rapid and sustainable economic and social development by building an economy based upon an ICT enabled and networked SME sector capable of applying affordable yet effective ICT solutions (UNDP, 2004)

In their handbook Heeks and Duncombe (2001) discuss the opportunities that Information Communication Technologies (ICT) provide for SMEs in developing countries. SMEs, vital to the economy in any country, are very often recognized as an economy growth engine (Brouthers et al, 1998). They often occupy strategic positions in the economy, positions for which large companies lack the flexibility. The SME sector plays a significant role in its contribution to the national economy in terms of the wealth created and the number of labour employed (Rashid et.al, 2001). With the development of ICT and shift to the knowledge based economy e-transformation and introduction of ICT is becoming an increasingly important tool for the SMEs both to reinvigorate corporate management and promote growth of the national economy (UNDP, 2004).

It is widely accepted that it is important for business to embrace ICT and e-commerce technologies. This is more pronounced in the case of SMEs. The adoption of ICT and e-commerce technologies is vital for their on-going survival. It gives them the competitive advantage and enables them to compete with larger organisations, and to operate on an international scale.

Despite advances in information technology and acceptance by large organisations of such technologies, the same level of adoption is not evident among SMEs (Bode & Burn, 2002; Knol & Stroeken, 2001). This low level of adoption, often impede the SMEs in developing countries in particular, the much needed exposure or leverage. This also suggests that SMEs face significant and unique challenges in adopting ICT and e-commerce (Marshall. N. et.al, 2000).

This paper focuses on the barriers for adoption of ICT and e-commerce more specifically to the SMEs in developing countries. This, it does, with a description of a pilot study carried out in Sri Lanka to investigate the barriers and inhibitors to adopt ICT and e-commerce by SMEs. Sri Lanka was chosen as the test bed as it is a developing country struggling with its economy but on its way to an e-society. The research findings from Sri Lanka could prove to be useful with other developing countries of a similar nature.

This paper will first outline a review of current research about ICT and e-commerce technology adoption. Then a framework summarising current knowledge is presented. A brief description of the research methodology used to carry out the empirical study is given next. This is followed by the analysis of results. The paper concludes with discussions on conclusions, limitations, and future areas of research.

LITERATURE REVIEW

SMEs in Sri Lanka

SMEs are extremely important to the economy of any country. They play a critical role in economic development. This is more pronounced in the case of developing countries. Sri Lanka is no exception.

Different countries use different parameters to define SMEs. Some use the number of persons employed, amount of capital invested, amount of turnover or nature of the business (Gamage, 2003). In Sri Lanka there is no clear definition of an SME as government agencies use various different criteria to define SMEs (Cooray, 2003; Gamage, 2003). The main criteria used are the number of employees, the size of fixed investment, and the nature of the business and the sector (Cooray, 2003). In Sri Lanka, The National Development Bank (NDB), the Export Development Board (EDB), and Industrial Development Board (IDB) use value of fixed assets as the criterion for definition, whereas the Department of Census and Statistics (DCS), Small and Medium Enterprise Development (SMED), and the Federation of Chambers of Commerce and Industry (FDCCI) use the number of employees as the criteria. The World Bank defines enterprise size in Sri Lanka based on the number of employees: those with fewer than 49 employees are small; those with 50-99 employees are medium-sized; and those with more than 100 employees are large.

For this study we consider the following criteria to define the SMEs in Sri Lanka.

Table 1: Definition of Terms: SMEs and MEs (adapted from European Union, 2003;Gamage, 2003)

Micro Enterprises	Within SMEs Category, micro enterprises are enterprises with fewer than 10 employees
Small Enterprises	Small enterprises have between10 and 49 employees
Medium sized enterprises	Medium sized enterprises have fewer than 250 employees

The SMEs in Sri Lanka perform a strategic role, by accounting for a very high percentage of the total number of business establishments. The 2004 mission statement of the International Labour Organisation (ILO) reported that 75% of Sri Lanka's labour force was employed in the SME sector (including the agricultural sector). A more recent World Bank survey on Sri Lanka records approximately one million people working in the manufacturing sector while the Survey of Industries finds around 400,000 working in establishments employing 25 or more workers. These numbers suggest that SMEs contribute substantially to employment and income generation (World Bank Survey, 2003).

While the domestic market is the main outlet for SMEs in Sri Lanka, they make a significant contribution towards exports. Although direct exports from this sector may not be large, SMEs play an important role as indirect exporters. There are a large number of SMEs that manufacture export products or parts, with larger entrepreneurs coordinating such arrangements and handling the direct exports. Coir based products, wood, handicrafts, leather products, plants and foliage are examples of such arrangements involving SMEs which are sub-contracted by large-scale exporters (World Bank Survey, 2003).

An ambition of the Sri Lankan is to enjoy a leading position on the electronic highway in the year 2005. This ambition is implemented in the 'e-Sri Lanka' program, which should provide Sri Lankan SMEs 'a ramp to the digital highway' and stimulate e-commerce and e-business.

Barriers to ICT and E-commerce Adoption in SMEs

This section outlines recent literature on the barriers/inhibitors for adopting ICT and e-commerce by SMEs.

Research works investigating the barriers that affect SMEs adoption of ICT and e-commerce have identified a variety of factors which can be grouped into several categories. A number of authors (eg: Chau, 2001; Mehrtens et al., 2001) identify factors relating to three major categories, owner/manager characteristics, firm characteristics, or to costs and return on investment.

The owner/managers play an important role in decision making in SME organisations. Hence it can be concluded that a number of factors that affect adoption of e-commerce have to do with the owner/manger characteristics. Iacovou et al. (1995) found that the owner's lack of awareness of the technology and perceived benefits is a major barrier to a take up of e-commerce. The lack of knowledge of how to use the technology and low computer literacy are other contributory factors for not adopting e-commerce (Kirby and Turner, 1993). Julien and Raymond (1994) discuss how the owner's level of assertiveness in decision making would affect the adoption of e-commerce. If the owner is subjective and refers to the opinions of experienced people who recommend the adoption of e-commerce into the organisation, then he is also more likely to accept their opinions (Harrison et al., 1997). Mistrust of the IT industry and lack of time are two other factors that affect the decision to adopt e-commerce (Akkeren and Cavaye, 1999).

SME owners are concerned about a return on investments, reluctant to make substantial investments when short-term returns are not guaranteed (Akkeren and Cavaye, 1999).

There are some other factors related to the characteristics of the organisation, which affect adoption of e-commerce. Iacovou et al., 1995 found that the current level of technology usage within the organisation affects process of adoption.

In another study by OECD (1998), it has identified lack of awareness, uncertainty about the benefits of electronic commerce, concerns about lack of human resources and skills, set-up costs and pricing issues, concerns about security, as the most significant barriers of e-commerce for SMEs in the OECD countries.

As summarised by Courtney and Fintz, (2001) there are several other factors that affect the adoption of e- commerce. These are, low use of E-commerce by Customers and Suppliers, concerns about security aspects, concerns about legal and liability aspects, high costs of development and computer and networking technologies towards e-commerce, limited knowledge of e-commerce models and methodologies, and accrued benefits to the company of which the trade is unconvinced.

SMEs have limited resources (e.g. financial, time, personnel). This "resource poverty" has an effect on the adoption of e-commerce. They cannot afford to experiment with these technologies and make expensive mistakes (EBPG, 2002).

Barriers for e-commerce in Developing Countries

It is revealed that less attention with SME e-commerce research has been paid to developing countries with different economic, political, and cultural circumstances. Identifying the differences is an initial step to understand the process of technology adoption. This is particularly important if governments believe that electronic commerce can foster economic development (Martha Garcia-Murillo, 2004).

SME studies of electronic commerce issues in developed countries [Corbitt et al; Huff and Yoong, Mehrtens et al.] indicate that electronic commerce issues faced by SMEs in developed countries can be totally different from those experienced by SMEs in developing countries. Organisations adopting ICT and e-commerce in developing countries face a number of challenges that are specific to them and are more pronounced than would be the case in developed countries. Some of the inhibitors for e-commerce efforts in developing countries are the lack of telecommunications infrastructure, lack of qualified staff to develop and support e-commerce sites, lack of skills among consumers needed in order to use the Internet, lack of timely and reliable systems for the delivery of physical goods, low bank account and credit card penetration, low income, and low computer and Internet penetration (Anigan,1999; Bingi et al.,2000; Panagariya, 2000), Kapuruandara et.al, 2004). Lack of telecommunications infrastructure includes poor Internet connectivity, lack of fixed telephone lines for end user dial-up access, and the underdeveloped state of the Internet Service Providers.

Cultural barriers in some countries may also exist to reject the acceptance of e-commerce as a way of doing business (Bingi et al, 2000). In countries like Sri Lanka, India for instance, shopping is a social activity, and personal, face-to-face contacts with sellers is an important part of the shopping experience. Distrust of what businesses do with personal and credit card information is an e-commerce issue in any country, but, in countries where there may be good justification for such distrust, it could become a serious obstacle to e-commerce growth (Anigan, 1999, Elkin, 2001). Lack of developed legal and regulatory systems also would inhibit the development of e-commerce in developing countries.

Cloete, Courtney, and Fintz (2002) in their study of SME adoption of e-commerce in South Africa, found that adoption is heavily influenced by factors within the organization. There, lack of access to computers, software, other hardware, and telecommunications at a reasonable cost; low e-commerce use by competitors and supply chain partners; concerns with security and legal issues; low knowledge level of management and employees; and unclear benefits from e-commerce were found to be the major factors that inhibit adoption.

Dedrick and Kraemer (2001) in their study of e-commerce in China, found that there are many significant barriers to adopt e-commerce. Limited diffusion of computers, high cost of Internet access, and a lack of online payment processes were found to directly inhibit e-commerce. Inadequate transportation and delivery networks, limited availability of banking services, and uncertain taxation rules indirectly inhibit e-commerce.

El-Nawawy and Ismail (1999) in their study of electronic commerce adoption by SMEs in Egypt found that the main factors contributing to the non-adoption of electronic commerce in Egypt are awareness and education, market size, electronic commerce infrastructure, telecommunication infrastructure, financial infrastructure, legal system, government role, pricing structure, and• social and psychological factors.

Schmid et al. (2001) suggest that the main electronic commerce issues facing SMEs in Argentina are awareness, access to hardware, infrastructure, organisational culture, financial issues. A comparison of the two studies in Argentina and Egypt, (both developing countries) suggests that the key factors of electronic commerce adoption in developing countries are: awareness, telecommunication infrastructure, and cost. It also suggests that SMEs in developing countries share similar issues. The Internet and electronic commerce issues of SMEs in Samoa are consistent with the studies conducted in other developing countries (Schmid et al., 2001; El-Nawawy and Ismail, 1999).

In a study of Sri Lankan SME capability to adopt electronic commerce conducted by the Sri Lankan Business Development Centre (2002), key factors inhibiting the adoption of electronic commerce by SMEs were identified as; lack of knowledge and awareness about the benefits of electronic commerce, current unprepared ness on the part of SMEs to adopt electronic commerce as a serious business concept, lack of exposure to IT products and services, language barrier and lack of staff with IT capability, Web-based selling not seen as practical, limited use of Internet banking and web portals, inadequate Telecommunications infrastructure. Added to these the online payments remain an obstacle as the online credit card payments in Sri Lanka are governed by a limit of US$500 – US$1000. Therefore, it is difficult to purchase high value goods and services online or to purchase large quantities of goods and services online.

The barriers discussed above are summarised in the Table 2.

Table 2: Summary of E-commerce Adoption Barriers

Barriers to E-commerce Adoption	**Reported by**
High cost of E-commerce implementation; Internet technologies too expensive to implement	Iacovou et al (1995); Fielding (1996); Lawrence (1997); Purao & Campbell (1998); Van Akkeren & Cavaye (1999); Riquelme (2002); Quayle (2002)
E-commerce too complex to implement	Fielding (1996) ; Quayle (2002)
Low level of existing hardware technology incorporated into the business	Lawrence (1997)
SMEs need to see immediate ROI and E-commerce is a long-term investment	Lawrence (1997); McGowan & Madey (1998)
Organisational resistance to change because of the fear of new technology amongst employees	Lawrence (1997); Van Akkeren & Cavaye (1999)

Preference for and satisfaction with traditional manual methods, such as phone, fax and face-to-face	Lawrence (1997); Poon & Swatman (1999); Venkatesan & Fink (2002)
Lack of technical skills and IT knowledge amongst employees; Lack of computer literate/specialised staff	Iacovou (1995); Lawrence (1997); Damsgaard & Lyytinen (1998); Van Akkeren & Cavaye (1999); Quayle (2002); Riquelme (2002); Chau & Turner (2002)
Lack of time to implement E-commerce	Lawrence (1997); Van Akkeren & Cavaye (1999); Walczuch et al (2000)
E-commerce is not deemed to be suited to the way the SME does business	Iacovou et al (1995); Abell & Limm (1996); Poon & Swatman (1997); Hadjimanolis (1999);
E-commerce is not deemed to be suited to the products/services offered by the SME	Hadjimanolis (1999); Walczuch et al (2000); Kendall & Kendall (2001)
E-commerce is perceived as a technology lacking direction	Lawrence (1997)
Lack of awareness about business opportunities/benefits that E-commerce can provide	Iacovou et al (1995); Quayle (2002)
Lack of available information about E-commerce	Lawrence (1997)
Concern about security of E-commerce	Abell and Limm (1996); Purao & Campbell (1998); Hadjimanolis (1999); Van Akkeren & Cavaye (1999); Poon & Swatman (1999); Quayle (2002); Riquelme (2002)
Lack of critical mass among customers, suppliers and business partners to implement E-commerce	Abell and Limm (1996); Hadjimanolis (1999)
Lack of awareness of perceived benefits	(Iacovau et al, 1995; Kirby and Turner, 1993).
Lack of awareness of the technology	(Kirby and Turner, 1993; Thong and Yap, 1995).
Lack of Assertiveness	(Julien and Raymond, 1994; Harrison et al, 1997)
subjective norm	(Harrison et al, 1997).
Lack of Organisational readiness	(Iacovau et al, 1995).
external pressure to adopt IT	(Thong and Yap, 1995; Iacovau et al, 1995).
Dependency on customer/supplier	(Kirby and Turner, 1993)
cost of adopting / using IT short to medium term return	Akkeren. J, Cavaye A.L.M, (1999).
The lack of knowledge of how to use the technology, low computer literacy	. (Kirby and Turner, 1993).
Mistrust of the IT industry, lack of time	(Akkeren and Cavaye, 1999).
lack of awareness, uncertainty about the benefits of electronic commerce, concerns about lack of human resources and skills, set-up costs and pricing issues, concerns about security	(OECD, 1998)

Low use of E-commerce by Customers and Suppliers. Concerns about security aspects, Concerns about legal and liability aspects, high costs of development, Limited knowledge of e-commerce models and methodologies. Unconvinced of benefits to the company.	(Courtney and Fintz, 2001)
limited resources (e.g. financial, time, personnel	(EBPG, 2002).

The above literature survey on the SME barriers for e-commerce adoption reveal that there are many significant factors which affect the adoption of e-commerce technologies. These factors can be grouped to develop a framework for investigations. This paper proposes the following framework for this purpose. The barriers for the SMEs in adopting ICT and e-commerce can be broadly categorised into Internal and External issues.

Internal: SME has control over and the ability to change the internal factors within the organisation. Eg. lack of time or resources, lack of awareness on the part of the owner/ manager. Internal barriers could be further categorised into Individual (owner/manager) and Organisational (firm) barriers. These can be addressed

External: Barriers that cannot be resolved by the SME Company. They have no control over these, and are compelled work within the constraints. Eg. Inadequate telecommunication infrastructure etc. Some of these barriers could be addressed by SMEs working together. The SMEs can get together irrespective of the industry sector they are in, to form clusters to share the expenses, resources, facilities etc. or the SMEs from the same industry sector can work together. To address certain other external barriers governmental intervention may be required.

The pilot study reported in this paper, attempts to provide more in-depth information about the factors/barriers in the context of SMEs in Sri Lanka. By identifying the major barriers the second stage of this study proposes to develop a methodology that would address the internal factors and would work within the external constraints.

RESEARCH METHODOLOGY

As very little research has been carried out on SMEs in Sri Lanka, we chose an exploratory pilot study to help find out the barriers and inhibitors to adoption of ICT and e-commerce among the SMEs. Face to face interviews are most suitable for this type of exploratory study because the researcher can adapt the questions as necessary, clarify doubts, and ensure that the responses are properly understood (Sekaran, 2000). This is of particular importance as the potential interview participants may not have a clear understanding about ICT, e-commerce and its applications. Furthermore, face to face interviews allow the researcher to explore issues raised by the respondents, which generally is not possible through questionnaires or telephone interviews.

This study was carried out in July 2005, through a face to face, semi structured interviews of 17 SMEs. This random sample which was chosen from a list maintained by the Export Development Board of Sri Lanka is representative of SMEs from various industry sectors, engaged in different businesses in the western province of Sri Lanka. The western province was selected as this province boasts the highest density of companies using ICT. No restriction based on type of industry was made. It was also important that the organisation had adopted at least one of the following parts of ICT: phone, fax, packages, and basic office packages (Word, Excel etc), Internet searching and browsing, e-mail, or a Web site. The interview guide line was developed mainly based on primary literature and the proposed framework.

The owner/managers were interviewed as they are the decision makers of the SME organisations. They were interviewed to get first hand information about their use of ICT and e-commerce, their perceptions and the barriers for adopting it. Each interview lasted nearly one hour. Each of the interviews was recorded, subsequently transcribed and subjected to qualitative content analysis. This involved categorisation of responses into certain themes, some of which were inherent in the interview questions, and some of which emerged through the interviewees' responses to questions posed.

DISCUSSION OF RESULTS

87% percent of the participating companies were categorized as service companies, and the remaining 23% as manufacturing companies. 35 % of the participants have a static website mainly used for advertising purposes. Most of these Websites are not updated regularly. None of the participants uses the Website for buying or selling. 767% of the participants have Internet and e-mail facilities in their organisations. 52 % use the Internet for browsing and searching, and also use e-mail as the preferred medium of overseas communication. 62% of them have ADSL connections and 25% dial-up connections. While majority of participants have Internet access, only 43% allow every user to access the Internet.

The following factors which surfaced as barriers for adoption of ICT and e-commerce are categorised according to the framework discussed.

Internal *Individual/ owner manager*	Lack of awareness about business opportunities/benefits that E-commerce can provide, Lack of available information about E-commerce, Lack of awareness of the technology, Lack of knowledge and awareness of available systems, Lack of access to expert help and reliable advise, lack of time and resources, E-commerce is not deemed to be suited to the products/services offered by the SME , SMEs need to see immediate ROI and E-commerce is a long-term investment, E-commerce is perceived as a technology lacking direction, Lack of Assertiveness, Concern about security of E-commerce, Fear of expansion, Unconvinced of benefits to the company, subjective norm, Lack of trust of the technologies, lack of skills
Internal *Firm*	Lack of technical skills and IT knowledge amongst employees, Lack of time to implement E-commerce, Lack of Organisational readiness, E-commerce is not deemed to be suited to the way the SME does business, customers, suppliers not connected (channel conflict), Lack of staff, time to investigate technology and new systems, Low use of E-commerce by Customers and Suppliers, lack of awareness of IT and minimal use of it, In-house IT expertise, Language barriers, Company culture, Lack of computer literate/specialised staff , Low level of existing hardware, technology incorporated into the business, Preference for and satisfaction with traditional manual methods such as phone fax and face-to-face , Organisational resistance to change because of the fear of new technology amongst employees, Limited knowledge of e-commerce models and methodologies
External	Costly internet, Inadequate telecom infrastructure, limited teledensity, unreelable telecommunications and power supply, Limited band width, Unstable economy, Cultural and linguistic differences, Political uncertainty, High cost of E-commerce implementation, High cost of telephone facilities, Insufficient number of available telephone lines

Lack of awareness is the most important barrier, cited by almost 88% of respondents. This is significant for majority of owner/managers, who described themselves as having basic computer literacy but had no idea what technologies were available, how e-commerce technology could benefit their businesses or if it would suit their business needs. It was a major concern, as they are the decision makers of their organisation. A process of education and training of the owner/manager can overcome these barriers. Education and training is viewed as being crucial to addressing the lack of readiness of SMEs in adopting and developing their electronic business capabilities (Chau, 2001).

The second most cited reason that could inhibit the adoption of e-commerce is the cost. Respondents in our study were concerned about the cost of internet, cost of equipment and cost of e-commerce implementation. SMEs who are poor in resources cannot afford to experiment with these technologies and make expensive mistakes (EBPG, 2002). It may be possible to overcome this by working together and forming clusters.

The inadequate telecom infrastructure was the third most frequently cited barrier, chosen by 83% of respondents. Respondents that chose this barrier were more advanced in their usage of ICT and used e-mail and internet hence may be more likely to have experienced this problem.

The remaining barriers was cited by less than 70% of respondents. Even though they are considerably important, it was evident that the above three are amongst the key factors affecting adoption. These support the findings of previous research and also suggests that SME's in developing countries share similar issues (Schmid et al.; El-Nawawy and Ismail; Boalch and Bazar). It also indicate that these key issues are totally different from that of developed countries (Corbitt et al.,1997; Huff and Yoong,2000; Mehrtens et al,2001.; Poon and Swatman,1999) where organisational readiness and external pressure is considered as two of the main factors.

Thus the author's research has identified the existence of significant barriers to adoption of ICT and e-commerce for SMEs, which is supported by academic literature on the topic. However, this requires thorough and further investigation of the barriers, recognition of the significance of each barrier and a methodology to alleviate them.

CONCLUSION:

Most participant SMEs in the pilot study are in the early stages of ICT and e-commerce adoption and the majority expressed their eagerness to embrace it further. As commonly found in previous studies (Schmid et al., 2001; El-Nawawy and Ismail, 1999) in developing countries, lack of awareness appears to be amongst the major barriers of SMEs in adopting e-commerce. They were very keen to be aware and knowledgeable about the benefits ICT and e-commerce can bring to their organisations, how and where ICT and e- commerce can be used. We found that most SMEs are aware the market will become more dynamic, competitive, and global, and that there is an increased need to use the ICT and e-commerce in their business.

There has been a lack of research undertaken in Sri Lanka to determine SMEs' barriers for adoption of ICT and e-commerce. This pilot study has shown that while the SMEs agree that adoption is essential in today's business, they are hindered from adoption as they are plagued with many constraints, some of which are more specific to developing countries like Sri Lanka. The next stage of the study outlined in this paper is to investigate some of the issues that have emerged in more detail to analyse their significance further. This pilot study will serve as a framework and a guideline in designing further research. Furthermore, it is felt that there is a need to find methodologies which address these barriers to help successfully e-transform the SMEs in developing countries.

This paper contributes towards knowledge in the following areas. Firstly, it reviews the existing barriers for SMEs in developing countries found in various literatures. Secondly, it categorizes barriers in two ways: factors that can be solved within the organisation and the factors that cannot be resolved within the organisation. Finally, the paper recognises that the study requires further investigation to find solutions to the research question in discussion.

REFERENCES:

Abell, W. & Limm, L. (1996). Business Use of the Internet in New Zealand: An Exploratory Study. Proceedings of AUSWeb. URL: Retrieved August, 2004

Anigan, G. (1999). "Views on Electronic Commerce." International Trade Forum, Issue 2, 23–27.

Bingi, P., A. Mir, and J. Khamalah. (2000). "The Challenges Facing Global E-Commerce." *Information systems Management* 17(4), 26–35.

Boalch, G. and B. Bazar. (1997). "A Preliminary Model of Internet Diffusion in Developing Countries." In Proceedings of the Third Australian World Wide Web Conference, Southern Cross University, Australia, 5–9 July 1997.

Bode, S & Burn, JM, 2002, 'Strategies for consultancy engagement for e-business development -a case analysis of Australian SMEs', in Managing Information Technology in Small Business:challenges and solutions, ed. S. Burgess, Idea Group, Melbourne, Australia, pp. 227-245.

Chau, S.B. & Turner, P. (2001). A Four Phase Model of EC Business Transformation Amongst Small to Medium Sized Enterprises. Proceedings of the 12th Australasian Conference on Information Systems. Coffs Harbour: Australia.

Charalambos L. Iacovou , Izak Benbasat , Albert S. Dexter, Electronic data interchange and small organizations: adoption and impact of technology, MIS Quarterly, v.19 n.4, p.465-485, Dec. 1995

Cloete, E., S. Courtney, and J. Fintz (2002). "Small Businesses' Acceptance and Adoption of E-Commerce in the Western-Cape Province of South-Africa." *Electronic Journal on Information Systems in Developing Countries* 10(4), 1–13, http://www.ejisdc.org.

Cooray, M.N.R. (2003) *Walk through Cleaner Production Assessment in SME's - A Case*

Study from Sri Lanka, Small to Medium Enterprise Developers, http://www.aprcp.org/roundtables/4th/Papers/Cooray_W05.pdf.

Corbitt, B., G. Behrendorf, and J. Brown-Parker. (1997). "Small- and Medium-Sized Enterprises and Electronic Commerce." *The Australian Institute of Management* 14, 204–222.

Courtney, S. and Fintz, J. (2001) Small Businesses' Acceptance and Adoption of e-

Commerce in the Western-Cape Province of South-Africa, Empirical Research Project, Department of Information Systems, UCT.

Damsgaard, J. & Lyytinen, K. (1998). Contours of Diffusion of Electronic Data Interchange in Finland: Overcoming Technological Barriers and Collaborating to Make it Happen. Journal of Strategic Information Systems 7, 275-297.

Dedrick, J. and K.L. Kraemer (2001). "China IT Report." *Electronic Journal on Information Systems in Developing Countries* 6(2), 1–10, ttp://www.ejisdc.org.

Domaracki, G.S. (2001) The Dynamics of B2B e-Commerce, AFP Exchange, **21**, 4, 50-57.

Duncombe, R., & Heeks, R. (2001). Handbook for Entrepreneurs in Developing Countries, University of Manchester, UK: IDPM.

EBPG (2002). eEurope go digital: Benchmarking national and regional e-business policies for SMEs. Final report of the EBusiness Policy Group, 28 June 2002.

Elkin, N. (2001). "Online Privacy and Security in Latin America."

El-Nawawy, M.A. and M.M. Ismail. (1999). "Overcoming Deterrents and Impediments to Electronic Commerce in Light of Globalisation: The Case of Egypt." In *Proceedings of the 9th Annual Conference of the Internet Society, INET 99*, San Jose, USA, 22–25 June 1999.

Fielding, J. (1996). Getting Out in Front with EDI. Inform 10(9), 12-18.

Gamage, A.S. (2003) Small and Medium Enterprise Development in Sri Lanka*:* A Review, http://wwwbiz.meijo-u.ac.jp/SEBM/ronso/no3_4/aruna.pdf

Hadjimonolis, A. (1999). Barriers to Innovation for SMEs in a Small Less Developed Country (Cyprus). Technovation 19(9), 561-570.

Harrison, D.A., Mykytyn, P.P. and Rienenschneider, C.K. (1997) Executive Decisions About IT Adoption in Small Business: Theory and Empirical Tests, Information Systems Research, **8**, 2, 171-195.

Huff, S. and P. Yoong. (2000). "SMEs and E-commerce: Current Issues and Concerns. A Preliminary Report." In *Proceedings of the International Conference on E-commerce*, Kuala Lumpur, Malaysia, 21 November 2000, pp. 1–5.

Iacovou, C.L., Benbasat, I., and Dexter, A.A. (1995) Electronic Data Interchange and Small Organisations: Adoption and Impact of Technology, MIS Quarterly, **19**, 4, 465-485.

James T. Perry , Gary P. Schneider, Electronic Commerce, Course Technology Press, 2001

Julien, P.A. and Raymond, L. (1994) Factors of New Technology Adoption in the Retail Sector. Entrepreneurship: Theory and Practice,**18**, 5, 79-90.

Kapurubandata ,M., Arunatileke S., Ginige, A.,(2004) Application of eBusiness Strategies for SMEs in Developing Countries, IEEE 2004, pp 49-59**,** Taipei, Taiwan

Kirby, D. and Turner, M. (1993) IT and the Small Retail Business, International Journal of Retail and Distribution Management, **21**, 7, 20-27.

Kendall, J.E. & Kendall, K.E. (2001). A Paradoxically Peaceful Coexistence Between Commerce and Ecommerce. Journal of Information Technology, Theory and Application 3(4), 1-6.

Knol, WHC & Stroeken, JHM 2001, 'The diffusion and adoption of information technolo gy in small- and medium –sized enterprises through IT Scenarios', *Technology Analysis & Strategic Management,* vol.13, no.2, pp. 227-246.

Lawrence, K.L. (1997). Factors Inhibiting the Utilisation of Electronic Commerce

Facilities in Tasmanian Small to Medium Sized Enterprises. Proceedings of the 8th Australasian Conference on Information Systems 587-597.

MacKay, N., Gemino, A., Igbaria, M. & Reich, B. (2001). The Impact of electronic commerce: A survey of small and medium–sized enterprises in British Columbia. Research Results, Simon Fraser University, British Columbia.

Marshall, P., Sor, R. & McKay, J. (2000). An industry case study of the impacts of electronic commerce on car dealerships in Western Australia. Journal of Electronic Commerce Research, 1, 1, February 2000.

Martha Garcia-Murillo9 2004), Institutions and the Adoption of Electronic Commerce in Mexico). ELECTRONIC Commerce Research, 4, 201-219

McGowan, M.K. & Madey, G.R. (1998). The Influence of Organization Structure and Organizational Learning Factors on the Extent of EDI Implementation in U. S. Firms. Information Resources Management Journal 11(3), 17-27.

Mehrtens, J., Cragg, P.B. & Mills, A.M. (2001). A model of Internet adoption by SMEs. Information and Management, 39.165-176.

OECD (1998). SMEs and electronic commerce, Ministerial Conference on Electronic Commerce, Ottawa, Canada, 7-9 October 1998.

Panagariya, A. (2000). "E-Commerce, WTO, and Developing Countries." *World Economy* 23(8), 959–979.

Poon, S. & Swatman, P. (1997). The Internet for Small Businesses: An Enabling Infrastructure. Fifth Internet Society Conference 221-231.

Poon, S. & Swatman, P.M.C. (1999). An Exploratory Study of Small Business Internet Commerce Issues. Information & Management 35, 9-18.

Purao, S. & Campbell, B. (1998). Critical Concerns for Small Business Electronic Commerce: Some Reflections Based on Interviews of Small Business Owners.

Proceedings of the Association for Information Systems Americas Conference. Baltimore, 325-327.

Quayle, M. (2002). E-commerce: the Challenge for UK SMEs in the Twenty-first Century. International Journal of Operations and Production Management. 22(10), 1148-1161.

Riquelme, H. (2002). Commercial Internet Adoption in China: Comparing the Experience of Small, Medium and Large Businesses. Internet Research: Electronic Networking Applications and Policy 12(3), 276-286.

Schmid, B., K. Stanoevska-Slabeva, and V. Tschammer. (2001). "Towards the E-Society: E-Commerce, E-Business, E-Government." Zurich, Switzerland, 13 October.

Sekaran. U. (2000), Research Methods for Business – A Skill Building Approach, 3rdEdition, John Wiley & Sons, Inc.

Thong, J. & Yap, C. S. (1995) CEO characteristics, organizational characteristics, and information technology adoption in small business. Omega, August, 23 (4), 429-442.

Van Akkeren, J. & Cavaye, A.L.M. (1999). Factors Affecting Entry-Level Internet Technology Adoption by Small Business in Australia: An Empirical Study. Proceedings of the 10th Australasian Conference on Information Systems. Wellington, New Zealand, 1-3 December.

Venkatesan, V.S. & Fink, D. (2002). Adoption of Internet Technologies and E-Commerce by Small and Medium Enterprises (SMEs) in Western Australia. Proceedings of the Information Resource Management Association International Conference 1136-1137.

Walczuch, R., Van Braven, G. & Lundgren, H. (2000). Internet Adoption Barriers for Small Firms in The Netherlands. European Management Journal 18(5), 561-572.

European Union (2003) Communication From The Commission: Sample Statement On The Information Relating To The Status Of An Enterprise As A SME.

Asia Foundation (2002) Survey on E-Commerce Implementation in the SME Sector of Sri Lanka, Sri Lanka Business Development Centre, www.asiafoundation.org/pdf/SLreport.pdf.

Evolution of eBusiness Adoption in the Western Sydney Region

Ms. Ana Hol, Prof. Athula Ginige, Dr. Robyn Lawson
University of Western Sydney, AeIMS Research Group
Sydney, Australia

Email: a.hol@uws.edu.au, a.ginige@uws.edu.au, r.lawson@uws.edu.au

Abstract

The research topic takes up the issue of group dynamics within a Global Information System, and a relating question of human rationality. The interests in the area of information systems have concentrated mainly on technological aspects so far. If the human component were taken into account, it has been analyzed from the level of an individual. So have all new concepts of rationality. The detailed project objectives arise from the following research questions: Is collective behavior a phenomenon of exclusively qualitative nature, impossible to structure? Does irrationality of individual actions determine irrationality and unpredictability of the whole social system? A model exemplification of a Global Information System is a modern, electronic stock exchange. The anticipated findings should help formulate a new, adaptive, paradigm of rationality.

Keywords:

Global Information System, Global IS dynamics, IS social subsystem, group behavior, rationality.

INTRODUCTION

Western Sydney is considered one of the fastest growing regions in Australia. Currently, about 1.8 million people, representing 43% of Sydney's population reside within the region. The region has experienced the highest rate of population growth of any region in Australia over the past decade (0.9%). It is Australia's third largest economic region which generates over $58 billion a year, and turnover represents 10% of the Australian Economy (Western Sydney Budget Statement 2003-2004). Correspondingly, over the last 20 years, the region's population growth has been over one-and-a-half times that of the whole Sydney Statistical Division. Today, more than 72,000 businesses are located in the region. In addition, 150 of Australia's top 500 companies are located in the region, which has considerably enhanced the region's growth. It is expected that by 2015, major expansion in the region will see the arrival of 220,000 new residents, 100,000 additional workers and an expected gross turnover of $80 billion (Western Sydney Budget Statement 2003-2004).

World wide, businesses are faced with intense and open competition in the quest to meet the new demands and challenges of the global market. To meet the demands, businesses are compelled to implement ICT to become globally competitive. The use of ICT has changed the way businesses in today's information society operate, and has significantly decreased the cost of transactions. It has also influenced how tasks within organizations are performed, which has given rise to business re-engineering, and the emergence of new more efficient business processes. Moreover, pricing models have changed noticeably as manufacturers or designers can now easily directly deliver goods and services. In today's society, electronic means of communication and the presence of web portals have enabled businesses to function without the need for distributors or retailers. Furthermore, new pricing and business models have made an impact on production and services. Mass production and mass marketing have been changed and modified in line with rising demands and needs of customers, resulting in mass customization and niche marketing. Consequently, ICT has enabled these changes, including the ability for fast processing of small orders. Therefore, the effect of economy of scale has been decreased.

As a result, the new dynamic business environment strengthens the importance of knowledge, virtualization, innovation, immediacy and globalization. The properties and characteristics of the new environment are constantly changing, requiring organizations to be able to sense and respond without delay. They also need to be able to change quickly to remain competitive in the global arena (Tapscott, 1996). The use of ICT to enable businesses to embrace the new global competitive environment using eBusiness technologies requires more than the application of the various technologies (Lawson et al, 2003). A whole organization approach is needed that examines the internal as well as the external processes.

This paper reports on a study that investigated adoption of ICT among businesses in Western Sydney. In addition, it highlights the status of ICT infrastructure, and identifies skills and requirements needed for today's businesses. Previous research indicated that ICT could be an important enabler for business growth, as it can easily open new business avenues and frontiers (Brynjolfsson & Hill, 2000). This study analyses results from two major surveys carried out by researchers at the University of Western Sydney (UWS) in 2000 (Ginige et al. 2001) and 2003 (Khandelwal et al. 2003). These surveys examined how organizations used ICT to carry out their business operations in 2000 and 2003. Further analysis was undertaken, which compared results from the surveys with the Australia Bureau of Statistics surveys (ABS, 2004).

DATA COLLECTION

In 2000 and 2003 the Advanced Enterprise Information Management Systems (AeIMS) Research Group of the UWS conducted two surveys titled "Information Technology: Status and Potential". The aim of the two surveys was to establish the level of ICT usage, and determine the growth and potential of the region. In addition, the study assessed ICT skills, and examined the requirements and needs of the businesses. It was expected that the results of the two surveys would be helpful to business, government and industry sectors across the region. Additionally, it was expected that the surveys would identify business readiness to take up ICT and pinpoint issues in ICT adoption, use and acceptance.

A number of research data collection techniques were evaluated, such as structured interviews (Bullen & Rockart, 1981), focus groups, group interviews (Khandelwal, 1992) and survey approach (Galliers et al., 1994). As the objective was to collect a large amount of data, the survey method was selected. Limitations were identified and steps taken to ensure a maximum response rate. Each survey was mailed with a covering letter personally addressed to company's Chief Executive Officer (CEO). Furthermore, to ensure the quality of the responses, the letter indicated that one or more appropriate individuals (Head of Organisation, CEO, IT Manager, and Production Manager) could complete the survey.

Using standard surveying procedures a sample of the region's businesses was selected. The sample covered 19 industries and 14 local government areas. Adaptive sampling (Thompson, 1992) was used to assure appropriate businesses and parts of the regions were covered in the surveys. In total 590 (13.5%) surveys were returned in the 2000 Survey (Ginige et. al. 2001) and 307 (7.2%) in 2003 Survey (Khandelwal et al, 2003).

To meet the objectives of the survey a range of questions were developed. The questions were clustered into the following groups to form the questionnaire:

- Details of the organisation – included general company information such as location, size, industry and information about main business processes.
- IT infrastructure – included questions about staff, their skills and knowledge, current and future ICT trends, general hardware and software information such as a number of computers used, types of productivity tools packages and applications.
- Network infrastructure – included questions about the organisations' networking capabilities and infrastructure.
- Internet infrastructure – included questions about the Internet, its use, connections and applications and electronic transactions.
- Website and eCommerce capabilities – included questions about websites, its development, hosting and potential site marketing. In addition, respondents who already had a website were asked to state what online services they were providing.
- Future Plans – this section looked at views and plans of the company including ICT visions and initiatives.

In addition to the above question clusters, an extra cluster was included to determine if the respondents were in ICT industries within the region that supported ICT implementation and maintenance.

eTRANSFORMATION ROADMAP

To analyse, compare and measure adoption of ICT among businesses within the region, the eTransformation Roadmap was developed from results of the 2000 Survey (Ginige et al, 2001). The roadmap is depicted in Figure 1. Subsequently, the eTransformation Roadmap was used to analyse the results of the 2003 Survey (Khandelwal et al, 2003) and compare the adoption of ICT with the 2000 Survey (Ginige et al, 2001).

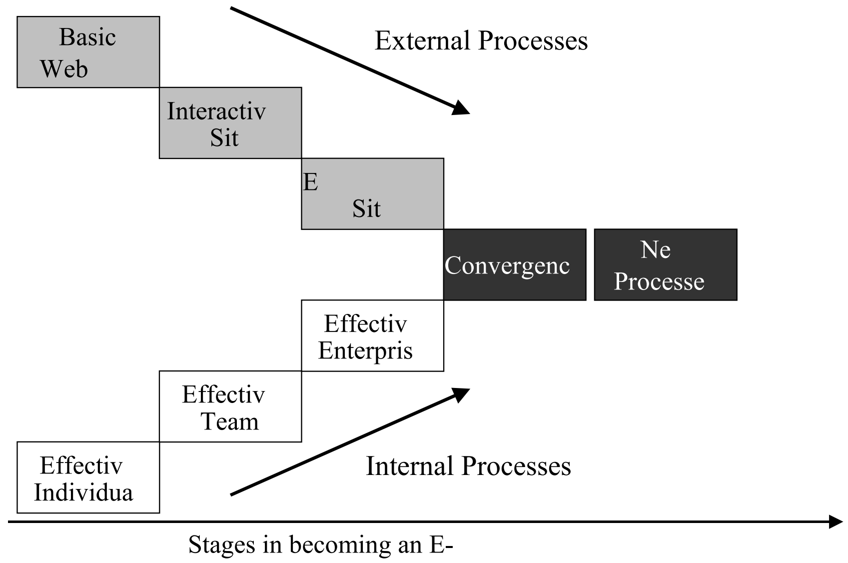


Figure 1: eTransformation Roadmap

The eTransformation Roadmap indicates that there are two major types of processes (internal and external) for a typical 'bricks and mortar' company that influence eTransformation. Furthermore, the map can help explain various stages through which a business needs to pass through to successfully eTransform. The stages are detailed in Table 1.

eTransformation Stages and Accompanying Technologies

Stage	**Hardware, Software and Networking infrastructure**
Effective Individual	Individuals with stand-alone computers having productivity software such as accounting packages spreadsheets, word processors etc installed. Possibly dial up connection to the Internet for individual email accounts.
Effective Team	Computers are networked. People can work in teams using networked applications. Providing email and Intranet capabilities can enhance productivity of a team.
Effective Organisation	Organisation now uses enterprise wide applications; a single application that supports different sections of the organisation such as purchasing, sales, accounting, manufacturing etc. thus enabling information integration and sharing across enterprise.
Basic Web Site	Organisation having its own domain name and "Brochure ware" type web site hosted with an ISP.
Interactive Web Site	Organisations having Web sites that provide two-way flow of information. From these Web sites uses should be able to get immediate responses to structured queries such as a quotation for a particular product configuration user has selected. Also another feature would be to provide personalised information to frequent visitors. These types of Web sites though can be hosted with an ISP are better to host on site. This requires a Web server and a high-speed dedicated connection to the Internet.
E Commerce Site	At this stage the organisation should have secure Web servers to facilitate financial transactions or a link to a payment get way to get this facility.
Convergence	Organisation has now achieved integration of information that needs to support all its business processes. The flow of information of an organisation that has reached the level of convergence.

Table 1: eTransformation Stages and Accompanying Technologies (Ginige, et al, 2001, pg. 7)

DATA ANALYSIS

Data collected from the 2000 and 2003 surveys were compared and analysed based on the eTransformation Roadmap. Results of this analysis are shown in Table 2

Use of IT to enhance Internal and External Business Processes

Stage	Business at a Given Stage	2000	2003
Effective Individual	Businesses that have computers	88%	97%
	Businesses that use productivity tools for:		
	Accounting	58%	65%
	Payroll	28%	30%
	Inventory and stock control	5%	3%
	Web Design	3%	28%
Effective Team	Communication Infrastructure: Organizations that:		
	Have LANs	54%	63%
	Are connected to the Internet	56%	70%
	Have email facility	47%	74%
	Have intranets	8%	30%
Effective Organization	ERP-, MRP-type applications	11%	15%
	CRM		13%
Basic Web Site	Organizations that have Web sites	41%	59%
	Web sites that are hosted:		
	In-house	25%	20%
	With an ISP	71%	77%
	Not specified	4%	3%
Interactive Web Site	Websites that have:		
	Online electronic catalog	18%	44%
	Handle customer enquiries	30%	71%
eCommerce Site	Websites that have online capabilities to:		
	Receive orders	10%	19%
	Provide after-sales service	6%	18%
	Offer follow up service	4%	14%
	Track order progress	2%	5%
	Deliver of goods (information products)	1%	8%

Table 2: Survey 2000 and 2003 Results

From Table 2 it can be noted that between 2000 and 2003 considerable changes have occurred for both internal and external business processes. The 2003 data indicates that 97% of companies have computers compared to 88% of companies in 2000. In addition, the 2003 data shows that companies are making use of various applications and tools particularly those that are required for accounting at 65%, compared to 54% in 2000. Furthermore, in 2003 59% of companies had websites compared to 41% in 2000.

Data indicates that in 2003 almost all companies have reached the “Effective Individual” Stage by the use of computers, and a range of productivity software. It was also found that majority of companies have reached the “Basic Website” Stage with web sites or plans for the development of web sites.

The “Effective Team” and “Interactive Site” Stages seem to have been concern for companies in 2003, as they were looking into developing and enhancing their communication infrastructures and networks. For example, in 2003 74% of organisations were communicating via e-mails compared to only 47% in 2000. Also, the number of organisations with Intranets has increased from 8% in 2000 to 30% in 2003. In addition, in 2003 companies were making use of their websites with the introduction of online catalogues and the handling of customer enquiries. For example, in 2000 only 18% of companies had an electronic catalogue while in 2003 44% of companies had introduced catalogues. The number of online customer enquiries sharply increased from 30% in 2000 to 71% in 2003.

Results indicate that not many companies have reached the "Effective Organisation" Stage or "eCommerce" Stage. Only a very few companies have introduced Enterprise Resource Planning in 2003 (15%) and Customer Relationship management tools (13%). Data collected about "eCommerce" Stage indicated that a few companies have started receiving orders electronically. For example 10% of companies were doing this in 2000 while in 2003 the figure only rose to 19%. Data corresponding to payment options and transaction processing are still not significant enough to be reported. However, increases in these types of activities are expected to increase in the future.

DATA COMPARISON WITH ABS STUDY

For further analysis of the region, the data from the two surveys was compared with data from the Australian Bureau of Statistics yearly survey on "Business Use of Information Technology". The ABS mailed their survey to a random sample of Australia-wide registered businesses. The random sample consisted of approximately 9,000 businesses across various industry sectors, states and territories, business sizes and type, and number of employees. The response rate for their study was approximately 93%. The ABS warn that due to the nature of their study, the gathered data should be taken with caution as results may be indicative of a particular time or an event prior to the data collection (ABS, 2004). The comparison of the data from the UWS 2000 and 2003 surveys with the ABS data is represented in the Table 3.

<table>
<tr><th></th><th>Processes</th><th>UWS 2000 Survey</th><th>UWS 2003 Survey</th><th>Growth</th><th>ABS 2000</th><th>ABS 2003</th><th>ABS Growth</th></tr>
<tr><td rowspan="7">Internal Processes</td><td>Use of Computers</td><td>89%</td><td>97%</td><td>8%</td><td>84%</td><td>83%</td><td>-1%</td></tr>
<tr><td>Accounting Packages</td><td>58%</td><td>65%</td><td>7%</td><td></td><td></td><td></td></tr>
<tr><td>Productivity Tools</td><td>7%</td><td>12%</td><td>5%</td><td></td><td></td><td></td></tr>
<tr><td>Internet</td><td>56%</td><td>70%</td><td>14%</td><td>69%</td><td>71%</td><td>2%</td></tr>
<tr><td>e-mail</td><td>47%</td><td>74%</td><td>27%</td><td></td><td></td><td></td></tr>
<tr><td>LAN</td><td>54%</td><td>63%</td><td>9%</td><td></td><td></td><td></td></tr>
<tr><td>Intranet</td><td>26%</td><td>30%</td><td>4%</td><td></td><td></td><td></td></tr>
<tr><td rowspan="5">External Processes</td><td>Companies with Websites</td><td>41%</td><td>59%</td><td>18%</td><td>22%</td><td>23%</td><td>1%</td></tr>
<tr><td>Online Catalogues</td><td>18%</td><td>44%</td><td>26%</td><td></td><td></td><td></td></tr>
<tr><td>Customer Inquiry</td><td>30%</td><td>71%</td><td>41%</td><td></td><td></td><td></td></tr>
<tr><td>Receiving Orders</td><td>10%</td><td>19%</td><td>9%</td><td>9%</td><td>13%</td><td>4%</td></tr>
<tr><td>Tracking and status</td><td>2%</td><td>5%</td><td>3%</td><td>2%</td><td>5%</td><td>3%</td></tr>
</table>

Table 3: Comparison Table

To more comprehensively analyse the data and identify trends, the data was split into internal and external processes. An examination of the data for Internal Processes, the ABS (2004) data indicates that in 2000 84% of Australian-based companies had computers while in 2003 the figure was 83%. Data collected in the 2000 and 2003 UWS surveys reveal a different trend for businesses in the region. It appears from the above data, that use of computers within the region was steadily rising (from 89% to 97%), and no fluctuations were apparent.

Furthermore, ABS data indicates that in 2000 69% of companies were connected to the Internet, while in 2003 the percentage reached 71%. The UWS surveys show a slightly sharper increase (56% to 70%) in Western Sydney. An examination of the data that corresponds to the External Processes reveals that the ABS data shows that in 2000 22% of businesses Australia-wide had websites, while in 2003 23% of companies had websites. For the Western Sydney region however the increase was much sharper, with 41% in 2000 and 59% in 2003.

In addition, the ABS data indicates that a number of companies that were receiving orders increased from 9% in 2000 to 13% in 2003. When comparing ABS findings to the findings for Western Sydney it can be seen that in 2000, 10% of companies were receiving online orders, and this increased to 19% in 2003. Both ABS and UWS surveys indicate that Tracking and Order Status has risen from 2% in 2000 to 5% in 2003.

GROWTH ANALYSIS

To complete the analysis of the survey data study of the growth for both Internal and External Processes was undertaken. Furthermore, to analyse current trends and predict future trends growth findings were graphed against the eTransformation Roadmap. Internal Growth Analysis consisted of the following measures (See Figure 2, noting that percentage within brackets shows the growth from 2000 to 2003):

- Effective Individual – three measures namely: Use of Computers (8%), Use of Accounting Packages (7%) and Use of Productivity Tools (5%); bars 1, 2, and 3 respectively.
- Effective Team – three measures namely: Use of Internet (14%), Use of e-mail (27%) and Use and applications of Intranets (4%); bars 4, 5, and 6 respectively.
- Effective Enterprise – only one measure, Use and application of Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) tools (9%); bar 7.

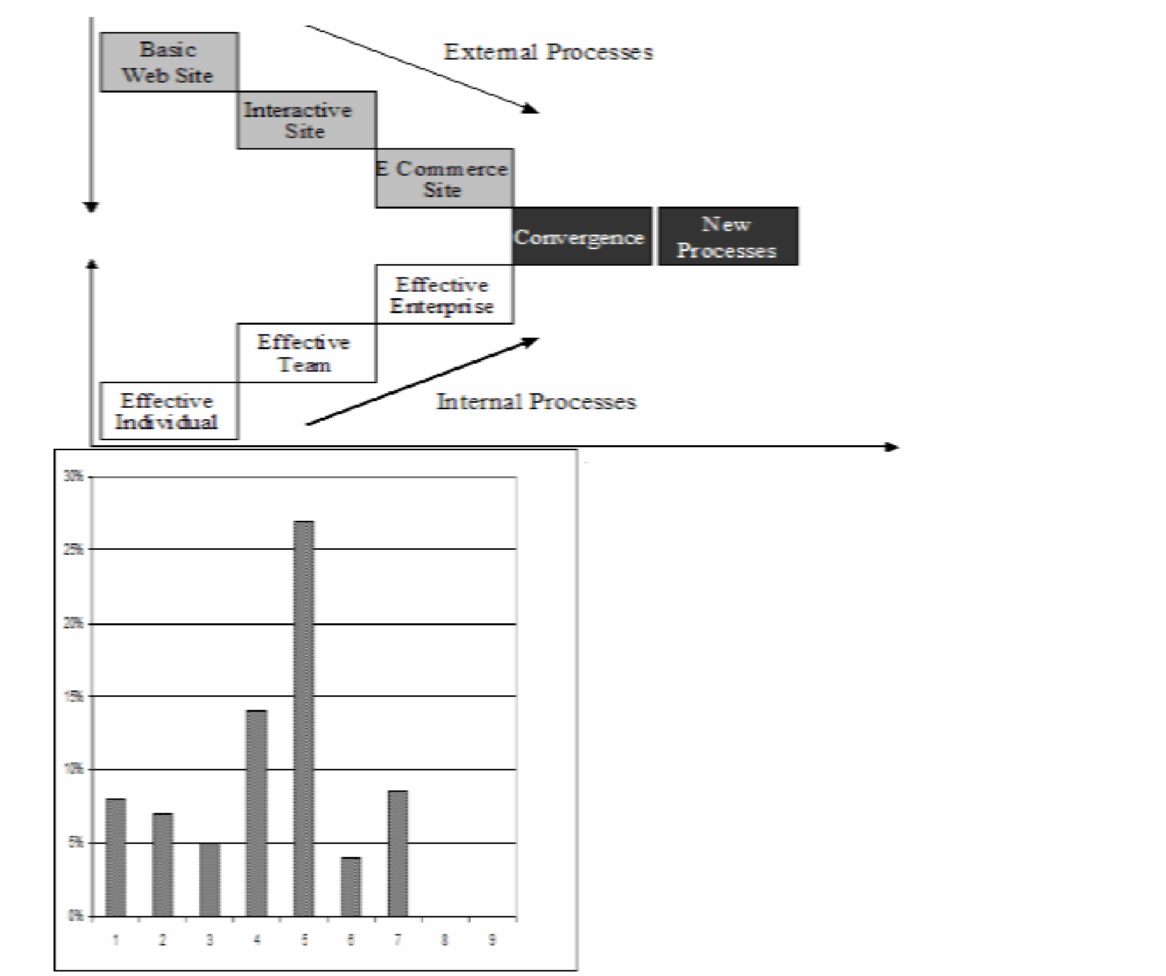


Figure 2: Internal Growth Analysis

Analysis of the Internal Processes shown in Figure 2 indicates that most of the growth corresponds to activities related to creation of internal effective teams. Furthermore, from Figure 2 it can be seen that the use of the Internet has grown by 14% and that the use of e-mail was the highest growing factor for Internal Processes overall (27%) between 2000 and 2003.

Growth for the "Effective Individual" Stage appears to be reasonably low. The reason for this is that "Effective Individual" Stage has almost reached saturation. From Table 3 it can be seen that almost all companies (97%) use computers. In addition, data from the table shows that Western Sydney has experienced a much sharper growth than the rest of the Australia. The fluctuation in the ABS findings (in 2000 84% of businesses used computers, while in 2003 the figure was 83%) might have been influenced by factors external to the business such as the introduction of GST (Goods and Services Tax) in 2000.

It is clear that "Effective Enterprise" Stage (Figure 2) in 2003 was in its early stages of development. In 2003 companies were still far from the state of an effective organisation as there were no enterprise-wide applications reported to support different sections of the organisation, and enable information sharing and integration across the enterprise. It should be noted however that at the time a small number of organisations are using Enterprise Resource Planning and Customer Relationship Management tools (growth of 9%).

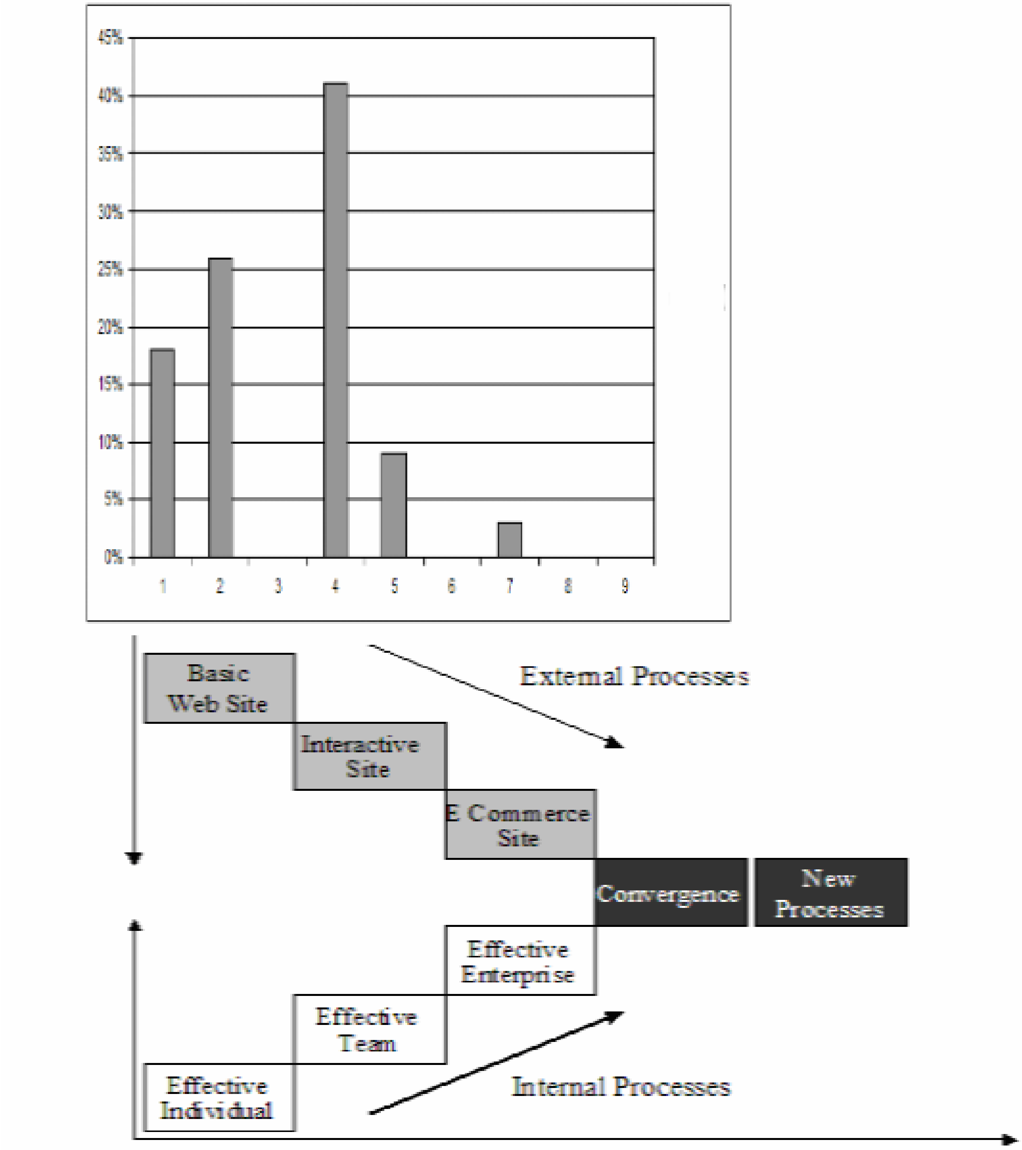


Figure 3: External Growth Analysis

Details about External Growth measures were mapped as following (see Figure 3):

- Basic Website - two measures namely: Companies with Websites (18%) and Companies with Online Catalogues (26%); bars 1 and 2 respectively.
- Interactive Site - two measures namely: Online Customer Inquiry (41%) and Number of Received Orders (9%); bars 4 and 5 respectively.
- eCommerce Site - only one measure: Ability to Track and Get Order Status (3%); bar 7.

From Figure 3 it can be noted that between 2000 and 2003 the number of companies with a website increased by 18%, which was followed by an increase in website development and the introduction of online catalogues (26%). This is the "Basic Website" Stage. The "Interactive Site" Stage was the fastest growing stage between 2000 and 2003. From the data it can be seen that Online Customer Inquiries increased by 41%. In addition, it was noted that a number of the companies receiving online orders increased by 9%.

By 2003 not many businesses had embraced eCommerce, although there were attempts by organisations to get close to having eCommerce website. It appears that companies have first started receiving inquires online, and then followed with online orders. It is also observed that some organisations have provided facilities for customers to track their orders using the organisations website (3%). Very few companies have reached convergence and have started establishing new business processes. Studies of both internal and external processes highlight a gradual eTransformation trend and indicate changes that have occurred within the region between 2000 and 2003.

From this analysis it can be seen that the Western Sydney region has commenced its eTransformation journey. The progress is slow but steady and progressing more rapidly than in other parts of Australia. Most companies within the region have reached the "Effective Individual" and "Basic Website" Stages.

Likewise, most companies within the region have started moving towards the stages of "Effective Team" and "Interactive Site". The progress towards these stages is still quite slow although from the data gathered it seems that in 2003 a very large number of organisations have started seeing the potential that websites offer and have commenced introducing online catalogues (26%) and communicating with staff and customers online (41%). For the companies to reach this stage it was essential to understand and learn what websites can offer and how a company could make full use of such technologies. Increase in the Internet access, use of e-mail and other communication tools, as well as development of interactive websites are all indicative that companies are progressing and learning more about available technologies.

Not many companies have yet reached "eCommerce" and "Effective Organisation" Stages, however they are slowly making progress by adopting required technologies to ensure that they are able to communicate electronically, share data and carry out day-to-day business processes. To assure further growth, businesses within the region will need to study factors that they perceive as hindering the company, before they can fully eTransform. In 2000 38% of companies found finances to be a problem when considering adoption of ICT, and in 2003 this figure rose to 60%. Furthermore, 53% of businesses indicated that lack of time was a problem adopting ICT, as was the need to assure employees have necessary education and skills to implement and maintain ICT (34%).

It is expected that ICT adoption will continue to rise in the years to come and that by 2006 most companies will have full eCommerce websites and will work within Effective Organisations, which will use enterprise-wide applications. In addition, to assure companies have gained the full benefit of eTransformation, it is expected that in the future ICT adoption will be closely linked to product innovation, education, training and organisational change and learning (Productivity Commission, 2004). This can also be supported by data gathered from the 2003 Survey where 70% of the respondents indicated that if they had necessary resources, such as appropriate skills and infrastructure, their business competitiveness would have improved dramatically.

ISSUES IMPACTING TAKE UP OF eTRANSFORMATION

ICT can have a strong influence on how business processes within the organisation are carried out. It is now clear that ICT alone does not function as an element of eTransformation, but needs to be embedded in the organization, people, and knowledge.

From the findings it can be noted that businesses within the Western Sydney Region have realised that there is a need to either retrain existing staff or recruit new people, so that their organisations can have the necessary skills and knowledge to be sustainable in the global environment. The 2003 Survey data indicates that due to such demands 57% of organisations have decided to train their existing staff to become competent users of ICT. Other organisations decided either to outsource or get expertise elsewhere (41%), hire contractors (31%) or employ new staff (20%) that will have knowledge and expertise to use ICT (Khandelwal et al, 2003).

For businesses to adapt there is a need to experience environmental demands, which constantly impact on the way organisations deal with changed circumstances. The Western Sydney region has been experiencing changes in the past decade. These changes have been noticed on a number of frontiers, such as cultural, economical and environmental. Western Sydney initiatives and motivations by WSROC have also aided eTransformation. To assure a growing region's population is able to prosper, it was noted by WSROC that it is important to recognise a need to invest in both social and knowledge capitals (Western Sydney Initiative, 2005). Furthermore, one of the statements explaining the region's vision states: "Greater Western Sydney's role as a major region for growth, change and innovation within Sydney will be recognised as critical to the sustainability and success of the metropolitan area as a whole" (Greater Western Sydney Initiative, 2005).

To facilitate and sustain the eTransformation of organisations within the region, it is important to have an adequate number of ICT organisations within the region to provide hardware, software and ongoing support. Due to such demands the AeIMS Research Group has studied ICT organisations within the region. It was noted that ICT business demands within the region are on the rise. Below is Table 4, which indicates the top five activities of such organisations.

Business Activity of ICT Organisations	**2000**	**2003**
Software development	38%	69%
Computer networking	38%	45%
Computer/ICT consultancy services	25%	45%
Web site and e-commerce system development	33%	38%
Computer hardware assembly/manufacturing	15%	38%

Table 4: Business Activity of ICT Organisations

This indicates that organisations within the region require assistance to effectively eTransform and adopt ICT. Predominately, it seems that companies need appropriate software to do the accounting and essential computer networking so that they can be connected to the Internet and share and use data. From Table 4 it can also be seen that more ICT businesses within the region are specialising in website and eCommerce system development as well as the required hardware manufacturing and assembly.

Once Western Sydney's businesses grasp knowledge and it becomes an integral part of their business components, they will be able to respond quickly, meet deadlines and act swiftly. Moreover, this will allow businesses to appropriately select, implement, and make use of ICT resources so that necessary business processes, transactions and tasks can be completed effectively (Hol & Ginige 2004).

CONCLUSION

The impact that the use of ICT has on businesses within the Western Sydney Region is very strong and apparent compared to the Australia-wide ICT adoption. Businesses within the region are clearly changing, re-engineering their processes and readily making use of available technologies.

Based on the eTransformation Roadmap Model it can be seen that the "Effective Individual" Stage is reaching saturation as is "Basic Website" stage. In the period between 2000 and 2003 it appears that most of the growth has been in the "Interactive Site" and "Effective Team" Stages. In addition, data shows a systematic and staged adoption of ICT across businesses in the Western Sydney Region.

eTransforming businesses need time, knowledge and effort to be able to successfully adopt appropriate ICT. Furthermore, such requirements open up new avenues and encourage new ICT companies to open their doors and provide hardware, software and maintenance support to those eTransforming organisations.

Further studies have been undertaken to study use of ICT within the region. Once data has been collected it will be possible to make deeper comparisons and analysis. In addition, action research is currently undertaken by a number of AeIMS researchers to study use of ICT within the region and observe, study and guide business undertaking eTransformation. Furthermore, a repeat survey is planned for 2006. We hope to find that growth has moved to "eCommerce Site" and "Effective Organisation" Stages. The Western Sydney Region has the potential to be Australia's first eRegion.

REFERENCES

A Greater Western Sydney Initiative (2005) Future Report: Greater Western Sydney Regional Planning and Management Framework, WSROC Western Sydney Regional Organisation of Council, Australia

ABS (Australia Bureau of Statistics), (2004) Research No: 8129.0 Business Use of Information Technology, *Canberra Time*, Australia

Brynjolfsson, E. & Hill, L.M. (2000) Beyond Computation: Information Technology, Organisational Transformation and Business Performance, Journal of Economic Perspectives, Vol14, No4, pp.23-48.

Bullen, C.V. & Rockart, J.F. (1981) A Prime Success Factors, Centre for Information Systems Research, *Stolan School of Management*, Working Paper No. 69. June 1981

Galliers R. D., Merali, Y. & Spearling L. (1994) Coping with Information Technology? How British Executives perceive the key information systems management issues in the mid 1990s, *Journal of Information Technology*, 9, 223-238.

Ginige, A., Murugesan, S., Khandelwal, V., Pollard, T., Costadopolous, N. & Kazanis, P. (2001) Information Technology in Western Sydney - Status and Potential, School of Computing and IT, University of Western Sydney, Australia

Ginige, A., Murugesan, S., & P. Kazanis, (2001) A Roadmap for Successfully Transforming SMEs into E-Businesses, *Cutter IT Journal*. Vol. 14, No. 5, pp. 39-51.

Hol, A & Ginige, A (2004) Super-Symbolic Economy: Modified ROI for SMEs, Proceedings of IADIS e- Commerce 2004 International Conference, Lisbon, Portugal 14-16 December, 2004.

Khandelwal, V., Ginige, A., Curry, J., Bajaj, K., Lan, Y. & Schmid, R. (2003) Information Technology in Western Sydney - Status and Potential, School of Computing and IT, University of Western Sydney, Australia

Khandelwal, V. K. (1992) Information Systems Study, Opportunity Management Program, IMB Corporation, New York

Lawson, R, Alcock, C, Cooper, J and L Burgess (2003) Factors affecting the adoption of electronic commerce technologies by SMEs: an Australian Study, *Journal of Small Business and Economic Development, 10(3), 265-276.*

Productivity Commission (2004) ICT *Use and Productivity: A Synthesis from Studies of Australian Firms,* Commission Research Paper, Canberra

Tapscott, D., (1996) The Digital Economy: Promise and peril in the Age of Networked Intelligence, McGraw Hill; Australia, Sydney

Toffler, A. (1990). *PowerShift: Knowledge, Wealth, and Violence at the Edge of the 21 Century*, Bantam Books: USA, New York.

Western Sydney Budget Statement 2003-04, *New South Wales Government*, Australia

Jordan E-Government: Success Factors

Ibrahim Tadros1 Asim El Shiekh2 A. Abdali Rashed3
2 &3Arab Academy for Banking and Financial Sciences, 2AlBatra University

Abstract
E-government is considered as hot topic tackled by many researchers as it is considered as future fact especially for the developing countries. There are different definitions of e-readiness and the successes factors which are assessed by using many different tools, depending on the results and goals. It very important to gauge the e-readiness state especially in developing countries, as many barriers would be considered as a big challenge. This research introduces a case study of developing countries to explore and discover the success factors that would make the e-government project viable.
The results show that it is necessary to enhance community awareness about e-literacy. Moreover the problem of professional shortage should be taken into consideration.

Keywords:
e-government, e-readiness, e-government successes factor

1. INTRODUCTION

Government services are provided through a variety of channels including retailers, banks and post offices. It is critical that the technology solutions which sit on top of an e-Government infrastructure are within the reach of all citizens .
E-Government is much more than building a web site, it is the infrastructure that governments today are building to transform the way they complete their missions.
E-government services are supposed to be presented electronically . E-Government presents a new and innovative approach to address the traditional problems of government services utilizing the Internet and World-Wide-Web.
Direct effects of e-Government include cost effectiveness in government and public operations, significant savings in areas such as public procurement, tax collection and customs operations, with better and continuous contact with citizens, especially those living in remote or less densely populated areas.
This study addresses the challenges that should be taken into account to facilitate the complex relationships between government and its constituencies to enable success interaction, transaction and delivery of government services. It is organized as following: in section 2, previous studies are reviewed. In section 3, story of Jordan is discussed. In section 4, hypotheses and analysis are discussed. Section 5 presents conclusion and recommendations.

2. PREVIOUS STUDIES

One of the local studies 2, discusses e-government perspectives in Jordan. It studied the effect of e-government on quality of the services, which are supposed to be presented electronically. It focused on the need of e-government then discussed the e-government goals, the situation and barriers of e-government in Jordan. The study concluded that there is a relation between e-government and quality of the electronic services. However some studies tried to tackle e-government problem to discover whether the public sector realizes the meaning of e-government . Furthermore the study listed the barriers that might face the e-government implementation then it suggested solutions to make it work successfully. On other hand, e-government would affect the strategic position of the business organization .

Azam studied the possibility of applying e-government. He concluded that e-government is very important to support the democracy principles. The results showed that it is possible to have an e-government if the laws and legislations are found.
The recommendations of e-government conference (that held in Dubai) focused on :

- The role of public sectors to aware their employees about the importance of e-government.
- Training employees to be qualified enough to used new technologies.
- The importance of public and private sectors co-operations to learn from other countries experiment.

Wescott focused on the efficiency of e-government principles and its effect in monitoring, making government processes easier, questioning and reducing the managerial corruption in public sectors. These can be applied if employees participate in both decisions making and accessing information.

Another study was carried out to cover all aspects of e-government. 250 organizations from 5 countries US, Australia, Canada and UK were conducted. The study introduced a vision of the strategic management level and their ambitions and plans to face e-government modern principles. The study found that the strategic management succeeded in to have service in easier way, increasing the productivity level and reducing the number of compliments and shining picture of the organization.

Two studies (and) that can be classified as Jordanian studies tried to present an assessment model and analysis..
In 10 authors presented a general framework model for e-government readiness assessment. The model consists of six factors: organizational, governance and leadership, customer, competency, technology and legal readiness. The paper concluded that e-service does not need special legislation as it requires assurance of written instructions by the authority. Also the paper suggested improving the legal system to support e-government.

Paper 11 overviewed worldwide experiences such as USA, UK, Singapore, UAE, Egypt and e-government in Jordan. It presented a model of e-government and overviewed the e-readiness in Jordan focusing on infrastructure back office management, policy and legal, community and education and how they could be built. The paper discussed the challenges facing e-government in Jordan such as shortcomings in internet infrastructure, digital geography, privacy, and security, limit IT, legal system and awareness.

Another Arab study   viewed Sudan experience as Sudan started applying e-government. The project is supervised by National Information Center (NIC) and ministry of Science and Technology. The findings: e-government main problems can be stated in culture and legal system. In addition, the most important challenge that faces the conversion from bureaucratic to e-government is employees' conversion acceptance.

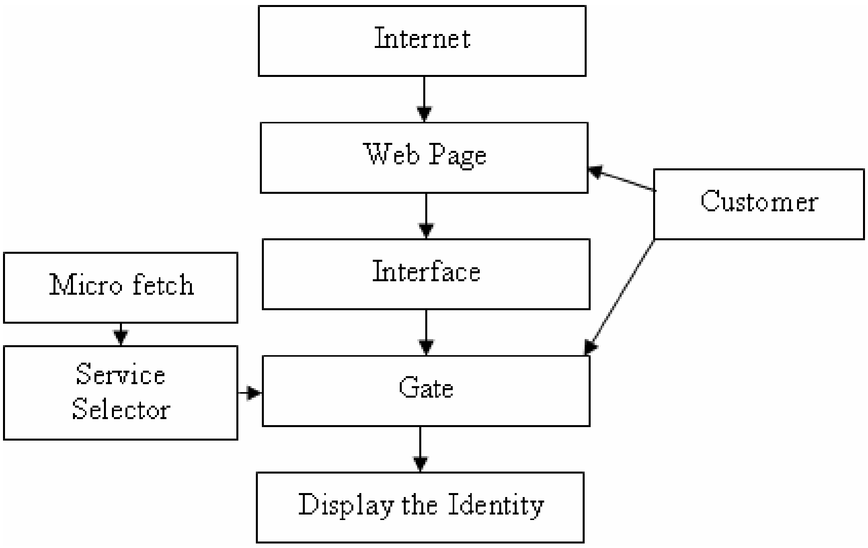


Figure 1: E-government model 11

Success factors:
We can summarize the success factors as following:
- Most citizens should be familiar with internet   as in Dubai and Jordan case.
- Re-engineering management and process and hierarchy.
- Providing essential infrastructure.
- Providing technical support.
- Providing security.
- Providing suitable budget.
- Educating the community.
- Government leadership
- Comprehensive e-government promotion.

Paper 10 recommended the following points to grantee successful conversion:
- E-government implementation should cover all activities and departments.
- Strong support for the conversion via authority.
- Citizens should be involved in the conversion (awareness).
- Strong infrastructure preparation especially national network

- Control should be done by government authority.
- Legal system should be updated to meet the new security requirements.
- Cooperation with other countries that have good experience in this field.
- Studies and conferences should be held to over elaborating discussion.

3. THE STORY OF JORDAN:

Jordan needs to apply e-government to take the advantage of the opportunities offered by all trade agreements; Jordan would need more efficient, market-oriented customs regime in compliance with world trade organization (WTO) requirements, capable of handling increased traffic at the borders while at the same time preventing the entry of pirated software . The following subsections will demonstrate the

Jordan Software Industry

Many research papers have been carried out to overview the software aspects in Jordan to give an indication for the growing state of software industry in developing countries and . Many aspects can be considered as vital points that should be highlighted such as factors of strength, weaknesses, the difficulties and challenges Jordanian companies suffered from. In addition, it clarified the "hopes" that Jordanian companies dream to achieve with comparison to their limited capability. A questionnaire was distributed to Jordanian companies and data was collected and analyzed (13 and). However other researches focus on outsourcing and how could Jordan compete in the international market (17 and).

As a result of increasing the number of companies that invest in the software development sector, an amendment to the Copyright Law was introduced to acknowledge that copyright ownership of all works created by employees shall rest with the employer provided that such works are related to the business of the employer and provided that employees are utilizing the knowledge, tools, and resources made available by the employer .

In 1999, Jordan's parliament amended the country's 1992 Copyright Law and passed various regulations to better protect intellectual properties and . Two years later, King Abdullah received a special award from the Business Software Alliance (BSA) for his efforts to enforce the country's copyright and trademark laws. Largely due to these efforts, software piracy in Jordan has seen a steady decline since 1994, when rates reached 87%. By 2002, piracy rates had dropped to 64%, although the total losses of the software industry had risen, from US$2.2 million in 1994 to US$3.5 million in 2002 .

4. THE SAMPLE DESCRIPTIVE ANALYSIS AND HYPOTHESES TESTING:

Data are collected using a questionnaire and observation on e-government success factors, conducted for three different groups: The public sector employee, citizen and University instructors

4.1 The Sample Descriptive analysis And Hypotheses Testing

The Sample Descriptive

Table1: The sample destitution according to gender

Gender	Frequency	Percent
Male	211	54.5
Female	176	45.5
Total	387	100.0

Table 1 shows the destitution according to gender of the sample.

Figure 2: The sample destitution according to gender

Table 2: The sample destitution according to Experience

Experience	Frequency	Percent
Less than 5 year	124	32.0
From 10-5	208	53.7
From 11-15	28	7.2
More than 15	27	7.0

Total	387	100.0

Table 2 and figure 2 demonstarate the sample destitution according to Experience
Table 3 the sample destitution according to Experience

Social Status	Frequency	Percent
Single	142	36.7
Married	222	57.4
Widow	6	1.6
divorced	17	4.4
Total	387	100.0

Figure 2: sample destitution according to Experience.

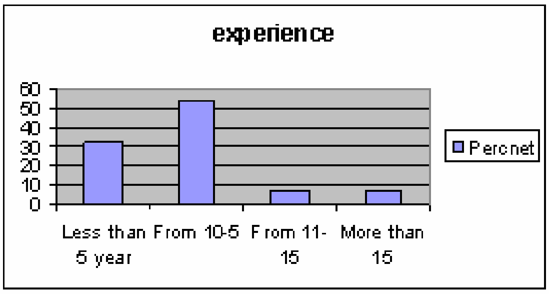


Table 3: The sample destitution according to Social status

Table 3 and figure 3 show that 36.7% are single, 57.4% are married, 1.6% are widow and 4.4% are divorced.

Figure 3: The sample destitution according to Social status

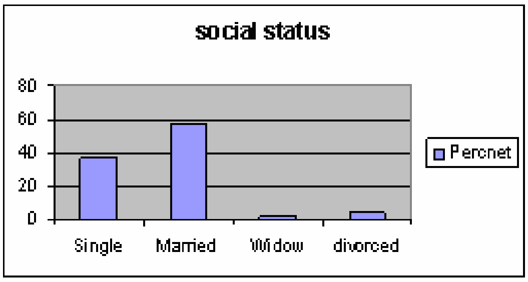


Table 4: The sample destitution according to age

Age	Frequency	Percent
Less than 30	219	56.6
30 – 39	98	25.3
40 – 49	43	11.1
More than 50	27	7.0
Total	387	100.0

From table 4 and figure 3, 56.6% less than 30 years 25.3% between 30 – 39 years 11.1%, between 40 – 49 and 7% are more than 50 year with percent 7%.

Figure 4: The sample destitution according to age

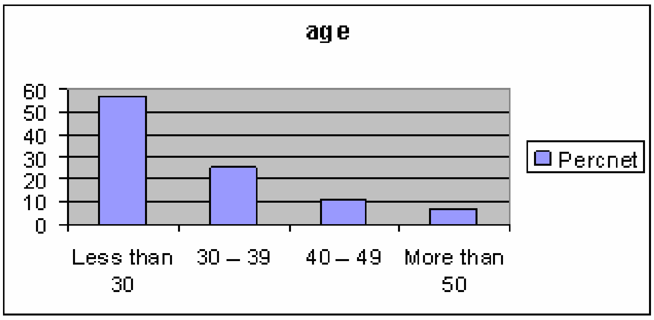


Table 5: The sample destitution according to Job Description

Job description	Frequency	Percent
High level	43	11.1%
Meddle level	78	20.2%
Low level	266	68.7

Table 5 shows the destitution according to the job description.

Table 6: The sample destitution according to the needed to accomplish any governmental transactions

Answer	Frequency	Percent
Yes	387	100.0
No	0	0
Total	387	100.0

Table 6 shows destitution according to the need to accomplish any governmental transactions, the sample answers yes. And Table 7 and figure 5 show the evaluation for the governmental service provided.

Table 7: evaluation for the governmental service provided

Answer	Frequency	Percent
Very Excellent	20	5.2
Excellent	85	22.0
No opinion	112	28.9
Bad	153	39.5
Very Bad	17	4.4
Total	387	100.0

Figure 5 governmental service provided

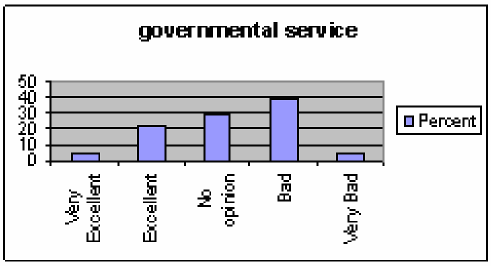


Table 8: use of the internet in accomplishing any governmental papers?

Answer	Frequency	Percent
Yes	7	1.8
No	380	98.2
Total	387	100.0

Most of the respondents has not used internet to accomplish any governmental papers.

Table 9: The sample destitution according to the household income?

Answer	Frequency	Percent
Less 150	45	11.6
Between 150 and 299	288	88.27
Between 300 and 349	24	0.06
Above 450	30	0.07

From the table above the sample of study contains 45 individuals their salary is less than 150 JD with percent 11.6% , 288 individuals their salary is between 150 and 299 JD with percent 88.27%,24 individuals their salary is between 300 and 349 JD with percent 0.06% and 30 individuals their salary is more above 450 JD with percent 0.07%

Hypotheses Testing

Citizens' awareness

Ho: There is No awareness from the citizens of the electronic government existence

Variable	Mean	S.D	T. Value	Sig
Awareness	2.5116	0.88555	-10.849	0.00

Table 10: hypotheses 1 test

From table 10 the mean of attitude about the role of e-government in improving the Jordanian government service equals to 2.5116 with S.D equals to .88555 to test this hypotheses we use one sample T. test that compares between the actual mean with 5 likert scale (3). T value equals to 10.849 with Sig equals to 0.000 so we reject the hypotheses H1 and accept the hypotheses H0. There is No awareness from the citizens of the electronic government existence.

Internet Usage

Ho: Internet usage is not considered as obstacle of implementing e-government

Variable	Mean	S.D	T. Value	Sig
Internet usage	3.666	0.785	16.690	0.000

Table 11: internet usage hypothesis test

From table 11, the mean of attitude about the role of e-government in improving the Jordanian government service equals to 3.666 with S.D equals to .785 to test this hypotheses we use one sample T. test that compares between the actual mean with 5 likert scale (3). T value equals to 16.690 with Sig equals to 0.000 so we reject the null hypothesis H0 and accept the alternative hypotheses that Internet usage consider as obstacle of implementing e-government.

Communication Cost

The Eigen values equals to 1.097 with percentage of Variance 8.436 of this study which contains the following statement: The high cost of the telecommunications through the internet reduces the desire in using the electronic government, the degree of loading (0.929) and degree of extraction equals to (0.870).

Lack of Professional People
H0: There is no Lack of professional people

Variable	Mean	S.D	T. Value	Sig
difficulties	3.254	.371 2	51.24	0.0 0

Table 12: Lack of Professional People test

From the table above the mean of attitude about the role of e-government in improving the Jordanian government service equals to 3.254 with S.D equals to .3712 to test this hypotheses we use one sample T. test that compares between the actual mean with 5 likert scale (3). T value equals to 52.24 with Sig equals to 0.000 so we reject the null hypotheses H0 and accept the alternative hypotheses there is Lack of professional people.

4. Factors Analysis

Statement							Extraction
Equality in treatment	Factor1						
E-government reduces the routing and bureaucracy in finalizing the transactions.	0.484						0.575
E-government will help in reducing the dissent of the affective systems.	0.406						0.483
E-government will help in reducing the courtesy phenomena.	0.823						0.746
E-government will help in reducing the effect of the personal relationships on accomplishing the work.	0.824						0.763
Technology usage		Factor 2					
E-government participates in the fast spreading of the important decisions and rules.		0.523					0.414
The transactions done though e-government equals to the transactions done manually in its accuracy and officially.		0.593					0.681
E-government helps in supporting the technical development.		0.622					0.509
E-government facilities the conditions for the citizen to cope with the modern tetchiness in the field of IT.		0.572					0.527
e-government will help in encouraging the development of the employees efficiency in using the computers		0.675					0.559
The use of e-government reduces the unnecessary and justified communications between the employee and the citizen.		0.437					0.491
Improving services			Factor 3				
E-government participates in improving the services provided in the ministries (after applying the international standards).			0.407				0.353
The usage of e-government is distinct in increasing the accuracy of finalizing the transactions.			0.651				0.541
E-government facilities the personnel works.			0.824				0.697

E-government helps in accelerating finalizing of transactions.			0.430				0.688

Table 13:The factor analysis result has divided the statement into six factors(continue)

E-government participates in providing a feedback to any matters concerns the citizen easily and fast.			0.648				0.494
E-government helps in responding to the problems and needs of the citizens efficiently and fast.			0.456				0.561
reduces costs to finalize the transactions				Factor 4			
The usage of e-government reduces the required costs to finalize the transactions				0.815			0.724
E-government helps in minimizing the follow up in order to finalize the transactions.				0.745			0.637
Internet facilities					Factor 5		
E-government participates in spreading the internet lecture.					0.672		0.745
E-government helps the employees in providing the services for the citizens at all times.					0.383		0.552
E-government helps in rising it's acceptance for the suggestions and the problems of the citizens.					0.514		
Meeting the citizens needs who resident in the other district beside the main city.						Factor 6	
E-government helps in meeting the citizens needs who resident in the other district beside the main city.						0.799	0.689
Eigenvalues	5.263	1.877	1.701	1.678	1.383	1.100	13.002
% of Variance	23.92 %	8.53%	7.73%	7.63%	6.28%	5.00%	59.09%

Table 13: The factor analysis result has divided the statement into six factors

Factor 1Equality in Treatment
The Eigenvalues equals to 5.263 with percentage of Variance equals to 23.92 of this study which contains the following statement: (The e-government reduces the routing and bureaucracy in finalizing the transactions, The electronic government will help in reducing the dissent of the affective systems, e-government will help in reducing the courtesy phenomena, and e-government will help in reducing the effect of the personal relationships on accomplishing the work , the degree of loading (0.484, 0.406, 0.823, 0.824)and degree of extraction equal (0.575,0.483,0.746,0.763).

Factor 2 Technology Usage
The Eigenvalues equals to 1.877 with percentage of Variance equals to 8.53 of this study which contains the following statement: e-government participates in the fast spreading of the important decisions and rules, the transactions done though the electronic government equals the transactions done manually in it's accuracy and officially, e-government helps in supporting the technical development, e-government facilities the conditions for the citizen to cope with the modern tetchiness in the field of information technology, e-government will help in encouraging the development of the employees efficiency in using the computers, the uses of the electronic government reduces the unnecessary and justified communications between the employee and the citizen, the degree of loading (0.523,0.593,0.622,0.572,0.675,0.437) and degrees of extraction equal (0.414,0.681,0.509,0.527,0.559,0.491).

Factor 3 Improving Services:

The Eigenvalues equals to 1.701 with percentage of Variance equals to 7.73 of this study which contains the following statement: The electronic government participate in improving the services provided in the ministries (after applying the international standards),The usage of e-government services is distinct in increasing the accuracy of finalizing the transactions, e-government facilities the personnel works, e-government helps in accelerating finalizing of transactions, e-government participate in providing a feedback to any matters concerns the citizen easily and fast, and e-government helps in responding to the problems and needs of the citizens efficiently and fast. The degrees of loading (0.407,0.651,0.824,0.430,0.648,0.456) and degree of extraction equal (0.353,0.541,0.697,0.688,0.494,0.561).

Factor 4 Reduces Costs to Finalize the Transactions
The Eigenvalues equals to 1.678 with percentage of Variance equals to 7.63 of this study which contains the following statement: The usage of e-government reduces the required costs to finalize the transactions and e-government helps in minimizing the follow up in order to finalize the transactions. The degrees of loading equal to (0.815, 0.745) and degree of extraction equal to (0.724, 0.637).

Factor 5 Internet facilities
The Eigenvalues equals to 1.383 with percentage of Variance equals to 6.28 of this study which contains the following statement: e-government participates in spreading the internet lecture and e-government helps the employees in providing the services for them. The degrees of loading are (0.6720.383) and degrees of extraction equal (0.745, 0.552).

Factor 6 (meeting the citizens needs who resident in the other district beside the main city)

The Eigenvalues equals to 1.100 with percentage of Variance equals to 5.00 of this study which contain the following statement: e-government helps in meeting the citizens needs who resident in the other district beside the main city. The degrees of loading (0.799) and degree of extraction equal (0.689).

5. CONCLUSION

The transition to a successful e-Government requires visionary executive leadership, broad vision with a tactical plan, Culture change, Partnerships, A supportive public policy environment, strong infrastructure, financial resources, skilled human element, deliver electronic and integrated public services, Bridge the digital divide, achieve lifelong learning, rebuild government-customer relationship,
This findings showed that the Jordan e-government is on the way and has gained clear recognition of its achievements in doing so.

Nonetheless there were several obstacles that still need to be dealt with in the near future, these include:

- The Jordanian Community suffers from e-literacy.
- There is no sufficient awareness by the citizens in applying the electronic government
- There is a concern of citizens' privacy.
- There is Lack of professional people.

Recommendations
The main areas of further study directly relevant to this work and recommendations for the implementation of successful e-government are the following:

- Re-enforcing the national training capacity to boost the ICT sector with guarantee Quality of the training (standards) and Accessibility (affordable fees).
- Expanding the number of free public Internet access points.
- Provide aides at access points who can train citizens in basic computer skills.
- Creating programs that include traditional media, like radio programs or newspaper columns, where citizens can learn about e-Government.
- These should be timely and when someone sends government an e-mail message, that message should be promptly acknowledged and responded to in a timely manner.
- Professional workers should remain in Jordan or ensure their resources and successes help benefit Jordan Incentives need to be in place to retain them, and to identify and attract those Jordanians living abroad
- Liberalizing telecommunications industry this will effect on lowered costs, enabled IT industry; increased company competitiveness; increased foreign investment.
- Legalizing voice over Internet protocol (VOIP) by private telecommunications providers this will affect on lower cost telecommunications.

REFERENCES

Tadros Ibrahim, Success Factors in Jordan e-government, PHD thesis, Arab Academy for Banking and Financial Sciences, may 2005.

AL-Honity Mohammad, E-government Impact on e-Services: Case Study, The First Conference on Business Administration Emerging Issues in the Third Millennium: Opportunities and Challenges Facing Arab Business Organizations, 3-5 May 2005, Amman, Jordan, pp: 155-187.

Al-Shawa, Nisreen, Public Sector Employees and E-Government, Masters Thesis, Yarmouk University, Irbed, Jordan, 2004.

Alkhatib Fahad and Alhuseini Falah, E-Government Impalct on Organizational Strategic Management: Case Study on Jordanian Organizations, Dirasat Jornal, Issue 1, pp: 163-182.

Azzam, Ahmad, e-Government in Jordan: Implemntation Possibility, Masters Thesis, Yarmouk University, Irbed, Jordan, 2001.

E-government conference, Dubai, UAE, 2000.

Wescott Clay, On the International Policy, Roadmap for E-Government in the Developing World, 2001, E-Government Enabling Asia-Pacific Governments and Citizens to Do Public Bossiness Differently, Asian Development Bank, Bangkok, Thailand.

Deloitte Research, At the Dawn of E-Governments: the Citizens as in Customers, Washington, 16th June 2000.

Al-Omari, Ahmad, Al-Omari, Hussein, A Framework Model for Assessment of E-Government Readiness, Proceedings of International Conference on Telecomputing and Information Technology 2004, 22-24 September, 2004, Applied Science Private University, Amman Jordan, pp:232-237.

Al-Omari, Hussein, E-Government Architecture: a Comparison Analysis, Proceedings of International Conference on Telecomputing and Information Technology 2004, 22-24 September, 2004, Applied Science Private University, Amman Jordan, pp: 215-223.

Badrdin Merghani, Appying Ideal Bearucratic methods or E-Government,The First Conference on Business Administration Emerging Issues in the Third Millennium: Opportunities and Challenges Facing Arab Business Organizations, 3-5 May 2005, Amman, Jordan, pp: 237-259..

Byoomi Hijazi, Legal System to Protect Electronic Processes, Dar AlFeker, Eygypt, 2003.

AMIR , Achievement of Market-Friendly Interactive and Results Programs Building a Modern Customs Administration for Jordan, Jordan is making significant strides in implementing customs reforms and improves status as international trading partner, retrieved from http://www.amir-jordan.org/success_case_pspi.shtm

Rashed A. Abdali, Oliamat Mohammad, Software industry in Jordan the Prospects and Hopes, The 2nd International Conference on Information Technology "Human Computer Interaction" May 3-5, 2005, Al-Zaytoonah University, Amman, Jordan .

Rashed A. Abdali and Oliamat Mohammad, Software Industry in Jordan the Ambitions and Barriers, Arab Academy for Banking and Financial Sciences, 2005.

El-Sheikh Asim, Rashed A. Abdali and Abu Samaha Ala, The Prospects of GSO in Arab Countries: Case Study of Jordan, submitted as a chapter in refereed book, titled with Outsourcing and offshoring in 21st Century: a Socio-Economic Perspective, Edited by Dr. H. kehalm, University of Western Sydney, Australia, and V. Singh, publisher IdeaGroup.

El-Sheikh Asim and Rashed A. Abdali (2004), The Prospects of GSO in Jordan, International Conference on Telecomputing and Information Technology/ ICTIT 2004, Applied Science Private University, Amman Jordan, 22-24 September 2004.

REACH, Jordan's REACH Initiative Legislative Changes, retrieved from http://www.reach.jo/legislative_changes.htm

El-Sheikh Asim, Rashed A. Abdali, Peace A. Graham (2005): Software Piracy: Possible Causes and Cures, chapter of refereed book: L. Freeman G. Peace, Information Ethics: Privacy and Intellectual Property, Idea, pp: 84-98.

El-Sheikh Asim, Rashed A. Abdali and Bal Beed Majed (2004), The Affect of Applying Laws and Legislation in Reducing Software Piracy, , Journal of Applied Sciences, 2005,vol. 7, No. 1, pp: 11-33.

Business Software Alliance (BSA) (2003). Eighth Annual BSA Global Software Piracy Study, Washington, DC: Business Software Alliance, retrieved from http://www.bsa.org

Modeling Group Dynamics in a Global Information Systems

Jacek Unold
The Wroclaw University of Economics, Wroclaw, Poland
unold@han.ae.wroc.pl

This research was supported by a *Marie Curie International Fellowship* within the 6th European Community Framework Program.

Abstract

The research topic takes up the issue of group dynamics within a Global Information System, and a relating question of human rationality. The interests in the area of information systems have concentrated mainly on technological aspects so far. If the human component were taken into account, it has been analyzed from the level of an individual. So have all new concepts of rationality. The detailed project objectives arise from the following research questions: Is collective behavior a phenomenon of exclusively qualitative nature, impossible to structure? Does irrationality of individual actions determine irrationality and unpredictability of the whole social system? A model exemplification of a Global Information System is a modern, electronic stock exchange. The anticipated findings should help formulate a new, adaptive, paradigm of rationality.

Keywords:

Global Information System, Global IS dynamics, IS social subsystem, group behavior, rationality.

INTRODUCTION

One of the main issues both in the theory and practice of Social and Economic Sciences is the question of integration, i.e., how actions and interactions of individuals lead to the emergence of phenomena which characterize social entireties. This topic acquires particular importance in the light of the dynamic integration processes of societies, e.g., the next stage of the European Union enlargement. The integration processes, aided with the most recent achievements in information technology, harmonize with globalization and virtualization of human activity, from social to political to business one.

The outlined issue is a background of two basic research threads proposed in the project. The first part refers to the question of human behavior within an information system (IS). A fully integrated society of the future will make a fundamental, subjective element of a Global Information System. That Global IS will be either a ubiquitous and wireless Internet or some totally different, unknown yet, technological platform. And it is crucial that so far the interests in the area of IS have concentrated mainly on technological aspects. The issue of human behavior within an IS has been generally omitted, as one belonging to other disciplines. Admittedly, since the mid-1990s we have observed some growth of interest in the domain of social aspects of the IS development (Avison and Fitzgerald, 2003), but those interests have concentrated on the specificity of individual behavior. However, the nature of global phenomena and the features of dispersed collectivities denote a necessity of a new perspective on the society and organization. No longer can we perceive the human component of an IS as independent individuals. The users of local, regional and global telecommunications networks create a specific form of a "virtual crowd", accessing the same sources of information and reacting to the same sets of stimuli. These users, through their interactions, compose the phenomenon of IS dynamics. Dynamics not understood as one referring to the flow of energy (e.g. electrical impulses), but dynamics based on the collective information processes, reflected in collective actions. The phenomenon named by H. Simon "a collective mind".

And here we find the other research thread of the research proposal: the issue of human rationality. The Western organizational culture is still based on the three main determinants: individualism, competition, and a mechanistic-reductionist perspective. As a result, the essential body of scientific achievements in the area of human behavior concerns individuals (Nelson and Quick, 2000), and this is reflected in the paradigm of rationality. This depiction, known as Rational Choice Theory (Halpern and Stern, 1998), assumes that individuals are perfectly rational, with clearly defined preferences, and optimizing their behavior at all levels of a decision-making process. Reductionism, which is related to it, postulates that collective behavior is composed of the sum of rational behavior of all individuals. Since this "sum" is purely theoretical and abstract, it is generally accepted that all phenomena concerning a collectivity are exclusively qualitative and cannot be structured.

The deficiencies of the traditional, idealistic approach to rationality have been known and discussed for a long time. Admittedly, since Simon's idea of "bounded rationality" it has been allowed that human actions can be more "satisficing" than "optimizing", but all new concepts of rationality still refer only to individual behavior (Halpern and Stern, 1998). At the same time, it has been emphasized that there is a need for such a formulation of the rationality principle so that it can take into account the specificity of collective behavior, so different from the individual one.

The research project takes up this issue, aiming to model information processes of collectivity and to structure this phenomenon through the identification of its quantitative dimension. These findings should help formulate a new, wider approach to rationality, which could respond to the integration and globalization trends of modern societies.

PROJECT OBJECTIVES

An information system (IS) is a set of interacting components: *people, data/information, procedures, hardware, software, and communications* (Benson and Standing, 2002). In another exemplary approach, an IS is a system which assembles, stores, processes and delivers information relevant to an organization or to society (Avison and Fitzgerald 2003). The authors of this definition stress that an IS is a human activity (social) system which may or may not involve the use of computers. It is evident that, regardless of an approach, a social subsystem (*people*) makes a basic, subjective element of every IS.

The scope of application divides IS into micro- and macroeconomic categories, and the dynamic growth of the Internet and Internet-based information systems initiated the birth and development of a Global Information System (Global IS). Owing to the interest of a social subsystem, a model exemplification of a modern IS will be in this project an IS of a present-day electronic stock exchange, because the performance of financial markets is a typical example of group (crowd) reactions (Plummer 1998).

The issue of human behavior within an IS is the most recent research trend in the discussed area. In 1994, A. Morita (1994), the founder of Sony Corp., pointed to the constantly growing gap between the world of business, and generally a society, and the new world of information technology ("IT/business gap"). It was the first human aspect within the area of IS that was detected so clearly.

The problem identified by A. Morita a decade ago concerned individual attitudes and actions. This proposal takes up a new and unrecognized issue of collective behavior within an IS. The importance of this topic results from the fact that collective information processes and collective behavior compose the basic determinants of an IS dynamics, and this phenomenon plays a key role in the development and functioning of the Global Information Society. According to C. Eden and J.C. Spender (2003), the dynamics of an organization represents changes in the various types of knowledge, in the learning and unlearning processes. At the same time, the collective knowledge cannot be understood without paying attention to the communication processes going on among the group's members (Weick 2000). It follows that the basic determinants of the IS dynamics are: knowledge, learning and unlearning, and these phenomena relate to the information processes of collectivity. The operation of collective information processes determines organizational learning, which is a useful metaphor describing the way an organization, also a virtual and global one, adapts to its environment. Obviously, the individualistic and optimizing approach to rationality has become insufficient in that new, virtual environment, and one of the biggest challenges in the discussed area is an attempt to adapt the traditional paradigm of rationality to the new reality.

This research project proposes an innovative approach to the analysis and modeling of collective information processes and the mechanisms of collective behavior within an IS. This is the main objective of this project. This is also the first attempt of this kind in relation to the above-mentioned social issues and it should help define a new approach to rationality.

The following research questions have been formulated:

i. Are information processes of collectivity and the resulting collective behavior, which make the IS (Global IS) dynamics, phenomena of exclusively qualitative nature, impossible to structure?
ii. Does irrationality or non-rationality, and unpredictability of individual actions determine irrationality and unpredictability of the whole social system?
iii. Is there a research method allowing to identify and analyze a phenomenon of "social rationality", from its theoretical to methodological to empirical dimension?

The introductory stage of the research enables to propose the following theses:

i. Collective behavior, which is a basic determinant of the IS (Global IS) dynamics, does not proceed in a planned and intended manner, but is adaptive and follows certain patterns found in the world of nature.
ii. Collective behavior can be expressed in a model form, which enables to structure this phenomenon, otherwise considered purely qualitative so far.
iii. The identification of quantitative attributes of collective behavior will provide substantial theoretical and methodological premises for the extension of the optimizing and individualistic notion of rationality by the social and adaptive aspects.

The outlined background and chosen research objectives clearly indicate that the issues taken up in the project will be the subject of broad multidisciplinary and intersectorial investigation.

THEORETICAL BACKGROUND

The theoretical part of the research will begin with the identification of the determinants of an IS dynamics and relating this phenomenon to the behavior of a social subsystem of a given IS, thus to the issue of collective behavior. Generally, the point will be to identify a quantitative dimension of this, otherwise known as purely qualitative, phenomenon.

The next novelty will be an analysis of the genesis and evolution of the rationality principle and its influence in micro- and macroeconomic models. The limitations of the traditional approach to rationality have been discussed for a long time. One of the research objectives is a detailed analysis of all rationality concepts known in literature, from those "individualistic-optimizing" to the most recent ones, which try to take into account social and organizational aspects of human actions. Such a comprehensive elaboration on the rationality issue has not been found in literature so far. This part of the study will allow for the identification of the basic determinants of rationality in the Theory of Information Systems. Owing to the concentration of interests on the technological aspects of IS, this will be the first attempt to relate the issue of rationality to the IS field.

The research on collective mind and collective behavior will be conducted in a social subsystem of an IS of a stock exchange. This choice is motivated by the fact that, contrary to other collectivities, the behavior of a stock market collectivity (investors) is reflected by relatively simple and concrete indicators, i.e., price changes shown by an index chart and some "mechanical" indicators of collective activity, such as the volume (the number of shares changing hands during a session) and the total turnover (money engaged on either side during a session). In the introductory analysis we found that this social subsystem reveals all 4 attributes of a nonlinear, complex adaptive system. Such a system:

i. Consists of a network of agents (here: investors) acting in a self-managed way without centralized control.
ii. The environment in which the investors operate changes and evolves constantly, which is the result of continuous fluctuations in economy and the market situation, but it also is produced by the interactions among the agents.
iii. Competition among the agents (investors) leads to a consensus, reflected by a current market trend.
iv. This trend suggests the hidden existence of organized patterns of collective behavior, which is the result of the emergence of a natural dynamic structure of this social system.

The classification of a social subsystem among complex adaptive systems allows for the application of the most recent achievements of Complexity Theory and Chaos Theory. Complexity Theory can offer a range of new insights into the behavior of social and economic systems. The idea of self-organization and emergence can be used to identify and explain the dynamics of individual and collective behavior, e.g., on the stock market. Thousands of independent and difficult to observe transactions, carried out by individual participants of the market, generate an emergence of specific and predictable patterns of collective behavior. These phenomena can only be identified on the higher - collective, not individual - level of social organization. S. Kauffman's (1996) famous phrase "order for free" describes that process of "crystallization", also known as the emergence of complexity in complex adaptive systems. The fundamental challenge of this project will be finding a quantitative measure of that emergence.

To model and structure the behavior of this complex adaptive system we will use the elements of Environmental Economics - one of the most trendy areas in Economics. Further analysis will amplify the approach proposed by F. Capra (1995), who believed that the new paradigm of rationality should take into account the fact that "an economy is a living system, and one of many aspects of a large ecological and social structure." This assumption leads to the notion of an open system (contrary to the Newtonian isolated system) and entropy (Gray 1998). This enables the introduction of an innovative research concept. The project's concept is based on a 3-element system, *society-economy-nature*, and it replaces the 2-element system, *economy-nature,* which has been applied in Economics since the 1960s. The proposed model of collective behavior will be based on this 3-element system, thus allowing for the identification of dependencies in its 2-element subsystem, *society-nature*. This will help identify and describe phenomena, which are observed in the surrounding world of nature and, to the same extent, are expected to regulate behavior of the crowd. This is the actual, as opposed to only formal and declared, introduction of the ideas of open system and entropy to Economic and Social Sciences.

The initial results point to the possibility of a graphic representation of the analyzed phenomenon of collective behavior. The most common curve in the world of nature is a logarithmic spiral, which is isomorphic, self-similar, and based on the Fibonacci ratio $\Phi=1.618$. The preparatory stage of the research contains the successful application of the logarithmic spiral to an index chart. Each index chart, according to technical analysis, represents the crowd sentiment on the market. This is the first attempt of this kind with reference to the emerging Polish stock market. After the necessary adjustments and tests the spiral should turn out to be a new and powerful forecasting tool.

The next innovative aspects of the project are revealed with the interpretation of the identified phenomena, especially with reference to the scientific achievements of Quantum Mechanics (Heisenberg's uncertainty principle) and Chaos Theory.

RESEARCH METHOD

The outline of the proposed research method can be presented in several basic steps:

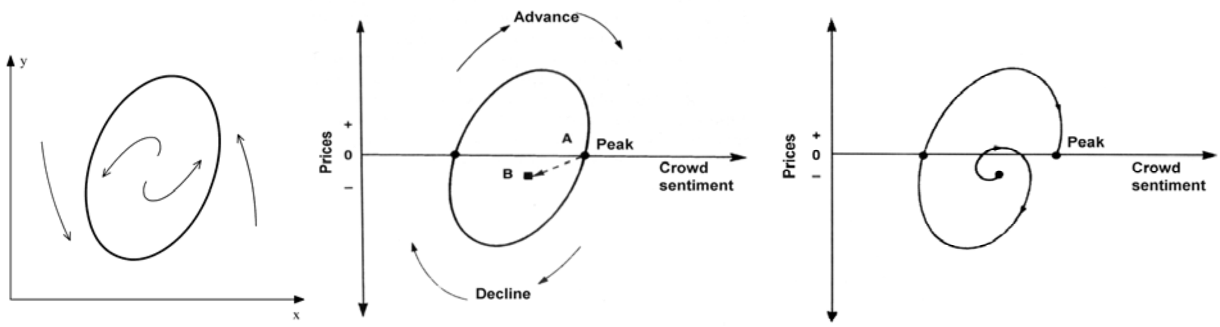
i. Identification of the research area, based on the Theory of Information Systems, Organizational Behavior, Societal Behavior, Complexity Theory.
ii. Identification of the mechanism of collective behavior (Organizational Behavior, Theory of Cycles).
iii. Analysis of the identified mechanism (Theory of Cycles, Chaos Theory).
iv. Synthesis: a mathematical description of the mechanism (Environmental Economics and 3-element model *society-economy-nature*).
v. Exemplification and verification of the mechanism (application of a logarithmic spiral on stock indexes charts).
vi. Interpretation and discussion: a new approach to rationality (Theory of Economics, Rational Choice Theory, Management, Organizational Behavior, Societal Behavior, Quantum Mechanics).

Below runs a brief description of these six stages.

i. The collectivity of investors compose a non-linear complex adaptive system.

ii. A collectivity is created by information capable of uniting single individuals into a group. The group, then, lives its own life, a life, which depends on the exchange of information with the environment. The most significant symptom of this phenomenon is the collectivity's fluctuation during this exchange, and it reflects its dynamics. According to the Theory of Cycles such stable fluctuations between a system and its subsystem can be presented in a model form as a bounded cycle (Jordon and Smith, 1999) - Fig.1a.

iii. According to Chaos Theory, the bounded cycle is one of the 3 possible forms of an attractor Peters 1996). It is also a basic mechanism through which complex adaptive systems react to the fluctuations of the environment. Because this cycle is stable, it does not represent all adaptive processes. In reality, the flow of information is not a continuous process. So, when unexpected information appears (*information shock*), the collectivity tries to conform to the new conditions by changing its dynamic structure. It is expressed by a sudden "jump" from the cycle path. As far as financial markets collectivities are concerned, a jump in both *prices* (y) and *moods* (x) occurs (point A to point B in Fig. 1b). Some time later, the collectivity tries to return to the basic cycle path and this phenomenon is expressed by a spiral of the adaptation process (Fig. 1c).



a b c

Figure 1. Formation mechanism of a spiral of the adaptation process of collectivity: a) bounded cycle; b) information shock and "jump" from the cycle path; c) spiral of the adaptation process.
Source: author's research based on (Plummer, 1998) and (Jordon and Smith, 1999)

iv. The key question is, What kind of a spiral represents these phenomena (because there are several different spiral movements)? This base can be found in the world of nature, because a collectivity also forms a natural system (Frost and Prechter, 2001). Thanks to the proposed 3-element model (*society-economy-nature*) we can look for analogies between its two natural subsystems: *society (collectivity)* and *nature*. The most common curve in nature is a logarithmic spiral.

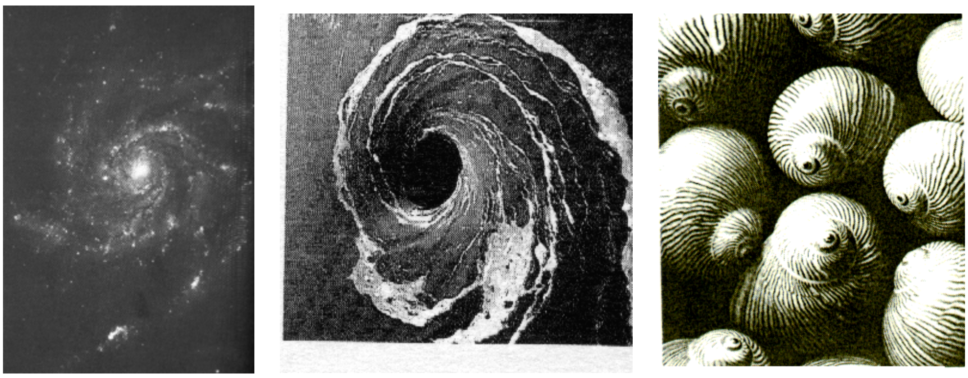


Figure 2. Logarithmic spirals in Nature: galaxies, whirlpools, snail shells.
Source: (Prechter i Frost 2001).

The tail of a comet curves away from the sun in the spiral. Distant galaxies, hurricane clouds, ocean waves and whirlpools swirl in spirals, as do many other natural phenomena. The construction of the logarithmic spiral is based on the Fibonacci ratio $\Phi=1.618$, known as the Golden Ratio or Golden Mean. It defines the ideal proportions.

v. According to the 3-element model, this natural law, permeating the Universe and described by the Fibonacci ratio $\Phi=1.618$, should refer to the dynamics of collective behavior as well. Since adaptations to the exchange of information spiral and financial markets reflect psychology and the dynamics of the crowd, the spiral identified in price formations also should be logarithmic. During the introductory stage we confirmed that, indeed, the top of each successive wave of higher degree on the index chart is the touch point of the logarithmic expansion.

Figure 3 presents how the golden spiral applies to the Polish emerging stock market. The spiral embraces all the vital turning points of the WIG chart, starting at the peak of the first wave of the bull period (March 1994), through a year-long bear market (1994/1995), the second, very extensive, bull market, to the end of the next corrective wave (July 1999).

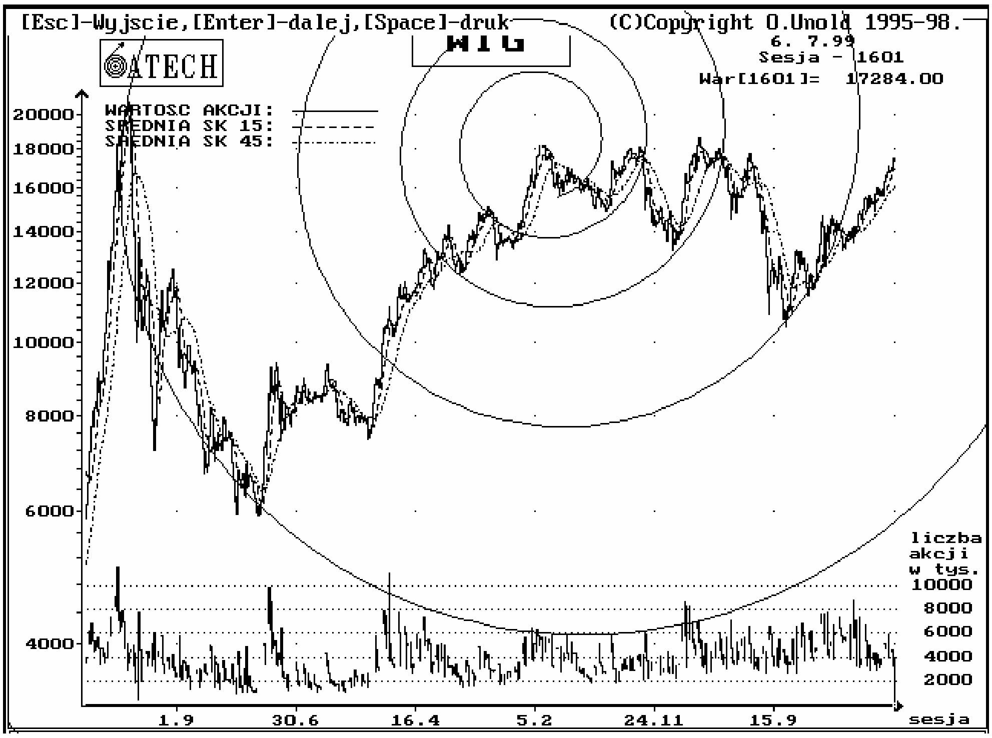


Figure 3. Logarithmic spiral on the Warsaw Stock Exchange Index (WIG)

Source: authors' research

Another figure (Fig. 4) confirms the logarithmic extension at the German DAX index (January 2005 – March 2003).

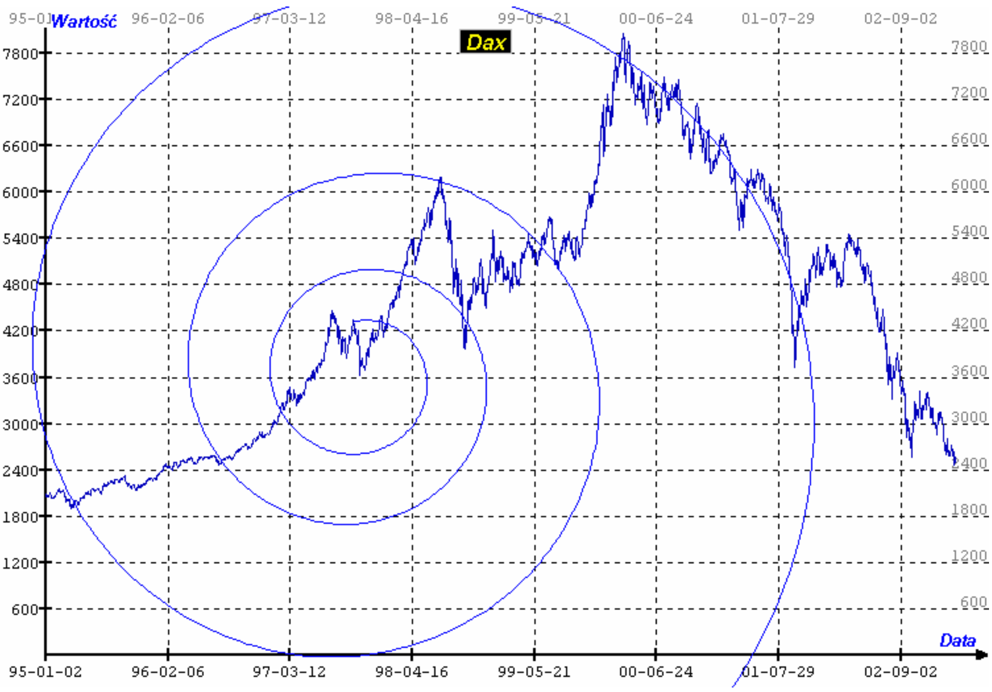


Figure 4. Logarithmic spiral on the German DAX

Source: authors' research

Yet another one (Fig. 5) reveals a similar pattern at the French CAC40, from January 1995 to March 2003.

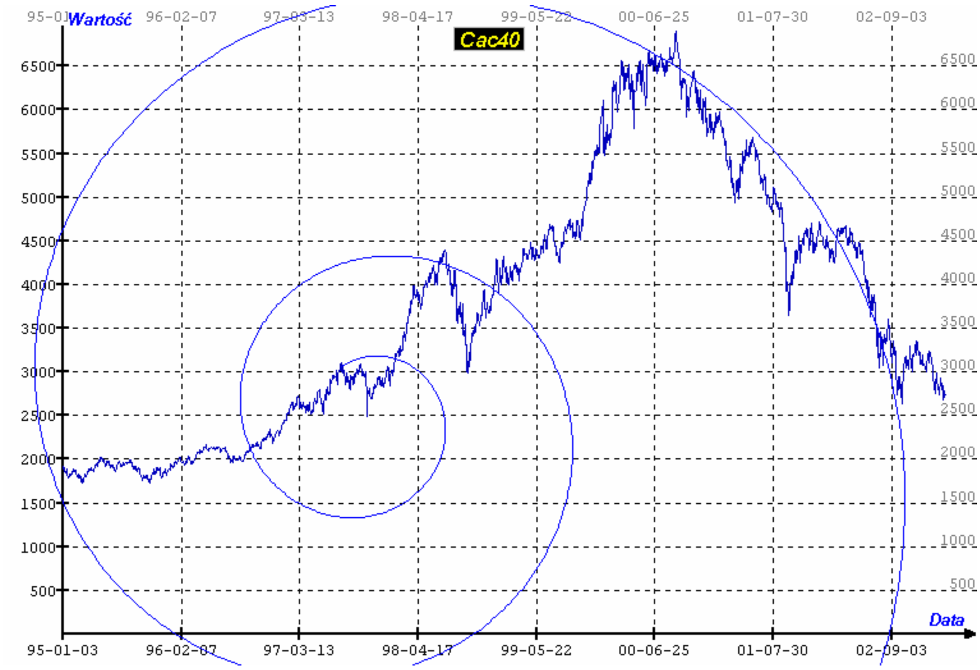


Figure 5. Logarithmic spiral on the French CAC40

Source: authors' research

Last but not least, the Hongkong Hanseng index, analyzed in the time frame June 1987 – April 1999 (Fig. 6).

Figure 6. Logarithmic spiral on the Hongkong Hanseng index

Source: authors' research

vi. The logarithmic spiral is self-similar and isomorphic. It follows that information processes of collectivity are also isomorphic. The identification of isomorphism and self-similarity in the analyzed system is of great importance in the proposed research procedure. The spiral in Fig. 1c represents a new, modified form of the attractor presented in Fig. 1a. This spiral is a metaphorical equivalent of a *fractal attractor (strange attractor)*. This metaphor has a deep theoretical grounds, as a logarithmic spiral actually is a fractal. Most importantly, however, it suggests the occurrence of a certain form of rationality of collective behavior (see the sequence: *fractal-recurrence-collective mind*). Also the identified phenomena of cyclical recurrence and adaptability can be tentatively recognized as an expression of collective rationality. Another research thread which will be used to theoretically explain the identified phenomena and justify the method incorporated in the project is Heisenberg's uncertainty principle, with the example of the dual nature of electrons (unpredictability of behavior on the individual level, predictability on the collective level of an "electron cloud").

INITIAL CONCLUSIONS

The initial observations should be confirmed in the world's most representative markets and this will allow to prove the universality of mechanisms controlling the behavior of complex adaptive systems.

The identification of a fractal attractor (strange attractor) in the model of a social system carries far reaching theoretical and methodological consequences. It implies self-similarity and recurrence of system behavior. Recurring patterns of behavior in an organization are called organizational culture, and the notion of "organizational culture" is used interchangeably with the concept of "collective mind" (Eden and Spender, 2002). Thus, the identification of a fractal attractor in the analyzed social system suggests, on the grounds of Chaos Theory, the occurrence of rationality of collective behavior and defines the model representation of the adaptability of collective behavior – a spiral movement. It is worth to stress that the concepts of *collective mind* and *organizational intelligence* add a crucial qualitative dimension to systems analysis. They add the missing internal social dimension to the technical or mechanistic dimension, which is the focus of the classical theory of systems and organizations.

The initial findings suggest that the decision-making process of collectivity within a model and exemplary Global Information System is adaptive and follows specific patterns found in nature. Therefore, unlike the decision-making process of an individual, this process can be expressed mathematically and ought to be predictable. In other words, individual behavior, which is often irrational and unpredictable, is expected to compose an adaptive, spiral and, thus, predictable process of collective decision-making.

If the proposal theses are confirmed, the main scientific result will be the formulation of a new paradigm of rationality. In the era of globalization and virtualization we shift our interest from traditionally perceived "physical collectivities" to a "dispersed, virtual crowd", which is a totally new social phenomenon. The realization of the project will allow to reach the theoretical grounds of a new paradigm, which will refer to the behavior of crowd and the notion of adaptation as a more natural reaction to information stimuli than optimization. Moreover, adaptation will not exclude traditional optimization. Optimization will remain a specific case of adaptation, applicable to strictly deterministic decision situations. This way, the project will contribute to an understanding of mutual interactions between societies and individuals. It will examine and structure the unique influence that social processes exert on the decision-making processes of an individual. In this sense we will be able to speak of *system rationality*, which should not depend on the rationality or irrationality of the system's components.

REFERENCES

Avison D., Fitzgerald G. (2003): Information Systems Development: Methodologies, Techniques and Tools. Mc-Graw Hill, New York.

Benson S., Standing C. (2002): Information Systems: A Business Approach. John Wiley & Sons, Milton.

Capra F. (1995): Turning Point. Ashgate Publishing Group, Aldershot, UK.

Eden C., Spender J.C. (2003): Managerial and Organizational Cognition. Theory, Methods and Research. Sage Publications, London.

Frost A.J., Prechter R. (2001): Elliot Wave Principle. John Wiley & Sons, New York.

Gray R.M. (1998): Entropy and Information Theory. Willey & Sons, New York.

Halpern J., Stern R. (1998): Debating Rationality: Nonrational Aspects of Organizational Decision Making. Cornell University Press, London.

Jordon D.W., Smith P. (1999): Nonlinear Ordinary Differential Equations. Oxford University Press, Oxford.

Kaufmann S. (1996): At Home in the Universe. Oxford University Press, Oxford.

Morita A. (1994): Made in Japan: Akio Morita and Sony. HarperCollins, London.

Nelson D. L., Quick J. C. (2000): Organizational Behavior: Foundations, Realities, and Challenges. South Western College Publishing, Cincinnati.

Peters E.E. (1996): Chaos and Order in the Capital Markets: A New View of Cycle, Prices, and Market Volatility. John Wiley & Sons, New York.

Plummer T. (1998): Forecasting Financial Markets. Kogan Page limited, London.

Russo N.L., Fitzgerald B., DeGross J.I. (2001): Realigning Research and Practice in Information Systems Development: The Social and Organizational Perspective. Kluwer Academic Publishers.

Weick K. E. (2000): Making Sense of the Organization. Blackwell Publications, New York.

Websites Recommended by Australian Healthcare Practitioners' for use by Overweight and Obese Adolescents.

Rachel Mahncke, Leisa Armstrong
School of Computer and Information Science, Edith Cowan University
rmahncke@student.ecu.edu.au
l.armstrong@ecu.edu.au

Abstract

The general public are increasingly using the Internet to access information to assist in making health related decisions. Given this user demand for electronic health (e-health) technologies, there is a new area of preventative self-care applications emerging globally. E-health is the use of the Internet, mobile devices and other electronic technologies to deliver health care services through "teaching, monitoring (e.g. physiological data), and interaction" (Eysenbach, 2001; Oh, Rizo, Enkin & Jadad, 2005).

An area of e-health of interest to this research is in regards to diet and exercise, as well as the ability to provide ongoing online support. Such e-health applications create opportunities to deliver a number of health benefits to users, including adolescents. Studies indicate that adolescents are actively seeking health related information and support online. This paper will discuss some Internet enabling applications available online, as well research results to ascertain the availability of e-health applications for use by overweight and obese adolescents in Australia.

Keywords
Websites, e-health, adolescents, health, WWW, Internet applications, online support.

INTRODUCTION

Obesity has become topical in the popular media as it has been declared as a worldwide epidemic (Popkin & Monteiro, 2003). The occurrence of overweight related problems in Australia is considered to be of a "great health, social and economic concern" (Sport Industry Australia, 2003). The main health related concerns to being overweight or obese, is the increased risk of developing serious preventable health related problems and diseases. Some of these health issues and diseases can begin in childhood or adolescence, which can increase the risk of mortality (Freedman, Dietz, Srinivasan & Berenson, 1999). This risk can be reduced by 90% if exercise is increased and weight is lost (O'Brien, 2004). The economic cost of treating obesity related diseases is expected to increase together with the incidence of obesity, "unless effective intervention is successful" (Sport Industry Australia, 2003). However, these health risk factors are often not evident in adolescents and it is therefore the "psychosocial problems" that are the most common problems associated with adolescent obesity (Australian Institute of health and Welfare, 2003). An adolescent for the purposes of this research is defined as a person between the ages of 10 and 19 years old (World Health Organisation (WHO), 2004).

Thus the general public are increasingly using the Internet to access information to assist in making health related decisions (Forkner-Dunn, 2003). One emerging solutions in healthcare are electronic health (e-health) initiatives, such as online and website applications made available via the Internet, mobile telephones or other electronic devices. This research has investigated whether environmental causes of obesity could be positively affected by using such e-health initiatives to re-educate adolescents about healthy eating and exercise behaviours (Snowball, 2004). As a response to this user demand for e-health technologies, there is a new area of preventative self-care applications emerging globally.

Numerous definitions exist for the term e-health (Pagliari, Sloan, Gregor, Sullivan, Detmer, Kahan, Oortwijn and MacGillivray, 2005). However, one definition which provides a commonly expressed theme is that it is "the use of information and communication technology to improve or enable health and health care" (Wysocki, 2001). E-health therefore, is the process of providing healthcare via electronic means, in particular over the Internet. It can include "teaching, monitoring (e.g. physiological data), and interaction with healthcare providers, as well as interaction with other patients afflicted with the same conditions" (Eysenbach, 2001). Another reason that people use electronic technologies for "health-related concerns" is to interact with others who have "similar conditions for the purpose of sharing experiences and emotional support" (Dickerson, Flaig and Kennedy, 2002).

This paper will discuss research undertaken to ascertain the availability of diet and exercise e-health applications available in Australia for use by overweight and obese adolescents. The aim of the research was to investigate, from a healthcare practitioners' perspective, whether online and e-health applications could provide support for and be beneficial to, overweight and obese adolescents. The research identified the online and electronic resources healthcare practitioners' recommended to their adolescent patients, in order to understand how information systems solutions could better assist these patients in achieving healthier lifestyle outcomes.

E-HEALTH EXPLAINED

A report prepared by Mitchell (1999) for the Department of Communications, Information Technology and the Arts (DOCITA) places e-health within the information economy context as shown in Figure 1. E-health is an expansion on what was formerly known as telemedicine (Eysenbach, 2001). E-health therefore is part of the e-commerce industry and encompasses, but is not limited to, telemedicine, telehealth and health information as listed in Table 1 below. E-health creates opportunities to deliver a number of health benefits to users, including adolescents.

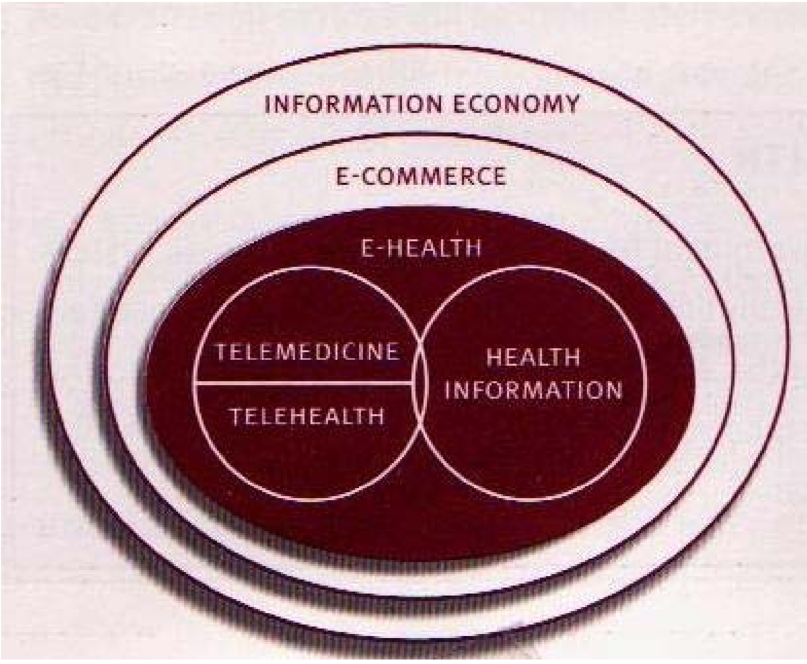


Figure 4: Relationship between e-commerce, e-health, telemedicine (Mitchell, 1999, p. 15).

Table 1 below, adapted from Mitchell (1999) lists the technologies and applications which are utilised by e-health applications.

Table 3: Technologies that are increasingly used in e-health (Adapted from Mitchell, 1999, p.13).

Examples of information and telecommunication technologies used in health	E-health examples utilising these technologies

1. World Wide Web 2. Mobile phone devices 3. Personal digital assistants (PDAs) and other hand held devices 4. Telephone 5. Fax 6. Email 7. Videoconferencing 8. Teleradiology technologies 9. Telepathology technologies 10. Computer servers, hosting databases of health information, for access via the Internet or an Intranet 11. Computers providing facilities for sorting, storage, retrieval of health information, including patients' records and medical images, for accessing via LANs or WANs 12. CD ROMs for storage of health information and educational data	1. Call centres 2. Websites and services to hand-held devices 3. Combination of Internet and CD ROMs for continuing medical and health education 4. Nurses using hand held computers or mobile phones used to access databases on servers 5. Telemonitoring devices in homes, linked by phone to health experts 6. Remote ECG analysis 7. Teleradiology combined with videoconferencing

ONLINE INFORMAL HEALTH EDUCATION.

The Australian Bureau of Statistics (2002) reports that up to forty percent of the Australian population have Internet access from home. The Internet is also being increasingly provided for free in most schools and public libraries. This means that a large percentage of the population, including adolescents, can gain access to this medium. A widely reported advantage of the Internet is that it extends across socio-economic and geographic boundaries. This creates opportunities for e-health to be delivered into cities and rural communities alike, as well as across all age and cultural demographics.

Available on the Internet there are numerous examples of e-health websites devoted to rare diseases such as; diabetes, depression, eating disorders and various forms of cancers. Their sustained existence suggests that consumers believe that there is a need for online information and support to be made available. Internet applications exist, which have been designed and developed to support and assist in adolescent weight loss. Currently adolescents can gain information about eating disorders, such as: anorexia nervosa, bulimia nervosa, and binge eating; as well as about obesity, nutrition and exercise from numerous global health-related websites on the Internet.

A study by Skinner, Biscope, Poland and Goldberg (2003) attempted to establish the emerging roles for healthcare practitioners by utilising e-health applications more effectively for the adolescent cohort. They have identified three emerging roles for healthcare practitioners in e-health: Firstly, it must provide an interface for adolescents that will assist them in quickly finding the information they are seeking (Skinner et al., 2003). Content that is designed and tailored towards adult users is considered too detailed for adolescents (Flicker, Goldberg, Read, Veinot, McClelland et al., 2004). The result is that they either do not seek this content or skim through without gaining the knowledge they require; secondly, websites must provide the means for practitioners to connect with adolescents by extending the ways and times in which they operate (Skinner et al., 2003). For example, making use of e-health applications to interact with adolescents either via e-visits, scheduled times in chat forums or by e-moderation. However, busy practitioners possibly lack the time to deliver these services to adolescents; thirdly, websites must provide the means for practitioners to assist adolescents in becoming critical about evaluating the quality of health-related information that is available (Skinner et al., 2003).

Lasker (2005) raises a number of concerns about the disadvantages of relying on the Internet for information and support. This research investigated 79 research studies relating to the "accuracy, completeness and comprehensibility" of online information and support. The research concluded that the health related websites evaluated, were not always reliable and were regarded negatively by the researchers (Lasker, 2005). The reasons being are that these sites often lacked the sensitivity of face-to-face communication, and produced large amounts of unrelated information and mail. The result being that the users needed to read large amounts of information before they gained any useful benefit from the resources.

However, another study which surveyed doctors concluded that there were benefits for patients that used the Internet for information, advice and social support (Potts & Wyatt, 2002). This study surveyed 800 "Web-savvy" doctors in the United Kingdom, to ascertain their perceptions about the actual benefits and possible harm to their patients by using the Internet (Potts and Wyatt, 2002). The majority of doctors reported that there were more benefits to patients using the

Internet than the possible harm it could cause (Potts & Wyatt, 2002). Weiner (2003) believes that adolescents are equally ready to learn in "cyberspace". The Internet offers an opportunity for adolescents to improve and develop their "health literacy" (Gray, Klein, Noyce, Sesselberg & Cantrill, 2005). However the issues of finding high quality information on the Internet that is presented in a quick and readable format, and targeted towards adolescents, remains the overriding issues with ensuring the usefulness of the Internet as a health tool for adolescents.

SURVEY OF HEALTHCARE PRACTITIONERS

This research either emailed or posed a fourteen question survey to a purposeful sample group of 1100 Australian healthcare practitioners across all states and territories, in both rural and metropolitan areas in August and September of 2005. The research had a multidisciplinary approach as practitioners from numerous professions were identified as potential respondents. The research predominantly targeted general practitioners (GP's), dietitians, and psychologists, but was also inclusive of: Aboriginal health experts, cardiologists, child health experts, exercise physiologists, health promotion researchers, homeopaths, medical practitioners, naturopaths, nutritionists, nurses, obesity experts, paediatricians, psychiatrists and sports dietitians. A total of n = 121 responses to the survey were received, equating to an 11% response rate. 75.1% of responses were received from the three targeted professions as is indicated in Figure 2. The remaining 24.8% of responses comprised of a variety of professions, listed above, and referred to in this research as "Other".

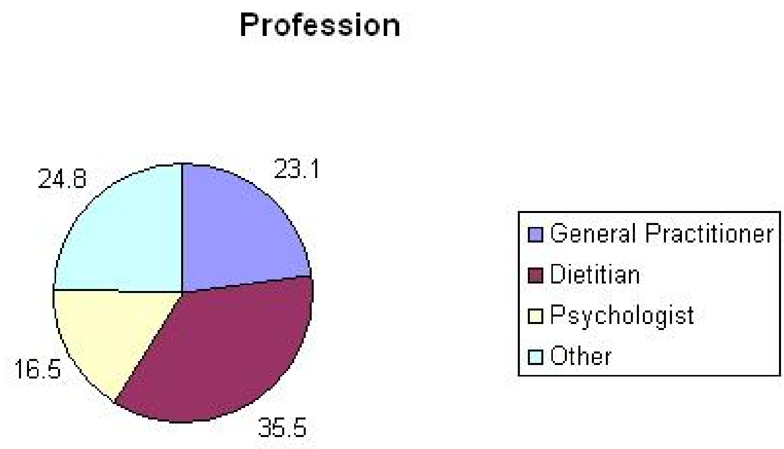


Figure 5: Percentage of responses per professions

PRACTITIONER RECOMMENDATIONS

The survey asked respondents what type of electronic resources they recommended to their overweight and obese adolescent patients. A list of possible electronic resources was provided as well as providing the ability for the respondent to indicate their own type of resource. Only half the 121 respondents answered this question. Of those that answered the question, 45.5% indicated that they did not recommend any electronic resources, including Internet resources as shown in Figure 3. Seventy percent of survey respondents were unsure as to whether there were sufficient online and electronic resources available for use by overweight and obese adolescents. The main reason for not recommending electronic resources included not being aware of available resources or not being confident with the resources that are available online. Therefore, approximately 25% of respondents to this survey indicated that they recommend electronic resources to their overweight and obese adolescent patients.

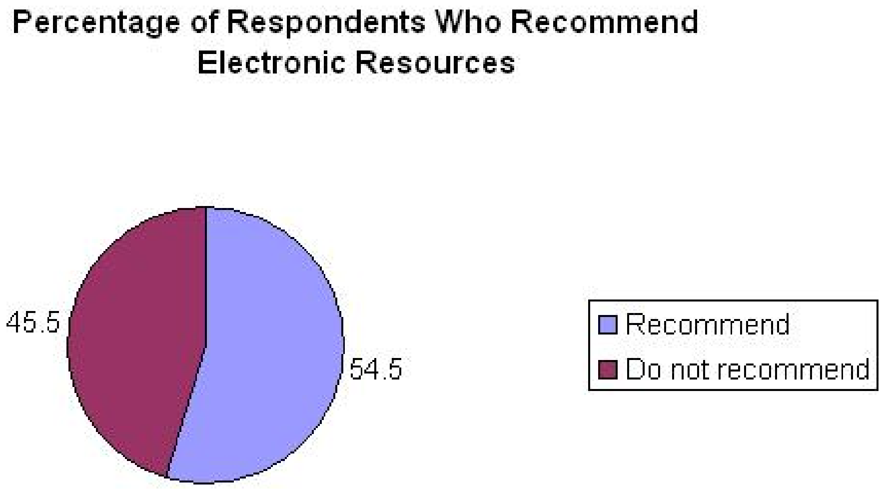


Figure 6: Percentage of respondents who recommend electronic resources

Of the 54.4% of respondents who recommend electronic resources to their overweight patients, 49.1% recommend websites. Information in electronic form was the second highest selection at 18.9%, and the remainder of the electronic resources are depicted in Figure 4.

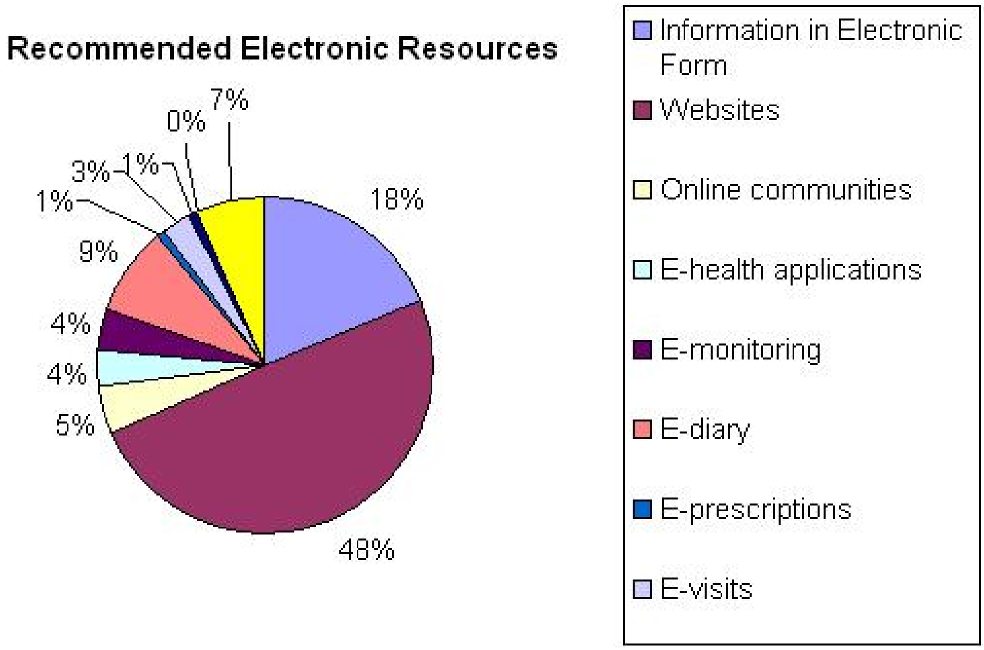


Figure 7: Recommended Electronic Resources

Respondents, who recommend electronic resources, were asked for the address and contact details of the resources they recommended to their overweight patients. The websites listed in Table 2, indicates the Internet location of the recommended electronic resources. The majority of the forty four recommended websites provide trusted factual health information. The websites cover a range of topics including; healthy eating, diet, exercise, mood and depression treatments and training, government websites, e-commerce websites, sales of health products, and weight loss surgery sites. Referring back to the definition of an e-health application, on evaluation of each of these sites, it was discovered that many of the sites fulfilled the necessary e-health requirements. These requirements include functionality beyond static factual content, such as interaction, support, online chat with both members and experts, message boards, online communities, recipes, games, services to mobile devices etcetera.

Table 4: Websites that Australian healthcare practitioner's recommend to their overweight and obese adolescents patients.

www.weightwatchers.com.au	www.activeforlife.com.au
www.nutritionaustralia.com.au	www.sportsdietitians.com
www.betterhealth.vic.gov.au	www.fitnessaustralia.com.au
www.woolworthsfresh.com.au	www.dairycorp.com.au
www.daa.com.au	www.sanitarium.com.au
www.ifnotdieting.com.au	www.foodwatch.com.au
www.beyondblue.com.au	www.dietclub.com.au
http://moodgym.anu.edu.au	www.goforyourlife.vic.gov.au
www.dhs.vic.gov.au	www.perthdietclinic.com.au
www.betterhealth.vic.gov.au	www.glycamicindex.com
www.chw.edu.au	www.weightloss-in-australia.com
www.obesityguidelines.gov.au	www.diabetesaustralia.com.au
www.asso.org.au	www.quackwatch.com
www.fitness2live.com	www.healthyeating.org
www.cyh.com	www.weightlosssurgery.com.au
www.dietworld.com	www.heartfoundation.com.au
www.gudhealth.gov.au	www.pharmacydirect.com.au
www.coeliac.org.au	www.diabetesaustralia.com.au
www.mypyramid.com	www.beyondblue.com.au
http://optimalhealth.com.au	www.healthinsite.gov.au
www.med.umich.edu/umim	www.eda.org.au
www.dairycorp.com.au	www.med.monash.edu.au/non-cms/mentalhealth

The literature review conducted prior to the administration of the survey revealed some additional websites that provided the required health related content and functionality, but which were not mentioned by the healthcare practitioners in the survey. These websites are listed in Table 3 below.

Table 5: Additional websites identified by the researcher but not mentioned as being recommended for use by healthcare practitioners.

www.healthybodyclub.com.au	www.onlineslimmingclub.com
www.mydr.com.au	www.blackdoginstitute.org.au
www.oa.org	www.healthyeatingclub.org
www.eatingdisorders.org.au	www.weightloss-in-australia.com
www.healthinsite.gov.au	www.blubberbuster.com
www.kidshealth.org	www.onlineslimmingclub.com
www.healthyactive.gov.au	www.click2quit.co.uk
www.healthboards.com	

There are numerous Australian government websites which provide excellent resources for adolescents on healthy lifestyle choices. For example, health information resources are available via the health department websites, with links to *Healthinsite*, Nutrition Australia and the Health Body Club. This indicates that the government has understood the need to provide these resources to the public and have attempted to meet this user need and demand. This research indicates that there is a sizable amount of factual information relating to health, diet and exercise available.

However, given that 45.5% of respondents indicated that they do not recommend any electronic resources, this may indicate the need to incorporate the opinions of, and include healthcare practitioners when developing such resources in order to ensure the success of such application and websites.

EVALUATION

The ideas relating to online support for this cohort of adolescents that were offered by the healthcare practitioners are already incorporated into numerous health websites. The healthcare practitioners in general did not indicated ideas that included more technologically advanced functionality, and which utilises trends such as adolescents' passion for mobile devices. Services to mobile devices are available from a number of websites in the form of SMS messages being delivered to teenagers. For example, *MyPyramid* a free e-health application made available through the United States Department of Agriculture provides services to mobile devices which enable users to input data and view the details graphically. Also, pharmaceutical companies, such as Galderma Australia, deliver text messages to subscribers in regards to acne information and reminders to administer the treatment correctly.

This research identified the application *MyPyramid* to be the epitome of a preventative self-care e-health application, such as does not exits within Australia. Once the user has entered their age, sex and level of physical activity, then the site provides the user an eating pyramid applicable to their requirements. The site then provides meal and exercise tracking worksheets and an invitation to anonymously log on to obtain the *MyPyramid* Tracker functionality. It is this functionality which is makes this site impressive.

The Tracker offers assessment of both the user's food intake and physical activity. Selecting either of these options offers the user access to analysis screens. The food intake option for example, once the calorie intake has been captured for the day, then allows analysis of energy intake, and relates this to required vitamin and minerals intake, correct food group choices and overall attempt to indicate where the user requires assistance. This application can be used on a daily basis, and with discipline could be used very effectively to implement healthy behaviour choices. Unfortunately, it appears that an Australian equivalent of the My Pyramid application is not available, specifically in regards to the metric system information.

The second site worthy of mention is Australian National Universities (ANU) MoodGYM training program (Groves, Griffiths, & Christensen, 2003). Research indicates that depression is one of the possible symptoms or causes of overweight and obesity. This web application, run within a browser, delivers cognitive behavioural therapy for the treatment of depression. The application is part of a research project into depression and first guides the user through a series of questions in order to deliver a customised program suited to the user's needs. The aim of this e-health application is to modify the user's behaviours in relation to their depressive thinking and to prevent and decrease symptoms of depression (Groves et al, 2003). The benefit of mood gym over the traditional method of reading books for self-help is that it is interactive in providing real time feedback and advice to users.

There are numerous other websites which offer excellent factual health related information such as *www.healthboards.com, www.mydr.com.au* and *www.healthinsite.gov.au*. These websites, which are often free, allow registered members to connecting with others who may be going through the same experience. There has been much discussion and research relating to the usefulness of online chat rooms in relation to making a real health contribution to users. Such facilities often provide moderators but still require the user, especially if they are feeling vulnerable, to be cautious.

CONCLUSION

This research has concluded that there is sufficient factual health related content available on the Internet, provided predominantly by websites. These websites offer healthy sensible eating and exercise related advice online. The information in electronic format, identified by the survey respondents, is suitable for use by overweight and obese adolescents. However, this research concludes that these resources are not provided in an interactive, age appropriate format. Adolescents need information to be presented to them in a manner in which they can easily locate and that is the relevant information which they seek. In addition rural respondents indicated that English is often not the adolescents' home language, although it is the language most often utilised by Australian websites to present this health related content. Rural respondents also indicated that their patients would appreciate more pictures instead of words.

The review of existing websites and survey of healthcare practitioners' knowledge of available Internet resources suggests that there is a need to develop online health resources specifically for adolescents in Australia. The related resources of interest to this study are preventative self-care applications that incorporate diet, health and exercise components. Current resources appear to provide mostly adult content in a predominantly static information-related environment. However, majority of websites reviewed did not provide the degree of interactivity, or customised interfaces that adolescent users require.

Provision of chat facilities remains controversial due to the possibility of misinformation and harm that can result in unmonitored environments. A concern that was raised in the literature and by a number of healthcare practitioners', relates to adolescents spending too much time in front of a computer, and not being outside playing and being physically active. Adolescents need to be made aware of appropriate time restrictions in relation to time spent on computers and watching television. Equally, educating adolescents about the benefits of exercise in terms of its ability to improve health outcomes and as a cure for mild depression would be beneficial to them. Given adolescents passion for technology it is anticipated that this cohort would be quick to adopt new e-health initiatives.

REFERENCES

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, 3(2), e20. Retrieved June 29, 2005 from http://www.jmir.org/2001/2/e20

Dickerson, S. S., Flaig, D. M., & Kennedy, M. C. (2002). Therapeutic connection: help seeking on the Internet for persons with implantable cardioverter defibrillators. *Heart Lung, 29*(4), 248-255. Retrieved June 12, 2005 from ProQuest Database.

Forkner-Dunn, J. (2003). Internet-based Patient Self-care: The Next Generation of Health Care Delivery. *Journal of Medical Internet Research, 5*(2), e8. Retrieved Accessed April 27, 2005 from ProQuest Database.

Gray, N.J., Klein, J.D., Noyce, P.R., Sesselberg, T.S., & Cantrill, J.A. (2005). The Internet: a window on adolescent health literacy. Journal of Adolescent health,.37(3), 243.

Groves, C., Griffiths, K., & Christensen, H. (2003). Working out MoodGYM: A user's guide. Centre for Mental Health Research: Canberra, 2003.

Hansen, D. L., Derry, H. A., Resnick, P. J., & Richardson, C. R. (2003). Adolescents Searching for Health Information on the Internet: An Observational Study. *Journal of Medical Internet Research, 5*(4), e25. Retrieved Accessed April 27, 2005Database.

Mitchell, J. (1999). The unstoppable rise of e-health. Report prepared for the Department of Communications, Information Technology and the Arts (DOCITA). Retrieved September 22, 2005 from http://www.jma.com.au/tmed99.htm

Oh H, Rizo C, Enkin M, Jadad A (2005). What Is eHealth (3): A Systematic Review of Published Definitions. *Journal of Medical Internet Research*,7(1):e1. Retrieved June 29, 2005 from http://www.jmir.org/2005/1/e1

Skinner, H., Biscope, S., Poland, B., & Goldberg, E. (2003). How Adolescents Use Technology for Health Information: Implications for Health Professionals from Focus Group Studies. *Journal of Medical Internet Research, 5*(4), e32. Retrieved Accessed April 27, 2005Database.

United States Department of Agriculture (USDA). (2005). My Pyramid. Retrieved August 5, 2005 form

http://www.mypyramid.com

White, M. H., & Dorman, S. M. (2000). Online support for caregivers. Analysis of an Internet Alzheimer mailgroup. *Comput Nurs, 18*(4), 168-176. Retrieved June 12, 2005Database.

World Health Organisation. (2004). *Adolescent Health and Development*. Retrieved March 27, 2005, from http://www.who.int/child-adolescent-alth/OVERVIEW/AHD/adh_over.htm

Weiner, C. (2003). Key Ingredients to Online Learning: Adolescent Students Study in Cyberspace – The Nature of the Study. *International Journal on E-Learning 2*(3), 44-50. [Online]. Available: (CD-ROM), San Diego, CA, USA.

An Intelligent System for Diagnosing Tourette Syndrome

Dr. Alberto Ochoa O. Zezzatti
Zacatecas University
Zacatecas, Mexico
Email: megamax8@hotmail.com

Mr. Amol Wagholikar
Griffith University
Gold Coast, Australia
Email: a.wagholikar@griffith.edu.au

Abstract

The process of medical diagnosis is always complex. It demands the valuation of multiple interacting factors in a patient's case. One or more experienced doctors propose a corresponding treatment after observing the signs and symptoms of the patient. With the advent of Artificial Intelligence (AI) techniques such as Case-based reasoning (CBR), education of medical subjects has become effective. CBR plays an important role in building intelligent systems for disease prognosis and diagnosis. We discuss a CBR-based intelligent system built for diagnosing Tourette syndrome. The results of the proposed system justify its usefulness for diagnosing Tourette syndrome.

Keywords:

Tourette syndrome, Disease diagnosis, Artificial Intelligence, Case based reasoning, Intelligent systems

INTRODUCTION

There has always been a continuous evolution in the field of the health. National System of Mexican Health has developed a plan of action for increasing the quality of people's life by continuously updating the health sector professionals through innovative medical education systems. Human resource and its development are always important for a nation's growth. In the modern era, computer systems are an important ally for this purpose. Healthcare forms an integral part of improving quality of human life. Integration of modern intelligent systems for medical education plays a significant role. Tourette Syndrome (TS) is a widely known syndrome (Anderson, 1996). Hence an intelligent system built for educating health professionals about TS can be a very useful system. We term such intelligent systems as Intelligent Tutoring Systems (ITS). In this paper, we would use these two terms interchangeably. We believe such system would also assist the healthcare professionals in patient's health information management. We discuss our intelligent system in this paper.

This paper is mainly organized into various sections. In the next section, we describe the Tourette syndrome, its causes and importance to build an intelligent system for providing medical education of TS. The following section discusses the development of our system. After that we discuss the proposed software and results given by our system. We conclude our work in the last section.

TOURETTE SYNDROME DESCRIPTION

In 1825 the first case of Tourette Syndrome (TS) was registered in medical Literature (Burden, 1996). Marquise de Dampierre showed the symptoms such as involuntary tics in many parts of the body, several vocalizations including echolalia and coprolalia. In 1885, a French neurologist Dr. Georges Gilles de la Tourette described a disorder characterized by tics. In the present century, many studies have been undertaken, in search of an explanation and solution of the above disorder. These studies show that 1% of the general population is affected by this disorder (Pearce, 1996). The causes of this disorder can be organic, hereditary or psychological. The "tics" are short duration involuntary movements. They are non-rhythmic, abrupt, repetitive and painful (Wagaman et.al, 1995). It is possible to suspend these movements for a limited time. However, it requires great amount of will power from the patient. Usually they worsen under conditions of stress, anxiety, annoyance and fatigue. They can also occur when something unpleasant is anticipated (Bados, 1995). They can also be induced by the presence of certain stimuli such as cough or gestures of another person. However, its effect is reduced in the presence of strangers or certain physical activities like playing or doing repairs (Vera, 1995). The tics and the TS can be associated to many other problems such as obsessive-compulsive upheavals, atentiocional deficit disorder with or without hyperactivity, problems of language, difficulties in the control of the impulses and dream disorders (Scotti et. al, 1994). In order to prevent as well as make a correct diagnosis of this syndrome, it is important to have proper education among the healthcare professionals. This can be achieved through Intelligent Tutoring Systems, which were developed to overcome the deficiencies in the computer-based education systems. In the next section, we discuss about the computer-based education systems.

Computer-based education systems

The computer-based education systems (e.g. linear programs, graft programs, generative systems) are also known as Computer Assisted Instruction Systems (CAIS) (Lopez, 1993). The main deficiencies of the CAIS are:

- They try to include complete courses instead of limiting itself to concrete subjects.
- Barriers of communication between the tutor and the student, which restrict the interaction among them.
- Students do not have knowledge of how and why the tasks are executed. Also the programs reacts using a series of anticipated situations independent of student's answers.
- It is not possible to transfer these systems from one domain to other. They are domain incompatible.
- These systems tend to be static instead of evolving and dynamic.
- In these systems the knowledge is not updated regularly.

In summary these are expensive and repetitive programs in which there is no relation between what is taught and how it is taught. Due to these problems and efforts made by certain researchers in this area, Intelligent Tutorial Systems (ITS) are developed. The ITS combines techniques of Artificial Intelligence (AI) with psychological models of the student and the expert. It also involves application of theories of the education (Capponi, 1987). We are proposing application of ITS for the education of Tourette syndrome using an AI technique known as Case-based Reasoning (CBR). In next section, we present a brief survey of modern ITS.

Examples of ITS

In relation to Medical Sciences, the university cannot teach all the required knowledge, facts and abilities without informing new discoveries and changes in the medicinal science. Internet offers vast information on the new discoveries in the form of technical advances in the medical sciences, on-line educational sources with clinical histories, images and findings in different cases from various medical institutions. Users can surf all the existing advances in the world of telemedicine through a virtual university. With such advances in computer science, many applications are developed for medical education (Mann et. al, 1996).

A Paediatric Hospital in Germany has applied Case-based reasoning for knowledge acquisition (Fisher et. al, 1996). They developed a program called "Casus" to resolve the problem of education in medicine using a case library taken from real medical practice. The program assists an apprentice to determine a diagnosis using its in-built case-library. This program gave satisfactory results and hence it is believed that this system can be used daily in future. Another example of such case-based system is a system involving case library with psychosocial aspects of the patients with breast cancer. The students learn from this system using "game of rolls". This system represents "patient" and "specialist" using a special style of presentation and is applicable to other types of cancer as well. Also it offers useful recommendations to the doctor (Sleeman, 1982).

In Italy, CBR systems are applied for medical education especially for learning and diagnosis of congenital malformations (Ochoa et. al, 2005). From the above examples, it is clear that case-based systems are quite useful systems for medical education. We propose to extend the application of such systems for the education of Tourette syndrome. Hence we develop an approach of Intelligent Tutoring for the education of Tourette syndrome. We further justify the need of such system in the next section.

Justification

Although the Tourette Syndrome is not a worldwide problem of health, still proper care must be taken to avoid serious complications of this syndrome. Hence it would be useful to provide necessary information about the syndrome. Also CBR involves solving new problems using the solutions from the previous cases. It is a suitable approach for medical diagnosis as it involves proposing a corresponding treatment using past cases of the patients. This system is certainly valuable in the absence of a Psychiatrist or a Neurologist at the first level of attention in various societies (Polare, 2005). Medical education systems must have high quality. In the context of undertaken development, the system will involve compilation and retrospective revision of material sources i.e. clinical histories of the real cases from psychiatric medical practice (Mink, 1998). In order to define strategies and content of the system, the organization and analysis of the information will require the consultation of different specialists and experts, as well as the corresponding bibliographical revision. We discuss the development of our system in the next section.

Development of the System

The system will justify the solution from the similar cases. The knowledge acquisition process involved at least 35 cases mainly originating from the consultations of medical experts at the first level of attention in different areas. These cases were presented to the expert. The experts performed diagnosis, prognosis and the medical conduct. The resident specialist diagnoses these cases. The specialists are either from Psychiatry or Neurology area. The expert confirms the diagnosis and then only it is incorporated in the system.

Methodology

To formulate a diagnostic conclusion in Psychiatry is a complex clinical problem. Also it is difficult to obtain the same efficiency as an expert, with diagnosis of only few cases. Hence to obtain the efficiency of the expert, 87 % of the diagnosis given by the system must match with expert's diagnosis. Elements of observation by the expert must also be considered. Large number of observations must be conducted. Due to this complexity, there is always a scope of improvement to make the system as perfect as possible in future investigations.

Categorizing the diagnosis, prognosis and the medical conduct can test the specificity and sensitivity of the system. The proposed instrument will be a medical aid tool, which would be useful in orienting the medical conduct of TS. It would be more sensible and little specific to the variations in the phenomenon under study and the user's valuations. This is an important result, which justifies use of such tool. Also it is guaranteed that this tool will have a capacity of giving accurate results in more than 80 % of the diagnostic considerations.

The case library in the proposed tool would consist of use of interpretative type explanations and using archives with a similar extension (Kolodner, 1997). It is feasible and easily transportable between different domains. It uses the option of internal correlations of the program and estimation of the weights assigned to the predicting characteristics. It defines 25 predicting characteristics and 3 objectives of the system. Those objectives are:

i. The diagnosis in the 9 proposed groups

ii. Directions of the medical conduct to follow

iii. User's valuations with a total of at least 100 cases in its case library.

The rationale behind the above predicting characteristics and objectives is to achieve maximum agreement between the system's diagnosis with an expert's diagnosis. With this background discussion, we now present the development of the proposed intelligent system in the next section.

DEVELOPED INTELLIGENT TUTORING SYSTEM

The intelligent system for diagnosing Tourette Syndrome (SIDSTOU), showed in this paper, is a program developed in Java using Jbuilder 9.0 (Nieto, 2002). It uses extension files *.sto that constitutes the case library. Its application was very useful for the creation of the case-based system. It uses the option of internal correlations of the program and estimation of the weights of the predicting characteristics. It defines 25 predicting characteristics (See Figure 1) and 3 objectives of the system that are: the diagnosis in the 9 proposed groups, directions in the medical conduct to follow and user's valuations with a total of 355 cases in its case library. The following figure shows the software interface.

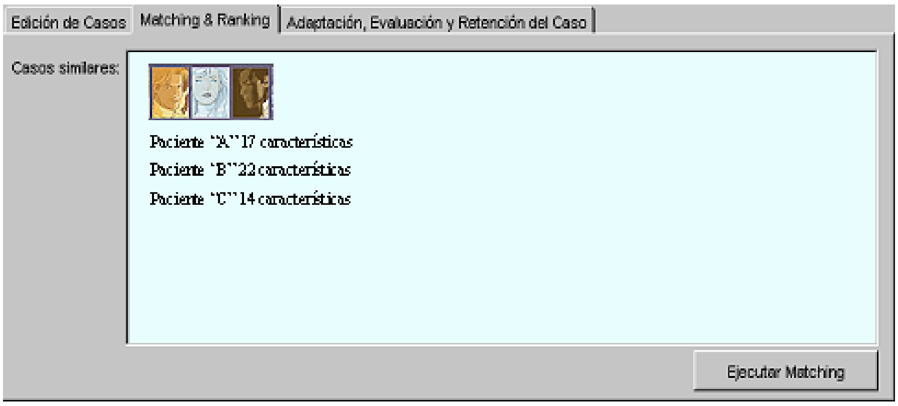


Fig. 1. Developed Software Interface

The system justifies its solution from the similar cases. It provides auto learning mechanism-using values of the predicting characteristics. It involved consultations of medical expert. After this initial knowledge acquisition, 100 more cases unknown to the system were added. These cases were associated with psychiatric upheavals to the TS in adults. After constructing the case-base, the software gave the following results:

- 87% of efficiency for the diagnosis
- 89% for the medical conduct to follow
- 92% for the user's valuations.

These results are considered good. 87 % of effectiveness is achieved after comparing the expert evaluations. These evaluations were obtained during the knowledge acquisition phase. It meets the aims of the medical aid. It assists the doctor in consultation. These results show the effectiveness of this software over the systems designed for study of the Tourette syndrome at the international level. We discuss the results in the next section.

RESULTS

The following results are obtained by the proposed ITS.

1. In relation to specificity and sensitivity of the system, sensitivity from the instrument was 88.88%. Specificity of 85.45% with a predictive value for the true positives of 83.33% was obtained. Specificity with a predictive value for the truly negative ones was 90.38%.

2. In relation with the treatment, sensitivity of system was 93.5%. 78.57 % specificity was obtained. Specificity with a predictive value for true positives was 91.78% and for truly negative ones was 81.48%.

3. In case of prognosis, one obtained a sensitivity of 98.90%. Specificity of 22.22%, with a predictive value for the true positives was 92.78% and for truly negatives was 66.6% [See Table 1].

4. With relation to the number of cases, a case library of 18, 36, 45, 72, 135 and 176 cases respectively was tested with the expert evaluations. The Chi-square test verified that the variables, number of cases and correct answers given by the system for the given three objectives had statistically significant relation for a smaller P of 0.001. The Spearman tests of correlation were not statistically significant. The exact criterion about the number of optimal cases is not decided at present for a greater effectiveness of the system. The amount of correct answers increases with increase in number of cases [See Table 2].

Table 6: **Performance evaluation by the expert for 100 cases**

Group	VP	VN	FP	FN	S%	E%	Predictive Value VP %	Predictive Value VN %	Efficiency
Personality T.	40	47	8	5	88.88	85.45	83.33	90.38	87%
Ambulatory T.	67	22	6	5	93.05	78.57	91.78	81.48	89%
Reserved Prognosis	90	2	7	1	98.90	22.22	92.78	66.66	92%

Where, Label:

VP: true positive. VN: true negative. FP: false positive.

FN: false negative. S: sensibility. E: specificity.

Predictive value of truly positives. Predictive value of truly negatives.

Table 7: **Comparison of agreement between the experts and the human specialist for each one of objective characteristics**

Number of Cases	Agreement in diagnosis	No agreement in diagnosis	Agreement in medical conduct	No agreement in medical conduct	Agreement in Prognosis	No agreement in prognosis
18	1	17	5	13	11	9
36	8	28	18	18	29	7
45	11	34	27	18	27	18
72	26	46	63	9	69	3
135	113	22	119	16	122	13
176	154	22	157	19	162	14

Chi-square for the diagnosis =175.641. Chi-square for medical conduct=81.669.

Chi-square for the prognosis=51.735. DF=5 Smaller P of 0.001.
Correlation Coefficient of Spearman for diagnosis=0.11595.
Correlation Coefficient of Spearman for medical conduct =0.34786.
Correlation Coefficient of Spearman for prognosis=0.144
Critical value -/+=0.81165 for a smaller P of 0.05

CONCLUSION

The proposal of using the case library for the education of the patient with Tourette syndrome is based on the set of aspects that are enunciated next and that constitute results of our investigations.

1. The proposed case library is integrated by 28 characteristics with 288 possible values or domains and will have to contain at least 100 cases. It should use the criteria of Moriyama for the validation of the content.

2. Relation of statistically significant association with the test of Chi-square between the numbers of cases of case library exists.

3. The created software has an effectiveness of 87% in his diagnostic considerations, 89% in his capacity of direction in medical conduct to follow and of 92% in its predicted valuations. These results are considered good in comparison with the results obtained by the other authors (Bichindaritz, 2003).

4. The expert who sets out to create tool of medical aid for the diagnosis must be more specific. The expert's capacity of direction in the medical conduct to follow should be more sensible.

5. It is tried to widely value the utility of the expert who plans to create a tool of medical aid for the boarding of the patient with Tourette syndrome having psychiatric upheaval.

6. This tool with an acceptable effectiveness conserves the integrity of the knowledge and provides better results. It is reliable and relatively simple for its application.

The above results are considered good and prove the usefulness of the CBR based system for diagnosing Tourette syndrome.

REFERENCES

Anderson, P., (May, 1996). Obsessive Compulsion and Tic Linked to Sore Throats. Medical The Post. http://www.mentalhealth.com/mag1/fr51.html.

Bados A. (1995) "The tics and their upheavals: Nature and treatment in the childhood and adolescence". Madrid: Editions Pyramid S. A.

Bichindaritz, I., (2003). Solving safety implications in a case based decision-support system in medicine. In *Workshop on* CBR *in the Health Sciences*, 9–18. ICCBR'03.

Burden, G. (July, 1996) Imperial The Gene. Medical The Post. http://www.mentalhealth.com/mag1.

Capponi, R. (1987) Psicopatología and Psychiatric Semiología Santiago: University Editorial.

Fisher, MR; Shaver, S; Grasel, C; Bachering, T; Handl, H; Gartner, R; Scherbau, W; Scriba, PC., (1996) "Casus-Model Trial: To Computer-Assisted Author for System Problem Oriented Learning in Medicine". Z-Artiz-Forbild-Jena, Aug 90[5]:385-9.

Kolodner, JL., (1997) "Educational implications of analogy: to view from marries-based reasoning, Am-Psychol, Jan 32[1]:57-66.

Lopez Ostio, J. Cols (1993), Tutorial Systems (ITS). Typed conference. San Sebastián Spain.

Mann, BD; Sarechdern, AK; Nieman, LZ et. Al., (1996) "Medical Teacher to students by role playing:a model integrating psychosocial issues with disease management". JI. J-Cancer-Educ, Summer 11[2]:65-72.

Mink. J. W. & Weinberger, D. R., (1998). http://www.mentalhealth.com./fr20.html

Nieto M. ; Ochoa A., (2002) "Applying dependences model to CBR software". CIIC-'02; Soto La Marina, México.

Ochoa A. ; Fernandini M. & Shingareva I., (2005) "Use of Oniric techniques for describing Pa'nar Syndrome". Central Asia CCBR; Astana, Kazakhstán.

Pearce, J. (1996) "Good Habits and Bad Habits: Of the life in family to the life in society". Madrid: Editions Paidós

Polare I.; Kaneshiro, Takeshi & Malashona, N. (2005) "Modelling human societies using CBR". Central Asia CCBR; Astana, Kazakhstán.

Scotti, J. R., Schulman, D. And & Hojnacki, R. M. (1994) Functional Analysis and Unsuccessful treatment of Tourette Syndrome in a Man With Mental Profound Retardation Behaviour Therapy. 25, 721-738.

Sleeman, D.; Brown, J.S., (1982). Intelligent Tutoring Systems. AcademicPressLondon.

Vera, M. N. & Vila, J. (1995). Techniques of Relaxation. In V. Caballo (Eds), Manual of therapy techniques: modification of the conduct. (pp. 161-181). Madrid: Veintiuno century of Spain publishing.

Wagaman, J. R., Miltenberger, R. G. & Williams, D. And, (1995). Treatment of to vocal tic by Differential Reinforcement. J. Behav. Ther. & Psychiat. 26 (1), 35-39.

A Survey of Recent Research to support Remote Neonatal Intensive Care via Mobile Devices

Carolyn McGregor, Brent Morris
Health Informatics Research,
School of Computing and Maths
University of Western Sydney
Email: c.mcgregor@uws.edu.au

Abstract
Premature and ill term babies born in metropolitan and regional Australia are monitored and supported by a range of medical devices within Neonatal Intensive Care Units (NICUs) or Special Care Nurseries. Information produced by these devices is not available on mobile devices and require that a Neonatologist be in a hospital in order to view this data. This paper presents a survey of recent research to support streaming medical data from a NICU to a mobile computing device for remote intensive care in near real-time utilizing standard residential-grade broadband infrastructure. A key benefit of this research is that the functional requirements of such a system are used to drive the assessment of recent research as it relates to these requirements.

Keywords:

Web services, XML, data streams, compression, neonatal intensive care unit

INTRODUCTION

Approximately eighteen percent (18%) of babies born in New South Wales (NSW), Australia require special care or neonatal intensive care admission (Midwives, 1994). Premature babies can be up to 17 weeks early and may only weigh 450gms; they can spend 3 or 4 months in intensive care and have dozens of specific diseases before discharge. In addition, fifteen percent of neonatal intensive care admissions are transferred after delivery from smaller regional hospitals without intensive care facilities. Similar geographical conditions apply within other countries such as New Zealand, Canada and the USA where small regional hospitals are spread throughout the country (McGregor, Heath et al., 2005).

Several medical monitoring devices monitor a baby's physiological parameters such as blood oxygen, blood pressure and heart rate. Other devices such as ventilators offer mechanical life support. All these devices have the ability to output the device readings via a serial port, however the data formats vary greatly from device to device. As such, efforts to make data from devices attached to local or remote babies accessible for viewing through centralized neonatal intensive care unit (NICU) consultant physician viewing stations are hampered by the myriad of formats required for transmission. Regional hospitals have equipment to provide limited NICU support within Special Care Nurseries, but without the ability for a Neonatologist to receive information from this equipment, the baby must be moved to a Referral Hospital with Neonatologist support. Given the critical requirement to maintain a consistent environment, moving a baby at this time can be life threatening. Transferred critically ill, term and pre-term babies have higher mortality rates and higher rates of long term disability than similar babies born in hospitals with intensive care facilities. A major limitation is that the attending physician at the Regional Hospital must contact a neonatal specialist via telephone, who may or may not be located at the NICU at that time, to describe the baby's symptoms and, where possible, relay any physiological information verbally. The consulting Neonatologist must then make decisions based on this verbal exchange (McGregor, Kneale et al., 2005).

It is very common for critically ill babies to have significantly abnormal variation in the measured parameters minute by minute and not all these variations are made available to the consulting Neonatologist. Frequent transient falls in blood pressure and blood oxygen content, often with swings into the high range, may be of critical importance in survival and quality of life free of significant disability (Lister, 2000).

Hence the Neonatologists located at Referral Hospitals require the ability to obtain information from the monitors attached to babies. Similarly, a Neonatologist need not be located at a PC within the hospital to view local or remote patient data, but should be free to view this information through any secure Internet/Intranet connected device.

The Bush Babies on Broadband research project (McGregor, Kneale et al., 2005), aims to address this need by significantly improving the quality of treatment for babies born in regional and remote areas. The project uses broadband to transmit real-time data collected from bedside medical monitors, audiovisual streams of the patient and static physiological data such as x-ray images to the consulting Neonatologist in order for the Neonatologist to gain a better picture of the patient's condition than is currently available.

A key benefit of the framework developed as part of that research, is that it is available to link remote hospitals with the supporting NICU Neonatologist 'on demand' eliminating the need to establish permanent point to point connections. While the clinical trials have commenced for that research project, the Neonatologist needs to be located at a PC to utilise the environment.
As a result, there is a need to extend the Bush Babies on Broadband research to incorporate the use of personal digital assistants (PDAs) by Neonatologists in addition to PC devices to receive the information relating to the patient's condition.
This paper presents a survey of recent research to support streaming medical data to a mobile computing device for remote intensive care in near real-time utilizing standard residential-grade broadband infrastructure. Throughout this paper the term *mobile device* will to be used to refer to a portable computing device with size dimensions similar to a PDA allowing for it to be easily portable. The term *user* will be used to refer to a consulting Neonatologist that makes use of the mobile device from remote monitoring of medical data.
We begin with the functional requirements to support remote NICU monitoring in order to adequately explain the proposed system functions. A short description of the Bush Babies Broadband and e-Baby Architectures then provide the context for the research problem of streaming physiological data to mobile devices. Current uses of mobile devices in Telehealth are then detailed. Issues relating to mobile data reception are then described. The importance of synchronization between the mobile device and the remote care facility will then be addressed. Wireless interference issues are then presented. The notion of near real-time is addressed. Human-computer interface issues are then presented. The paper concludes with an explanation of future work to be completed in relation to this research.

REMOTE NICU MONITORING FUNCTIONAL REQUIREMENTS

Any remote NICU monitoring solution must be a robust system to be implemented for medical care. In order to define this robustness the system must comply with a series of functional requirements. Dunn distinguishes four quality criteria for software: reliability, usability, maintainability and adaptability (Dunn, 1990). Kazman et al. use a different schema to classify the qualitative properties of software (Kazman et al., 1994):

- Quality system output. This includes correctness, security, reliability and availability.
- Quality aspects for system developers and administrators, such as maintainability, portability, adaptability and scalability.
- Quality system use, for example ease of use, predictability and learnability.

The Kazman schema will be used to classify the components of the system as it provides for easy categorization of the proposed system functions. The proposed system will be required to have quality system output. As the system will be used for the purposes of remote intensive medical care, correctness, security, reliability and availability are of dire importance. A failure within any of these components could result in the injury or death of a patient. The quality of the data sent to the mobile device is important as the consulting Neonatologist will need to make life saving decisions based on the data presented on the screen.
The data streams proposed for this system contain many parts. The web services based framework created to support the e-Baby Architecture (McGregor, Heath et al., 2005) allows medical data streams to be sent to a remote receiver. Data is captured from multiple medical devices and sent to a data collection unit (DCU) for data warehousing and further analysis on a consulting physician station (CPS). A CPS is a PC that allows a care provider to monitor a patient from a remote location over a residential-grade internet connection, such as an office or home. Attached to the DCU is a video camera to capture visual data of the patient and a microphone to capture audio data. Within the proposed system the CPS functions will be facilitated by a mobile device.
The medical waveform data in the e-Baby Architecture, that is the motivation for this research, is transmitted in an XML format between the web services that connect the components of the framework. This facilitates easy data sharing through web services. XML data is a verbose format compared with equivalent data sent in a binary format, research has been completed to test under what conditions transmission speeds can be increased through the use of XML-aware compression methods (McGregor, Purdy et al., 2005).
Based on the data stream information as detailed above, the proposed mobile device architecture would require the capability to accept a data stream composed of many kinds of data. The most important kind of data in a neonatal intensive care context is video stream data. Much information can be taken from a patient through visual cues, making a frequently updated or moving image a key piece of data to the proposed architecture. The data stream must also incorporate other kinds of data such as frequently updated medical waveforms, medical imaging, and static textual data pertaining to the patient.

The contents and size of the data stream is dependant on two key factors, one being the mobile device user's requirements in order to consult with the remote hospital adequately and secondly the bandwidth capacity of the network the device is connected to at that time.

For these data streams to be considered of a high quality, the packets must be sent via a correct, secure, reliable and available mechanism.

The mobile device will require the ability to function in a range of connection types and speeds. From a low bandwidth situation such as a residential-grade cellular network to a high bandwidth situation such as an 802.11b/g or a Bluetooth network. It has been suggested that some wireless networks cause interference with intensive care monitoring devices (McGregor, Purdy et al., 2005; Simmons et al., 2002). As a result the proposed mobile device architecture must address this issue.

The hardware design of the mobile device must allow for it to be in operation for at least 24 hours of working time (Ammenworth et al., 2000), anything less than this would lower the quality of reliability for the device.

If the proposed system is to be successful it must not only give quality outputs but the developers and administrators that work with it must be able maintain it easily. The mobile device receives its data in an XML format from web services located at the referral hospital. In this way the proposed system can fit into any pre-existing information system without undue difficulty for the administrators and developers.

In order to comply with the Kazman schema the proposed system must be predictable, easy to use, and learnable. The mobile device must operate in such a way that the consulting physician using it will not be caught off guard by any unexpected events. The device must be capable of operating in multiple network architectures. The device must handle switching between different connection types as seamlessly as possible. The proposed architecture needs to be robust enough to allow for the mobile user to move from a zone where cellular transmissions are permitted to a zone where they are not permitted, such as a hospital. The users of the mobile devices in the proposed architecture are medical specialists, and thus need to be reachable by a remote hospital even if they are in another hospital.

The user interface (UI) of a mobile device is important as the device must display the visual representations of the multiple data streams. The UI needs to be unobtrusive and simple to allow the consulting physician to focus on the care of the patient and not the operation of the device (Lu et al., 2005), however the UI needs to have within it the capability to handle complex operations such as managing the data streams being sent to the device.

For the proposed system use to be of a high quality the design of the mobile device and the remote monitoring software must be easy-to-use, predictable, and learnable.

BUSH BABIES BROADBAND AND E-BABY ARCHITECTURES

The research problem motivating this paper is the use of mobile devices to provide functionality similar to that of the Consulting Physician Station (CPS) within the Bush Babies on Broadband architecture and the overarching e-Baby architectures. The Bush Babies on Broadband architecture is shown in Figure 1 and is designed to provide remote Neonatologist support for the care of premature and ill term babies born in regional and remote Australia prior to or instead of their transfer to the Neonatologist's NICU.

The Bush Babies on Broadband project is designed to run on off-the-shelf PC technology, the project currently consists of four main components (McGregor, Kneale et al., 2005):

1. **The Bush Babies Remote Station (BRS)** resides within the maternity ward of a regional or remote hospital. The BRS is, in fact, any personal computer with Web browser software and access to the Internet. The BRS is used to initiate a Bush Babies on Broadband session and is used by the remote hospital care provider. The BRS allows access to the centralised Referral Database of available consultants.
2. **The Bush Babies Control Centre (BCC)** provides a centralised Referral Database of available consulting Neonatologists. A consulting Neonatologists can be selected, by using criteria such as specialty, humidicrib availability, location and time.
3. **The Consulting Physician Station (CPS**) is any personal computer running a Microsoft Windows operating system and with access to the Internet. Consulting Neonatologists use the CPS to display static and real-time physiological data as well as video images of the patient.
4. **The Data Collection Unit (DCU)** is situated in close proximity to the patient bed or cot. The DCU is a small-footprint PC running specialised software and is connected to bedside medical monitors, a video camera and image input device. The DCU is the information hub for the patient and packages the incoming data into multiple RTP streams for transmission across the Internet.

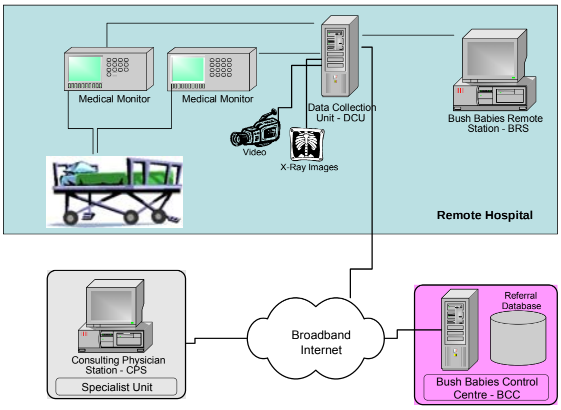


Figure 8 Bush Babies Broadband (McGregor, Kneale et al., 2005)

The main difference between the CPS and the proposed mobile devices architecture is that the PC component monitoring functions are instead received by a mobile device. This will remove the physical location constraints from the consulting physician, allowing them to be reachable if they are not in a position to access their personal computer easily.

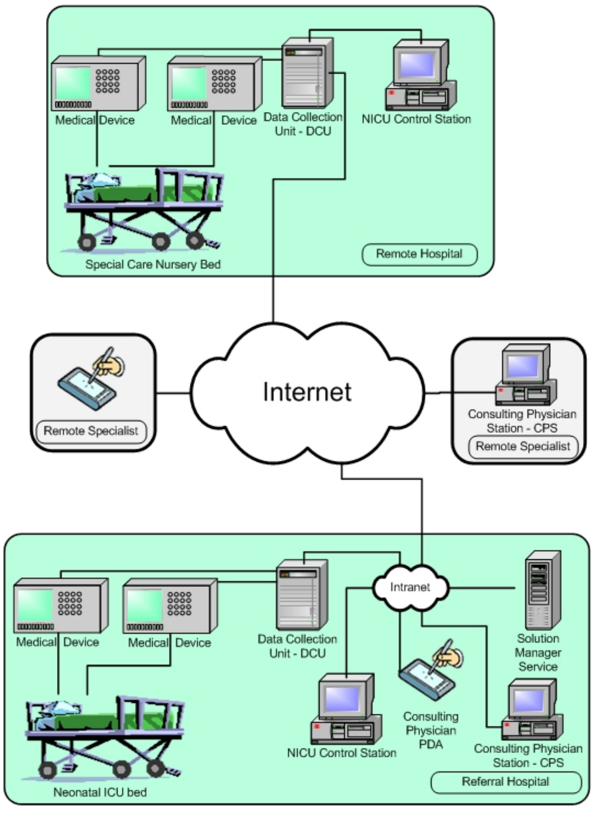


Figure 2: The e-Baby Architecture (McGregor, Heath et al., 2005)

In association with the Bush Babies project McGregor et al (McGregor, Heath et al., 2005) propose a web service based framework for the transmission of XML encoded physiological data output by medical monitoring and life support devices. That research, together with the previously mentioned Bush Babies are portions of the "e-Baby" research collaboration, that is researching new approaches to the application of computing and information technology to support mobility in healthcare through local and remote neonatal intensive care. The high level e-Baby architecture is shown in Figure 2.

The web service based framework for the transmission of XML encoded physiological data forms part of the Solution Manager Service (SMS). The SMS is situated in the referral hospital NICU and receives and stores data collected by the

Data Collection Units (DCUs) via the *physiological log web service* for near real time analysis and trend detection. The Consulting Physician Station (CPS) is used by the Neonatologists to access the physiological data located in SMS via a set of *analyse web services* or via a direct link to the data as it is streamed through the *physiological log web service.* That research enables patient and care provider mobility, as a result of the web serviced based data transmission. However, care provider access is only currently available through the PC based CPS. The architecture proposed in this paper will enable forwarding of data to consulting physician PDAs and other mobile devices as shown in Figure 2.

CURRENT USES OF MOBILE DEVICES IN TELEHEALTH

Telehealth is the transmission of images, voice and data via telecommunication channels to support medical care. Telemedicine connects patients and health care providers improving access to quality health care particularly in rural and remote parts of NSW. Telehealth and telemedicine initiatives to provide mainstream public health care delivery across rural and metropolitan areas, have mainly focused on video conferencing (McGregor, Heath et al., 2005).

Telemedicine systems have been used in the past to aid in disaster relief situations. In these situations audio, video and medical imaging outputs were exchanged with a remote hospital (Garshnek, 1999). The data transfer networks used were either space agency or military-owned satellite networks or dedicated high-speed land lines. While these networks provide a very high speed data transfer speed, they are inflexible and expensive. Dedicated high-speed land lines require a large investment and require stationary points of transmission and reception. The satellite networks provide mobility and high speed data transfer but require bulky equipment that is not suited to size limitations required by a PDA. A terrorism response exercise conducted in the United States in 2002 tested the effectiveness of transmission of medical data streams, video streams from a remote station to a hospital over satellites and the Internet via Internet2 (Simmons et al., 2002). This system was designed for medical teams consisting of multiple experts providing assistance to an area affected by terrorism. This system has portability limitations. For example, acquiring and maintaining a satellite connection required a 35lbs dish as well as numerous extra devices to create a wireless local area network (WLAN) to handle reception of data from the satellite. Another drawback to using the system easily was the choice of Internet2 to communicate over the Internet, a network not yet commercially available.

Most of the recent computing and IT research to support intensive care units (ICUs) has focused on clinical alerts (McGregor, 2006; McGregor, Heath et al., 2005). The information made available to these systems is limited to a small set of physiological data and/or clinical data from patient location within their ICUs. In addition, care provider access to these systems is limited to the receipt of alerts with minimal content via email and in some cases pagers. These alerts occurred after pre-defined conditions were met within a certain timeframe. The messages were terse and provided little more data than a couple of lines of text (Ammenworth et al., 2000).

Remote medical monitoring systems have been proposed that use sensors located on the body that connect to a PDA (Shih et al., 2004). The PDA is used to translate raw data into a streaming format that can be then sent over cellular networks to a remote computer monitored by a care provider via the Internet. This data could then be used for display and analysis. In this case there was a limited amount of data that could be captured as the system was designed for a patient to remain monitored while still being active. Only thermographic and electrocardiogram (ECG) data was captured and transmitted.

MOBILE DATA RECEPTION ISSUES

The system described in the Bush Babies Broadband Architecture is designed to send medical and video stream data over a residential-grade high-speed land line network such as Asymmetric Digital Subscriber Line (ADSL) (McGregor, Kneale et al., 2005). The current system sends approximately 100kbps of real-time physiological data and approximately 500kbps of video stream data. This system also supports the transmission of static physiological data files sporadically at approximately 300kbps. This bandwidth requirement of 600kbps with occasional bursts up to 900kbps is easily transmitted over a high bandwidth land line, but exceeds the bandwidth capabilities of most of the currently available residential cellular networks. Currently it appears that only third generation cellular (3G) networks are capable of fulfilling the high bandwidth requirements of the proposed system as it stands (Chandran & Valenti, 2001).

Bandwidth requirements of the proposed system may be reduced if investigations into the minimum required resolution for video streams and medical data streams are conducted to determine the minimum level of detail required in order for the user to confidently make care decisions.

In addition, there is a requirement for voice communication between a remote hospital and the mobile device user. Once the bandwidth requirement for the data streams has been addressed, it can be established if there is available bandwidth to add a voice data into the same stream as the data stream or if this must be facilitated by a separate connection.

SYNCHRONIZATION ISSUES

It is crucial for a mobile data receiver to remain in contact with its transmitter in order for it to receive new data and thus provide the user with recent enough data to make informed care decisions.

Synchronization in the case of the proposed architecture must be in near real-time and not through the process of hot syncing. The problem with hot syncing is that PC and PDA data only match immediately after hot syncing and hot syncing large volumes of data causes hot syncing to fail intermittently resulting in incomplete or duplicated information. Hot syncing allows only for periodic updates up to several times in an hour. The proposed architecture requires synchronization to occur several times in a minute.

As stated previously, the bandwidth requirement for the proposed system needs to be confirmed. Once this has been completed an investigation what wireless networks can keep up with the data rate must be performed.

There has been some research conducted on the ability of a mobile device to maintain a connection while switching between a cellular networking environment and a wireless local networking environment. The concept is known as interoperability and is currently feasible. A framework proposed by Joseph et al (Joseph et al., 2004) allows for interoperability between CDMA2000 cellular and WiFi local area networks. The framework addresses the situation of a mobile device existing in multiple networks by introducing the concept of user profiles. The profile relevant to the proposed system is "Class 3 – Glitch Conscious" profile, a device using a Class 3 profile will not switch from a low bandwidth network to a high bandwidth network unless specifically instructed to do so by the user. Using this profile will enable the user to maintain the connection with a referring hospital, until the user decides to switch. The user may decide that moving to a higher bandwidth network is worth the momentary pause in data transfer, or they may decide to remain in a low bandwidth network if the situation requires constant connectivity at the expense of less information.

WIRELESS INTERFERENCE

The mobile device user in the proposed architecture needs to be capable of connecting with a remote hospital even if they themselves are in a hospital. Recent research indicates that in some cases cellular and handheld phones interfere with pacemakers and intensive care monitoring devices (Simmons et al., 2002). One study found that mobile telephones on the Global System for Mobile communications (GSM) do not interfere with medical equipment when a safety margin of 2m is guarded (Ammenworth et al., 2000).

Further research needs to be completed in this area due to the proliferation of 802.11b/g and Bluetooth devices in recent years. These devices operate on the 2.4GHz frequency band and not the 26-1000MHz frequency range recommended for use by medical devices as outlined in the 1993 International Electrotechnical Commission (IEC) Standard IEC 60601-1-2. This is the prevailing international standard for radio frequency immunity of medical devices (Lin, 2000).

Amendments to the United States' Federal Communication Commission's (FCC) rules were released in 2000 to establish the Wireless Medical Telemetry Service (WMTS). The FCC reports that the most wireless medical telemetry devices currently operate on the 450-470MHz or on vacant TV channels in the range of channel 7 to channel 46 (Commission, 2000), it was reported that some medical devices conflicted with the Private Land Mobile Radio Service (PLMRS) in use by digital television broadcasting which operated on the 450-460MHz. The WMTS exists to help mediate conflicts over this frequency range.

Despite the research and standards recently made, research still needs to be completed to establish if the age of a medical device is factor. Some care facilities may be using equipment created before the formation of the WMTS and it needs to be established if these older devices are susceptible to interference from recent wireless networking developments.

REAL-TIME TRANSMISSION ISSUES

This section discusses the nature of network latency and how it may affect the proposed architecture. Latency is an inescapable component of any remote monitoring system. The only way to monitor a patient in real-time is to be in the same room as them and personally monitor the output of the medical devices connected to the patient.

Research needs to be completed to establish the minimum frequency of updates of medical data streams to give meaningful data to the mobile device user. Available bandwidth is the primary constraint in establishing the maximum frequency of updates capable. Recent research found that the processing power of a PDA device was not a constraint in the transmission of medical data (Shih et al., 2004). The research found that the PDA's processor was capable of transmitting up to 360 samples of ECG data per second, but sending data with this frequency was not reasonable. Sending data at such a frequency would saturate the data channel and cause packet loss. The research project resolved this problem by using high-rate sampling and low-rate data transmissions, sending a maximum of 20 seconds of ECG data every minute. While the mobile device in the proposed system does not send medical data to the remote hospital, the same issues can be expected in the receiving of medical data from the remote hospital.

As discussed earlier in the paper, compression of the medical data stream has been testing to support the e-Baby *physiological log web service*. The processing abilities of mobile devices needs to be investigated to see if they are capable of decompressing received medical data streams fast enough to improve any of the latency problems resulting from using the device in a low bandwidth networking environment.

USER INTERFACE ISSUES

The user interface (UI) of a mobile device is important as the device must display the visual representations of the multiple data streams. Display of this data is affected by the small size of the mobile device screen compared with the size of a normal medical monitoring device or a desktop computer (Ammenworth et al., 2000; Lu et al., 2005). The UI needs to be unobtrusive and simple to allow the device user to focus on the care of the patient and not the operation of the device (Lu et al., 2005). Ammenwerth et al (Ammenworth et al., 2000) suggest that due to the restricted screen size of a mobile device, "mobile information access should be limited to only relevant data", we agree with this suggestion. The proposed system must have in place methods for the user to select what data is to be sent to the device, these methods must be non distracting and easy-to-use and not overwhelm the user with complexity.

One way to mitigate some of these issues is through context-sensitive applications. Context-aware applications are applications that implicitly take their context of use into account by adapting to changes in a user's activities and environments (Anind et al., 2004). The mobile device user in the proposed system will operate in multiple environments. These environments will be defined by the physical location of the mobile device user, the wireless network capabilities in that location, and the current operation the device is performing. The basic UI provided with a PDA for example is adequate for providing basic PDA functions, but may prove to be inefficient for remote monitoring of a patient. When the device is placed into a context where remote monitoring of patient medical data is required, the UI should be altered allow the care provider to focus on the task of monitoring the patient.

FUTURE WORK

Further issues impacting the provision of remote NICU monitoring include: the hardware constraints of mobile devices, security over wireless networks, trust issues regarding the reliability of the proposed system, and further details regarding the minimum level of detail and frequency requirements for video and medical data streams received by the mobile device. These issues will be examined in detail in future research.

CONCLUSION

This paper presented a survey of recent research to support streaming medical data from a NICU to a mobile computing device for remote intensive care in near real-time utilizing standard residential-grade broadband infrastructure.

It demonstrates that there has been little research completed in remote ICU monitoring, or more specifically remote NICU monitoring. However, research in similar niches has been conducted. This research indicates that there are currently many constraints that may make implementation of the proposed system difficult for the near future. There needs to be further research into the bandwidth capabilities of wireless networks to determine if there exists a cellular network that can handle the bandwidth needs of the proposed system. The mobile device user must have software robust enough to mitigate travel between different network types while maintaining a connection and the remote hospital and receiving enough data to make informed care decisions. Research must also be conducted in the realm of wireless local area networks to determine if the hardware developments in recent years cause interference with medical monitoring devices. Hardware and software abilities of mobile devices needs to be assessed to determine their ability to process compressed XML data in order to determine if they can reduce network latency by receiving compressed data streams.

This research supports our future research investigations to develop an architecture to support mobile remote NICU monitoring. The development of that architecture will be based on extensions to previous e-Baby and Bush Babies research (McGregor, Heath et al., 2005; McGregor, Kneale et al., 2005; McGregor, Purdy et al., 2005).

REFERENCES

Ammenworth, E., Buchauer, A., Bludau, B., & Haux, R. (2000). Mobile information and communication tools in the hospital. *International Journal of Medical Informatics, 57*(1), 21-40.

Anind, K., Dey, R. H., Beckmann, C., Li, I., & Hsu, D. (2004). *a CAPpella: Programming by Demonstration of Context-Aware Applications.* Paper presented at the CHI 2004, Vienna, Austria.

Chandran, N., & Valenti, M. C. (2001). Three Generations of Cellular Wireless Systems. *IEEE Potentials, 20*(1), 32-35.

Commission, F. C. (2000). *Amendment of Parts 2 and 95 of the Commission's Rule to Create a Wireless Medical Telemetry Service*.

Dunn, R. H. (1990). *Software quality : concepts and plans*. Englewood Cliffs, NJ: Prentice Hall.

Garshnek, V. (1999). Applications of Telemedicine and Telecommunications to Disaster Medicine: Historical and Future Perspectives. *Journal of the American Medical Informatics Association, 6*(1), 12.

Joseph, D. A., Manoj, B. S., & Murthy, C. S. R. (2004). *Interoperability of Wi-Fi Hotspots and Cellular Networks.* Paper presented at the The Second ACM International Workshop on Wireless Mobile Applications and Servcices on WLAN Hotspots (WMASH '04), Philadelphia, PA, USA.

Kazman, R., Bass, L., Abowd, G., & Webb, M. (1994). *SAAM: A Method for Analyzing the Properties of Software Archietectures.* Paper presented at the International Conference on Software Engineering - ICSE'16, Sorrento, Italy.

Lin, J. C. (2000). Health Aspects of Wireless Communication: A real and Present Wireless Danger. *ACM SIGMOBILE Mobile Computing and Communciations Review, 4*(1), 17-18.

Lister, R., Bryan, G. and Tracy, M.,. (2000). *The e-Babies Project:Integrated Data Monitoring and Decision Making in Neonatal Intensive Care.* Paper presented at the European Conference of Information Systems (ECIS 2000).

Lu, Y. C., Xiao, Y., Sears, A., & Jacko, J. A. (2005). A Review and a Framework of Handheld Computer Adoption in Healthcare. *International Journal of Medical Informatics, 74*(5), 409-422.

McGregor, C. (2006). Mobility in Healthcare for Remote Intensive Care Unit Clinical Management. In B. Unhelkar (Ed.), *Handbook of Research in Mobile Business: Technical, Methodological and Social Perspective*. Hershey, PA, UWA: IDEA Group Publishing.

McGregor, C., Heath, J., & Wei, M. (2005). *A Web Service Based Framework for the Transmission of Physiological Data for Local and Remote Neonatal Intensive Care.* Paper presented at the IEEE International Conference on e-Technology, e-Commerce and e-Service, Hong Kong.

McGregor, C., Kneale, B., & Tracy, M. (2005). *Bush Babies Broadband: On-Demand Virtual Neonatal Intensive Care Unit Support for Regional Australia.* Paper presented at the 3rd International Conference on Information Technology and Applications (ICITA), Sydney, Australia.

McGregor, C., Purdy, M., & Kneale, B. (2005). *Compression of XML Physiological Data Streams to Support Neonatal Intensive Care Unit Web Services.* Paper presented at the IEEE International Conference on e-Technology, e-Commerce and e-Service, Hong Kong.

Midwives. (1994). NSW Midwives Data Collection.

Shih, E., V.B., Curtis, D., & Guttag, J. (2004). *Demo Abstract: Continuous Medical Monitoring Using Wireless Microsensors.* Paper presented at the SenSys'04.

Simmons, S. C., T.A.M., Blanarovich, A., Workman, F. T., Rosenthal, D. A., & Carbone, M. (2002). Telehealth Technologies and Applications for Terrorism Response: A Report of the 2002 Coastal North Carolina Domestic Preparedness Training Exercise. *Journal of the American Medical Informatics Association, 10*(2).

An Introduction to Telecommunication and Telemedicine

Sandra S Lazarus
University of Sydney, Australia
Sandra@cit.uws.edu.au

Abstract

The incorporation of telecommunication within a medical environment can be defined as 'telemedicine'. At the core of telemedicine is the provision of healthcare at a distance with the use of technology. This paper outlines current developments in telecommunication and telemedicine by introducing the applications of such technology in health settings. Furthermore, by addressing issues facing telemedicine and looking at the potential advances of how telecommunication technology can provide enhanced communication between several medical / health institutions. The paper will also review the implementation of mobile application, telemedicine standardisation, social development, economical developments and medical education in a medical / health environment.

Keywords:

Telemedicine, Telecommunication, Telehealth, Mobile Telemedicine, HL7.

INTRODUCTION

While there are many issues surrounding the research, development, and facilitation of telemedicine and telecommunication within health care infrastructures, it can be seen that the communicative dynamics of telecommunications technology and the implementation of telemedicine play a pivotal part in the success of integrating technology in modern medicine. This can be seen in the communication between two or more health facilities, the transfer of medical data, and the development of standards to further harness medical information and improve communication. By the development of telemedicine standardisation the application of any technology can available more extensively. This standard is being provided through the development of HL7 (Health Level 7), which is currently being used to communicate between different healthcare organisations. Telemedicine is also impacting the social and economical areas are being affected as the both telecommunication and telemedicine grow. Medicine is a particularly rich application area for web technology, recent congresses devoted to Internet in Medicine report many very interesting experiences [Kanoui H, al, 2000].

TELEMEDICINE AND TELECOMMUNICATION

The term telemedicine derives from the introduction of telecommunication, e-commerce and e-health in medicine, and has developed into a more specialised area of technology. One of the earliest methods of telemedicine through the use of telecommunication was via telephone networks. To this day it still remains a comprehensive and effective way of commutating medical data to and from rural areas. Today, telemedicine allows health care practitioners to deliver specialised care to locations with provider shortages, through visual and audio assessment using video conferencing equipment or videophones [Palsbo, S, 2004]. It also enables medical information and various data to be transmitted over long distances using DSL technology, including ECG readings, electronic stethoscope findings and a host of pathology results [Denes, S, 2003].

The transfer of medical information through telecommunication can be synchronous (real-time), or asynchronous (non-real time). Synchronous flow of information is preferred by medical institutions as it provides updated information as it occurs, which in hospital setting can be urgent. Synchronous information is also a more realistic option when it comes to multi-directional communication, where information is continually being updated by numerous users. Video-based telemedicine in particular, allow audio and visual information to be transmitted through telephone networks or digital connections from one physician to another [Brennan A, et al, 1999].

The asynchronous flow of information is uploaded into the system and then downloaded by the user which results in a loss of time. In this case information is not immediate; it is stored and then forwarded accordingly. Although this is not the preferred method of communication it is still broadly used by medical practitioners.

Health care providers are able to utilise the above communicative structures of telemedicine, to treat and diagnose patients in the emergency department where medical information is urgently required on a constant basis. By transmitting patient information via Wireless communication, practitioners in the emergency departments are able to gain assistance from specialist physicians who are located elsewhere. For instance, in rural areas where specialised health care expertise are limited, this form of communication can provide a network of imperative information for transfer between central health station to rural station.

Figure 1: The diagram illustrates the transfer of a patients' medical information to and from rural areas to central areas (e.g. metropolitan hospitals) for diagnosis and evaluation.

Multi-directional telemedicine can provide an exceptionally cost-effective method for communication between several medical organisations. The practicality of this type of communication means that colleagues are able to collate a vast array of information on patient diagnosis and treatment. In this way, telemedicine has enhanced the relationship between multi-professional teams by actively promoting a shared learning environment for the mutual benefit of the patient [Newton H, 2003]. Evidently, the idea of a shared learning environment signifies that medical professionals have a greater chance to exchange and share opinions and information, through the utilisation of telecommunication technology.

RECENT DEVELOPMENTS

Telemedicine has the ability to remove geographical limitations and can also remove several over-head costs involved in specialised health care. Previously, there were only two main types of networks: Local Area Network (LAN) and Wide Area Network (WAN), but now other communication protocol is gradually replacing these traditional networks, and this essentially because digital communication has greater speed and reduced error rates[Pande R, et al, 2003].

In using advanced digital imaging technology, users can not only send and receive written information but are now also able to view digitalised medical images such as CT scans, X-rays and Ultrasounds. This transfer of medical information is carried out with the use of integrated service digital networks (ISDN) and broadband ISDNs (B-ISDNs), which is chiefly based on digital communication [Pande R, et al, 2003].

Recent capabilities in Broadband technology have allowed for a great expediency in the delivery of medial information within metropolitan and rural communities. In fact, the 2000 Telecommunications Market Review and Forecast, cable and DSL services are expected to average a 24.2 percent growth annually through 2005. It is also expected that in 2005, spending will total an estimated $13.8 billion [Flanigan M, 2002].

The potential of telecommunications within health care is such that medical information is readily made available though electronic transfer. This availability of medical information has had a notable positive social impact, in that it provides for some equality between rural and metropolitan areas for health administration. While there is still some research being conducted to evaluate the suitability of networks, the incorporation of telecommunications in the medical environment indicates that mobile telecommunication provides solutions to difficulties that exist within a range of traditional communication systems.

This incorporation of telecommunication, together with telemedicine and health has produced telehealth, which can be described as the amalgamation of all three areas. By allowing practitioners to communicate medical information over long distances, telemedicine has provided a new avenue for medical diagnosis and treatment. Telemedicine and telehealth has provided links and improvements in the way in which information is communicated. The chief purpose of both telemedicine and telehealth is to provide an improved, managed and structured communication system linking health care providers.

POTENTIAL FOR APPLICATON

Mobile Telemedicine application

The potential for the application of such technology can be seen in two focal areas of medicine. Firstly, there is the introduction of mobile devices and mobile technology in medical processes such as patient care, decision making and drug administration, then there is the translation and the standardisation in the communication of patient data across the medical locale including all medical test results, which is not just limited to administrative data. Also a large percentage of that new staff use PDA device for assessing medical information. Accesses to drug information databases on PDAs provide a fast and easy way to reference medication details including drug interactions [Lazarus S, 2006]. With the development of mobile devices which are specially customised for specific medical purposes. Today there are two basis operating systems for PDA devices, Palm OS and Window CE for Pocket PC. But, it is just recently due to the evolution of PDA devices into more powerful tools that it's become possible to take the next step and utilise the full capacity of the PDA devices in a hospital environment. Mobile telecommunication requires wireless technology, communication which makes for the mobility of information as it is transferred across the network. Currently there are two forms of wireless communication being considered and utilised in medical institutions, these wireless technologies are Bluetooth and WiFi. The information flow that can be transferred consist of singular or multi directional communication. Singular direction communication is viewed by the mobile device user, but cannot be altered within the system or be updated by the user. Multi directional communication allows mobile device users to enter information into the medical systems which then allow medical staff to update information as it occurs at the patient bedside. In the same way a wireless network of information transfer can provide mobile information to where and when it is required using mobile devices to display information form medical databases.

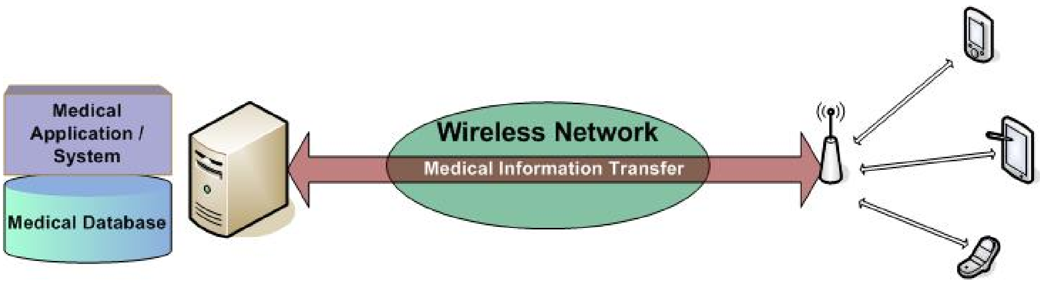


Figure 2: The diagram illustrates the Multi directional transfer of patient information using mobile devices allowing the mobile flow of information where and when it is required by the medical staff.

Recent trials conducted on a global scale reveal that the use of PDA devices makes a significant difference in the medical environment. Doctors seem to expect handheld computers to become increasingly useful, if not ubiquitous [McAlearney A, et al, 2004]. The use of PDA devices is becoming common practice in the Medical environment as it is proving to be a useful tool for physicians in their daily appointment scheduling. The overall use of hand-held devices is more common among physicians under the age of 45 with a 33% usage rate as opposed to a 12% usage rate amongst physicians above this age.

In-hospital studies with physicians utilizing PDA devices uploaded with medical programs for clinical assistant shows positive feedback. The data received following this trial showed that of all responders who used the PDA device as a clinical and diagnostic/decision-making tool, 30% used the PDA medical calculator, 78 % used the Medical Handbook, 82% used MIMS for drug referencing and 65% used Therapeutic Guideline. This in turn revealed that when given the opportunity, physicians and clinical staff are potentially given the chance to make full use of the information and data available on PDA devices [Lazarus S, 2006,].

Similar studies revealed evident indications that PDA devices are considered a useful intermediary tool for clinical decision making. Trainees reported more frequent PDA use in the hospital setting for direct patient care, while attending physicians reported more frequent PDA use in administrative settings and for calendar functions [McLead T, et al, 2003].

However, to allow PDA devices to be implemented in a clinical environment, the devices need to be fully integrated with medical systems. Therefore program developers need to consider the potential of programs being made available on PDA devices. Studies have shown that various computerized systems, such as reminder prompts and physician performance feedback have the potential improve physician performance and patient outcomes, however, if the operation of such systems is too time-consuming, physicians may not find them to be useful [Bodenheimer T & Grumbach K, 2003].

Telemedicine application Standardisation

For information to be universal, standards need to be developed to allow a variety of medical organisations and health facilities to share clinical data. HL7 (Health Level Seven) provides standards for interfacing and messaging protocol, it is developed from the application layers of ISO (Open System Interconnection) model. The ISO model is a networking of architecture for moving data from one software application to another application located on another computer [Hetenryck K, 2001]. HL7 is a standardised language which is internationally agreed upon to exchange medical information over a larger clinical demographic. The current purpose of the HL7 standards is to remove communication barriers between different medical environments by communicating with a number of systems for data-sharing. An example of this form of communication is the communication between other standard medical systems such as DICOM, CEN, EDIFACT and IEEE which are highly used in the communication patient information. Currently, New Zealand is operating a HL7 based National Patient Index. The health industry in Australia is similarly undergoing major changes, as until recently, most of the industry has relied on paper-based systems. Conversion to electronic communications offers significant improvements to the industry as a whole and provides gains in service efficiency and accuracy. This has prompted significant opportunities for IT and telecommunication providers. As illustrated in figure 3 HL7 sits between

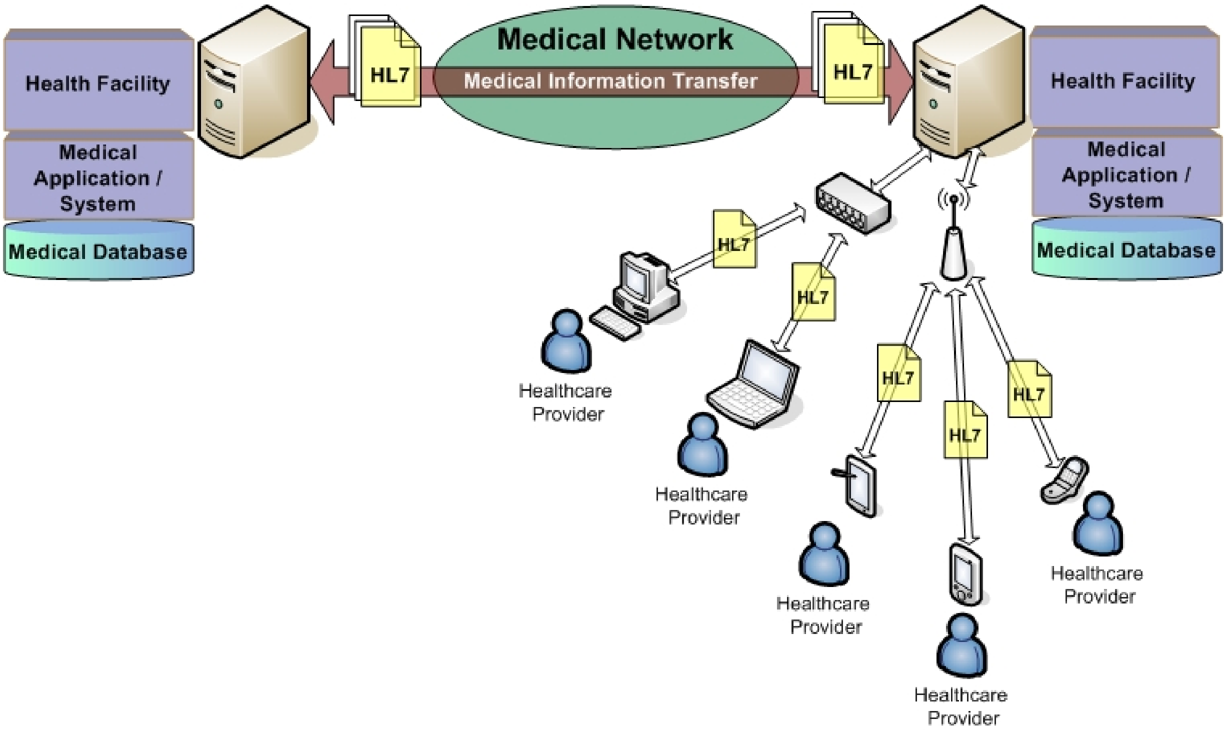


Figure 3: The diagram shows where in the transfer of medical information HL7 provides standardisation and makes it possible to share information.

Australia is currently moving gradually towards an integrated approach to sharing clinical information using HL7 and other medical based formats. In Australia some Pathology Labs and Radiology Clinics are sharing results and reports using HL7-files-based information for integration into local based clinical software. However other transactions such as admissions, discharge summaries, specialist reports, health referrals, and diagnostic images are not being communicated electronically. Few diagnostic companies utilise PACS (Picture Archiving Communication Systems) to share diagnostic images to consult radiologists. PACS systems enable the sharing of images such as X-rays, CT scans, Ultrasound and MRIs through high bandwidth telecommunication networks enabling radiologists to make diagnostic decisions in various locations away from the facilities that are the sites of such radiographs and sonagraphs. The reason for this is the high price tag on PACs system as they range from $250,000 to $500,000 upfront investment plus the requirement of high quality bandwidth networks [Lindsay A, 2005].

The above applications of telecommunication in medicine allow the medical information to flow using different methods of information transfer. Taking into consideration the number of health care providers and patients, this method of information transfer can provide integrated patient-centric clinical information that extends beyond the walls of any one health care setting or the boundaries of a provider's context of practice to support the process of clinical decision-making. The transfer of medical data and clinical information is necessary not only for the patient but for the diagnosing clinical staff and medical practitioners. A few trials have been conducted in regards to the communication of clinical information and patient referrals. However no significant progress has been made to move forward with development of systems for the purpose of electronic clinical information, communication and collaboration.

Medical Education

Due to the high availability and spread of the Internet, it is now being considered as an educational tool in medicine. The last several years have seen an increase in the number of medical sites. Many websites are being developed to educate patients about various medical conditions. These websites educate the patient in how to better manage their conduction, as well as how to monitor any changes in their conditions and act accordingly. This incorporation of the Internet with medicine can be related back to medical informatics and e-education in medicine. This will allow students to collaborate with other students, self assessment and peer evaluation, as well as give individuals time to reflect on individual progress.

Using the Internet as an educational tool allows updated information to travel further, and this provides instant information more readily to patients. In turn, patients can then access new, updated and relevant information about particular medical condition fast. By allowing patients and the general public to become more aware of diseases and their symptoms, individuals can react accordingly either by visiting their GP or Specialist. Some of the medical websites even have treatment methods on medical conditions. This enables patients to gain knowledge of the treatment they might be receiving, or outlines the treatment options they might have. This form of education is also being used to educate medical students by providing knowledge to the medical students about their subject matter. This once again removes distance barriers, so that, education can take place anywhere and at anytime. It also provides the student with the luxury of learning at a pace then they fell comfortable. Increasing better understanding of the material which can be referred to when required.

SOCIAL AND ECONOMICAL DEVELOPMENTS

One does not have to look very far to see the fundamental changes that are occurring in the health care industry, changes which are in turn forcing social and economic change. First and foremost, health care is becoming a more patient-driven industry. There is a palpable commitment to building a foundation that is grounded in the principles and values of patient and family focused care. Similarly, there is a demonstrated understanding of the need to shift the focus of health care efforts from the management of illness to the maintenance/promotion of wellness. As a result, we are seeing increased emphasis on the management of diseases across the continuum of care and along the lifecycle of the disease. To support this, the industry is experiencing a significant shift in how clinical decision-making occurs. Specifically, we are replacing the "lone ranger" decision-making practices of the past with truly collaborative, interdisciplinary, and evidence-based approaches. We are also seeing a move from generalised care to more specialised care [Lindsay A, 2005], as outlined in the table below. As health industry is change towards information management as broader spectrum of information is made available through improved communication.

Old Health Industry (without Telemedicine technology)	New Health Industry (with Telemedicine technology)
Health provider focused	Patient and Family Focused
Site of Care	Continuum of Care
Episode Management	Disease Management
Supply Management (supply on illness)	Demand Management (supply on demand)
Lone Decision Making (lone ranger)	Collaborative, Evidence-based decision
Efficiency	Effectiveness
Generalized Care	Specialized Care

Table 1: The table outlines the health industry with and without the implementation of Telemedicine technology. It is clear that the focus of the health industry changes after the introduction of telemedicine as management and communication is improved [Lindsay A, 2005].

While many of these changes are driven by advances in technology, they also require a capability from health information-structures – a capability that does not fully exist today. Though, with the proper implementation and application of technology and improvements in health information structures this can be achieved. Medical professionals require access to health information not only across different systems, but across different jurisdictions and technological boundaries. It's this capacity to view clinical information from all sources and to use the information-structure to initiate orders and referrals to a broader range of care and service providers which indicates the fluidity of information within health care structures. This happens by extending the capabilities of our integrated but tightly coupled information technologies to work within a national framework.

Like the broader economy, the health care sector is undergoing a boom in online services and communication technology, including a market for business-to-business e-commerce. Despite the somewhat slow take up of information technology in the General Practice environment, many believe the Internet health industry is set to explode in line with the high percentage of Internet use within the Australian community.

CONCLUSION

It is believed that the future of healthcare in Australia needs to progress towards a totally integrated platform for sharing clinical information. This will enable all health care providers and health facilities, in any location to share all patient information. The aim is that patients will be able to visit any health provider in Australia and have their clinical information and past history readily available, so as to ensure that the health provider has a holistic view and record of the patient's medical history.

Health care in Australia is, for many, the number one priority. Electronic health records (EHR) have been widely recognised as essential to modernising Australia's health care system as well as supporting the goals of health care providers by improving quality, accessibility and sustainability of medical information communication systems.

Telemedicine and telehealth has provided links and improvements in the way in which information is communicated. In particular, one of the most significant applications of telemedicine and telehealth in a medical environment is its implementation for an improved communication structure between different health care providers [Geisthoff U, et al, 2002].

REFERENCES

Bodenheimer T & Grumbach K, 2003, 'Electronic Techonology: A spark to Revitalise Primary Care?', JAMA, vol. 2, no. 290, pp. 259-164.

Brennan A, Kealy J, Gerardi L, Shih R, Allegra J, Sannipoli L & Lutz D, 1999, 'Telemedicine in the emergency department: a randomised trial', *Journal of Telemedicine and Telecare*, vol. 24, no. 3, pp. 149-164.

Denes S, 2003, 'Cap rock telephone open door to Telemedicine', *Rural Telecommunications*, vol. 22, no. 2, pp. 56-58.

Flanigan M, 2002, 'Long live Broadband', *Telecommunication Americas*, vol. 36, no. 8, pp. 16- 19.

Geisthoff U, Federspil p, Sittel C & Plinkert K, 2002, 'Telemedicine: Interaction between Clinic and General Practice', *HNO*, vol. 50, no. 9, pp. 812-821.

Hetenryck K, 2001, 'Health Level Seven: A standard for Health Information Systems', *Medical Computing Today.* www.medicalcomputingtoday.com view 03 / 10 / 2005.

Kanoui H, Joubert M & Maury G, 2000,'A Semantic-based Kernel for a Advanced Health Information Systems, *Medical Informatics*, vol. 25, no. 1, pp. 19-43.

Lazarus S, 2006, 'The Evaluation of Wireless Devices used by Clinical Staff at Westmead Hospital: Sydney', In B Unhelkar (Ed.), *Mobile Business: Technological, Methodological and Social Perspectives*. New York: IDEA Group Publishing.

Lindsay A, 2005, 'E-health and E-commerce', Birefing: *Medinexus*, Sydney, Australia

Lindsay A, 2005, Briefing: Telecommunication in Australia, Health Industry Today: *Medinexus*, Sydney, Australia

McAlearney A, et al, 2004, 'Doctors' experience with handheld computers in clinical practice: qualitative study', *BMJ*, vol. 328, no. 7449, pp. 1162-1169.

McLead T, et al, 2003, 'Survey Assessment of Personal Digital Assistant Use among Trainees and Attending Physicians', *AMIA*, vol. 10, pp. 605-607.

Newton H, 2003, 'Telemedicine in educational setting', *Nursing Standard*, vol. 17, no. 44, pp. 75-78.

Palsbo, S, 2004, 'Medicaid payment for Telerehabilitation', *Archives of Physical Medicine and Rehabilitation*, vol. 85, no. 7, pp. 1188-1191.

Pande R, Patel Y, Powrs C, D'Ancona G & Karamanoukian L, 2003, 'The Telecommunication Revolution in the Medical Field: Present application and Future', *Current Surgery*, vol. 60, no. 6, pp. 636-637.

An Efficient IP Traceback Approach

Dr. Wang-Jiunn Cheng, Dr. Maria R. Lee, Chung-Han Sheng
Shih Chien University
Taipei, Taiwan
Email: wjcheng@mail.usc.edu.tw
Email: maria.lee@mail.usc.edu.tw
Email: zaroftw@hotmail.com

Abstract

Tracing IP packets to their original sources is an important mechanism in defending against IP spoofing and Denial of Service attacks. Most proposed approaches, such as ICMP-based, link-testing, logging, or packet marking, are suffering from attacking which must be last long enough for a successful trace, generating more additional network traffic, increasing the size of IP header, or involving the cooperation among all AS routers. In this paper, we propose a new traceback approach by grouping the whole Internet end-users into a reasonable small area network with one and only one router named Internet keeper. The label switching tunnel mechanism will be applied among keepers to realize the traceback in a prevention manner to against IP source address spoofing and Denial of Service attacks promptly. We prove the keeper-based network is IP source address spoofing free and tracebackable for all unicast packets between any pair of hosts, and show that the time/space complexities of routing process are in O(1) at the keeper while the reasonable small area network is no more than 1000 end-users.

Keywords:

IP traceback, Denial of Service, IP source address spoofing, Label switching tunnel

I. INTRODUCTION

Since the stateless Internet routers forward packets only based on their destination IP addresses, attackers can easily flood a victim with spoofing source IP address such as DoS or DDos attacks which can exhaust resources of the victim computer or cause the traffic jam near the victim's network (Ferguson et al., 2000) (Lau et al., 2000). The best way to prevent these malicious packets traveled over the Internet is to promptly and reversely find out the path information of malicious packets to isolate the attackers. In other words, how to identify the malicious packets and start the traceback of these packets to their true source IP addresses has become the key technique to defeat Denial of Service attacks (Snoeren et al., 2002) (Waldvogel 2002) (Bala et al., 2002) (Aljifri 2003).

In this paper, we leave the identification of malicious packet problem to IDS or firewall systems and focus the problem of tracing attacks back toward their true sources. Unfortunately, most IP traceback approaches have suffered from lasting long enough time for reconstructing a successful trace such as packet marking (Song et al., 2000) (Goodrich 2002) (Lee et al., 2001) (Adler 2002), generating more additional network traffic to log the packet traveling information such as ICMP-based or hash-based (Bellovin 2000) (Stone 2000) (Snoeren et al., 2001), deriving more fragments while IP header size incremented to carry more trace information, or requiring the cooperative among routers with the whole ISP network or in AS level such as ingress filtering (Ferguson et al., 2000) (Stone 2000). Instead, we present a new approach for the traceback problem that has two different scenarios of the current Internet routing process.

The first scenario is using the tunnel mechanism to realize the traceback in a prevention way. Since both ends of a tunnel are easily to trace each other in nature, we propose a simple way to create a virtual tunnel in Layer 3 for each pair of hosts across the Internet before they start communicating with each other.

The second scenario is using the ARP address resolution protocol as the network ingress filter to realize the detection of source IP address spoofing at near the user site instead of ISP site. We propose a greedy process such that a local spoofing free of a small group of end-users can be used to achieve a global spoofing free Internet. In this manner, all Internet end-users are grouping into a large amount of reasonable small area networks and each small area network is assumed that it has one and only one single router to connect to the Internet. Throughout this paper, the router is named as the keeper of users of the small area network.

The rest of the paper is organized as follows. Section II describes the architecture of the keeper-based topology of small area networks and how to establish a virtual tunnel for each pair of hosts. The proof of spoofing free and traceback properties will be covered in Section III. The time and space complexities are analyzed in Section IV. Finally, the conclusions of the paper are discussed in Section V.

II. SCHEME TO CONSTRUCT IP TRACEBACK AMONG INTERNET KEEPERS

Due to security issues, most enterprises or organizations install the firewall or IDS between the Internet and their own internal networks to filter all possible malicious packets to/from the Internet. Based on the same scenario, it is reasonable to setup a boundary checkpoint to protect the internal network from Internet attacks. In this section, we propose a keeper-based Internet topology to realize the IP trackback capability among keepers. Some restrictions of keepers are described in subsection A. The topology of keeper-based internetworking is defined in subsection B. Finally, the scheme of constructing tracebackable tunnels among keepers is presented in subsection C and D.

A. Assumptions of keeper

As shown in Fig. 1, each small area network represents a group of end-users of an enterprise or organization. The keeper is the checkpoint between the internal network and the Internet. The relationships among keepers and groups of small area networks are based on the following assumptions:

- Each group of a small area network has no more than 1000 end-users.
- Each group has one and only one default router to connect to the Internet. The router dedicated to support IP traceback capability for end-users of the group is named as keeper. That is, each host of an Internet end-user will have one and only one corresponding keeper.
- Keepers are not nested with each other, that is, groups are not nested or intersected with each other.
- Keeper is considered as a well-protected and backdoor-free Internet router which supports the functions of IP traceback.

B. Keeper-based Internet

As shown in Fig. 1, the topology of keeper-based Internet illustrates that all traffic between any pair of hosts, such as the attacker host H_A and the victim host H_V in Fig. 1, will travel exactly through two keepers, K_A and K_V. Since the concept of IP traceback is not a well-defined network standard, we have defined the process of IP traceback throughout this paper as the following four steps:

Step 1. The attack host H_A generates attack packets to the victim host H_V.

Step 2. The victim host H_V detects that packets with the source IP address of H_A are attacking.

Step 3. On detecting attack packets, the victim host H_V should notify its own keeper K_V to directly traceback to the keeper K_A of the original attack host H_A.

Step 4. The keeper K_A can discard all subsequent attacking packets generated from H_A to H_V or prohibit all H_A's traffic

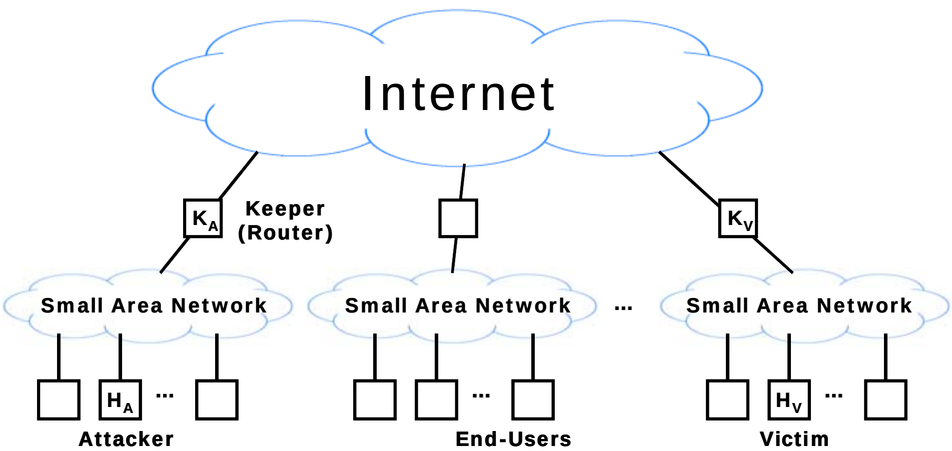


Figure 1. The topology of the keeper-based Internet.

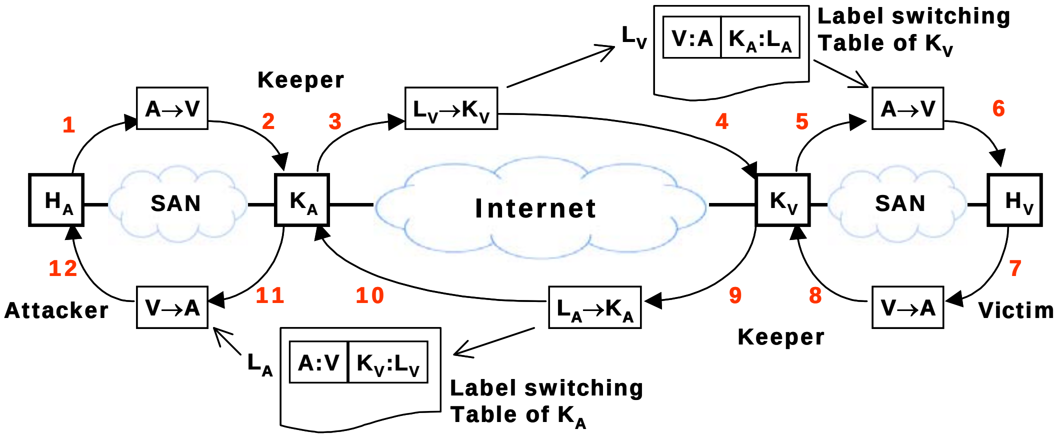


Figure 2. The label switching tunnel constructed by L_A and L_V between two keepers K_A and K_V for delivering packets between a pair of hosts H_A and H_V.

and redirect them to a special VLAN to verify its system availability.

The information of Step 1 and Step 2 should be supported from IP filtering devices such as IDS. The Step 3 is the critical section of traceback process between the two keepers of K_A and K_V. The last step is to perform all possible mechanisms to isolate the attack traffic from entering the Internet.

C. Virtual tunnel between a pair of keepers

As shown in Fig. 2, K_A is the keeper of host H_A and K_V is the keeper of H_V. Assume that a label switching tunnel has been already setup before H_A sends a packet to H_V. That is, the label switching table of K_A should have a record indexed as L_A corresponding to a record indexed as L_V in the label switching table of K_V. Each record of label switching table contains two fields: the key field and the next label field. The key field is used to lookup via a compound key of the destination IP address and the source IP address of all incoming packets which are delivering through the tunnel such as (H_A, H_V) of K_A and (H_V, H_A) of K_V. The next label field consists of the IP address of the other end of keeper and the index of the label switching table at the other end of keeper such as (K_V, L_V) of K_A and (K_A, L_A) of K_V. The pair of label switching records forms a virtual tunnel across the Internet such that all incoming packets generated from H_A with the destination IP address H_V will match the key field of label L_A and be routed to the other end of tunnel K_V.

In order to carry more label information in the IP header and keep the size of IP header unchanged, three fields of IP header are replaced and redefined. The reserved bit of flag of fragment field is set to 1 to indicate that the packet is routed with label switching tunnel protocol between a pair of keepers. The destination IP address field is set to the IP address of the other end of the tunnel. The source IP address field is modified as the switched label index of the other end of the tunnel. After recalculating the checksum field of the IP header, the packet can be forwarding to the other end of the tunnel across the Internet as a normal packet.

On arriving at the end of tunnel, the packet will be scanned by the keeper to verify whether the reserved bit is set or not. All incoming packets with reserved bit setting and destined to the keepers of the both ends of tunnel indicate that they must be applied the label switching protocol again. The keeper should restore both the destination and source IP addresses with the key field of the label record indexed as the label index of the source IP address field, set the reserved bit off, and recalculate the checksum fields before the packet is forwarding to their true destination.

In fact, the two L_A and L_V labels forms a virtual tunnel between K_A and K_V such that all packets between the pair of hosts H_A to H_V will travel through the virtual tunnel between K_A and K_V. In other words, it is easily and promptly for K_V to notify K_A that is H_V under attack by H_A and K_A can drop or stop all subsequent attack packets generated from H_A whenever K_A received the attack notification from K_V and vice-versus.

Ignoring the establishment of label switching tunnel procedure, the scheme has many advantages as follows:

- The packet size is unchanged. No more additional packet should be generated except the attack notification.
- The time complexity of hashed and direct index of the label switch table is negligible.

- The most important property is that no needs of lasting enough long time attack traffic to reconstruct the traceback path information.

Only one disadvantage is deduced that the source IP address is occupied by label between the pair of keepers such that the Internet routers can not correctly generate ICMP messages back to their original source IP address such as in the case of source quench.

D. Establishing a Virtual Tunnel

There are two ways of setting up a virtual tunnel between a pair of keepers. One is in piggyback mode and the other is in ICMP mode. Since the SYN, SYN-ACK, and ACK packets of the three-way handshaking protocol to establish a TCP connection are quite small size, we can add additional label switching information on the IP options and will never generate the fragment problem during the handshaking stage.

In additional, establishing a virtual tunnel, all subsequent traffic including both the connection-oriented and connectionless traffic can travel through the same one tunnel without any effort. That is, only the first time of communicating between a pair of hosts needs to perform the establishment process but not to construct tunnels per connection. If the first packet between a pair of hosts is not the SYN packet of a TCP connection, to issue an ICMP label switching binding request and wait for an ICMP label switching binding response is inevitable.

The binding request and binding response is used to exchange label information between the pair of keeps such as the binding request information of (H_A, H_V, L_A) issued from K_A and the binding response information of (H_V, H_A, L_V) replied by K_V.

III. THE SPOOFING FREE AND TRACEBACK PROPERTIES

In order to prove that the keeper-based Internet has the property of spoofing free is not trivial. If a network has no guarantee of spoofing free, the traceback problem will become more complex to find all attackers they have the same forged source IP addresses. In the worst case, such traceback processes may become one kind of flooding process while all attackers using the same forged source IP address to attack a specified victim simultaneously.

The proof of spoofing free property is two folds. The first is how to support the spoofing free property within the small area network of a keeper. The second is based on the spoofing free keeper-based small area network to realize the global spoofing free across the Internet.

Lemma 1. A Router R receives a packet delivered from host A with the source IP address and source MAC address matches one of the entries in R's ARP table if and only if R is the first hop router of host A.

Proof: Let R be the first hop router of host A. It means that R and A will have the same network number of IP address such that host A must perform the ARP process to lean the MAC address of R before directly forwarding the first packet to R. After the ARP process performed, both R and A will have a cache ARP entry with each other. In such a situation, whenever R receives a packet from A the source IP address and the source MAC address of the packet must be found in the ARP table.

On the contrary, let R receive a packet sent from A and the source IP address and MAC address of the packet match with an ARP entry. Since the IP address of an ARP entry indicates that its corresponding host is directly connected with the route R, R becomes the first hop router of host A.

Lemma 2. If router R receives a packet without matching its source IP address with the IP address of an ARP entry corresponding to its source MAC address, the packet must be indirectly forwarded from a router or directly delivered from a host with spoofing source IP address.

Proof: Assume that there exits a host A such that A can directly deliver a packet to R without matching its source IP address of an ARP entry. By Lemma 1, if router R is the first next hop of host A, the source IP address and the source MAC address of the packet will match with an ARP entry in the router R. That is, it is impossible for a host to directly deliver a packet to router R without matching its source IP address of an ARP entry in R. Since all packets are forwarded from either a host or a router in the Internet, only the router can store-and-forward packets to router R under the case of without matching or a router forged by a host with spoofing source IP address.

Lemma 3. If all routers within a keeper's small area network are trusted, then the keeper's small area network is spoofing free network.

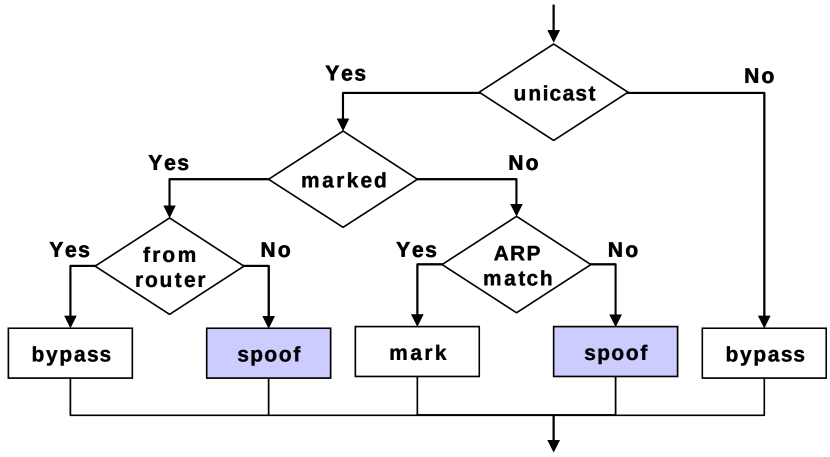


Figure 3. The flow chart of spoofing free verification mechanism for all routers within the small area network.

Proof: Let all routers within a keeper's small area network have performed the ARP verification algorithm and mark the reserved bit of the flag of fragment field to 1 whenever the source IP address and source MAC address matches an ARP entry of a trusted router. Under such conditions, if a router receives a packet with reserved bit mark to 1, it means that the packet has already been verified at its first hop router by Lemma 1. Therefore, it is spoofing free. Otherwise, if the packet matches an ARP entry in the current router, it is also spoofing free by Lemma 1.

The rest condition is a packet without setting reserved bit and without matching an ARP entry. It means that the router is the first hop but fails to match. That is, the packet is absolute a spoofing packet by Lemma 2.

As shown in Fig. 3, it is possible that the attacker knows the ARP verification algorithm and issues spoofing packets with the reserved bit marked. In such a situation, the next router will assume that the packet has already been verified at its first hop router. This is also implied that the packet should be passed from a router to this router. In other words, we should verify to see whether the corresponding IP of its source MAC address of the ARP table is a router or not. If the IP address is found in routing table, the packet is spoofing free, otherwise it is also a spoofing packet.

Theorem 1. If all keepers' small area networks are spoofing free networks, then the keeper-based Internet is spoofing free.

Proof: Let the victim host V receives a packet from the attack host A where A is spoofing free within the keeper K_A's small area network. Assume that there exists a packet generated from host A' with spoofing address A arriving at host V. This is, there will exists a keeper $K_{A'}$ such that host A' is spoofing free within the small area network of $K_{A'}$. Therefore, there will have two hosts A and A' using the same IP address. According to the label switching tunnel mechanism, both of their keepers K_A and $K_{A'}$ will set up a virtual tunnel with the keeper K_V of host V before packets are sent from A or A' to V. Since both A and A' have the same IP address, the label binding request in K_V will detect that both labels (K_A, L_A) and ($K_{A'}$, $L_{A'}$) have the same key field (A, V). It is conflict to assign a tunnel with two different keepers such that one of hosts A or A' is forged. Consequently, host V will never receive any on packet with spoofing source IP address. In other words, the keeper-based Internet not only can support local spoofing free within the small area network, but also support the global spoofing free among all keepers' small area networks.

Theorem 2. If the keeper-based Internet is spoofing free, the Internet is trackbackable.

Proof: Since the keeper-based Internet is spoofing free by theorem 1, it means that there always exists a virtual tunnel between K_A and K_V for any host A to send packets to host V without spoofing source IP address of host A. On detecting the host V under attack by host A, the virtual tunnel can be used to abort or stop all subsequent attack traffic at the end of K_A promptly. Consequently, the spoofing free keeper-based Internet has the traceback property in nature.

IV. ANALYSIS

The flow chart of new routing scheme for routers to support spoofing free property with the small area network is shown in Fig. 3. By lemma 1, a packet without reserved bit marked and without matching ARP entry is a spoof source IP address. Since only the router can forward a packet with reserved bit marked to the next hop router, a packet with reserved bit marked and its corresponding IP address of the ARP entry with the source MAC address not found in router table is also a spoof source IP address. The major time complexity of the scheme is the searching of an ARP table to verify whether the packet is spoofing free. In general, the size of an ARP table of a network interface is proportional to the size of a LAN directly attached to the interface. In fact, the implementation of dynamic cache-based ARP table is usually smaller than the real size of a LAN. Consequently, the time complexity of searching ARP table can be considered as the average size of a LAN, that is, a constant time.

If a packet is verified as a spoofing free one, the marking of the reserved bit of flag and the recalculating of checksum are also in a constant time. In fact, the time complexity of the ingress filtering of checking spoofing free is only a minor problem. How to deploy such a filtering scheme to all kinds of routers within the small area network would become the major deployment problem even only a minor modification of the routing algorithm.

The time and space complexities of establishing virtual tunnels among keepers are significant. The larger end-users of a small area network will cause the larger size of label switching table. The larger size of label switching table implies that the keeper requires more time to search the label for all outgoing traffic (to the Internet). On the contrary, the label switching time for incoming traffic (from the Internet) is negligible because the value of label is the direct index of label switching table. That is, only the search problem of the label switching table of an outgoing packet should be taken into considerations. In addition, the total outgoing bandwidth of a keeper and the server or client type of users are also the major factors of the size of label switching table.

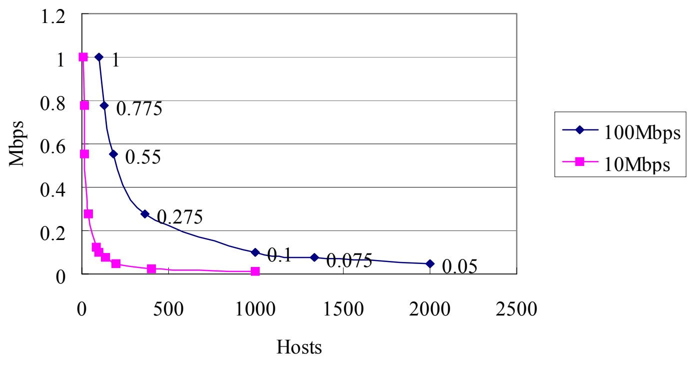


Figure 4. The relationship of the amount of hosts and the outgoing traffic of a host in a fixed outgoing bandwidth of 100Mbps and 10Mbps.

ince the total outgoing traffic will be concentrated to the keeper and the bandwidth of keeper is limited, the amount of users of a small area network can be easily estimated as shown in Fig. 4. In this paper, we will focus on analyzing the case of the following conditions. Assume that there are 1000 hosts with a small area network. Each host with 0.1Mbps outgoing traffic will produce 100Mbps traffic to the keeper. The average packet size is about 380 bytes per packet. Each label switching entry has 20 bytes long and the time to live of a tunnel is last at least 30 minutes long. The average packets per TCP connection is about 20.55 packets (Since the traffic of UDP and ICMP are quite small ratio of the overall unicast traffic, we assume that the TCP traffic dominates the outgoing traffic).

Because the label switching tunnel is a pair of hosts, each tunnel can be reused by subsequent TCP connections without allocating a new label. As shown in Fig 5, in the worst case, each tunnel corresponding to only one TCP connection without reusing, the label switching table is about 83.5MB, that is, there are 1216.5 distinct connections per second. During the time to live of 30 minutes, there will be 4.3M connections corresponding to 4.3M entries of the label switching table.

In the case of 1000 distinct users, two-level hash function can be used to speed up the searching time of label switching table. The first-level hash function uses the source IP address to find the label switching table for each user such that the 4.3M entries can be reduce to 4.3K entries per user. Subsequently, the second-level hash function uses the destination IP address to find the target entry within the size of 4.3K entries. Based on the locality of network applications, a virtual tunnel between a pair of hosts can be reused for two or more TCP connections. While the average times of reusing tunnel are 4 TCP connections, the size of label switching table will become more reasonable size of 1K entries at the second-level hash function. This is, the time complexity of comparing the source IP address and the destination IP address of a outgoing packet with the key field of label switching table is also in constant time by the two-level hash function in the case of 1000 users with 0.1Mbps outgoing traffic.

V. CONCLUSIONS

In our approach, the tunnel scheme has the advantage of the traceback property in nature, but how to embed the label information into each packet in the current IPv4 becomes the big problem in our proposed approach. Due to the replacement of the source IP address with a label, the router will cause a problem of sending ICMP packet back to its true source. Only both keepers of a tunnel know the true IP addresses of both end hosts. That is, if the router wants to issue an ICMP packet back to the source, the ICMP packet should be forwarding to the keeper as the reserved bit marked. In addition, we only consider the unicast traffic because the tunnel is an end-to-end mechanism. Many attacks may use the broadcast or multicast IP address.

To compare with other approaches, our approach has solved many advantages. First, it is no needs of attacking traffic which must be last long enough for a successful trace. If an attacker is identified by IDS system, the pair of keepers can promptly closed the virtual tunnel to drop all subsequent attaching packets. Second, it is no needs to involve the cooperation among ISP or AS routers. The keeper is used as a firewall or gateway for an organization. If both ends have the keeper function, the traceback function is supported. Otherwise, it is fully compatible with a traditional Internet router. Third, the label binding process can be piggybacked with TCP three-way handshaking process. The increasing of IP

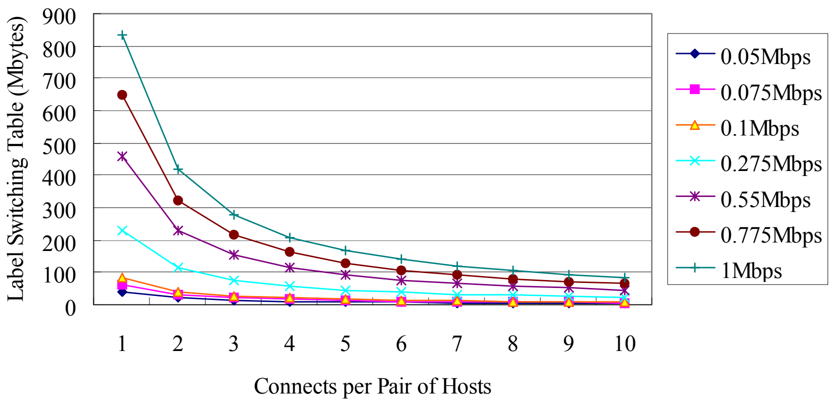


Figure 5. In the case of 1000 hosts of a keeper-based small area network, the size of label switching table related to the outgoing traffic from 1Mbps to 0.05Mbps and the size of table reduced as the reused connections of a pair of hosts increment.

header will never cause the problem of IP fragmentation. During tunnel mode, the size of IP header remains unchanged such that no additional network traffic is generated. Finally, the major advantage of our approach is that the whole Internet is spoofing free if all organizations are under the protection of a keeper in tunnel mode.

REFERENCES

(Ferguson et al., 2000) P. Ferguson and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing," *RFC2827, IETF*, May 2000.

(Lau et al., 2000) F. Lau, S. H. Rubin, M. H. Smith, and L. Trajovic, "Distributed Denial of Service Attacks," in *IEEE International Conference on Systems, Man, and Cybernetics*, pp 2275-2280, October 2000.

(Snoeren et al., 2002)A. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, B. Schwartz, S. T. Kent, and W. T. Strayer, "Single-Packet IP Traceback," *IEEE/ACM Transactions on Networking*, vol,10, no.. 6, pp. 721-734, December 2002.

(Waldvogel 2002) M. Waldvogel, "GOSSIB vs. IP Traceback Rumors," *Proc, 18th Ann. Computer Security Applicatons Conference* (ACSAC 2002), pp. 5-13, 2002.

(Song et al., 2000) D. X. Song and A. Perrig, "Advanced and Authenticated Marking Schemes for IP Traceback," *Proc. ACM SIGCOMM'00*, pp. 295-306, August 2000.

(Goodrich 2002) M. Goodrich, "Efficient Packet Marking for Large-Scale IP Traceback", *Proc. 9th ACM Conference Computer and Communication Security, ACM Press*, pp. 117-126, 2002.

(Lee et al., 2001) W. Lee and K. Park, "On the Effectiveness of Probabilistic Packet Marking for IP Traceback under Denial of Service Attack," *Proc. IEEE INFOCOM, IEEE CS Press*, pp. 338-347, 2001.

(Adler 2002) M. Adler, "Tradeoffs in Probabilistic Packer Marking for IP Traceback," *Proc. 34th ACM Symposium, Theory of computing, ACM Press*, pp. 407-418, 2002.

(Bellovin 2000) S. Bellovin, "ICMP Traceback Messages," *Internet draft: Draft-bellovin-itrace-00.txt*, March 2000.

(Stone 2000) R. Stone, "CenterTrack: An IP Overlay Network for Tracking DoS Floods," in *Proc. USENIX Security Symposium*, pp. 199-212, Aug. 2000.

(Bala et al., 2002) T. Bala and S. Matsuda, "Tracing Network Attacks to Their Sources," *IEEE Internet Computing*, vol. 6, no. 3, pp. 20-26, 2002.

(Aljifri 2003) H. Aljifri, "IP Traceback: A New Denial-of-Service Deterrent?," IEEE Security & Privacy, pp.24-31, 2003.

(Snoeren et al., 2001) A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Joines, F. Tchakountio, S. T. Kent, and W. T. Strayer, "Hash-Based IP Traceback," in *Proc. ACM SIGCOMM'01*, pp. 3-14, 2001.

An Improved System for Processing of Continuous Queries over Infinite Data Streams

Ehsan Vossough
School of Computing and Information Technology
University of Western Sydney, Campbelltown

e.vossough@uws.edu.au

Abstract

This paper describes an ongoing work for the development of system for executing continuous queries over several continuous data streams. It provides an improved solution for processing several streams of continuous data for associative and non-associative relational algebra operations and minimizing storage of the intermediate results between the operations. Processing of data streams are finding their way into many e- business applications such as web-based financial search engines that evaluate queries over real-time streaming financial data for stock tickers and news feeds. Some experimental results also demonstrate improvements in the performance of continuous query from the previous work done in this area.

Keywords:

Data stream management system (DSMS), continuous queries, infinite data streams, online databases, continuous DBMS.

1. INTRODUCTION

A data stream is a continuously growing and theoretically unlimited sequence of data elements. A data stream can be generated from repetitive measurements taken from various real time processes such as medical monitoring systems, financial stock markets, weather monitoring systems, network tracking systems and many other real-time sensing systems. Processing and management of data streams has emerged in the past few years as a new field of research to deal with new organization of data requiring new approaches to both data management and data processing. Current relational Database Management Systems (DBMSs) are not designed to operate the continuously expanding sequences of data. Instead, they process static data containers in more batch-oriented than online modes. A data stream needs to be processed by a modified form of a query that is called a continuous query and requires a special approach to both the management and processing of data. The development of a general purpose Data Stream Management Systems (DSMSs) has been proposed as a challenging research objective [1, 2, 3, 5, 9, 8].

Even if it may not be possible to achieve a similar performance with a DBMS as with a dedicated machine there are some added advantages. They include a reduction in the cost for a stream processing system by utilizing the existing DBMS within an organization and in house maintenance and customization of the system to meet the changing needs of the users.

Data stream processing techniques have its roots in the earlier developed methods for adaptive, incremental, and continuous query processing. A comprehensive review of the works and the results achieved can be found in [4, 6, 7, 11, 13, 14, 16, 17, 27, 29, 33, 35].

The recent work [4] reviews the research issues and unsolved problems in this area. The computational models suitable for data stream processing fit either into the categories of incremental or dataflow processing models. Continuous data processing model originates from the techniques developed for incremental [18, 22, 37], adaptive [3, 21, 23, 24, 25, 34, 39] and continuous query processing [12, 25, 30, 33].

Generally, stream-processing operations should be able to implement the principles of *on-line*, *spontaneous*, and *continuous computations*.

- On-line computations mean that algorithms implementing the computations see only a part of input and never operate on a complete input data set. The algorithms that have such a property are commonly called as *on-line* algorithms [28, 38].
- Spontaneous computations mean that whenever an input data streams is extended with an atomic data element or updated then such modification immediately triggers all applications that process this stream [30, 37].

- Continuous computations mean that the system should continuously re-compute the results of queries and make them available and up-to-date at every moment in time [4, 35, 36].

The remaining sections of the paper are organized as follows. Section 2 provides the motivation for the development of a data stream management system. Section 3 presents a brief background of data stream processing system using a tandem. Section 4 describes the elimination of intermediate repositories in associative and non-associative operations. Section 5 provides some experimental results. The paper concludes with a general vision and plans for future developments.

2. MOTIVATION

In the real world there are many examples of real time monitoring and data collection systems where data needs to be processed continuously. Large web sites that monitor web logs online to enable applications such as performance monitoring and load-balancing. Some web sites may be served by several widely distributed web servers (e.g., Yahoo [42]) that analyse many distributed web logs, e.g. track heavily accessed web pages as part of their real-time performance monitoring.

Web-based financial search engine that evaluates queries over real-time streaming financial data such as stock tickers and news feeds. For example, the *Traderbot* web site [40] supports both one-time and continuous queries posed by its customers.

Some examples in the scientific world include electronic sensors that continuously sample data for processing by a computer; automate record keepings, to ensure that some critical limits in the data are not exceeded and to minimize human errors. Data collected in this way must be analyzed immediately as they are collected. A typical example is a weather monitoring system that checks temperature, atmospheric pressure, rainfall, humidity, wind speed, wind direction and solar energy. A data logger collects the information to be processed, displays them on the monitor and archives them on a hard disk. Collected data could be logged either to a local or to a remote machine using a data acquisition system.

A traditional way of processing continuous data is to develop specialized machines designed to process real time collected data, such as a military application that monitors sensor readings worn by soldiers. Such hardware would require specialized software to run the machine and is usually very expensive to purchase.

Current research on DSMS is motivated by the inherent limitations of conventional DBMSs for the processing of data stream applications. However, many of the functional components of a conventional DBMS could quite easily be re-used for processing data streams. We showed this was possible through appropriate selection of transaction processing techniques, query optimizations, physical database design and fine tuning of the system parameters [36].

The traditional DBMSs target business-oriented data processing. DBMS operates on a passive repository that stores large amounts of data. The human operators perform manual data processing tasks on such a repository. This is normally referred to as Human Active DBMS Passive (HADP) approach. In the latter case, repository data is processed when all the information has been collected from various sources. Data stream applications, however, are systems that monitor and process continuous streams of data. They are capable of processing large amounts of data continuously from many sources.

Very frequent updates, insertions and continuously repeated queries have a very negative impact on the performance of DBMSs. Stock market monitoring applications have a low tolerance for missing data. They can overload the DBMS with real time data making it imperative for DBMS to employ intelligent resource management (e.g. scheduling) and graceful degradation strategies (e.g. load shedding) when data volume is too high. Logging of transactions on continuous streams can also overload the DBMS. This is significant when a DBMS must keep up with logging transactions and at the same time process large volumes of data transactions from several continuous streams.

In this paper we also look at a limitation of SQL for processing of data streams and suggest an alternative approach to process associative binary operations (such as join, union) and non-associative binary operations (such as difference).

3. A STREAM PROCESSING SYSTEM (BACKGROUND)

A model for processing of infinite data streams was presented in [35, 36] for three binary operations; join α_{join}, union α_{union} and difference α_{diff}. This model considers a building block called a *tandem* consisting of two stream operators that concurrently process two data streams. When a data element arrives from a stream it is first appended to a *sequential window* of data elements over that stream. It is called a sequential window because it contains data elements in the increasing sequence of their time stamp as they arrive. The size of a sequential window stays constant. From hereon, the term *window* will refer to a sequential window. When one data element is appended to a window, a last *expired* data element is removed from it as to maintain the window size constant. The new data element is then operated with the window over the second stream.

Operations in a tandem were performed with two threads (in SQL), one for each stream. In the first tandem, one thread runs a loop that operates a newly arrived data element from stream A with a window over the stream B and the results are inserted into a temporary table called a Joint Index Table (JIT_{AB}). Similarly, a single data element from stream B is operated with a widow of data element over stream A and results are inserted into JIT_{AB}.

In a succeeding tandem, the first thread *only* operates on the results in JIT_{AB} that were generated from a new data element, with all data elements in a window over stream C. This operation is performed one data element at a time. A second thread processes a new data element from stream C with *all* data elements in JIT_{AB}. For every new stream a new tandem is added in cascade [36].

Operations for a number of tandems connected in cascade can be described as follows;

Stream A: $(((\delta A < op_{AB} > W_{SB}) < op_C > W_{SC}) \,..\, < op_n > W_{SN}) \rightarrow JIT_{A..M}$
Stream B: $(((W_{SA} < op_{AB} > \delta B) < op_C > W_{SC}) \,..\, < op_n > W_{SN}) \rightarrow JIT_{A..M}$

… and so on for the remaining streams; where δA, δB are data elements from streams A, B respectively. W_{SA}, W_{SB} … W_{SN} are windows over streams A, B … N, respectively. Intermediate results from each tandem is stored in JIT_{AB}, JIT_{ABC}, … $JIT_{A..M}$, respectively. Hence, for N tandems there will be M = N-1 JITs.

Several solutions were proposed to synchronize operations of all tandems, based on the time stamp of each data element. The need for synchronization arises from different inter-arrival delays between data elements and the delayed results from output of one tandem into the next tandem [36]. It was also shown that order of associative operations on several input streams can be changed without affecting the final results. In the following sections associative operations refer to join and union operations and non-associative operations refer to difference operation. Aggregate and zip operations are left to future work.

4. ELIMINATION OF JIT IN ASSOCIATIVE AND NON-ASSOCIATIVE OPERATIONS

This paper looks into the removal of the intermediate and temporary storage for associative tandems to greatly improve performance. We also look at the non-associative operation and suggest a solution.

Consider stream A; a new data element δA is processed within a number of nested loops, one for each stream; first loop operates $\delta A < op_{AB} > W_{SB}$, results are processed immediately one element at a time with window C, $< op_C > W_{SC}$, and so on. To ensure synchronization, only those elements with a lower time stamp than δA, T_A, participate in the operation [36].

With associative operations the order of loops can be changed to accommodate the stream. For example, if a new data element δ*s* with time stamp $T_{\delta s}$ arrives from stream *s*, following nested loop join operations are performed, where δ_t is an element in W_{ST}, and so on;

Receive data element δ_s,
  For all elements δ_t in W_{ST}
    If $T_{\delta s} > T_{\delta t}$
      For all elements δ_a in W_{SA}
      If $T_{\delta s} > T_{\delta a}$
        …
        Join (δ_s, δ_t, δ_a, …) -> Final Result
      End if
    End if

Changing the order of loop for δ_t and δ_a would yield the same results. Similarly, a data element δ_t would initiate a nested loop operation as above; with $T_{\delta s}$ replaced with $T_{\delta t}$ and δ_t is operated with all windows W_{SA}, W_{SB} … W_{SN}, excluding W_{ST}. Operations on data element δ_t flows from outer most loop into the inner most loop until data elements in all windows are operated with δ_t. Results are then appended to a final repository. This operation clearly does not require storage of intermediate results.

Upon the arrival of a new data element from any stream the corresponding tandem initiates a spontaneous operation on the data element. Meanwhile other tandems also operate concurrently on data elements from other streams. The same operation is also applied to removal of an expired data element from the final result. Expired data elements are initially flagged for removal, and are then deleted at a suitable time when there is minimal load on the system.

In a non-associative operation, such as left difference and right difference, we cannot change the order of loops for each operation. For example, consider stream operations A-(B join C). We can remove intermediate results for A-(δB join C) and A-(B join δC) but not for δA-(B join C). Hence, there is a need for JIT_{BC}. The only improvement in the performance can be achieved by eliminating the operation A - read JIT_{BC} for every δA or δB. Ideally, we would like to reduced above expression to a form A-(B intersection C) or (A-B) union (A-C). However, this is not possible since schema for a join operation is different from schema for an intersection. Above method can also be applies to a block of data; when processing JIT, by considering operations on one data element at a time, until all data elements in the block are processed.

5. EXPERIMENTS

Experimental platform were based on windows XP operating system and a Pentium 4-M, 1.6GHz machine with 500 megabytes of memory. One implication of using SQL is the need for intermediate tables for every pair of operation and the lack of nested select statements for associative operations. C++ was considered as a suitable alternative for the implementations that would eliminate storage of intermediate results.

Each stream is made up of people with following data; first name, last name, address, and age. Operations on the streams were based on people with different ages. Every new person arriving on a stream also obtains a unique time stamp. A hash based approach was adopted for join, union and difference operations. The reason for latter two operations will be clarified later.

All windows were implemented as arrays residing in memory. Each window consists of hash buckets for people of different ages with the total size of all buckets the same as the window size. People in each bucket were treated as data sets. Before starting the experiment all buckets were empty. Newly arriving data elements populate the buckets. When the window capacity was reached, for each new person recorded in a bucket, a person with the oldest time stamp was removed, either from the same bucket or from a different one, depending on the person's age. The person with the oldest time stamp is referred to as an *expired* person. Any results associated with an expired person were also removed from the final result, including each JIT, at an appropriate time when the computer is at an idle state. A reference to an expired person was recorded in a temporary array for the later removal operation.

An indexed two dimensional array recorded bucket positions and their sizes. Generally, a bucket in a window is locked by tandem during an operation. This ensures integrity of operations if a bucket is being populated by one tandem while another buckets read from the same bucket, missing some newly added data elements. As one bucket is locked, other buckets in the same window may still be concurrently referenced by other tandems. This results in an improved system performance when different sections of a window are utilized by several tandems. Operations were synchronized as described in [36].

The following tasks were performed within a thread; generation of time stamps, recording of a person in a window, sliding of the window, operations on a person, removal and recording of expired people, keeping a track of system parameters, such as window sizes, and bucket sizes, management of expired data elements and JITs and measurements of test duration. Operations were performed along a left deep or right deep section of a syntax tree where all nodes perform a join or union were implemented with a number of nested loops. Operations were performed in the same order as the nodes traversal.

Figure 1 show the results from joining two streams with sliding window size of 2000 people. The lower curve represents time taken to join two streams with data size as shown on the horizontal scale.

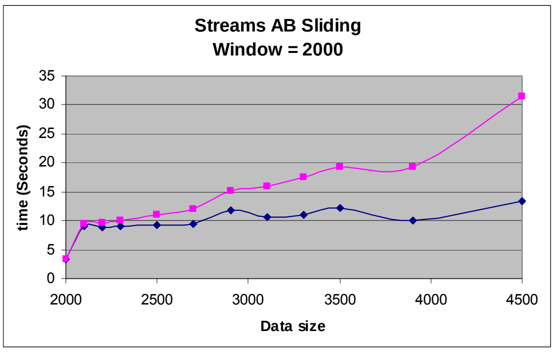


Figure 1: A single join tandem with sliding window of size 2000

The upper curve is the total time (including join time) required to remove expired people from the final results. Results show a considerable improvement using C++ over join operations using SQL [36]; latter time was measured at 28 seconds compared with 3.38 seconds including clean up time of expired data elements.

Experiments with a larger window of 5000 people show a similar proportion of time for cleaning up expired data elements; as can be seen in Figure 2. However, overall performance was still better than results using SQL; 146 seconds versus 5.8 seconds.

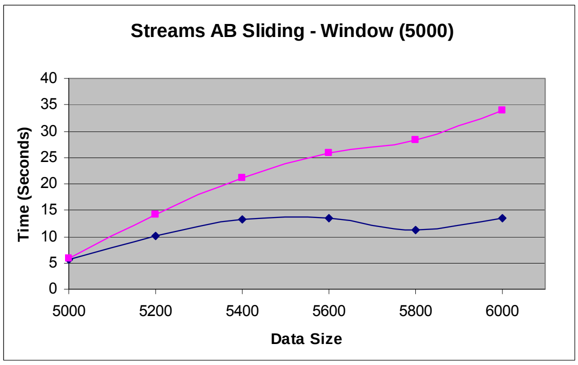


Figure 2: A single join tandem with sliding window of size 5000

Figure 3 shows the outcome for three cascading join tandems. Accumulated clean up time can of the order of five times the actual join, but considerably better than results obtained from two cascading tandems with intermediate storage of results [36].

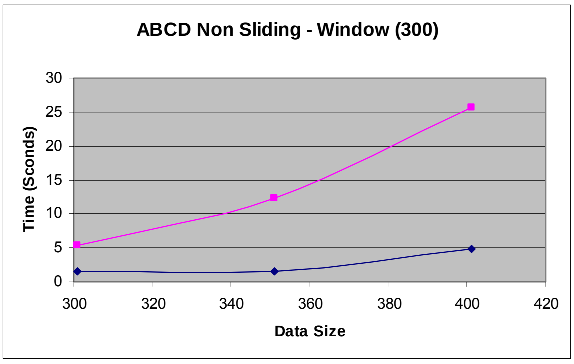


Figure 3: Three join tandems with sliding window of size 300

CONCLUSIONS

A need to process continuous queries over infinite data streams has become a reality. Traditional DBMS systems are geared toward the processing of finite stored data sets rather than a constant flow of data. General limitations of current DBMS is to cope with data elements that arrive quickly and in large numbers. A system described in this paper is an improvement to a system for processing of continuous queries [36]. It was described and tested for some associative and non-associative operations on continuous stream. Experiments showed several orders of magnitude improvement, enhanced by the removal of JITs as intermediate storage of data between tandems. A method of hashed based operations using sequential sliding windows for all streams was discussed.

Future research will look into a distributed network of tandems focusing on what method can best be used for synchronization of stream operations other than applying time stamps in the form of sequence numbers from a global table. Several issues may need to be closely investigated, as follows. What would be effects of network propagation delays on tandem operations? One would expect a larger propagation delay than having two tandems in the close proximity of each other, and the correctness of the final result would be delayed in proportion to the network delays.

System described here can be adapted to web based applications that use a modified form of existing DBMSs to deal with dynamic flow of data between elementary operators. Such a modification would require an introduction of additional SQL instructions and modules developed to manage operation of continuous queries.

REFERENCES

[1] Arasu, A. *et al*. (2003). STREAM: The Stanford Stream Data Manager, In Bulletin of technical committee on Data Engineering, IEEE Computer Society, 26(1), pp. 19-26.

[2] Abadi, D. *et al.* (2003). Aurora: A data stream management system. In Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 663–663.

[3] Avnur, R., Hellerstein, J. M. (1998). Eddies: Continuously Adaptive Query Processing, Proc. ACM-SIGMOD International Conference on Management of Data, No. 2, pp.106-117.

[4] Babcock, B. *et al.* (2004). Load Shedding for Aggregation Queries over Data Streams, Proceedings of the 20th International Conference on Data Engineering, pp.350.

[5] Babcock, B. *et al.* (2003). STREAM: The Stanford Stream Data Manager, In Bulletin of technical committee on Data Engineering, IEEE Computer Society, pp.19-26.

[6] Babcock, B. *et al*. (2003). Chain: Operator scheduling for memory minimization in data stream systems. In Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 253–264.

[7] Babcock, B. (2003). Maintaining variance and k-medians over data stream window. In Proceedings of the Twentieth ACM SIGACT-SIGMODSIGART Symposium on Principles of Database Systems, pp. 234–243.

[8] Babcock, B. *et al.* (2002). Models and Issues in Data Stream Systems, ACM PODS, Madison, Wisconsin, pp.1-16.

[9] Babu, S., Widom J., (2001). Continuous Queries over Data Streams, SIGMOD Record, 30(3), pp. 109-120.

[10] Carney, *et al*. (2002). Monitoring streams a new class of data management applications. In Proceedings of the 28th International Conference on Very Large Data Bases (VLDB), Hong Kong, China.

[11] Carney, *et al*. (2003). Operator Scheduling in a Data Stream Manager. In proceedings of the 29th International Conference on Very Large Data Bases (VLDB'03).

[12] Chen, J. (2000). NiagraCQ: A scalable continuous query system for Internet databases. In Proc. of the 2000 ACM SIGMOD Intl. Conf. on Management of Data, pp. 379–390.

[13] Cherniack, M. *et al.* (2003). Scalable Distributed Stream Processing. In proceedings of the First Biennial Conference on Innovative Database Systems (CIDR'03), Asilomar, CA

[14] Cranor, C. *et al.* (2003). The Gigascope Stream Database, In Bulletin of technical committee on Data Engineering, IEEE Computer Society, pp.27-32.

[15] Cola, J. L. and Pouzet, M. (2002). Type-based Initialization Analysis of a Synchronous Data-flow Language, Electronic Notes in Theoretical Computer Science 65 No. 5.

[16] Das, A. *et al*. (2003). Approximate join processing over data streams. In Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 40– 51.

[17] Domingos, P., Hulten, G. (2000). Mining high-speed data streams. In ACM KDD Conference Proceedings.

[18] Feldman AAM1997, (1997). Efficient algorithms for discovering frequent sets in incremental databases. In DMKD Workshop proceedings.

[19] Getta, J., Vossough, E. (2004). Optimization of data stream processing. Association of Computer Machinery-Special Interest Group on Management Of Data (ACM- SIGMOD) Record, SIGMOD Record, Volume 33, issue 3, pp. 34-39.

[20] Getta, J., Vossough, E. (2004). Grouped processing of relational algebra expressions over data streams. Advanced DataBaese and Information Systems (ADBIS)2004, Budapest, Hungary, Volume 3255, 22-25 September 2004(336-347), ISBN: 3-540-23243-5.

[21] Hellerstein, J.M. *et al.* (2000). Adaptive Query Processing: Technology in Evolution, Bulletin of the IEEE Computer Society Technical committee on Data Engineering, 23(2): pp.7-18.

[22] Lars, B., Leo Mark (1995). Incremental Computation of Time-Varying Query Expressions, IEEE Transactions on Knowledge and Data Engineering, vol. 7, No. 4, pp. 583-590.

[23] Madden, S. *et al.* (2002). Continuously adaptive continuous queries over streams. In Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, Madison, Wisconsin, pp. 49–60.

[24] Madden, S. *et al.* (2003). Telegraphcq: An architectural status report. Bulletin of the Technical Committee on Data Engineering, pp. 11–18.

[25] Olston, C. *et al*. (2003). Adaptive filters for continuous queries over distributed data streams. In Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 563–574.

[26] Srinivasan, V., Carey, M.J. (1991). Compensation-Based On-Line Query Processing, Proceedings of the 1992 ACM SIGMOD International Conference on Management of Data, pp. 331-340.

[27] Stonebraker, M. *et al.* (2003) The Aurora and Medusa Projects, In Bulletin of technical committee on Data Engineering, IEEE Computer Society, pp. 3-10.

[28] Sumii, E., Kobayashi, N. (2000). Online-and-Offline Partial Evaluation: A Mixed Approach, PEMP00, ACM 2000, pp. 12-21.

[29] Tatbul, *et al*. (2003). Load Shedding in a Data Stream Manager. In proceedings of the 29th International Conference on Very Large Data Bases (VLDB'03).

[30] Terry, D. *et al*. (1992). Continuous Queries over Append-Only Databases, Proceedings of the 1992 ACM SIGMOD International Conference on Management of Data, pp. 321-330.

[31] Tucker, P. A. *et al*. (2003). Applying punctuation schemes to queries over continuous data streams. Bulletin of the Technical Committee on Data Engineering, pp. 33–40.

[32] Vossough E., Getta, J. (2001). Block Level Query Scrambling Algorithm within Distributed Multi-Database Environment, Proceedings of the DEXA 2001 12th International Workshop on Database and Expert Systems Applications, pp. 123-127, 3-7.

[33] Vossough, E., Getta, J. (2002). Processing of Continuous Queries over Unlimited Data Streams, Proceedings of the DEXA 2002 14th International Workshop on Database and Expert Systems Applications, pp. 123-127, 3-7.

[34] Vossough, E. (2002). Adaptable Processing of Large Data Sets in Distributed Environments, Proceedings of the EDBT 8th International Conference on Extending Databases Technology, pp 101-110.

[35] Vossough, E. and Getta, J. (2003). Processing of Data Streams in Relational Database Systems, the 7th IASTED International Conference on Software Engineering and Applications, Marina del Rey, USA, pp. 341-346.

[36] Vossough, E. (2004). A System for processing continuous Queries over infinite data streams. Databases and Expert Systems Applications (DEXA) Conference, Zaragoza, Spain, September 2004, 720-729, ISSN: 0302-9743.

[37] Yellin, D. M. (1991). INC: A Language for Incremental Computations, IBM T. J. Watson Research Center, ACM Transactions on Programming Languages and Systems, No. 2, pp. 211-236.

[38] Yellin, D. M., Haas, P.J. (1999). Ripple Joins for Online Aggregation, In Proc. Of the 1999 ACM SIGMOD Intl. Conf. On Management of Data, pp. 287-298.

[39] Zachary G. *et al*. (2000). Adaptive query processing for Internet applications. Bulletin of the Technical Committee on Data Engineering, pp. 19-26

[40] http://www.traderbot.com.

[41] http://www.unitracking.com

[42] **http://www.yahoo.com**

Fingerprint Recognition with Pattern Matching Technique using Fuzzy Logic

D. M. Akbar Hussain, Zaki Ahmad*, Shaiq A. Haq**
Assistant Professor
Department of Computer Science and Engineering
Aalborg University Esbjerg
Niels Bohrs Vej 8, 6700 Esbjerg DENMARK
Phone: (+45) 7912 7729 Fax: (+45) 7512 7710
Email: akbar@cs.aue.auc.dk
* School of Electronic, Communication and Electrical Engineering
University of Hertfordshire
College Lane, Hatfield, Herts
AL10 9AB UK
Office Tel. +44 1707286279
Email: z.1.ahmad@herts.ac.uk
**Professor and Chief Consultant IT, IT Research Centre
University of Engineering & Technology, Lahore PAKISTAN
Office Tel. +92(42)682-9397.
Email: shaiq_haq@yahoo.com

Abstract:
A fingerprint is a pattern of ridges and valleys that exist on the surface of the finger. The uniqueness of a fingerprint is typically determined by the overall pattern of ridges and valleys as well as the local ridge structures e.g., a ridge bifurcation or a ridge ending, which are called minutiae points. Designing a reliable automatic fingerprint matching algorithm is quite challenging. However, the popularity of fingerprint sensors as they are becoming smaller and cheaper, automatic identification based on fingerprints is becoming not only attractive but an alternative complement to the traditional methods of identification. The critical factor in the widespread use of fingerprints identification is, satisfying the performance e.g., speed of matching and accuracy requirements of the application. The widely used minutiae-based representation utilizes this discriminatory information available in a fingerprint for identification. However, we propose a fingerprint recognition technique using fuzzy logic. This approach utilizes two types of features; ridge ending and bifurcation to construct triangles and the fuzzy logic determine the degree of membership for matching. The unique difference in our implementation is that fuzzy concept is used for matching rather than the feature extraction. The laboratory results obtained indicate that matching process efficiency can be improved using this technique.

Keywords:
Fingerprint, Recognition, Biometrics, Fuzzy, Gabor Filter, Identification, Minutiae.

1. INTRODUCTION

Fingerprint Identification System is an important biometric technology. Feature extraction (minutiae) is a key step in a fingerprint matching system. Some time because of imperfections in the image acquisition process minutiae extraction methods could miss certain information for example error in the location of the minutiae [1][2]. However, feature extractions are the most widely used attributes for fingerprint identification and recognition.

The following section discusses in detail the entire technique for extracting these features. After having these features extracted the system creates triangles of reasonable sizes with these minutiae (bifurcation/ending) as shown in figure 1. The matching process uses a fuzzy logic. Fuzzy logic has been used in discovering vagueness and complexity as found on a fingerprint image [19]. However, our implementation of fuzzy logic is used for matching instead of uncovering feature attributes.

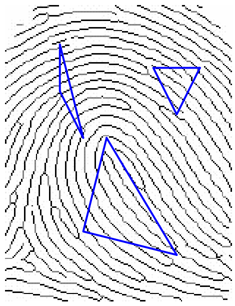


Figure 1: Example Triangles

1.1 Background of Minutiae Based Method

Typically, each minutia is described by its location in the image ***(x, y)***, orientation θ and the type ***t*** meaning if it is a ridge ending or bifurcation etc., as shown in figure 2. In general some other attributes can also be included in the minutiae. Suppose ***T*** represents the template (database image) and ***I*** (query or suspect image) represent the input fingerprint with location, orientation and type:

$$T = \{m_1, m_2, \ldots\ldots m_m\} \quad m_i = \{x_i, y_i, \theta_i, t_i\} \quad i = 1 \ldots m \quad (1) \quad I = \{n_1, n_2, \ldots\ldots n_n\} \quad n_j = \{x_j, y_j, \theta_j, t_j\} \quad j = 1 \ldots n \quad (2)$$

Where ***m*** and ***n*** are numbers of minutiae in ***T*** and ***I*** respectively. A minutia is considered matched with another minutia if the spatial distance D_s, between them is equal or smaller than a given threshold d_0 and direction difference θ_d is equal or smaller than a given threshold θ_0 [3].

$$D_s(n_j, m_i) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \leq d_0 \quad (3) \quad \theta_d(n_j, m_i) = \min(|\theta_j - \theta_i|, \quad 360^0 - |\theta_j - \theta_i|) \leq \theta_0 \quad (4)$$

In order to maximize the number of minutiae matching ***T*** and ***I,*** must be aligned. However, correct alignment some times requires, displacement in ***x, y*** and θ to be recovered. Another factor which could be significant is the scale in case the fingerprints are obtained at different resolutions. Tolerating a higher number of transformations results in additional degrees of freedom to the fingerprint matcher, therefore, this issue has to be carefully evaluated as each degree of freedom bring whole set of possible alignments and may result in false matching. In our study we are not considering this issue.

2. FINGERPRINT MATCHING ALGORITHM

A fingerprint matching algorithm has to execute a number of intermediate steps before the real matching process begins. A flowchart for these processes is shown in figure 3. In the following paragraphs a background for these processes in relation to our implementation is being provided [3][4][5][16][17]. It can be seen that whole methodology looks computationally expensive due to the required processing. Therefore, not only innovation in these algorithms for various blocks is required but matching process also needs more ideas for example based on fuzzy logic. This was our motivation for investigating and proposing a matching algorithm based on fuzzy logic.

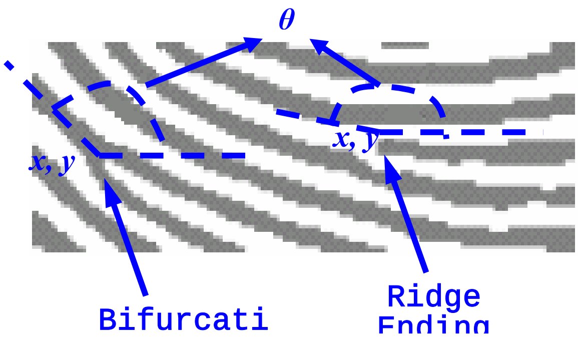


Figure 2: Minutiae for Ridge Bifurcation and Ending

2.1 NORMALIZATION

The image has to be normalized to a pre-specified mean and variance. Normalization process reduce variations in grey level values along ridges and valleys, the normalized image is given as follows [3];

$$\mathrm{N}(i,j) = \begin{cases} \mu_0 + \sqrt{\dfrac{\sigma_0 [P(i,j) - \mu]^2}{\sigma}} & \text{if } P(i,j) > \mu \\ \mu_0 - \sqrt{\dfrac{\sigma_0 [P(i,j) - \mu]^2}{\sigma}} & \text{otherwise} \end{cases} \quad (5)$$

Where ***P(i, j)*** is the grey level value at pixel location ***(i, j)***, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the estimated mean and variance of ***P*** respectively. ***N(i, j)*** is the normalized grey level value at pixel location ***(i, j)***, $\boldsymbol{\mu_0}$ and $\boldsymbol{\sigma_0}$ are the desired mean and variance values respectively. The estimated mean and variance can be computed as;

$$\mu = \frac{1}{height*width} \sum_{i=0}^{height} \sum_{j=0}^{width} P(i,j) \quad (6)$$

$$\sigma = \frac{1}{height*width} \sum_{i=0}^{height} \sum_{j=0}^{width} [P(i,j) - \mu]^2 \quad (7)$$

2.2 Orientation

For the orientation, image is divided into blocks of size *(h * w)* e.g., in our case (3 * 3) and then gradients ∂_x***(i, j)*** and ∂_y***(i, j)*** at each pixel are calculated using Sobel operator. Sobel edge detection mask detects vertical and horizontal edges separately. And these directional edges are combined finally [6]. Sobel gradient 3x3 mask is shown below.

a_0	a_1	a_2
a_3	a_4	a_5
a_6	a_7	a_8

$$\partial_x = (a_2+2a_5+a_8)-(a_0+2a_3+a_6)$$

$$\partial_y = (a_0+2a_1+a_2)-(a_6+2a_7+a_8)$$

Sobel masks for ∂_x and ∂_y are defined as respectively.

-1	0	1		1	2	1
-2	0	2		0	0	0
-1	0	1		-1	-2	-1
	∂_x				∂_y	

The least mean square estimate of the local ridge orientation centered at pixel location *(i, j)* is computed as [7][8];

$$\theta(i, j) = \frac{1}{2} \tan^{-1}\left(\frac{\partial_y(i, j)}{\partial_x(i, j)} \right) \qquad (8)$$

2.3 Ridge Frequency

The ridge frequency is made by the ridges and valleys in the local neighbourhood as a sinusoidal shaped plane wave. Determination of the frequency is made by dividing the normalized image ***'N'*** into blocks of certain width and length *(w * w)*, in our implementation it is (10 * 10).

For every block centered at *(i, j)* an orientation window of size (*l * w)* is created as shown in figure 4, for our implementation this window is of size (32 * 10). Now *x* signature *x[0], x[1], ……x[l-1]*, of the ridges and valleys is calculated for each block centered at *(i, j)* [7] from the following expression:

$$x[k] = \frac{1}{w} \sum_{n=0}^{w-1} N(u, v) \qquad k = 0,1,..(l-1) \qquad (9)$$

$$u = i + \left(n - \frac{w}{2}\right) \cos\theta(i, j) + \left(k - \frac{l}{2}\right) \sin\theta(i, j) \qquad (10)$$

$$v = j + \left(n - \frac{w}{2}\right) \sin\theta(i, j) + \left(\frac{l}{2} - k\right) \cos\theta(i, j) \qquad (11)$$

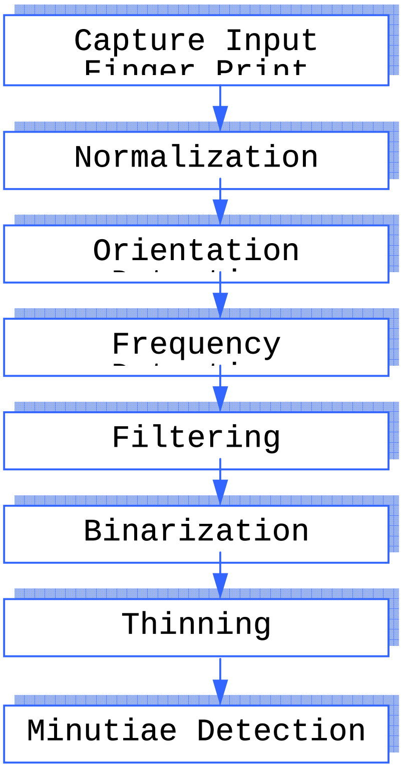


Figure 3: Steps Leading to Minutiae Detection

It can be seen that when there are no minutiae points in the oriented window then x signature will form a discrete sinusoidal shape wave having the same frequency as of ridges and valleys in the oriented window. The frequency ***f(i, j)*** can be computed by finding the average number of pixels ***γ(i, j)*** between the two peaks. The frequency is

$$f(i, j) = \frac{1}{\gamma(i, j)} \qquad (12)$$

For a fingerprint image scanned at 500 dpi typically, has frequency range between 1/3 - 1/25 [7].

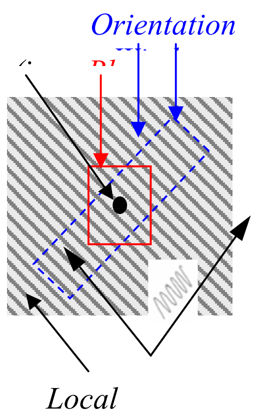


Figure 4: Block and Oriented Window Description

2.4 Filtering

Filtering is necessary to remove undesired noise from the fingerprint image [9][12][13]. Gabor filter [10][11] is typically used as a band-pass filter to remove this undesired noise and to keep true ridges and valleys structures preserved [7]. It is defined as:

$$h(x, y : \varphi, f) = \exp\left\{-\frac{1}{2}\left[\frac{x_\varphi^2}{\delta_x^2} + \frac{y_\varphi^2}{\delta_y^2}\right]\right\}\cos(2\pi f x_\varphi) \quad (13)$$

$$x_\varphi = x\cos\varphi + y\sin\varphi \quad (14)$$

$$y_\varphi = -x\sin\varphi + y\cos\varphi \quad (15)$$

Where φ is the orientation of the Gabor filter, f is the frequency of a sinusoidal plane wave, δ_x and δ_y are the space constants of the Gaussian envelope along x and y axes, respectively.

In order to implement the Gabor filter three parameters frequency, orientation and the space constants have to be provided. Therefore, two parameters frequency and orientation are obviously obtained from the local orientation and the local ridge frequency just described. However, the choice of space constants involves a compromise. Large value means more rectification of noise, but on the other hand more chances of creating spurious ridges and valleys. Small value means it may create less spurious ridges and valleys but filtering of noise is affected. Therefore, best option could be to obtain an average value by empirically from a set of data images. In our implementation we have taken a value of 3 for both constants. The enhanced image $E(i, j)$, can be obtained by:

$$E(i, j) = \sum_{u=-\frac{w_g}{2}}^{\frac{w_g}{2}} \sum_{v=-\frac{w_g}{2}}^{\frac{w_g}{2}} h(u, v : \theta(i, j), \varphi(i, j)), N(i-u, j-v) \quad (16)$$

where θ, φ, N and w_g are the orientation, frequency, normalized image and the size of the Gabor filter respectively. In our implementation the size is 11.

2.5 Binarization

Binarization of the image basically, facilitates further processing. Because, true information is binary; ridges against background. This process takes the greyscale image and converts it into the binary image. However, as the fingerprint image do not pose same contrast characteristics, using a single intensity threshold could be inappropriate. The contrast characteristics even can vary if the finger is pressed more at the center or sides. Therefore, typically a locally adaptive thresholding is often used. The implementation here assumes a uniform contrast therefore, an average color value is determined from the image which is used as a threshold to transform the pixels.

2.6 Thinning

Before the minutiae detection the binary image is passed through a thinning process which converts the widths of ridges to a single pixel. Importantly, a reasonable thinning process should maintain connectivity and also produce minimum artefacts. These artefacts are primarily from erroneous bifurcations with one very small offshoot.

3. MINUTIAE EXTRACTION

Minutiae detection process takes the thinned image and extracts the minutiae points from that image. Ridge ending could be found where a thin line terminates and bifurcation could be found at the junction of three lines. As it is said in the previous paragraph there will be extraneous minutiae points because of the processing so far especially due to filtering and thinning. These extra points are eliminated here using a threshold which for example in the case of bifurcation determines if any of the two branches of it is shorter than a certain value so it is not considered a valid minutia point. Also, for example in the case of two endings found opposite to each other are most likely to be the same line. Endings at the boundary of the image are also not valid minutiae candidates.

4. FUZZY SET DEFINITION

In typical fuzzy minutiae extraction system the feature extraction is made by uncovering the grey pixel values for fuzzy computation. These grey level pixels are mapped against two membership functions having numerical values of 0 to 1, representing the degree of membership [19]. Similar concept is used in our study but the unique difference is that fuzzy set is used for matching specific regions of fingerprint image created by these features as opposed to feature extraction itself. The membership function chosen to represent these fuzzy sets is shown in figure 5 where 1 indicates a complete membership and 0 obviously denote a complete exclusion. The slope represents degree of membership. There are three variables in the fuzzy set, firstly, area of the triangle, secondly, 3 angles of triangle and thirdly, the type of minutiae used to construct the triangle.

The fuzzy logic model these three variables in the fuzzy set with appropriate fuzzy rules to find if a match exists between a query fingerprint and a database fingerprint image.

5. Implementation and Fuzzification of Data

The matching process starts by utilizing fuzzy conditional (if, then & else) rules to obtain membership function for each triangle pair (query/suspect and database triangle). These conditional statements give a qualitative expression of the rule resulting in a classification for numerical representation. The membership degree is computed by fuzzifying this classification data. In the implementation there are three levels of conditional rules, one level determine the full membership second determine the exclusion and the third level is for all those marginal cases. This process continues until all the triangles are exhausted in the query image. While this processing is being performed each level also has a unique tag index which is incremented each time it passes through that level. Now at the end if the tag index percentage corresponding to the full membership level is 90 % or more then there is no further processing required as matching (identification) is declared. On the other hand if tag index percentage corresponding to exclusion level is 11 % or more a mismatch result is declared. The marginal cases are handled differently, for each triangle side the number of ridges are counted and compared, if the difference is less than a threshold it is assumed to be a matching triangle. This case is now moved into the full membership level and the tag index is incremented. Similarly, all the cases are dealt and if at the end the tag index percentage corresponding to full membership level is 90 % or more a matching decision is made.

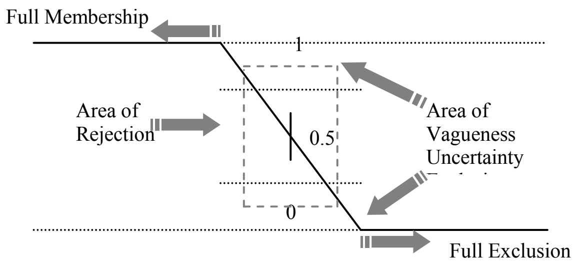


Figure 5: Membership Function

6. CONCLUSION

The complete system was implemented using a fingerprint scanner FDC-FPC1010 daughter card with its platform support board DSK6713 from Texas Instruments. The resolution for this scanner is ~360 less than for example used typically by FBI or other law enforcing agencies. Laboratory tests with a moderate number of samples were conducted and it shows an average improvement of about 10 % compared with if we use only a standard minutiae fingerprint matching method. One reason of this improvement performance is because of fuzzy set concept basically modelling marginal cases more realistically, which in classical system would be rejected. However, the expected performance should be a bit higher to make it more significant. In future we would like to investigate couple of things; implementing the same system without a matching sensor data image, currently our query image and the database images are scanned with the same sensor. Also we would like to consider all kinds of image prints (ARCH, LOOP and WHORL). The processing time is another factor we would like to investigate more with all these additional changes. The concept of using triangles was inspired from a very novel approach for identification provided in references [14][15][18], where triangles are formed using 3 minutiae points. After forming the triangles a number of indexing components relating to these triangles are evaluated and an index score is computed for a final decision. Our implementation is more of a simpler nature of comparison based on fuzzy logic.

7. References

[1] Jain, A., Hong, L. and Bolle, R.: On-Line Fingerprint Verification. IEEE-PAMI, Vol.19, No.4, pp. 302-314, Apr. 1997.

[2] Ratha, N., Karu, K., Chen, S., and Jain, A.: A Real Time Matching System for Large Fingerprint Databases. IEEE-PAMI, Vol. 18, No. 8, pp. 799-813, Aug. 1996.

[3] U. Halici, L.C. Jain and A. Erol: Introduction to fingerprint recognition, Intelligent Biometric Techniques In Fingerprint and Face Recognition, CRC Press 1999.

[4] X. Jiang and W. Y. Yau: Fingerprint minutiae matching based on the local and global structures, Proc. Int. Conference on Pattern Recognition, Vol. 3, pp 1038-1041, 2000.

[5] A. K. Jain, S. Prabhakar: Fingerprint matching using minutiae and texture feature, Proc. Int. Conference on Image Processing, Vol. 3, pp 282-285, 2001.

[6] A. R. Rao: A: Taxonomy for Texture Description and Identification, New York: Springer-Verlag, 1990.

[7] Hong, L., Wan, Y.and Jain, A.: Fingerprint Image Enhancement: Algorithm and Performance Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 20: 777 -789. 1998.

[8] Marious Tico, Pauli Kuosmanen: Fingerprint Matching Using an Orientation Based Minutia Descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 8, 2003.

[9] Greenberg,S. Aladjem,M. and Kogan,D.: Fingerprint Image Enhancement using Filtering Techniques, Real-Time Imaging 8 ,227-236 (2002).

[10] J. G. Daugman: High Confidence Recognition of Persons by a Test of Statistical Independence, IEEE Transactions Pattern Analysis Machine Intelligence, Vol. 15, No. 11, pp 1148-1161, 1993.

[11] J. G. Daugman: Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two Dimensional Visual Cortical Filters, Journal Opt. Soc. Amer. A, Vol. 2, pp 1160-1169, 1985.

[12] Anil K. Jain, Salil Prabhakar, Lin Hong and Sharath Pankanti: Filterbank-Based Fingerprint Matching, IEEE Transactions on Image Processing Vol. 9, pp 846-859.

[13] J. H. Chang and K. C. Fan; Fingerprint Ridge Allocation in Direct Grey Scale Domain, Pattern Recognition, Vol. 34, pp 1907-1925, 2001.

[14] B. Bhanu and X. Tan: A Triplet Based Approach For Indexing of Fingerprint Database for Identification: Proc. Int. Conference on Audio and Video Based Biometric Person Authentication, pp 205-210, 2001.

[15] B. Bhanu and X. Tan: Fingerprint Indexing Based on Novel Features of Minutiae Triplets, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5), pp 616-622, 2003.

[16] A.K. Jain and S. Minut: Hierarchical kernel fitting for fingerprint classification and alignment, Proc. International Conference on Pattern Recognition, Vol. 2, pp 469-473, 2002.

[17] C.T. Hsieh, Z.Y. Lu, T.C. Li and K.C Mei: An effective method to extract fingerprint singular point, Proc. International Conference on high performance computing in the Asia-Pacific Region, vol. 2, pp 696-699, 2000.

[18] Z. M. Kovacs-Vajna: A fingerprint verification system based on triangular matching and dynamic time warping, IEEE Transactions Pattern Analysis and Machine Intelligence, 22 (11), pp 1266-1276, 2000.

[19] Sagar V. K., Ngo D. B. L., Foo K. C. K: Fuzzy control for fingerprint feature selection, ACCV'95 Conference, Vol. 3, pp. 767-771, Dec. 5-8 1995.

8. Acknowledgement

The author likes to thank CTIF (Center for Teleinfrastruktur), Aalborg University for providing the hardware, FDC-FPC1010 daughter card and DSK6713 board.

A Flexible Context-Aware Scheme for Mobile e-Services

Kuen-Liang Sue, Wen-Jr Wu and Chung-Hsien Tsai
National Central University
Jhongli, Taiwan
Email: klsue@mgt.ncu.edu.tw, {93423037, 93423002}@cc.ncu.edu.tw

Abstract

The mobile communication technologies contribute to the development of the novel mobile commerce. The context-awareness is one of the most potential mobile services. It collects the data about mobile device users and their surroundings, and provides these users with the contents in which they are interested. By using the fuzzy logical processing model in our research, we are able to make basic data correspond to their high-level information so that the recognition of users' behaviors and situations could be inferred. We also illustrate our scheme in detail with the numerical examples. Many other mobile applications could depend on the results of our activity recognition mechanism, and then make it possible to transmit the contents that users demand in the present or in the near futures. Briefly, we propose a scheme with flexibility and interoperability to promote the progress of context-aware technologies, and this scheme could make our lives more convenient.

Keywords:

Mobile commerce, Context-aware applications, Fuzzy logic, Activity recognition

1. INTRODUCTION

Due to the rapid development of the Third Generation (3G) and other mobile technologies, mobile phones and personal digital assistants (PDA) turn into more and more available to all; meanwhile, the public gradually cannot have a modern living without mobile communication networks. People may need the mobile devices anytime and anywhere, so the use of services which are concerned with the mobile communication networks become a new trend. The commercial service via the mobile devices and wireless communication networks is generally referred to as mobile commerce or m-commerce. M-commerce will produce large-scale commercial opportunities along with very high monetary value.

As mobile computing becomes widespread, mobile users could widely received mobile applications which is becoming a popular business and creating the opportunities for mobile commerce. The context-aware service is one of the important applications. Context-aware applications can discover and take advantage of contextual information that is used to characterize the situation of people. Once the context-aware application develops maturely, more immediate users' demands could be satisfied. For example, when people stroll around the streets, they can automatically get the promotion information or digital coupon from near bookstores. When driving on the road, drivers are able to get the information about neighboring gas stations, not about unessential advertisement from far. When the audiences watch a baseball game on television, they can have the related statistics from the baseball announcer; however, they cannot have it when sitting at the field. Soon, as long as a user subscribes to a specific service, the hand-held device can receive interesting data about this game immediately.

To manage the context-aware information, the context-aware attributes and its framework should be well-organized. Hence, the principal goal of this research is exploiting basic sensor's measurement through the mobile communication to recognize a hand-held device owner's scenario. Our mechanism designed in this research possesses high flexibility for adjustment, so that it could be easily modified by setting the appropriate input or output items to fit the actual demand in the future implementation. Most importantly, the paper provide a method to realize the concept of context awareness.

The rest of this paper is organized as follows. The previous works related to this paper is introduced in section 2. In section 3, the proposed framework is described. Section 4 introduces how the proposed scheme works in detail. The numerical examples are examined in section 5. Finally, we show the possible applicable aspects and draw our conclusions in section 6.

2. RELATED WORKS

The location-based service is an essential class of context-aware services. Both the location-based and other context-aware services would employ the location information supplied by the communication infrastructure or satellite positioning. Here the context-aware is widely collecting surroundings information, and then providing the contents that appeal to users (Chalmers et al., 2004; Lee et al., 2002). Whether a novel business model of mobile service can succeed or not will be partially decided by the context-aware and location information. Therefore, how to design this kind of framework would be a key factor. Ubiquitous computing or pervasive computing is one of the visions about information technologies and wireless communication services in the present generation. Pervasive computing makes appropriate network technologies to support human daily life anytime and anywhere. This is an important field to develop the oncoming mobile technologies and information systems (Kindberg and Fox, 2002). The above-mentioned affects human life greatly, and we cannot reach the perspective without location information applications.

In order to manage the context-aware information systematically, Korpipaa et al. (2003) designed the context-aware attributes as well as the designed structure, and they also measured and classified the factors in sound, light, temperature and humidity in an environment. Thus we are able to obtain the high-level information from those basic factors. In brief, their research conceived a software and hardware framework, and the server side mainly included application, resource server, context recognition service, and so on. If users would like to know high-level contents, the context manager would reply to them what they want according the situation from basic sensing value of sensors. However, the process to recognize the high-level contents from sensing value is still not explored.

Besides, Mühlenbrock et al. (2004) utilized the personal computer, telephone, personal digital assistant (PDA), ambient sound, and user feedback to gain basic sensor data; furthermore, they defined activity, availability and inputs, and then looked for higher-level information like people’s activity and availability at the moment by using Bayesian approach. They tried to develop applications for the support of workplace interactions. Furthermore, an unsupervised recognition technique that helps users to label anonymous activity episodes by displaying contextual information collected by the sensors at home was surveyed (Wilson et al., 2005).

Considering the related works given above, this research believes that surrounding sensor measures are raw data and that an approach to recognize high-level contents from the basic measurement of sensors will be a very significant mechanism. It needs to combine various factors related well with the surroundings. In other words, an inferential method is necessary for deciding a user’s uncertain situation. The uncertain nature is applicale to be solved by means of fuzzy logic. Accordingly, exploiting fuzzy logic to reason context-awareness is a feasible scheme. Fuzzy theory is originated from fuzzy sets brought out by L.A. Zadeh at UC Berkeley in 1965. The process model with fuzzy sets can quantify abstract real-world meanings and fit human thought and behavior (Ross, 1997). As we know, the fuzzy logic applications have been successful in certain technological products such as the controllers for air-conditioners, washing machines, and so forth.

The previous works indicate that fuzzy logic is good at dealing with the handoff in mobile communication and also useful for the uncertain location. A fuzzy logic-based scheme for selection of base station at the time of handoff is presented. The scheme considers three criteria to make a fuzzy handoff decision regarding handoff to any particular base station. The fuzzy logic is also useful for a location tracking strategy based on spatial reasoning decision support (Dang et al., 2000; Jamadagni, 2000). Mohanty and Bhasker (2005) take into account the multiple attributes of the product, and analyze them with respect to the consumer’s desire by a fuzzy approach. Finally they classified these products into different levels of preference. Similarly, a fuzzy approach will be a good solution for our research. Our research uses fuzzy logic to know flexible high-level contents from homologous raw data. We focus on the important research issue about the activity recognition that has emerged in a mobile environment. The activity recognition is meant to determine what activity a person engages in. We introduce the fuzzy logic approach for modeling these user states, and we present an overall mechanism for achieving the activity recognition with limited sensing data.

3. DESCRIPTION OF THE SERVICE FRAMEWORK

This research is built on a context-aware framework. A context server is set up with a base station controller (BSC) to deal with context-aware services. A mobile node (MN) acts under some cell belonged to a base station and may stride across different cells. The sources of information about location can derive from MN’s located cell or from Global Positioning System (GPS) that makes a mobile device have its own geometric location, if necessary.

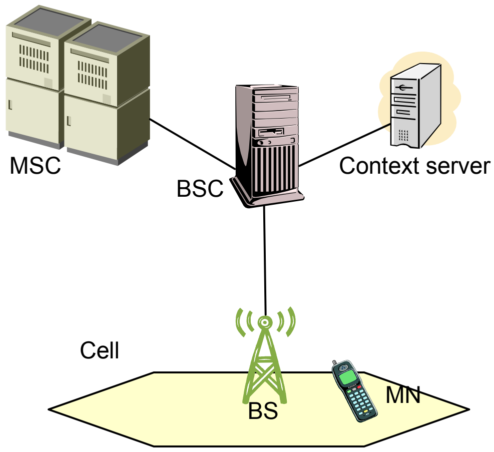


Figure 1. Overview of service framework

A context server collects data of basic context types, and then transfers these values of the context type data into corresponding membership degrees. Afterward, the context server engaged in recognizing and handling for providing suitable services. This research brings out a context-aware framework shown as Figure 1. For recognizing users' activity in a mobile environment, it is necessary to define the input attributes. A user's condition such as movement, background and so on could be reasoned via the following attributes: background sound, positioning signal level, user speed, working state, and use frequency of device. Moreover, if adopting a user feedback mechanism, it can rebuild and adjust the membership degrees as well as the concrete rule base. Hence, the context server can build a more accurate model through continuous training and adjustment.

In this paper, we propose a mechanism of recognizing context-aware information from the basic sensors data. The research tries to show the possibility to recognize a user's situation by utilizing the fussy logic theorem. Some measurable data related to the attributes of user's situation are supposed to be collected by some basic sensor. Through such mechanism, mobile service provider can recognize a user's scenario and provide proper context-aware applications. For the succinctness of description and discussion in this paper, we will use the following notation in Table 1.

Table 1. Description of the notation

Notation	Description
u_1	Background sound
u_2	Positioning signal level
u_3	User speed
u_4	Working state
u_5	Use frequency of device
v_1	Low membership degree
v_2	Moderate membership degree
v_3	High membership degree
W_1	Location weight sets
W_2	Movement weight sets
W_3	User task weight sets

W_4	User situation weight sets

4. PROPOSED SCHEME PROCESS

This paper adopts processes from basic context types to membership functions and fuzzy sets, and also from the quantitative method of background values in the real world to the qualitative method of complex messages, and finally determines the information contents to apply to mobile services.

4.1. Basic Context Types

By means of sensors laid out in mobile communication systems and personal mobile devices, we are able to know many basic context types. The inputs of our mechanism are these values of basic context types, including background sound, positioning signal level, user speed, working state, and use frequency of device. We use a set U to indicate this:

$$U = \{u_1, u_2, u_3, u_4, u_5\} \tag{4.1}$$

These basic context types are very important for our scheme. As utilizing the above basic type data, we shall recognize the high-level context types subsequently.

4.2. Membership Functions

Each input of basic context type has its corresponding membership function. A membership function is a curve that defines how each point in the input space is mapped to a degree of membership between 0 and 1. We are concerned more about the relation between each membership degree, not the pure number of membership degree.

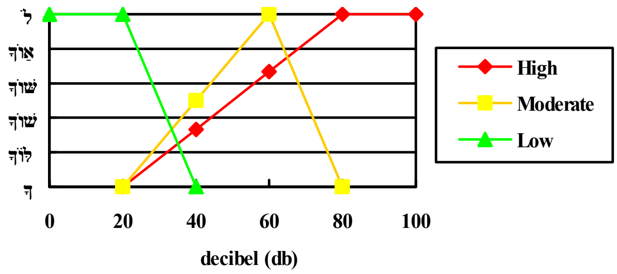


Figure 2. Background sound membership function $\mu_1(x)$

With the step of membership function, there is standardization which avoids the effect of the extreme large or small value. For instance, background sound membership function $\mu_1(x)$ shown in Figure 2. The boundary of membership degree in fuzzy logic is [0, 1]. We define high, moderate and low membership degrees as v_1, v_2 and v_3 in light of every membership function. They form fuzzy set V:

$$V = \{v_1, v_2, v_3\} \tag{4.2}$$

Low membership degree v_1 represents the weak tendency toward the basic context type, and moderate membership degree v_2 stands for the fair tendency toward the basic context type, and high membership degree v_3 stands for having strong tendency.

4.3. High-level Context Types and Weights

The high-level context types are drawn from basic context data. These high-level context types are the outputs after the process of our mechanism. For high-level context type, four categories of characteristics about a user's situation are classified. For example, the second one is movement, including still, walk, fast walk, run, vehicle and fast vehicle. About these types of outputs, each category has a weight set. In the weight set, the first one is x_1 that indicates the related level, that is, the relationship near or far for u_1, and the second one is x_2 that indicates the related level for u_2 (i.e., the relationship near or far for u_2). The rest are deduced by analogy, such as x_3 for u_3, and x_4 for u_4. That is, weights are decided according to the levels of relationship between set U and high-level context types. Thus we define the general formula W_i:

$$W_i = \{x_1, x_2, x_3, x_4, x_5\} \quad \text{where } i = 1, 2, 3, 4 \text{ and } x = a, b, c, d \tag{4.3}$$

Also, the specific formulas for each high-level context type are shown as follows:

$$W_1 = \{a_1, a_2, a_3, a_4, a_5\} \tag{4.3.1}$$
$$W_2 = \{b_1, b_2, b_3, b_4, b_5\} \tag{4.3.2}$$
$$W_3 = \{c_1, c_2, c_3, c_4, c_5\} \tag{4.3.3}$$
$$W_4 = \{d_1, d_2, d_3, d_4, d_5\} \tag{4.3.4}$$

4.4. Fuzzy Logical Operation

By the definitions of formula (4.1) and (4.2), the set U and set V are now attained. In regard to the following operation step, the matrix R is listed for the step:

$$R = \begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \end{bmatrix} \end{array} \tag{4.4}$$

The foregoing r_{kj} indicates a value which some basic context element u_k (e.g., u_1) corresponds with some v_j (e.g., moderate membership degree v_2) in the membership function. There is an operation called composition on fuzzy relations. The composition can combine two or more sets that are related to each other. It operates with a given fact to produce an output that represents the decision in a fuzzy way. The formula is as (4.5).

$$o_{ij} = \bigvee_{k=1}^{n} \left(x_{ik} \wedge r_{kj} \right) \tag{4.5}$$

4.5. Outputs

Therefore, when we would like to know some characteristic like user situation, the three formulas (4.3), (4.4) and (4.5) can be used. Let the set W_i and matrix R have a composition, and then an output set is produced:

$$\begin{aligned} O_i &= W_i \circ R \\ &= \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix} \circ \begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \end{bmatrix} \end{array} \\ &= \begin{bmatrix} o_{11} & o_{12} & o_{13} \end{bmatrix} \end{aligned} \tag{4.6}$$

Through the mechanism we can easily get the clusters. The value o_{ij} has something do with the conditional tendency. So if some mobile nodes have the similar o_{ij}, these device holders may engage in the similar activity or in the same situation. We will go into more details on the strength of those examples in Section 5.

By the mechanism we even can obtain the classification. To conclude the classification, a rule base shown in Table 2 is necessary and then we get a classification result about that characteristic. The value o_{ij} has something do with the composition of inputs. The classification can be realized since they conform to some tendencies. In the following, we build a prototype of rule base. As applying a system implementation, the adjustment of classification could be emended by the user feedback mechanism.

Table 2. Rule base

Types	Logical expression	Location (O_1)	Movement (O_2)	User task (O_3)	User situation (O_4)
I	$o_{11}\geq o_{12} \wedge o_{12}\geq o_{13}$	deep indoor	still	deep rest	very fine
II	$o_{11}\geq o_{13} \wedge o_{13}>o_{12}$	indoor	walk	rest	fine
III	$o_{12}>o_{11} \wedge o_{11}>o_{13}$	half-indoor	fast walk	available	ordinary
IV	$o_{12}>o_{13} \wedge o_{13}>o_{11}$	half-outdoor	run	active	complex
V	$o_{13}>o_{11} \wedge o_{11}>o_{12}$	open	vehicle	busy	disorder
VI	$o_{13}>o_{12} \wedge o_{12}>o_{11}$	very open	fast vehicle	extremely busy	extremely disorder

The o_{ij} value is related with the composition of inputs. For example, suppose $W_2 = \{0.2, 0.5, 0.8, 0, 0\}$ suitable to be used for a crowd. When O_2 conforms to a logical expression: $o_{13}>o_{12} \wedge o_{12}>o_{11}$, it means that the background sound, user speed, and use frequency of device have a strong level tendency. Thus the user might take a vehicle at high speed. If we know all weight set W_i and R, the full result $O = \{O_1, O_2, O_3, O_4\}$ can be resolved.

4.6. Training

Assume that there are training data; we can get excellent weights as relying on the weight decision approach. Considering the weight decision method, we introduce the approaching degree which is used for computing similarity, and its meaning is defined as (4.7).

$$(A,B) = 1/2[(A \bullet B) + (\overline{A \oplus B})] \tag{4.7}$$

For formula (4.4), A and B are two fuzzy sets, $\bullet$ is inner product, and $\oplus$ is outer product. If some output $O_i = [o_{11}, o_{12}, o_{13}]$ and matrix R are known, we can draft n kinds of possible weight sets $(W_i)_{t,\ t=1\sim n}$. Each of them respectively does fuzzy composition with R, and we obtain $(O_i)_t$, and then we compute the approaching degree between $(O_i)_t$ and O_i. The approaching degree 1 is most similar; on the other hand, degree 0 is most dissimilar. The best weight set candidate (a weight set with highest approaching degree) should be selected.

5. EVALUATION

For easily understandability of the meaning of this scheme, the numerical examples that we apply approach to our research are given. Meanwhile, the time complexity and performance are available too. This research observed the people's work schedule and use frequencies of device in the light of thirty students of the department of management at National Central University. The background sound and positioning signal level referred to conditions in the campus and its neighborhood. Also, the possible speed range of going on foot and vehicle is measured.

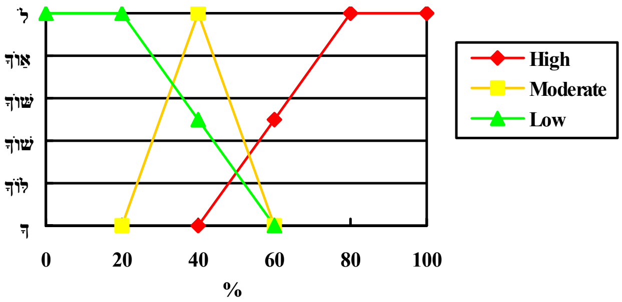


Figure 3. Positioning signal level membership function $\mu_2(x)$

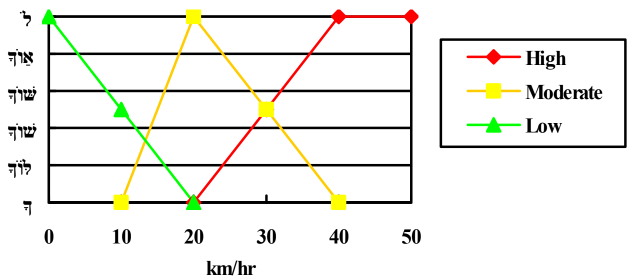


Figure 4. User speed membership function $\mu_3(x)$

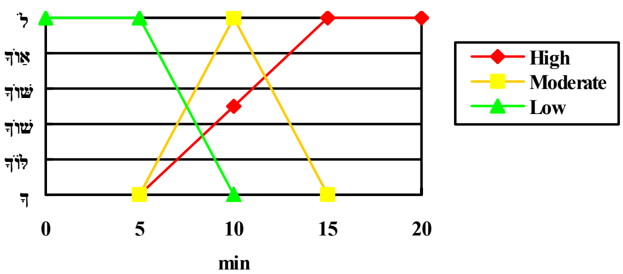


Figure 5. Working state membership function $\mu_4(x)$

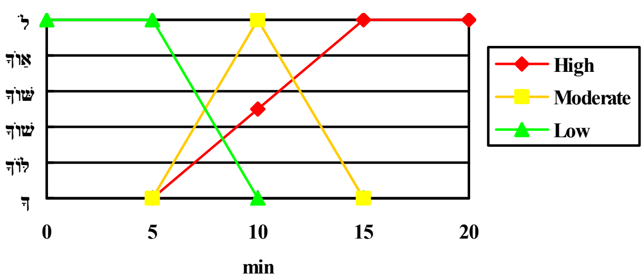


Figure 6. Use frequency of device membership function $\mu_5(x)$

The membership functions from $\mu_1(x)$ to $\mu_5(x)$ accordingly are designed and it is shown in Figure 2 to Figure 6. Please note that the use frequency of device indicates the minutes in the latest one hour. This membership function can be brought forth in the implementation of realistic data through the approach of Cano and Nava (2002).

Table 3. Numerical examples

	u_1 (db)	u_2 (%)	u_3 (km/hr)	u_4 (o'clock)	u_5 (min)
MN1	80	40	16	16	15
MN2	40	40	10	12	10
MN3	80	70	36	20	5
MN4	20	20	0	22	0
MN5	50	60	5	8	0

5.1. Numerical examples for classification

For the weights according the characteristic classification, for instance, we set up weight sets as follows:

- Background weight sets $W_1 = \{0.1, 0.8, 0.1, 0, 0\}$
- Movement weight sets $W_2 = \{0.2, 0, 0.8, 0, 0.5\}$
- User task weight sets $W_3 = \{0.2, 0, 0.1, 0.8, 0.3\}$
- User situation weight sets $W_4 = \{0.7, 0, 0.3, 0.1, 0.2\}$

Now if we would like to know the background around a user, so we work out the composition of W_1 and R, i.e.O_1. The rest can be deduced by analogy, and the final result is $O = \{O_1, O_2, O_3, O_4\}$. Here, for instances, there are complete data of five mobile nodes given in Table 3.

We take O_1 of the mobile node MN1 for example. Its u_1 is 80 decibel, so membership degree v_1 is 0, v_2 is 0, and v_3 is 1 as corresponding to the membership function of background sound. The rest can be known in the same reason. We now have the R from those five membership functions, and then compute it by formula (4.6). List this operation as the follows:

$$\begin{aligned}
&\text{MN1's } O_1 = W_1 \circ R \\
&= \begin{bmatrix} 0.1 & 0.8 & 0.1 & 0 & 0 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 & 1 \\ 0.5 & 1 & 0 \\ 0.2 & 0.6 & 0 \\ 0 & 0.5 & 1 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0.5 & 0.8 & 0.1 \end{bmatrix}
\end{aligned}$$

So we have the output that o_{11} is 0.5, o_{12} is 0.8, and o_{13} is 0. The logical expression of user's background in this example is $o_{12}>o_{11} \wedge o_{11}>o_{13}$. The result after comparing with the rule base is the "half-indoor". Similarly, the results of O_2, O_3 and O_4 are run/slow vehicle, extremely busy, and extremely disorder respectively. So the overall result is as follows.

O = {half-indoor, run/slow vehicle, extremely busy, extremely disorder} = {III, IV, VI, VI}

The result indicates that the hand-held device user runs half-indoors and has a disorder state, and for the student crowd they could attend class of physical education. If a student is still in an indoor and quiet place, he may read a book at the moment.

We just illustrate the example of a single mobile node, and now we talk about the multi-nodes example. If a mobile service provider would like to know which one has a disorder tendency so that the entertainment message may not be suitable at present, it can be resolved. So we are now concerned about the user situation (O_4). After the process of the membership functions and formula (4.6) in our mechanism, the results are gained and we take MN3 for instance.

$$\begin{aligned} MN3's\, O_4 &= W_4 \circ R \\ &= \begin{bmatrix} 0.7 & 0 & 0.3 & 0.1 & 0.2 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0.75 \\ 0 & 0.2 & 0.8 \\ 0.5 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0.2 & 0.2 & 0.7 \end{bmatrix} \end{aligned}$$

The all results about five mobile nodes are as follows:

MN1's O_4 = [0.2, 0.3, 0.7].

MN2's O_4 = [0.3, 0.5, 0.33].

MN3's O_4 = [0.2, 0.2, 0.7].

MN4's O_4 = [0.7, 0, 0].

MN5's O_4 = [0.3, 0.7, 0.5].

Because MN1 and MN3 have the lower o_{11} value and higher o_{13} value, they should be in a similar situation. In addition, the lower o_{11} value and higher o_{13} value imply the disorder tendency, so the MN1 and MN3 possibly are those ones that should not receive the entertainment advertising messages in the present. We also can tell that MN4 may be at a fine situation without disturbance.

5.2. Illustration for weight decision

In Section 4.3, I have introduced the weight decision method. When we some training data, we can get the weight which is more reliable. If we obtain that a matrix R and the movement (O_2) was by vehicle, for example, O_2 is defined as [0.2, 0.2, 0.8] that conforms to the logical expression $o_{13}>o_{11} \wedge o_{11}>o_{12}$ obviously. Here this R is as follows.

$$R = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0.75 \\ 0 & 0.2 & 0.8 \\ 0.5 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Suppose that there are some weight alternatives for W_2: {0.2, 0, 0.8, 0, 0.5}, {0, 0.1, 0.2, 0.1, 0.3}, and {0.3, 0, 0.5, 0.4, 0.6}. Each of them does fuzzy composition with known R, and we obtain [0.5, 0.2, 0.8], [0.3, 0.2, 0.2], and [0.6, 0.2, 0.5]. It can be observed that [0.3, 0.2, 0.2] does not conform to the logical expression o13>o11 o11>o12, so {0, 0.1, 0.2, 0.1, 0.3} should not be considered. By formula (4.4), we compute the approaching degree between $(O_i)_t$ and O_i. These respectively are 0.8 and 0.65. The degree 0.8 is larger than 0.65, so the weight set candidate {0.2, 0, 0.8, 0, 0.5} should be selected.

5.3. Efficiency for computation

Although using matrix multiplication, the load for our scheme is reasonable because its time complexity is O(n). If a system of the mobile service provider prove n MNs, For processing each MN, we multiply 1*i matrix and i*j matrix where i = 5 and j = 3 since there are five weights and three membership degrees (low, moderate and high). The time complexity of matrix multiplication for each MN is i*j which is 15, a constant. Thus for n separate MNs, the time complexity of matrix multiplication is O(n).

5.4. Characteristics and applications

The research by M. Mühlenbrock et al. (2004) is to decide whether someone is available or not. In contrast with their research, our mechanism discovers more facts about users' behavior. This proposed mechanism in our work can be applied to some mobile commerce applications, like mobile advertising. The mobile advertising can adopt the information which is extended from our mechanism in both the pull marketing, i.e. user-driven and the push marketing, i.e. not user-driven (Aalto et al., 2004). To employ the information derived from our mechanism, users' needs are probably satisfied more and rightly for some services. The activity recognition has emerged in the mobile context-awareness, and helps those systems to give the proper services. Let us say that it is improper to give a user the gas station message if he is walking indoors; on the other hand, it may be appropriate to provide latest news.

We just show the examples of a single mobile node and of multiple nodes. It is known that one belongs to a certain classification, and it can be told that two or more users are in some similar situation. We are able to add or remove the output items in light of requirement. The merit in this mechanism is high flexibility. It can be adjusted the context-aware input/output as well as the sensor and process according the actual application. Also, all the basic context types as inputs in this mechanism come to standardization by means of the membership function. Each input of basic context type has its corresponding membership function. The range of membership degree in fuzzy logic is between 0 and 1 so that the effect of extreme large or small values has been avoided. Another question is about the erection of weight set values. In fact, the true weights can be obtained as relying on the weight decision approach if there are training data. We have introduced the approaching degree used for computing similarity in the previous section.

Moreover, by using our scheme, the context server will soon be able to collect real-world context data from multiple users. Analyzing these collected data will be helpful for developing a more reliable higher-level context recognition mechanism which is capable of delivering more accurate context-aware information to applications. With the training data and weight decision approach, the trustworthy weight sets can be decided for implementation. Since fuzzy logic does not require precise, noise-free inputs, the fuzzy logic approach offers unique features that make it a particularly good performance for the chaotic decision problems like the activity recognition. Hence the applications that use our proposed mechanism will have high feasibility and flexibility.

6. CONCLUSIOINS

Our proposed scheme would discover the data of mobile device users and their surroundings. We develop the signal processing techniques and classification mechanism. So we make it possible that the advanced context-aware application could be provided to users according to their situations. The fuzzy logical process model is a good tool for such application. It collects the basic data by specific sensors and reason users' behaviors and situations out.

Accurate recognition of human activities is an important requirement in mobile commerce. Such technology helps service providers to determine a potential user's activity and provide their proper context-aware services to the user. The toughest challenge about the context-aware is how to compute and express the complex and fickle users' state in the environment. Our contribution is to provide a technical mechanism for the realization of the context awareness. Moreover, its low operation cost is confirmed through a complexity analysis. It can be used to recognize the basic information so that the characteristic identical with a user's situation is inferred. As long as this mechanism is used well and other related technologies develop maturely, many amazing dream may come true. Not only people receive the instant promotion message when passing through a department store, but also the interesting digital content as viewing works of art in the gallery.

This research links up an important part in context-aware services. The inputs are under standardization and the outputs reach classification in the proposed mechanism. Hence, many applications that satisfy users' immediate need could depend on the concept revealed in our research in the near future. To make the context-aware service more acceptable, some urgent future works are observed during our research. In addition to more precise context detection skills, privacy protection is becoming an important issue. Personal and confidential data like the location of a user might be sensitive and have profound implications for individual private information. We must use such data and also protect it from revealing privacy. It is the obvious critical point we should focus our effort on to realize m-services.

7. REFERENCES

Aalto, L. et al. (2004) Bluetooth and WAP Push Based Location-Aware Mobile Advertising System, *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services*, pp. 49–58

Cano, J.C. and Nava, P.A. (2002) A Fuzzy Method for Automatic Generation of Membership Function Using Fuzzy Relations from Training Examples, *Proceedings of the 2002 Annual Meeting of the North American Fuzzy Information Processing Society*, pp. 158–162

Chalmers, D., Dulay, N. and Sloman, M. (2004) Meta Data to Support Context Aware Mobile Applications, *Proceedings of the 2004 IEEE International Conference on Mobile Data Management*, pp. 199–210

Dang, M.S. et al. (2000) Fuzzy Logic Based Handoff in Wireless Networks, *Proceedings of IEEE 51st Vehicular Technology Conference*, vol. 3, pp. 2375–2379

Kindberg, T. and Fox, A. (2002) System Software for Ubiquitous Computing, *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 70–81

Korpipaa, P. et al. (2003) Managing Context Information in Mobile Devices, *IEEE Pervasive Computing* , vol. 2, no. 3, pp. 42–51

Lee, D.L. et al. (2002) Data Management in Location-Dependent Information Services, *IEEE Pervasive Computing*, vol. 1, no. 3, pp. 65–72

Mohanty, B.K. and Bhasker, B. (2005) Product Classification in the Internet Business—A Fuzzy Approach, *Decision Support Systems*, vol. 38, no. 4, pp. 611–619

Mühlenbrock, M. et al., (2004) Learning to Detect User Activity and Availability from a Variety of Sensor Data, *Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications*, pp. 13–22

Ross, T.J. (1997) *Fuzzy Logic with Engineering Applications*, McGraw-Hill

Satish Jamadagni, N.S. (2000) Dealing with Location Uncertainty in Mobile Networks Using Contextual Fuzzy Cognitive Maps as Spatial Decision Support Systems, *Proceedings of IEEE 52nd Vehicular Technology Conference*, vol. 3, pp. 1489–1492

Wilson, D.H., Long, A.C. and Atkeson, C. (2005) A Context-Aware Recognition Survey for Data Collection Using Ubiquitous Sensors in the Home, *CHI Extended Abstracts on Human Factors in Computing Systems*, pp. 1865–1868

Prioritizing and Partitioning of Business Requirements for Incremental Web Development

Diana Goerge, Khaled M. Khan
School of Computing and Information Technology
University of Western Sydney
Locked Bag 1797 Penrith South DC NSW 1797

Australia

k.khan@uws.edu.au

Abstract

This paper proposes a conceptual framework for web application development. The framework is based on the concept of prioritizing and partitioning requirements at the systems analysis phase. The theory of the proposed method is centered on partitioning requirements based on dependency and importance. We have applied our approach in an example to illustrate the applicability of the approach.

Keywords:
Web development, requirements, prioritizing, partitioning web development, iterative development.

1. INTRODUCTION

There is a need for new development methods for web applications. Existing development methods used for non-web application development cannot address the unique characteristics of web based systems. Web application developers have taken two approaches to addressing this dilemma --they either battle with existing traditional methodologies, or upon having these methodologies fail turn to ad hoc development. As with any area of application development, if the methodology does not suit the development environment then serious shortcomings in project outcomes are inevitable. There is a serious variance of how development is undertaken and how it should be undertaken to get the full potential of the project. This paper proposes a method for partitioning requirements for web development. The aim of this paper is twofold: (1) to present a method to assist in web based application development; and (2) to illustrate the applicability of the proposed method in an example.

Some may argue that "traditional" applications and web applications are no different and so we can use the same development strategies for both types of applications. Others consider web applications so different that traditional development methodologies are irrelevant and simply cobble web applications together. However, some research such as published in (Offutt, 2002, Reifer, 2000, Reifer 2002) indicate that there are certainly differences between the web based systems and non-web based systems in terms of their development characteristics. One reason, suggested by Issakowitz (1998), is that the web requires new approaches to development because it has the potential of reaching a wider audience.

It seems that many practitioners are gaining understanding and acquiring knowledge in the new dimension of web application development and beginning to recognise that there is a distinction between traditional development and web application development. "Web based application characteristics appear to challenge the appropriateness and relevance of methods and techniques followed for conventional systems." (Poulymenakou, 1999).

Capturing the requirements for web applications development is somewhat different to capturing requirements for traditional applications. There are different issues. The requirements may not be readily discovered, due to the underlying business processes being new. The requirements must then be created rather then discovered (Gordijn, 2000). In traditional system development, the approach has been to develop all of the requirements, then specify and construct the system and then bring the system into use – a single iteration of a development life cycle. Iterative development, in which increasingly more functional applications are developed, as described in some existing methodologies such as the Unified Process described by Jacobson (1999), is necessary in web development.
There is a need to complete web applications faster because of the demand for applications. They must be initiated, planned and developed in record time. Also, because of the changing technology, what is relevant today may not be useful in six months time. Therefore we need a method that allows a developer to keep up with the evolution and the speed of the changes.

The concept of lightweight and adaptive methodologies is that they are fundamentally different to traditional methodologies as they are less structured. Yet they still provide guidance and boundaries to the development, without the need for heavy documentation at every activity. Fowler (2001) describes lightweight methods as being more adaptive rather than predictive, whereas heavy methods try to plan the whole process step by step in great detail for a longer time period. As a result, light methods embrace change, as they adapt to the dynamics of the project. Lightweight and adaptive methodologies are most relevant to web application development.

In this paper we attempt to prioritise and partition the business requirements of the system for web based application development. The paper is organised as follows. Section 2 presents the proposed method and its relationship with the systems development life cycle. The section overviews the method, its major characteristics, and mechanics of the method. We apply our approach in an example to show its applicability in section 3. Section 4 concludes the paper.

2. PROPOSED METHOD

Based on current literature, observation and our own experience, the following are three fundamental issues that our work focuses on:

- *Evolutionary web applications development* - development of web applications should embrace an incremental approach where one function (or a small set of functions) is completed at one time, rather than having the complete set of functions developed in one step.
- *Dynamic content, functionality and structure* - web applications should be dynamic because of the constant changing of requirements; systems need to be flexible in a way that allows quick changes to the requirements of the system.
- *Rapid web application development* - web applications must be developed at a faster speed by breaking up complex problems into smaller, more manageable partitions.

Based on the above characteristics, we propose a technique, using the requirements for a web based application, to partition the requirements. Each partition is then developed in turn, providing a development approach that incrementally delivers an operational system, albeit with minimal functionality, earlier than by using traditional development techniques. The proposed technique for development of web applications uses the requirements of the application as its starting point. The requirements are prioritised and then development of the system follows the prioritised requirements step by step. This is described in detail in the following sections.

2.1 Overview

The analysis phase of a web application development process, applying our partitioning technique, has two definitive steps which are outlined as follows:

- *Prioritise requirements* – identified requirements are examined and prioritised
- *Create and prioritise sub domains* – from the prioritised requirements, the requirements are grouped into sub domains. Each sub domain forms a group of requirements that are developed as a separate iteration.

These processes are seen in Figure 1. The figure shows where this approach fits into the analysis of the development life cycle. As it shows, the method allows a feedback mechanism.

As the process is evolutionary and based on incremental development, we need to be able to differentiate the types of iterations that occur when using this method. There are two different types of iterations that have been identified, namely as:

- *Core iteration(s):* The core iteration is the very first and most complex iteration as it is the initiation of the project. This iteration is the fundamental structure of the entire web application. Due to its complexity, the core iteration(s) may require more than one iteration to complete.
- *Satellite iterations*: These are the subsequent iterations, which occur after the core has been developed. These iterations can be seen as cycles where extra functional components are created and plugged into existing components that have been created in previous iterations.

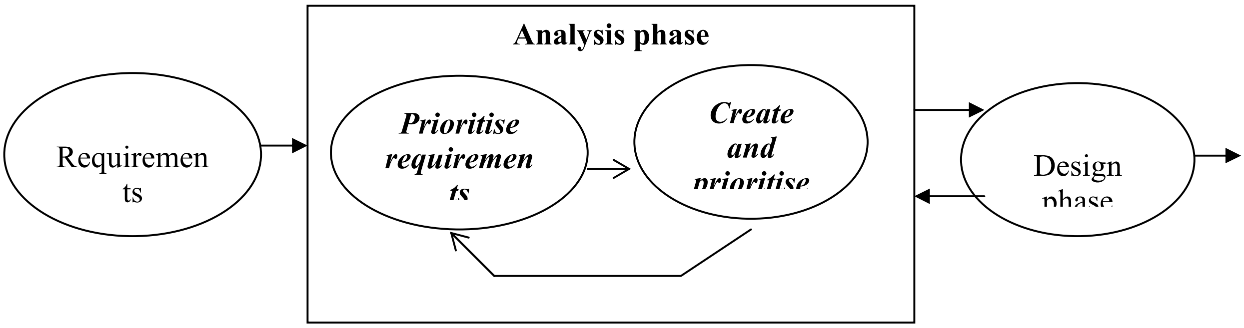


Figure 1: The proposed portioning method within the system development life cycle

2.1.1 Prioritise requirements

The focus of this paper is not to describe how requirements should be captured. A developer can use any of the appropriate tools or techniques to capture the requirements for the system application. Once a collection of system requirements has been drawn up, the requirements are examined in an effort to prioritise them. This is done on the basis of two attributes of each requirements: *importance* and *dependency*. In an effort to show this, the following example is used throughout the remainder of this section: Let's assume we have found a collection of nine requirements in no particular order from the "requirements elicitation" phase. We now need to prioritise these requirements so that we end up with a list of ranked requirements, such as R1 to R9 as shown in Figure 2. Nonetheless, how do we initially go about prioritising the requirements when they are currently without order?

Requirement Priority	Requirement Name
R1	Product Database
R2	Data Layer Functions
R3	Order Tracking
R4	Product Listing
R5	Sales Procurement
R6	Stock Availability
R7	Stock Alerts
R8	Sales History Listing
R9	Secure Login

Figure 2: Prioritised Requirements – An example

'Important' requirements are those that are the fundamental business objectives of the system and are normally non-technical in nature. However, the 'dependency' relationships are much harder to depict due to the fact that all requirements must be examined together rather then individually as with the important requirements. These are the requirements which cannot function without other requirements first being developed and include data, architectural and security issues as examples. A further example of dependency would be an area where we may have a secure environment, the user is granted access to the application via a secure login. Therefore other requirements may be dependant on this requirement being fully developed, as there must be a security check done before the user can log in and utilise any other modules that have been developed.

2.1.2 Create and prioritise sub domains

In this step, we take the system requirements and group the requirements into *sub domains*. We group the requirements so that the system can evolve over time. Unlike prioritising user requirements, which is quite separate, we create and prioritise sub domains in one step rather than two. The first sub domain that requires our attention is the core iteration, the first iteration for development in the system. It has the highest importance, is the most time consuming and is the foundation and the core of the development project. When considering the core iteration we must consider the whole project rather than separate components of it. We have a strong focus on the overall system and its components. We commence with the list of requirements obtained from the prioritise requirements step. Our objective is to assign the requirements into groups called sub domains.

We take the prioritised system requirements and group the requirements into what is called a sub domain. Here we are creating groups of requirements and working out what and when they need to be implemented as one unit. Since this method is based on evolution, there will be many iterations of the proposed method. The idea is to group the requirements so that the system can evolve over time. This step allows us to even further refine the requirements, because we are trying to maximise the output. If we implemented the requirements list one by one it would take a greater amount of time because we are not gaining in areas such as reducing developer overhead. Our purpose is to improve the requirements in such a fashion that we can gain maximum performance. In the step, it is how we partition the development by constructing the core of the system and then once this is completed, incrementally disclosing the rest. By first developing the core of the system, we are provided with a way of plugging components into a web application. 'Plugging in' is achieved by a team of designers and developers in an order dictated by the requirements rankings.

The 'Prioritise requirements' step may not greatly differ from the traditional system development lifecycle. Though, where our method is different is in the 'Create and prioritise sub domains' step.

2.2 Priority by dependency

Continuing with our example we identify relationships of our nine requirements that have now been prioritised in the order of R1 to R9. From this prioritised list of requirements, we need to determine which requirements have an interrelationship with other requirements within the list. This is an important step, as it will affect the way in which we group the sub domains. Let us assume that R1 is our important requirement from the previous step. In this example we also assume that R1 is a data requirement and for every other requirement, R2 to R9 is dependant on that data in some kind of way. The following shows that every requirement from R2 to R9 has a dependency relationship with R1 as depicted: R1 → R2, R3…R9

Normally, when we prioritise requirements, if a requirement exists that is dependant on another, then the latter has higher priority. However, during the 'importance' ranking process, as requirements are ranked on '*importance*' first and then on '*dependency*', it may have a higher priority then its dependence. If this occurs, we make a note of this issue by documenting the artifact, or we can change our priority ranking. The reason why we may want to change this, is as we gain a greater understanding of the requirements, we want to change the priorities. For example, let's take a closer look at R9, *Secure* Login, we discover it has many dependencies as shown:

R1 → R2, R3 …R9
R7 → R6, R8

In Table 1, the table on the right shows the original priority ranking based on 'importance' only. However, whilst looking at the dependencies within our list, we have discovered that *Secure Login* importance has increased. There are two reasons, namely there are many dependants on this requirement and though more prominently its importance was overlooked. Therefore we change its priority.

Requirement Priority	Requirement Name
R1	Product Database
R2	Data Layer Functions
R3	Order Tracking
R4	Product Listing
R5	Sales Procurement
R6	Stock Availability
R7	Stock Alerts
R8	Sales History Listing
R9	*Secure Login*

Priority based on importance only

Requirement Priority	Requirement Name
R1	Product Database
R2	Data Layer Functions
R3	*Secure Login*
R4	Order Tracking
R5	Product Listing
R6	Sales Procurement
R7	Stock Availability
R8	Stock Alerts
R9	Sales History Listing

Priority based on importance and dependence

Table 1: Changing *Secure Login* requirement based on importance and dependence

As another example of finding a dependant relationship, we can say that R6, *Sales Procurement* and R8, *Stock Alerts* require R7, *Stock Availability* before they can exist. R6, which was ranked on importance rather than dependency, must be developed after R7. Finally, we have found all the existing dependencies as shown below from our priority table based on importance.

R1 → R2, R3 …R9
R3 → R4, R5 …R9
R7 → R6, R8

The parent requirements are positioned left of the arrow and have relationships with other requirements. The requirements right of the arrow cannot exist without the parent requirement being developed. We have placed all the nine requirements and its dependant ant relationships in a table. We have now found all the dependant relationships that occur with the nine requirements, as shown in the Table 2.

Requirement Priority	Requirement Name	Dependence
R1	Product Database	
R2	Data Layer Functions	R1
R3	Secure Login	R1
R4	Order Tracking	R1, R3
R5	Product Listing	R1, R3
R6	Sales Procurement	R1, R3, R7
R7	Stock Availability	R1, R3
R8	Stock Alerts	R1, R3, R7
R9	Sales History Listing	R1, R3

Table 2: Online Ordering system with complete prioritised requirements

2.3 Iteration and partitioning requirements

Once we have the requirements and prioritised them, we will then need to create our sub domains. However, how we create them will also depend on the type of iteration being performed. We need to work out what iteration we are performing and then create the sub domains accordingly.

2.3.1 Core Iteration

The first sub domain, also known as the core iteration, requires more of our attention, as it is the first development iteration. It has highest importance, is generally the most time consuming, and the foundation of the development project. When considering the core iteration we must consider the whole project rather than separate components. The core iteration/first sub domain can be described as a central unit. Later into the development, the components will be plugged into the central unit, coming from the satellite iterations.

Our objective in this step is to assign the requirements obtained from 'Prioritise requirements' into groups called *sub domains*. We have already worked out which requirements have associations with other requirements. The method in which the previous step was completed, will affect the way in which we group the sub domains. We create our sub domains by looking at the core objectives of the business; secondly, the requirements that carry the business core objectives. From the 'Prioritise requirements' process, we know that R1, the *Product Database*, is the parent that must precede every other requirement that is dependant on R1. Therefore, R1 as a fundamental requirement we can safely assume that it is a part of our first sub domain, S1 as depicted in Figure 3.

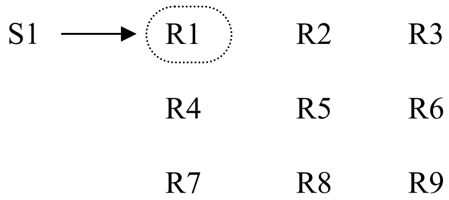


Figure 3: Preliminary sub domain

We know that nothing else can function without R1; therefore R1 belongs in S1, the core, because all the other requirements depend on the data. We now examine the other requirements to see what else needs belong to S1. R1 cannot be the only requirement in S1. It serves no meaningful purpose if it were to be implemented on its own. From our knowledge of R2, *Data Layer Functions*, we know that it must be implemented straight after R1. Therefore R1 and R2 are brought together into S1 in Figure 4.

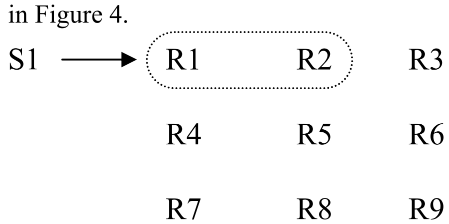


Figure 4: R2 may also be added

However, S1 is still not complete. Again we look at the remaining requirements and find what is needed to get the core application up and produce a functional deliverable. Lets say that R1 and R2 are based on data requirements; we will then need to access these data requirements. In our example, we discovered that R3 and R5 should be a part of S1. As all these requirements now make it possible to develop a useful application.

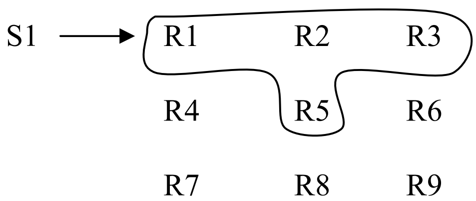


Figure 5: The first sub domain S1

Our first sub domain, S1, therefore consists of four requirements as shown in Figure 5, namely, R1, R2, R3, and R5. Therefore, our Online Ordering System example shows that we can have a simple system up and running with just four requirements. Requirements that have higher priorities, such as R4, may have to be deferred until the next iteration. Even though it is a lesser priority, R5 needs to be implemented before R4 as it helps make up the core requirement of the web application and therefore is suited to the first sub domain.

If the formation of our first sub domain, S1, leads to all nine requirements being within S1 then it may either be that we have not carefully thought out how we are going to break down the requirements into single functional units, or that this method or process may not be suitable to the system. This may be because the problem the developer is dealing with is a simplistic web application that may only need one iteration. It is important that the core is developed first by itself. Then other sub domains, as components will plug into the rest of the application through later iterations.

2.3.2 Satellite Iterations

The second and subsequent iterations, or 'satellite iterations', can only be commenced once the first or preceding sub domain is completed. The approach we take in developing within the satellite iterations is quite different. The subsequent sub domains are more likely to be small components, which act as enhancements to an overall central unit. In our example, we have assumed that only one iteration is required for the core of the web application. Once we have fully completed the first iteration our next step is to work on the proceeding sub domain, namely S2.
S2 shown in Figure 6 is created by; taking the next set of highest ranked requirements in importance and satisfying our dependencies, and placing them into the sub domain. Firstly we take R4, *Order Tracking* and put it into S2. R4 is dependant on R1 and R3, but these dependences were completed in S1, therefore affecting R4. Our next requirement on the list is R6, *Sales Procurement.* It has a third dependency R7. If we would like to develop these requirements we will also need to include R7, *Stock Availability*.

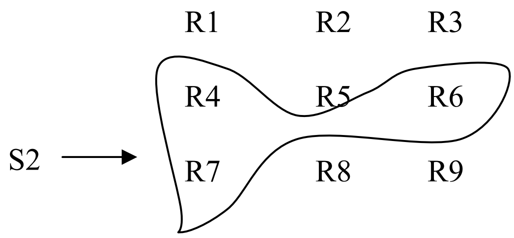


Figure 6: S2, part of the second iteration

Thus our second sub domain, S2 consists of R4, R6 and R7 as shown in Figure 6. It is desirable to keep the sub domains small so that time from one delivery of the application to the next is kept to a minimum. It is also desirable to keep the sub domains evenly based in terms of the time it takes to implement the sub domain. This process is continued until all requirements are allocated to a sub domain. Our final sub domain, S3, will therefore consist of the remaining two requirements, R8 and R9 as shown in Figure 7.

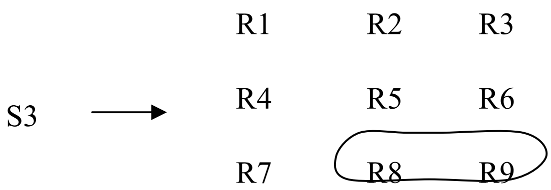


Figure 7: The third sub domain, S3

2.4 Revisiting requirements

We revisit the requirements each time a sub domain has been implemented. Although we go back to 'Create Requirements', we do not look at the requirements from scratch. Instead, we analyse what we have currently implemented, our web application and review the remaining requirements.
Static Requirements
If we are satisfied with the requirements and there are no changes, we keep the requirements as they are and move on to the remaining steps 'Create Sub Domains' and 'Prioritise Sub Domains'. We also analyse the sub domains that are yet to be implemented. Whether we make changes to the sub domains or not, we proceed with the implementation of the next iteration. Therefore if we assume that there are no new requirements after completing S1 then the developer can immediately proceed with S2.

Changing Requirements
If there are changes in the requirements, for example; let's assume that S1 has been implemented and developers now see a need to add an extra requirement. We revisit the 'Prioritise Requirements' step. The new requirement will then need to be prioritised within the remaining requirements, of R4, R6…R9. Let us assume that after implementing and rolling out

the first sub domain our development team revisited the requirements and found a need for a module to authorize the refund of sales from retailer before they send the product back to the warehouse. This new requirement, *Refund Sale*, would be prioritised according to the same criteria as we applied in the 'Prioritise requirements' step. In this case we find it to be ranked 10^{th} in our new set of requirements. We must now revisit the steps 'Create and prioritise sub domains' to reorganize our new set of requirements, incorporating the new item.

3. AN EXAMPLE

We are undertaking the following investigation in an attempt to evaluate the proposed partitioning method. In this section we explore and examine aspects of the method being used.

3.1 Problem Description

The example is set in the computing department within a university. It is a student project allocation system. The aim was to produce an online web based allocation system, which allocated students into appropriate teams and teams into projects with supervisors. All the business rules of the allocation system will not be discussed in this paper; simply a brief overview will be given in order to give the general idea of what the system is all about. The system has three types of users, they are: course coordinators, supervisors, and students. Each user has a different role within the system.

One of the reasons why this example was chosen as the focus of our investigative research was that the software development methodology being used is the generally agreed object-oriented methodology. The developers had been trained in object-oriented techniques and had no previous experience in web development; therefore it gave us a good ground to base our research upon. Thus, we could study the approach used by our subjects with a view to determining if any improvements could have been made during their development using the method. Following is the results of this method investigation.

3.2 *Application of the method*

This section presents a brief illustration of the proposed method. It is not however, a full investigation of the example, rather a summary which presents the ideas behind the method. The requirements that have been identified in the project. In this representation they are not in any particular order, however the next step will be to prioritise these requirements. Table 3 shows a list of ranked requirements from R1 to R15. The requirements have been taken from table 1 and have been ranked by importance and dependence. Importance has been ranked by business criteria, where a scale of high, medium and low has been used to display the importance of the requirements. The scale is not significant to this discussion as any scale can be used to rank the requirements.

Priority	Requirement	Importance	Dependence
R1	Manage Student Data Extract	High	R8
R2	Manage Student Details	High	R1
R3	Manage Project Details	High	R8
R4	Manage Client Details	High	R8
R5	Manage Supervisor Details	High	R8
R6	Generate Reports	High	R2..R5,R7,R10,R13
R7	Assign project to client	Medium	R3,R4,R8
R8	Login	Medium	
R9	Create Team	Medium	R2,R8
R10	Assign project to team	Medium	R7,R8,R9
R11	Join Team	Medium	R2,R8
R12	Manage Supervisor meetings	Medium	R5,R8
R13	Assign Supervisor Meeting	Medium	R8,R9,R12
R14	Exit Team	Low	R8,R9
R15	Assign Student to Team	Low	R2,R8,R9

Table 3: List of ranked requirements

The core iteration contains R1 through R5 and R8. These are the highest ranked requirements in importance and dependencies are satisfied. Satellite iterations 1 and 2 contain the medium importance requirements, again satisfying dependencies. The requirements are separated into two sub domains, as it is desirable to keep sub domains small so that time from one delivery of the application to the next is kept short. Satellite iteration 3 contains the two lowest ranked requirements. The sub domains are illustrated in Figure 8.

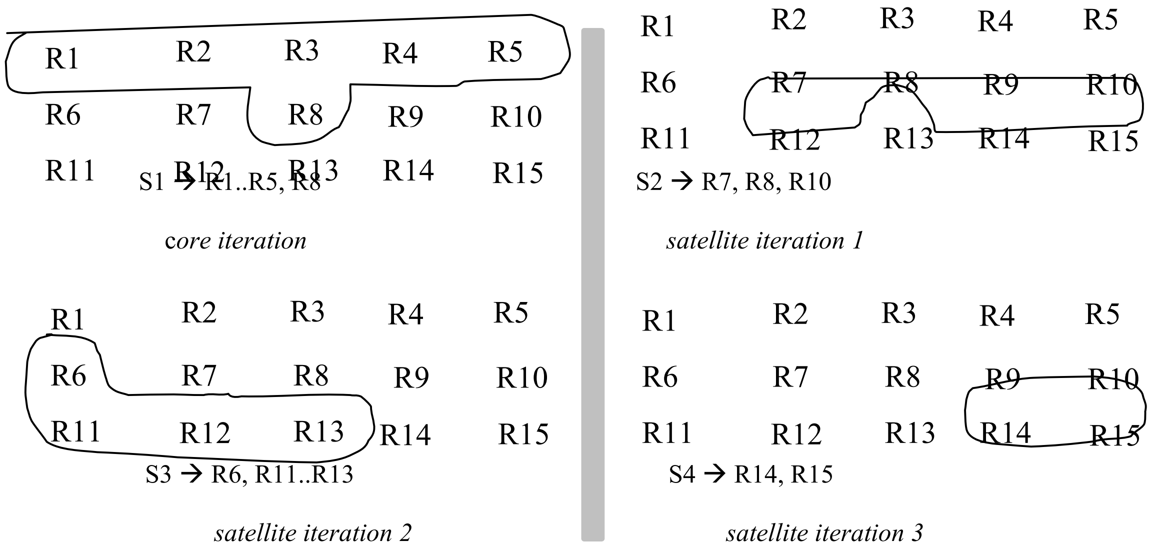


Figure 8: Sub domain in two iterations

Let us consider the dependence of three requirements and the significance of dependencies in partitioning the requirements.

- Manage Student Data Extract – This is the highest ranked requirement out of the fifteen requirements, the reason being that it is the requirement that extracts student data from an external source and hence permits other requirements to carry out their subsequent requirements.
- Generate Reports – This requirement has a high importance and therefore is ranked 6th. Yet, table 2 shows that there are requirements ranked lower then generate reports upon which R6 is dependant. Therefore, we have a conflict between importance and dependence, which would force the inclusion of R6 into a later satellite iteration. This indicates a need to review this requirement. It may suggest that the requirement requires decomposition or that the requirement will be delivered later then client expectations. In the former, the create and prioritise sub domains step must be repeated. In the latter the client must be informed and agreement reached.
- Assign project to team - This requirement has three dependencies, namely R7 to R9. We do not include 'Manage Project Details, R3 as a dependency, because R7 already includes R3 and we do not chain the dependencies.

4. CONCLUSIONS

In this paper we have examined and presented a method of partitioning business requirements for web based development. This partitioning method allows the developer to continue using development tools that are appropriate for their system and to integrate the partitioning method with their tools to improve the software development process for web applications. The strength of the partitioning method is that it welcomes change to the development process because it allows developers to break down the business requirements as short term goals and into controllable iterations. The nature of the process allows better control over planning which should lead to better estimation about development time and cost. The iteration development process acts as a feedback mechanism as the web application is developed incrementally and is being constantly reviewed. We acknolwledge that more research need to be done in this area to explore the possibility for the application of the method in a large system development environment.

REFERENCES

Fowler, M. (2001). http://www.martinfowler.com/articles/newMethodolgy.html [13/07/01]

Gordijn, J., Akkermans, H. & Vliet, H. (2000). Value Based Requirements Creation for Electronic Commerce Applications. *Proceedings of the 33rd Hawaii International Conference on System Sciences*.

Isakowitz, T., Bieber, M. & Vitali, F. (1998). Web Information Systems. *Communication of ACM*, 41(7), 78-80.

Jacobson, I., Booch, G. & Rumbaugh, J. (1999). The unified software development process. USA: Addison Wesley.

Larman, C. (1998). Applying UML and Patterns, An Introduction to Object-Oriented Analysis and Design. New Jersey: PH PTR.

Offutt, J. (2002). Quality Attributes of Web Software Applications, *IEEE Software*, 19 (2), 25-32.

Poulymenakou, A., Drakos, W., Papzafeiropoulou, A. & Doukidis, G. (1999). Towards New Web Application Development Practices. *Australian Journal of Information Systems*, 1, 107-113.

Reifer, D. (2000). Web Development: Estimating Quick-to-Market Software, *IEEE Software*, 17(6), 57-64.

Reifer D. (2002). Ten Deadly Ricks in Internet and Intranet Software Development. *IEEE Software*, 19 (2) 12-15.

Standing, C. (1999). Managing and Developing Internet Commerce Systems with ICDM. Proceedings of the *10th Australasian Conference on Information Systems*.

Standing, C. (2000). Internet Commerce Development. Norwood: Artech House.

Yourdon, E. (2000) Just Enough Software Requirements. http://www.yourdon.com/seminars/RM.HTML [21/02/01]

Data and Information Quality: an Information-theoretic Perspective

Wei Hu
Junkang Feng
University of Paisley, United Kingdom
Email: {wei.hu, junkang.feng} @paisley.ac.uk

Abstract

The evaluation of data quality (DQ) and information quality (IQ) has been treated as challenging issues in the field of information systems management for the last twenty years. However, we observe that the definitions of DQ and IQ in the literature are not necessarily convincing, which seems to have hampered the development of deep and sound understanding of the issue, and of practically applicable and effective measures and techniques for their evaluation. Through our review of relevant information systems literature, we believe that a rigorous and theoretically sound foundation is highly desirable to provide an insight into specifying and distinguishing the terms 'data quality' and 'information quality'. This paper presents a data-info quality model based upon theories of semantic information. We provide a set of definitions, compare data quality with information quality, and outline the objective and subjective aspects involved in addressing this problem. This model forms a basis for our further research in data and information quality assessment.

Keywords:

Data Quality, Information Quality, Theories of Semantic Information, S-B-R Framework, Semiotics

INTRODUCTION

Information systems (IS) play a key role in organizations for decision-making and efficient business flow for years. Issues regarding the evaluation of data quality (DQ) and information quality (IQ) have been noticed and identified increasingly within the field of information systems management in recent years. Numerous research efforts have been made in this area from different disciplines and using different research approaches for the purpose of developing data and information quality concepts and methods (Ballou&Pazer, 85; Burgess et al., 04; Dedeke, 00; English, 99; Eppler, 01; Hill, 04; Lee et al., 02; Liu & Chi, 02; Price & Shanks, 04; Redman, 01; Wand & Wang, 96; Wang & Strong, 96; etc). Hundreds of tools have been produced for evaluating quality in practice since 1996 (English, 99).

Research and practice indicates that data or information quality should be defined accurately and is taken as encompassing multiple dimensions. Many data or information quality frameworks have been presented in the literature. They contain some quality dimensions or categories derived normally based upon some research method in a specific domain with a set of quality metrics, criteria, components, items, or attributes. Eppler (01) gives five future directions for information quality research. The quest for more generic framework and the development of frameworks that show interdependencies between different quality criteria are emphasised. It appears, however, there is still lack of theoretical underpinnings of the exploration of interdependencies or inter-relationships among those quality indicators proposed. It leads to the difficulties for the professional who needs to decide an appropriate framework with a large set of criteria given for a task in hand within an organization. When investigating further, we observe that the terms 'data quality' and 'information quality' are considered synonyms by many if not all. They are usually interchangeable in relevant quality literature. The very concept of IQ is somewhat nebulous (Ballou et al., 03-04). It makes the discussion of the aforementioned question difficult and ambiguous. It seems to us that the notions of DQ and IQ are yet to be defined adequately in a grounded way.

In this paper, we wish to argue that a set of well-established theories, including Dretske's semantic theory of information (Dretske, 81), Devlin's 'infon' theory (Devlin, 91), Stamper's Organizational Semiotics (Stamper, 97) and Floridi's revised standard definition of information (Floridi, 05), would provide a novel insight into investigating DQ and IQ and shed light to the interdependencies among different quality indicators. The specific aim of this paper is to explore how this problem might be approached from another perspective, namely an information-theoretic perspective, and then further research might be pursued to develop a quality framework thereby to analyze and rearrange existing or derive new quality categories for quality assessment in practice.

This paper is organized as follows. We first review existing studies about DQ and IQ and the limitations of them that we notice particularly in Section 2. In Section 3, we present the basic notions of the theories referenced in this paper, and then introduce an information-centric framework for information systems and information flow. In Section 4, we propose a data-info quality model for understanding 'data', 'data quality', 'information quality' and their inter-relationships. And then we use this model to analyze quality categories from some existing approaches. Finally, in Section 5 we give conclusions and indicate future work.

LITERATURE REVIEW

Existing studies have reached a consensus that DQ and IQ is a multi-dimensional concept. Research efforts have been made to derive quality indicators for the development of different quality frameworks. Wang and some other researchers (Lee et al., 02; Wang & Strong, 96; etc], following the methods developed in marketing research for determining the quality characteristics of products, present a framework of information quality (IQ) from information consumer's perspective. They group all of their IQ dimensions into four IQ categories, Intrinsic IQ, Contextual IQ, Representational IQ, and Accessibility IQ. English gives three reasons to measuring information quality and two definitions of IQ (English, 99). One is its inherent quality, and the other is its pragmatic quality. His approach to quality includes three components, namely data definition quality, data content quality, and data presentation quality. DeLone and McLean's review of the MIS literature during the 1980's reports twenty-three IQ measures from nine previous studies (DeLone & McLean, 92; DeLone & McLean, 03). D&M (DeLone and McLean) say: "understandably, most measures of information quality are from the perspective of the user of this information and are thus fairly subjective in character" (DeLone & McLean, 92).

Furthermore, we find that there are different classifications for existing approaches to DQ and IQ in terms of different perspectives. We illustrate them in Table 1.

Perspective	Classifications
Research Approaches (Price & Shanks, 04)	Empirical research
	Practitioner-based approach
	Theoretical approach
	Literature-based approach
	Integrated approach
Communities (Lee et al., 02)	Academics' view
	Practitioners' view
Subject Domains (Burgess et al., 04)	Software Quality
	Data Quality
	Information Quality
	Web Quality

Table 1. Classifications of existing approaches to DQ/IQ

In addition, Eppler (01) reviews and finds out twenty information quality frameworks appearing in the literature from 1989 to 1999 in sixteen various application contexts. Many approaches to the quality problems in his findings, however, are proposed from a management, manufacturing, or technology perspective. He claims that the majority of frameworks they studied are context-specific rather than generic and widely applicable. He evaluates the frameworks according to two dimensions: analytic and pragmatic criteria respectively.

Through reviewing the literature, it seems to us that there is a lack of overarching theoretical perspectives or approaches for classifying existing quality frameworks with respect to their quality indicators delivered. Fundamental questions still remain as to how quality should be defined and the specific criteria that should be used to evaluate information quality (Price & Shanks, 04). As mentioned in Section 1, we are arguing that defining and distinguishing DQ and IQ should be addressed as a priority. However, from the work of Price and Shanks (04), the authors indicate that "due to the lack of agreement on the precise definition of information in the literature, we choose to restrict our usage of the term information to informal discussion and avoid its use in formal definitions". It is difficult to achieve an agreement on the definitions of the terms Data and Information. To this end, we attempt to use an information-theoretical perspective for seeking a solution and providing a fresh insight as it would seem necessary to construct a formal and theoretically sound quality framework under which we derive quality criteria and categories.

Existing studies normally consider data or information as a type of products or output of an information system and use the analogy between data and products to develop measurement models of DQ and IQ (Kahn, 97; Lee et al., 02; Price & Shanks, 04; etc). In the literature, the definitions of data quality and information quality are distinguished depending on whether information is considered to be a product or a service. However, the analogical approach is still limited because data are after all different from products (Liu & Chi, 02).

Theoretical approaches do appear in the literature. Wand and Wang drive quality definitions by anchoring them in ontological foundations and base on the notion that the role of an information system is to provide a representation of an application domain as perceived by the user. For the information system to function properly, both the representation and interpretation transformations, involved in the development and use of an information system, need to be performed flawlessly (Wand & Wang, 96). It results in a set of four intrinsic data quality dimensions: complete, unambiguous, meaningful, and correct. A semiotic information quality framework (Price & Shanks, 04) is presented to define information quality and corresponding quality categories in terms of three semiotic levels, namely syntactic, semantic, and pragmatic, defined by Morris (38) and in terms of definitions for data, information and meaning by Mingers (95). Hill (04) proposes an information-theoretic model based upon Shannon & Weaver's information theory for the purpose of considering customer information quality in an organization. It provides a quantitative assessment of proposed information quality improvements. However, there seems a lack of knowledge and attempt of using an information-theoretic perspective for investigating both terms of 'data' and 'information' and DQ and IQ. For example, Wand and Wang (96) derived four DQ attributes, which is only a small sample of the attributes in assessing intrinsic DQ. This might be due to the lack of an understanding of the subjective and objective nature of the domain.

AN INFORMATION-THEORETIC APPROACH TO QUALITY

Through reviewing the literature, we believe that an information-theoretical underpinning for the terms of 'data', 'information', DQ, and IQ should shed light to the quest of a generic quality model for the purpose of exploring interdependencies or inter-relationships among quality indicators proposed in various quality frameworks. In this paper, we present an overall model for such a purpose that is based upon a set of well-established theories.

THEORIES OF SEMANTIC INFORMATION

Information is still an 'explicandum' (Floridi, 05) in academic community today. Numerous attempts have been made to define it, but many of them are 'merry-go-round' definitions (Stamper, 97). Shannon and Weaver's paper (49) over half a century ago gives a mathematical model of communication, in which they use probability to define the amount of information that is caused by 'reduction in uncertainty'. This covers only the engineering aspect of information creation and transmission. Dretske (81) makes a profound paradigm shift from engineering aspect to semantic aspect of information. We take Dretske's account of the relationship between information and knowledge to be an important insight, which we intend to use as a way of incorporating epistemological considerations into the theory of information.

Following Dretske, information will be taken as created by or associated with a state of affairs among a set of possibilities of a situation, the occurrence or realization of which reduces the uncertainty of the situation. We focus on claims of the form 'a's being F carries the information that b is G'. From the point of view of semiotics, which has been used in developing a science for information systems, we say that one signal, a's being F, carries information about a state of affairs, b is G. Relevant to this, Dretske establishes the following definition of information content:

Let k be prior knowledge about a specific information source, r being F carries the information that s is G if and only if the conditional probability of s being G given that r is F is 1 (and less than 1 given k alone).

Following above definition, we proposed our first basic notion called 'data bears information' (Hu & Feng, 02; Xu & Feng, 02) which is now re-illustrated in Figure 1.

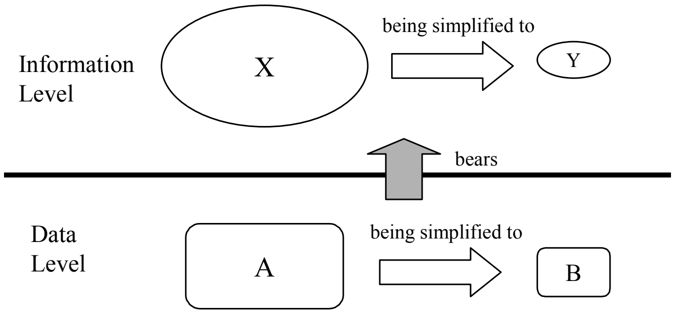


Figure 1. Simplification on information level and data level

The main point relevant to this paper that this diagram illustrates is that a representation/signal is considered to represent/carry part of information existing in the real world. When the source of information, namely that part of real world, is changed or simplified, a new representation/signal could be used to replace the old one. For example, in the database area, we could use (entity-relationship) ER schemas to design a conceptual representation for a university (a part of real world). With the modification made on the information requirements of the university's information systems, the representation used to bear the information source, namely ER schemas in this case, would be rearranged accordingly.

Information can be transmitted. A state of affairs, say r_1, is a particular case or an instantiation of a general situation, say r. The reduction in uncertainty at r due to the occurrence of r_1 may be accounted for by one or more events, say s_1, s_2,...,s_n, that occur at another general situation, say s. This gives rise to a special kind of relationship - '*informational relationship*' (Dretske, 81) - between these two general situations r and s. An informational relationship captures certain degree of dependency between a state of affairs r_1 of a general situation r and what takes place in another general situation s. This dependency can be demonstrated by the fact that r_1's appearance alters the distribution of probabilities of the various possibilities at s. The dependency is a type of regularity concerning different general situations based upon nomic dependencies (Dretske, 81), logic, or norms, etc. in a social setting. Due to this relationship, information created at s is transmitted to r. We will call s the 'information source', and r 'the bearer of information about s'. Moreover, a state of affairs r_1 at r can be seen as a signal that carries information about s. A sign/signal carries information about states of affairs in the world – what it signifies, even though the sign/signal may never be actually observed by anyone. Besides, if it is recorded, r_1 becomes a piece of data. Thus data carry information. In general, data in a database system are a collection of recorded signals or events, which bear certain information about the source within a process of information transmission. Information is carried by a sign and is objective and in analogue form.

Therefore we believe that it would be beneficial to look into problems regarding data and information from the perspectives adopted by various semantic information theories, which might help reach the root and reveal the essence of the problems. In order to introduce a theoretically sound foundation for the notions of information and our data-info quality model, we start with the ontological assumption that information is objective. In the beginning there was information. The word came later (Dretske, 81). The existence of information is independent of its interpreters or receivers (agents). We notice that Floridi defines four types of data: primary data; metadata; operational data; derivative data (Floridi, 05). He revises the 'standard definition of information' and adds a fourth condition to it. His work will be discussed further in Section 4.

THE S-B-R FRAMEWORK

To facilitate further studies of information within the context of information systems, that is, to gain insight and to be able to explain various phenomena in human communication, information creation and transformation, and the development of information systems, an overarching framework seems highly desirable even necessary. Aforementioned various theories and semiotics can be seen, among other things, address the issue of information and information flow in different ways and emphasize different aspects of it. We find that all these may be incorporated within a framework, which would help make sense of them, and make good use of them in understanding information and information flow. We believe that such a framework should be formulated from the point of view of how information is created, carried and finally received. Therefore we have created a framework consisting of Information Source, Information Bearer and Information Receiver, and the links between them. We call such an abstract model the 'S-B-R Framework' (illustrated in Figure 2).

We use a simple example to show how this framework might work. As illustrated in Figure 2, some information is created due to reduction in uncertainty, for example, the tree is 80 years old, rather than it is possible that the tree is 40 years old or 80 years old among many other possibilities at an information source. This information can be carried by an information bearer due to an informational relationship between the source and the bearer, which may be based upon some 'nomic dependencies' (Dretske, 81). An information bearer provides an opportunity for an information receiver, for example a human agent, to receive information about the information source. By consulting an information bearer, an information receiver can acquire information (illustrated by dotted line in Figure 2) if the receiver is aware of and attuned to some constraints (Devlin, 91), which formulates the dependency and therefore the informational relationship between the bearer and the source.

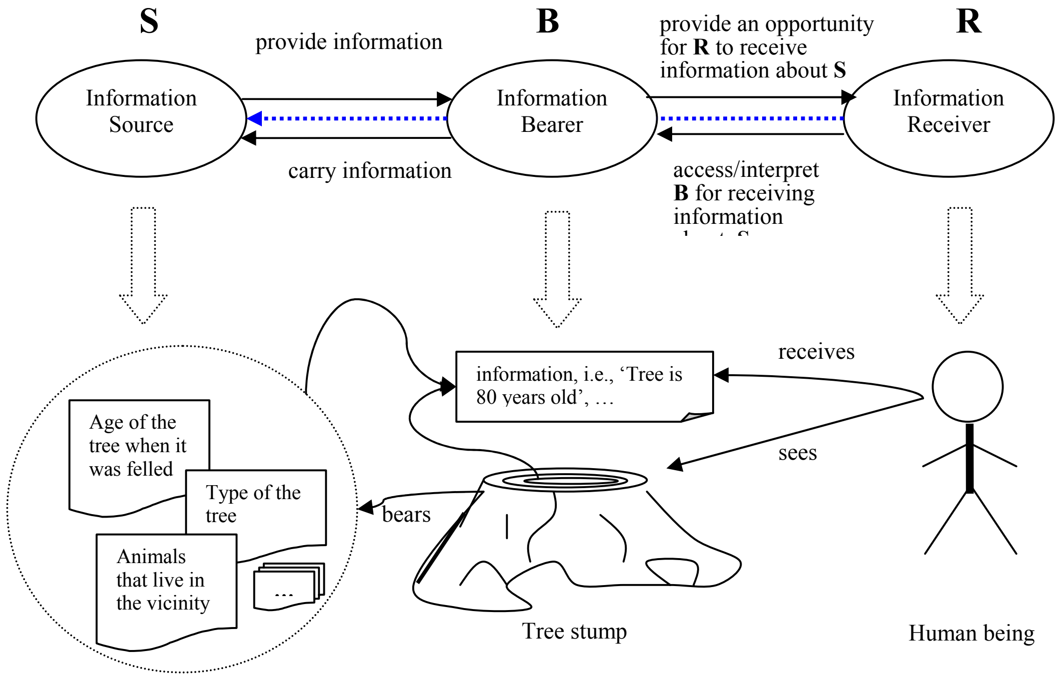


Figure 2. S-B-R **Framework**

Information Source

Information must be created in the first place. Following Dretsk, any situation may be regarded as a source of information as long as reduction in uncertainty takes place. It could be a Universe of Discourse, a particular situation (Devlin, 91), a relation, an event with uncertain outcomes, and so on. For example, the situation 'choosing one from eight employees to do an unpleasant job' can be an information source S.

From the point of view of semiotics, an information source S can be seen as the 'sign object' (Falkenberg, 98) that conforms to the definition of 'sign' given by Charles Sanders Peirce. It is a thing that the sign alludes or refers to.

Information Bearer

Information flow requires, as necessity, some representation of information, which we call the bearer. An Information Bearer can be a traffic light or signal, a physical sign or an IT system. Following Stamper (97), anything, say x, can function as a sign if it can stand for something else, say y, for the people in some community. Here, x is an information bearer for y. With our S-B-R framework, our ontological assumptions are that information may or may not be carried by a bearer; information can be conveyed only through a bearer; and information is independent of whether one receives it or can receive it or not. For example, if a book were written in ancient Chinese, we would consider that it carries certain information no matter whether we can read it or not.

In addition, we maintain that the literal meaning, if any, of a bearer is independent of the information that it bears. It is only accidental that the former is (part) of the latter.

Considering the structure of a sign given by Peirce, we agree that the 'representamen', which is a thing serving as the 'carrier' of the sign, is independent of its meaning (Falkenberg, 98). For example, an entity in an Entity Relationship data schema might refer to something that has no semantic correspondence with the meaning of the name given to that entity.

Information Receiver

To be able to receive information carried by a bearer, following Devlin, we maintain that an information receiver must be aware of and actually invokes some relevant 'constraints' (Devlin, 91) in order to receive information that is borne by a bearer. Different receivers may receive different information from the same bearer. The users of an information system are information receivers. In a system integration environment, an agent or a mediator can be an information receiver, which may process information further.

A DATA-INFO QUALITY MODEL

Information Quality is critical in organizations (Ballou et al., 03-04; DeLone & McLean, 2003). Early research efforts in Data Quality at MIT led to the development of the Total Data Quality Management (TDQM) cycle: Define, Measure, Analyze, and Improve (Wang, 04). Tu & Wang worked on ER extensions at the attribute level via modeling data quality of the original schemas (Tu & Wang, 93). Brodie (80) places the role of data quality within the life-cycle framework with an emphasis on database constraints. We believe that data quality has a close relationship with the tasks of information systems design and information quality has an inter-relationship with data quality of an information system.

In this section we put forward an observation, namely, it might be helpful to go back to the basics of information systems development. A similar perspective has been utilized by Wand and Wang (96). In this paper, we use another perspective, namely, information-theoretic perspective to look at Information Systems from the point of view of information flow from the source of information to the receiver of the information via some information bearer for the purpose of forming a data-info quality model. This idea comes from a seemingly widely accepted opinion that an information system is designed to store data (including multi-media data) and provide information to the information consumers. It is an 'information-bearing' media for the purpose of serving business processes and performance within an organization. Furthermore, it appears that there is a lack of a practical, theoretical-grounded information-centric model in the literature thereby to explore and analyze an inevitable phenomenon, namely, information flow, in IS development and IS evaluation, in particular, DQ evaluation and IQ evaluation. The motivation of our work is that we aim to bring some contribution on the theoretical level through our model and address relevant issues mentioned in Section 1.

DEFINITIONS

Many definitions of the terms 'data quality' and 'information quality' have been proposed in the literature. Eppler lists seven definitions of information quality from reviewing existing literature on information quality published from 1989 to 1999 (Eppler, 01). It seems that many of them are defined from a management, manufacturing, or technology perspective. Some definitions for both of terms are ambiguous and overlap. We wish to argue that this might be caused by the lack of a sound theoretical foundation. The S-B-R framework described above might fill in this gap by providing a fresh insight into the problem and help define 'data' and 'information' for studying 'data quality' and 'information quality'. Drawing on relevant literature regarding data quality and information quality and under the S-B-R framework, we generalize a conceptual model for considering these two terms as illustrated in Figure 3. We call it the 'data-info quality model'.

In the diagram, **S** normally contains three parts in the context of Information Systems Development. They are 'original user requirements', 'user expectations', and 'organizational needs'. The latter two change due to the dynamic nature of organizational goals, business strategies and performance. In the middle of the diagram, **B** is an information system that is a carrier or a mediator of information source **S**. It can be an ERP system, a CRM system, and so on, in the core of which lies a data engine, such as a database or a data warehouse. **R**, the information receiver, receives information, which is part of **S**, by accessing and interpreting **B**.

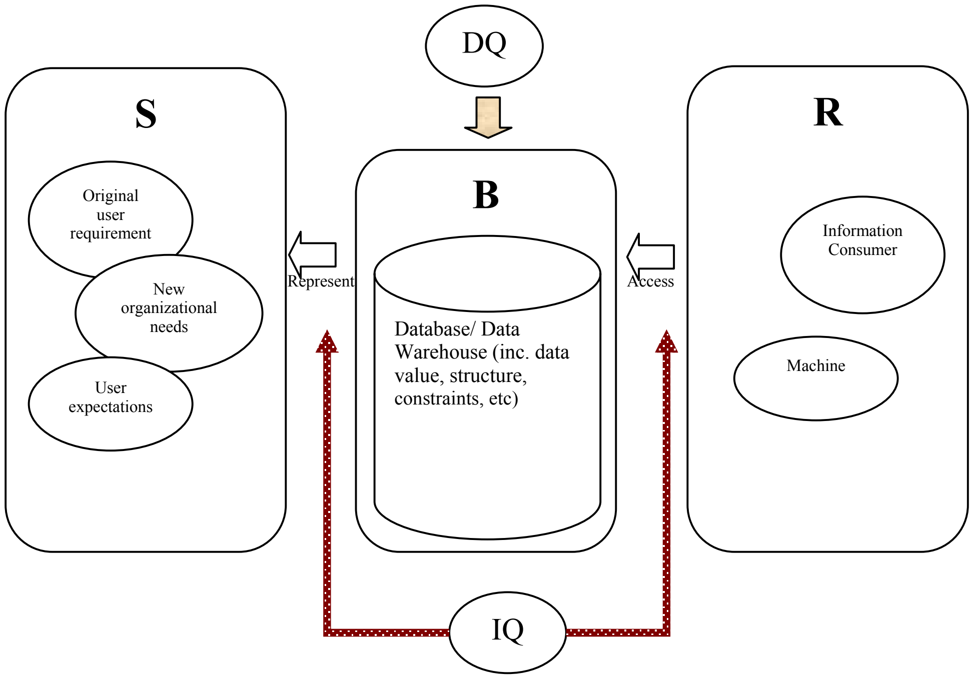


Figure 3. A Data-info Quality Model

Following the notion of 'data bears information' discussed in Section 3 and the objectives of data quality and information quality evaluation appearing in context (Wang et al., 95) we look at Information Bearer (B) for assessing 'data quality'. In the other words, the assessment of data quality is a task to define the quality of an information bearer. For assessing 'information quality' of an information system, we examine the linkage between Information Source (S) and Information Bearer (B), and the linkage between Information Receiver (R) and Information Bearer (B). In the other words, to assess information quality, we have to take the whole chain from S through B to R into consideration. We examine how well the information bearer represents the information source, and how well the information bearer supports the information receiver. That is to say, we look at how good the bearer is at conveying information to the receiver who would use perception and other cognitive means for this purpose. To enable such assessment, we present the information-theoretic definitions of data, information, data quality and information quality below.

Definition 1. *Data* is a set of values recorded in an information system, which are collected from the real world, generated from some pre-defined procedures, indicating the nature of stored values, or regarding usage of stored values themselves; or, a model for the purpose of organizing, constraining, representing those values in an information system for its consumers.

We define *data* here in a broad sense to cover values and structures existing in an information system. Following Floridi, the first type above can be of four types (namely primary data, metadata, operational data, and derivative data) according to their sources and purposes. The second type has a direct impact on the organization of data of the first type in terms of requirements.

Definition 2. *Information*, carried by non-empty, well-formed, meaningful, and truthful data (Floridi, 05), is a set of states of affairs, which are part of the real world and independent of its receivers.

We define *information* in an objective way following Dretske and Floridi. Floridi (05) revises standard definition of information with adding a fourth condition that information must be truthful. As explained by Florid, 'Truthful' is used here as synonymous for 'true', to mean 'representing or conveying true contents about the referred situation or topic'.

Definition 3. *Data Quality* is the intrinsic quality of data (a type of information bearer) itself.

This definition reveals the objective characteristics of the task of evaluating the quality of data, such as, representation, precision, and etc. It is in conformity with the discussion of the 'syntactic quality criteria' reported by the work of Price and Shanks (04) and the 'inherent information quality characteristics' defined by English (99), and the 'intrinsic' and 'contextual' data quality category proposed by Wang and Strong (96).

Definition 4. *Information Quality* is the degree to which the information is represented and to which the information can be perceived and accessed.

The term 'information quality' is defined from two directions in our data-info quality model. It is not a one-array concept; rather it is *the degree of some relevant correspondence* between the information source and the information bearer, and between the information bearer and the information receiver respectively. From a semiotic perspective, our work on this level is also in conformity with the 'semantic quality criteria' and the 'pragmatic quality criteria' reported by Price and Shanks (04), the 'pragmatic information quality characteristics' defined by English (99), and the 'representational' and 'accessibility' data quality categories proposed by Wang and Strong (96).

DATA QUALITY VS. INFORMATION QUALITY

According to Floridi (05), nonempty, well-formed and meaningful data may be of poor quality. Data that are incorrect, imprecise or inaccurate are still data and they are often recoverable, but, if they are not truthful, they can only constitute misinformation, which is not information at all. Following Floridi and considering our data-info quality model, we believe that high data quality is a necessary condition for seeking high information quality within an information system. It is not, however, a sufficient condition. For example, a well-organized database using Chinese characters that has recorded accurate and timely stock information does not have high information quality if its users include some non-Chinese speakers even though the system has high data quality. Take another example, a decision-maker is provided a stock report with a set of complete, readable, and well-formatted data. He/she will not obtain any information if data is not true or inaccurate to reflect real situation. Therefore, high information quality should be based upon high data quality, and the data must be appropriately presented and accessible to the information consumer.

Based upon our above thinking and definitions regarding data quality and info quality, we can rearrange existing quality dimensions and criteria in the literature into a new framework, as shown in Table 2 and Table 3 respectively. It is intuitively organized based upon our experience and corresponding description of the selected dimensions from the literature.

	Data Quality	Information Quality
Price and Shanks (04)	Syntactic quality	Semantic quality, pragmatic quality
English (99)	Inherent characteristics	Pragmatic characteristics
Wang et al. (96)	Intrinsic, contextual	Representational, accessibility
Dedeke (00)	Ergonomic, accessibility	Representation

Table 2. Some existing quality dimensions rearranged within a data-info quality framework

Data Quality	Information Quality
Accuracy, format, timeliness, precision, amount of data, etc.	Relevancy, accessibility, usefulness, readability, completeness, consistency, reliability, importance, truthfulness, etc.

Table 3. Some existing quality criteria rearranged within a data-info quality framework

Interdependencies among quality dimensions and criteria can be further explored and studied from the point of the view of the inter-relationship between data quality and information quality. The quality criteria for the former will clearly have impact on the latter. Distinguishing information quality from data quality will help IS professionals and organizations derive required and appropriate quality criteria for the task in hand. Further analysis and validation on aforementioned issues will be reported in our future publications.

OBJECTIVITY VS. SUBJECTIVITY

In the relevant literature, the notion of data or information quality depends on the actual use of data. They are normally investigated from the viewpoint of information consumers. From the work of Wand and Wang (96), a design-oriented approach is proposed to define data quality based upon a concept called 'possible data deficiencies' in a system context. Ballou and Pazer's study focuses primarily on intrinsic dimensions that can be measured objectively (Ballou & Pazer, 85; Ballou & Pazer, 95). However, it would appear that the issue of subjectivity versus objectivity that are involved in data and information quality evaluation in information systems are hardly addressed adequately. We believe that to address this issue is important - not only can an insight of the problem be gained, but also it should benefit the selection of research methods for the development of a methodology for assessing data quality and information quality.

Our preliminary thinking about this philosophical issue is that it can be looked at with the 'S-B-R' perspective. In Figure 3, we have shown that information quality is concerned with two linkages between S and B, and between B and R separately. The first linkage embodies the objective aspect of the problem following our ontological assumption on information. It depends on design-oriented or system-oriented. Therefore, theoretical techniques (i.e., SQL query design, schema transformation, and etc) and quantitative research methods will contribute to detecting and providing solutions to the problems. The second linkage should be looked at within a social setting, and therefore predominately inter-subjective or subjective (Mingers, 95). For example, different groups of information consumers may have different qualifications and different knowledge background, and therefore may receive different information from accessing the same information bearer. Qualitative research methods may contribute to identifying problems, reaching conclusions and obtaining solutions. Much more work should be carried out along this avenue.

SUMMARY AND FUTURE WORK

In this paper, we have examined some fundamental issues concerning data and information quality evaluation from an information-theoretic perspective that is informed by a set of well-established theories. We have proposed a data-info quality model based upon an information-centric framework to provide a rigorous theoretical foundation for (1) defining and distinguishing the terms of 'data quality' and 'information quality'; (2) discussing the inter-relationships between two terms; (3) studying the subjective and objective characteristics of data quality and information quality. A more generic framework for data and information quality and a set of quality categories and criteria with their interdependencies articulated will be reported in future publications.

This model is being validated through a two-stage survey. First, a series of interviews will be organized with selected organizations, enterprises, and institutions in the UK and China. The goal is to elaborate the model using a qualitative research method and to generate a data-info quality framework. Then, a questionnaire will be used to test in the real world the proposed quality framework and to categorize quality criteria.

REFERENCES

Ballou, D. P. and H. L. Pazer, "Modeling Data and Process Quality in Multi-input, Multi-output Information Systems", *Management Science*, 31(2) 1985, pp. 150-162.

Ballou, D. P. and H. L. Pazer, "Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff", *Information Systems Research*, 6(1) 1995, pp. 51-72.

Ballou, D., Madnick, S., and Wang, R. Y., "Special Section: Assuring Information Quality", *Journal of Management Information Systems*, Winter 2003-4, Vol. 20, No. 3, pp. 9-11.

Brodie, M. L., "Data quality information systems, information, and management," vol. 3, pp. 245-258, 1980.

Burgess, M., Fiddian, N. J., and Gray, W, "Quality measures and the information consumer", *ICIQ 2004*

Dedeke, A. "A Conceptual Framework for Developing Quality Measures for Information Systems", *Proceedings of the 2000 Conference on Information Quality (IQ-2000)*, Cambridge, MA, USA, 2000, pp.126-128.

DeLone, W. H. and McLean, E. R. "Information Systems Success: The Quest for the Dependent Variable", *Information Systems Research*, Volume 3, No. 1, March 1992, pp. 60-95.

DeLone, W. H., & McLean, E. R. "The DeLone and McLean model of information systems success: A ten-year update". *Journal of Management Information Systems*, 19(4), 2003, pp. 9-30.

Devlin, K. *Logic and Information*. Cambridge University Press, Cambridge, 1991.

Dretske, F. I. *Knowledge and the Flow of Information*, Basil Blackwell, Oxford, 1981.

English, L. P., *Improving Data Warehouse and Business Information Quality*. Wiley & Sons, New York, 1999.

Eppler, M. J., "The concept of information quality: an interdisciplinary evaluation of recent information quality frameworks", *Studies in Communication Sciences*, 1 (2001) p.167-182.

Falkenberg, D. E., Hesse, W., Stamper, R., et al. *A Framework of Information Systems Concepts – The FRISCO Report (web edition)*, IFIP, 1998.

Floridi, L., "Is Semantic Information Meaningful Data?", *Philosophy and Phenomenological Research*, Vol. LXX, No. 2, March 2005.

Hill, G., “An information-theoretic model of customer information quality”, *Proc. IFIP Int’l Conf. on Decision Support Systems*, Italy, 2004.
Hu, W. and Feng, J. 2002. “Some considerations for a semantic analysis of conceptual data schemata”, In *Systems Theory and Practice in the Knowledge Age*, (E. Ragsdell et al.), Kluwer Academic/Plenum Publishers. New York. ISBN 0-306-47247-3.
Kahn, B. K., Strong, D. M. and Wang, R. Y., “A Model for Delivering Quality Information as Product and Service”, in *Conference on Information Quality*, Cambridge, MA, pp. 80-94, 1997.
Lee, Y. W., Strong, D., Kahn, B., and Wang, R., “AIMQ: a methodology for information quality assessment”, *Information and Management*, 40(2) pp. 133-146, 2002.
Liu, L. and Chi, L., “Evolutional Data Quality: A Theory-specific View”, *ICIQ 2002*.
Mingers, J. “Information and meaning: foundations for an intersugjective account”, *Information Systems Journal*, 1995; 5:285 – 306
Morris, C., “Foundations of the Theory of Signs”, in *International Encyclopedia of Unified Science*, vol.1, University of Chicago Press, London, 1938.
Price, R.J., Shanks, G.A., “A semiotic information quality framework”, In R. Meredith, G. Shanks, D. Arnott and S. Carlsson (eds.) *Proceedings of the 2004 IFIP International Conference on Decision Support Systems (DSS2004): Decision Support in an Uncertain and Complex World*, Prato, Italy, 1-3 July: 658-672.
Redman, T., *Data Quality: The Field Guide*, New Jersey: Digital Press, 2001.
Shannon, C. E. and Weaver, W. The mathematical theory of communication. Urbana: University of Illinois, 1949.
Stamper, R. “Organisational Semiotics”, In *Information Systems: An Emerging Discipline?*, Mingers, J and Stowell, F. ed. The McGraw-Hill Companies, London, 1997.
Tu, S.Y. and Wang, R. Y., “Modeling Data Quality and Context Through Extension of the ER Model”, Massachusetts Institute of Technology (MIT) Sloan School of Management, Cambridge, MA, TDQM-93-13, 1993.
Wand, Y. and Wang, R. Y., “Anchoring Data Quality Dimensions in Ontological Foundations”, Communications of the ACM, 39(11): 86-95, 1996
Wang, R.Y., Kon, H.B., and Madnick, S.E., “Data quality requirements analysis and modeling”, *Proc. Ninth Int’l Conf. on Data Engineering*, pp. 670-677, Vienna, 1993.
Wang, R. Y., Storey, V. C., and Firth, C. P., 1995, “A Framework for Analysis of Data Quality Research”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7, No. 4, 1995.
Wang, R.Y. and Strong. D.M. (1996) “Beyond Accuracy: What Data Quality Means to Data Consumers”, *Journal of Management Information Systems*, 12(4): 5-34.
Wang, R. Y., “Data Quality: Theory in Practice”, *EPA 23rd Annual Conference*, April 2004.
Xu, H. and Feng, J., “Towards a Definition of the ‘Information Bearing Capability’ of a Conceptual Data Schema”, In *Systems Theory and Practice in the Knowledge Age*, (E. Ragsdell et al.), Kluwer Academic/Plenum Publishers. New York. ISBN 0-306-47247-3, 2002.

An Object-Oriented Spatial Decision Support System for Emergency Management

Tsai-Yun Liao
Chaoyang University of Technology, Taichung, Taiwan
Email: tyliao@cyut.edu.tw
Ta-Yin Hu
National Cheng Kung University, Tainan, Taiwan
Email: tyhu@mail.ncku.edu.tw

Abstract

The central part of Taiwan area was rocked by a magnitude 7.3 earthquake on September 21, 1999. The earthquake knocked out several major provincial highways and caused the collapse of hundreds of thousands of buildings and structures. In order to relieve damages caused by emergency incidents, decision makers need to respond quickly to bring traffic operations under control. The need for a spatial decision support system (SDSS) to provide path information with realistic traffic characteristics arises in several stages during emergency management, such as evacuation route and alternative routes for transporting commodities. In this research, an object-oriented decision support system is proposed and implemented to illustrate possible benefits from the SDSS. The object-oriented approach, which represents physical entities of the problem domain by program modules (object classes) and utilizes UML as its tool, allows the user to control the levels of details and summarize the aggregate data during decision-making.

Keywords:

Emergency management, decision makers, SDSS, Object-oriented approach, UML

INTRODUCTION

The central part of Taiwan area was rocked by a magnitude 7.3 earthquake on September 21, 1999. The earthquake knocked out several major provincial highways and caused the collapse of hundreds of thousands of buildings and structures. During and after the earthquake, including evacuation, response, and recovery stages, several problems are identified and summarized:

(1) Traffic control strategies are not consistent with traffic flow patterns.

(2) The routes for procurement and logistics do not consider possible road damages.

(3) The emergency management center could not provide alternative routes for evacuation.

(4) During the recovery stage, available roads are not utilized and cooperated with traffic control strategies.

In spite of different possible classifications, common characteristics of disasters are their rareness and serious damages to the public. In other words, proper planning for evacuation and procurement might reduce the damage to the minimum. In order to relieve damages caused by emergency incidents, decision makers need to respond quickly to bring traffic operations under control. The need for a spatial decision support system (SDSS) to provide path information with realistic traffic characteristics arises in several stages during emergency management, such as evacuation route and alternative routes for transporting commodities. A good emergency SDSS must have graphical information display abilities to show the location of emergency incidents and direction of possible evacuation routes. GIS (Geographical Information Systems) technology has been developed as a means for organizing and analyzing spatial data to display graphical information to users (decision makers).

In this research, a two-tier object-oriented approach is proposed and implemented to illustrate possible benefits from the SDSS to support decision makers in responding to disasters more effectively. The object-oriented approach, which represents physical entities of the problem domain by program modules (object classes) and utilizes UML (Unified Modeling Language) as its tool, allows the user to control the levels of details and summarize the aggregate data during decision-making.

The two-tier object-oriented SDSS integrates model objects, GIS objects, data objects, and display interfaces. Thus, the system is easy to maintain and could be extended to include other objects easily, such as real-time traffic information objects.

There are four modules in the object-oriented SDSS: Data Base Management Systems (DBMS), Model Base management Systems (MBMS), GIS, and User Interface Management Systems (UIMS). In GIS part, the system incorporates MapX for managing GIS-related objects to display maps as well as numerical results. Decision criteria for evacuation, logistics, maintenance, and procurement, are established to reflect decision making processes. In the MBMS, two major computation objects, including static and dynamic computation objects, are constructed to provide path information. The static path computation object based on Floyd-Warshall Algorithm provides all node-to-node path calculations; the dynamic path computation object based on DYNSMART provides dynamic simulation of traffic flows [7,16]. These paths are then applied in the SDSS depending on decision criteria.

The paper is organized as follows. A brief review for the related issues such as transportation impacts under earthquake, SDSS, and object-oriented analysis and design is given in Section 2. An object-oriented SDSS architecture is proposed and explored in Section 3. The object-oriented analysis and design using UML are discussed in Section 4, followed by system implementation and illustrations. Conclusions and further enhancements are summarized in the last section.

LITERATURE REVIEW

In this section, several related issues are briefly described to provide a fundamental background for this research. Main topics include the impacts of earthquake, SDSS, DYNASMART, and object-oriented analysis and design.

Impacts of Earthquake

The 1994 Northridge earthquake in the U.S. disrupted four major highway routes in Southern California. Willson [19] examined the impacts of the earthquake on Los Angeles County trucking firms and found that congestion delay and circuitous routing were the most common impacts. Willson [19] also indicates that rerouting and rescheduling strategies were ad hoc, rather than part of prearranged earthquake responses. Following the Northridge earthquake, a strategy was needed to bring the traffic operations under one net and to point it in the right direction. This came about in the form of the Traffic Management Plan (TMP) that was developed by a team of traffic engineers from Caltrans.

At 1:47 on the morning of September 21, 1999, a 7.3 magnitude earthquake shook Taiwan. The earthquake caused in excess of 2,100 fatalities and 8,700 injuries. The government mobilized rescue units immediately after the earthquake occurred, but people from all quarters have criticized the relief effort. Therefore, it is essential to develop an effective SDSS for the emergency management. The different phases of impacts of earthquake related to the development of SDSS are outlined as follows:

Phase 1 – Rapid response and emergency relief (10-day)

- 1.1 Rapid assessment of impacted areas
- 1.2 Establishment of residence sites in damaged areas such as Puli and Tungshih
- 1.3 Procurement and distribution of basic commodities
- 1.4 Distribution of personal hygiene kits
- 1.5 Medical assistance to affected people

Phase 2 – Meeting basic needs (2 months)

- 2.1 Continued presence at Tungshih residence site
- 2.2 Routing of commodities
- 2.2.1 Needs assessment of most needy populations
- 2.2.2 Inventory and distribution tracking system

Phase 3 – Rehabilitation and reconstruction (12 months)

- 3.1 Unmet housing needs
- 3.2 Assistance to vulnerable groups
- 3.3 Assistance to sponsored children
- 3.4 Infrastructure improvement

SDSS for Emergency Management

A decision support system (DSS) is an interactive computer-based information system providing decision models and data to solve semistructured and unstructured problems for supporting decision makers [18]. A good emergency decision support system must have graphical information display abilities to show the location of emergency incidents and direction of possible evacuation routes. GIS technology has been developed as a means for organizing and analyzing spatial data to display graphical information to users (decision makers). Based on GIS, spatial decision support systems (SDSS) can be viewed as a linkage between GIS and modeling approaches such as operations research [5].

Tufekci [17] presents a conceptual framework for an effective, integrated, and modular DSS for hurricane emergency management. Pidd et al. [13] describes a prototype SDSS for emergency planners in developing contingency plans for evacuations from disaster areas. It links a GIS (ARC/INFO) with a specially written object-oriented micro-simulator. Zografos et al. [20] proposed a real-time DSS for roadway network incident response logistics.

Traffic Simulation-Assignment Model DYNASMART

DYNASMART uses an assignment and simulation modeling approach to assign time-varying traffic demands, model the corresponding traffic patterns, and evaluate network performance [10]. One of the key features of DYNASMART is its ability to model vehicle paths in the network as the explicit outcome of route assignment decisions at origins or at nodes of the network. Traffic flow is represented using a hybrid approach where vehicles are tracked individually or in macro particles, and moved consistently with macroscopic traffic flow relations.

Object-Oriented Analysis and Design

The object-oriented approach, which represents physical entities of the problem domain by program modules (object classes), has created considerable interest in recent years for the development of new information systems [4]. The object-oriented approach provides the capabilities to enhance the correspondence between problem areas and their semantic representation and to allow the user to control the levels of detail and summarize the aggregate data during decision-making.

UML is a visual modeling language adopted as a standard for object-oriented modeling and design in software development by the Object Management Group (OMG) [6]. UML, unifies the methods of Booch, Rumbaugh (OMT), and Jacobson, consists of a set of diagrams used to represent models of a system in diagrammatic notation through stages of development [2]. There have been several researches related to the object-oriented analysis and design and the adoption of UML in object-oriented modeling [1,3,8,9,12]. Generally, there are nine types of UML diagrams: use case diagrams, sequence diagrams, collaboration diagrams, state diagrams, activity diagrams, class diagrams, object diagrams, component diagrams, and deployment diagrams. In this paper, four diagrams including use case diagrams, class diagrams, sequence diagrams, and collaboration diagrams are deployed in order to construct an object-oriented SDSS for emergency management.

FRAMEWORK FOR OBJECT-ORIENTED SDSS

The object-oriented SDSS for emergency management aims at enhancing the interaction between decision makers and useful data as well as information during earthquake disasters; however, the SDSS could also help decision makers to evaluate possible routing strategies during any emergency management. The decision-making process for emergency management is discussed in order to establish the fundamental reasoning for the proposed SDSS, and then the proposed structure for emergency management is explored. The SDSS structure then is analyzed through a two-tier object-oriented analysis.

The decision making process for emergency management, as illustrated in Figure 1, is classified into five stages.

(1) Problem identification

Problem identification needs to consider emergency characteristics and identifies major impacts to the public.

(2) Decision making parameters

Decision makers provide possible parameters as well as weights for different evaluation criteria.

(3) Alternative generation and evaluation

According to emergency characteristics and decision parameters, possible alternatives are generated and evaluated.

(4) Decision

Alternatives are evaluated according decision criteria, and possible decisions are generated for decision makers.

(5) Implementation

Implementation stage executes the decisions and provides a feedback mechanism for decision makers to modify decision parameters for the next mission.

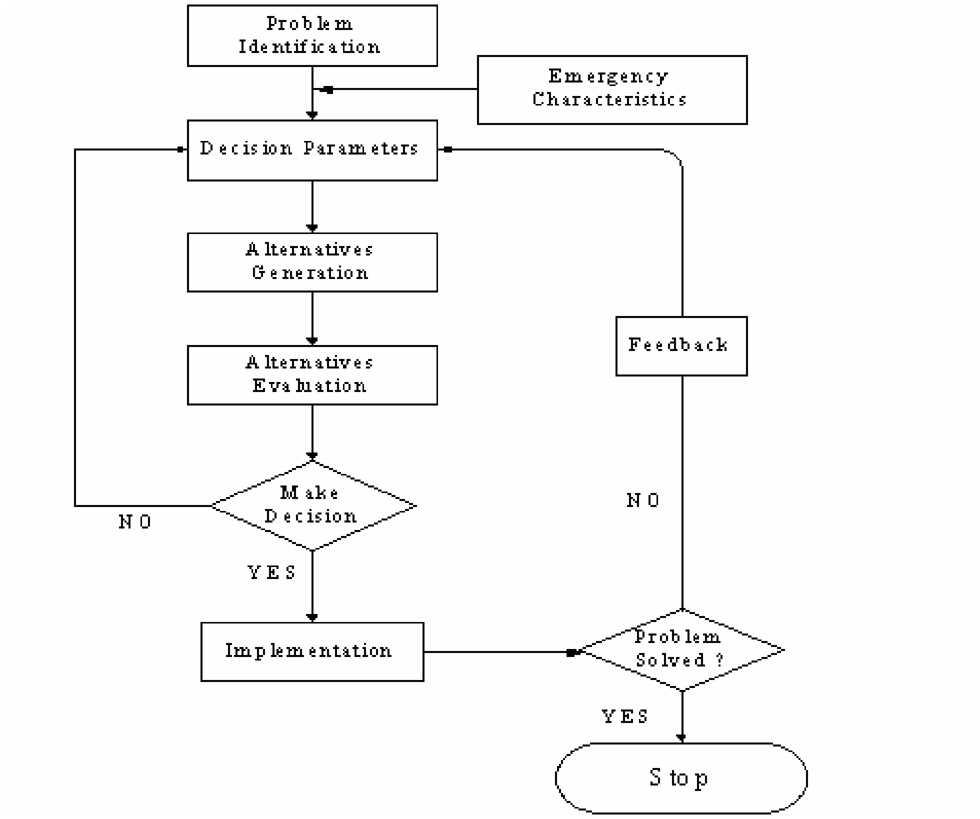


Figure 1. Decision Making Process for Emergency Management

A conceptual framework of SDSS for emergency management is proposed in Figure 2. The framework integrates possible features from use case analysis of UML and traditional DSS framework. The SDSS is classified into DSS tools, DSS Generator, and specific DSS functions. Several actors in this SDSS are Toolsmith, Technical Support, DSS builder, Intermediary, and Manager (user). DSS tools considered in the system are MapX, DYNASMART [7], operating systems, Microsoft Excel, PC and other devices. Three main components in DSS Generators are User Interface Management Systems (UIMS), Data Base Management Systems (DBMS), Model Base Management Systems (MBMS), and Geographic Information Systems (GIS). Four major specific DSS functions in this system include Procurement, Routing of Commodities, Road Maintenance, and Evacuation. Each function represents a use case in the UML diagrams.

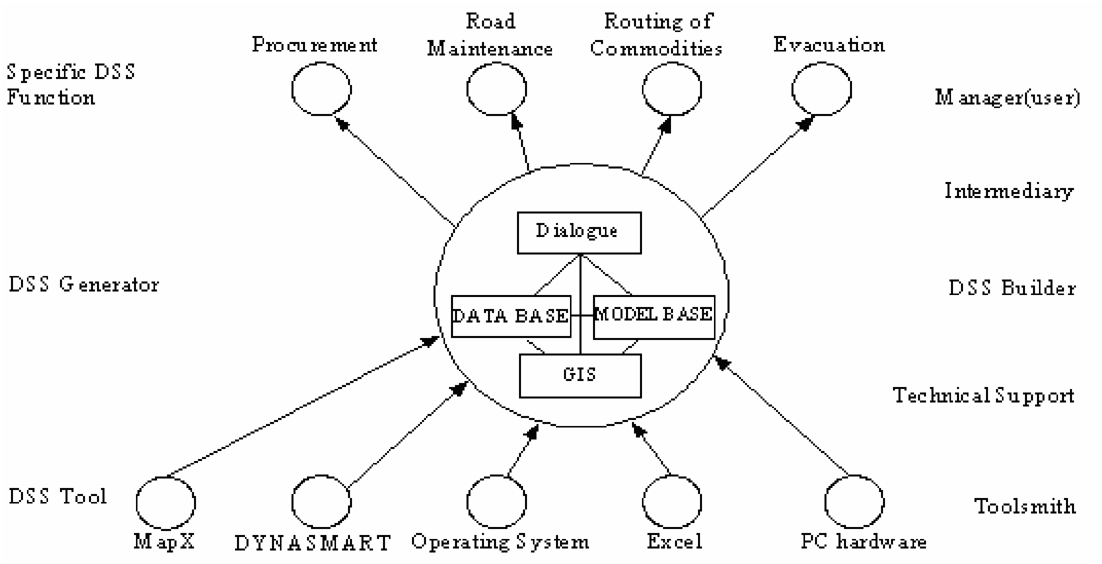


Figure 2. Conceptual SDSS Framework for Emergency Management

According to the above SDSS framework, an object-oriented SDSS for emergency management is proposed and shown in Figure 3. The object-oriented SDSS is constructed into two-layer models; the first layer deals with user-interface and the second layer is for data and model management. The object-oriented approach presents physical entities of the problem domain by program modules (object classes). These objects are described as follows:

(1) Text Process Object

Text process object reads the user input through files and/or keyboards, transfers the text into useful information, and stores the information into proper databases. The input includes real-time emergency situations, traffic conditions, population distribution pattern, and other possible factors.

(2) Graphic Object

Graphic object takes into account of graphic data and different types of information through different layers' constructions. The graphic object can display digitized maps as well as other information, such as population distribution, local administrations, and locations for possible rescue units, medical units, and engineering units. The graphic object plays an important role in helping decision makers to evaluate alternatives.

The graphic object is implemented based on MapInfo MapX, a cost-effective, robust, ActiveX component [11]. MapInfo MapX enables developers to add mapping functionality to any application. The result is a highly visual way to display and analyze location-based data to better serve customers, make better decisions, and manage assets and operations more effectively.

(3) Database and DataMgr Objects

Database object and DataMgr Object are two objects for data management. The system uses Microsoft Excel to manage data. DataMgr object retrieves and modifies data directly.

(4) Decision Analysis Object

Decision Analysis Object provides some scenarios under emergency management through possible damage conditions, and associated weights are pre-specified through users.

(5) Route Analysis Object

One primary goal in this research is to provide timely route for procurement as well as evacuation. The route analysis can be accomplished through static shortest path calculations [16] or through dynamic traffic simulation.

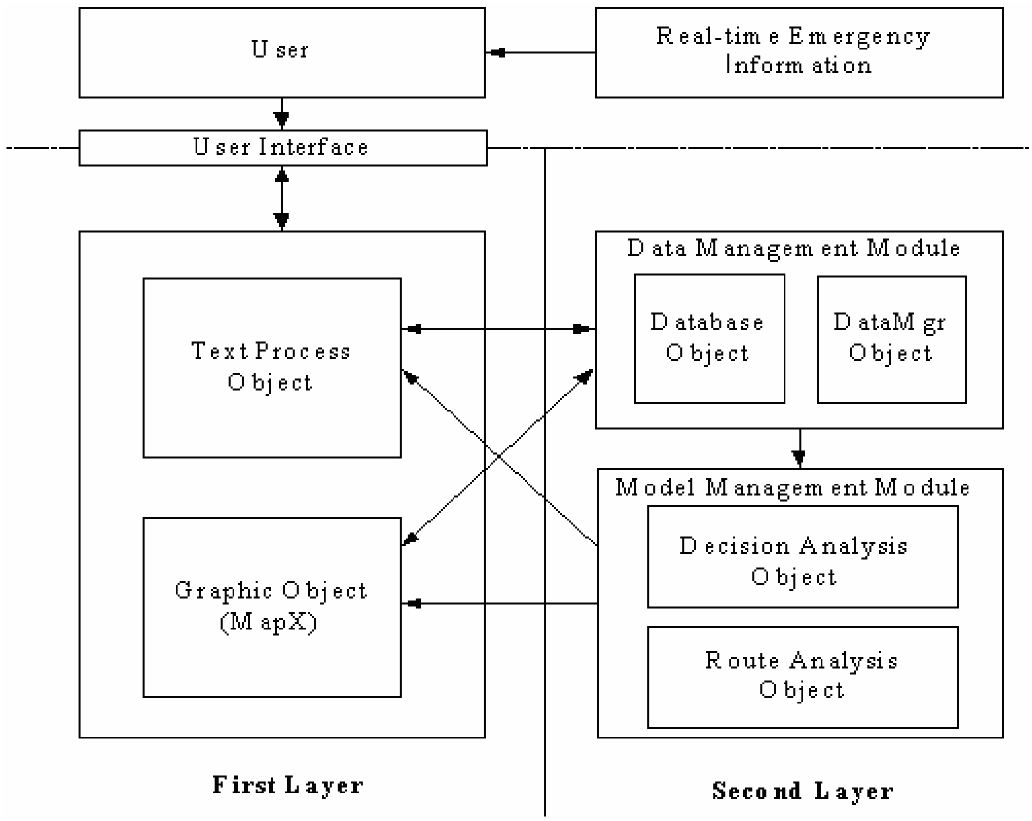


Figure 3. An Object-oriented SDSS for Emergency Management

OBJECT-ORIENTED MODELING FOR EMERGENCY MANAGEMENT SDSS

According to the proposed SDSS for emergency management in Figure 3, several UML diagrams are deployed in order to construct the object-oriented SDSS for emergency management:

(1) Use case diagrams

(2) Class diagrams

(3) Sequence diagrams

(4) Collaboration diagrams

Use case diagrams

Use case diagrams are models that describe the system functions of an application and show how users of the system (actors) interact with use cases [15]. Each use case encapsulates a sequence of operations of the system, typically triggered by interaction with an actor or another use case. Figure 4 shows a use case diagram that indicates the decision maker and required five use cases. The 'extend' relationship represents an optional operation after the originated use case has been performed [14]. The 'include' relationship indicates that the use case 'Data Management' is included when the user (decision maker) interacts with the three use cases 'Order Possible Mission', 'Locate and Display Resources', and Route Analysis and Display'.

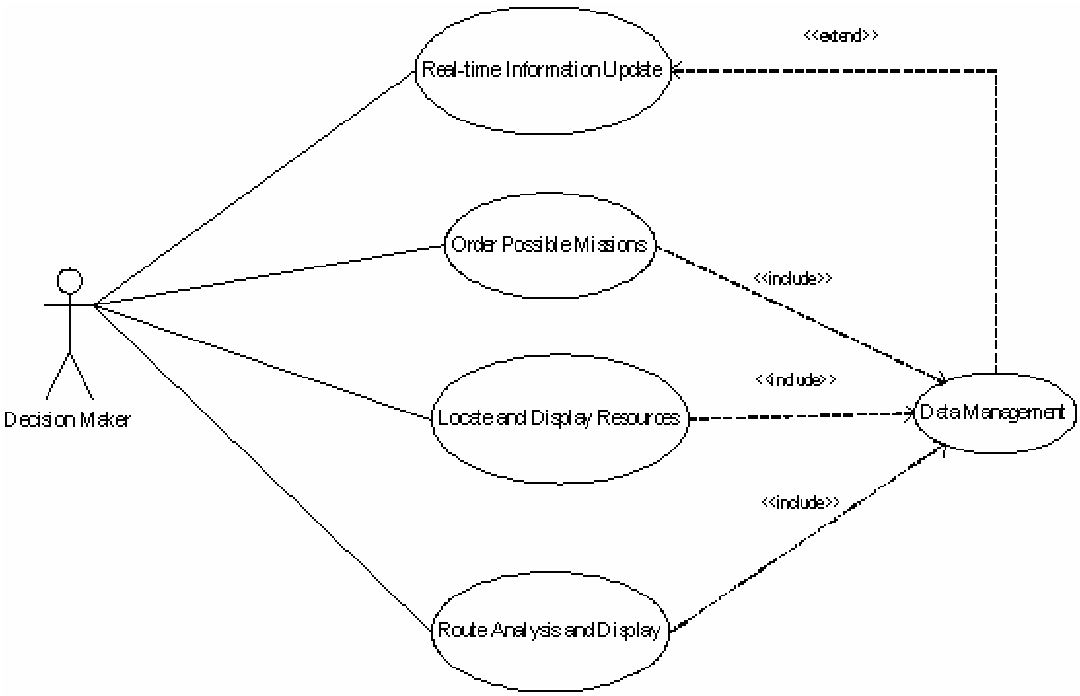


Figure 4. The Use Case Diagram

Sequence diagrams

Sequence diagrams are models that describe how groups of objects collaborate in some behavior and present the flow of messages between instances in time. An example of sequence diagrams, as shown in Figures 5, is generated to illustrate several SDSS functions, such as real-time information updating, emergency incidents evaluation and display, available resources display, and optimal path calculation and generation. The detail contents of the sequence diagrams are illustrated on left side of the figures.

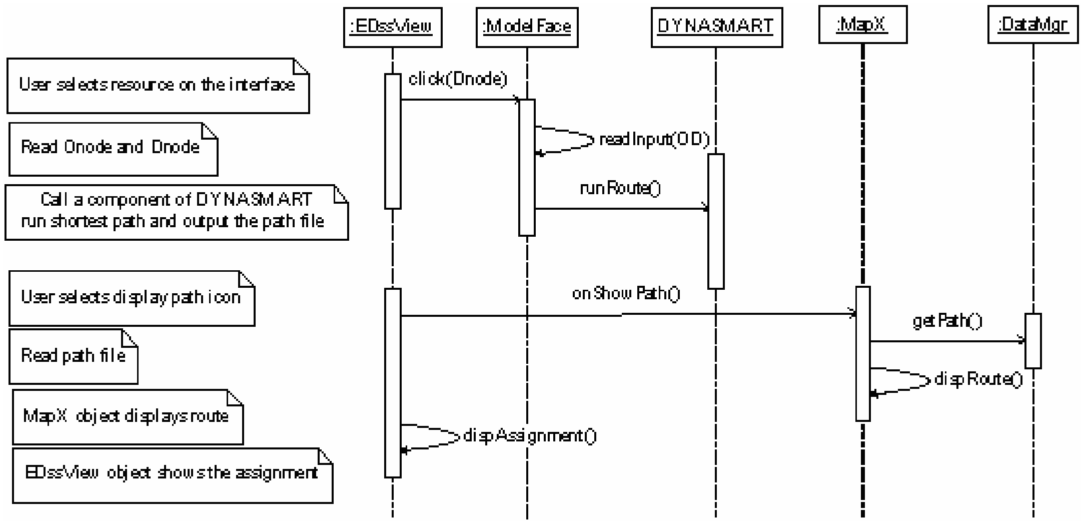


Figure 5. Route Analysis and Display

Collaboration diagrams

The collaboration diagrams show how objects work with each other by sending messages and exchanging data. Sequence diagrams and collaboration diagrams are two different views of object interactions. For example, the 'Route Analysis and Display' function shown in Figure 5 can be represented as collaboration diagrams in Figure 6 to show how objects work with each other. Detailed description of the collaboration diagram is presented below. Comparing with sequence diagrams, collaboration diagrams are better in modeling the logic of the implementation of a complex operation, particularly one that interacts with a large number of other objects.

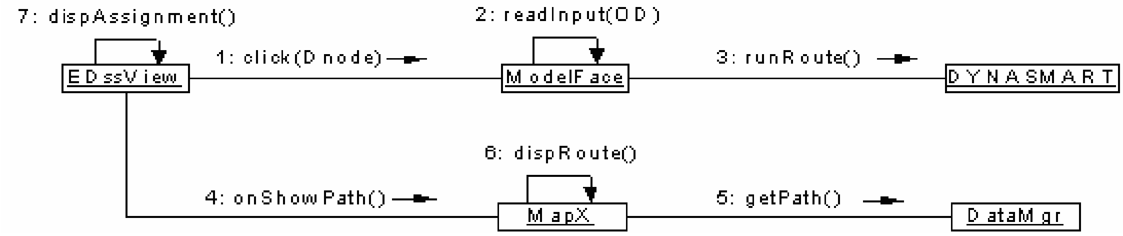


Figure 6. Route Analysis and Display (Collaboration Diagram)

(1) User selects resource on the interface.

(2) Read Onode (Origin node) and Dnode (Destination Node)

(3) Call a component of DYNASMART run shortest path and output the path file.

(4) User selects display path icon.

(5) Read path file.

(6) MapX object displays route.

(7) EDssView object shows the assignment.

Class diagrams

The class diagram technique has become truly central within object-oriented methods. The class diagram describes the types of objects in the system and the various kinds of static relationships that exist among them. The class diagram is shown in Figure 7. Class Data is an abstract data class, and is inherited by two classes, Class DataMgr and Class Database. Class Database provides maintenance functions for data base management, such as delete(), modify(), and maintain(). Class MapX, Class ModelFace, and Class DAM operate based on data provided through Class DataMgr. Different data sources and data formats are integrated through Class DataMgr, so other classes can focus solely on manipulation as well as optimization.

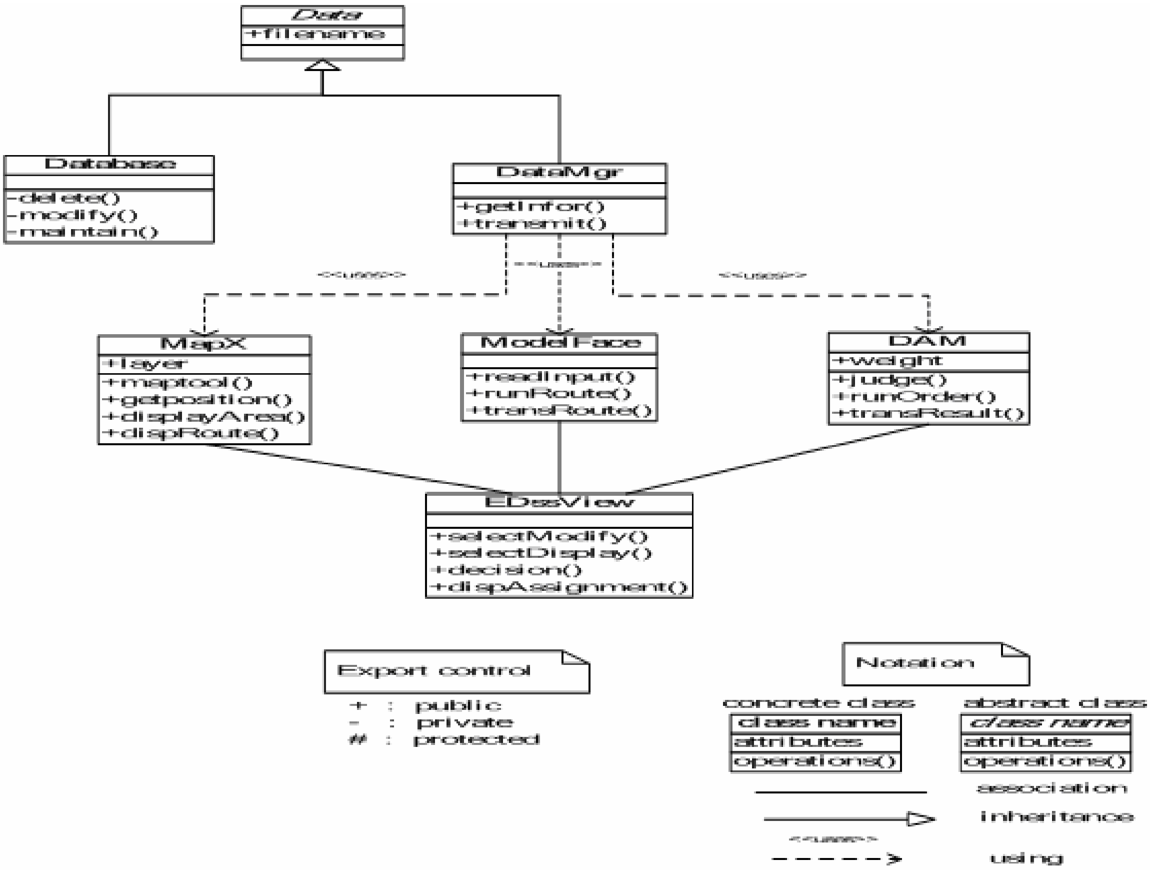


Figure 7. Class Diagram

SYSTEM IMPLEMENTATION AND ILLUSTRATIONS

The object-oriented SDSS is implemented by Microsoft Visual C++6.0 and MapX3.5. The system architecture is shown in Figure 8. The input/output and interrelationships among variety of objects are illustrated in the figure. The SDSS provides an interface between decision makers and specified application objects, and the requested information is sent to the SDSS to process optimal route generation according to user specified criteria. The route calculation algorithm then generates an integrated network with associated information attributes and calculates the optimal routes. The integrated network includes a network topology, link characteristics, and possible traffic conditions. The network is established according to an up-to-date digital map, in which area, geographic attributes, and traffic attributes are generated.

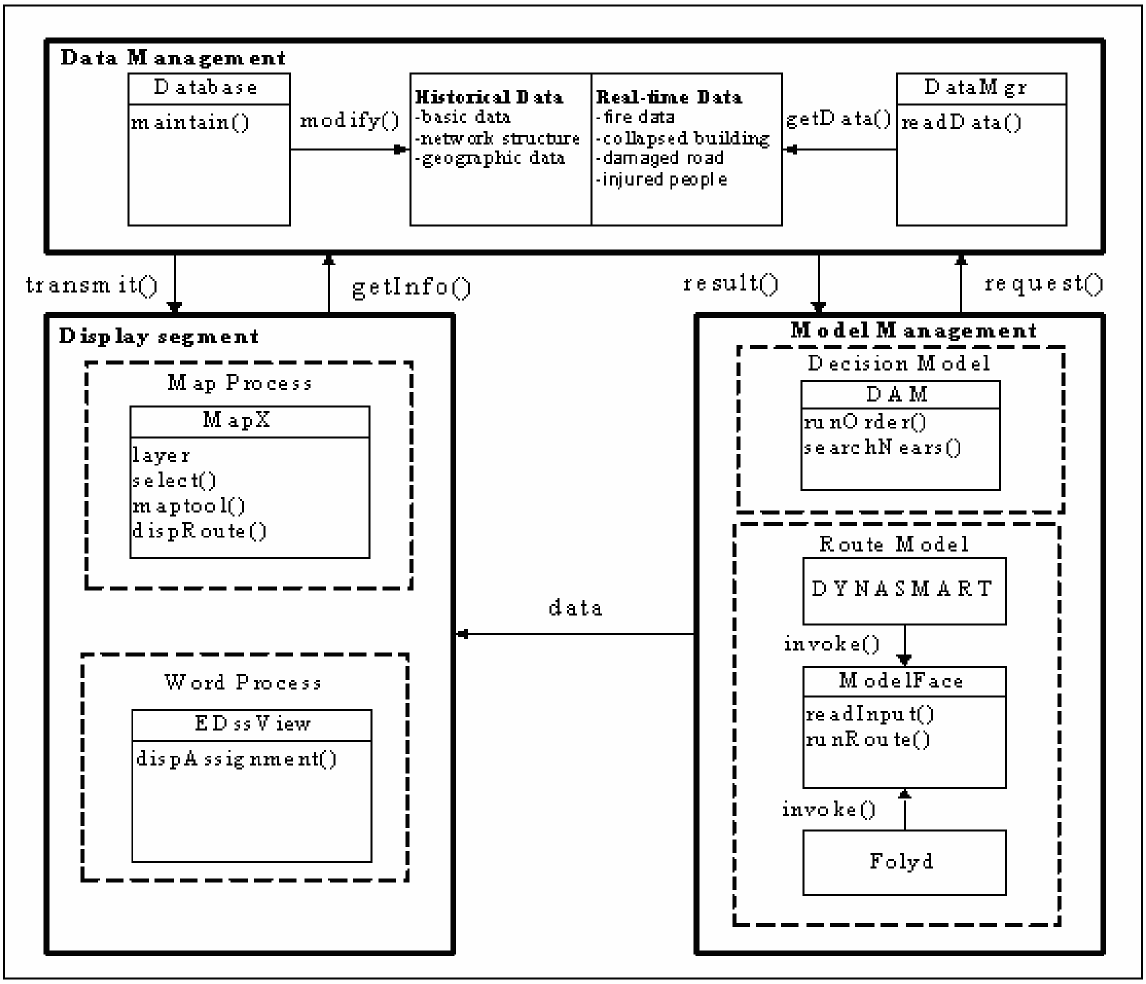


Figure 8. Program Structure for SDSS under Emergency Management

In order to illustrate the computational performance of the proposed framework, a GIS-T for Taichung City Area is constructed. There are 4291 nodes and 12664 links in the GIS-T. A high-end PC workstation (P4-1.7G, 768MB RAM) is used to develop and illustrate the SDSS. All applications and algorithms are written in C++, and complied by Microsoft Visual C++ 6.0. The user-interface, as illustrated in Figure 9, provides several functions. These functions are described in the following:

(1) File: provides file-related management functions.

(2) Edit: provides editing functions for digitized map and/or data.

(3) View: provides different views to resources, maps, and resources.

(4) Option: provides options for system operations.

(5) Decision: provides decision rules as well as weights for decision makers.

(6) Help: provides basic explanations and help functions.

The system is illustrated through two figures, and these figures are described as follows:

(1) Available resources and associated location information

As shown in Figure 9, once the type of disasters is identified, the SDSS displays available resources and locations onto the GIS map. Several layers in the SDSS include population, administration range (or location, boundary), and other fixed facilities, such as signals, can also be shown on the map.

(2) Decision and Associated Route Display

As shown in Figure 10, the system displays possible alternatives for decision making according to numerical weights of decision criteria. Possible routes are also generated and displayed.

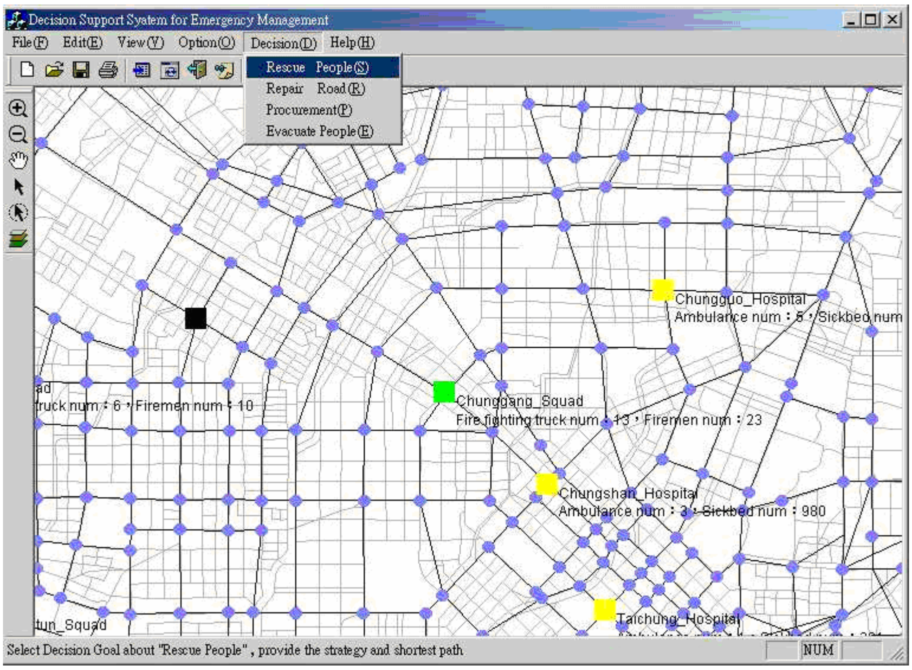


Figure 9. Available Resources and Associated Information

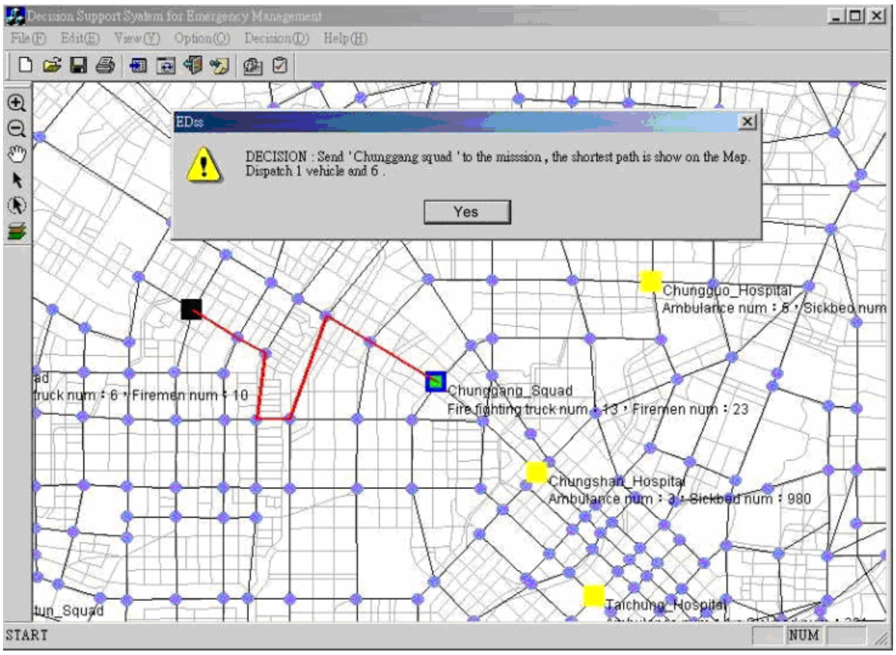


Figure 10. Decision Generated Through the SDSS

CONCLUSIONS

In this paper, an object-oriented approach is adopted to develop a SDSS for emergency management, especially for earthquake. Numerical experiments based on a city network are illustrated. The major advantage of the SDSS is its flexibility and extendibility to incorporate other objects. The next stage is to develop a distributed CORBA-based SDSS, which can be managed and accessed through the web interface. CORBA object technology could be deployed in the communication layer to achieve efficient, transparent, and scalable communication between the components involved. Several performance indices, including execution time, memory efficiency, and communication speed, might be considered to evaluate the proposed architecture.

ACKNOWLEDGEMENTS

This paper is based on work partially supported by National Science Council, Taiwan, R.O.C. under the projects NSC 91-2211-E-324-023- and NSC90-2211-E-035-017. The author is grateful to Hwei-Ru Hsiao for her help in preparing the artwork of this paper. Of course, the authors are solely responsible for the contents of this paper.

REFERENCES

[1] B. Bauer, J.P. Muller, and J. Odell, An extension of UML by protocols for multiagent interaction, International Conference on MultiAgent Systems (2000) 207-214.

[2] G. Booch, J. Rumbaugh, and I. Jacobson, The Unified Modeling Language: User Guide (Addison-Wesley, MA, 1999).

[3] H. Chabrol and D. Sarramia, Object Oriented Methodology Based on UML for Urban Traffic System Modeling, Third International Conference on the Unified Modeling Language (UML2000) (2000) 425-439.

[4] H.G. Chen and D. Sinha, An Inventory Decision Support System Using the Object-Oriented Approach, Computer Ops Res. Vol. 23, No. 2 (1996) 153-170.

[5] M.D. Crossland, B.E. Wynne, and W.C. Perkins, Spatial Decision Support System An overview of technology and a test of efficacy, Decision Support Systems Vol. 14, No. 3 (1995) 219-235.

[6] M. Fowler and K. Scott, UML Distilled: A Brief Guide to the Standard Object Modeling Language, Addison Wesley Longman, Inc. (2000).

[7] T.Y. Hu, Dynamic Analysis of Network Flows Under Advanced Information and Control Systems, Ph.D. Dissertation, The University of Texas at Austin (1995).

[8] P.B. Kruchten, The 4+1 View Model of Architecture, Software IEEE (1995) 42-50.

[9] J. Lin, Y. Chu, and H. Tsao, An Architecture and Development Method for Networked Human Resource Information Systems, International Journal of Computers and Their Applications (2002) 38-53.

[10] H. S. Mahmassani, Development of DYNASMART-X Software for Real-Time Dynamic Traffic Assignment, Technical Report ST067-85-TASK E, Center for Transportation, The University of Texas at Austin (1998).

[11] MapInfo Corporation, Map X Reference Manual, Version 3.5 (Troy, New York, 1999).

[12] S. Park, V. Sugumaran, and S. Lee, An architecture-centric approach for multi-agent system development and application, Third International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems (2001) 160-169.

[13] M. Pidd, F.N. de Silva, and R.W. Eglese, A simulation model for emergency evacuation, European Journal of Operational Research 90 (1996) 413-419.

[14] T. Quatrani, Visual Modeling with Rational Rose 2000 and UML (Addison Wesley, MA, 2000).

[15] T. Rowlett, Building an Object Process Around Use Cases, JOOP (March/April 1998) 53-58.

[16] D. R. Shier, On Algorithms for Finding the K Shortest Paths in a Network, Networks, Vol. 9 (1979) 195-214.

[17] S. Tufekci, An integrated emergency management decision support system for hurricane emergencies, Safety Science 20 (1995) 39-48.

[18] H.J. Watson, G. Houdeshel, and R.K.R. JR., Building Executive Information System and other Decision Support Applications, (John Wiley & Sons Publisher, 1997).

[19] R. Willson, Impacts and Responses: Goods Movement After the Northridge Earthquake, Journal of Transportation and Statistics, Vol. 1, No. 2 (May 1998).

[20] K.G., Zografos, K.N. Androutsopoulos, and G.M. Vasilakis, A real-time decision support system for roadway network incident response logistics, Transportation Research Vol. 10C, No. 1 (2002) 1-18.

A General Analysis of Trade Data in the Context of ICT Infrastructure Variability

Dr. Vijay Vemuri
C.W. Post Campus, Long Island University
Brookville, NY 11548
Email: vvemuri@liu.edu

Dr. Shahid Siddiqi
C.W. Post Campus, Long Island University
Brookville, NY 11548
Email: ssiddiqi@liu.edu

Abstract

The chief concern of this paper was to examine the effect of ICT infrastructure on international trade. A modified version of the traditional gravity model was used to study bilateral trade. The data on 64 countries was obtained from IMF Direction of Trade statistics (1990-2000). GDPs and ICT statistics were obtained from United Nations Common Database. The extended gravity model was confirmed by the regression analysis. The main research question was answered affirmatively: The ICT infrastructure had positive effect on the trading volume between countries. Distance had a negative effect on the amount of international trade between countries. However, Internet itself had a negative effect on the volume of trade. Possible explanations could be its delayed effect on volume of trade or rate of technology diffusion or the "productivity paradox".

Keywords:

International trade, Internet, Gravity model, ICT, Globalization

INTRODUCTION

The discovery of the compass galvanized trans-oceanic trade besides leading to the discovery of America. The privatization of the Internet, it is tacitly assumed, would result in the stimulation of international trade. In other words, the availability of information is critical to risky ventures. Relative to the domestic market, international commerce falls in this category on account of the strategic challenge of' discontinuity' of environment coupled with the issue of territorial sovereignty. In this sense the history of supportive information generation, processing, consolidation and dissemination is the history of globalization. The prerequisite for the Internet to become the vehicle for such processes is the ICT (information and communication technology) infrastructure. In the paper we are directly concerned with ICT infrastructure and its effect on international trade.

BACKGROUND

The ability to exploit ICT has been found to be a strong correlate to a country's GDP per head. According to the World Bank (2004), all types of infrastructure are critical to firms while telecommunications services are particularly vital to the investment climate (129). Though transport infrastructure has acquired particular value in the context of multilaterally negotiated reduction of tariff and non-tariff barriers, this research has not considered these in the analysis. Since our concern is with macro aspects of trade and ICT, we have not looked at the continued spearheading of strategic intervention by the multinationals in this research. However, this factor has critical consequences for the globalization of foreign economies.

PRIOR RESEARCH

Empirical evidence suggests that in recent years international trade in goods and services has been on the rise. The causes for the increase are varied. These include: 1) liberalization of foreign economic policies, 2) growing international economic activity, and 3) technological innovations lowering costs of communication, collaboration, and search for trading opportunities and partners (Garrett, 2000). The latter includes the issue of increasing efficiencies in transportation of goods and information services Specifically, considerable anecdotal and trade literature suggests that privatization of the Internet is a major impetus for stimulating the growth in international trade (Freund and Weinhold, 2002). However, it is not clear if, in fact, the cheap, ubiquitous and reliable communication channels created by the universal accessibility to the Internet *per se* is sufficient for increase in international trade and consequent economic growth. Nevertheless, the linkage between telecommunications and economic activity was found to be established by Dutta (1991) in his empirical analysis of Granger causality. His time series data evaluated levels of telecommunications and economic activity for thirty countries. He concluded that causality was stronger from levels of telecommunications to economic activity relative to the opposite direction.

The traditional and popular model for explaining the international trade has been the gravity model. The simplest version of the gravity model suggests that trade between two countries is directly proportional to the product of the gross domestic product of the two countries and inversely proportional to the distance between the two countries. The intuition behind the model is that the size of both economies together determines the opportunities for trade. Distance between the two countries increases the transportation costs and introduces a wedge in opportunities for trade (McCallum, 1995). From this arises the issue of market imperfection, creating the opportunity for foreign direct investment (FDI) entry strategy for multinationals. The basic gravity model is embellished with additional factors to further explain the level of trade between two countries. The most common additional factors explaining the volume of bilateral trade include population of trading countries, proximity of countries, common language, colonial linkage, historical trade patterns, and membership in free trade areas (Freund and Weinhold, 2002). In recent years measures of ICT have been incorporated into the gravity models (Freund and Weinhold, 2004; Dutta, 2001; and Clarke 2002).

The effect of Internet on international trade cannot be studied in isolation. We incorporate various measures of ICT infrastructure into our model to reevaluate the effect of Internet on international trade. Our additional measures of ICT infrastructure include number of telephones, number of telephones for 100 inhabitants, number of computers, number of computers per 100 inhabitants, number of Internet users, and number of internet users per 100 inhabitants.

RESEARCH QUESTIONS

The prime focus of this research is to investigate the effect of privatization of the Internet and ICT infrastructure on the levels of bilateral trade between countries.

THEORETICAL FRAMEWORK

We use a modified gravity model to answer our research questions. The estimation model we propose is:

$$\log(\text{trade}_{ijt}) = \beta_0 + \beta_1 \log(\text{gdp}_{it} \times \text{gdp}_{jt}) + \beta_2 \log(\text{pop}_{it} \times \text{pop}_{jt}) + \beta_3 \log(\text{distance}_{ij}) + \beta_4 \text{Min}(\text{phones}_{it}, \text{phones}_{jt}) + \beta_5 \text{Min}(\text{computers}_{it}, \text{computers}_{jt}) + \beta_6 (\text{Internet Users}_{it}, \text{Internet Users}_{jt}) + \beta_7 \text{Internet}_t + \varepsilon_{ijt},$$

where, trade_{ijt} is the total of exports + imports between countries i and j in period t.
gdp_{it} is the gross domestic product of country i in year t.
pop_{it} is the population of country i in year t.
distance_{ij} is the distance in miles between country i and j. To avoid mathematical complications (to avoid logarithm of 0), distance between neighboring countries is fixed at 1.
phone_{it} is the number of telephones in country i in year t.
computers_{it} is the number of computers in of country i in year t.
$\text{Internet Users}_{it}$ is the number of Internet users in county i in year t.
Internet_t is a dummy variable denoting the accessibility of Internet for conducting commercial trade in year t. The variable takes a value of 0 for 1994 and before, and 1 otherwise.
ε_{ijt} is the residual of regression estimation.

Minimum function variables used with phones, Internet users and computers reflect the fixed proportions of production technologies observed in communications. Fixed proportions technology with "L"-shaped isoquants are relevant for determining the number of simultaneous communication channels that can be established at any given time. Number of simultaneous phone connections that can be established between two countries with one million and 100 million phones is at most one million – smaller of the two numbers. A priori, it is likely that number of telephones, computers and Internet users are highly correlated. In this paper, our intention is not to include all of the variables but to build a parsimonious model representing ICT infrastructure.

To see if the level of penetration of ICT is relevant in determining international trade, the estimation equation is run again using phones, computers, and Internet users per 100 inhabitants in each county. In addition to the data mentioned above, we included fixed effects for the country-pairs. We also ran regressions to see if the fixed effects model further explains the relationship between trade and ICT infrastructure of trading partners.

DATA COLLECTION

The trade data for the above estimation equation is obtained from IMF Direction of Trade statistics published annually. We obtained import and export statistics for 64 countries for the years between 1990 and 2000. Our sample of countries is 64 consists of 21 developed, 20 middle-income and 23 developing economies. Their 2000 GDPs range from 9.76 trillion (US) to 3.95 billion (Nicaragua) US dollars. gdp, and ICT statistics are obtained from United Nations Common Database, which in turn obtains its data from 30 specialized international data sources. Distances between any pair of countries is estimated from the world map. NSFNET backbone services' Acceptable Use Policy prohibited the use of NSFNET for purposes other than research and education. Retirement of NSFNET in April 30, 1995 was considered as the time of Internet privatization (Kesavan and Shah, 2001). Hence, the Internet variable is coded as 0 for the years 1990 to 1994 and as 1 for 1995 to 2000.

The data we collected is extensive. The sample of 64 countries resulted in 2,016 county pairs and for each of the 11 years studies resulted in 22,176 observations of dependent variable. About another 250,000 observations are obtained for the independent variable.

RESULTS AND DISCUSSION

Table 1 shows the results of our regressions. The regression results show that our model has significant explanatory power.

Our extended model confirms the basic gravity model which explains trade between countries**.** GDP and distance significantly affect the volume of trade, though in opposite directions. Population, however, has negative effect on the volume of trade. A version of the gravity model predicts that population size of the trading countries has positive impact on the level of trade. Our results however, do not confirm it. This could be an artifact of our model in the sense that we required the coefficient of logarithm of populations to be the same for both countries. Our estimation equation may have to be expanded to further explain the effect of population on trade. A variety of factors could be at play, for instance, age distribution and urbanization. It should also be noted that no attempt has been made to include literacy in general or technical literacy or technical education in particular.

Our main research question is answered affirmatively: The ICT infrastructure has positive effect on the trading volume between countries. This result persists when various surrogates of ICT infrastructure are used. Further, use of number of computers, telephone, or Internet users representing ICT infrastructure or the alternative use of their corresponding levels per 100 inhabitants does not change our conclusion.

A priori, we expected that the dummy variable representing the use of Internet for commercial purpose would have significant effect on trade. Surprisingly, this variable has negative effect on the volume of trade. We do not have a complete explanation for it, except that we hypothesize that the Internet use for trading may have delayed effect on volume of trade. The question of technology diffusion could be relevant (Keller, 2001). The "productivity paradox" linking high levels of computer use by businesses and lack of observable improvements in productivity at the macro level may have much to explain here as well (Brynjolfsson, Erik and Lorin M. Hitt (1998). Slow adjustment to new technology, changes in business processes, learning, slow cultivation of trust among trading partners can all be instrumental in non observing "before and after" Internet effect. With passage of time, estimation equations may reflect the slow adjustments to trading opportunities afforded by the Internet connections. It may well be that depending upon the stage of economic development and the distribution of relevant variables like the extent and type of educational context, there may be different degrees of lagged effect on different countries.

Every one of the regressions reported here show that the coefficient for the distance is significant and negative suggesting that distance has a negative effect on the amount of international trade between countries. This suggests that Ghemawat's (2001) argument is valid in a much broader context. We were concerned more with distance in the sense of geography rather than in the sense of other factors like culture or mindset. This does not, however, suggest that distance hampers communication between trading partners, or "world has not shrunk". The coefficient of regression merely suggests that in the time periods studied, distance still appear to affect trading activity.

CONLCUSIONS

Our chief concern in this paper was to examine the effect of ICT infrastructure on international trade. For this purpose we used a modified version of the traditional gravity model. In order to investigate the bilateral trade effect of privatization of the Internet and ICT infrastructure we obtained data on 64 countries from IMF Direction of Trade statistics (1990-2000). We also collected GDPs and ICT statistics from United Nations Common Database. Our extended gravity model was confirmed by the regression analysis. The main research question was answered affirmatively: Indeed the ICT infrastructure had positive effect on the trading volume between countries. Distance, in the traditional geographic sense rather than in the sense of other variables like culture, had a negative effect on the amount of international trade between countries. However, Internet *per se* had a negative effect on the volume of trade. Possible explanations could be its delayed effect on volume of trade or rate of technology diffusion or the "productivity paradox". Another anomaly was the population factor which, contrary to other studies, showed a negative effect on the volume of trade. This could be an artifact of our model in the sense that we required the coefficient of logarithm of populations to be the same for both countries. Our estimation equation may have to be expanded to further explain the effect of population on trade. A variety of factors could be at play, for instance, age distribution and urbanization. It should also be noted that no attempt has been made to include literacy in general or technical literacy or technical education in particular. In future studies we may need to add other refinements to our parsimonious model.

Table 1: Regression Results

Dependent Variable: Log ($Trade_{ij}$)						
Regression No.	1	2	3	4	5	6
Constant	-39.882	-39.887	-39.692	-40.299	-39.797	-39.9597
	(-109.969)*	(55.29)*	(-163.216)	(-79.002)*	(-110.391)*	(-55.83)*
log (gdp_i, gdp_j)	0.9552	1.002	0.880202	0.926997	0.90322	0.940861
	(125.349)*	(107.26) *	-182.865*	(148.28)*	(-102.276)*	(91.548)*
log (Distance)	-0.111	-0.1409	-0.110116	-0.1393	-0.11203	-0.14208
	(-26.917)*	(25.23)*	(-36.462)*	(-33.758)	(-27.26)*	(-25.656)*
Internet			-0.186757	-0.20007		
			(-8.985)*	(-10.255)		
log(pop_i, pop_j)	-0.10699	-0.12686			-0.03573	-0.03684
	(-13.486)*	(14.30)*			(-3.558)*	(3.358)*
Min(Computers per 100)			0.040392	0.039681		
			(27.258)*	(27.678)*		
Min (Internet users)	8.07E-08	7.17E-08			3.31E-08	2.01E-08
	(7.069)*	(6.72)*			(2.743)*	(1.790)
Min (Telephone)					0.007	0.008868
					(11.46)*	(13.722)*
Fixed Effects	Not Included	Included	Not Included	Included	Not Included	Included
R^2	0.684	0.735	0.691	0.733	0.689	0.74

* indicates statistical significance at 1%.

REFERENCES

Brynjolfsson, E. and Hitt, L.M. (1998). Beyond the Productivity Paradox. *Communications of the ACM*, 41(8), 49-55.

Clarke, G.R.G. (2002). Does Internet Connectivity Affect Export Performance?. *Discussion Paper No. 200274*, World Institute for Development Economics Research, United Nations University

Clarke, G.R.G., and Wallstern, S.J. (2004). Has the Internet Increased Trade? Evidence from Industrial and Developing Countries. *Policy Research Working Paper Series 3286*, the World Bank.

Dutta, A. (2001). Telecommunications and Economic Activity: An Analysis of Granger Causality. *Journal of Management Information Systems.* 17(4), 71-95.

Freund, C., Weinhold, D. (2002). The Internet and International Trade in Services. *The Economic Review,* 92(2), 236-240.

Freund, C., Weinhold, D. (2004). On the Effect of the Internet on International Trade. *Journal of International Economics* 62(1), 171-189.

Garrett, G. (2000). The Causes of Globalization. *Comparative Political Studies,* 33, 941-991.

Ghemawat, P. (2001). Distance Still Matters, The Hard Reality of Global Expansion. *Harvard Business Review,* 79(8), 3-11.

Keller, W. (2001). International Technology Diffusion. *NBER Working Paper No. 8573*, October.

Kesan, J. P. and Shah, R. C. (2001). Fool Us Once Shame on You - Fool Us Twice Shame on Us: What We Can Learn from the Privatizations of the Internet Backbone Network and the Domain Name System. *Washington University Law Quarterly,* 79(1), 89-220.

McCallum, J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *American Economic Review,* 85, 615-623.

Rauch, J.E. (1999). Networks Versus Markets in International Trade. *Journal of International Economics*, 48. 7-35.

Roller, L-H and Waverman, L. (2001). Telecommunications Infrastructure and Economic Development. *American Economic Review*. 91(44), 909-23.

World Bank (2004). A Better Investment Climate for Everyone, *World Development Report*, Washington, D.C., 2004.

An HMM-based Approach for Mining User Moving Patterns in Cellular Communications

Ming-Chuan Hung, Jain-Chung Chang, Don-Lin Yang and Jungpin Wu*
Department of Information Engineering and Computer Science
*Department of Statistics
Feng Chia University, Taichung, Taiwan
Email: **mchong@fcu.edu.tw**, jain@soft.iecs.fcu.edu.tw, dlyang@fcu.edu.tw, cwu@fcu.edu.tw

Abstract

With the advance of information technology and wireless network, an increasing number of mobile users rely heavily on the cellular communication to carry out their daily activities. However, due to the limitation of present mobile network architecture, a constantly moving mobile user may encounter significant delay when requesting data or services. In this paper, we study the process of knowledge discovery to find the most frequent moving patterns of mobile users for further analysis. Using proposed methods, we can feedback the discovered results to the service providers of cellular network for improvement. We use data mining techniques and Hidden Markov probability model to mine the frequent moving nodes of users. Then, they were merged to find out the most frequent moving patterns. Our approach is scalable and is able to process data incrementally. When new moving records are collected, previous results can be exploited to speedup the process of mining new moving patterns.

Keywords

Personal communication servers system, Hidden Markov model, Data mining, User moving pattern, Association rule

1. INTRODUCTION

With the rapid computerization of businesses and organizations, a huge amount of data has been collected in databases. While these are historical data containing information from the past, we would like to use them for future prediction. At the same time, users expect more useful information coming out of these databases. As a result, traditional ad hoc mixtures of statistical techniques and data management tools are no longer adequate for analyzing this vast collection of data. Data mining is defined as the extraction of implicit, previously unknown, and potentially useful patterns from data (Bayardo R.J., 1998; Dunkel B. & Soparkar N., 1999; Han J. & Kamber M., 2001). Today data mining (or knowledge discovery in databases, KDD in short) has emerged as a growing field of multidisciplinary research for discovering interesting, useful knowledge from large databases. For example, a marketing manager is not going to be satisfied with a simple listing of sale contacts without detailed information about customers' past purchases as well as prediction of future purchases. As a result, new tools are required.

With the increasing demand of portable computers, such as mobile phones, tablet computers, palmtops, users can carry their computing power around anywhere they want. In addition to cellular phones, these portable computers can be equipped with wireless communication devices that enable users to access global data services from any location. Today, the mobile devices are beginning to support many applications such as multimedia and World Wide Web. Users are able to surf the Internet and read their emails when they are travelling. And there are more and more users using different information systems via wireless communication. Such information systems provide investment and finance services like stock trading, banking, and wireless conferencing. Thus, the mobile user can use these information systems to access data anywhere at any time by wireless communication (Bernd B. & Bennington B., 1996; Davies N., Blair G.S. & Cheverst K., 1996). However, mobile devices equipped with wireless communication facility do not always connect to the network due to the high cost of wireless communication or the unavailability of the cellular network. The bandwidth is a limitation on the wireless communication medium and induces rather high communication cost for mobile users. Beside cost-performance consideration, a mobile computing system is usually of a distributed server architecture (Elmagarmid A., Jain J., & Furukawa T., 1995; Krishnakumar N. & Jain R., 1997), in which a service area referring to the converge area where the server can provide services to mobile users. If we assume there is a mobile user in the cellular network, and he/she usually moves around and changes the location continuously. After the user made a request for data or services from an information system, the cellular network must find out his/her location first. Second, the cellular network provides services from the information system to the user. In this scenario, the cellular network works with low efficiency because of the latency in delivering the result to the user in the right location.

In the past, there are many researches on World Wide Web and cellular network using traditional data mining methods. Since there is no significant improvement yet, we try to combine the idea of data mining and the technique of Hidden Markov model to develop more efficient methods for mining user moving patterns. Since the architecture of a mobile computing system is distributed in nature, finding user moving patterns is helpful because it is able to improve the execution performance of servers and facilitate the location lookup of mobile users. We focus on finding user moving patterns in the cellular network, and the aim of our work is to decrease the latency and cost.

2. RELATED WORK

In this section, we first describe the framework of the personal communication servers system (PCS) in detail. From PCS we can obtain required data for mining user moving patterns. After that, we introduce some techniques of data mining and the Hidden Markov model.

Today, most systems that use the personal communication servers system are the Global system for cellular communication of cellular and the digital communication system-1800. Thereinafter, we describe the framework and services of personal communication servers system. The network of personal communication servers system is divided into two sections, the first section is switching system (SS). The switching system provides the function of exchange of communication lines and the function of roaming management. Another section is the base station system (BBS) which connects cellular phones to the switching system, as shown in Figure 1.

The switching system consists of mobile switch center (MSC), visitor location register (VLR), home location register (HLR), authentication center (AUC), and equipment identity register (EIR). The mobile switch center provides exchanging of communication lines, controlling the subsystem of stations, and connecting with other networks. The visitor location register stores the user information of cellular phones roaming in the network and supplies the communication services.

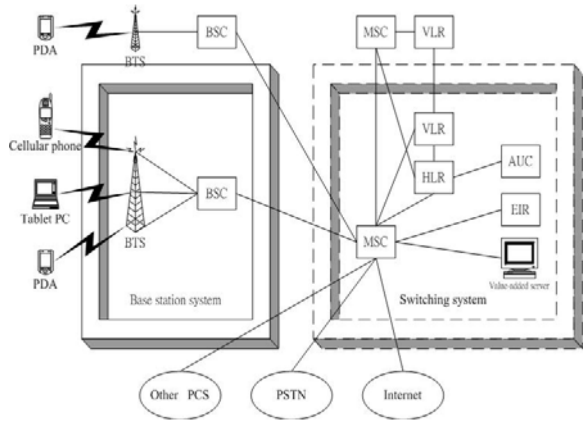


Figure 1. The framework of PCS network

The home location register stores the information of users and the location of cellular phones. All of the users must register in the static HLR. The authentication center and equipment identity register are responsible for the process of identifying users and cellular phones. The base station system (BBS) consists of base station controller (BSC) and base transceiver station (BTS). The base station system controls the base transceiver station, distributes the channel of network, exchanges the lines of subsystem, and connects to the MSC. The base transceiver station connects cellular phones and acts as the center of maintenance and accounting.

The first technique of data mining used in this research is association rule (Agrawal R., Imilienski T., & Swami A., 1993; Agrawal R. & Srikant R., 1994), which indicates that there exist relationships between itemsets in databases, like supermarket transactions. Time series is a set of attribute values over a period of time, and time series data may be continuous or discrete. Typical data mining applications for time series include determining the similarity between two different time series and predicting future values for an attribute. Time series analysis is viewed as finding patterns from the data and predicting future value. A sequence is viewed as an ordered list of attribute values from any domain. In fact, the terms sequence and time series are often used interchangeably. The only difference might be that the sequence may not have explicit relationship with time (Han J. & Kamber M., 2001).

We describe two algorithms: one is Apriori-like algorithm and the other is the Scal algorithm of statistics. For the former, the original idea of LM algorithm (see Peng W.C. & Chen M.S., 2000) comes from the MF (maximal forward) algorithm, which can find the maximal user forward paths but not backward paths (Peng W.C. & Chen M.S., 2000; Liu G.Y. & Maguire G.Q., 1995). The LM algorithm improves the MF algorithm by including backward moving paths. For mining user's backward moving paths the LM algorithm uses the MM (Maximal moving sequences) algorithm to extend user's moving paths. Using the MM approach, the LM algorithm counts user's moving patterns from the user maximal moving path. If the number of nodes of the path is more than the threshold specified by the user, then these nodes become the user moving patterns. Different from the Apriori-like algorithms, the Scal algorithm uses statistical method. The Scal algorithm uses the relation of groups, and it can be used to mine user moving patterns. But if the subsequences which divided by user moving patterns are not correct, we can not obtain the user moving patterns correctly. So the divide phase still needs help of experts.

We will use the Hidden Markov model (HMM) which differs from Markov model. A basic Markov model is a model with each state corresponding to an observable event and its state transition probability where the probability depends on the current state and predecessor state. Extending from Markov model by considering more conditions of observable events, the HMM can apply to more complex processes including speech recognition and computational gene finding, and it is more feasible than Markov model. A generalized Hidden Markov Model consists of a finite set of states, an alphabet of output symbols, a set of state transition probabilities and a set of emission probabilities. The emission probabilities specify the distribution of output symbols that may be emitted from each state. Therefore in the Hidden model, there are two stochastic processes; the process of moving between states and the process of emitting an output sequence. The sequence of state transitions is a hidden process and is observed through the sequence of emitted symbols. For a formal definition of HMM, please refer to an HMM tutorial by Rabiner L. (1989).

3. PROPOSED METHOD: MUMP

As stated above, we found that most previous works have focused on counting the support of nodes in databases. In other words, they just count of appearing nodes without taking the probability of user traveling into account (Chen M.S., Park J.S., & Yu P.S., 1998). Previous studies also did not consider the architecture of cellular network; therefore Hidden Markov model was applied in our work. In our proposed method of Mining User Moving Patterns (MUMP), the user moving records are preprocessed by time period in the first step. This is intended to focus on the users who move to some cellular nodes more frequently in the same time period. Second, a Hidden Markov model is built to calculate the probability of user traveling in the cellular network, and it is used to predict the node that the user might move to next. After calculating these probabilities, we can find all frequent pairs of visited-nodes which are like large 2-itemsets in the association rule mining by using the Hidden Markov model. The user moving patterns can therefore be obtained by merging frequent visited-nodes. Additionally, neighbor nodes and connectivity are defined below to make sure the relation of connection in cellular network.

Definition 1: neighbor node

In a cellular network, a node *x* can connect to one or more nodes based on the network architecture. A neighbor node of *x* is any node that can connect to *x*.

Definition 2: connectivity

If one node can connect to another node in a cellular network, we say that there exists connectivity between these two nodes.

For example, in the network shown in Figure 2, node A connects to nodes B, E, and F. Therefore, we say nodes B, E, and F are neighbor nodes of node A by *Definition 1*. Since node A connects to node B, we say node A and node B have connective relation by *Definition 2*.

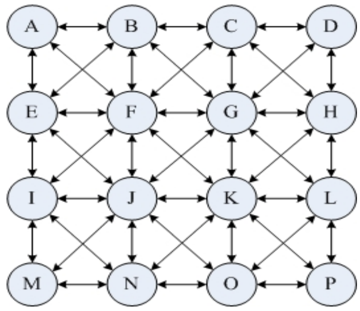


Figure 2. The cellular network

Now we define three terms: moving records, moving nodes, and moving patterns. Moving records are the logs that record the movement of users in a cellular network. Moving nodes are the nodes that users move to. And moving patterns are the patterns that users frequently move and these patterns consist of moving nodes.

Next, we illustrate the three steps of constructing the Hidden Markov model: (1) construct the initial state distribution matrix, (2) construct the state transition probability matrix, and (3) construct the observation probability matrix.

For the initial state distribution matrix, we can find the probability of user being at the initial starting node by using the techniques of statistics or consulting the experts of cellular network. We use a simple example to illustrate our proposed method MUMP. In the beginning, the cellular network architecture is assumed to be the same as the one in Figure 2. The network has 16 nodes and each node connects to its neighbor nodes. Therefore, there are 16 initial probabilities of user starting nodes. We assume the initial probability of each user starting node is the same, so the probability for each node is 1/16 = 6.25%.

After finding the probability of initial user starting node, we construct the state transition probability matrix. This probability matrix comes from user moving records. We use statistical methods to compute the probability with the formula:

$$P_o^n = X / Y$$

where X is the total number of old-nodes connecting to new-node and Y is the total number of old-nodes connection to other nodes.

For example, given 10 user moving records in Table 1, we count the number of old-nodes connecting to new-nodes. Table 1 shows the user moving records and the above mentioned relation while Table 2 displays their probabilities in the cellular network. Take node A for example, from Table 2 node A has ten connections to other nodes, where node A connects to node B three times, node E two times, and node F five times. Therefore, we can compute the probability of an old-node connecting to a new-node P_A^B : meaning connecting node A to node B with the probability of 3/10. Similarly, P_A^E is the probability of connecting node A to node E with a value of 2/10 and the value of P_A^F for connecting node A to node F is 5/10.

Table 1. Example of user moving records

Record	Moving pattern
1	AFKP
2	AFKLP
3	AFKOP
4	AFGKP
5	AFJKP
6	ABFGKP
7	AEFKP
8	ABCDHLP
9	AEIMNOP
10	ABCDGJMNOP

Old node	New node, the number of times		
A	B, 3	E, 2	F, 5
B	C, 2	F, 1	
C	D, 2		
D	G, 1	H, 1	
E	F, 1	I, 1	
F	G, 3	J, 1	K, 3
G	J, 1	K, 3	
H	L, 1		
I	M, 1		
J	M, 1		
K	L, 1	O, 1	P, 5
L	P, 2		
M	N, 2		
N	O, 2		
O	P, 3		
P			

The observation probability matrix is like the transition probability matrix for the relation of connecting old-nodes to new-nodes, except it is established according to the cellular network architecture and the direction in the cellular network is also taken into consideration. For example, in Figure 2, we find node A connects to node B, the direction is *right*. So the observation probability of node A connecting to node B is 1/3 (33%). The nodes E and F are similar to node B. The observation probability of node A connecting to node E is also 1/3 (33%), the direction is *down*. And the same for node A connecting to node F (33%), the direction is *right-down*. For simplification, we assume uniformity and show the transition probability matrix in Table 3.

Table 2. Show all of these probabilities of old-nodes connecting to new-nodes

Old node	Connect to new node	Probability (%)	Old node	Connect to new node	Probability (%)
A	AB	30	G	GJ	25
	AE	20		GK	75
	AF	50	H	HL	100
B	BC	67	I	IM	100
	BF	33	J	JM	100
C	CD	100	K	KL	14
D	DG	50		KO	14
	DH	50		KP	72
E	EF	50	L	LP	100
	EI	50	M	MN	100
F	FG	43	N	NO	100
	FJ	14	O	OP	100
	FK	43			

Table 3. Example transition probability matrix

Old node	Connect to new node	Probability (%)	Old node	Connect to new node	Probability (%)	Old node	Connect to new node	Probability (%)
A	AB	33	F	FK	12.5	K	KG	12.5
	AE	33	G	GB	12.5		KH	12.5
	AF	33		GC	12.5		KJ	12.5
B	BA	20		GD	12.5		KL	12.5
	BC	20		GF	12.5		KN	12.5
	BE	20		GH	12.5		JO	12.5
	BF	20		GJ	12.5		KP	12.5
	BG	20		GK	12.5	L	LG	20
C	CB	20		GL	12.5		JH	20
	CD	20	H	HC	20		LK	20
	CF	20		HD	20		LO	20
	CG	20		HG	20		LP	20
	CH	20		HK	20	M	MI	33
D	DC	33		HL	20		MJ	33
	DG	33	I	IE	20		MN	33
	DH	33		IF	20	N	NI	20
E	EA	20		IJ	20		NJ	20
	EB	20		IM	20		NK	20
	EF	20		IN	20		NM	20
	EI	20	J	JE	12.5		NO	20
	EJ	20		JF	12.5	O	OJ	20
F	FA	12.5		JG	12.5		OK	20
	FB	12.5		JI	12.5		OL	20
	FC	12.5		JK	12.5		ON	20
	FE	12.5		JM	12.5		OP	20
	FG	12.5		JN	12.5	P	PK	33
	FI	12.5		JO	12.5		PL	33
	FJ	12.5	K	KF	12.5		PO	33

After setting the three matrixes, the initial state distribution matrix, the state transition probability matrix and the observation probability matrix, we calculate the probability of mobile user's moving node by using the formula $\alpha * \beta * \pi_i * a_i * b_i$. This formula of probability differs from the one used by HMM model. The factor **α** denotes the probability of a node being close to some special locations. If we know some nodes of a cellular network are near any special location, like department stores, we can set this factor α larger than one. And the factor β stands for the probability of a node being affected by a special event. However, in our study we don't discuss these factors because we don't have the real data from cellular networks and event providers. In our experiments, α and β were set to one. If one wants to estimate α and β, we can suggest the use of statistics to set these parameters. For example, we can count the number of nodes in a cellular network and identify the nodes close to stores or buildings of public interests. Then we can set the value of α parameter. Finally, π_i is the mean element of the initial state distribution matrix, a_i is the mean element of the state transition probability matrix, and b_i is the mean element of the observation probability matrix.

To cite an instance from Table 1, we compute the probability of a mobile user moving from node A to node B. First, the mobile user starts from node A and the probability of initial state distribution matrix is 1/16 (6.25%). The mobile user indeed moves from node A to node B, so the probability of state transition probability matrix is 3/10 (30%). The cellular network architecture can provide the probability of observation probability matrix connecting node A to node B, which is 1/3 (33%). Therefore the result probability of the mobile user moving from A node to B node is 1*1*1/16*3/10*1/3 (0.625%). In this way, we can calculate all the probabilities of mobile user moving nodes, as shown in Table 4.

After selecting the threshold from the result of calculating probabilities, we can find the frequent moving nodes for those with probabilities over the threshold. We obtain the frequent visited-nodes in mobile user moving records. Since finding frequent visited-nodes is only concerned with the relation of connecting the old-node to the new-node, there are too many numbers of these relations.

Therefore, we need to merge these nodes to become user moving patterns. This merging phase is different from the Apriori algorithm, because the frequent visited-nodes must be considered with characteristics of cellular network: neighboring and connectivity. When we merge these frequent visited-nodes, we must follow the principle of merge operation.

Principle of merge operation

We use *fn* to denote frequent visited-nodes, and each *fn* has two nodes denoted as <$node_1$, $node_2$>. If there are more than two *fn'*s, we can use the following principle to merge these frequent nodes. Given *fn1*=<$node_{1,1}$, $node_{1,2}$> and the *fn2*=<$node_{2,1}$, $node_{2,2}$>, if $node_{1,2}$ of *fn1* is the same with $node_{2,1}$ of *fn2*, we merge *fn1* and *fn2* to become the pattern <$node_{1,1}$, $node_{1,2}$, $node_{2,2}$>. We apply this principle to merge all of frequent nodes to obtain the user moving patterns.

If the threshold was set to 0.005 in Table 4, we can find the frequent visited-nodes: AB, AF, BC, CD, DG, DH, EF, EI,GK, HL, IM, JM, KP, LP, MN, NO and OP. Using the above principle, we merge the frequent visited-nodes one at a time. We can find all the moving patterns (paths) ABCDGKP, AF, DHLP, EF, EIMNOP, and JM. After finding the user moving patterns, we can use these patterns to predict the next nodes where the user may move to.

Table 4. Probabilities of mobile user moving nodes

Old node	Moving path	The probability of initial state distribution (%)	The probability of state transition (%)	The probability of observation (%)	The probability of moving path (%)
A	AB	6.25	30	33	0.61875
	AE	6.25	20	33	0.4125
	AF	6.25	50	33	1.03125
B	BC	6.25	67	20	0.8375
	BF	6.25	33	20	0.4125
C	CD	6.25	100	20	1.25
D	DG	6.25	50	33	1.03125
	DH	6.25	50	33	1.03125
E	EF	6.25	50	20	0.625
	EI	6.25	50	20	0.625
F	FG	6.25	43	12.5	0.3359375
	FJ	6.25	14	12.5	0.109375
	FK	6.25	43	12.5	0.3359375
G	GJ	6.25	25	12.5	0.1953125
	GK	6.25	75	12.5	0.5889375
H	HL	6.25	100	20	1.25
I	IM	6.25	100	20	1.25
J	JM	6.25	100	12.5	0.78125
K	KL	6.25	14	12.5	0.109375
	KO	6.25	14	12.5	0.109375
	KP	6.25	72	12.5	0.5625
L	LP	6.25	100	20	1.25
M	MN	6.25	100	33	2.0625
N	NO	6.25	100	20	1.25
O	OP	6.25	100	20	1.25

Because not all services or data are readily available at every node in a cellular network, the system must locate the node where the requester resides such that proper delivery can be made quickly enough. This can be accomplished by predicting the next node a user is going to move to. When user moving patterns are available, we can use these patterns to predict these next nodes. For example, given patterns ABCDGKP, AF, DHLP, EF, EIMNOP, and JM, if a user requests services or data at node A, we can provide services or data to nodes B and F in advance. Using such user moving patterns, we can reduce latency and lower cost. Therefore, our method can help the cellular network to provide more efficient and useful services.

With the use of the Hidden Markov Model, it can continuously adjust itself. After locating all the user moving patterns, we can use them to adjust the observation probability matrix. Given the new user moving records, we would like to retain the discovered information of mobile user moving patterns.

To get better result, we apply a weighted parameter to adjust the observation probability matrix. Keeping adjusting the observation probability matrix can be viewed as the continuous learning of the Hidden Markov model.

We take the above mentioned example for a detail explanation. When we obtain the user moving patterns ABCDGKP, AF, DHLP, EF, EIMNOP, and JM, we increase the probabilities of AB, BC, CD, DG, GK, KP, AF DH, HL, LP, EF, EI, IM, MN, NO, and OP and decrease the probabilities of other nodes in the observation probability matrix. By using this weighted parameter to adjust the probability of a node in the observation probability matrix, we can say that we add intelligence to our method. When there are new user records, our method can deal with it intelligently.

However, there are many special situations in today's cellular network that can provide more flexible services. With the architecture of current cellular network, it is unable to handle this problem very well. A possible way to solve the problem is using the register operation. After entering a cellular network, a user must register to obtain services from the system which can deal with special situations. During registering, we can acquire user data in terms of user location and requested information. Then we can use these data to provide users with requested services more efficiently.

4. EXPERIMENTS

We have done many experiments to show that the proposed MUMP method is feasible and effective. According to the characteristics of the assumed cellular network architecture, we run our experiments with different parameters including probability, time, number of users, and hit ratio. We code the programs by using Visual C++ and .NET. All the experiments were performed on a personal computer of Intel Pentium 4 processor with a clock rate of 2.4AGHz and 512MB DDR266MHz main memory.

A Simulation of user records

Because of the consideration of business interests and trade secrete, real user moving records and cellular network architectures are difficult to obtain. Therefore, we use simulated data in our experiments. First, the assumed cellular network architecture looks like the one in Figure 2, which has 4×4 nodes. In this cellular network, each node can connect to any of the other nodes, thus a user can move between any two nodes in the cellular network. Table 5 depicts the parameters used in our experiments. Under the parameter settings of $|D|=10$, $|T|=8$, and $|N|=16$, we can generate 10 user records with an average length of 8. Table 6 displays the generated data for our experiments.

Table 5. Parameters used in the experiments

Parameter	Meaning
$\|D\|$	Number of user records
$\|T\|$	Average length of user moving nodes
$\|N\|$	Number of nodes for mobile network

Table 6. The generated sample data

No.	User record
1	ABCDEFGHIJKLMNP
2	ABEFIMN
3	BCDEFGKP
4	JK
5	BCEDFGIKM
6	BCEFJM
7	BEJM
8	JKLMN
9	KNOP
10	EFKN

Different thresholds for the probability of user frequent visited-nodes were used in the experiments. Figure 3 shows the experiment result of D1kT3N9 and D1kT4N16 by using the thresholds of probability: 0.001, 0.005, and 0.01. Here the numbers of user moving patterns for D1kT3N9 are 14, 7, and 5, respectively. We can see that the number of user moving patterns decreases as the probability of user frequent visited-nodes increases. We can see that the higher the threshold the fewer the number of qualified patterns.

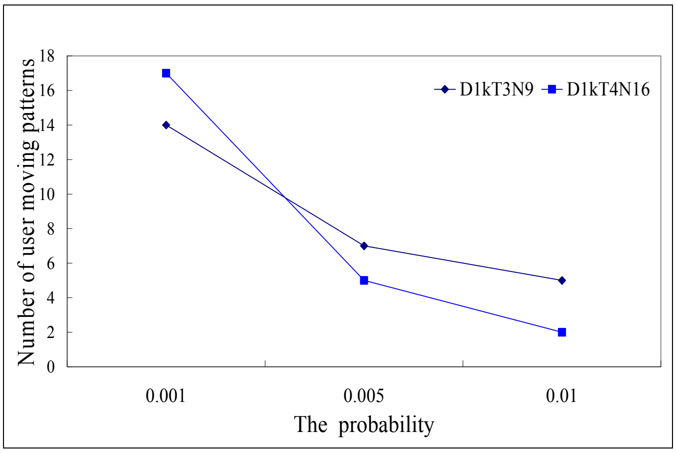


Figure 3. Results of different probabilities

Figure 4 shows the numbers of patterns discovered from our proposed method MUMP for three different numbers of user records under three different architectures. For architecture (N=3×3=9) with an average length of user moving nodes of T=3, the numbers of user records are D=100, D=1k, and D=10k for the datasets D100T3N9, D1kT3N9, and D10kT3N9, respectively. The threshold probability of user frequent visited-nodes is 0.001. We can see that the number of user moving patterns for D100T3N9 is 10 and 14 for D1kT3N9 as well as D10kT3N9. For the (N=4x4=16) architecture, the average length of user moving nodes is T=4. The datasets are D100T4N16, D1kT4N16, and D10kT4N16 and the threshold probability is also 0.001. The number of user moving patterns is 15 for D100T4N16, 17 for D1kT4N16, and 16 for D10kT4N16. For the cases of D100T5N25, D1kT5N25 and D10kT5N25, the result numbers of user moving patterns are 18, 16, and 18, respectively. We expect that the number of mined user moving patterns is approximately proportional to the size of the user records. There is a discrepancy for the case of D1kT5N25 in Figure 4. Our observation is that the generated data has the feature of randomness.

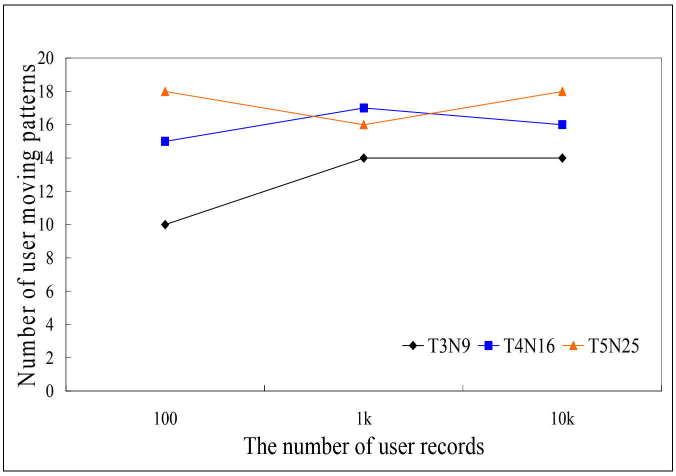


Figure 4. Results of different number of user records

Figure 5 shows the execution time of the MUMP. We can see that the execution time is proportional to the number of user records when the features of the cellular network architecture do not change. In this figure, the time unit is mini-second. As the number of users increases, the executing time also increases. For example, for the case where T=3 and N=3x3, the executing time is 1,082 mini-seconds for D100T3N9, 1,807 mini-seconds for D1kT3N9, and 5,610 mini-seconds for D10kT3N9. In the case T=4 and N=4x4, the executing time is 1,088 mini-seconds for D100T4N16, 1,753 mini-seconds for D1kT4N16, and 6,926 mini-seconds for D10kT4N16. And in the case T=5 and N=5x5, the executing time is 113 mini-seconds for D100T5N25, 1,890 mini-seconds for D1kT5N25, and the executing time is 7,570 mini-seconds for D10kT5N25.

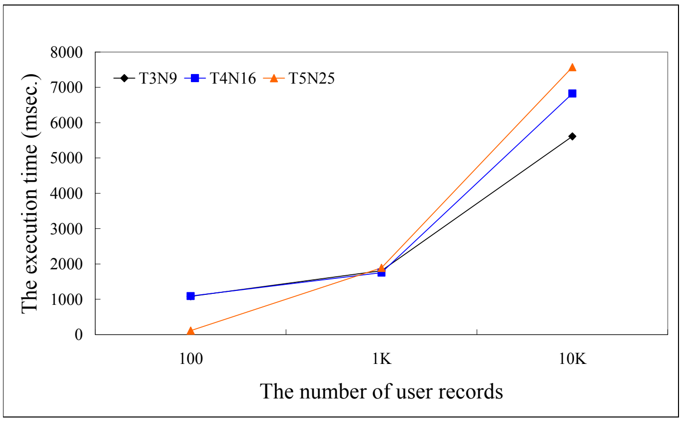


Figure 5. Execution time of MUMP

Next, we discuss the hit ratio of our method MUMP. We divide the user moving records into two parts for training and testing. The results are shown in Figure 6. The source data of user moving records are D10kT3N9, D10kT4N16, and D10kT5N25. Figure 6 shows that the hit ratio increases when the threshold decreases. For the dataset D10kT3N9 the hit ratio for threshold 0.005 is 12.02%, 34.65% for threshold 0.001, and 36.78% for threshold 0.0005. But the hit ratios are not the same for D10kT4N16 and D10kT5N25. Since the size of the cellular network of N=9 is smaller than the cellular networks of N=16 and N=25, we mined fewer user moving patterns. For the dataset D10kT4N16, we can find that the hit ratio for threshold 0.005 is 33.74%, 65.76% for threshold 0.001, and 75.91% for threshold 0.0005. And for the dataset D10kT5N25, we can find that the hit ratio for threshold 0.005 is 25.25%, 53.77% for threshold 0.001, and 76.67% for threshold 0.0005. Although the dataset D10kT5N25 has a lower hit ratio in average, the trend of increasing results is as expected.

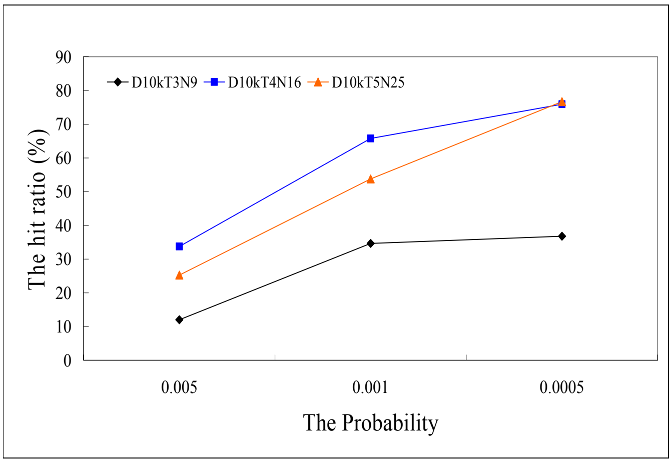


Figure 6. The hit ratio of MUMP

5. CONCLUSIONS

In our research, we found that the Hidden Markov model can be applied to the service study of cellular networks. By using the Hidden Markov model we can calculate the probability of users' moving nodes, and then obtain their moving patterns in the cellular network. We have demonstrated that our MUMP method is workable and effective. MUMP is based on the Hidden Markov model with the use of data mining techniques. In addition, we take the relation of connecting an old-node with new-nodes as our key approach in this research. We treat this relation as large 2-itemsets in the association rule mining process.

Using the proposed MUMP method, we can find user's moving patterns by merging the frequent visited-nodes. These user moving patterns can help network administrators to improve service efficiency and decrease the waiting time for users. Besides, the proposed approach can retain previously mined information of user moving patterns. After acquiring new moving records, we can calculate the probability of frequent visited-nodes more accurately. Our MUMP is a flexible method and is able to adjust itself according to the network architecture. In our experiments, we set the initial state distribution matrix and the observation probability matrix on the condition of fair probability. When applying our MUMP to real cellular network, the initial state distribution matrix and the observation probability matrix can be built with real data of the target cellular network.

Our study focuses mainly on mining mobile users' moving patterns. However, in the cellular network there are other behaviour patterns of mobile users. For example, if we apply association rule mining to find the relation between mobile users' behaviour and their corresponding location, we can acquire more business information about the need of these target customers. It can help us understand users' behaviour patterns in the business application or traffic flow control and management, etc. With the same idea, data mining can also help us perform better personalized services. As for improving the efficiency of our method, we will use parallel distributed process in our future work. In addition, using our approach can result in many more interesting applications of information services, such as tour guiding, shortest travel routes, and associated commercial promotions.

REFERENCES

Agrawal R., Imilienski T., & Swami A. (1993), Mining association rules between sets of items in large databases, P*roceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207-216.

Agrawal R. & Srikant R. (1994), Fast algorithm for mining association rules in large databases, *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487-499.

Bayardo R.J. (1998), Efficiently mining long patterns from databases, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 85-93.

Bernd B. & Bennington B. (1996), Applications of mobile computing and communication, *IEEE Personal Communication,* 3(1), 64-71.

Chen M.S., Park J.S., & Yu P.S. (1998), Efficient data mining for path traversal patterns, *IEEE Transactions on Knowledge and Data Engineering,* 10(2), 209-221.

Davies N., Blair G.S. & Cheverst K. (1996), Supporting collaborative application in a heterogeneous mobile environment, *Computer Communication: Special Issues on Mobile Computing*, 1-20.

Dunkel B. & Soparkar N. (1999), Data organization and access for efficient data mining, *Proceedings of the 15th International Conference on Data Engineering*, pp. 522-529.

Elmagarmid A., Jain J., & Furukawa T. (1995), Wireless client/server computing for personal information services and applications, *ACM SIGMOD Record,* 24, 16-21.

Han J. & Kamber M. (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.

Krishnakumar N. & Jain R. (1997), Escrow techniques for mobile sales and inventory applications, *ACM Journal of Wireless Network,* 3(3), 235-246.

Liu G.Y. & Maguire G.Q. (1995), A predictive mobility management scheme for supporting wireless mobile computing, *Walkstation Project Technical Report*.

Peng W.C. & Chen M.S. (2000), Mining user moving patterns for personal data allocation in mobile computing system, *Proceedings of the 29th International Conference on Parallel Processing*, pp. 573-580.

Rabiner L. (1989), A Tutorial on hidden markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77(2), 257-286.

Ontology-Driven Semantic Information Retrieval

Huang Hai
Institute of Intelligent Machines
Chinese Academy of Science Email: huanghai@ustc.edu
Wang Rujing
Institute of Intelligent Machines
Chinese Academy of Science Email: rujingw@iim.ac.cn
Huang He
Institute of Intelligent Machines
Chinese Academy of Science Email: huanghe324@126.com
Xiao BO
Institute of Intelligent Machines
Chinese Academy of Science Email: xiaobo@iim.ac.cn

Abstract
In this paper we describe the method of an ontology-driven semantic Information retrieval which is based on measuring the similarity of concepts in the ontology and matching the query with the documents represented with RDF graph. This method permits to search the information based on the document intentions and produces a ranking of those documents closer to the user's intention.

Keywords:
Information system, Ontology, Information retrieval, RDF, Concept Similarity

1. INTRODUCTION

Currently, with the information explosion caused by internet, increasingly diverse information is available. It is necessary to retrieve such great amount of information. Traditional retrieval methods are based on keywords matching. Keyword-based search methods have limitations as follows: Keywords in a document do not necessarily mean that the document is relevant, and relevant documents may not contain such words. Semantic relations such a hyponymy, meronymy, antonymy are not exploited. What is more, this approach is problematic when the user does not know what is there in database and what kind of semantic concepts are involved in the domain.

Recently, an emphasis has been placed on the use of ontologies for representing application domain knowledge. Ontology is a specification of a shared conceptualization [Gruber, T, R, 1993]. Ontology specifies a shared vocabulary to model a domain of interest. This vocabulary describes the type of objects and concepts that exist [Valerie Cross, 2004]. Meantime there is a tendency to begin use more complicate representations than using keywords.

In this paper, we discuss an ontology-driven semantic Information retrieval method. We use the domain ontology to specify the objects or concepts and their properties and relations between concepts. Then, we present a brief overview of semantic similarity, semantic distance and relatedness measures and talk about applying these measures with respect to matching ontological concepts. We use RDF graphs to represent the content of documents or web pages. As we know, RDF (RDFS) is a primitive ontology language, which can make the semantics unambiguous and machine accessible. Also, it provides the reasoning support for semantic with which we increase the precision of results. We are developing an information retrieval system which is based on RDF graph matching and concepts similarity measuring in the ontology.

The presentation begins with section 2 introducing ontology, Ontological representation and using RDF a primitive ontology language to represent the resources such as texts or web pages. Section 3 overviews several approaches for measuring semantic similarity and relatedness between concepts within an ontology. Section 4 discusses a semantic information retrieval system which is based on RDF graphs matching and concept similarity and relatedness measuring. Section 5 presents conclusions and discusses future work.

2. ONTOLOGY

The philosopher Willard Van Quine observed that the fundamental question of ontology can be expressed in the word:" what is there". It can be answered in just one word:" everything". The term ontology means a "particular theory about the nature of being or the kinds of existents". Now, ontology has been used in so many areas such as artificial intelligence, knowledge engineering, Information retrieval, linguistics, database theory and so on[John F. Sowa, 1999].

The methods of representing the ontology are so different and depend on details of the problem domain. For example, CYC [D. Lenat, 1990] is an axiomatized ontology whose categories are distinguished my axioms and whose definitions are stated in logic. Word Net is constructed as a semantic network with sets of synonymous terms, or synsets constituting its basic organization.

RDF (RDFS) [W3C, 1998] is a primitive ontology language which is domain-independent. RDF, Resource description framework, is a collection of triples. Each expression in RDF consists of a subject a predicate and an object (subject, predicate, and object). A set of such triples is called a RDF graph. The assertion of an RDF triple says that some relationship, indicated by predicate, holds between the things denoted by subject and object of the triple. For example, we could assert that (fish, a type of, animal)

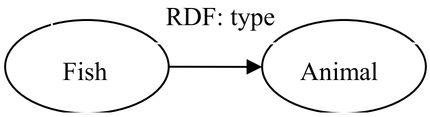


Figure 2.1 a RDF graph

RDF does not make assumptions about any particular application domain, nor does it define the semantic of any domain. RDF Schema (RDFS) define some core classes and core properties, such as RDFs: class, RDFs: property, RDFs: type, RDFs: subclass of, RDFs: subproperty of, through which we can capture or infer semantics of the resources described in RDF(S) or RDF. For example, from RDF graphs we can infer that:

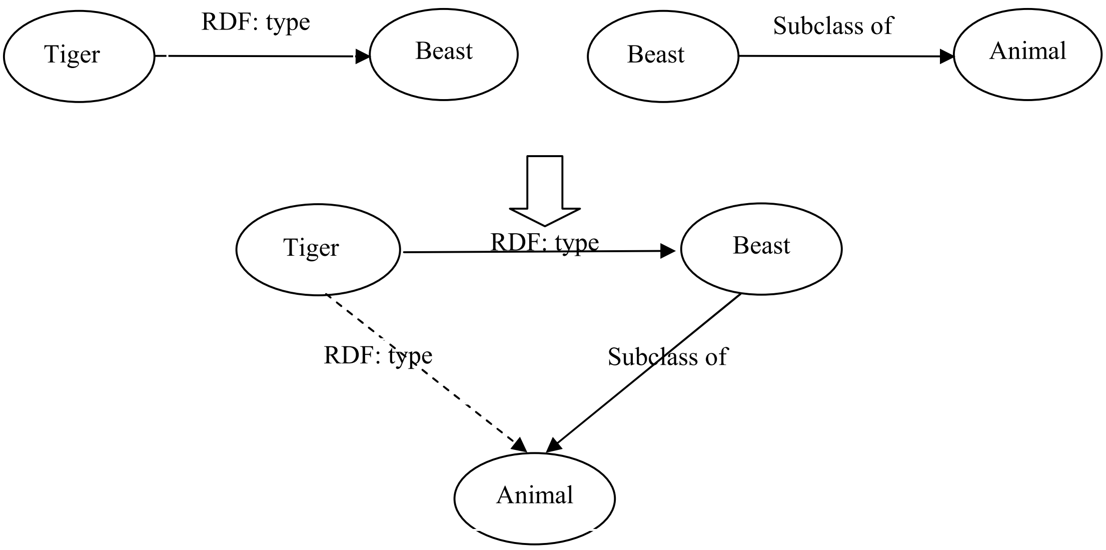


Figure2.2 reasoning on RDF graphs

We use RDF and RDF(S) to represent documents or web pages and we also represent the domain ontology in RDF and RDF(S).

3. CONCEPT SIMILARITY AND RELATEDNESS

3.1 Concept Similarity

Similarity is an important and widely used concept. One of the most natural approaches [Rada, Roy, 1989] to determine semantic similarity in the ontology is to use its graphical representation and measure the distance between the modes corresponding to the words or concepts being compared the number of edges in the shortest path between them. The shorter the distance the more similar the concepts are semantically. But in common, concepts in the ontology are more general than those lower in the hierarchy. One example in [Richardson, 1994] illustrates the problem. The distance between plant and animal is 2 in WordNet and the distance between zebra and horse is 2 too. But, almost all people think that zebra and horse are more similar than plant and animal. So it is obvious that only counting edges is not enough to measure the similarity between concepts.

In order to overcome the limitation of simple edge counting, some methods [Kim, Y and Kim, 1990][Lee , J and Kim, 2003] weight each edge. But because the weight of each edge has to be signed by hand, these methods are not practical for large ontologies. Another method [Richardson, 1994] presented a way of automatically weighting each link by the depth, the density of edges at depth, and the strength of connotation between parent and child nodes.

One method is to measure the minimum edge count distance between C_1 and C_2 by the maximum depth D of a taxonomic hierarchy the only use hyponymy, is a, type links between concepts.

$$Sim(C_1, C_2) = \max[-\log(\min_{c_1,c_2}[length(C_1, C_2)] / 2 * D)] \qquad \text{(L1)}$$

The method measures the depth the least common super concept C_3 between C_1 and C_2.

$$Sim(C_1, C_2) = 2 * N_3 / (N_1 + N_2 + 2 * N_3) \qquad \text{(L2)}$$

Where N_1 is the length of path from C_1 to C_3, N_2 is the length of the path from C_2 to C_3, and N_3 is the length of the path from C_3 to the root of the hierarchy. We use this method to measure the concept similarity in the ontology. The value of the similarity is between [0, 1]. For example:

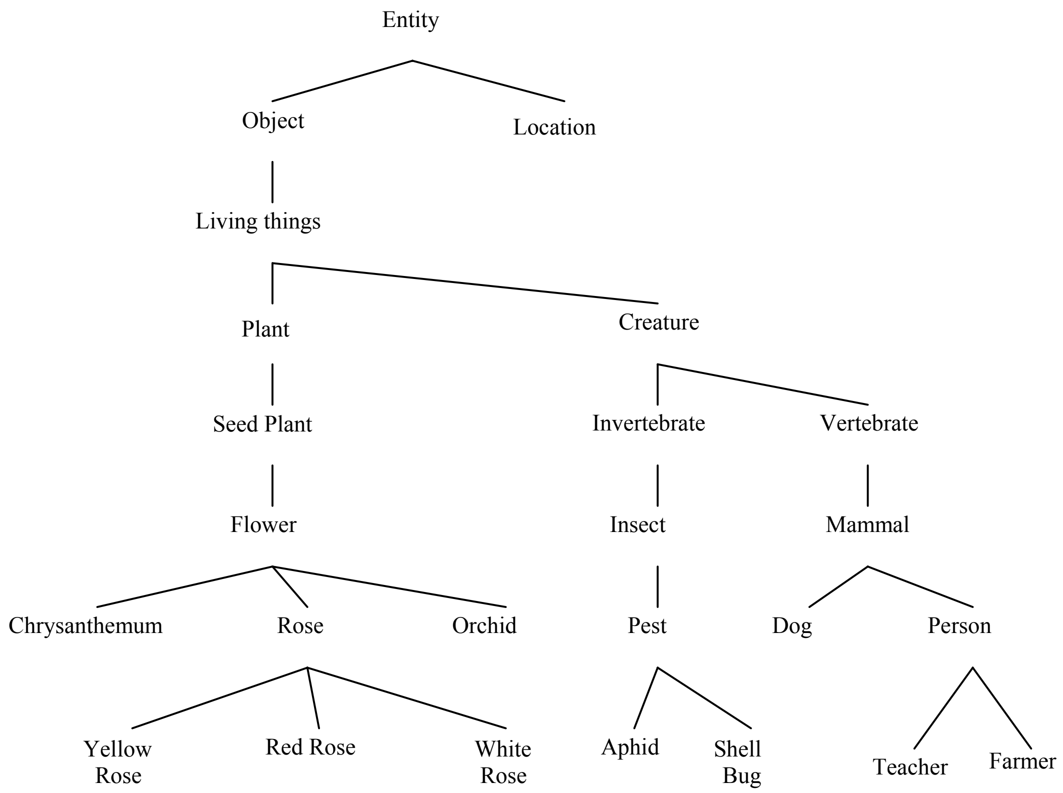


Figure 3.1 the structure of concepts

We can acquire the semantic similarity between concepts from formula L2:

$$sim("YellowRose","RedRose") = (2\times 7) \div (1+1+2\times 7) = 0.875$$

$$sim("Rose","Pest") = (2\times 2) \div (5+4+2\times 2) = 0.308$$

$$sim("Rose","Aphid") = (2\times 2) \div (5+5+2\times 2) = 0.286$$

3.2 Semantic Relatedness

We know that similarity is a special case of relatedness. That two concepts are not similar does not mean that they are not related. For example, the terms *computer* and *software* are more related than the terms *computer* and *calculator* because this kind of relatedness is based on relationship such as “computer uses software”.

The paper [Li Jun Zhu, 2004] divide the relation between concepts into two types (the obvious relatedness and the latent relatedness) .The obvious type includes direct relation and indirect relation. The latent includes the inherited direct /indirect relation. Direct relation means that the two terms have direct relation. For example, Aphid jeopardises the Rose, so they have the direct relation. (Figure3.2)

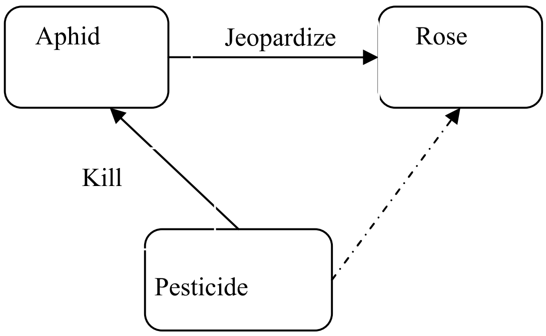
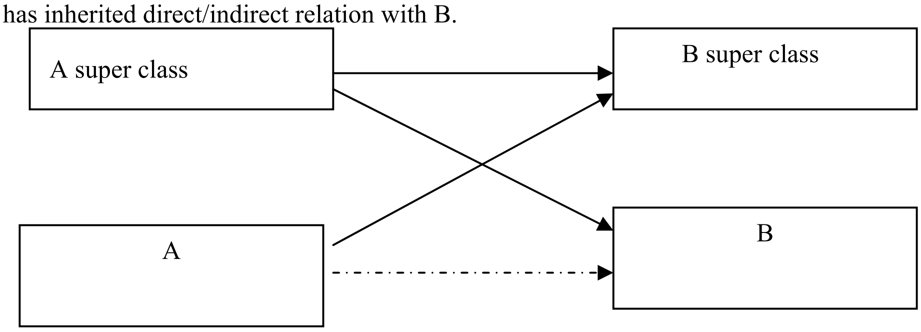


Figure 3.2 indirectly related

Indirect relation means that the two terms are indirectly related, Such as, the relation between the flower Rose and the pesticide. The Rose and the pesticide are related through Aphid.

Inherited direct/indirect relation mean two terms such as A and B, if the superclass of A is directly (indirectly) related with B, then A has inherited direct/indirect relation with B.



Direct or indirect related Direct or indirect inherited related

Figure 3.3 the inherited related

The value of the relatedness is between [0, 1] .if the two terms have the direct relation, the relatedness is 1.if they have none relation, the relatedness is 0.

Definition: the relatedness between term A and term B is Rel (A, B)

Rel (A, B) = 1; if A and B have the direct relation.

Rel (A, B) = $\frac{1}{n+1}$; if A is connected with B through n terms.

Rel (A, B) =0.75 ; if A and B have the direct inherited relation.

Rel (A, B) = $\frac{3}{4(n+1)}$; if the super class of A is connected with B through n terms.

4. INFORMATION RETRIEVAL SYSTEM

The main steps of our information retrieval system are:

1. Convert the query to the RDF graph.

2. Load the domain ontology and resources.

3. Compare the query graph with the RDFs Closure Graphs (not Raw Resource Graphs).

4. Return the user those documents which have the highest relatedness values with the query.

The architecture of our system is shown in Figure 4.1.

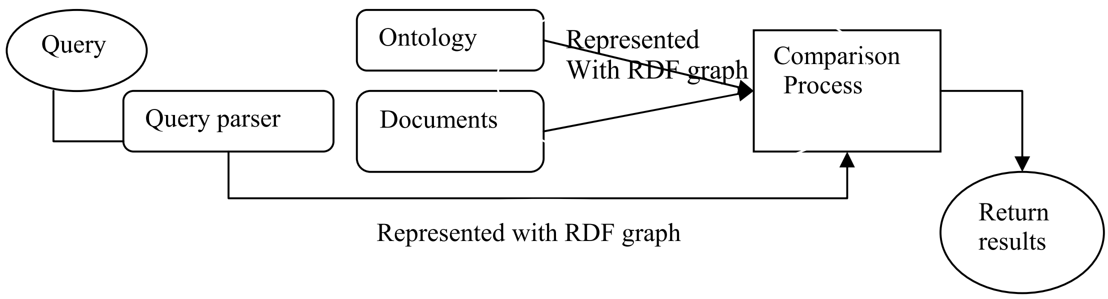


Figure 4.1 the architecture of system

4.1 Convert the query to the RDF query

The first step of our information retrieval system is to accept the users' query and convert it to the RDF graph which will be matched with documents also represented with RDF graph. So often we query the information such as "which documents are related with "Rose"?" It can be described with RDQL [HPLabs, 2002] and also can be represented in RDF graph as Figure 4.2:

```
SELECT ? X,  ?Object
WHERE (? x RDF: type AO: pageURL)
          (? page, AO: about, object)
            (? Object, RDF: type, Rose)
  USING
       RDF   FOR http:// http://www.w3c.org/1999/02/22-rdf-syntax-ns#
        AO   FOR http://127.0.0.1/ao-schema#
```

Query:

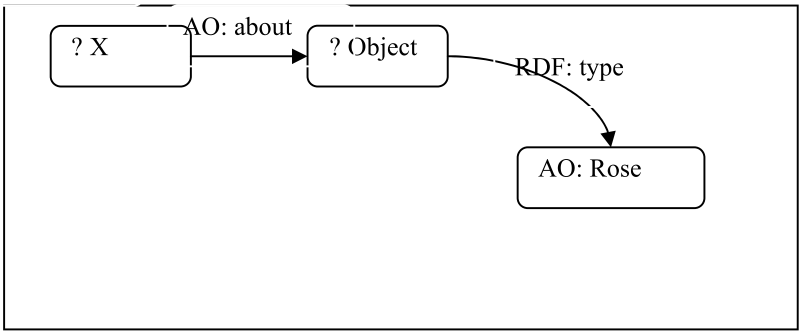


Figure 4.2 convert query to RDF graph

4.2 Load the domain ontology and resources

We use RDF (RDFs) to represent the resources (texts or web pages) which permit to search the information based on the document intentions. And RDF (RDFs) is a widely used ontology language; the documents which are represented with it are conveniently to share with each other.

We can represent the content of the documents with RDF graphs. And also we create the domain ontology which specifies the shared concepts and their relationship and properties between concepts. Figure 4.3 shows ontology and documents represented in RDF graph. When the domain ontology is loaded, more useful information about concepts and relationships not described in documents will be added. Thus, in fact the query graph will not match the Raw RDF Graphs but RDF Closure Graphs.

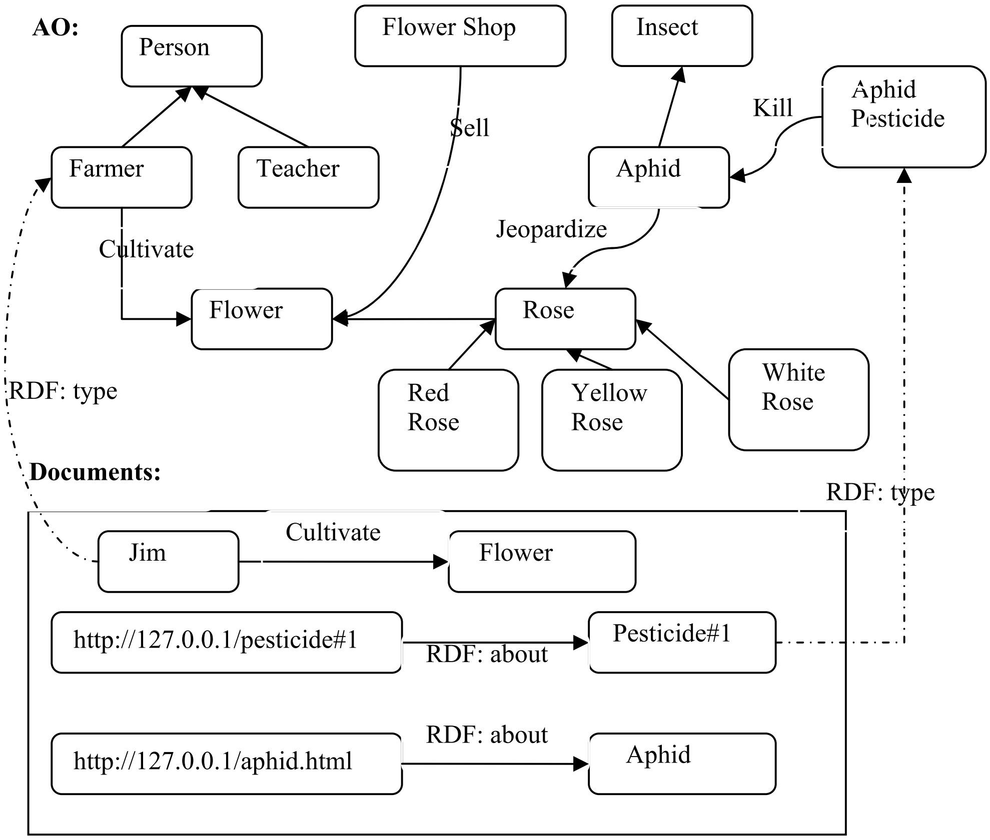


Figure 4.3 documents loaded with ontology and represented with RDF graph

4.3 Compare the query graph with the RDFs Closure Graphs (not Raw Resource Graphs)

As we know, the user's query and the resources are both represented in RDF graph. Our algorithm for the comparison of two RDF graph consists of two main parts: 1) Find the intersection of two graphs. 2) Measure the similarity between the two graphs.

The intersection consists of the two elements: 1) All concept nodes that appear in both two RDF Graphs .2) All relation nodes that appear in both two RDF graphs. In this case, the query is completely matched with the document. But sometimes the query cannot be completely matched with the document. At that time, we measure the similarity between the query graph and the document graph. Our algorithm is presented as follows:

1. Find the concept C_0 in the query, and computing Sim (C_0, C_i), Rel (C_0, C_i); ($C_i \in C$, C is the set of all concepts)

2. $SR_{0,i} = \lambda \times$ Sim (C_0, C_i) + $(1-\lambda)\times$ Rel (C_0, C_i); $\lambda \in [0,1]$

3. $\forall$ $C_i \in C$, if $SR_{0,i} > \xi$ then put C_i in the extended concept set D ($\xi \in [0,1]$ is the critical value we set)

4. Replace C_0 with $D_i \in D$ (extended concept set), if the query (with the concept D_i) can be matched with certain documents then return them and rank them by $SR_{0,i}$.

4.4 Return the user those documents which have the highest similarity values with the query

The documents are ordered by the value of SR. Thus it produces a better ranking of those documents closer to the user needs. For example: given the query “which documents are related with “Rose”?” (described in Figure 4.2) , from the documents and ontology presented in Figure 4.3, we can acquire the results ranked:

Page	Object	SR	λ
http://127.0.0.1/aphid.htm	Aphid	0.643	1/2
http://127.0.0.1/pesticide#1	Pesticide#1	0.25	1/2

Table1: the results of returning

5. CONCLUSION

We have described an ontology-driven semantic information retrieval system which is based on the comparison of the query and the document represented with RDF Graphs. We also talk about the method of measuring the concept similarity in ontology. We use this method in that the ontology is often created by domain experts. Thus, it is more accurate to represent the category of domain concepts and measure the similarity between concepts. The comparison of the query and the document represented with RDF Graphs allows improving the retrieval of information in two main directions: 1) it permits to search the information based on the document intentions. 2) It produces a ranking of those documents closer to the users’ intention.

Currently, we are still adapting this system to the practical application. Now we mainly describe documents with RDF(s) by hand, which is a tedious work. Thus we are designing and implementing a system semi-automatically transferring documents into RDF(s). The method that we propose is not limited in information retrieval. Other uses of the method include text mining, text classification, web extraction and so on.

6. REFERENCES

1. [Gruber, T, R, 1993] Gruber, T, R. "*A translation approach to portable ontology specifications*", knowledge acquisition, 5(2) 1993, pp.199-220

2. [Valerie Cross, 2004] Valerie Cross “Fuzzy Semantic Distance Measures between Ontological Concepts” IEEE Translation on systems, 2004

3. [John F. Sowa, 1999] Knowledge *Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, CA, ©2000. Actual publication date, 16 August 1999.

4. [D. Lenat, 1990] D. Lenat and R. Guha, *Building Large Knowledge Based Systems: Represented and inference in the CYC Project.* Reading, Mass.: Addison Wesley Publishing Company, 1990

5. [W3C, 1998]http://www.w3.org/TR/RDF-primer/

6. [Rada, Roy, 1989] Rada, Roy, Hafedh Mili, Ellen Bicknell, and Maria Blettner *.Development and Application of a Metric on Semantic Nets*. IEEE Translation on systems, Man and Cybernetics 19:17-30.1989

7. [Richardson, 1994] Richardson, R, A. Smeaton, and J.Murphy *Using Wordnet as a knowledge Base for Measuring Semantic similarity between words*. Technical Report, Working Paper CA-1294, school of Computer Applications, Dublin City University: Dublin, 1994

8.[Li Jun Zhu, 2004] Research of Domain Knowledge Based Information Resource Management PatternIn World Wide Web Environment Doctoral Dissertation of Chinese Agricultural University, 2004.

9. [Kim, Y and Kim, J., 1990] Kim, Y and Kim, J." A Model of knowledge based information retrieval with hierachical Concept graph" Journal of Document, 46, 1990, 113-116

10. [Lee, J and Kim, M., 2003] Lee, J and Kim, M." *Information retrieval based on Conceptual distance in a is-a Hierarchy* "Journal of Documentation, 49, 2003, 188-207

11. [V. Cross and T. Sudkamp, 2002] V. Cross and T. Sudkamp, Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications, New York: Physica-Verlag, ISBN 3-7908-1458, 2002

12.[HPLabs,2002]A. Seaborne, RDQL – RDF Data Query Language, part of the Jena RDF Toolkit, HPLabs Semantic Web activity, http://hpl.hp.com/semweb/, 2002

An evaluation of Adaptive Educational Information Systems and its impact on University students' learning performance

Amal Al-Dujaily
Massey University
Auckland, New Zealand
Email: a.abas@massey.ac.nz

Dr. Hokyoung, Ryu
Massey University
Auckland, New Zealand
Email: h.ryu@massey.ac.nz

Abstract
Adaptive hypermedia systems (AHS) are believed to effectively support the user's learning process, adapting to the current level of user's knowledge. A model of a user's knowledge of an application domain needs to be firstly established. And in return, the learner's user model specifies what to adapt and how to support the user's learning process for obtaining the application domain, and then suggests the most relevant contents from the application domain for the learners to enhance their learning in an effective way. The user models that were employed in most of the AHS do not seem to be consistently spotted what to be incorporated and how it can support the student's learning in terms of their different knowledge acquisition strategy. We noted that individual personality would have effects on each knowledge acquisition strategy, and consecutively it would facilitate the different use of AHS. We are pursuing to investigate how to embody the personality features into the evaluation terms, which has been overlooked in the previous AHSs studies. Experiments conducted in two different countries revealed a significant impact of individual personality on learning performance with AHSs.

Keywords:

Adaptive Hypermedia System, User Model, MBTI, Personality, Introversion, Extraversion.

INTRODUCTION

It is generally believed that the Internet or Web-based systems have increased the effectiveness of educational applications, particularly for on-line teaching and distant learning. Further, historically, their applicability has been much extended by the development of Adaptive Hypermedia Systems (AHS) in higher educational applications, such as AHA (De Bra & Calvi, 1998), ELM-ART (Brusilovsky, Schwarz, & Weber, 1996) and Interbook (Brusilovsky et al., 1996; Eklund & Brusilovsky, 1999). Indeed, they provided better authoring tools for developing web-based courses, facilitating the student's learning experience with the support of adaptivity that they offer (Brusilovsky, 2001).

AHSs are a more advanced style of tutoring systems, in conjunction with the benefits of both Intelligent Tutoring Systems (ITSs) and Hypermedia Systems (HSs). It allowed a directive tutor style, mimicking one-to-one tutoring situations in the general educational context, and a flexible student-centred approach in computer-assisted instructions. In practice, AHSs have been extended to overcome the problems that were often associated with on-line educational applications, which were delivered over traditional hypermedia environment. For example, in the traditional hypermedia systems, the learners tended to lose their orientation in hyperspace while navigating within the complex hypertext structure (Brusilovsky & Eklund, 1998; Hammond & Allinson, 1990), and the lack of guidance from the teachers or system supports seemed to make effects on learners' worse learning experience (Brusilovsky & Eklund, 1998; Papasalouros, 2002). As a consequence, the comprehension of the contents that the systems intended to deliver appeared to be less effective to novice or inexperienced learners. Perhaps, these problems are due to fact that the traditional hypermedia systems have not accounted for the individual characteristics of learners, providing all students with the same information regardless of their difference (Papasalouros, 2002). In effect, AHSs were intended to negotiate these problems, thus enhancing online learning performance.

Although there have been many studies on the adaptive techniques of AHSs, less investigation for evaluating AHSs has been made. This may be partially because the techniques themselves are the most important features that should be maintained in the design of AHSs, and partially because there are a too wide spectrum of the aspects that the evaluation should be aimed at in the evaluation terms

(Brusilovsky et al., 1996). On the grounds of this brief understanding of the current evaluation issues of AHSs, I am aiming at filling the gap which will lead to enrich educational information systems for the actual users, e.g., University students.

Research question

The primary research question is to evaluate an adaptive educational information systems and its impact on University students' learning performance while they are asked to employed the systems in the delivery of a particular course. In fact, this research concerns arose from the author's teaching experience in an Asian university in which most of the students had the different learning style against that of the students in a Western university. Based on this capital question, the first research agenda was whether the different learning styles (personality) would have a consequence of learning performance; and, if so, how this phenomenon should be incorporated in an evaluation framework of adaptive educational systems will be followed.

METHODOLOGY

To investigate these research questions, we firstly established a methodology that was expected to reveal the relationship between personality characteristics and the learning performance, as shown in Figure 1.

Stage 1	Stage 2	Stage 3
Generating a framework	Exploring understanding of the frame work	Evaluating the framework
- Identify knowledge acquisition style - Characterising adaptivity of the adaptive education system - Setting out the application domain - Case study 1 (Asian students)	- Comparing the effectiveness of each system (Adaptive system vs. non-adaptive systems) - Case study 2 - (Asian vs. Western) students)	- Evaluating the framework with grounded theory - Case study 3

Figure 1. The Research framework

Stage 1. Generating a framework for understanding students' individual differences

In the initial stage of the research, we intended to generate a framework to understand how we could review the effectiveness of the adaptive education systems. Three aspects were considered as follows: (i) identifying each student's knowledge acquisition style, (ii) characterising adaptivity of the adaptive education system, and (iii) setting out the application domain for the first case study. To deal with the first aspect, Myer-Briggs Type Indicator (MBTI) questionnaire test was employed. The MBTI test was originally developed to measure people's personality type (Carolyn & al., 2001), however, it has also been used for developing different teaching methods that meet different students' learning styles. It provides four sets of preferences, which has two opposite styles for each of the four sets of preferences, as follows:

Source and direction: *Extraversion (E) vs. Introversion (I)*

Extraverts are oriented primarily toward the outer world; thus they tend to focus their perception and judgment on people and objects, their personality normally tends to act first, and think or reflect later. In contrast, Introverts are oriented primarily toward the inner world; thus they tend to focus their perception and judgment upon concepts and ideas', thus thinking or reflecting is followed by acting.

Method of perception: *Sensitive (S) vs. Intuitive (N)*

A preference between two opposite ways of perceiving; Sensitive feature is characterized by a preference for direct, concrete experiences; therefore they are practical, immediate, and uncomfortable with abstract idea. By contrast, the intuitive learners tend to abstract theoretical accounts, using imagination rather than on concrete realities.

Decision making: *Thinking (T) vs. Feeling (F)*

A preference between two contrasting ways of judgment, a person may rely primarily through thinking to decide impersonally on the basis of logical consequences. In contrast, a person may rely primarily on feelings (F) to decide primarily on the basis of personal or social values.

Implementation: *Judgment (J) vs. Perception (P)*

Judging people are decisive, planful and self restricted. They focus on completing the task, only want to know the essentials, and take action quickly. They plan their work and work their plan. Deadlines are sacred. In contrast, Perceptive people are curious, adaptable, and spontaneous. They start many tasks, want to know everything about each task, and often find it difficult to complete a task

According to MBTI, the personality types, which are closely related to user's information processing strategy, are any combination of these four aspects of personality features, establishing 16 different styles of learning. The most common MBTI types are ESTJ, the Extraverted-Sensing-Thinking-Judger versus ISTJ, Introverted-Sensing-Thinking-Judger personality type (Myers, 1998).

Secondly, we reviewed the current adaptive education systems to see what features they would like to deliver. From our understanding, all AHSs have been designed to include the same basic components of the domain model, the user model, and the techniques to adapt both contents and links with respect to the user model (Cannataro, 2001). Hence adaptive education systems can accommodate different students to achieve different educational goals with the same course material (Henze & Nejdl, 1999). For example, Interbook integrated the history-based, knowledge-based and prerequisite-based method (Eklund & Brusilovsky, 1999), in order to provide adaptive navigation support. The application domain used in Interbook was applied in the first case study, undertaken in an Asian University (Al-Dujaily, Hokyoung, & Kamal, 2005). It identified that introverted students more concentrated on theoretical aspects and extroverted students were particularly interested in practical applications.

Stage 2. Exploring understanding of the framework

Based on the understanding of Stage 1, the main purpose of this stage is to extend the accounts from the previous stage with an empirical study. For this end, we carried out a major experimentation to see if the personality types could actually have effects on the use of adaptive education systems. Two different systems (non-adaptive system and adaptive system), which acted as independent variables, were considered as the students were asked to learn a computer programming language. To examine if the individual personality would affect the learning performance, MBTI tests of all the participants were performed. The first half of the experiment was carried out in an Asian university, and currently the other half to evaluate the performance of Western students is being performed.

PRELIMINARY RESULTS

In the first half of the main experiment, we firstly found that Asian students had benefits from adaptive educational information systems, enhancing their understanding of the contents that the systems intended to deliver. It made less navigation and repetitions. The MBTI results showed that 70 out of 100 students were introverted. In detail, the ISTJ (Introversion, Sensing, Thinking, and Judgment) learning style was dominant (82.85%), while the other learning styles, i.e., ISFJ, and ISFP, were not common. In summary, the results from the half of the case study showed that there was the significant difference between the adaptive information systems and the non-adaptive information systems, but it did not explicitly demonstrate whether this difference came from the system characteristics or from the user's side. Therefore, we carried out the other half of the experiment in the different country of which the students were expected to have the different learning style against the Asian students. Small scales of pilot tests were performed, and then the main experimentation was performed. Similar results had been found in terms of western students, where students tend to benefit more from adaptive educational information systems. Moreover statistical results for both countries showed that introverts gained an average of 7.07% more in their learning performance while extraverts gained 6.13%.

Conclusion

The conclusion drawn from above results justifies the usefulness of adaptive educational information systems as a learning tool for higher education for both Asian and Western students. Since the type of personality has directive impact on learning performance as an outcome, adaptive educational information systems are useful for both introverted and extroverted students.

Future research plan focusing on Stage 3: Evaluating the framework

In accordance with our results, we will need a new framework to encompass the personality aspect into the specification of the user model, as shown in Figure 2.

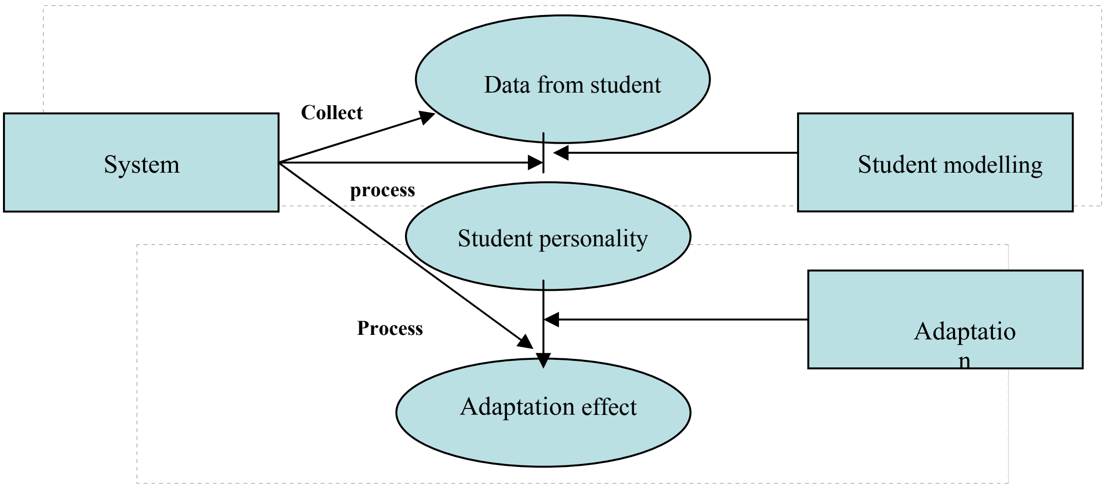


Figure 2. User model adapting personality

References

Al-Dujaily, A., Hokyoung, R., & Kamal, J. (2005, July 5-8). *Introversion and the implication of intelligent tutoring systems: A lesson from Arab Students.* Paper presented at the ICALT, Kaohsiung,Taiwan.

Brusilovsky, P. (2001, June 23-26, 2001). *Adaptive Educational Hypermedia:"A summary of an invited talk at PEG'01 conference".* Paper presented at the Proceedings of Tenth International PEG conference, Tampere, Finland.

Brusilovsky, P., & Eklund, J. (1998). A study of user-model based link annotation in educational hypermedia. *Journal for Universal Computer Science, 4*(4).

Brusilovsky, P., Schwarz, E., & Weber, G. (1996, June 12-14). *"ELM-ART: An intelligent tutoring system on World Wide Web".* Paper presented at the Third International Conference on Intelligent Tutoring Systems.

Cannataro, M., Cuzzocrea, A and Pugliese , A. (2001). *A probabilistic approach to model adaptive hypermedia system.* Paper presented at the International workshop for web Dynamic.

Carolyn, S., & al., e. (2001). Myers Briggs Type Preferences in Distance Learning Education. *International Journal of Educational Technology, 2*(2).

De Bra, P., & Calvi, L. (1998, June 20-24). *"AHA open adaptive hypermedia architecture. Towards a generic adaptive hypermedia system (AHA).* Paper presented at the 2nd workshop on adaptive hypertext and hypermedia . Pittsburgh, PA.

Eklund, J., & Brusilovsky, p. (1999). Intebook, "An adaptive tutoring system". *Uniscience news, 12*.

Hammond, N., & Allinson, L. (1990). *Extending hypertext for learning: an investigation of access and guidance tools.* Paper presented at the The fifth conference of the British Computer Society, Human-Computer Interaction Specialist Group on People and computers V table of contents.

Henze, N., & Nejdl, W. (1999). *" Bayesian Modeling for Adaptive Hypermedia Systems".*

Myers, I. (1998). Guide to the Development and Use of the Myers-Briggs Type Indicator. *CPP, Inc, 3rd edition*.

Papasalouros, A. (2002). Ob-AHEM: AUML-enabled model for adaptive education hypermedia application. 76-88.

Wu, H., & Miller, l. (2001). Design of data warehouses using metadata "information and amp software technology". *43*(2), 109-119,.

Under Benefits of Bio-KMS, the Knowledge Management Process Effects on Knowledge Management Performance

Li-Ling Hsu

Department of Information Management
National Kaohsiung First University of Science and Technology
karenhsu@ccms.nkfust.edu.tw

Abstract

The aim of this paper is to examine how bioscience industry exploits KMS to effectively manage intellectual property, especially patent documents. A majority of Taiwan's bioscience firms are small and medium-sized start-up companies that possess many valuable intellectual property and patents/licenses. Nevertheless, due to lack or distortion of information, they are more likely to be charged of patent infringement by overseas companies who own the invention or technology. A case study on a Taiwan's leading bioscience firm is conducted to analyze the effect of KMS on patent management. This study findings show that KMS helps the firm avoid a conflict of interest, ensures rapid publication while allowing the firm to protect R&D results, and enhances competitive advantage. Three constructs are developed in this research: KMS, business innovation performance, and Bio-KMS benefits. Five propositions are developed and supported by case analysis results. The main contributions of this paper can be classified in: knowledge-based and technology -based aspects. Implications for theory and practice are offered.

Keywords:

Knowledge management system (KMS), Knowledge management process (KMP), Business innovation performance (BIP)

1. INTRODUCTION

1.1. Research Background and Motivation

Biotechnology industry has grown rapidly since 1978. Nowadays it has become the focus of high-tech development for developed countries. Nowadays, over sixty firms develop new products and other businesses by use of Bioinformatics all over the world. According to Front Line, the Bioinformatics market in 2000 is US$1.1 hundred million dollars. Till 2005, the profit increased to US$4 hundred millions, and is estimated to increase 8 hundred millions in 2010. It shows that Bioinformatics market has a great potential, the market is not big, though. A large majority of Taiwan's biotechnology firms are small and medium enterprises, with its capital less than 5 hundred million dollars, and employees under 50 people. In particular, biotechnology firms possess a high percentage of knowledge workers (Chao, 2004). The ability for these knowledge workers to effectively acquire and use expertise will be a determinant of the success of biotechnology firms (Drucker, 1999; Liebowitz, 2000).

As Liebowitz (2000) points out in his research, managers believe knowledge assets or intellectual capital is the most important asset to separate their firms from other competitors. The organizational knowledge is carried out through multiple forms/entities, including organization culture and identity, routines, policies, systems, and documents, as well as individual employees (Alavi & Leidner, 2001; Grant 1996a, 1996b; Liebowitz, 2000, 2002, 2004). Because knowledge-based perspectives are difficult to imitate or duplicate, and will be able to produce long-term sustainable competitive advantage (Liebowitz, 2000, 2002, 2004; Sunassee & Sewry, 2002, 2003). Meanwhile, they further posit that information technology is a force to effectuate the knowledge-based perspectives of the firm (Alavi & Leidner, 1999; Chou & Lin, 2002; Liebowitz, 2002). Advanced information technologies, such as the Internet, intranets, extranets, browsers, data warehouse, data mining techniques, and software agents, can be used to systemize, enhance, and expedite knowledge management (Alavi & Leidner, 2001; Huber, 2001). Nevertheless, new technology implementation may not effectuate the knowledge management or knowledge resources.

1.2. Research Purpose

This paper adopts 'Organizational Knowledge Creation' theory as the theoretical framework. In this paper we propose a conceptual framework by taking Taiwan's small and medium biotechnology firms as targets to examine if the implementation of knowledge management systems helps increase the performance of business innovation. The purpose of this paper is stated as follows:

(a) To examine whether the implementation of knowledge management processes has positive impact on the firm's 'innovation performance'.

(b) To examine whether the effect of Bioinformatics management systems enables the knowledge management activities and promotes the firm's innovation performance.

2. LITERATURE REVIEW

2.1. Knowledge Management Process, KMP

Knowledge assets have become one of the most important assets in today's enterprises. Knowledge management is a key component in an organization's ability to realize the full potential of its intellectual assets in strategic and tactical decision making and in creating a competitive advantage (Silver, 2000; Hsieh et al., 2002). The transformation of knowledge into assets relies on knowledge management processes. Nonaka (1994) explicates two kinds of knowledge in organizations: explicit and tacit knowledge. Explicit knowledge is easier to explicate, so knowledge can be externalized in the form of documentation. On the other hand, if tacit knowledge is to be made explicit, then it must be extracted through socialization (Nonaka, 1995). Knowledge can range from being highly tacit to easily codifiable. Codifiability of explicit knowledge is relatively easier when compared with that of tacit knowledge (Laura et al., 2001; Yang & Wan, 2004). Knowledge codifiability involves message creation, model creation, and language representation (Robin, 2001; Gardan & Gardan, 2003; Yang & Wan, 2004).

Nevertheless, knowledge is created and held not by an organization itself, but by individuals; it can be disseminated through individual's interactions, sharing and learning. Nonaka and Takeuchi (1995) also propose 'Organizational Knowledge Creation Theory' for knowledge creation process. Organizational knowledge creation is thought of as the knowledge capabilities of the enterprises, capabilities of integrating novel knowledge into the product design, services and management systems. Organizational knowledge creation spirally flows through individual level, group level, and organizational level, and undergoes four stages of knowledge integration process, i.e. socialization, internalization, combination, and externalization. In other words, knowledge creation refers to the process of replacing or innovating new tacit and explicit knowledge in an organization (Pentland, 1995), and the process of knowledge creating, sharing, distribution, and organizational evaluation through social activities, interactions, and individual's reflections on the experiences of others (Nonaka, 1994).

2.2. Business Innovation Performance, BIP

Previous studies have shown that a majority of senior IT managers consider creativity and innovation as the most important tasks in IT management (Couger, 1998; Liebowitz, 2000; Niederman et al., 1991; Zawacki, 1993), since it is the prerequisite for a manager having the ability to transform individual knowledge into organizational resource, and manage intellectual assets (Sunassee & Sewry, 2002, 2003). The resource-based view is grounded in the perspective that innovation is the change that impacts resource output. Economic theory postulates business innovation changes consumers' value and satisfaction with resources. Industrious and revolutionary view identifies business innovation in terms of innovative activities needed and the extent to which innovation takes place as: architectural innovation, evolutionary, market niche, and regular innovation (Abernathy & Clark, 1985). In terms of historical perspective, business innovation can be divided into three phases: innovation, growth, and maturity (Tushman & Nadler, 1986; Holt, 1988; Utterback, 1994).

Business innovation refers to the change of organizational resource output, and the activities that are innovative and may bring profits to the firm (Souder, 1988). Business innovation can also refer to a series of innovative representation or activities contributory to resource output through interactions with/between individuals, groups or organizations (Gattiker & Larwood, 1990). Hence, both internal and external forces drive business innovation. Meanwhile, Utterback (1994) postulates that product development will also affect business innovation, since organizations evaluate the maturity of product development to re-adjust innovation activities. In addition, Afuah (1998) emphasizes that innovation will become a motivator only when top management realizes the importance of it, since their perception to innovation is dependent upon individual experiences in management, their consciousness and concerns over organizations and industries.

2.3. Techniques and Benefits of Knowledge-oriented Bioinformatics Systems

The application of Biotechnology is also shifting its function from storing and managing explicit knowledge to converting explicit into tacit knowledge. Hieter and Boguski (1997) refer that in post-genomics era, biological knowledge is evolving from structural genomics to functional genomics. When taken from information technology perspective, structural genomics aims to collect, store, analyze and manage data, while functional genomics analyzes the data sets with intellectual knowledge and data mining techniques. In terms of knowledge management, the former is explicit, while the latter is tacit which requires individual experiences and competency to extract and interpret. Nowak (1995) also indicates in post-genomics era it is rather important to extract tacit knowledge from the tremendous amount of gene sequence information.

The knowledge-based perspective takes knowledge as the intangible assets of organizations (Liebowitz, 2000; Sunassee & Sewry, 2002, 2003). Hence, Assessing the quantitative or qualitative, tangible or intangible assets has become an essential task for CIO and CEO (King et al., 2002). Previous studies have offered relevant indicators for assessing the benefits of KMS, such as the ability to improve decision-making and productivity (Stata, 1997). Alavi and Leidner (1999) classify the benefits of KMS into process results and organizational results. The process improvements can be further divided into communication improvements and efficiency gains. The perceived organizational benefits can be classified into financial, marketing and general nature (Alavi & Leidner, 1999). A number of suggested measures for the perceived benefits of KMS include avoidance of problems, sharing of best practices, faster cycle times to problem solution, faster new product cycles, faster customer response time, increased number of innovative products/services, improved customer satisfaction, reduced risks (Skyrme, 1999). Nevertheless, these measures are still under examination, and require further investigations to examine respective influences on organizational performance and knowledge management processes.

3. RESEARCH PROPOSITIONS

3.1. Research Model

With 'KM' as theoretical fundamentals, we develop a conceptual framework shown in Figure 1.

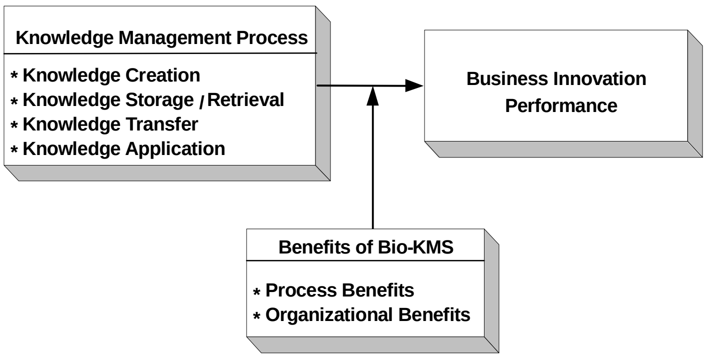


Figure 1. This Research Model

This model consists of three constructs: KMS, business innovation performance, and Bio-KMS benefits. KMS is divided into knowledge creation, storage/retrieval, transfer and application. Bio-KMS effects will be thought of having process and organizational improvements. This research model is developed in the attempt to examine:

(a) The impact of BIP on organizational management process.

(b) Bio-KMS is an enabler to improved organizational management process and business innovation performance.

3.2. Research Propositions

3.2.1. Correlation between Knowledge Management Processes (KMP) and Business Innovation performance (BIP)

Chattell (1998) postulates the firm's core competency must be developed through (a) instant knowledge acquisition, (b) knowledge screening and application, (c) knowledge regeneration, (d) knowledge applied to work, and (e) performance. There are also scholars who postulate the importance of knowledge flow. For organizational learning, when the organizations are more capable of interpreting and acquiring knowledge, their core competencies will be enhanced (Liebowitz, 2000). Nevertheless, there is no direct and obvious connection between importance and future capacity, therefore organizations must build formal and informal process for knowledge interpretation and acquisition that enhances individual's ability to operate machines. As knowledge accumulates, individual's experiences and knowledge transition will be crucial to the importance and future competencies of the firm. Nevertheless, knowledge accumulation is the last stage of knowledge management, and is relative less important in the firm's competitive advantage, in comparison with knowledge interpretation and acquisition. A higher level of knowledge expansion helps elevate individual and organizational competence.

Liebowitz (2002) also addresses that knowledge sharing improves innovation. Hence, knowledge sharing, acquisition, and application are the core elements of innovation. Cohen and Levinthal (1990) Nonaka and Konno (1998) and Yang & Wan (2004) agree that learning process and knowledge interpretation, accumulation, and expansion are not only crucial to organization's innovative capabilities, but also to the continual product and process improvement and competitive advantage (Chuang, 2004). In a word, knowledge development and accumulation have positive impact on innovation. As far as Taiwan's hi-tech industry as concerned, shortened product life cycles and dynamic technological knowledge have changed the firm's way of acquiring knowledge. With highly confidentiality of internal information, the process of knowledge accumulation and knowledge flows will be determined by each department. The more explicit the technological knowledge is, the faster for knowledge accumulation. We develop the following propositions:

- P1: Knowledge creation has a positive impact on BIP.
- P2: Knowledge accumulation and retrieval have a positive impact on BIP.
- P3: Knowledge transfer has a positive impact on BIP.
- P4: Knowledge application has a positive impact on BIP.

3.2.2. Correlation between Bio-KMS benefits and Knowledge Management Process (KMP) and Business Innovation performance (BIP)

The biggest value of Bioinformatics is its ability to integrate information science and biology into data, which was transformed into organizational knowledge. Take NCBI (National Center for Biotechnology Information), which maintains GenBank as an instance, the database has been doubled about every 18 months, which present challenges to computational scientists, for the accelerated rate of genomic databases outstripped the speed of computation (Moore's Law). Hence, many scholars in post-genomics era suggest that a better understanding of the function of information technology (data mining, ontology, and knowledge maps) to mining tacit knowledge, which increases the efficiency of problem solving and competitive advantage, will be needed (Nowak, 1995; Wagner et al., 2003). Information technology is designed to process flood of data, integrate heterogeneous databases, analyze complex data, and makes data extraction a lot easier. Based on the above rationale, we develop the following proposition:

- P5: Bio-KMS benefits may improve knowledge management processes and enhance BIP.

3.3. Research Design

Previous studies on knowledge management processes have concentrated on hi-tech information or electronic industry. Little has been found in the research of biotechnology industry. Hence, this paper aims to investigate the small and medium-sized biotechnology firms in Taiwan, adopting knowledge management process, bio-KMS effects and BIP as major constructs. According to Yin (1989), descriptive or exploratory research techniques only apply to case study analysis. Hence, we will use case study as the approach, and interview members in biotechnology firms in the attempt to investigate the characteristics of knowledge innovation and management, and operations modes of these firms. The business scope of Firm A is the research and development transgenic fish.

4. CASE STUDY ANALYSIS

4.1. Case Background

Firm A was established in 1977, and is the first aquarium biotech corporation that has been ISO-9001 and ISO-14001 certified. Its aim is to establish a biotech industry integrating life science, biological education, art, and pressure relief. In 1996, firm A established a 'Biotech Research Center' and worked with science institutes around the world to develop advanced biotech products. With quality products delivered to almost 50 countries, firm A positions itself as "The Aqua Leader of New Age". With twenty years experiences in the field, firm A is devoted to satisfy customers by producing quality products. In 2001, firm A earned its global fame with its first neon fish, and in the same year, established the first transgenic fish genus center in Asia, which made great contributions to the study of transgenic fish genus. In 2004, firm A created 'Transgenic Fish Technical Platform' and was awarded 'Outstanding Award' by CAITA (Chinese Association for Industrial Technology Advancement).

4.2. KMS Structure

The KMS structure of firm A is shown in Figure 2.

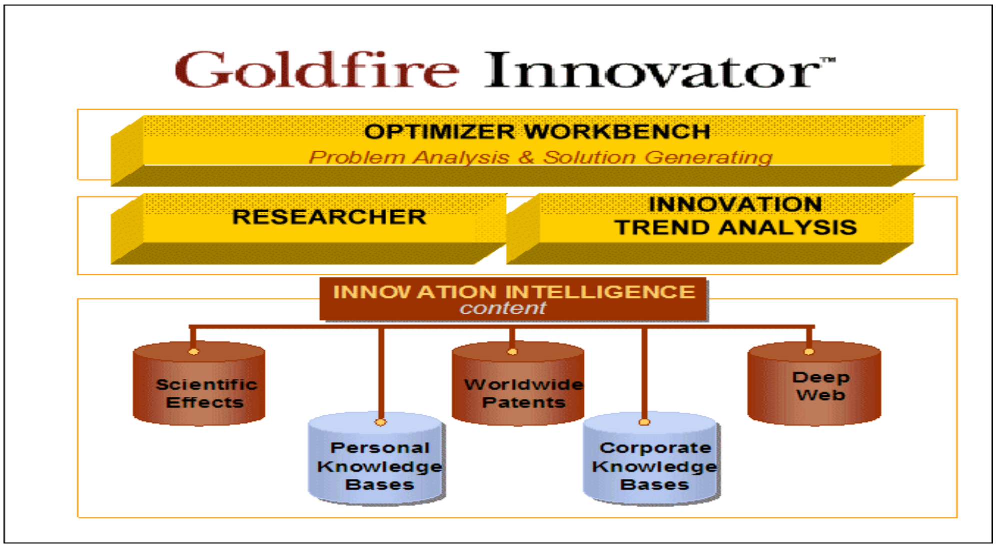


Figure 2. KMS Structure of Firm A

5. CONCLUSION & IMPLICATIONS

5.1. Conclusion

Summarizing the discussions, we conclude as follows:

(a) Knowledge-based Perspective: A very large percentage of people in the firm are R&D researchers, while marketing, accounting, laws and general assistants the next. These people have professional expertise, so they value information and knowledge very highly. Their information or knowledge, which conveys conceptual problems, solutions, even the experimental procedures, is mainly written in free-style, and is put in simple words. Meanwhile, knowledge stored in KMS is immediate information. To reduce the complexity of information, the information system will filter massive data according to the experiences (individual, work group, or organization) or business rules, and turn it into organizational knowledge as the reference for the practice of future project or experiments.

(b) Technology-based Perspective: Biotechnology firms give much attention to patents and IP rights, and searches for information such as genetic data, patents, prescriptions, even competitors' information become relatively essential. Hence, firms employed information technology to assist KM processes implementation (knowledge creation, storage/retrieval, transfer and application), by which knowledge can be distributed among individual, group and organizational levels. The IT tools the firms employed involve Internet (such as on-line database of patents /genetic data), databases (storage), data mining (gene mapping, decision making systems), and knowledge maps. With these tools, the firms not only systemize and codify information, but also enhance knowledge sharing and R&D and innovation capabilities.

5.2. Managerial Implications

Biotechnology is a synthetic technology comprising of natural science and engineering, and requires multiple techniques and research for product improvement or research. Spalding (1986) postulates that the enormous research expenditure of biotechnology needs individual or public's financial support. Roberts & Mizouchi (1989) posit that biotechnology industry has four characteristics. Firstly, it requires huge investment in resources; then, biotechnology synthesizes multidimensional knowledge; It requires long-term and stable financial support; small-sized biotechnology firms, in particular, must rely on sales channels, marketing techniques to promote their products. Finally, biotechnology industry can proceed with research contracts, technology authorization, alliances, or joint ventures. Oakey (1993) points out small-sized firms often create key techniques, and government should provide funding to ensure they have long-term research resources. When the new techniques have the protection of patents, government should enact a law to protect the product rights. With the liberalization of trade (Gibbs, 2000), which creates a large free trade market, products not only earn a high rate of return, but also a long product life cycles. Nevertheless, bioscience products are strongly related to our health, and require non-compromised quality control, a strict regulatory governance and supervision of nation's government. In a word, biotechnology is a technology-intensive, integrate, and multi-dimensional knowledge industry (Hsieh, 2002). Previous literature and empirical studies in this paper shows that 20% of the topics concern information technology, while 80% of the topics are concentrated on organizational context. In other words, when the organization deems knowledge the most important intangible assets of the firm, it should start reforming the existing organizational structure, culture in compliance with the industrial characteristics and market structure. More importantly, it is suggested that the set up of an environment that supports knowledge sharing and transfer will play an important role. Under the supportive environment employees would love to share and communicate, which improves the knowledge innovation and transfer and corporate competence.

According to the above statements, we conclude two implications:

(a) The use of KMS may enable knowledge-intensive industries to improve business innovation competence.

(b) The implementation of KMS may help the biotechnology firm to establish its unique knowledge assets and overcome the brain drain of professional personnel.

Moreover, organizational context factors should be added to the further exploration that may dig out the influences of more interdependent variables for this topic.

REFERENCES

1. Alavi, M. & Leidner, D. E. (2001). Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly, 25*(1), 107-136.
2. Chou, D. C. & Lin, B. (2002). Development of Web-Based Knowledge Management Systems. *Human Systems Management, 21*(3), 153-158.
3. Chuang, S. H. (2004). A Resource-Based Perspective on Knowledge Management Capability and Competitive Advantage: An Empirical Investigation. *Expert Systems with Applications, 27*(3), 459-465.
4. Cohen, W. M. & Levinthal, D. A. (1990). Absorptive Capacity: a New Perspective on Learning and Innovation. *Administrative Science Quarterly, 35*(1), 128-152.
5. Couger, J. D. (1988). Key Human Resource Issues in the 1990s: Views of IS Executives versus Human Resource Executives. *Information & Management, 14*(4), 161-174.
6. Cross, R. & Baird, L. (2000). Technology is not Enough: Improving Performance by Building Organizational Memory, *Sloan Management Review*, 69-78.
7. Gardan, N. & Gardan, Y. (2003). An application of knowledge based modelling using scripts, *Expert Systems with Applications, 25*(4), 555-568.
8. Grant, R. M. (1996). Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration, *Organization Science, 7*(4), 375-387.
9. Huber, G. P. (2001). Transfer of Knowledge in Knowledge Management Systems: Unexplored Issues and Suggested Studies. *European Journal of Information Systems, 10*(2), 72-79.
10. King, W. R., Marks, P. V. & McCoy, S. (2002). The Most Important Issues in Knowledge Management. *Communications of the ACM, 45*(9), 93-97.
11. Laura, B. C., Todd, M. A. & Scott, F. T. (2001). Knowledge Codifiability, Resources, and Science-Based Innovation. *Journal of Knowledge Management, 5*(2), 195-204.
12. Liebowitz, J. (2002). Facilitating Innovation Through Knowledge Sharing: A Look at the US Naval Surface Warfare Center-Carderock Division, Special Issue on Knowledge Management in E-Commerce. *Journal of Computer Information Systems, 42*(5), 1.
13. Liebowitz, J. (2004). *Addressing the Human Capital Crisis in the Federal Government: A Knowledge Management Perspective*. Butterworth-Heinemann Business Books / Elsevier (Knowledge Management Series).
14. Niederman, F., Brancheau, J. & Wetherbe, J. (1991). Information Systems Management Issues for the 1990s. *MIS Quarterly, 15*(4), 475-500.
15. Nightingale, P. (2000). Economies of Scale in Experimentation: Knowledge and Technology in Pharmaceutical R&D. *Industrial and Corporate Change, 9*, 315-359.
16. Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science, 5*(1), 14-37.
17. Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the*

Dynamics of Innovation. New York: Oxford University Press.
18. Nowak, R. (1995). Entering the Post Genome Era. *Science, 270*, 368-371.
19. Oakey, R. P. (1993). Predatory Networking: the Role of Small Firms in the Development of the British Biotechnology Industry. *International Small Business Journal, 11*(4), 9-22.
20. Souder, W. E. (1988). Management Relations Between R & D and Marketing in New Product Development Project. *Journal of Product Innovation Management, 5*, 6-19.
21. Stata, R. (1997). Organizational Learning- The Key to Management Innovation. *Sloan Management Review, 30*, 63-73.
22. Sunassee, N. N. & Sewry, D. A. (2002). The Theoretical Framework for Knowledge Management Implementation. in *Proceedings of the 2002 South African Institute of Computer Scientists and Information Technologists Annual Conference*, Pretoria, South Africs, Devilliers, A., Conradie, M. and Singh, S. (Eds.), 235-245.
23. Tushman, M. & Nadler, D. (1986). Organizing for Innovation. *California Management Review, 28*(3), 74-93.
24. Yin, R. K. (1989). Case study Research: Design and Method. California: Sage Publication.
25. Zawacki, R. A. (1993). Key Issues in Human Resource Management. *Information System Management, 10*(1), 72-75.

HSM: A Hierarchical Spiral Model for Knowledge Management

Zhaohao Sun, Gang Hao

School of Economics and Information Systems, University of Wollongong
Wollongong NSW 2522 Australia
Email: zsun@uow.edu.au, gh997@uow.edu.au

Abstract

Knowledge is the most important asset for organisations today. Knowledge management (KM) has become one of the most moving research and development fields in information systems (IS), business management and commerce development. There are many researches in the technical and strategic aspect of KM. However, how to model KM is still a big issue. This paper will fill this gap by providing a hierarchical spiral model (HSM) for KM. The proposed model will facilitate the research and development of KM and IS.

Keywords:
Knowledge management, Process models, Knowledge-based systems, Information Systems.

INTRODUCTION

The flourishing interests on knowledge management (KM) have recently led to a rush pursue on organizational knowledge in the business world (Coakes, 2003). For example, there are already a large number of KM initiatives implemented in organizations. Typical approaches to these initiatives were using information technologies (IT) for managing organization-wide knowledge resources. Building a KM system with internet technologies or creating a knowledge repository system with database technologies is very common (Radding, 1998). In order to better use IT in KM, it is necessary to model the KM process in a more systematic way. Process orientation (Ahn & Chang, 2004) is a perspective widely accepted in organization science. Recently, there have been a number of attempts to integrate KM and process orientation (Liebowitz, 1999). However, there are few KM process models that are based on system development methodology. Whereas, KM software must be embedded in stage model, in order to attain a seamless design, regarding the special qualities and requirements of knowledge work, detailed studies of the existing processes and analysis of the used knowledge are necessary. Therefore, this paper will propose a hierarchical spiral model (HSM) for KM in organisations, based on the waterfall model and the spiral model for KM.

The rest of the paper is structured as follows: Section 2 discusses KM as a process. Section 3 reviews three kinds of process models for KM. Section 4 proposes a new hierarchical and spiral model for KM: HSM, and discusses the processes in the HSM and their relationships. Section 5 concludes the paper with some concluding remarks.

KNOWLEDGE MANAGEMENT AS A PROCESS

Knowledge management (KM) implies that knowledge can be managed. KM is a set of processes directed at "creating, capturing, storing, sharing, applying, reusing" knowledge (Wiig, 1997). This definition is criticized for making KM to involve somewhat mechanistic and sequential process steps and for focusing attention on explicit knowledge artefacts as opposed to tacit knowledge, although knowledge engineering (KE) reflects this view of KM. There are also other definitions of KM. For example, KM is the systemic and organizationally specified process for acquiring, organizing, and communicating knowledge of employees so that other employees may make use of it to be more effective and productive in their work (Davenport, et al, 1998). KM provides the perspectives and approaches to put investments made in data and information to better use where it is needed most. KM is the science of collecting organizational knowledge and, by recognizing and understanding relationships and patterns, turning it into usable, accessible information and valuable knowledge (Loshin, 2001). KM can be also considered the process of "delivering the right knowledge to the right persons at the right time" (Coakes, 2003), although it is criticised to be applied to an outdated business model (Hildebrand, 1999). Alternative definitions have been attempted to better capture the complexities of knowledge and KM. The effective KM is the creation of management processes and infrastructure to bring together knowledge and communities in a common ecology that will sustain the creation, utilization and retention of knowledge. Nonaka (1994) recognises that KM must address both explicit and tacit knowledge as well as their interaction, and begins to address some of the mechanism for doing this. It does not, however, capture all aspects of KM, nor does it address how knowledge will be used or how a knowledge-based enterprise will ultimately function. More specifically, KM is developing and managing integrated, well-configured knowledge systems with embedding work systems (Barth, 2000). It is clear that both knowledge systems and their embedding work systems can be managed. Finally, this definition is broad enough to capture all aspects of KM but is not overly vague – one can define, with some precision, what a knowledge system is. One can also articulate how work systems can become embedded within knowledge systems.

It should be noted that there are two main motivations for an organisation to implement KM. The first one is to generate, communicate, preserve and share knowledge internally and externally. This means that the organisation can prevent knowledge loss and develop asset. Furthermore, good KM enables organization to retain critical expertise and prevent critical knowledge loss resulting from retirement, downsizing, employee departures and changes by building an organizational economy (Liebowitz, 1999). It also improves the organization's ability to capitalize on legal protection for intellectual property. The second one is to ensure a vital workforce and promote human capital. The basic benefits of KM are to improve productivity and gain competitive advantage through embedding KM processes into daily work activities. By capturing, sharing, and generating knowledge to stimulate innovation and achieve success, KM enables the organization to fully understand its customers, prospects, the market, and the competition (Yildizoglu, 2001). Effective KM also helps to keep good relationships with clients by increasing customer knowledge, expediting response to customer queries, suggestions, and complaints. It also ensures improved consistency and quality when serving customers.

PROCESS MODELS FOR KNOWLEDGE MANAGEMENT

As IT and Information system (IS) play an increasing important role in KM, it is necessary to deploy IT and IS to develop KM. During the development of KM, there have been many models for KM (Chua, 2004). Blumentritt and Johnston (1999) propose a cyclic model for KM, while Sun (2004) proposes a waterfall model for KM. Nonaka (1994) proposes a spiral model of KM. In what follows, this section will review these three kinds of process models.

A Cyclic Model of Knowledge Management

Blumentritt and Johnston (1999) attempt to draw together key thinking in the area of KM to enable a distinction to be drawn between information and knowledge, and provide a framework encapsulating the various strands of thinking about knowledge into an overarching structure. They believe that KM has been concerned with knowledge creation, acquisition and flow, and develop their own framework to categorise knowledge. Within this framework there is an increasing level of difficulty associated with transforming knowledge into information. They challenge the contention that IS may be used interchangeably with KM systems. Instead, they argue that knowledge can only exist within the mind of the individual. By creating a clear distinction between knowledge and information they attempt to develop a model which shows the links between knowledge and information. One of the models that they have developed bears great similarity to Nonaka's knowledge spiral model (Nonaka, 1994), showing it within the context of the transformation of knowledge into information and back through the cycle to knowledge.

Nissen, et al (2000) has developed another life cycle model which makes flow time explicit and supports a multidimensional framework that enables a new approach to analysis and visualization of diverse knowledge-flow patterns in the organisation. This model describes "a continuous cycle with six phases of knowledge flowing through the organisation: 1) creation, 2) organization, 3) formalization, 4) distribution, 5) application, and 6) evolution" (Nissen & Levitt, 2002). Briefly, the creation phase begins the life cycle, as new knowledge is generated within an organisation, similar terms from other models include capture and acquisition. The second phase pertains to the organization, mapping or bundling of knowledge, often employing systems such as taxonomies, ontology and repositories. The third phase addresses mechanisms for making knowledge formal or explicit; similar terms from other models include store and codify. The fourth phase concerns the ability to share or distribute knowledge in the enterprise; this also includes terms such as transfer and access in other models. Knowledge use and application for problem solving or decision making in the organization constitutes Phase 5, and 6 and covers knowledge refinement and evolution, which reflects organizational learning, and thus a return to knowledge creation through time. It should be noted that the above models are generally iterative and involve feedback loops between stages.

A Waterfall Model of Knowledge Management

One of the concerns of KM should be the management of the KM process itself, which has been ignored in the existing models to some extent. In order to resolve this issue, Sun (2004) divides the main stages of KM into more specific processes. KM goes far beyond the storage and manipulation of the data, or even of information. KM is a discipline that focuses on knowledge processing and corresponding management which permeates each of following processing stages:

- Understand and use knowledge
- Discover and expand knowledge
- Capture, and acquire knowledge from a variety of sources
- Select, filter and classify the existing knowledge
- Define storage structures for saving knowledge

- Design ontology of knowledge
- Generate, adapt and/create new knowledge
- Measure and/or evaluate knowledge
- Visualized knowledge
- Distribute and/or transfer knowledge to other organization or individuals
- Recommend, share, utilize/apply and sell knowledge
- Retain and maintain knowledge as an asset

The management of knowledge processing for each processing stage includes analysis, planning, support, collaboration, coordination and all management functions. Based on all these aspects of KM, Sun (2004) states a waterfall model to explain the function of KM. From this model, it is easy to see the processing stages in KM, while each processing stage can be regarded as each component in the KM. The sequence of the processing stages is also pragmatic rather than precise; the processing stages may also not be executed sequentially, because some later processing stages sometimes are a basis for some former process stages. For example, knowledge ontology is the basis for further processing knowledge including storage, generation, creation, and classification. Knowledge creation sometimes can be the result of knowledge sharing, while knowledge sharing also could lead to new knowledge creation. Therefore, these processing stages in this model are interrelated logically and systematically. This model can help to overview the overall functions of KM and understand the interrelationship between process stages. However, this waterfall model can only cover some aspects of KM, an iterative or spiral processing models of KM is needed to reflect the interrelationship between the processing stages in the waterfall model.

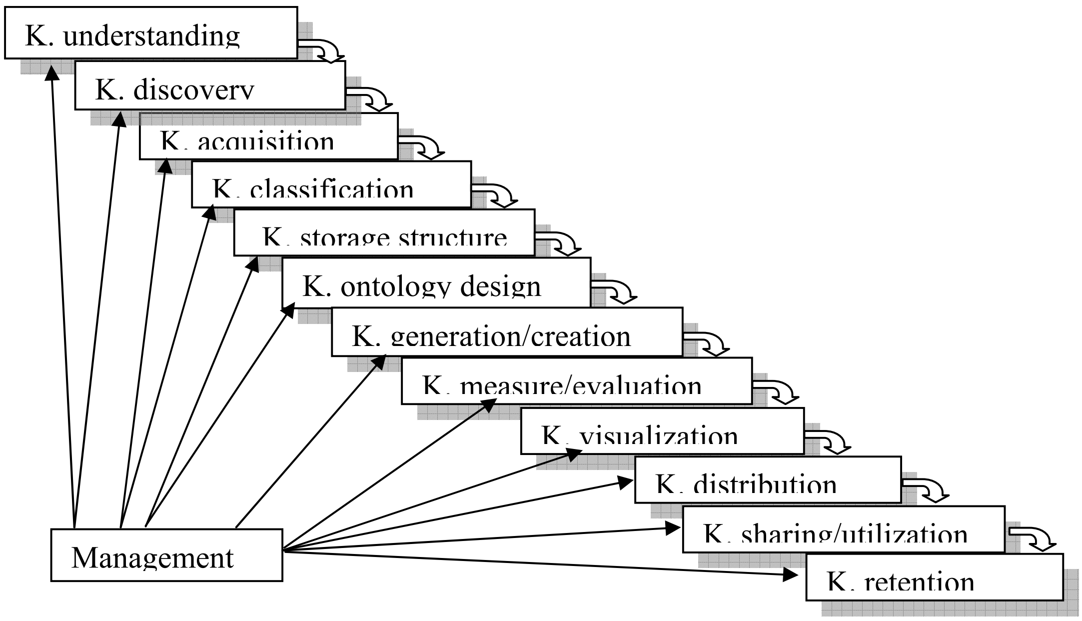


Fig. 1. A waterfall model of knowledge management

A Spiral Model of Knowledge Management

Nissen (2002) and Nonaka (2004) propose a spiral model of dynamic interaction between tacit and explicit knowledge and characterize four processes (socialization, externalization, combination and integration) that enable individual knowledge to be amplified and effect organizational knowledge crystallization. Nissen and Levitt (2002) adapt Nonaka's spiral model to a dynamic model of knowledge flow. This model intends to describe how knowledge flow through the modern organisation, and what kinds of managerial interventions can be made to enhance the flow of knowledge. The model embeds the knowledge flow to organisation's everyday work flow to distribute knowledge through the organization. Nissen and Levitt (2002) believe that KM is unlikely conducted effectively before the phenomenon of knowledge flow can be understood. Comparing with Nonaka's work (Nonaka et al, 1996), Nissen and Levitt (2002) introduce dynamic representation of knowledge flow dynamics taking into account computational organization theory so that their model provides much greater fidelity and insight into knowledge flow dynamics, and enables the execution and performance of diverse knowledge-work processes to be simulated for analysis and comparison. This dynamic model illustrates how to enhance the flow of knowledge through the modern enterprise, as shown in Fig. 2.

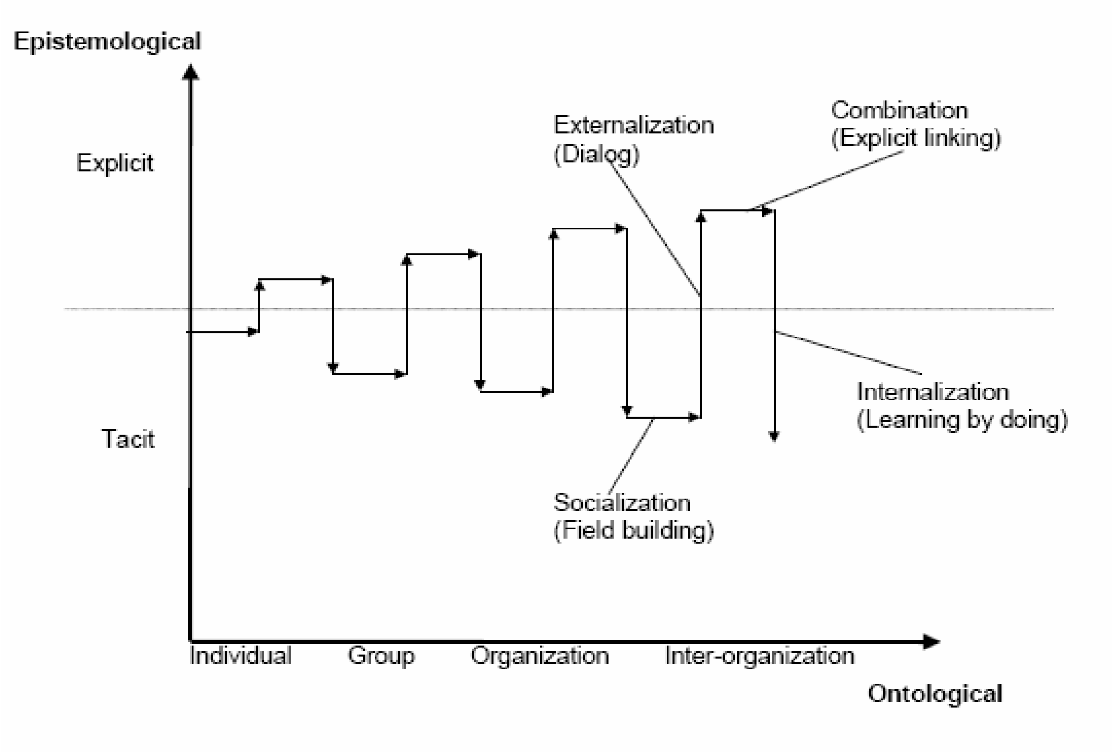


Fig 2. A spiral model of KM (Nonaka, 2004)

This model describes continuous and routine flows that comprise the bulk of organizational knowledge work. It delineates the interaction between epistemological and ontological dimensions used by Nonaka as the principal means for describing knowledge as it flows through the enterprise (Nissen & Levitt, 2002). As noted above, this flow is characterized by four enterprise processes: socialization (tacit to tacit), externalization (tacit to explicit), combination (explicit to explicit), and internalization (explicit to tacit). The related trigger concept (Nonaka, et al, 1996) is also integrated into the figure to show where each knowledge conversion process is induced by one of four triggers: field building, dialog, linking explicit knowledge, and learning by doing. Briefly, socialization denotes members of a team sharing experiences and perspectives, much as one anticipates through tightly connecting workgroups and communities of practice. Externalization denotes the use of metaphors through dialog that leads to articulation of tacit knowledge and its subsequent formalization to make it concrete and explicit; such dialog or what Nonaka refers to as collective reflection is described as inducing externalization. Combination denotes coordination between different groups in the organization, along with documentation of existing knowledge to link and combine new intra-team concepts with other explicit knowledge in the organisation. Internalization means that diverse members in the organization apply the combined knowledge from above, often through trial and error, and in turn translate such knowledge into tacit form at the organization level. As suggested by the repeating pattern delineated in the figure, such interaction between “triggers” and conversions enables a continuous “spiral” development of knowledge.

The above examination of related work on process models of KM shows that there is no single existing model that addresses all the important processes of knowledge. Further, the components or stages in KM have not been classified in such categories that they can be treated in a hierarchical way, which is useful and practical in organisational practice. For example, in the waterfall model of Sun (2004), the knowledge processes have the same importance from a KM viewpoint without taking into account hierarchy of the knowledge processes. Spiral development of KM has also not examined this issue in detail in that model. In what follows, we will propose a new process model for KM in order to resolve these issues.

HSM: A HIERARCHICAL SPIRAL MODEL FOR KNOWLEDGE MANAGEMENT

There are some indispensable processes of knowledge management (KM). These processes have appeared in a number of existing process models in KM. Different models have different emphases on certain amount of the process stages of KM. For convenience, the hierarchical spiral model (HSM) proposed here is mainly based on Sun's waterfall model (Sun, 2004) and Nonaka's spiral model of KM (Nonaka, 1994). The major process stages of the HSM are extracted from the waterfall model. The processes are divided and categorized to five main processes and eleven sub-processes from a hierarchical viewpoint, which form a spiral within three main processes. The proposed HSM provides the guidance between the different phases of KM activities. The main processes present the operational process stages carrying out in real KM activities such as knowledge selection, sharing, update, etc. The sub-processes demonstrate how these main processes been done in detail and show the relationship between each other. Fig. 3 shows the proposed model by illustrating its processes and their basic relationships. The main operational processes consist of knowledge selection, knowledge creation, knowledge sharing, knowledge preservation and retention, and knowledge update (Chua 2004). Three of the main processes consist of several sub-processes. The arrows connecting the processes denote the interaction of knowledge flows. In what follows, we will examine the main processes and the sub-processes in the model respectively in some detail.

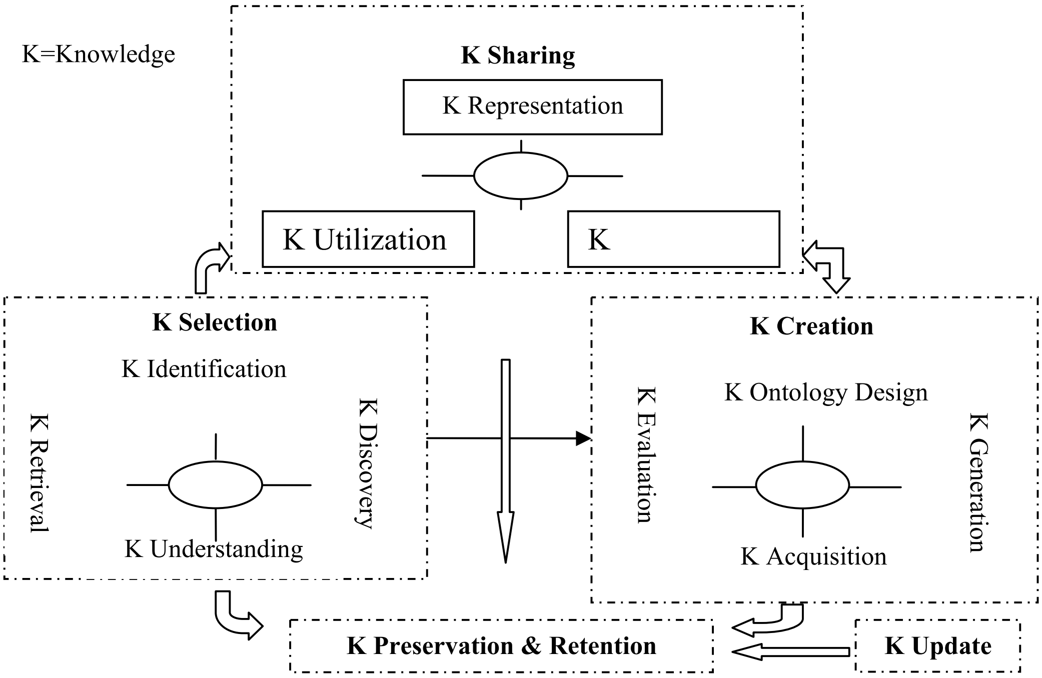


Fig. 3. A hierarchical spiral model for knowledge management

Main Knowledge Processes in HSM

As the first main process of KM, *knowledge selection* identifies knowledge needs by understanding, and select useful knowledge from the existing repository. This process supports to make knowledge reachable, and filtrate the useful knowledge from an organization's existing knowledge and make it easy to search and find. The right knowledge has to be found, retrieved and absorbed. Searching is a significant issue if there is a large amount of knowledge available and the right knowledge becomes difficult to find. Identifying potentially valuable knowledge is of importance, since a huge amount of knowledge stored will eventually lack trust of people (Blumentritt & Johnston, 1999). Knowledge selection involves several sub-processes, which refer to as knowledge identifying, understanding, discovering, and retrieving. These four sub-processes are not necessarily followed in a strict sequence, but rather there can be overlaps and iterations.

If the identified knowledge needs can not be satisfied with the existing knowledge or when the required knowledge is unavailable in an organization, then the knowledge flow goes to another main process of KM: *Knowledge creation*, as shown in Fig. 3. This process supports generation and creation of knowledge. Development of new knowledge in an organization focuses on creating new products, better ideas, more efficient services or new skills. The knowledge creation is desirable if the existing knowledge does not fit needs or is basically too expensive. New knowledge is usually generated in the research department in this process. If the identified knowledge is reachable outside the organization, then the knowledge will be acquired from other organisations. These two ways of creating knowledge are described as knowledge generation and acquisition. Before generating and acquiring new knowledge, the frame of knowledge should be established by ontology design, which is another sub-process supporting knowledge creation. The importance of knowledge creation depends on organization culture, organization objectives and innovation/research efforts (Davenport & Prusak, 2000). Approaches to knowledge creation are techniques and tools of data mining, knowledge discovery, knowledge-based systems, and machine learning (Becerra-Fernandez et al, 2004).

The third main process in HSM is *knowledge sharing*, which is fulfilled after existing knowledge has been identified or the new knowledge has been created. This main process is regarded as the core process of KM, because one of the primary objectives of KM research and practice is to foster the flow of knowledge among organization members (Chua, 2004; Shin, 2004). Knowledge sharing is performed by distribution and utilization of the knowledge that has been selected or generated from the organisation and acquired outside. When sharing knowledge, new knowledge is often created by combining the shared knowledge and existing knowledge (Davenport & Prusak, 2000). The purpose of the first two main processes is to provide useful knowledge for sharing knowledge. They contribute to knowledge sharing by selecting existing knowledge and creating new knowledge.

The new knowledge has to be stored in the fourth main process: Knowledge preservation and retention. Knowledge preservation aims at retention of knowledge assets. The new valuable knowledge has to be stored from time to time. This has to be accomplished by efficient storage media to access knowledge, to prevent valuable expertise from disappearing. The importance of knowledge preservation depends on the viscosity of knowledge to store, amount of knowledge accruing, organisation objective, infrastructure and culture (Huber, 1991). Finally, the knowledge needs to be updated frequently because the knowledge becomes obsolete rapidly in the knowledge society. Nowadays, the transformation of the world and technology forces the organization to renew and update knowledge in time. Otherwise, the use of obsolete knowledge would mislead and cause negative influence on organization.

Sub-Processes of Knowledge Management in HSM

There are four sub-processes in the main process of knowledge selection; that is, knowledge identification, knowledge discovery, knowledge understanding and knowledge retrieval.

Knowledge identification is the first sub-process of knowledge selection. This sub-process identifies the need for knowledge, and determines it. Before knowledge can be created or shared, the need for knowledge has to be identified. Further requirements have to be determined to find the right knowledge in the case of sharing and to enable the creation of the right knowledge in the case of creation (Radding, 1998). The importance of knowledge identification in an organisation depends on organisation objectives, infrastructure and organisation culture (Davenport & Prusak, 2000). Knowledge identification also refers to recognize and realize the needs of knowledge within an organisation. When a knowledge seeker with a clearly formulated request pulls knowledge selection, knowledge identification primarily involves locating the resources from which the knowledge is to be captured (Coakes, 2003). As knowledge identification progresses the undesired resources, knowledge is filtered.

Knowledge discovery refers to finding valuable knowledge existing in the organization. After the needs are identified, it should first find the needed knowledge within organisation. There is much useful knowledge that is immerged within the organization. Knowledge discovery is to mine this valuable intellectual capital from documentation, database and mind of experts. Some tools and techniques like data mining and interview are useful for knowledge discovery. Data mining can help knowledge seeker to discover desired knowledge or maybe some unexpected useful knowledge from the huge databases of the organization (Becerra, etc, 2004). Interview with incentive encourages knowledge and experience holders to express the knowledge they possess.

Knowledge understanding refers to comprehending the discovered knowledge. People cannot use knowledge well if they do not understand what the knowledge is. The more employees know the knowledge well the better the knowledge can be used for the organisation. Understanding knowledge comes from a number of attributes, both of the knowledge itself and the format in which it is used, and the previous experience of the user. When accumulating knowledge there is a conscious choice to learn or discard the knowledge of others. Very often, you choose what you agree with and what does not challenge your own assumptions. This is a form of self-prejudice (Ahn & Chang, 2004). In fact, the mind always follows a certain pattern – knowledge can therefore gradually become formulaic.

Knowledge retrieval refers to extracting knowledge from the existing knowledge resources such as a document, database, a data warehouse, a computer system or an employee. In cases where knowledge to be selected resides in multiple identified resources, knowledge retrieval involves coordinated collection or gathering from a variety of resources such as databases or data warehouses. The sequencing and timing of retrievals within a collection effort can have a significant effect on what knowledge is captured. It is important that the correct knowledge be retrieved completely.

It should be noted knowledge recommendation is also involved in knowledge selection. Knowledge recommendation usually occurs after knowledge retrieval and before knowledge is accepted (Sun and Finnie, 2004). Just as product recommendation plays an important role in business management, knowledge recommendation will become an important part of KM, although no such study has been found so far, to the knowledge of the author.

The main process of knowledge creation consists of the following sub-processes: Knowledge ontology design, knowledge generation, knowledge acquisition, and knowledge evaluation.

Knowledge ontology design is to formalize the existing knowledge and offer a format for adding new knowledge. Ontology is an explicit specification of conceptualization. In Artificial Intelligence (AI), it refers to providing definitions for the vocabulary used to represent knowledge in a given domain. Knowledge ontology design formalizes the semantics of objects and relations in a knowledge universe of discourse and provides a set of terms, which can be used to examine these objects and their relationships (Coakes, 2003). This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents the knowledge in the knowledge universe of discourse such as business negotiation.

Knowledge generation is concerned with producing new knowledge. According to Nonaka (1994), in a strict sense knowledge can only be generated in people's minds. Knowledge generation can and should also take place in an uncontrolled manner. After the need for knowledge has been identified and it cannot be discovered internally, this represents the possibilities of acquiring or generating the desired knowledge. Knowledge generation also occurs when the knowledge cannot be acquired from outside organization. There are several ways of knowledge generation, for example, by sending people to training, have them read books or assigning a consultant that provides knowledge. Data mining and machine learning are intelligent techniques for knowledge generation. Knowledge sharing can also generate new knowledge by combining shared knowledge with the user's existing knowledge (Huber, 1991).

Knowledge acquisition supports to obtain knowledge. No organisation is able to produce all needed knowledge itself. So the knowledge that an organisation can not develop itself has to be acquired. Acquiring knowledge can be accomplished by acquiring innovative organisations, recruiting experts, buying documents from outside sources, hiring consultants, buying patents, and retrieving the Web using search engines such as google.com, and so forth. Relationships with customers, suppliers, competitors and partners do also serve potential external sources for knowledge. The importance of knowledge acquisition depends on organisation culture and objectives. Methods to implement knowledge acquisition are i.e. data bases containing indices of external sources potentially valuable for the organisation.

Knowledge evaluation needs to be conducted after the knowledge has been generated internally or acquired from outside. The new knowledge that has been generated or acquired needs to be evaluated to make sure that the new knowledge is correct and valuable before it can be shared in the next main process. Usually the acquired knowledge is often correct, because it is the knowledge used in other organisations. However, this does not mean it is unnecessary to be evaluated, because knowledge good for other organisation may not fit the acquired organization. For the generated new knowledge, it is often the immature knowledge with many mistakes, the evaluation should be conducted to identify whether this new knowledge is worth further development.

The main process of "Knowledge Sharing" contains the following sub-processes: knowledge representation, knowledge distribution and knowledge utilization.

Knowledge representation (KR) means to represent the knowledge in a more clear and storable way. KR includes activities such as documentation, conversion, exhibition and formalization. A willingness of documentation helps those unknowns who follow later. According to Tsoukas (2002), explicit knowledge is most easily documented as writings, or in electronically, digitalized form. Tacit knowledge, on the other hand, remains in the mind of the holder and must be exchanged live with another person via a dialog. KR is an important research field in AI (Nilsson, 1998). Semantic network, frames, concept mapping, and predicate logic, to name a few, are KR techniques for representing knowledge in intelligent systems.

Knowledge distribution supports the spread of knowledge. Knowledge has to be made available throughout the organisation. That is spreading and sharing know-how which is already presented within the organization. This function goes hand-in-hand with knowledge representation. A distribution process may incorporate a variety of techniques ranging from books, reports, visual identity, correspondence, and electronic communications (Kikoshi & Kikoshi, 2004). No matter what techniques are used for spreading the knowledge, distribution process should make sure that the message is received by the knowledge seeker, but it does not guarantee the spreading knowledge is well understood, which is done by knowledge understanding in the main process of knowledge selection.

Knowledge utilization supports knowledge application. Simple availability does not guarantee that present knowledge is indeed used. Knowledge representation and distribution is a precondition to successfully apply knowledge. This still does not ensure utilization, but the chance of usage of highly available and distributed knowledge does increase. Moreover, utilization means to assist knowledge workers to apply implemented knowledge. The importance of knowledge utilization depends on the complexity of problems, organisation culture, trust of knowledge sources and organisation infrastructure. Furthermore, the design of system interfaces may greatly affect utilization.

Knowledge Flowing in HSM

In the previous sections, the main processes, sub-processes and their functions have been discussed. This subsection will examine the relationships among the processes by looking at the knowledge flowing among the processes in the HSM. First of all, it will look at the knowledge flowing among the main processes.

The knowledge management (KM) flow starts with *knowledge selection*. In this process, the knowledge needs is identified and the useful knowledge is understood and retrieved within organization. If the needs of knowledge can be satisfied by the discovered and retrieved knowledge in the organization, then the management process can go to the next stage: knowledge sharing. Otherwise, the knowledge flow goes to the process: knowledge creation. In this process, new knowledge is created by generation within the organization or acquired from outside. The created knowledge is also evaluated in the process. After that, the new knowledge can be shared in the knowledge sharing process. In the sharing process, the selected or created knowledge is firstly represented to let people know. Consequently, knowledge is distributed by various ways. Finally, the distributed knowledge can be utilized by the knowledge seekers. The processes of knowledge selection, creation and sharing are all linked to knowledge preservation and retention. All the retrieved and generated knowledge is stored and maintained in this process. Knowledge is then updated and saved again, because it has to be adapted to the needs and context of the organisation.

The sub-processes in the main process of knowledge selection are knowledge identification, knowledge discovery, knowledge understanding and knowledge retrieval. The knowledge flowing starts with knowledge identification. After the needs are identified, the required knowledge needs to be found in the organization. This is called knowledge discovery. The discovered knowledge cannot be used if it can not be understood by employees. So the next sub-process is to understand the discovered knowledge. At the end of this main process, useful knowledge is retrieved and organized for sharing and preserving. In practice, these four sub-processes are interrelated. For example, when retrieving identified knowledge, some related knowledge needs to be discovered. Then the process goes back to the knowledge discovery to find the important knowledge.

Knowledge creation has also four sub-processes: Knowledge ontology design, knowledge generation, knowledge acquisition and knowledge evaluation. Knowledge ontology design can be considered as a bridge between knowledge selection and knowledge creation, because it formalizes the existing knowledge and establishes a format for adding new knowledge. As the format is set, it is ready for knowledge generation. This means to encourage employees to produce new knowledge for the knowledge needs. If the required knowledge cannot be generated within the organization or cost too much, then the knowledge flow goes to the knowledge acquisition. However, the sub-process of knowledge acquisition does not have to be followed by knowledge generation. It may follow the ontology design process. Actually, the relationship between knowledge generation and knowledge acquisition process is parallel. These two sub-processes are linked to knowledge evaluation. The new knowledge is evaluated in this process for the next main process: knowledge sharing.

There are three sub-processes in knowledge sharing. As the first sub-process of this main process, knowledge representation is to reveal and demonstrate selected and/or created knowledge. This makes the knowledge seeker find the required knowledge easier. The knowledge distribution process continues to deliver knowledge to the employees that need it. Finally, knowledge can be used in the knowledge utilization process.

CONCLUSIONS

The compelling knowledge needs require the organization to pursue knowledge as much as possible. However, the problems of generating and transferring knowledge between individuals, across organisations and from each generation to the next are substantial and complex. Pursuing knowledge without effective management may lead to overlap in work and assignments, taking into account the fact that it requires similar processes for knowledge creation, documentation, storage, and dissemination. This paper reviewed three different types of process models: a cyclic model, a waterfall model and a spiral model. Then it proposed the HSM: a hierarchical spiral process model based on the waterfall model and the spiral model for KM. The HSM provides an approach to managing knowledge processes in a hierarchical and spiral way so that the importance of each knowledge process in the KM can be emphasised in a different way. Future research could be related to integration of management as a process into the HSM and the economic analysis based on the proposed model.

We will also compare the proposed model with other existing models, validate and verify the proposed model in real business practice, and apply the proposed model to organizational learning. This could help the understanding of process modeling of KM.

REFERENCES

Ahn, J.-H., & Chang, S.-G. (2004). Assessing the contribution of knowledge to business performance: The KP3 Methodology. *Decision Support System,* 36, 403-416.

Avison, D. E., & Fitzgerald, G. (2003). *Information Systems Development: Methodologies, techniques and tools*, 3rd ed, McGraw-Hill International (UK), London.

Barth, S. (2000). Defining knowledge management. *CRM Magazine*, 4 July 2000.

Becerra-Fernandez, I., Gonzalez, A, & Sabherwal, R. (2004). *Knowledge management: Challenges, solutions, and technologies*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Bergeron, B. (2003). *Essentials of Knowledge Management,* John Wiley. & Sons, Inc. New Jersey.

Blumentritt R., & Johnston, R. (1999). Towards a strategy for knowledge management. *Technology*, *Analysis and Strategic Management*, 11(3), 287–300.

Brule, J.F., & Blount, A. (1989). *Knowledge Acquisition*, McGraw-Hill. USA.

Chua A. (2004). Knowledge management systems architecture: A bridge between KM consultants and technologies. *International Journal of Information Management*, 24, 87-98

Coakes, E. (2003) Knowledge Management: Current issues and challenges, IRM Press. London

Davenport, T. H., De Long, D. W., & Beers, M. C. (1998). Successful knowledge management projects, *Sloan Management Review*, Winter 1998, 43-57.

Davenport, T.H., & Prusak, L. (2000) *Working Knowledge: How organizations manage what they know*. Boston, Harvard Business School Press.

Firestone, J.M., & McElroy M.W. (2003). *Key issues in the new knowledge management*. Elsevier Science. USA.

Hildebrand, C. (1999). Does KM = IT ? Sept 15, *1999 Issue of CIO Enterprise Magazine*. Available at http://www.cio.com/archive/enterprise/091599_ic_content.html.

Huber, G. (1991). Organizational learning: The contributing processes and the literatures, *Organization Science* 1, 88-115.

Kankanhalli, A., & Tan, B.C.Y. (2004). A review of metrics for knowledge management systems and knowledge management initiatives. In *Proc. of the 37th Hawaii Intl Conf on System Sciences*.

Kikoshi, C.K., & Kikoshi, J.F. (2004). The inquiring organization: Tacit knowledge, conversation, and knowledge creation: Skills for 21st-Century Organizations" London

Lehaney, B., & Clarke, S. (2004). *Beyond Knowledge Management*. Idea Group Inc. London

Liebowitz, J. (1999). *Knowledge Management Handbook*. CRC Press.

Loshin, P. (2001). Quick study of knowledge management. http://www.computerworld.com/databasetopics/ data/story/0,10801,64911,00.html. accessed 25 April 2005.

Nilsson, N.J. (1998). *Artificial intelligence: A new synthesis*. San Francisco, California: Morgan Kaufmann Publishers, Inc.

Nissen, M., Kamel, M.N., &Sengupta, K.C. (2000). Integrated analysis and design of knowledge systems and processes. *Information Resources Management Journal* 13 (1), 24-43.

Nissen, M., & Levitt, R. (2002). Dynamic models of knowledge-Flow dynamics, CIFE Working Paper No.76. Stanford University.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*. 5(1), 14-37.

Nonaka, I. (2004). *Hitotsubashi on knowledge management*, John Wiley. Singapore.

Nonaka, I., Takeuchi, H., & Umemoto, K. (1996). A theory of organizational knowledge creation, *International Journal of Technology Management*, 11(7/8), 833-845.

Probst, G., Raub, S., & Romhardt, K. (1999). *Managing knowledge*. Springer -Verlag, Berlin Heidelberg.

Radding, A. (1998). Knowledge management: Succeeding in the information based global economy, Computer Technology Research Corp.

Shin, M. (2004). Framework for evaluating economics of knowledge management systems, *Information & Management,* 42, 179-196

Sun, Z. (2004). A waterfall model for knowledge management and experience management. In: *Proc of 4th Iinternational Conference on Hybrid Intelligent Systems*, Japan, IEEE Press, pp 472-475.

Sun Z, Finnie G. (2004) *Intelligent Techniques in E-Commerce: A Case Based Reasoning Perspective*. Heidelberg, Berlin

Springer Verlag.
Tsoukas, H. (2002). Do we really understand tacit knowledge? http://is.lse.ac.uk/Events/ESRCseminars/ tsoukas.pdf. Accessed 30 June 2005.
Wiig, K. (1997). Knowledge management: Where did it come from and where will it go?" *Journal of Expert Systems with Applications*, 13(1) 1-14.
Yildizoglu, M. (2001). Connecting adaptive behaviour and expectations in models of innovation: The potential role of artificial neural networks. *The European Journal of Economics and Social Systems* 15 (3), 203-220.

BPM - Bridging the gap between Business Processes and technologies for Process Management

Dr. Olivera Marjanovic
University of New South Wales
Sydney, Australia
Email: o.marjanovic@unsw.edu.au

Abstract

The field of BPM (Business Process Management) is becoming increasingly relevant. However, there is a wide gap between the current understanding of BPM by business practitioners on one side, and technology developers on the other side. Unless this problem is addressed, developers will continue to provide business process support solutions that may not bring the expected business value to their adopters. The main objective of this paper is to contribute to a better understanding of this widening gap. The paper argues that it is necessary to use the knowledge dimension of a business process (rather than its structure) to assess applicability of different BPM technologies. Even more, this knowledge dimension will also determine possible end-user involvement and, most importantly, possible sources of competitive differentiation related to business process implementation in an organisation. The paper uses three representative types of business processes to demonstrate the proposed approach.

Keywords:

Business Process Management, Knowledge Perspective, Competitive advantage, Procedure and practices

INTRODUCTION

In recent years, BPM (Business Process Management) is emerging as a new field of applied research. However, a comprehensive review of very recent literature in this area confirms the existence of a wide gap between the current understanding of BPM by business practitioners on one side, and technology developers on the other side. For example, some of the leading IT practitioners in the area of business process technologies use the following definition: "BPM includes methods, techniques and tools to support the design, enactment, management and analysis of operational business processes… Business process management system is a generic software system that is driven by explicit process design to enact and manage operational business processes". (van der Aalst, ter Hofstede and Weske, 2003, pg 1.). Consequently, "many business analysts do not differentiate between BPM and workflow systems, or treat BPM systems as providing workflow management capabilities combined with enterprise application integration" (McGovern, 2004).

On the other hand, business practitioners have a completely different view of BPM. They see it as a management theory or strategy (i.e. how to create competitive advantage through business process integration). For example "BPM is commitment to expressing understanding, representing and managing a business (or the portion to which the theory is applied) in terms of an interdependent collection of business processes responsive to an environment of internal and external events" (McGovern, 2004). This particular definition sees business processes as dynamic and BPM as a knowledge discovery process. Furthermore, it is a business view that BPM practices should not force an organization to behave in a certain way. BPM should facilitate better understanding of business processes through application of the underlying theories and principles. Thus, BPM becomes a combination of relevant theories and the associated group of methods developed to facilitate management of business from the process perspective as well as management of business processes.

Obviously, the above two examples illustrate a wide gap between current understandings by two different groups. On one side business processes are viewed as predefined and routine operational processes, while on the other side the emphasis is on their dynamic nature and link with business strategy. However, it is important to point out that not all technology developers see business processes as routine and predefined. On the other hand, not all business people see them as evolving organisational processes, thanks mainly to the Business Process Reengineering era that has resulted in "very mechanistic" view of processes. However, the gap exists and is evident in the current literature as well as BPM practice.

Why it is important to attempt to bridge this gap? On one side, it is necessary to initiate development of different types of process management systems. On the other side to facilitate better understanding of current and existing business processes technologies both in terms of their limitations and opportunities. This will enable business community to better contribute to new developments of process-related technologies. Most importantly, bridging this gap will contribute to better alignment between business process management systems and company's strategy. This, in turn, will enable business practitioners to better understand how to create a sustainable competitive differentiation on the basis of their business processes and BPM.

The main objective of this paper is to contribute to a better understanding of this widening gap. The paper argues that a possible way to narrow this gap comes from better understanding of the knowledge dimension of business processes. This knowledge dimension enables us to evaluate possible use of the existing and emergent business process technologies, much better than the process structure that is widely used by technology developers. On the other hand, this knowledge dimension is also important to help us to better understand how to use business process support technologies to identify possible sources of competitive differentiation. The paper is based on the analysis of the state-of-the art developments in the field of BPM as well as author's extensive experience in this area.

RELATED WORK

This section describes common classifications of business processes (BP). It then illustrates how these classifications have been used by developers of various BPM systems.

Business process classifications

The following widely used classifications of business processes are often quoted by BPM's IT community and used as a basis for development of various business process support solutions.

The first relevant classification starts from the process structure. Thus Keen and Scott-Morton (1978) described *highly structured*, *semi-structured* and *unstructured* business processes. Highly structured processes are routine BP often found at the operational level. Other categories of BP are usually found at the higher organisational levels (tactical and strategic). Examples of "semi-structured" processes include cash management and brand management while unstructured processes include basic research and concept definition phase of new product development. Pava (1983) made a further distinction between unstructured and *unstructurable* business processes. In unstructured processes, like cash management, a degree of process structure can increase over time. On the other hand, in "unstructurable" processes, increased structure is neither possible nor desirable. This is because it will increase process rigidity, and at the same time, prevent creativity in problem solving. These processes involve building knowledge in a recursive and participatory manner through human sense making (Boland and Tenkasi, 1995). Because the term unstructured suggests that structuring is possible and desirable, while emergent does not, thus, the term *emergent* is a better label for many knowledge processes (Markus et. al. 2002).

Another widely used classification was introduced by (Mohan, 1997). Is also based on process structure. According to this classification all processes are categorised into four different types based on their *repetitiveness* and *business value.*

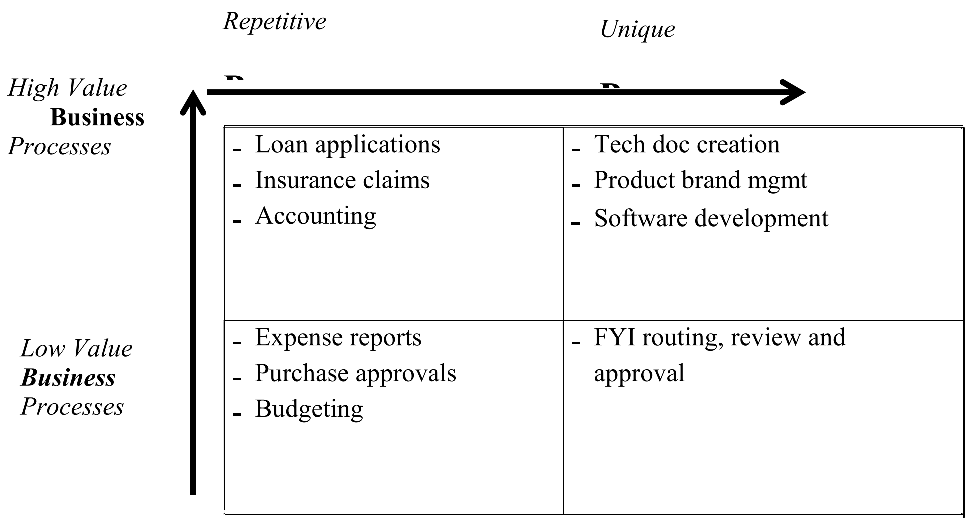


Figure 1. Business process classification (Mohan, 1997)

This classification emphasises the ability to predefine process structure i.e. to pre-define tasks and their order. This leads to a distinct separation between process modeling and process execution phases.

Another relevant classification, for the fist time, takes into account the knowledge perspective of business processes. Thus (More, 2000) classifies business processes on the basis of their *knowledge intensity* (as depicted by Figure 2). Knowledge intensity relates to the level of knowledge sharing, collection and reuse that is required during processes execution. According to this classification, the models of knowledge-intensive processes cannot be predefined as they evolve during process execution. Consequently, for those processes, it is not possible to separate process modeling from process execution phases as it is done in highly repetitive, routine processes.

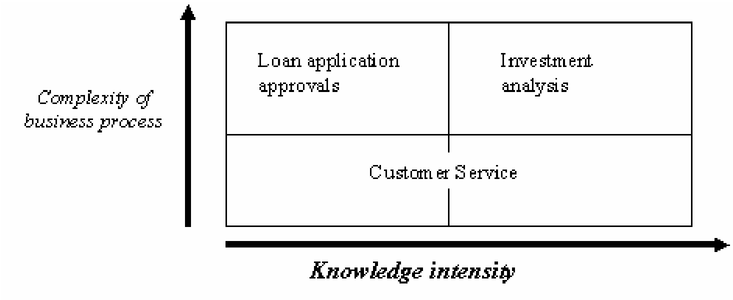


Figure 2. Knowledge intensity/complexity of business processes (Moore, 2000)

However, apart from knowledge intensity, this particular classification does not identify the type or possible source of this knowledge. Therefore, it does not provide enough details for possible differentiation of competitive advantage.

Technologies for business process support

This section gives a brief overview of the existing business support technologies and explains their current limitations.

Workflow technology

Workflow technology has been widely recognised as one of the most influential business technologies and for many years as the leading process-oriented technology (WFMC, 2001). In essence, workflows are designed to specify, execute, manage, monitor and streamline business processes by allocating the right task to the right person at the right point of time along with the resources needed to perform the assigned task. Furthermore, this technology enables integration of different tools and technologies used to support the individual tasks. For this reason, it is also considered to be one of the leading process *integration* technologies (WFMC, 2001).

Workflow design and implementation involves two distinct phases: *modeling* (also called built-time) and *execution* (run-time) phases. Briefly, during workflow modeling, individual tasks, roles and tasks dependencies are identified and combined in a business process model according to the corresponding business rules. It is important to note that workflow models are created by workflow analysts (i.e. process experts), not end-users. Furthermore, these models are executable models. This means that once such a model is created by using a graphical modeling language, it is stored in the workflow repository and is ready to be used (without any other programming involved). This is one of the reasons why workflow technology has became so popular.

During run-time, a number of different process instances are created on the basis of the same model. For example, a model of the business process “Home Loan Application” is created during built-time and stored in workflow repository. Then during run time, for each customer applying for a home loan, a system will create a different process instance. Therefore, a single process model will have hundreds or even thousands of corresponding process instances, all running in parallel.

It is important to observe that workflow technology requires a process model to be fully predefined during built-time, before it could be used during run-time. This means that at any point during process execution it is possible to use this model to determine what should happen next. This property makes this technology highly suitable for repetitive, highly structured business processes. However, not all processes fit this description. Consequently, possible application of workflow technology to any other BP type is likely to be very problematic.

Even though task coordination is fully automated, workflows significantly differ from job-shop scheduling systems. During workflow execution, human agents still *participate* in decision making related to possible outcomes of individual tasks. For example, a home-loan officer will decide whether to approve a particular home loan application. However, all possible outcomes (decisions) are known in advance and pre-defined by the workflow model.

Only recently, the workflow community has started to concentrate more on flexible business processes where models (including coordination structures) cannot be fully specified in advance. Notable examples include adaptive workflows (Reichert et. al. 1998) and emergent workflows (Carlsen and Jorensen, 1998). However, flexibility of the existing adaptive and emergent workflows is still very limited.

Enterprise Resource Planning Systems (ERPs)

ERP systems emerged in mid 90s and for the first time, offered the integrated, off-the-shelf solutions to support BP integration in an organization. ERPs have an embedded workflow system to support coordination, but workflow models are not directly available to users as in workflow management systems. Rather, business processes are implemented via various modules that incorporate standardised, benchmark models of the core organisational business processes. Thus, in order to implement these processes, an organization has to select and configure the required modules.

Although there are examples of successful implementation of ERP systems, there are also numerous counter-examples that illustrate various problems with this technology and its adoption in an organisation. First of all, ERPs are based on an assumption that generic, benchmark examples of processes, embedded within these systems, would suit all businesses. All that ERP users need to do is to configure and adopt these pre-defined models. However, as (Miers, 2004) points out: "..developing a deep understanding of the business and its interactions in the value chain follows from contrasting alternative perspectives, rather than slavishly following one "true" method". When people develop new ways of doing things (within the normative and managerial boundaries of what is possible) their organisations learn too".

Furthermore, standardised processes cannot be used to create competitive advantage for a company, at least not directly. Indirectly, they could contribute through, for example, improved efficiency of standardized business processes. This point is also confirmed by (McGovern, 2005) "… the non-proprietor, idealized business processes tend to evolve towards a public definition and don't directly afford competitive differentiation" (pg. 6).

As ERPs incorporate generic process definitions, inevitably they are very costly to configure, maintain and upgrade. (Ghalimi, 2003) as quoted by (Miers, 2004, pg. 23) offers a very good insight into the relationship between the cost of ERP software upgrade and Return of Investment (ROI): "Around half of SAP's customer base bought its product between 1998 and 2000. It has been estimated that it will cost each of them around $5 million to upgrade to the latest version, reapplying their customizations – estimated ROI approaching zero!"

Another equally challenging problem with this technology is its implementation (i.e. technology roll out process). ERPs often require very radical changes and the "big bang" approach very similar to what was promoted by Business Process Reengineering. This way of implementing an IT solution always implies an increased risk. "We saw what happened to the out-of-control ERP implementation where the core business process was petrified inside the core application. The IT function needs the capability to roll out revolutionary change – but to do that incrementally, in evolutionary steps." (Miers, 2005).

Web Services

Web services are another important example of business process support technology. They are self-contained, modular, Internet based applications, offered by different providers that have standard interface to enable efficient integration and implementation of complex business applications. Recent reports by various leading industry analysts and practitioners claim that web-services are going to revolutionise the existing IT applications as they enable easy integration of different platforms, tools and resources (Zhang, L-J, 2004), (Yang, 2003). *Composite* web services enable flexible, on-demand integration of individual services offered by different providers to meet a specific business objective. This integration is made possible by the fact that web services are platform neutral so as long as they comply with the common standard they can be integrated into a more complex structure. From the business perspective, composite web services enable integration and implementation of complex, dynamic inter-organisational business processes. This appears to be very promising direction for SMEs in certain industries where they could offer their services as web-services.

The field of web-services is emerging and there are many research and developmental challenges yet to be solved, including the coordination problem also called web service orchestration (Zhang, 2004). For example, currently available standards for service composition are based on the coordination models inherited from their predecessors - workflows. Consequently, they are equally inflexible, because they require all process components to be specified and pre-ordered in advance.

In summary, although web-services are more flexible technology than workflows, their current coordination paradigm is not flexible, nor dynamic enough to suit the needs of business process integration. This is the case with dynamic markets where all business partners and the process models can easily change. The same point is illustrated by (Hanson at. all, 2001), as cited in (Peltz, 2003)) by an interesting analogy. They compare the currently available models for web service orchestration to a vending machine with a set number of buttons that can be only pressed in a predefined order. What is needed is more flexible model analogous to a telephone call that involves a series of exchanges between parties at each end in a more flexible dynamic fashion. Only in this way it will be possible to assemble dynamic business processes.

UNDERSTANDING THE KNOWLEDGE DIMENSION OF BUSINESS PROCESSES

In order to understand the knowledge dimension of business processes, it is necessary to start from the basic concepts of explicit and tacit knowledge. *Explicit knowledge* can be written down or drawn and described to other people. Consequently, it can be organised, distributed and managed by technology. Good examples of explicit knowledge are various organisational procedures (i.e. how things should be done). Procedures are important for two different reasons. In the case of standardised processes, procedures are important to guide people performing individual tasks. They are also very important from the normative (legal) perspective, to define responsibilities and protect the rights of all parties involved. So, if something goes wrong in the process, it is possible to trace back process execution, compare it with prescribed procedures and detect what went wrong and who was responsible. However, it is also important to acknowledge the negative aspect of pre-defined procedures. They are often associated with the concept of organisational control. Thus, procedures are often imposed upon workers "to ensure control and compliance with pre-ordained approach and technology" (McGovern, 2004, pg. 3). Thus, technology developers rely on standardization and predictability of organisational procedures to design technical solutions that will make processes as efficient as possible and minimize the resources needed.

On the other hand, *tacit knowledge* are things known by people but usually not documented anywhere such as the know-how, understanding mental models and insights of an individual or disciplines (Alavi and Leinder, 2001). Very often, tacit knowledge is very difficult to easily communicate. This is why it cannot be automated, organized and managed by technology as in the case of explicit knowledge. However, this knowledge can be externalized, to some degree, trough "working things out". Externalisation of tacit knowledge in an organization results in development of organisational *practices*. They evolve from the accumulated experience and reflection-in-action. They are created by empowered knowledge workers through their ability to make decisions. Here the main emphasis is on quality, on satisfying customer's needs, finding creative solutions (within the normative boundaries) and on creating better customer-relationships. To illustrate the main difference between procedures and practices, (Miers, 2004) uses a very interesting example: "Consider an international bank – everyone would agree that the teller should not get creative with bank drafts. Tellers have very strict procedures (and systems) for handling money directly. On the other hand, try applying a rigorous procedural approach to those developing new trading instruments or those managing the bank's largest corporate customers (or even its executives). They would rebel if everything they did had to fit within rigid computerized procedures developed beforehand" (Miers, 2004, pg. 4).

In reality, all business processes combine to some degree both procedures and practices. So before designing or applying any existing or new BP support technology, it is necessary to determine the nature and the extent of these two components. For example, if business process support is developed to support procedures and is then applied to a BP where a practice component is more prominent, obviously such technology would not be very useful. Unfortunately, this is a typical problem with workflow technology using predefined models designed to capture BP's procedural aspects. On the other hand, creativity, often associated with practices cannot be captured by pre-defined models. Consequently, BPs with the more prominent practice aspect will need a different type of technology support.

A proper understanding of a possible relation between procedure and practice component in a particular business process, will also determine possible end-user involvement in design and implementation of business process. Recall that in traditional information systems, roles are clearly defined. Business analysts will define a model of an information system, developers will implement it based on the given specification and end-users will then proceed to use the system. However, this traditional division of roles and responsibilities is very different when it comes to business process support. Obviously practice-oriented business process, involve tacit knowledge that is hard to extenalise. Consequently, they require end-user involvement in a different form than what is currently the case with procedure-oriented business processes.

Furthermore, it is important to point out that better understanding of the knowledge dimension of a particular business process (its procedures and practices) will also help an organization to determine a possible source of competitive differentiation. The current business practice shows that organisations create competitive differentiation through their core competencies. Consequently, there is a direct link between core competencies and organisational (experiential) knowledge. For example, Stein and Zwass (2000) argue that the core competencies of an organization are derived from *episodic* knowledge (contextually situated decisions and their outcomes) rather than *semantic knowledge that is widely available in an organization.*

In summary, it is important to point out that in order to fully understand how to support business processes to create possible competitive advantage through BPM it is necessary to consider the knowledge perspective (procedures and practices) of a particular BP. However, this is not the case with currently available BP support technologies that predominately focus on process structure to determine what kind of support is needed. The same point will be further elaborated in the next section.

USING THE KNOWLEDGE DIMENSION OF BUSINESS PROCESSES

This paper argues that it is possible to use the knowledge dimension of business processes to determine suitability of different BPM technologies, the nature of the required end-user involvement and most importantly, possible sources of competitive differentiation related to business process execution.

To demonstrate this argument, this section will analyse three representative types of business processes from the following perspectives:

- Knowledge: procedural and practice aspects
- Technology: possible support by existing and emergent BPM
- End-users: required involvement of end-users in business process design and implementation
- Competitive advantage: possible sources of competitive differentiation

Procedure-oriented business processes

- Knowledge

Processes that demonstrate mainly the procedural aspect can be found at the operational level. An example is a post-graduate student admission process. This particular process is normatively regulated by a number of rules designed to govern the process and ensure its legality. Its procedural aspect makes it highly structured and repetitive because the same set of procedures is applied over and over again. Therefore, the predefined structure of this type of business process is, in fact, the main consequence of its procedural nature, not its cause. This is one of the reasons why it is not appropriate to consider process structure to determine what kind of business process supported is needed. Another reason will be identified later in the paper.

Due to their procedural nature, these processes involve predominantly explicit knowledge. Consequently, this explicit knowledge can be externalised to identify, for example, different tasks, their properties (such as roles, resources, temporal attributes) and coordination requirements (i.e. if individual tasks should be sequential, parallel and alternative). There are many available process models that can be used to represent this externalised knowledge (e.g. workflow models). Furthermore, procedures could also define how individual tasks should be performed. If this is the case, these tasks are considered to be routine.

However, in reality, these processes are not completely guided by rigid procedures. There are also aspects of these processes that involve tacit knowledge. For example, even in a well-structured process, people make decisions at various points during process execution. In this type of business processes all decision outcomes are known in advance (i.e. a post-graduate student will be admitted or not). If a complete decision making process could be easily described by enlisting all possible cases and precise instructions how to make a required decision, then this process can be fully automated by technology. But, here, this is not the case. The decision making process always require professional expertise (tacit knowledge). For example, a potential supervisor will use his/her professional expertise to assess the quality of a particular PhD proposal.

Another aspect of tacit knowledge involves possible exceptions from pre-defined models. “Almost all processes have non-deterministic elements such as those activities improvised by experienced professionals in resolving exception conditions” (Miers, 2004; pg. 6). The number and nature of possible exceptions will determine to what degree a particular process involves the exception handling practices.

- Technology

Once the explicit knowledge is externalised and their models are created, these BPs can be supported by technology such as workflows, ERPs or web-services. In fact, the current BPM technologies are suitable to support, to some degree, the procedural aspect of a given BP.

However, the practice-oriented aspect of this type of business processes will appear in the form of exceptions from pre-defined process model. Obviously, they will require the efficient exception handling support that is not always provided by existing BPM technology.

For example, effective exception handling is still an open problem in the area of workflows. Some workflow solutions require workflow analysts to enlist all possible exceptions in advance. However, in dynamic environments this method is quite limited and inflexible. Other workflow systems provide very limited support to guide users when dealing with an exception. However, they are usually structural exceptions i.e. deviations from pre-defined structure.

It is important to point out that degree and nature of exceptions in a particular BP are very good indicators of its practice aspect. In fact, the more this aspect is prominent, current BPM solutions become less and less suitable.

- End-user

Process support technologies suitable for this type of processes (such as workflow) change the roles involved in design and implementation of business processes. These technologies provide executable business process models. Once a BP model is designed by a workflow analyst, it can be directly executed by workflow engine. Therefore, process modeling is done by workflow or process analysts, its execution is delegated to workflow engine that will schedule different tasks to different roles as specified by the model.

- Competitive advantage

As already pointed out, externalisation of tacit knowledge can be used to create possible competitive advantage related to business process execution. In the case of business processes with the predominately procedural aspect, models standardised business processes cannot be used as a possible source of competitive advantage. However, technology can contribute to make processes more efficient. Consequently a possible source of competitive advantage could be better customer-support service provided by more efficient processes. However, a degree of caution is required here because efficiency cannot be always equated to quality, especially if these processes involve a degree of tacit knowledge.

Another very important source of competitive advantage for this type of processes are in fact exceptions and exception handling processes. The same point is also illustrated by (Miers, 2004): “The value of exception processing is usually underestimated… the exception handling associated with non-proprietary business processes is often extremely proprietary and determines the competitive success of the business.” In some cases, exception-handling practices can be documented and turned into organisational procedures. They, in turn, can be used to improve business process models in the future.

Combined business processes

- Knowledge

These processes involve both practices and procedures. They are also called the *Case Handling* processes where individual instances are customer-oriented and always treated as unique cases. There are numerous professions that operate on cases including medicine, legal services, accounting. Possible examples include a medical doctor making decisions about possible treatment for each patient, or an insurance officer handling different insurance claims. Although there are commonalities, each medical procedure or each insurance claim involve some specific aspects that require professional expertise. As (McGovern, 2005) pointed out “humans interpret rules and do what is best”. Practices get developed through the accumulated experience in dealing with these unique aspects. The rules that need to be interpreted are in fact explicit knowledge (i.e. the procedural aspect of case handling procedures). The procedural aspect is there to define the rights and responsibilities of all people and to ensure that work is carried out in consistent and compliant manner. Examples include medical guidelines that all medical professionals are expected to comply with.

Furthermore, based on its procedural aspect, it is possible to predefine a process model to some extent and only for the most common cases. However, a person with professional expertise is in charge of process design, selection of individual tasks and their combination into a process for each individual customer. They are also responsible for monitoring of each individual case and possible modifications, if required.

- Technology

Obviously, process support provided by traditional workflows and ERPs is not suitable here. More flexibility is required to ensure easy composition and execution of personalised process instances. Current solutions in the area of adaptive workflows (Lent and Reichart, 2005), emergent workflows (Carlsen and Jorgensen, 1998) and component-based workflows (Marjanovic, 2005) could be used to support, to some extent, case-handling BP. However, it is important to observe that any future solution is this area needs to move away from the business process structure and enable on demand assembly of process-components. Although web-service technology enables integration of individual services into complex structures that can be used to implement a component-based BP, this technology does not still meet the requirements of case-handling BP support. One of the main problems is inflexibility of the web-service coordination mechanism inherited from traditional workflows.

- End users

Compared to the other identified categories of BP, case-handling processes have very specific requirements for end-user support. In fact, here we have one category (type) of end-users (professional with domain expertise) designing models for other category of end-users (clients). This creates new research challenges related to end-user modeling and verification of business processes, monitoring and run-time modification.

- Competitive advantage

As already pointed out, every case-handling process model is, to some extent, unique. This uniqueness creates opportunities for competitive differentiation. They include the ability to select the best individual process components (tasks, services etc.) to suit the needs of a particular customer, personalisation of process models and superior customer service. However, the main source of competitive advantage is the expertise of professionals involved in design and implementation of personalised BPs. The role of technology is to support this processes.

Practice-oriented business processes

- Knowledge

These are high-value business processes that predominately involve organisational practices. Implementation of these processes is usually associated with organisational agility and their models are always proprietary solutions. Although these processes involve a procedural aspect (to ensure compliance with the normative environment), this aspect is not prominent.

It is interesting to observe that, in some cases, this procedural aspect can be even used to determine the structure of these processes. For example, in the BP “Marketing a new product” the same set of steps is followed each time a new product is launched (benchmarking competitors, deciding pricing strategy, planning promotions, etc.). “But in essence knowledge and intuition will drive these processes to success”. (Cervera, L.R, 2004). Thus, the value of this process is not created by, or associated with its structure.

- Technology

It is obvious that traditional process support technologies cannot be used to support this type of processes. Even though, in some cases, their structure can be to some extent predefined, the application of workflow or web service technology would be very inappropriate. Here, people determine when they are ready to move to the next task and again that’s experiential knowledge that cannot be easily delegated to workflow engine. They are also very likely to move forward and backwards between tasks depending on their progress, or even invent new tasks along the way. Even, the business process management systems designed to support emergent workflows (such as for example (Carsen and Jorgensen, 1998)) cannot be used in this case as they concentrate on business process structure.

Obviously, very different type of business process support is required. Participants need help for situated decision-making that will enable them to progress though the process. Obviously, coordination becomes human-driven rather than technology-driven. In essence, these processes involve creativity and it is a well-known fact that creativity cannot be ever automated. Possible support involves on-demand integration of different tools (e.g. for collaborative work, data analysis and visualization) but also support for mining and analysis of instances of previous processes. Process-mining support makes sense only within the boundaries of one organization, as these processes are proprietary solutions never to be reviled to competitors.

- End-users

Due to the nature of these processes, there is not a process model that can be pre-defined by a process specialist. Thus, end-users are process-designers and process-users. Obviously this dual role delegated to end-users require very user-friendly BPM system possibly in the form of plug and play components that could be easily assembled as required by a particular solution.

- Competitive advantage

In the case of this type of BP, the evolving process models as well as the whole experience of creation and execution of these processes represents the source of possible competitive differentiation. This certainly includes knowledge and abilities of all participants in the process to complete the individual tasks by using their tacit knowledge.

CONCLUSION

Current research and practice in the field of BPM (Business Process Management) confirms a wide gap between the current understanding of this term by business managers on one side, and technology developers on the other side. The main objective of this paper is to contribute to better understanding of this gap. The paper argues that the first step towards understanding of this gap is to analyse the knowledge perspective of business processes rather than their structure (as it is currently done by technology developers). To illustrate the proposed approach, the paper uses this knowledge dimension to analyse three representative types of business processes. For each type, it demonstrates how its procedural and practice oriented aspects can be used to determine possible technology support, requirements for end-user involvement and most importantly, possible sources of competitive differentiation. Current and future work involve further efforts to bridge the identified gap though development of theoretical frameworks as well as technical BPM solutions.

REFERENCES

Alavi, M. and Leidner, D.E. (2001) “Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues”, *MIS Quarterly*, Vol. 25, No. 1, March, pp.107-136.

Boland, R. and Tenkasi, R. (1995), “Perspective Making and Perspective Taking in Communities of Knowing”, *Organisational Science* (6:4), pp.350-372.

Carlsen, S. and Jorgensen, H. (1998), “Emergent Workflows: The AIS Workware Demonstrator”, *Proc. of the CSCW-98 Workshop: Towards Adaptive Workflow Systems*, Seatle.

Crevera L.R. (2004), “Knowledge Process Management”, available online from www.nevant.com.

Keen, P.G. and Scott-Morton, M. (1978) *Decision Support Systems: An Organisational Perspective*, Addison-Wesley, Reading, MA.

Lenz, R. and Reichert, M. (2005), “IT Support for Healthcare Processes”, Proc. of the 3rd International Conference on Business Process Management, BPM 2005, Nancy, France, September 2005.

Markus, M. L, Majchrzak, A. and Gasser, L. (2002), “A Design Theory for Systems that support Emergent Knowledge Processes”, *MIS Quarterly Vol. 26*, N.3, pp.179-212, September.

Marjanovic, O. (2005), “Process-oriented CRM enabled by component-based workflow technology” Proc. of the 18th Bled eCommerce Conference eIntegration in Action, Bled, Slovenia, June 6-8 2005.

McGoven, D. (2004), “An Introduction to BPM and BPMS”, Business Integration Journal, April 2004 issue (available from: www.bijonline.com).

Miers, D. (2004), “The Split Personality of BPM”, BPTrends, February, 2004 (available from: www.bptrends.com).

Mohan, C. (1997), “Tutorial: State of the Art in Workflow Management Systems Research and Products”, *Proc. of the 5th International Conference on Database Systems for Advanced Applications (DASFAA ’97)*, Melbourne Australia.

Moore, C. (2000), “Process Knowledge”, *Workflow Management Coalition* (available from : http://www.wfmc.org)

Pava, C (1983), *Managing New Office Technology: An Organisational Strategy*, Free Press, New York.

Peltz, C. (2003), “Web Service Orchestration – a Review of Emerging Technologies, Tools and Standards, Hewlett-Packard Company.

Reichert, M., Hensinger, C. and Dadam, P. (1998): Supporting Adaptive Workflows in Advanced Application Environments, *Proc. EDBT-Workshop on Workflow Management Systems*, Valencia, March 1998, pp.100-109.

Stein, E.W. and Zwass, V. (1995), “Actualising Organizational Memory with Information Systems,” *Information Systems Research*, 6 (2), pp. 85-117.

Van der Aalst, W. M. P., ter Hofstede, H.M. and Weske, M. (2003) “ Business Process management: A Survey”, Proceedings of the International Business Process Management Conference BPM 2003, Lecture Notes in Computer Science LNCS, 2678, pp.1-12, Springer.

Workflow Management Coalition (2001), *WFMC Interface: Process Definition Interchange Process Model.* (available from: http://www.wfmc.org)

Yang, J. (2003) “Web Service Componentization”*, Communications of the ACM,* October, Vol. 46, No. 10.

Zhang, L-J. (2004) “Challenges and Opportunities for Web Service Research”, *International Journal of Web Service Research*, Vol 1, No.1, 41-57, Jan-Mar, 2004.

Automatic construction of a concept hierarchy to assist Web document classification

Woo Chul Cho
Macquarie University
Sydney, Australia
Email: wccho@ics.mq.edu.au

Dr Debbie Richards
Macquarie University
Sydney, Australia
Email: richards@ics.mq.edu.au

Abstract

In this paper, we present a new technique which is the Admixture MCRDR-FCA (AMF) algorithm for Web document classification. The technique offers a practical approach to new Web document classification by combining and extending a number of the current techniques. The AMF algorithm has a number of noteworthy features: firstly, it provides a structured conceptual correlation between keywords and secondly it is optimised. Finally, the algorithm creates refined multiple new rules in order to achieve higher accuracy in the conclusions relating to document classification. This is achieved by clarifying the relationships between one concept and another concept before going on to provide a final classification to some category. To evaluate the AMF algorithm, we have developed a demonstration system that permits easy comparison with a number of other classification techniques.

Keywords:

Ontologies, Knowledge Acquisition, Web document classification, Multiple Classification Ripple Down Rules, Formal Concept Analysis

1. INTRODUCTION

Nowadays, almost all information is Web-accessible due to the remarkable evolution of information technology. The main purpose of the Web is to provide information. The Internet is the medium that provides at minimal cost a variety of information and does so more quickly than other mediums. Keeping pace with the growth of the computer industry is the growth in information posted on the Internet by both organizations and individual users. The rate of information creation on the Internet is around two billion items each day (Choi *et al*. 2004). Currently, this information exists on the Internet in the form of Web-based documents using the HTML format which are classified by document classification engines.

The purpose of this study is to develop an improved technique and a supporting demonstration system to provide more accurate document classification. We have developed the Web document classification system using the AMF (Admixture of MCRDR-FCA) algorithm which combines techniques from MCRDR (Kang 1995) and FCA (Wille 1982). In short, our system first deletes stop-words from Web documents. The remaining terms are stemmed using Porter's (1980) Stemming algorithm. Next, we extract word features from the Web documents by using Information Gain (Shannon 1948, Yang, Y. and Pedersen 1997). We use the MCRDR algorithm and the FCA algorithm individually to create an initial temporary knowledge base which is used to filter and generate a new knowledge base allowing final classification. This combining process is known as the AMF algorithm. To evaluate the AMF algorithm, we have developed a demonstration system that permits easy comparison with a number of well-known document classification techniques: naïve Bayesian. Naïve Bayesian with Threshold, Term Frequency-Inverse Document Frequency (TFIDF) and two Support Vector Machine techniques. The Naïve Bayesian classification algorithm (Mitchell 1997) uses probability based on Bayes theorem. The Naïve Bayesian with Threshold has been developed to overcome the lower classification precision of documents with low conditional probability that results due to the use of a fixed threshold in the naïve Bayesian algorithm. We compare our results with term frequency-inverse document frequency (TFIDF) (Ricardo and Berthier 1999), which works by expressing a weight vector based on word frequency of the given document 'd ', due to its traditional use in information retrieval. Finally, Support Vector Machines have been found to produce good results, so we compare our approach with the two main versions Support Vector Machines Pairwise (SVMPR) (Liau and Noble 2003) and Support Vector Machines Winner-Take-All (SVMWTA) (Aiolli and Sperduti 2005).

The paper is organised as follows. In Section 2, we introduce the concepts behind MCRDR and FCA and then describe the AMF algorithm for Web document classification. The subsequent section introduces the demonstration system, describes the experiments and discusses the corresponding results. The last section concludes this paper and suggests some paths for further work.

2. DOCUMENT CLASSIFICATION ALGORITHMS

In this section we briefly introduce the key concepts underlying the AMF (Admixture of MCRDR and FCA) algorithm. The final subsection describes how we have combined the latter two algorithms.

2.1. MCRDR (Multiple Classification Ripple Down Rules)

Kang (1995) developed the MCRDR algorithm. MCRDR overcomes a major limitation in Ripple Down Rules (RDR), which only permitted single classification of a set of data. That is MCRDR allows multiple independent classifications. An MCRDR knowledge base is represented by an n-ary tree (Kang 1995). The tree consists of a set of production rules in the form "*If Condition Then Conclusion*".

2.1.1. Creation of new rule

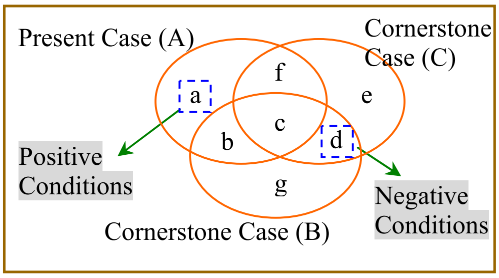


Figure1. *Difference List {a, not d} are found distinguish the Present Case (A) from two Cornerstones Cases (B) and (C)*

Formula (1):

Positive Condition

Present Case (A)-(Cornerstone Case (B))∪(Cornerstone Case C))

Negative Condition

(Cornerstone Case (B)∩(Cornerstone Case-Present Case (A)

We consider a new *case (present case) A* and two *cornerstone cases B* and *cornerstone cases C*. The cornerstone case is the case that prompted the rule being modified to be originally added. That is, the present case has been run and a rule has fired but the domain expert does not agree with the conclusion given. There must be some features in the present case which are different to the cornerstone case which merit a different conclusion. The present case will become the cornerstone case for the new (exception) rule. To generate conditions for the new rule, the system has to look up the cornerstone cases in the parent rule. When a case is misclassified, the rule giving the wrong conclusion must be modified. The system will add an exception rule at this location and use the cornerstone cases in the parent rule to determine what is different between the previously seen cases and the present case. These differences will form the rule condition and may include positive and negative conditions (see Formula (1)).

2.1.2 Inference

The inference process of MCRDR allows for multiple independent conclusions with the validation and verification of multiple paths (Kang 1995). This can be achieved by validating the children of all rules which evaluate to true. An example of the MCRDR inference process is illustrated in Figure 2. In this example, a case has attributes {a, c, d, e, f, h, k} and three classifications (conclusion 3, 5 and 6) are produced by the inference. Rule 1 does not fire. Rule 2 is validated as true as both 'a' and 'c' are found in our case. Now we should consider the children (rules 6, 7, and 10) of rule 2. From comparison of the conditions in children rules with our case attributes, only rule 6 is evaluated as true. Hence, rule 6 would fire to get a conclusion 6 which is our case classification. This process is applied to the complete MCRDR rule structure in Figure 2. As a result, rule 3 and 5 can also fire, so that conclusion 3 and conclusion 5 are also our case classifications.

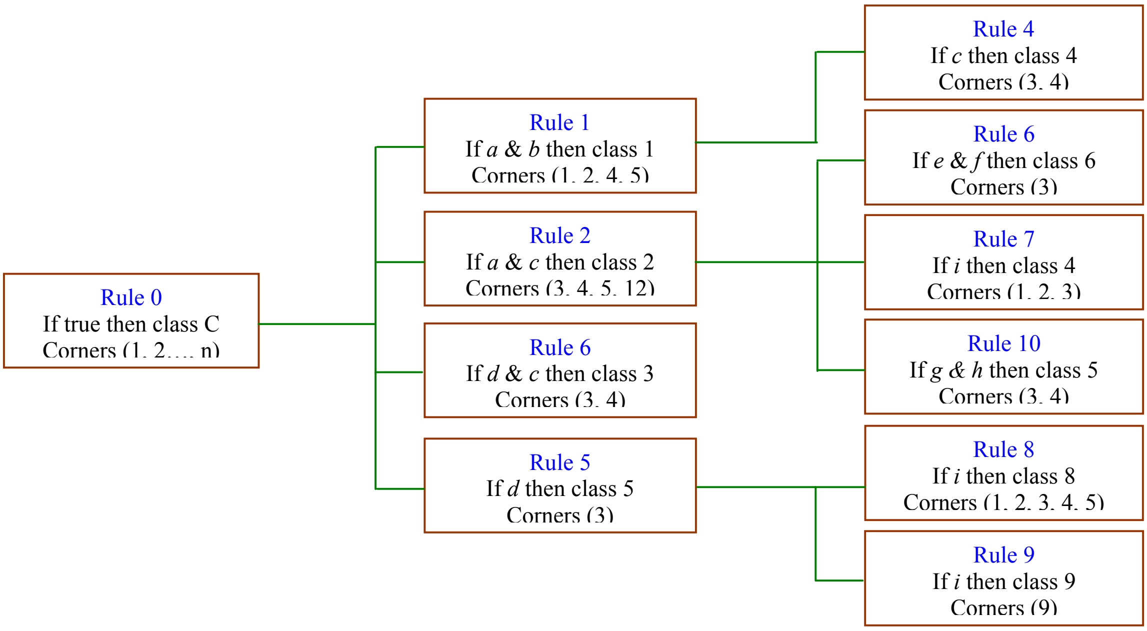


Figure 2. *Knowledge Base and Inference in MCRDR, Attributes are {a, c, d, e, f, h, k}*

2.2. FCA (Formal Concept Analysis)

Wille (1982) developed Formal Concept Analysis (FCA). FCA is based on the understanding of a concept as an entity of thought, consisting of an extension an intension, and representing this information using lattice theory (Birkhoff 1993, Ganter and Wille 1998). FCA provides an alternative graphical representation of tabular data that has been found to be instinctive to navigate and use (Ganter and Wille 1999).

2.2.1. Formal Context

The fundamental conceptual structure of FCA is the Formal Context (FCT). Definition of FCT follows:
Definition 1. A formal context is comprised from a set of objects and their attributes. A formal context constitutes a triple (G, M, I). 'G' is the set of objects, 'M' is the set of attributes and 'I' is a binary relation defined between 'G' and 'M'. Therefore $I \subseteq G \times M$. We can define a formal context (FCT) as: FCT := (G, M, I). If an object '*g'* has an attributes '*m'* then $g \in G$ is related I to '*m'* which is indicated by the relationship $(g,m) \in I$ or gIm. There are mean that '*g*' includes '*m*'. In FCT := (G, M, I), A is the set of objects ($A \subseteq G$) and B is set of attributes ($B \subseteq M$). About the set '*A*' and the set '*B*' satisfied '$A \subseteq G$' and '$B \subseteq M$', its two operators are (see Formula (2)):

Formula (2):

$$A \subseteq G: \quad A' = \{m \in M \mid (g,m) \in I \quad \text{for all} \quad g \in A\}$$
$$B \subseteq M: \quad B' = \{g \in G \mid (g,m) \in I \quad \text{for all} \quad m \in B\}$$

Definition 2. These two operators are used to formalize the notion of a formal context. In the above operators, A' and B' are set of objects and attributes respectively. A' is included by all objects in A. B' is included by all attributes in B. By finding all intersections of the primitive concepts given in the formal context we can generate all concepts and using the subsumption operator $\subseteq$ we order all concepts to form a concept lattice of all concepts.

2.2.2. Formal Concept

Commonly, Formal Concepts (FC) represents a relationship between objects and attributes. The FC is defined as a pair (A, B). The pair (A, B) satisfies $A' = B$ and $B' = A$ when '$A \subseteq G$' and '$B \subseteq M$' in formal context (G, M, I). In FC, 'A' is the set of objects and called *extent*, and 'B' is the set of attributes and called *intent*.

Formula (3):

$$\text{(a): } A' = \bigcap_{g \in X} \{g\}' \quad , \quad \text{(b): } B' = \bigcap_{m \in Y} \{m\}'$$

We can construct all formal concepts of a formal context by the formulas (3(a)) and (3(b)). We can obtain all extents A' by determining all row-intents $\{g\}'$ with $g \in G$ (see Formula (3(a))). Also, all intents B' can be obtained by determining all column-extents $\{m\}'$ with $m \in M$ (see Formula (3(b))). Then, all their intersections are found. For example, *table 1* presents a cross-table of the formal context to which formula (2) has been applied. The *table 2* shows how formal concepts can be derived from the formal context (FCT := (G, M, I)) in *table 1* using formula (3).

Table 1. *A cross-table shoes a formal context for a part of "Animal Kingdom". The column A1-A6 represents has-four legs, has-wings, has-long nose, has-hair, has-scales and has-webfoot. An 'X' indicates that the object has the corresponding attribute.*

	A1	A2	A3	A4	A5	A6
Anteater	X		X	X		
Elephant	X		X			
Beaver	X			X		X
Duck		X				X
Flying Fish		X			X	

Table 2. *This table shows process of finding formal concepts for example of the table 1. The table 2(a) shows results of process 1, 2 and 3. The table 2(b) presents result of process 4.*

Step	Intend	Extend
1		{1, 2, 3, 4, 5}
2	A1	{1, 2, 3}
3	A2	{4,5} { }
4	A3	{1, 2}
5	A4	{1, 3} {1}
6	A5	{5}
7	A6	{3, 4} {3} {4}

(a)

Step	Intend	Extend
1	{ }	{1, 2, 3, 4, 5}
2	{A1}	{1, 2, 3}
3	{A2} {A1, A2, A3, A4, A5, A6}	{4,5} { }
4	{A1, A3}	{1, 2}
5	{A1, A4} {A1, A3, A4}	{1, 3} {1}
6	{A2, A5}	{5}
7	{A6} {A1, A4, A6} {A2, A6}	{3, 4} {3} {4}

(b)

2.3. AMF (Admixture of MCRDR and FCA)

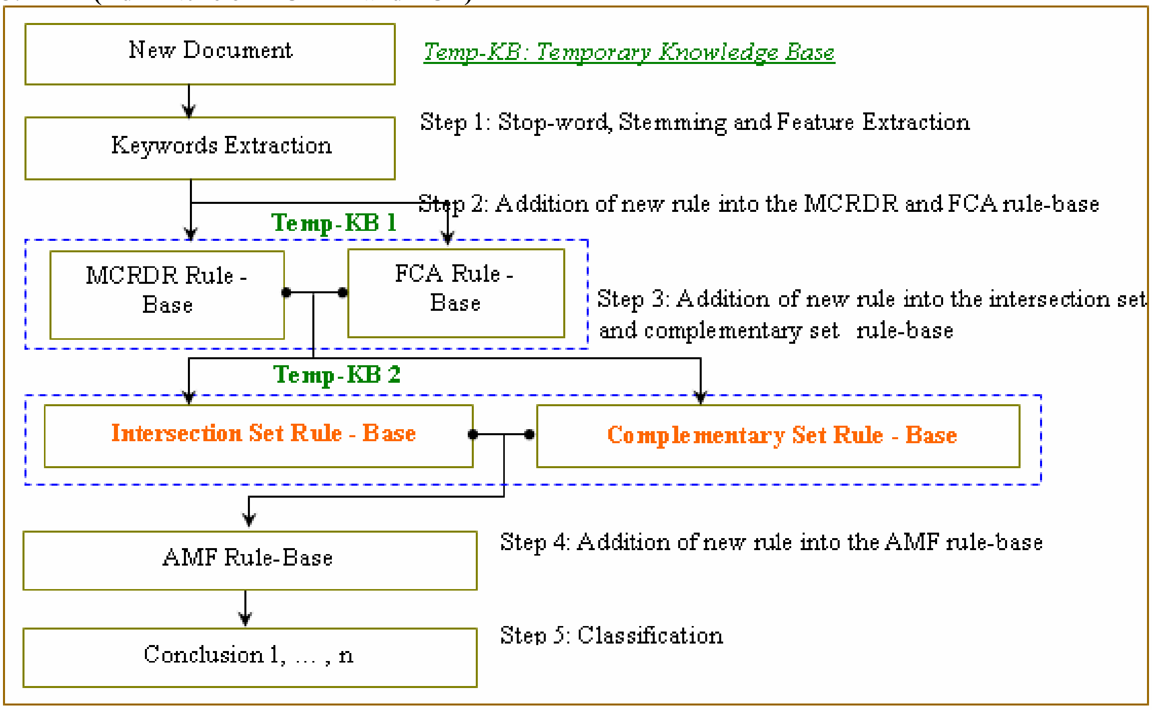


Figure 3. *Diagram of AMF Algorithm for document classification*

The AMF algorithm combines the merits of both the MCRDR and FCA algorithms, as shown in Figure 3. The AMF algorithm obtains new rules for accurate classification based on the feature keywords in the document. The feature keywords are independent of one another in the AMF algorithm. A prominent characteristic of the AMF algorithm is the ability to represent the semantic relationships between feature keywords and to allow conclusion (a Class) to have several rules.

2.3.1. Procedure for Creation New Rule

A process of document classification using the AMF algorithm, shown in Figure 3, is described as follows:

Step 1: As part of our data preprocessing and in order to increase the accuracy of document classification we perform typical stop-word removal, stemming and feature extraction using Information Gain for each document to produce a set of refined keywords.

Step 2: A Temporary Knowledge Base 1 (Temp-KB1) is created which comprises both an MCRDR Rule-Base and an FCA Rule-Base. This is achieved using the refined keywords from step 1 and applying separately the MCRDR algorithm (formula 1) and FCA algorithm (formulas 2 & 3). The algorithms can be applied automatically or manually. For instance in the case of MCRDR the system can select some feature that distinguishes the current case from the cornerstone cases or the rules can be specified by a domain expert, as is more common. The formal context from which the FCA rules/concepts are generated can be specified by a human or automatically populated from cases.

Step 3: The Temporary Knowledge Base 2 (Temp-KB2) contains two Rule-Bases which are the Intersection Set Rule-Base and the Complementary Set Rule-Base. Its two operations are:

Formula (4):

$$\text{Intersection Set} = \{\, MCRDR \cap FCA \,\}$$

$$\text{Complementary Set} = \{\, (MCRDR \cap FCA)^C \,\}$$

Using these operators we find the intersection and complementary sets between the rules in the MCRDR Rule-Base and the rules in the FCA Rule-Base in Temp-KB1. Then, the results (new rules) are added to the Intersection Set Rule-Base and Complementary Set Rule-Base in Temp-KB2 respectively.

Step 4: The AMF Rule-Base inherits all intersection rules from the Intersection Set Rule-Base in Temp-KB2. Next, the sets of rules in the Intersection and Complementary Set Rule-Bases are combined using the MCRDR algorithm, therefore creating new rules. Then, the new rules are validated using the FCA algorithm and added to the AMF Rule-Base. The new rules form a conceptual hierarchy of formal abstractions. If any rules could not be validated by the FCA algorithm, a 'stopping-rule' is added by applying the MCRDR algorithm. This means that the rule will not be used anymore.

Step 5: The document is classified into some class according to the inference process described above (see Section 2.4.2).

For example, Figure 4 shows the progress of Web document classification. We assume the keywords (called 'a', 'b', 'c', 'd', 'e') have been extracted from a Web document through applying Step 1. The keywords {a, b, c, d, e} form the new rules {(a&b), (a&c), (d&e)} and {(a&c), (c&d), (b&e)} by applying the MCRDR algorithm and FCA algorithm, respectively. The new rules are saved to the MCRDR Rule-Base and the FCA Rule-Base in the Temp-KB1 separately (Step 2). Next, we find the set of intersecting rule/s {(a&c)} and complementary rule/s {(a&b), (d&e), (c&d), (b&e)} from MCRDR Rule-Base and FCA Rule-Base in the Temp-KB1. Thereafter, the intersection set rule {(a&c)} is added to the Intersection Set Rule-Base and the complementary set rules {(a&b), (d&e), (c&d), (b&e)} are added to the Complementary Set Rule Base in Temp-KB2 (Step 3). The intersection {(a&c)} in the Intersection Set Rule-Base is added to the AMF Rule-Base through inheritance. The Intersection (a&c) is combined with the complementary set rules {(a&b), (d&e), (c&d), (b&e)} in the Complementary Set Rule-Base to create new rules {(a&b&c), (a&b&e), (a&c&d), (a&d&e)} using MCRDR algorithm. Next, the combined new rules {(a&b&c), (a&b&e), (a&c&d), (a&d&e)} are verified by FCA algorithm. Then, the rules separate veridical rule (correct rule) {(a&b&c), (a&c&d)}, and stop-rule {(a&b&e), (a&b&c)} and the rules {(a&b&c), (a&c&d)} are moved to the AMF Rule-Base (Step 4).

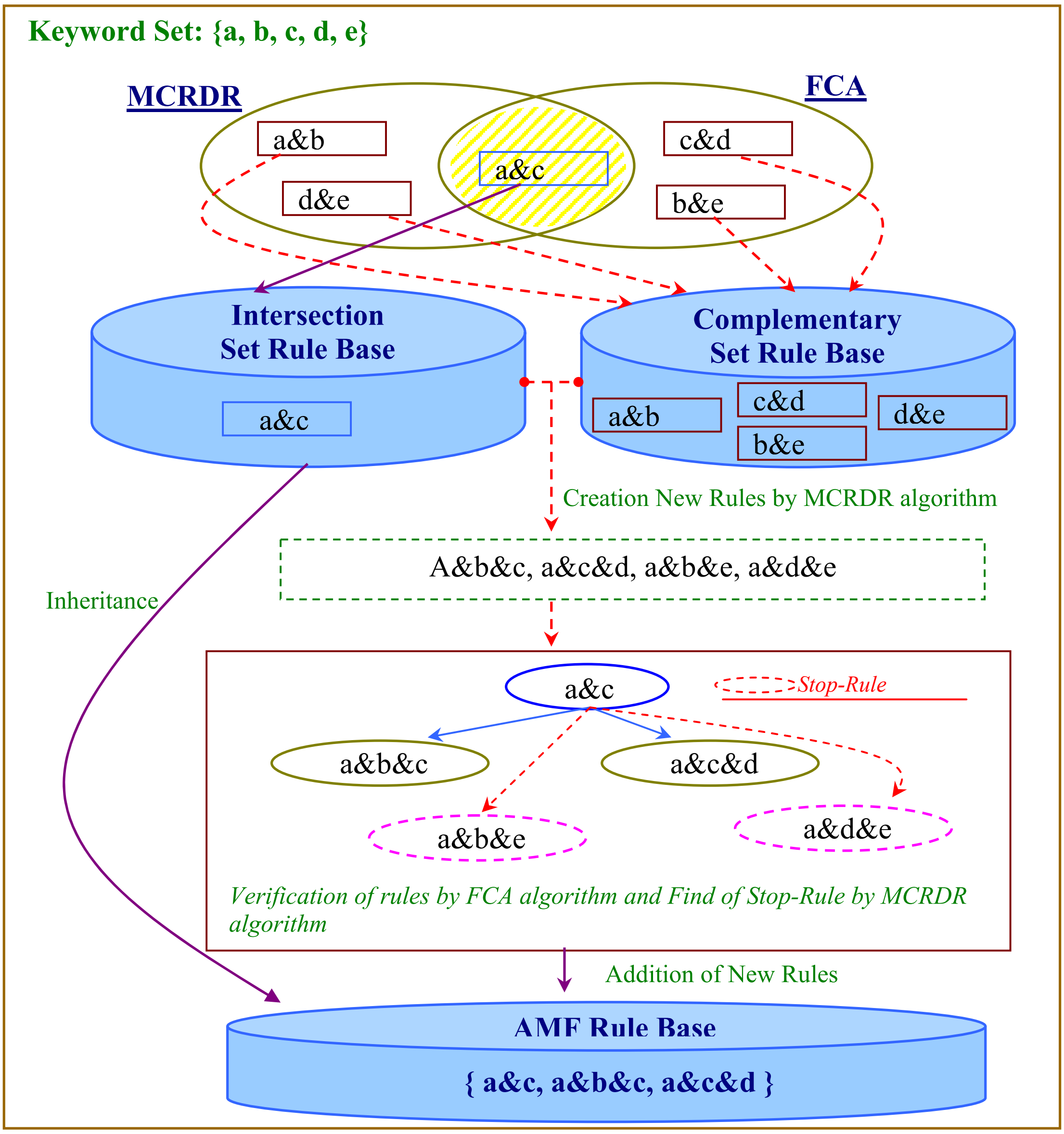


Figure 4. *Exemplifying the process of creating new rules using the AMF algorithm*

3. EXPERIMENT

3.1. Aims and Data Set

We developed the AMF algorithm for Web document classification. In order to validate the performance of the AMF technique, we compared its precision with a number of classification techniques. For the experiments, we used a data set of Web documents provided by Sinka and Corne (2002) which has been used for validating clustering of Web document algorithms. Sinka and Corne (2002) provide four themes, namely, “Banking & Finance”, “Programming & Language”, “Science” and “Sport”. The “Banking & Finance” and “Programming & Language” themes each contain three sub-categories, and “Science” and “Sport” themes each contain two sub-categories (see Table 3).

We extracted data sets containing 100 experimental data elements in each of the ten sub-classes, comprising a total of 1000 elements of experimental data. Each set of 100 was separated into a learning data set (60) and an evaluation data set

(40). We conducted our experiment three times. Initially we trained with 20 elements, each time increasing the input learning data by 20 data until the third run contained 60 elements. The same evaluation data set was used each time (see Table 4).

Table 3. Four themes with ten sub-categories

Theme	Sub-Category
Banking & Finance	*Commercial Banks*
	Building Societies
	Insurance Agencies
Programming Language	*Java*
	C / C++
	Visual Basic
Science	*Astronomy*
	Biology
Sport	*Soccer*

Table 4. Data Set for Experiment

Each of a Sub-Class		
Experiment Frequency	**Learning Data Set**	**Evaluation Data Set**
1 time	20	40
2 time	+ 20	= 40
3 time	+ 20	= 40
Total	**60**	**40**
All Sub-Classes		
Amount	**600**	**400**

3.2 Development System

The screen dump in Figure 5 displays the key elements of our system, which has been developed to evaluate the performance of the implemented algorithms. The screen consists of three parts; the top panel allows us to choose which classification rule to apply to the set of Web documents, the second panel allows selection of the 10 classes (Commercial Banks, Java, Biology, Soccer and so on) of the data and whether training (learning) or testing (evaluation) data is to be used. The third section on the screen (large lower panel) is used to display the contents of the data for the purposes of evaluating and confirming that the data has been classified into the correct class.

In order to improve the performance of the experiment, we performed pre-processing on the data. Data preparation is a key step in the data mining process. As described previously, for us this step involved deletion of stop-words, stemming and feature extraction (see Figure 6) in this system.
The meaning of 'Stop-words' refers to common words like 'a', 'the', 'an', 'to', which have high frequency but no value as an index word. These words show high frequencies in all documents. If we can remove these words at the start of indexation, we can obtain higher speeds of calculation and fewer words needing to be indexed. The common method to remove these 'Stop-words' is to make a 'Stop-words' dictionary in the beginning of indexation and to get rid of those words. This system follows that technique.

For stemming, we used the Porter Stemming algorithm (Porter 1980). Porter's Stemming algorithm is currently the most popular technique for this purpose. The Porter's stemmer does not remove prefixes. The Porter's stemmer has various rules for stemming. Each rule is defined by four sub-items which are Rule Number, Suffix, Replacement String and Stem State which allow the suffix to be replaced by the specified replacement string.

For feature extraction, our system uses the well-known Information Gain approach (Shannon 1948, Yang and Pedersen 1997) that selects words that have a large entropy difference as word features based on information theory. When the complete set of vocabulary (V) consists of rules (formula 5(a)) and n words, formula 5(b) shows the calculation of the information gain for each word ' w_k '. Those words which have the largest information gain are included in the optimized set of word features (K) as in formula 5(c).

<u>Formula (5):</u>

(a) $V = \{w_1, w_2, w_3, w_4, w_5, \ldots\ldots, w_n\}$

(b) $$InforGain(w_k) = P(w_k)\sum_i P(c_i \mid w_k)\log\frac{P(c_i \mid w_k)}{P(c_i)}$$

$$+P(\overline{w_k})\sum_i P(c_i \mid w_k)\log\frac{P(c_i \mid \overline{w_k})}{P(c_i)}$$

(c) $K = \{w_1, w_2, w_3, w_4, w_5, \ldots\ldots, w_L\}, K \subset V$

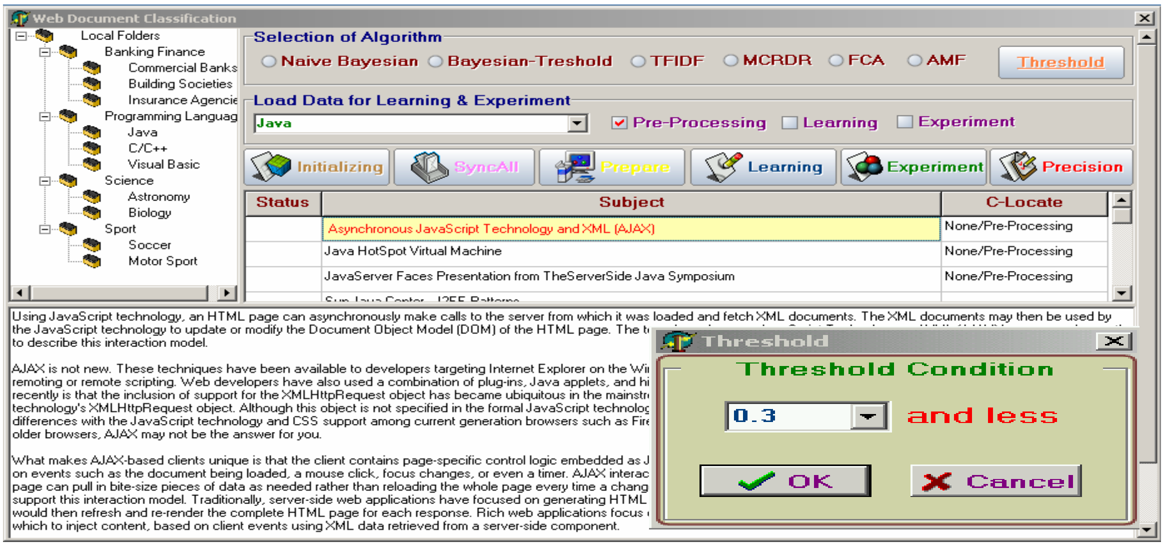


Figure 5. *Web document Classification system for experiment and control of Threshold value*

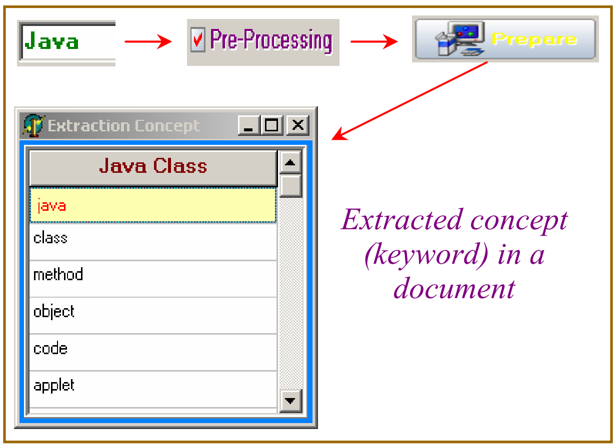


Figure 6. *Example, progress of extracted concepts (keywords) about a document in the system*

3.3. Results

Figure 7 displays the variation across algorithms and datasets for the experimental data. In Figures 7 and 8 and Table 5, columns A1-A8 represent the algorithms: *naïve Bayesian (A1), naïve Bayesian with Threshold (A2), TFIDF (A3), MCRDR (A4), FCA (A5), AMF (A6), SVMPR (A7) and SVMWTA (A8). The* rows C1-C10 represent the data sub-classes: *Commercial Banks (C1), Building Societies (C2), Insurance Agencies (C3), Java (C4), C/C++ (C5), Visual Basic (C6), Astronomy (C7), Biology (C8), Soccer (C9)* and *Motor Sport (C10).* We can see that AMF performs significantly better than all the other algorithms on the Visual Basic dataset, but is outperformed by the SVMWTA on some datasets. Figure 8 aggregates the results to reveal that AMF performs the best overall, even though it is only marginally better than SVMWTA.

Looking at the actual numbers, Table 5 show high overall precision 81% - 89% for all algorithms. As expected, the more documents used in training the higher the classification accuracy. There are clear differences in the classification accuracy of the methods. The AMF algorithm shows the highest precision 91.7% and average precision is 88.61% in the system. On the contrary, TFIDF shows the lowest average precision of 80.74%. Individually, the MCRDR and FCA algorithm achieve similar precision 85.86% and 86.12% respectively in the system. These results provide an initial benchmark and encouragement to continue our investigation in this direction. Future research should investigate and maximize the differences between Web documents and general text documents. In particular the structure offered by HTML and XML offer possibilities for incorporating alternative and multiple methods to achieve improved document classification.

Table 5. *Results of average precision of algorithms for sub-classes in Table 3*

	A1	A2	A3	A4	A5	A6	A7	A8
C1	80.3	80.8	78.3	84.2	85	85.8	85.7	86.3
C2	78.3	81.7	80	85	86.7	86.7	86	89.2
C3	83.3	84.8	81.7	86.7	85.8	87.5	82.8	87.6
C4	79.2	83.3	81.7	85.8	84.2	89.2	85.9	86.7
C5	78.3	82.5	79.2	82.6	85.5	87.7	89.6	91.5
C6	80.8	82.5	80	86.7	87.5	91.7	82.5	84.6
C7	81.7	85.2	81.5	85.4	86.2	87.5	86.6	88.9
C8	79.8	85	84.2	87.5	85.3	90	90.2	91.2
C9	85	81.7	78.3	88.3	86.5	89.2	86.4	87.8
C10	84.2	86.4	82.5	86.4	88.5	90.8	89.9	91.8
Ave	**82.39**	**83.98**	**80.74**	**85.86**	**86.12**	**88.61**	**86.56**	**88.56**

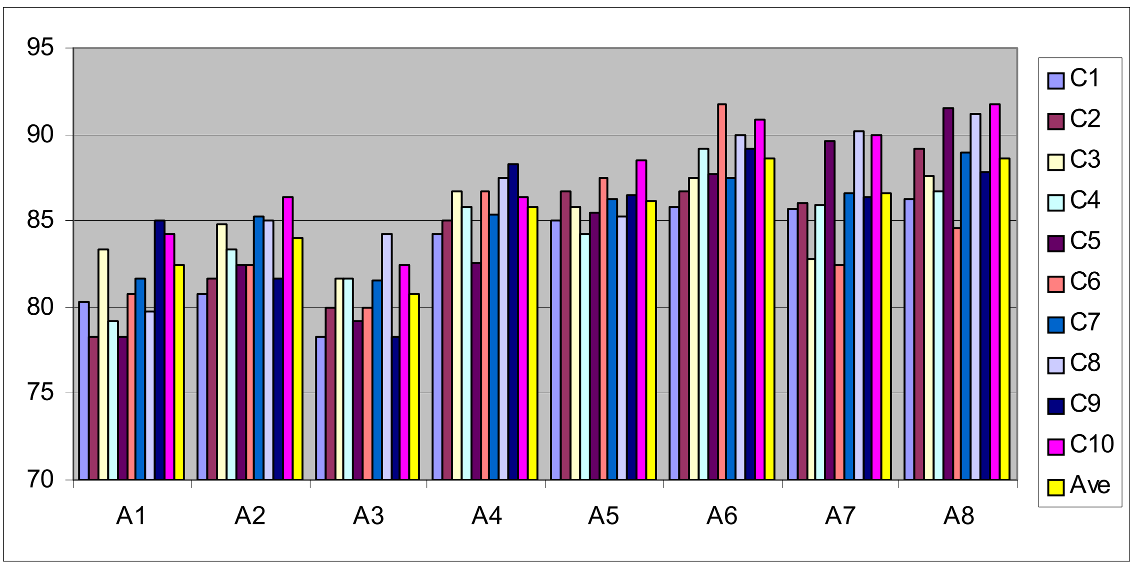


Figure 7. Results of average precision of algorithms for each sub-class in Table 5

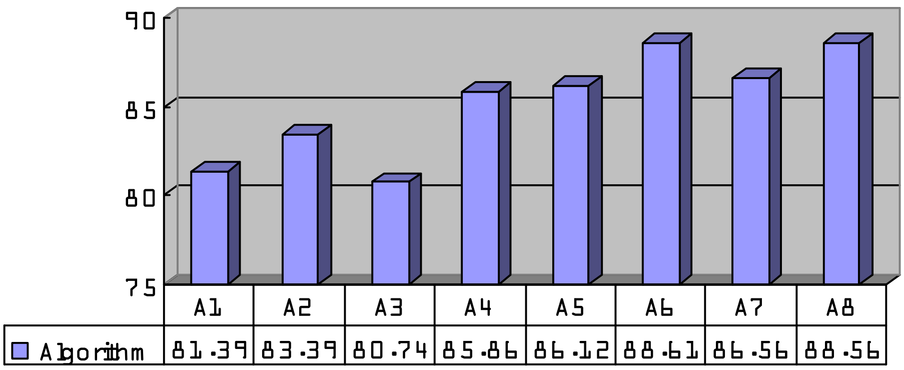


Figure 8. Results of average precision of algorithms in Table 5

4. CONCLUSIONS AND FUTURE WORK

As presented in the paper, we have achieved higher accuracy by using the AMF algorithm than existing classification algorithm like naïve Bayesian, TFIDF, SVMPR, SVMWTA, MCRDR and FCA classification algorithm. Future research should investigate and maximize the differences between Web documents and general text documents. In particular the structure offered by HTML and XML offer possibilities for incorporating alternative and multiple methods to achieve improved document classification. As supported by the results given in this paper, better accuracy in document classification can be achieved by improvements to the pre-processing techniques applied to document classification. The techniques employed for preprocessing, that is stop word removal, stemming and feature extraction, are common and were applied before all the algorithms. We propose that the use of a concept hierarchy provided the necessary structure to focus and organize the data better to achieve higher precision. In this paper we have focused on precision (usefulness of a hitlist) rather than the recall (completeness of the hitlist) as our goal was not to receive the most complete list of relevant web documents, which would take too long to read through, but that the ones we do retrieve are useful.

References

Birkhoff, G., (1993) *Lattice Theory 3rd edition*, American Mathematical Society, Incremental Clustering for Dynamic Information Processing, ACM Trans. on Information Processing Systems, pp. 143-164.

Choi, J.H., Seo, H.S., Noh, S.G., Choi, K.H. and Jung G.h, (2004) "Web Page Classification System based upon Ontology", *KIPS-2004*, Vol. 11, No. 06, South Korea, pp. 723-734.

Ganter, B. and Wille, R., (1998) *General lattice theory 2nd edition*. Birkhauser Verlag, Basel, pp. 591-605.

Ganter, B. and Wille, R., (1999) "Formal Concept Analysis – Mathematical Foundations", Springer-Verlag, Berlin

Kang, B.H., (1995) *Validating Knowledge Acquisition: Multiple Classification Ripple Down Rules*, PhD dissertation, Computer Science, University of New South Wales.

Porter, M.F., (1980) *An algorithm for suffix stripping*, Program 14 (3), pp.130-137.

Shannon, C.E., (1948) “A mathematical theory of communication”, *Bell System Technical Journal*, Vol. 27, pp.379-423 and pp.623-656.

Sinka, M.P. and Corne, D.W., (2002 “A large benchmark dataset for web document clustering”, *Soft computing Systems: Design Management and Applications, Frontiers in Artificial Intelligence and Application*, Vol.87, pp.881-890.

Yang, Y. and Pedersen, J.O., (1997) “A Comparative Study on Feature Selection in Text Categorization”, *14th International Conference on Machine Learning*, pp.412-420.

Wille, R., (1982) *Restructuring lattice theory: an approach based on hierarchies of concepts*, in: Ivan Rival(ed.), Ordered sets, Reidel, Dordrecht-Boston, pp.445-470.

Global Knowledge Management Systems: Key Cross-Cultural Issues And The Implications Of Sarbanes-Oxley Act

Mahesh S. Raisinghani
Pamela Smith Baker
Texas Woman's University, School of Management
Denton, TX, USA
Email: mraisinghani@mail.twu.edu

Abstract

In the current and previous decade, companies and their managers are relying on technology such as telecommunications, computers, and Internet to share knowledge and manage their businesses more efficiently and responsively (Galup, 2002; Dattero, 2002; Hicks, 2002). *Particularly with today's emphasis on competitiveness, team-based organizations, and responsiveness, top management cannot separate their responsibilities between people management, traditional/e-business management, and the demands of regulation, since all are interrelated in knowledge management systems (KMS). Understanding how to manage under conditions of rapid change is also now a critical skill. Today, work units in organizations with KMS are increasingly organized with team-based instead of the traditional top-down organization charts. On the top of teamwork-based condition, diversity must also be managed. Today's team based, geographically dispersed employees are increasingly guided by a network of values and tradition as part of an organizational culture in KMS. Traditions and values must be managed while managers deal with ever increasing demands of compliance and regulation. This paper explores, describes, and assesses the integration of KMS, the impact of KMS, the construction of KMS, acknowledging regulation's effect upon the way in which information technology is used to create new forms of knowledge creation, sharing, and sourcing.*

Keywords:
Knowledge Management Systems, Sarbanes Oxley Act, Regulation, Implementation, Cultural Issues

INTRODUCTION

All healthy organizations generate and use knowledge. As organizations interact with their environments, they absorb information, turn it into knowledge and take action based on it in combination with their experiences, values and internal rules. Without knowledge, an organization could not organize itself; it would be unable to maintain itself as a functioning enterprise (Capshaw, 1999). KMS must create knowledge to assist in strategic initiatives, operations, and compliance with mandatory regulation.

On the past decade of the business trends, perhaps none is more closely associated with technology than knowledge management. Technology likely will play an enabling part in these activities, but it is not to the extent occasioned by knowledge management. Knowledge management cannot be done without technology (Curley, 2001). Duhon (1998) defines Knowledge Management(KM) as "a combination of technology supporting a strategy for sharing and using both the brainpower resident within an organization's employees and internal and external information found in 'information containers'." The goal of KM is to simultaneously manage data, information, and explicit knowledge while leveraging the information resident in people's heads through a combination of technology and management practices (Duhon, 1998). Such integration is necessary to fulfill the diverse needs of management.
Human intelligence and experience can be stored and distributed by information. Because people come and go, technologies must be stored in databases for central access. It is not possible to scan a person's mind and store its content directly into a database; he or she has to put bits into a database (Hamblen, 2000). Technology, human experience, and cross-cultural data is combined to provide KMS satisfying the constant demands for planning, control, and making decisions while proving compliance with regulation.

CHARACTERIZATION OF IMPLEMENTATION SITUATIONS: REGULATION

KMS through combination systems purposely introduces complexity and even conflict to create new synergy (Linstone, 2000; Zhu, 2000). People who disagree in terms of their goals, values, or perspectives are more likely to find themselves arguing than are those with the same fundamental objectives. Other differences in goals, values, or perspectives that led to conflicts include those between flexibility and stability, and between short-run and long-run performance.
However, according to Fried (2000), KMS brings together people with different perspectives to work on a problem or project, forcing them to come up with a joint answer. The created new synergy can mean a joint answer. One of the most familiar and sensible ways of handling conflicts is to find some common ground on which people can agree. People

generally begin their work by finding some point on which both sides can agree and then building a solution from that point of agreement.

Of integral importance to all managers, and therefore to KMS, are governmental and other regulatory requirements applicable to each particular entity. Although some regulatory requirements are similar for all firms, firms in different industries often have differing rules to which they must conform. Wider still is the variation in regulations to which companies in diverse countries are bound. Effective KMS provides managers with access to all regulatory requirements, training as to the specifics of those regulations as they affect individual and corporate responsibilities and timely information as to particulars of the latest requirements. KMS should allow managers the ability to project the impact of new and proposed regulations. Further, the mandatory nature of regulation often aids in aligning purpose and effort.

The complexity of organization in KMS is more unpredictable and is constantly faced with the need to invent new decisions and respond quickly to emergencies. Creativity and entrepreneurial activities are emphasized. In many of today's boundary-less and team-based organization, management does not urge employees to stick to the rules or to the chain of command. Gilleland (2001) explains that the most important adaptive resources are employees who can acquire new knowledge and skills easily. Since the best predictor of mental nimbleness is proven experience in taking on new tasks, firms should seek out employees who have already mastered a variety of roles and skills. After they have been hired, employees should also be encouraged to change jobs often, to build and manage their own skill portfolios, and to master new work-related disciplines (Gilleland, 2001).

Although combination systems can lead to powerful results that are unobtainable in other ways, it is not a shortcut to KMS generation (Slywotzky, 2000). To get the powerful results, the compromise and collaboration is required. In collaboration, the both sides on the same side of the table to work out the agreement are often the best approach, especially when conflict is confronted in problem solving manner. Thus, a significant commitment of time and effort is required to give group members enough shared knowledge and shared language to be able to work together (Bonde, 2000). Vanhoucke (2001), Demeulemeester (2001), and Herroelen (2001) explain that careful KMS implementation is also necessary to make sure that the collaboration of different styles and ideas is positive, not merely confrontational. However, since KMS have members from different subsidiaries that typically compete with one another for scarce corporate resources, they tend to have a high degree of internal conflict - in addition to a low level of trust. As a result, it is generally best to form the team's group in terms of the company's position vis-a-vis the external marketplace instead of emphasizing internal dynamics (Scott, 1997). Among the many factors that determine how much trust people feel, three important ones are individual characteristics, quality of communication and the broader institutional context. People trust one another more when they share similarities, communicate frequently, and operate in a common cultural context that imposes tough sanctions for behaving in an untrustworthy manner. KMS, by their very nature, suffer from severe limitations on all three dimensions. Not surprisingly, when KMS fail, it is often because the team process did not emphasize cultivating trust (Rabey, 2001). New, complex regulations must be addressed with accuracy and completeness, requiring technology and human mental agility combined in teamwork.

Differences in regulatory requirements from country to country can widen the gap of trust in information produced, in that information users may perceive that information originated in lower regulatory environments may be of less quality. Managers of companies in lesser regulatory environments can alter KMS as they implement new technologies, adopt best practices in corporate governance, and provide better access to vital information in order to earn the trust of potential stakeholders worldwide (Valdez, 2005).

KMS REQUIREMENTS AND SARBANES-OXLEY

Several sections of Sarbanes-Oxley affect the manner in which companies do business as well as elucidate new requirements for internal control and management responsibility, heralding the need for expanded KMS. Sarbanes-Oxley requires that companies keep detailed records related to its financial systems, including electronic records and paper records. The KMS role in logging these records and making them available to management for assessment is critical. Document retention is governed by three provisions: a seven-year period for all documents that support a firm's financial report; a five-year period in the criminal provisions of the Act that covers all audit trails and audit workpapers, both hard and soft copies; and an undefined category of documents in connection with each board of director's inspection program which awaits board rulemaking as to its nature and scope (Miller, 2005).

Section 201 prohibits managers from using the same firm for auditing services and other types of accounting and management advisory services. Prior to the passage of SOX, companies could allow the accounting firm that provided their audit services to also provide information technology design and implementation, along with a host of other consulting services in aid to management. Now managers must deal with one firm for their audit while having other firms assist with information technology services. Thus, SOX expands the requirements for more participants in the inspection and creation of the KMS.

More important to managers are Sections 302 and 404. Section 302 requires that each CEO and CFO working for a publicly traded company prepare a statement certifying the "appropriateness of the financial statements and disclosures contained in the periodic report, and that those financial statements and disclosures fairly present, in all material respects, the operations and financial condition" of the company. Fraudulent statements are punishable by fines and prison terms, and ignorance of financial and internal control data is no excuse. To comply with SOX with comfort, KMS must provide managers with accurate, timely, and reliable information

A critical subset of KMS, internal control refers to "all activities and information services intended to protect assets and enhance the accuracy of financial information" (Kimmel, et al, 2004). Further, internal control pertains to "maintenance of records…in reasonable detail and reasonable assurance that information is properly recorded and … prevention or timely determination of unauthorized acquisition of protected information" is ensured (Green 2005). Per Section 404 of Sarbanes-Oxley, management must also provide a report discussing the adequacy of internal control. In this report, management must acknowledge in writing that the establishment and maintenance of adequate internal control is the responsibility of management. Further, this report must contain an assessment of the "effectiveness of the internal control structure and procedures" for the firm as a whole. Per SOX, auditors must perform an annual assessment of the adequacy of internal control. Failing this assessment is not an option, as consequences of failure would be devastating. These immense responsibilities for internal control systems require managers to have stronger control systems and more access to important data via KMS in real time.

As a result of the Sarbanes-Oxley Act, the certifying top management of the company and those middle managers who assist those officers must have, on a daily basis, adequate information to ensure that critical systems are set up and sufficiently monitored to be satisfied that all disclosure procedures and controls are operating effectively. Further, it is the responsibility of management to develop the systems that ensure the safeguarding of assets and important financial records.

KMS AND SARBANES-OXLEY ADOPTION

Managers are realizing the enormity of Sarbanes-Oxley adoption and maintenance and its implications for KMS. At Viacom, staff and consultants conducted almost 20,000 tests on 1,560 business controls and 540 IT controls last year; similarly, Time Warner committed 350,000 person hours examining internal controls in efforts to comply with Sarbanes-Oxley (Huffman 2005). Results of long hours reveal that problems have formed and will form that require continuous assessment and improvement. Many companies acknowledge that more training is needed for compliance to be complete. Lachnit (2001) suggests that companies will find that training provides a great return on investment, the most effective method for globally-dispersed firms being interactive and multimedia-based learning. Such training allows diverse and geographically dispersed team members to the training they need to do their jobs. In order to overcome the unique challenges confronting KMS and create a high-performing, effective unit satisfying the regulatory requirements of Sarbanes-Oxley as well as other needs, top managements must carefully craft the team's group, composition, and process. All three areas demand equal attention. A clear group without the appropriate mix of team members can lead to failure. Similarly, when a team's composition is sound but its process is not, effectiveness suffers. In short, the three work as a classification to determine overall team effectiveness. Further, in effective teams employees take an active interest in their colleagues and in improving the outcome of their mutual efforts (Eisenhardt (1997). Responsibility boundaries blur, and employees are more inclined to pitch in and help each other.

Because Section 404 of Sarbanes-Oxley consumes so much time and resources, many companies may find that, with scarce resources heavily consumed by Sarbanes-Oxley, other important strategic IT projects must be placed on the back burner. The issue becomes personal liability; thus senior managers are by necessity channeling IT spending into areas such as business intelligence in order to provide senior management with increased visibility into organizational performance and operations (Hoffman, 2005). This required use of resources could accentuate the demand for scarce resources and create conflict.

GEOGRAPHICAL IMPLICATIONS OF SARBANES-OXLEY AND KMS

Schottmiller (2000) suggests that hindrances to communication barriers resulting from differences in geography, language and culture also can impact KMS. As the world shrinks due to improved communication and travel technology, information requirements and information expectations collide. Recent difficulties in travel intensifies the need for better technology. Managers of companies around the globe fighting for the scarce resources supplied by global debt and equity investors must consider the impact of regulation in other countries upon perceptions of information. Acknowledging this globalization of markets, worldwide response to regulations in the United States is growing. Multinational companies whose stock is traded over U.S. public exchanges must comply with all regulations that U.S. companies must comply, one of which is Sarbanes-Oxley. SOX applies to any non-U.S. company registered on U.S. exchanges under the Securities Act or the Exchange Act, regardless of the company's country of incorporation. Over 1,300 non-U.S. corporations from 59 different countries file reports with the Securities and Exchange Commission, and many more would like to have the ability to sell stock on U.S. exchanges, capturing the economic advantages of expanded debt and equity markets. Managers of firms whose stock is not traded over U.S. exchanges are not responsible for ensuring their firms' compliance with U.S. regulations, many realize the importance of voluntary compliance. Thus, many managers of foreign companies whose stock is not traded on public exchanges in the United States based in countries with a lesser regulatory environment require KMS assistance in de facto adoption of regulations such as Sarbanes-Oxley. Financial managers in Europe, Asia, and Latin America see that the expansion of international relationships forces a "trickle-down" effect of Sarbanes-Oxley, pushing managers in foreign companies to set higher standards for corporate compliance and governance (Diaz, 2004)

With members living in different countries, separated by time zones and conflicting schedules, arranging team meetings can pose logistical challenges. Undoubtedly, technology (e-mail, teleconferencing and video conferencing) can enable members to work together despite geographical distance. But technology should be viewed as a complement to, not as a substitute for, team meetings. Face-to-face meetings foster familiarity and trust - not easily established through virtual meetings. When members cannot see one another's body language and directly experience one another's reactions, the emotional dimension critical to a team's success suffers. Moreover, certain types of deliberations demand face-to-face meetings. Brainstorming, for instance, requires unstructured free-form interaction over an extended period that is not suited to a virtual meeting (McKinsey, 2001).

LANGUAGE AND CULTURAL BARRIERS

The inability to understand what another person is saying is always a potential barrier to communication in cross-cultural settings. If language barriers are not adequately addressed, the likelihood of creating an atmosphere conducive to candid sharing of different viewpoints - and hence conducive to achieving creative solutions - is greatly diminished. Even in the case of KMS whose members speak the same language, differences in semantics, accents, tone, pitch and dialects can be impediments (Bonde, 2000). Although language may impede understanding, the specifics of regulation may improve communication through KMS. Further, the desire for compliance may motivate toward better understanding.

Members of KMS typically come from diverse cultures and, as a result, may bring different values, norms, assumptions and patterns of behavior to the group. Unless the differences in assumptions and beliefs inherent in that diversity are explicitly addressed, the cohesiveness of the group is likely to suffer and impede effectiveness (Long, 2000; Fahey, 2000). Eisenhardt (1997) mentions that changing the organization's culture can be important if top management spends all their time fine-tuning new tasks, instead of concentrating on the people side of the problem. They cannot succeed to see that collective responsibility is an attitude, a value, and a concern. It means taking an interest in one's employees and in improving the outcome of mutual (as opposed to individual) efforts. People who feel collectively responsible are willing to work especially hard to not let the team down.

KMS can also impact the organization culture through Internet. All an employee needs is access to a computer with a browser, a modem, and a telephone line or other Internet connection. The Internet can be used both to help create knowledge and to share that knowledge among dispersed users. Internet applications make it significantly easier to administer a selection process of knowledge sharing. Internet sources have become invaluable tools for sharing knowledge and ensuring compliance of Sarbanes-Oxley.

IMPLEMENTATION OF GLOBAL KNOWLEDGE MANAGEMENT AND REGULATION

In considering the plan to expand into the global marketplace, ideally, the expanding business would set forth a "no expense spared", where in management would focus on the potential benefits of the ensuing information technology implementation, and would worry less about its actual financial costs (Renkema, 1998). Unfortunately, this is not an ideal world, and the existing infrastructures, both technological and social, within each economy and each culture must be considered when planning for process updates. Looking at the developing countries first will give an "address the weakest chain in the link" approach toward planning for efficient global business processes.

In researching the existing infrastructures within developing countries, it has been hypothesized that "the lack of reliable transport and communication systems [which] compounds the general misery of disease, ignorance and poverty... [can be] attributed to a fundamental lack of information..." (Splettstoesser and Towry-Coker, 1999). This statement clearly reveals the real-world effects of the lack of information, not just in terms of disease, ignorance and poverty, but also in terms of the potential economic opportunities and financial markets that these dreaded maladies provide. As more of the developed countries realize the potential of connecting with the developing countries, a mutually beneficial partnership between the two economic imbalances is forming. The existing information platforms that have been enjoyed by the world's developed countries is helping the developing countries to progress into updated information technology without having to go through the growing pains of the developed countries' hard-earned learning curves.

Regardless of any types of cooperatives that might be forming, regardless of the motivations from which these cooperatives are born, the birth and maintenance of technical communicative information systems is quite complex. The willingness to acknowledge that organizational autonomy is actually an isolating weakness allows for a more interactive cooperation in these developed countries (Van den Berg and Mantelaers, 1999). The fact that this type of interactive cooperation is indeed possible serves as driver for these developed countries to consider the potential opportunities of underdeveloped marketplace.

In order to expand into the global arena, the developed countries must consider the political and economic conditions that exist in the developing countries, and the less developed countries must find the means to fund the costs of increased regulation and expanded KMS. In some of the developing countries, the access and acceptance information technology has historically been closely monitored by the respective governments. The concern that any individual citizen might have access to technology serves as one of the drivers for these developing countries' leaders to attempt to suppress the growth of information technology. Therefore, access to information technology has been restricted only to the highest in political power. However, in other developing countries, the governments have actually take of the potential development opportunities that accompany information technology, and have opted to facilitate their own inclusion into the global communications network (Splettstoesser and Towry-Coker, 1999). Obviously, each developing country appears to have its own unique set of challenges from the viewpoints of the developed countries.

In order for true cross-cultural cooperation to occur, information technology expansionists must research potential solutions to overcoming the inherent expansion barriers of hardware and software challenges, political, cultural, legal and language challenges. Expansionists must also take clear note that it is a part of human nature to be resistant to change, regardless of the cultural differences that exist from country to country. Offering a specific example, the aforementioned intersection of the human-based, perception-run, interfacing world intersecting with the implementation of information technology is still quite apparent in the African countries today. Although initiatives for progress are already in place, widespread acceptance of the technology is still an outstanding issue in most African developing countries (Splettstoesser and Towry-Coker, 1999).

Nelson and Clark (1994) contend that "the development, operation, use, and impact of information in different cultures" must be considered in the development of cross-cultural systems. With the development of such systems, engineers must consider how the proposed system[s] will be accepted by any given culture, the structure of the organizational environment, the basic architecture of the organizational process, how the organization considers the variables involved with decision-making, and performance outcomes, combining all these components into a measurable conclusion.

IMPLICATIONS FOR MANAGEMENT AND RESEARCH

The implications of KMS support managerial decision-making and control purposes. KMS are more than computers. KMS also include employees and top management, since it is their specific needs for knowledge sharing that KMS is designed to serve (Copeland, 2001). A KMS provides decision support for top management by producing periodic standardized, summarized reports, and diverse information for decision support, while helping top management make decisions that are relatively unstructured (Wu, 2000). KMS is also knowledge sharing systems designed to help top-level management acquire, manipulate and use the knowledge they need to maintain the overall strategic effectiveness of the companies, keying strengths and weaknesses, opportunities and challenges. Top management also uses KMS to quickly identify and understand evolving competitive situations (Mellander, 2001).

CONCLUSION

This paper explores, describes, and assesses possible issues of relevance include integration of KMS, impact of KMS, construction of KMS, acknowledging the importance and overarching nature of regulation. As the workforce becomes more diverse, it becomes more important to manage diversity so that the benefits of diversity outweigh any potential drawbacks. Managing diversity in KMS involves taking steps such as providing strong leadership, assessing the situation, providing training and education, changing the culture and systems, and evaluating the program.

KMS refers to any processes practices, or systems that facilitate the processing and transportation of data and information. The increasingly rapid deployment of KMS in organizations has speeded the transformation of many businesses today into knowledge sharing-based organizations. Shared knowledge is data presented in a form that is meaningful to the recipient. Knowledge, on the other hand, has been defined as information distilled via study or research and augmented by judgment and experience. Good-quality information must be pertinent, timely, and accurate and reduce uncertainty

It is safe to state that the world is moving more and more toward being a singular, global community. The developed countries, typically capitalist, are home to highly competitive capitalist businesses that are constantly seeking to expand. Therefore, these businesses constantly have to adapt to the ever-changing technological advances and accompanying external demands. With the advent, acceptance and exponential growth of technological advances (such as the internet and similar communications technology), both the developing countries and the developed countries are being forced to see the potential reciprocal benefits of the implementation of information technology and communications, understanding the costs of cross technological adaptations. However, each of these two groups of countries needs to take the time to research the cultural issues that will inevitably present themselves. Political barriers, legal considerations, language barriers, and the ever-present human trait of resistance to change must all be researched. All of these intangible real-world concerns that are a part of human life and human culture must be considered while the tangible, technical side of communication continues to advance at accelerating rates.

REFERENCES ARE AVAILABLE FROM THE AUTHORS UPON REQUEST

Long Term Information Needs In the Architecture, Engineering, and Construction (AEC) industries

Refad Bader
University of Western Sydney
Sydney, Australia
Email: refad1980@gmail.com

Howard Leslie
Sydney, Australia
Email: h.leslie@pacific.net.au

Dr.Yogesh Deshpande
University of Western Sydney
Sydney, Australia
Email: y.deshpande@uws.edu.au

Abstract
This paper is a contribution to the growing body of research work and literature that focus on the AEC sector. The appearance of information technologies has contributed dramatically in changing the business of almost every industry in the world. However, the Architectural, Engineering, and Construction (AEC) industry has had a mixed record in adopting and adapting to the new Internet technologies. Considering the huge amount of information that results from AEC projects, a need arises for an efficient way to save and access this information in the long term and use it for myriad reasons over the lifetime of a facility. The paper describes and analyses the information needs and its handling in the AEC sector.

Keywords:
AEC, Long-Term Information, Standard Business Vocabulary, Open communication Technology, Project Life Cycle.

1. INTRODUCTION

The appearance of information technologies has contributed dramatically in changing the business of almost every industry in the world. The rapid growth of Internet technologies is opening new possibilities for businesses to enhance their processes. However, the Architectural, Engineering, and Construction (AEC) industry, in spite of being one of the largest businesses in the world, has had a mixed record in adopting and adapting to the Internet technologies [12]. As Stephens [10] remark: "… it is still not very easy to find information about products, components or professional services when using the Internet. In many cases there is no option but to consult paper based catalogues and trade indexes. When such catalogue information is available via an Internet based service it is unlikely that the catalogue information can be easily re-used by other software applications. This means that various actors: Contractors, web portals, and product manufactures cannot exchange such catalogue information easily and therefore support business-to-business communication.

At the same time, AEC as a whole is facing increasing pressure to improve the effectiveness of its processes and the accountability of its decision makers. New techniques of handling increasing information load are needed to improve the transformation of data into information and then into knowledge [9]. One of the characteristics of information handling in the AEC sector that is in marked contrast to the general information management is the long-term relevance of a substantial amount of AEC-related information. A building or a facility, for example, typically will last for anything from 10 to 100 or more years and the decisions made at any stage of managing such facilities could have an impact many years down the line. This paper concentrates on this aspect of the information needs of the AEC sector.

The AEC industry deals with huge amount of information, some of it relevant for short periods of specific stages of a project, e.g. product enquiries, and some for the longer term, e.g. product characteristics which could affect the maintenance (replacement) or even demolition stage (as with asbestos) down the line. The availability of information that influenced the early decisions will make it easier to maintain or renovate or even demolish the facilities. The ability to access and exchange information of AEC projects between its actors in the future is a very important factor in the success of this sector [1]. Considering the huge amount of information that will result from each project, a need arises for an efficient way to save and access this information in the long term and use it for myriad reasons over the lifetime of a facility.

This paper is a contribution to the growing body of research work and literature that focus on the AEC sector, its practices and problems as well as the solutions made possible by the advent of new, especially Web-based technologies. The analysis and development of an efficient way to save and access AEC information in the long term will be the main goal of this research. The paper is structured as follows. Section 2 briefly explores the AEC sector and its contribution in the national economy and hence the need to address its problems. Section 3 covers the sector's long term information needs and its importance as well as the ways in which the information is managed, highlighting in the process, the problems, and deficiencies that various experts have explored and are addressing. Moreover, this section expands on this topic, comparing and linking it to one of the research priorities, viz. memories for life, identified by the British Computer Society and the ACM.. Finally, section four will present the direction of this research in the future.

2. ARCHITECTURE, ENGINEERING, AND CONSTRUCTION INDUSTRIES

The AEC sector plays a very important role in the world economy today. According to Bloomfield [14], the construction industry remains the largest industrial sector ahead of the foodstuffs and chemicals industries, and it is crucial in producing investments goods. It employs a large percentage of work forces, e.g. 5% in Australia in the years 2002–2003 and is characterized by the dominance of small and medium enterprises (SMEs) [7].

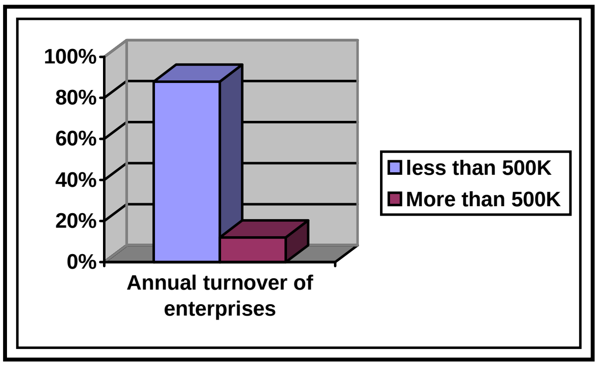


Figure1. Annual Turnover of Enterprises in NSW [7]

Figure 1 shows a rough distribution of the enterprises in the AEC sector in the state of NSW. From the total of about 25,000 enterprises in this industry, subcontracting enterprises, characterized by the total turnover of less than $500,000 per annum, deliver between 75% and 85% of the value of the industry.[8]. The SMEs have generally lagged behind the bigger enterprises in their uptake of IT [15] but, in keeping with the trends, they face growing pressures to improve the effectiveness of their processes and the accountability in their decision-making. New techniques of handling information are therefore needed to improve the transformation of data into information and then into knowledge [9].

Projects in the AEC sector include a variety of actors such as clients, architects, contractors, and construction mangers, all exchanging large amount of complex information at different times and locations [1]. Moreover, AEC project evolves over a project life cycle through several stages from design and construction to operation, maintenance and demolition, and many tasks during the life cycle, typically require the collaboration of those different parties involved this information generally includes the management of time, cost, quality, health and safety, environmental impact, and the exchange of communication [3].

3. AEC LONG TERM INFORMATION

As explained in the previous section, a huge amount of information is needed and recorded in AEC projects and maintained afterwards for facility management. This information is increasing in size and importance everyday. As a result, AEC has moved from paper-based systems, in which documents were physically exchanged, to an electronic version of the same systems [1].

At the same time, the ability to access and exchange information of AEC projects between its actors is a very important factor in the success of this sector [1]. Considering the huge amount of information that is resulting from each project, a need for an efficient way to save and access AEC information in the long term and use it for many reason such as maintenance, modification, or learning is arising [2]. The developments during the last decade in IT show that software and hardware are changing rapidly.

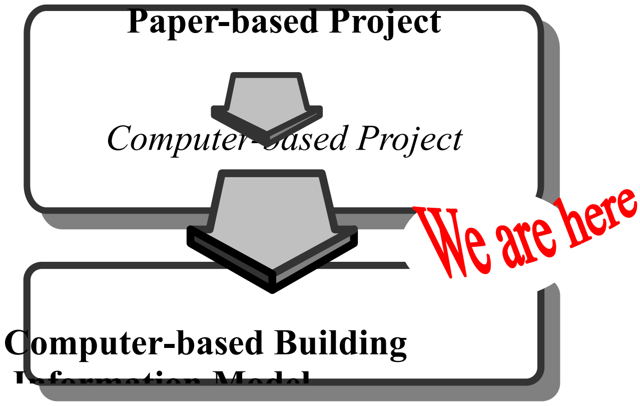


Figure2. Bridging the gap [1]

The international alliance of interoperability (IAI) worked in a model that is funded by industry leaders in 19 countries to define a comprehensive building information frame work as one authoritative semantic definitions of building elements which is called the IFC model [16]. IFC aims to fulfill the gap in AEC sectors for moving towards a computer based information model by using industry foundation classes to facilitate interoperability in the building industry [17].

The question naturally arises as to how AEC Information should be stored over a long period of time, allowing efficient access in the future whenever users including those who may not have been part of the initial project need it for modifying a part of a specific project.

3.1 AEC LONG TERM INFORMATION NEEDS

Memories of life [4] is a project that is currently being studied by a group of researchers in order to find efficient techniques for saving huge amount of digital data for a long time and access it efficiently in the long term future. In discussing the project, Fitzgibbon and Reiter [4] observe that “People are capturing and storing an ever-increasing amount of information about them selves, including emails, web browsing histories, digital images, and audio recordings. This tsunami of data presents numerous challenges to computer science, including: how to physically store such “digital memories” over decades; how to extract useful knowledge from this rich library of information”.

While Memories for life arose from concerns about managing digitally captured data by individuals and hence concentrates on them, similar concerns exist for more organized, industrial and business sectors as well.

As mentioned before, AEC projects capture and contain huge amounts of information through their cycles. Nevertheless, AEC sector still lacks an efficient way for saving this information as also the ability to access and retrieve the required information in the long-term future [2]. Consider, for example, a wall built as part of a specific room in a construction project. After twenty years, if some maintenance work is required to be done on that wall, it may become difficult to retrieve relevant information about the exact components and materials of that wall or the record of maintenance operation that was done on that wall. In such cases, the only way to carry out any meaningful activity is to ask engineers and expert builders to examine the wall and trust their judgment. A way for handling long-term information in AEC projects has to be established.

It is not realistic to expect the participants in the AEC to give up their existing tools and start using a homogeneous and overly complex single building model. Moreover, there is no single set of tools that can include all aspects of a sophisticated facility in a single information model [17].

AEC projects members are interested in building elements that are relevant to their role as architect, engineer, contractor, or designer. Details that are relevant to one project member may be nothing more than needless information to another member. As a result, AEC inherited ‘islands of information’ with project data tied up in proprietary file from a number of different actors. In order to achieve greater results in productivity, it is important to unlock the value of the information trapped on those isolated islands and allow it to be openly shared between software applications across the entire project team [17].

This information, which will be shared between project team members, will be also saved for future usage. In order to share and save this valuable project information, the management of AEC projects needs to rely on effective type of communication that comprises a large number of different professions in these sectors. This needs standardization of terminology, as a starting point. XML holds a promise in this area and has been employed by several groups to create new mark-up languages. The reason for using XML (Extensible Markup Language) is that XML is a text-based data format that can easily be parsed and understood by a variety of different platforms [13]. A number of projects are using xml in the same area such as gbxml, NewForma, and eConstruct.

Green building XML (gbxml) is a leading industry example of how the project information can be exchanged and shared between existing software applications [18]. Experience on live projects has demonstrated that using gbxml can dramatically streamline the transfer of building information between architectural and engineering models eliminating the time consuming processes for exchanging the information. [17]

NewForma – a venture funded software development company - is trying to increase the effectiveness and productivity of AEC sector by developing software that enables the flow of information between every building project team member in support of both project and business processes [19].

E-Construct is a European funded research project that has developed a new communication technology for the European building and construction industry, called building and construction extensible markup language (bcxml) in which it provides the European building and construction industry with powerful but low cost communication infrastructure [10]. bcXML is designed to support the communications needed between clients, architects, suppliers, and contractors when they are searching and procuring components [1].

eConstruct Project started in January 2000 and was expected to be finished by January 2002. According to [10], the main goals for eConstruct project were to Supporting the communication between Clients, Architects and Engineers, Suppliers of construction products (components and systems), and Contractors (i.e. eBusiness), enabling e-Procurement software to search for products and services from a number of Marketplaces or Portals, enabling manufacturers to provide catalogue content in a neutral format rather than a different format for each Marketplace or Portal, linking product data technology (PDT) with e-Procurement, integrating e-Procurement with eCommerce applications for invoices and payments, supporting the virtual construction enterprise over the borders of the individual European member states.
In eConstruct Project, two generations of applications were built. The first generation of application focuses on allowing the users to find a specific product and the best price for it. The second generation of applications focuses on helping to increase the competitiveness of the European building construction industry by supporting business-to-business transactions [10].

Another great need to facilitate the long-term information is to have a unified communication standard. It is a common consensus that common standards are vital in order for businesses to be able to communicate over the Internet efficiently.

A standard specification for representing components involved in AEC projects needs to be created, accepted, and used by parties involved in AEC projects [10], because this standard specification will guide the transaction of information between enterprises and their actors before saving them. EConstruct project [6], created a taxonomy, based on XML, in order to allow the users to save, retrieve, and transact their data smoothly in a clear, formal, and agreed definition of the industry objects and their properties and their relationships.

A standard business vocabulary is being developed as part of the ongoing research. The purpose of this glossary is to serve as a template that guides the exchanging of business information in AEC projects; this glossary is being developed using XML. Gbxml, eConstruct, and NewForma are examples of large projects, which are addressing similar opinion in this area. In addiction to their work, which is covered in the current research, there will be a deep concentration on ensuring the accessibility of AEC information in the long term, developing the vocabulary in such a way that it can grow in order to allow the addition of any new building components, and giving the individuals the ability to categorize their information into public and private information, where only the public information can be accessed and shared with other parties included in the project.

Moreover, this standard business vocabulary will be able to handle electronic business trading. Data can be read and processed by the machine so a computer could place a purchase order with the supplier's computer and make the payment without a person being involved at either end. At the moment, this is already possible with only large companies, but this project will be targeting small and medium enterprises to allow them to electronically trade their products.

In order to demonstrate the feasibility of the AEC for using a standardized vocabulary for communication and saving information for long-term use, a prototype is being developed. This prototype demonstrates how small and medium enterprises in the AEC can easily create standardized XML documents from business vocabulary templates, and use these to exchange information.

The SMEs can use their existing computers and software to smooth the transition to this proposed application. No huge investments are necessary in the initial phase of the transition. The main functionality of this prototype is to have a repository in which industry agreed XML business vocabularies; business collaborations and data couplets are identified, described and held.

The repository of long-term information will serve as the collaborator that provides the trading partners with the resources to prepare and exchange information. Once the required template is downloaded by an enterprise, it is manipulated internally with the use of the proposed prototype software. The processed XML document is then transmitted to the trading partners. At the end of each stage in the project, an xml document that includes the agreed components used in the project and their attributes will be saved in the long-term repository for future usage. The repository will thus have the agreed glossary, standardized and documented means of communicating data and information among the actors and the actual data. The repository will also identify and label information which will be needed for long term as opposed to, say, financial transactions which are generally not required beyond four or five years, depending upon the legal and audit requirements of such transactions.

4. RESEARCH APPROACH

While the focus is on the AEC industry – due to the background of those involved and the fact that the issues facing this industry are representative of the wider community – the approach adopted seeks to identify generic issues and arrive at universally applicable solutions.

Having identified a representative set of data access and exchange scenarios, the aim is to create demonstrable prototypes which will be used as proofs of concepts to discuss and explore the issues and the possible solutions with the appropriate industry organizations.

Three categories of information are currently under analysis. The first deals with the search, selection and specification of components to be included in an AEC project. Within the individual phases of this process the information can be considered 'short-term'. However selected data must be passed from one decision-maker to another (selector to specifier to purchaser to installer to operator and maintainer). The time scale involved is commonly measured in years.

The second part deals with the exchange of business information. This part will concentrate on information about transactions, such as purchase orders, invoices, delivery notes, payment records and so on. There are legal requirements of how and how long financial transactions must be kept for accounting and tax purposes. Such information may be considered as medium term information that must be archived and may be needed by various agencies in the future. A complete architectural project itself will require and generate a tremendous amount of information that may have to be accessed over a long period, as briefly discussed earlier.

The third part of our approach looks at identifying and aggregating relevant medium to long-term information. Such data includes planning requirement, various building codes and standards, client brief and user requirements, product manufactures instillation and operating instructions. Maintenance of and access to long-term information are currently seen as the province of people who are responsible for project archives. Policies and procedures are therefore formulated to facilitate these functions. In effect, longevity of information is specified in the process side of maintenance. We suggest that this longevity should be an attribute of the information item itself. In some ways, usability experts have already made Web site developers and administrators aware of this association through the recommendation that each Web page carry both 'last updated' details and meta tags which would tell the browsers when the page is set to expire. Apart from the logical association of an information item with its life expectancy, such a strategy will also help to overcome the implications of proprietary software not being available or unable to deal with information created by earlier versions of a software package. The problem is one of identifying relevant information (which the practitioner may or may not know exists), bringing it together at a time and in a format that will facilitate understanding and application as well as flag conflicts.

4.1 FUTURE WORK

This research is still in an early stage. There will be communication with other potential projects such as eConstruct, NewForma, and gbxml to examine their work in depth.

Long-term information needs will be examined in depth in order to define all the requirements for an effective long-term storage for AEC project information. Moreover, further development is needed for the completion of an application that uses the standard business vocabularies for transferring the information in AEC projects, which will be saved for long-term purposes.

This proposed application has to take into consideration the type of actors who are processing the information. For example, designers will be interested in specific information that is different from the type of information that a contractor might be interested in. The prototype has to insure the delivery of the right information for the right user.

In the Product Enquiry Search, in which users will be able to search for products they need and find the best price for them, the information will be valid and required only until a selection is made, after which only the details of the selected product need to be maintained. The 'life' of this information will depend on the types of the products and may range from short to long term. The second part, which handles the business information exchange will concentrate on information about transactions, such as purchase orders, invoices, delivery notes, payment records and so on. There are legal requirements of how and how long financial transactions must be kept for accounting and tax purposes. Such information may be considered as medium term information that must be archived and may be needed by various agencies in the future. A complete architectural project itself will require and generate a tremendous amount of information that may have to be accessed over a long period, as briefly discussed earlier. The third part of our approach is to identify a specific area where architects, planners, builders and facility managers will all ask, years down the line, for information about specifications of what exists and even how some decisions were made.

All these issues mentioned in this section, would make the SMEs in the AEC industry become more efficient, and lead to a smaller level of redundant information than before. It would also make the SMEs getting a higher level of IT sophistication on their way to making the business processes more efficient because this proposed long-term repository will stand as a valuable source for knowledge and experience generated from previous accomplished projects that will be used in new projects in order to improve the progress of the work.

5. REFERENCES

[1] H. Leslie, "Information Management in the Building and Construction industry", Private Communication, 2004

[2] H. Leslie, "Performance-Based Product Access in the Building and Construction Industry", Private Communication, 2005

[3] D. Bouchlaghem, A. Kimmance, Integrating Product and process information in the construction sector." Industrial Management and Data Systems 104, 2004, pp218-233

[4] A. Fitzgibbon, E. Reiter, " Memories of life, Managing information over human life time", 2003, URL: http://www.nesc.ac.uk/esi/events/Grand_Challenges/proposals/Memories.pdf

[5] M. Hinton, "'Towards a pattern language for information centred business change', International", Journal of Information Management, 2002, vol. 22, no. 5, pp. 325-341

[6] C. Lima, and M. Bohms, "The bcXML prototype: the eConstruct approach supporting eCommerce in the Building and Construction Industry", 2004, URL: http://www.itcon. org/data/works/att/ 2003_22. content. 08439.pdf

[7] NOHSC National Occupational Health and Safety Commission, "National Data Set for compensation statistics", 2004, URL: http://www.nohsc.gov.au

[8] NSW Government, "An overview of Construct NSW". Construction Agency Coordination Committee, 2003, URL: http://www.construction.nsw.gov.au/ConstructNSW/overview.html

[9] S. Shabban, H. Elkadi, “Information visualization for the architecture practice”, 2002, URL: www.doi.ieeecomputersociety.org/10.1109/IV.2001.942038

[10] D. Stephens, C. Lima, A. Ton, R. Steinmann, K. Woesteneknk, M. Bohms, and F. Tolman, “eCommerce and eBusiness in the building and construction industry”, 2002, URL: http://www.econstruct.org/6-Public/bcXML_CD/PublicDeliverables/pfr_v3b.pdf

[11] F. Tolman, R. Rees, R. Beheshti, M. Bohms, P. Debras, A. Zarli, and R. Steinmann, “eConstruct public deliverables”, 2000, URL: http://www.bcxml.org/6-Public/bcXML_CD/PublicDeliverables/d101_v4.pdf

[12] E. Gustavsson, “Electronic Marketplace Solution for the Construction Industry - Content, Challengers and Solutions”, Royal Institute of Technology, Stockholm, Sweden, 2004, http://www248.indek.kth.se/forskning/publikationer/ examensarbeten/2001/407.pdf.

[13] B. Harvey, "The Role of XML in E-Business, 2002, "URL: http://www.eccnet.com/papers/ssgr/ssgr.pdf

[14] R. Bloomfield, “An Internet Gateway to European Construction Resources", 2001, URL: www.buildnet.csir.co.za/constuctitafrica/authors/Papers/w78015.

[15] A. Ginige. Re Engineering Software Development Process for eBusiness Application Development, 2003, http://www.uws.edu.au

[16] J. Foster, I. Howell, “Facilitating Interoperability in the Building Industry - A Different Approach to Using IFCs”, April 2005, URL: http://www.aecbytes.com/ viewpoint/issue_15.htm

[17] DBIA, “Information Technology: What are the opportunities for design –builders?”, June 2005, URL: http://www.dbia.org/1206batcheler.pdf

[18] J. Kennedy, “Green Building XML”, 2000, URL: www. Gbxml.org/

[19] NewForma, “Building information modeling two years later – huge potential, some success and several limitations”, may 2005, URL: www.newforma.com

On the ontological aspects of Knowledge Management

Sai Lakkaraju
Central Queensland University – Sydney International Campus
333 Kent Street, Sydney, NSW 2000
Ph +61 2 8295 5992
s.lakkaraju@syd.cqu.edu.au

Vijay Khandelwal
University of Western Sydney
Locked Bag 1797, Penrith South DC NSW 1797, Australia
Ph +61 2 96859236 Fax +61 2 9685 9245
v.khandelwal@uws.edu.au

Abstract

Over time three generations of Knowledge Management have been identified. With phenomenal advances in Information and Communication Technology (ICT) the adoption of Knowledge Management in organisations is growing rapidly. At the same time the ontology of an organisation in view of Knowledge Managers is changing too. Several ontological perspectives have been proposed in the past, but we believe none of them is complete. The incompleteness is because of the ambiguous definition of knowledge. This paper revisits various definitions of knowledge, and defines knowledge as information rationalised through reflection, learning and logical reasoning. Finally, a model that can be easily adopted by any organisation in defining its ontology is presented. In this model, observations and wisdom are subjective and data, information and knowledge are objective.

Keywords:

Knowledge Management, Ontology

INTRODUCTION

Gruber (1993) has defined ontology as an explicit specification of conceptualisation of objects, concepts and other entities that are assumed to exist in some area of interest, and the relationship among them (Genesereth et al. 1987). He also pointed out that the term ontology in philosophy has a broader definition as a "systematic account of existence". The former is objective whereas the latter is subjective. In the context of organisations Knowledge Management systems started to evolve based on Gruber's definition, resulting in the Knowledge Managers looking at the organisation from a different perspective. In general an organisation consists of people, processes and technology working towards a common goal. We believe that Gruber's definition of ontology is not sufficient to look at the organisation as a whole, because the conceptualisation of human factors in an organisation is quite complicated. Therefore throughout this paper the definition of ontology refers to the latter unless stated explicitly.

Over time, three generations of Knowledge Management (Knowledge Management) have been identified (Deeper, 2004; Snowden, 2002). In the first generation (prior to 1995) the ontology of an organisation, being objective, consisted of data, information, and knowledge. The focus of Knowledge Managers was on appropriate structuring and flow of information to decision makers, and in the automation of major business applications. Because of its objectiveness the first generation of Knowledge Management never took into account the human elements, and as a result was proven inefficient and financially disastrous in most of the cases by 1985. In the second generation of Knowledge Management which started with the popularisation of SECI (Socialisation, Externalisation, Combination and Internalisation) model (Nonaka et al. 1995), the focus was on the movement of knowledge between tacit and explicit states through the four SECI processes. The second generation acknowledged the human elements but instead of incorporating those into the organisational structure it tried to eliminate the human aspect by capturing the subjective knowledge and then converting it into objective knowledge. The ontology of an organisation, according to the second generation Knowledge Managers was based on the works of Polanyi (1983) and Nonaka et al. (1995), which consisted of data, information and knowledge- both explicit and tacit). In this generation knowledge was considered subjective as well as objective. But this model was proven expensive and incomplete. The third generation started with the argument proposed by Stacey (2001) that, "knowledge is not a thing, or a system, but an ephemeral, active process of relating". This is supported by Sveiby's (2001) statement that," knowledge can be considered as an object as well as a process", and that the focus is on human resources as well as Information and Communication Technology (ICT) infrastructure. The ontology of an organisation, according to the third generation Knowledge Managers, consists of data, information, knowledge and wisdom. Being subjective it is not tangible and ignores (sometimes even undermines) the objective nature. The definition of knowledge thus became ambiguous which can be seen clearly in Sveiby's definition. Moreover the third generation Knowledge Management practitioners could not clearly draw a line between knowledge and wisdom.

From the above, one can clearly see an ontological shift in the definition of knowledge from objectivist to subjectivist. Objectivists define knowledge as a thing. An observer perceives it in the same way as any other observer. Subjectivists define knowledge as personal. Thus knowledge becomes a process of relating, and the pursuit of knowledge is based on individual experiences. It can be concluded that in the first generation knowledge was considered as a thing following objectivist ontology. The second generation considered the subjective nature of the knowledge in the form of tacit knowledge (Polanyi, 1983), and attempted to convert the subjective (tacit) knowledge to the objective (explicit) knowledge. In the third generation, knowledge was considered subjective. Similarly the ontology of an organisation shifted from objective (data, information and knowledge) in nature to subjective (data, information, knowledge and wisdom) in nature.

The remaining part of this paper is organised as follows. In the next section we discuss three ontological perspectives of an organisation developed by Spiegler (2000), Kakabadse et al. (2003) and Awad et al. (2004) as these are widely used by the third generation Knowledge Managers. We also present the drawbacks of those models. We then define terms used in our proposed model namely, observation, data, information, knowledge and wisdom. Finally, we propose a model that has five phases: observation, data, information, knowledge and wisdom, in which observation and wisdom are of subjective nature, and data, information and knowledge are of an objective nature, thereby incorporating the subjective (people) and objective (processes and technology) nature of an organisation.

From reality to wisdom

Spiegler (2000) proposed a model that has four states: data, information, knowledge and wisdom. According to this model, for an observer, the process of creating wisdom from data starts with reality, which is related to the entities that are related to the reality. The four states then are described as follows:

- Data are the attributes of the entities. Data (bases) represent, record, store, and maintain those attributes.
- Information is "knowing-that" and is the result of data processing operations such as organising, storing, calculating, retrieving and reporting.
- Knowledge is defined as "knowing-how" and is a consequence of information processing operations such as reformatting, quantification, qualification, clustering, learning, and dissemination.
- Wisdom is "knowing when" and/or "if". Knowledge contributes to wisdom through activities such as discovery, inference, value, experience, judgement, intuition and abstraction.

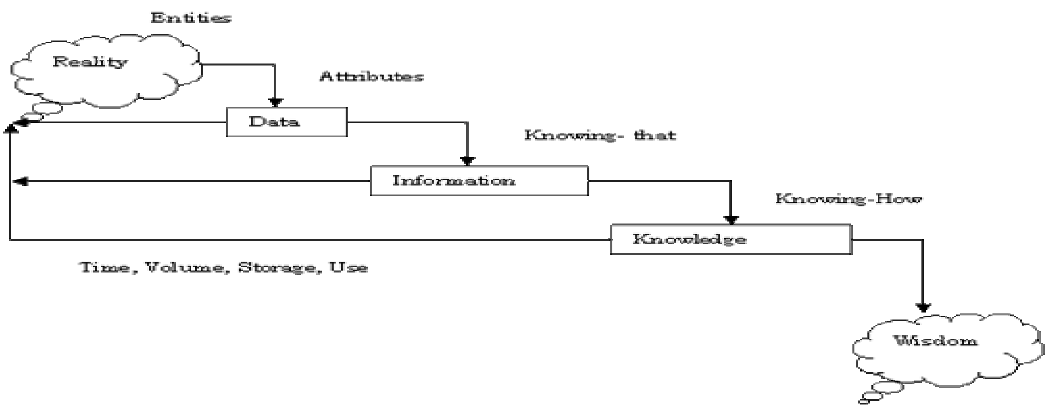
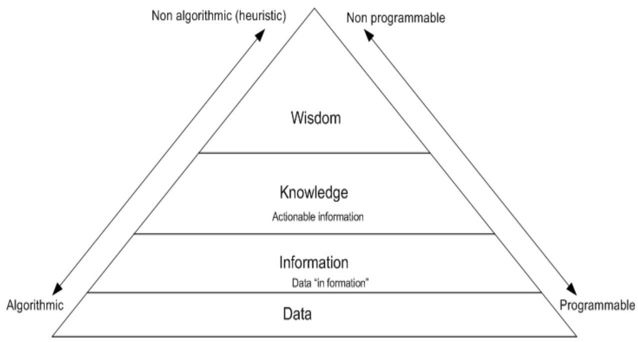


Figure 9: From Reality to Wisdom (Spiegler, 2000)

According to Spiegler (2000) knowledge is the process of knowing, a reflexive process that takes data and information, in a social (organisational) context, together with values and beliefs, needs, emotions, desires and socialising into a culture. It generates new data, information and/or knowledge. He further explained that knowledge and information turn into data with time, updates, reuse, application and more.

We can thus clearly see that wisdom is subjective, as the factors or knowledge processing operations that convert knowledge to wisdom, such as values, judgement, intuition, insight, and creativity are subjective in nature.



However this model neither explains the function of wisdom nor its effect on the reality. As discussed earlier an organisation consists of people, processes and technology. If we consider an organisation as reality, then because of the people factor one can conclude that there exist tangible attributes as well as intangible attributes. But this model does not discuss how to handle the intangible attributes. The only way to do so is to convert the intangibles to tangibles, which Sveiby (2001) claims has been done since 1985, and which is followed by many Knowledge Management practitioners.

From data to wisdom

Kakabadse et al. (2003) proposed a model that has five states: data, information, realisation, action/reflection, and wisdom. According to this model, for an observer the process of creating wisdom from data follows the following sequence.

- Data that are events observed explicitly, represented and organised, and are context free.
- Information is the data converted through manipulation, evaluation and interpretation.
- Realisation is the information converted through validation, internalisation and codification.
- Action/Reflection is the will applied to realisation.
- Wisdom is the experience of action/reflection.

Figure 10, shown above depicts the model proposed by Kakabadse et al. (2003). It is evident that action/reflection and wisdom are subjective in nature. One can classify data either as subjective or objective. If we consider data as objective in nature, then observation relates to the observations made by the processes, or the technology such as data mining tools. In such a case this model possesses the same drawback as the model (Spiegler, 2000) shown in Figure 9. If one considers data as subjective by incorporating human considerations, then this model fails to explain questions such as, what triggers such observations. Is data subjective or objective?

Figure 10: The Knowledge Triangle (Kakabadse et al., 2003)

From data to wisdom

Awad et al. (2004) proposed a model that has four states: data, information, knowledge, and wisdom. According to this model, for an observer, the process of creating wisdom from data takes the following sequence.

- Data that is unrecognised and unprocessed facts. Data is algorithmic and programmable.
- Information is the data “in formation”, which is an aggregation of data.

- Knowledge is the actionable information, human understanding of a specialised field of interest that has been acquired through study and experience.
- Wisdom is non-algorithmic and non-programmable.

From the definitions and explanations given by Awad et al. (2004, p. 40), it is evident that they considered only the objective nature of data. Rather than maintaining a balance, the authors leaned more towards Knowledge Management technologies than Knowledge Management processes. Now, every Knowledge Management practitioner knows that both the technologies and processes play a vital role in creating a Knowledge Organisation. This model fails to explain how knowledge (actionable information) turns into wisdom. The major draw back of this model is that it does not discuss the subjective nature of data or observations.

From Observations to Wisdom

Before presenting our model, we will discuss the terms observation, data, information, knowledge, and wisdom.

Observation

According to Davis et al. (1988) observations are the basis by which one recognises facts. They are perceptions of the observer about reality. They are experimental in nature and deal with the act of noting some object, or occurrence of some phenomenon in reality. Facts are those things or phenomena which the observer believes to be true. Facts are generally consensual in nature. The observers, who have observed the same phenomena, agree to their existence.

Data

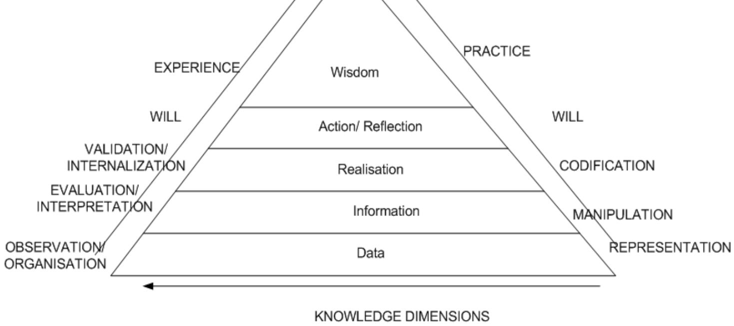
Figure 11: Data, Information, Knowledge, and Wisdom (Awad et al., 2004)

When thinking of data most people seem to emphasise the quantitative form that is easily stored and manipulated in computer systems. In our view data is observations organised for analysis. Data refers to chunks of facts about the state of universe of discourse. Data may be either quantitative or qualitative in nature. Quantitative data is data in numerical form. Qualitative data includes virtually any information that can be captured that is not numerical in nature. Observations made by researchers who follow qualitative research methods like participant observation, direct observation, unstructured interviews and case studies, generate qualitative data. Quantitative data is 'hard', 'rigorous', 'credible', and 'scientific' whereas the qualitative data is 'sensitive', 'nuanced', 'detailed', and 'contextual' (William, 2004). The processes and the technology available in an organisation help people in the organisation to convert observations into data. In our view data is objective in nature.

Information

Information generation from data requires four steps if it is to be useful beyond observation, or to be useful to individuals and groups other than the observer (Galliers, 1987; Giese, 1962; Nichols, 1969). These steps involve:

1. Classification of data, that is, relating the observations to anticipated situations of all classes. This classification reduces the complexity of the material, provides a means of identification by grouping like things together, provides a record of experience, and orders and relates classes of events. Three major characteristics of any classification system are:



   a. Classes must not overlap.
   b. The classification system must be exhaustive. Each item to be classified must be placed in some distinct category.
   c. The basis of classification must be significant (related to some specific goal), and in accordance with some predetermined pattern.
2. Establishment of procedures for recording data in a manner facilitating recall yet simplified to enable the operation to be routinised.
3. Summarisation of data classified and recorded.
4. Specification of the collection procedure of the system.

Knowledge

Knowledge Management practitioners have defined knowledge as an object or a process based on their ontological and or epistemological stand. It can be observed that knowledge was treated as an object until the emergence of the third generation of Knowledge Management which argued that knowledge is a process (Stacey, 2001). Sveiby (2001) took it further stating that, "Knowledge can be a process or an object". Keeping in view of the purpose and scope of this paper, we define knowledge as the rationalisation of information through reflection, learning and reasoning.

In the second generation of Knowledge Management, following Polanyi (1983) most Knowledge Management practitioners (Awad et al., 2004; Davenport et al., 2000; Gamble et al., 2004; Nonaka et al., 1995; Tiwana, 2000) believed that knowledge could be classified into two broad categories: tacit and explicit knowledge. These categories could be further subdivided into several types. Also, each category consists of various components such as intuition, experience, ground truth, judgment, values, assumptions, beliefs, and intelligence.

Explicit knowledge is that component of knowledge that can be codified and transmitted in a systematic and formal language: documents, databases, internets, intranets, e-mails, charts, etc.

Tacit knowledge is personal, context-specific knowledge that is difficult to formalise, record, or articulate. Tacit Knowledge is stored in the mind. The tacit component is mainly developed through a process of trial and error encountered in practice.

Wilson (2002) argued that the concept of "tacit knowledge" was wrongly understood by the Knowledge Management practitioners. The term "tacit knowledge" originated by Polanyi (1983), with his statement, "We know more than we can tell", pointed out that there is some knowledge which cannot be made explicit. Tacit knowledge is hidden knowledge, hidden even from the consciousness of the knower. Polanyi equates tacit knowledge with "acts of comprehension" by his statement that, "tacit knowing achieves comprehension by indwelling, and… all knowledge consists of or is rooted in such acts of comprehension" (Polanyi, 1958).

In other words, "tacit knowledge" involves the process of comprehension, a process which is, itself, little understood. Consequently, tacit knowledge is an inexpressible process that enables an assessment of phenomena in the course of becoming knowledgeable about the world. In what sense, then, can it be captured? The answer, of course, is that it cannot be captured. It can only be demonstrated through our expressible knowledge and through our acts (Wilson, 2002).

It may be argued that Knowledge Management practitioners wrongly understood the concept of tacit knowledge, or perhaps assuming there is a way to externalise tacit knowledge.

To keep us away from the ambiguity of tacit knowledge, explicit knowledge, and questions like "can tacit knowledge be captured?" We define knowledge as the rationalisation of information through reflection, learning and logical reasoning following Kantian epistemology, in which knowledge is perceived as a thing, something absolute, awaiting discovery through scientific reasoning. Following this definition of knowledge two categories of knowledge can be identified: explicit knowledge and implicit knowledge.

As discussed above explicit knowledge is that component of knowledge that can be codified and transmitted in a systematic and formal language, such as, documents, databases, internets, intranets, e-mails, charts etc.

Implicit knowledge is the knowledge that can be possibly articulated but has not been done yet.

Though our definition of knowledge, being objective contradicts with the definitions of knowledge by the subjectivists like Davenport et al. (2000), we have incorporated the subjective qualities of knowledge in the paradigm of wisdom which we think is embodied in it.

Sveiby (2001) defined knowledge as “a capacity to act” and referred to Wittingstein and Polanyi’s quote, “Knowledge is an activity best described as a process-of-knowing” as the source. If the word “act” describes a combination of reflection, learning and reasoning both externally as well as internally, then our definition coincides with that definition, Else, our definition of knowledge includes it.

Wisdom

Knowledge applied repeatedly to justify the logical basis for an action or belief becomes wisdom. Wisdom is the intellectual capacity to consistently distinguish what is important, true, correct and fundamental. In the state of wisdom there will be no processes of reasoning. You just know that you are aware.

Though various definitions of wisdom and knowledge sound similar to the definitions of tacit and explicit knowledge, there is significant difference between them. Firstly, wisdom is embodied and cannot be transferred. A wise being can transfer knowledge to another being or a machine. A knowledgeable being or machine can transfer information to a being or a machine. But one can only be wise by exercising repeated application of knowledge to justify the logical basis for an action or belief.

In our view, for an observer the process of creating wisdom from observations can be considered as the systematic and self-consistent creation of an intellectual and experimental structure through a hierarchical process following the sequence given below:

- Observations, which are perceptions (that may or may not be true) of an observer, and which based on the wisdom of the observer are converted into data that are facts.
- Data is then placed into one or more structures (processes and technology), and thus converted into information.
- Information is then rationalised through the processes of reflection, learning, and logical reasoning to become knowledge.
- Knowledge is then applied repeatedly to justify the logical basis of an action or belief and becomes wisdom.

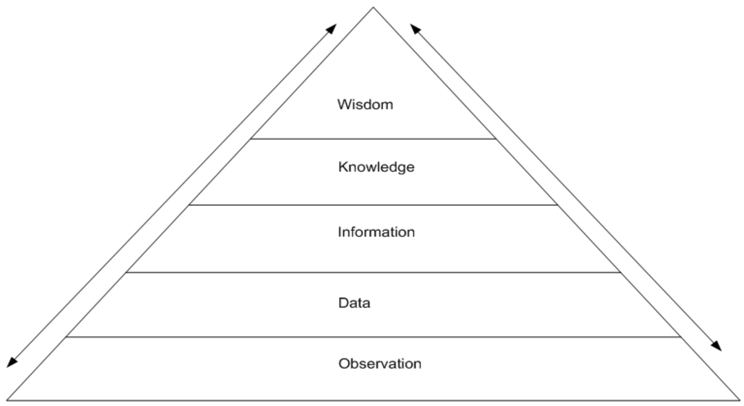


Figure 12: From Observation to Wisdom

Figure 12 depicts our model. In our model observation and wisdom are subjective and thus possessed by people. Data, information and knowledge are objective and thus possessed by people, contained in the processes and technology.

After making observations, new processes or technology will be developed by the people in the organisation over a period of time. By repeatedly using those processes and technologies people gain wisdom, which contributes to making new observations.

SUMMARY

Any organisation consists of people, process and technology. The ontology of an organisation in view of Knowledge Managers has been changing rapidly with the growth of ICT. Currently, we are in the third generation of Knowledge Management. In this paper, we have presented three widely used ontological perspectives and shown that they are not complete. We have also shown that the incompleteness in those models is because of the ambiguous definition of knowledge. Then we have defined knowledge as information rationalised through reflection, learning and logical reasoning. Finally, we have presented a model that can be easily adopted by any organisation in defining its ontology. In our model, observations and wisdom are subjective and data, information and knowledge are objective. Based on this model, we are currently investigating the epistemological, taxonomical aspects, and processes and technologies for Knowledge Management in organisations.

REFERENCES

Awad, E. M., & Ghaziri, H. M. (2004). *Knowledge Management*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Davenport, T. H., & Prusak, L. (2000). *Working Knowledge*. Boston: Harvard Business School Press.

Davis, D., & Cosenza, R. M. (1988). *Business Research for Decision Making*: PWS-KENT.

Deeper. (2004). *Business Knowledge Management: A study on market prospects, business needs and technological trends*: IBM Business Consulting Services.

Galliers, R. (Ed.). (1987). *Information Analysis:Selected Readings*: Addison-Wesley.

Gamble, P. R., & Blackwell, J. (2004). *Knowledge Management: A state of the art guide*. London: Kogan Page Limited.

Genesereth, M. R., & Nislsson, N. J. (1987). *Logical Foundations of Artificial Intelligence*: Morgan Kaufmann Publishers.

Giese, J. W. (1962). *Classification of economic data in accounting.* University of Illinois.

Gruber, T. R. (1993). *Toward principles for the Design of Ontologies Used for Knowledge Sharing* (No. KSL 93-04): Stanford University.

Kakabadse, N. K., Kakabadse, A., & Kouzmin, A. (2003). Reviewing the Knowledge Management literature: towards a taxonomy. *Journal of knowledge Management, 7*(4), 75-91.

Murdick, R. (1969). *Business Research: Concepts and Practice*: International Textbook Company.

Nichols, G. E. (1969). On the nature of management information. *Management Accounting* (April), 9-13.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York, NY.: Oxford University Press.

Polanyi, M. (1958). *Personal knowledge: towards a post-critical philosophy*: University of Chicago Press.

Polanyi, M. (1983). *The Tacit Dimension*: Peter Smith.

Snowden, D. (2002). Complex Acts of Knowing: paradox and descriptive self-awareness. *Journal of knowledge Management, 6*(2).

Spiegler, I. (2000). Knowledge Management: A new idea or a recycled concept? *Communications of the AIS, 3*.

Stacey, R. D. (2001). *Complex Responsive Processes in Organisations: Learning and Knowledge Creation (Complexity and Emergence in Organisations)*: Routledge.

Sveiby, K. E. (2001). *What is Knowledge Management?* Retrieved 25th February, 2005, from http://www.sveiby.com/articles/km-lessons.doc

Tiwana, A. (2000). *The Knowledge Management Tool Kit: Practical Techniques for building a Knowledge Management system*. Upper Saddle River, New Jersey: Prentice-Hall.

William, T. (2004). *The Research Methods Knowledge Base* (2nd ed.): Atomic Dog Publishing.

Wilson, T. D. (2002). The nonsense of 'Knowledge Management'. *Information Research, 8*(1).

The Effects of Knowledge Management on Information Sharing Practices: A Case Study on Nortel Networks

Dr. Katina Michael
University of Wollongong
Wollongong, Australia
Email: katina@uow.edu.au

Abstract

In 1999 Nortel Networks deployed the Open Text Livelink knowledge management system (KMS). Livelink allowed for the centralization of key corporate applications and associated content at a global, regional, line-of-business and departmental level. Prior to the implementation of Livelink on an enterprise scale, the corporation's 80,000 employees relied on fragmented departmental web pages which were scattered across 11 different Web servers making the task of finding information very difficult. This paper describes how the process of knowledge transfer at Nortel Networks changed with the deployment of Livelink and how it enabled the automation of workflows through the company's Web-based Intranet. The paper also provides an insight into how KMS empowered employees, acted to increase productivity and encouraged innovation. The importance of this paper is in highlighting the significant role of people in the success of KMS and to provide examples of the dynamics at play in such a large-scale operation.

Keywords:

Knowledge Management, Knowledge Transfer, Nortel Networks

INTRODUCTION

Knowledge management is defined as "the systemic and organizationally specified process for acquiring, organizing, and communicating knowledge of employees so that other employees may make use of it to be more effective and productive in their work" (Hahn & Subramani 2000, p. 302). It then follows that a knowledge management system (KMS) are all those components (software, hardware, people and processes) that support knowledge management initiatives. These may include but are not limited to work flow maps, web sites, portals, document management systems, customer relationship management (CRM), data warehousing, data mining processes, virtual teams, contact lists, databases, collaboration tools, applications and news (Davenport & Prusak 1998; Jashapara 2004). Although a term that is often used interchangeably with document management and information management, knowledge management seeks the higher ideal of *wisdom*, the tying together of the tacit and explicit realms. As Agostini et al. (2003) add: "knowledge is not important per se, instead the process of knowing, learning, and creating knowledge is the relevant aspect" (Nonaka & Takeuchi 1995). In this paper, knowledge management is seen through the eyes of a multinational telecommunications corporation, Nortel Networks. The main objective of the paper is to tell the story of the impact that Open Text Livelink had on 80,000 employees and their information sharing practices in a corporation that spanned a presence in over 150 countries. The before, during, and after KMS snapshots are presented to bring to the fore those overriding challenges, struggles and subsequent successes that follow an implementation of a large-scale eBusiness solution. In addition knowledge transfer dynamics in the company are explored, as are the effects of downsizing on the value of knowledge management.

METHODOLOGY

Data collected for this research was done through focused and unstructured interviews, observation and participation from 1996 to 2001, typical of a combinative ethnographic and action learning approach. The researcher was a business and network planner who was also given the role of the departmental chief knowledge officer. The business unit the researcher worked in spanned four countries in the Asia-Pacific including Australia, Singapore, India, China. While the majority of the study was based on data collected in that region, the researcher was seconded to international offices (including the United States and Canada) for periods of up to four months at a time, which aided in a more global understanding of information sharing practices pre and post-KMS implementation throughout Nortel Networks. The main actors in the study were employees the researcher worked with on a daily basis, ranging from director, senior manager, architect, and graduate level positions. By studying the interaction between these employees, who depended on each other for the completion of over one hundred and fifty projects in five years, and studying the way the KMS was used post implementation it became apparent that people would play an integral part in the success of the KMS. Focused interviews involved top level management including directors and senior managers who had been mandated with making the KMS successful; while individual unstructured interviews were conducted with over twenty employees to assess knowledge management practices and employee attitudes toward the KMS. Apart from interviews, the researcher conducted cyclic observations, of which the richest source of information was gathered during formal KMS focus group sessions and departmental post-mortem analysis sessions of account activities. These were mainly documented in the form of minutes of meetings and personal diary entries. Data was analyzed using a historical approach with a distinct narrative style of writing. The paper serves to fill a gap in the literature by addressing the need to present empirical evidence with respect to knowledge management in a social-technical context. While a great number of knowledge management case studies appear in popular business literature and Internet-based customer testimonials, very few studies have had access to the actual knowledge management system in place.

THE FRAGMENTED INTRANET

It was not that long ago that employees of global companies relied on fragmented web pages scattered across a multitude of servers throughout the world for their corporate information needs. Conducting searches for key pieces of data integral to the successful completion of a task was often a hopeless exercise. More often than not one had to pour through completely irrelevant hits on the company's search engine only to walk away without locating anything of value. Nortel employees working in offices throughout Asia in 1996 for instance, had limited access to information generated by employees in other regions. Even within Asia the practice of uploading data onto servers as a means of making documents available to others was uncommon, save for small software development teams whose work by its very nature demanded accessibility. File transfer protocol (FTP) servers was the closest anyone got to sharing data and these were often analogous to garbage dumps that were cleared periodically to free up server space. File naming conventions were absent, as was version control, and any other form of metadata describing individual documents.

Beyond FTP servers, emailing attachments was heavily relied upon to the detriment of increasing operational costs. Employees on occasion were warned by superiors- even as high up as region presidents- to pick up the telephone instead of emailing, as the costs for transporting megabytes of data throughout Asia were becoming exorbitant. A good PowerPoint package for instance could make the rounds of your inbox even as often as five times from five different sources. It was also a topic of amusement and debate that original work completed by an employee in one office (e.g. a newly created Excel model) would make its way across the CORWAN and come back several months later by email from another employee unknown to the original author. Sometimes authorship was even overridden and credit given to another individual who had simply adapted a few bits and pieces. This could be considered a form of internal plagiarism, save for the fact that the company theoretically owned the intellectual property. In other cases, information was emailed without an audit trail of recipients; there was no concept of privileges (save for hardcopy documents that contained a front cover distribution list), and "confidentiality" became a word that had different meanings to different people. In fact, releasing a document to the account team, independent of the level of security clearance placed on it, meant that it would end up in the hands of customers within days, if not hours, even if the contents had not yet been fully discussed internally. Stating that something was "confidential" was like placing a magnet on it for unauthorized disclosure.

On other occasions employees would receive large email attachments that had little- if anything- to do with their daily work tasks. As 'downloading-on-demand' was still unfeasible given the lack of infrastructure availability and adequate web training, broadcasting messages would ensure blanket coverage of the employee base and thus not miss any of the key recipients it was meant for originally. It was not on a few instances, however, that commercial Nortel product pricing lists (including margins and discount rates for different countries) would make the rounds of everyone's email inbox. This was not only a careless practice but competitively foolish. The telecoms sector is a small world, many employees working at Nortel at the time, had extended family working in opposition vendors. Even worse was that this type of practice was never identified as strategically perilous by upper management.

At the time the typical departmental setting was one where the majority of working information was stored on local hard-drives instead of a common server with employees responsible for making their own back-ups of data. At Nortel Networks Wollongong key project member's computers often fell victim to viruses or worms. And to make matters worse, as if the loss of files was not detrimental enough, back-up storage procedures for laptops and notebooks were non-existent. Project team members were often left scrambling to locate older versions of files to meet customer deadlines, on occasion losing days and even weeks of coherent work and research. Employees also required numerous passwords for a variety of applications, most of which would expire or be forgotten. The absence of a central login to company applications was always a contentious issue as company employees lost valuable time waiting for IT personnel to reset passwords when they could have been working on important documents for customer engagements. Smart card secure ID badges were even deployed to staff in 1997 for remote access but due to synchronization problems they were abandoned some time later.

THE PROBLEM OF KNOWLEDGE TRANSFER

The absence of a central portal for employees also meant that individual web-based applications were unknown unless the universal resource locator (URL) was launched via email or some other general communications forum. Different departments within the company may have had their own web pages but again these were inadequate, poorly maintained and updated, and had a very small audience with little or no access security on the intranet. Up until 1998, subordinates relied on the ingenuity, good name, and goodwill of their supervisors and managers to gather and socialize important information. For this reason, a good supervisor could fast-track your career giving you access to more. Other supervisors would hoard information, keep it to themselves for the purposes of self-promotion, and then tell their subordinates that they did not wish to overload them with unnecessary information. Employees in teams were expected to share their findings with one another in order to complete tasks but this was not always the company culture. It was impossible to know who the key experts in the company were for collaborating on projects, unless an employee was introduced by word-of-mouth or chain-style emails. To be good at one's work, more often than not, meant that an individual had to have a good network of colleagues- knowing the right people could save you a great deal of time, not to mention raising the accuracy of the actual results or solutions proposed.

Expatriates were often brought in to enhance the transference of knowledge between Canada, the US, France and other more isolated or newly established regions like Australasia. But no matter how good and strong these internal networks were there was always a question mark surrounding the vintage of the information sent by key contacts. Was it the latest product information for instance? Was the plan-of-records (PoR) the most recently defined? There was not always enough time to check these very important questions- especially given the time zone differences between Asia and other Nortel offices. One had to go on what they had as it was better than nothing and at least more accurate than a guesstimate. Sometimes consultants were seconded to projects for a short time, their access to company information was even more limited, and so they spent time reinventing the wheel, separated physically from the rest of the organization.

Collaboration was mainly insular, within project teams, and there was no manner to denote who did what in the corporation. Even up until 1998, the online corporate directory only noted the employee's name, telephone number, location, and reporting manager. Thus, the problems were not only physical in nature with regard to the CORWAN infrastructure but were application-centric as well. Simple Word documents like company policies were even difficult to locate on local servers. These factors altogether contributed to a loss of productivity and propagated inefficiencies across departments. The problem however was not isolated to Nortel Networks- all the other telecommunication giants were suffering likewise. Companies were struggling with how to manage "knowledge" in large corporations with complex product and service mixes. The answer was to evolve to a better working environment that took advantage of internet protocol (IP).

Doing business at web-speed caused dramatic changes not only to the way information was exchanged but to the way people themselves worked. At about the same time that Nortel merged with Bay Networks and announced the "right-angle" turn from circuit-switched telephony to IP, the company CEO John Roth, decided to invest in knowledge management (KM) to help facilitate the merger process, promote knowledge sharing among employees, encourage refinement of business processes through workflow management leading toward ISO9000 certification, offer a central login for employees, and assist in employee communications from the top down and vice versa. Roth was a CEO who sought opinions on particular issues directly from his employees. He purposefully showed by example, crafting global memos that were pages long of heart-to-heart reflection. He made employees feel special, even if they were grass-roots installers. If he shared his knowledge, the mandate was that employees should also. He urged employees to think together, and create an environment of openness to help win more business. His decision to implement Open Text Livelink fundamentally came from a corporate need to remain competitive, even though the decision to specifically purchase the Livelink system was decided on a golf course.

THE PROCESS OF CHANGE

Livelink did more than just enforce a technical change in infrastructure layout. It changed the way people worked and it challenged individual beliefs about ownership of information. The process of implementing a knowledge management system (KMS) was more than just about allowing the centralization of information and enabling the collaboration between individuals in different regions. It was to strike at the very core of departmental and global business practices. In fact, the implementation of Livelink coincided with the company's efforts to attain ISO9000 certification for as many different functional areas as possible. Some departments, like the Network and Systems Solution (NSS) department in Asia, found the challenge almost impossible. There was no defined workflow to how employees in designated roles conducted their studies, and studies varied in time, size, complexity and resource requirements. Some employees continually insisted to work with local files and share only a selection of documents with their project team. One of the key architects in the NSS team in Wollongong believed that their sophisticated models, if placed in the wrong hands, could have major repercussions on Australian business. He argued that his tools contained a great deal of sensitive customer and proprietary information, and if used inappropriately would mislead other customers or give competing companies unfair advantage. During the dot.com bubble, it was reported widely by the internal security team, that industrial espionage was a common happenstance. This is at a time that Human Resources (HR) were on a recruitment drive, offering between two and five thousand dollars for the successful recommendation of local scare talent. The competition for knowledge, especially that implicit knowledge residing in the brain of an employee, was fierce.

Chief knowledge officers were appointed in departments as well as ISO9000 team leaders to help the process overcome initial teething problems. However the use of an ISO 'policeman' in each department put some individuals in some very difficult situations. Some employees clashed with the ISO mandate which made them automatically rebel against the use of Livelink when in actual fact the two were separate requirements. More generally, there was resistance to change toward Livelink, and it seems that this was primary due to the lack of information provided to employees before it was rolled out worldwide. For instance, some employees complained that it was too time consuming to upload and download relevant data from the KMS and have to record the adequate metadata for every single document. While other employees saw the advantages of downloading-on-demand and the right to access useful information that could help make them more effective employees. Chief knowledge managers for some of the larger departments were also appointed to promote the use and benefits of Livelink but these employees were often ostracized by others who did not respect their work. Most employees viewed these individuals as an unnecessary company overhead however, claiming that they knew little about how the business worked and were restricted in what they could bring to teams as they were not involved in the initial creation of knowledge. Subsequently, the perception was that the need for chief knowledge managers to come up to speed meant that employees would be disrupted by incessant questioning.

There were a number of shortcomings related to the KMS in 2000 many of which were linked to the capacity for Livelink to handle multiple file types, especially object programming code and geographic information systems (GIS) extensions. Other problems were practical in nature, such as: where did the given department fit in the global organization structure, how would the department segment their server space to provide a repository of information that was meaningful in nature into the future, how could all the features of Livelink be utilized effectively etc. Timely training of how to use Livelink was also lacking and many of these courses came post-implementation. Colleagues first heard about Livelink through the grapevine, in an almost organic fashion, but when some groups had access and others did not it became a little confusing. The phased deployment plan was never communicated properly to employees, if at all, in some lines-of-business (LOB). Livelink definitely required top-level management support but even with this backing it still took some time for the skeptics to be converted. In essence nobody was saying that files could not be saved locally but that all working files had to uploaded in a timely manner. When it came to deciding what kind of sensitive data/ models to upload and who could and should be able to view it, there were some interesting confrontations. In essence key personnel who were experts in a given area and were generating their own models to support their work tasks, did not wish to give up what made them special for others to easily mimic or learn from with time. There continued to be some resistance until these same employees began to use access privileges for their uploaded files and essentially block everybody but themselves from using particular files. This was not in the true spirit of the KMS but at least this promoted another level of back-up storage. The facility for a personal (i.e. private) workspace on Livelink was available but few took advantage of it, opting to place work documents on the enterprise workspace or store things locally.

ENTER A KNOWLEDGE INFRA-"STRUCTURE"

For some departments, the KMS was the answer to gaining timely access to internal and external intelligence information. For other departments, the KMS would help them in their quest to raise their profile by providing an avenue to showcase their work. But before launching any such site, groups had to work together to map out the layout and structure of their virtual space on the KMS. This was not an easy task especially for those who had been working with ill-defined processes in the past- they were not embarking on building a pretty web site but to some degree on aiming for best practice. The initial brainstorming period raised questions about how work was being carried out, the type of work being completed in some departments, and the level of quality and quantity of work being produced in other departments. In effect, this gave birth to project management demands in the company, and encouraged visibility and transparency throughout the organization.

The knowledge infrastructure did not appear overnight. Groups worked for weeks and in some instances, months, to define what they believed would be a "future-proof" layout. After all, this space was to be the interface between them and the rest of the Nortel world. For the NSS team, this required a lot of collaboration, consultation, and reviews (exhibit 1). It also had to be decided which documents, past and present, would be ported onto Livelink, how this would be done, and who would have the responsibility. In the end the manager and chief knowledge officers did the great majority of this work, believing in the system, and made it voluntary for other employees in the team to follow suit. It took some months before the whole group had bought-into the idea but the group finally was proficient at using the KMS. During projects only the crucial documentation would be uploaded to Livelink and the URL shared with other collaborators. At the conclusion of projects, all the inputs, processing, and outputs would be uploaded to Livelink. For the NSS team, it was hoped that one day they would be able to use the KMS to automate their Bill-of-Materials (BoM) sheet for customers' Request for Quotation (RFQ). The proposal was to create Adobe PDF forms for "inputting" and use extensible markup language (XML) to facilitate the end-to-end calculations in Microsoft Excel or Access. In this way it was hoped that reusable content and repeatable processes could save employees time and allow them to take on more projects than in the past, as well as decreasing their time-to-market (TTM). The idea was to be able to source data that was usually all over the place (in essence distributed), and to put it into some structured context, where it had an invaluable role. Too often market researchers and financial analysts in the corporation would spend hours if not days searching for the right value- the KMS was about to change things.

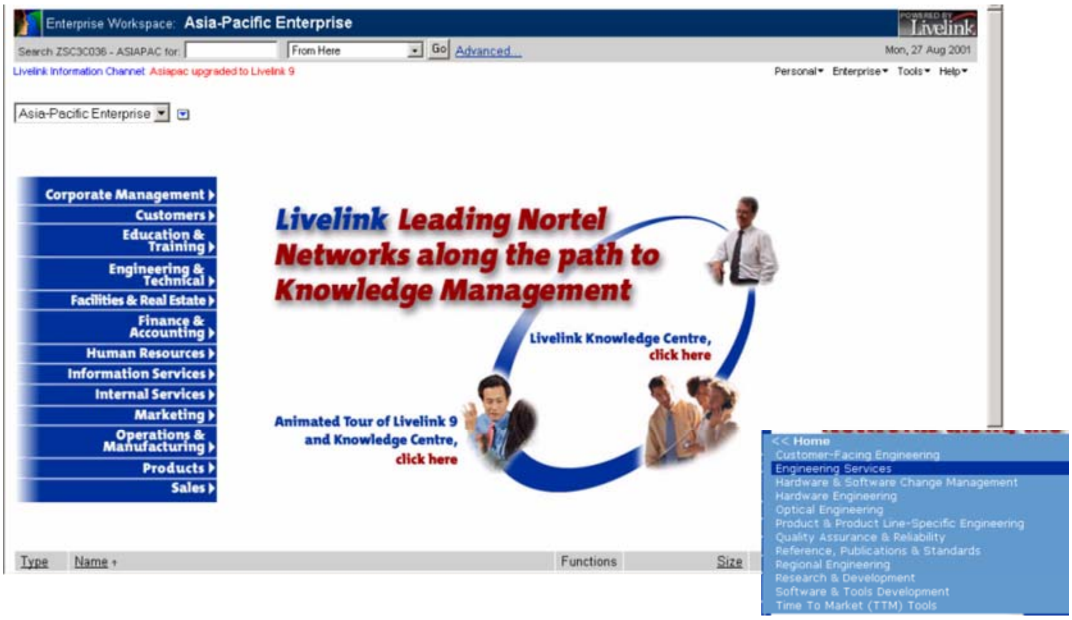


Exhibit 1. The Nortel Networks Livelink Portal in 2001

Initially only a small number of features were being used on Livelink from those available. For example, although it was possible to "check-in" and "check-out" documents, hardly anyone ever did. This would have ideally suited employees who were using databases and financial systems. Employees could also take advantage of creating metadata for their documentation or define access privileges but very few ever did. Naming conventions were specified, usually at the department level, but some documents complied while others did not. In brief, those practices that were pre-Livelink were difficult to break after the implementation of Livelink, although bit-by-bit, change did occur. For example, in the NSS team, filenames were made up of country telephone area codes, the type of task, the initials of the author and a date and version number. Livelink also allowed for the creation of workflows for specific projects, allocating tasks and their duration, and other dependencies. It was not that the feature was not useful but that project management skills were lacking.

To some degree, the majority of the corporation was using Livelink as a Document Management System (DMS) in the beginning, but this changed as time went on and new applications were specifically created to help the employees of the corporation communicate and collaborate better. The notion of virtual teams became prevalent around 2000, and this is when the KMS became very important. Customers also, for example, were given access to an extranet space in Livelink where they could upload and share files with their supplier. Account teams also made use of this capability to gather as much intelligence from clients and provide commensurate returns to them with product knowledge that was not accessible to the public via www.nortelnetworks.com. Livelink helped consolidate and strengthen business relationships. Within six months of its introduction the benefits of the KMS were evident. Remote dial-up access users especially found it much easier to send around a URL embedded in an email than having to wait over one hour to attach a file to an email and then send it to a list of recipients. It saved time and made employees more productive. One manager even credited the KMS to a growth in the number of patents generated by Nortel Networks. In the past the organization was very customer responsive at the expense of their knowledge creation- Livelink facilitated this business process (Perna 2001).

KNOWLEDGE MANAGEMENT APPLICATIONS

By the end of 2000, the knowledge management system was increasingly being touted as Nortel's most important corporate tool. It not only brought teams closer together that were previously geographically disparate but it formed the basis for the launch of the company's key communication and collaboration applications including: WorldOnline, StrategyNet, Market Research, Service Provider and Carrier Information (SP&C), Sales.com, Database Marketing Services (DMS), Corporate ID, Organization Structure, PeopleSearch, Building Locator, EmployeeConnect, CareerNet, Employee Training and Development, Information Services, Meeting Manager, Purchase Online, Travel Online and Stock Price, among others (exhibit 2). WorldOnline let the CEO directly broadcast multimedia messages to all the employees in the corporation. Employees could watch the broadcast live, or download a broadcast using Media Player. WorldOnline also reported the latest customer wins, highlighted key account and product strategies, and identified key employees and groups in the corporation. There were also a number of applications that supported sales-technical and marketing activities including Sales.com, SP&C, Market Research, and StrategyNet. Employees could use these portals to search for information about products, target markets, client backgrounds, and business case examples from across the globe. This knowledge empowered individuals and groups to produce higher quality output.

Exhibit 2. A Collage of Nortel Networks Knowledge Management Applications

Although some would argue that it was information overload at times, this was better than days gone by, when no data whatsoever was available. The assumed information overload problem could also help employees by allowing them to compare facts from a variety of sources, and grant them the ability to make a decision on which data was the most useful for a given project. Employees could also quickly ascertain who their counterparts were in other regions and who would be a likely collaborator for advice on technical matters. While collaborative tools like MeetingManager and NetMeeting were not a consequence of Livelink, they were taken advantage of more, because employees were made aware of their services via the KMS. The applications were paramount to those employees who made use of them every single day and multiple times a day. Livelink had become so embedded in practice that when the Code Red worm infected servers it knocked out two days of productivity for most groups. Without access to the KMS, people could only use the telephone to communicate (if they had the name of the person they wished to contact), read printed matter and or use electronic resources on their local desktop.

THE EFFECTS OF DOWNSIZING

In 2001, after dozens upon dozens of acquisitions, the company began to downsize as a direct consequence of the dot.com crash. At the time the share price of the company had reached some seventy-five US dollars at its peak, and at its lowest fell to below one US dollar. The regime to downsize, in some cases meant that whole departments were made redundant, irrespective of the top talent within it, and this had a major repercussion on the value of the KMS in the organization. Members from one school of thought could argue that the introduction of Livelink was "just-in-time", that it had taken root as an important tool before the downsizing was announced. Members from another school of thought could argue that the value of the KMS decreased after the rapid downsizing measures were enacted. Independent of the view taken the reality was that the KMS did help to retain corporate knowledge but it did not do so without end.

The challenge for those still employed by the company was first to know about the knowledge (in some instances it had been made private or read access only), and second if obtainable to know what to do with the information and how to use it. While some documentation was still used after the departure of the document creator, in most instances, employees still employed by the company felt they had to generate a lot of new material. In Nortel what became clear was that there was an intrinsic link between knowledge management and collaboration; take the collaborator away and the knowledge available to you lessens in worth significantly. The company continued to take drastic downsizing measures from 90,000 employees in 2001 to some 30,000 employees in 2004. Having cut about 60,000 jobs in three years the KMS could no longer expect to work miracles. The KMS once alive and used by so many, no longer had the same number of employees "feeding" it with information. Some employees, desperate to remain employed, even retreated to pre-Livelink practices, refusing to share their information with others, hoping that that would maintain their employability through the downsizing trend.

CONCLUSION

The value of knowledge management to large multinational corporations is undisputed. KMS is integral in organizations today that work at web-speed and require the creation of virtual teams who rely on reusable content and repeatable processes. However it needs to be emphasized that a KMS is not just a technology that can be implemented and can succeed on its own. It is people who will ultimately drive its success or failure. If used correctly the benefits are manifold including a dynamic working and learning environment that fosters information sharing and new value creation. Knowledge management helps employees build a collaborative culture, and create and extend their own personal business networks. Indeed there is a social side to this e-business solution. At no other time does this "socio-cultural" phenomenon become most obvious than when it is disrupted by necessary corrective actions to an organization's size. In the case of Nortel Networks employees, it was wonderful to share and create together until the threat of redundancies loomed and subsequently affected social practice. In times of downsizing the "one big happy family" culture is quickly overtaken by the "everyone for himself" reality and this has the effect of stifling the value of a KMS in the short-term, especially as employee morale is generally low during these times of substantial change. It does not mean that the KMS loses its value completely, to some degree if becomes increasingly important because there are less heads working together to solve the same number of problems. As the organization again reaches equilibrium the KMS can be used as a catalyst to re-build, re-create, and re-store.

REFERENCES

Agostini, A. et al. (2003). Stimulating knowledge discovery and sharing. Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work (pp. 248-257), Florida, USA.

APQC. (2001). Building and sustaining communities of practice- DaimlerChrysler AG case study. Knowledge Sharing Network. <Online> http://www.apqc.org/portal/apqc/ksn?paf_gear_id=contentgearhome&paf_dm=full&pageselect=detail&docid=101894 [Accessed 6 July 2005].

Davenport, T. H. and Prusak, L. (1998). Working knowledge: how organisations manage what they know. Harvard Business School Press. Boston, MA.

Hahn, J. and Subramani, M.R. (2000). A framework of knowledge management systems: issues and challenges for theory and practice. Proceedings of the Twenty First International Conference on Information Systems, (pp. 302-312) Brisbane, Queensland.

Jashapara, A. (2004). Knowledge management: an integrated approach. England: Prentice Hall.

Jennex, M. (2005). Case studies in knowledge management. IDG Press. New York.

Nonaka, I. and Takeuchi, H. (1995). The knowledge creating company. New York: Oxford University Press.

Perna, J. (2001). Reinventing how we do business. Vital Speeches of the Day. July 15th, 67(19), pp. 587-591.

Robertson, J. (2004). 'Knowledge sharing' should be avoided. Content Management Briefing. 16, p. 1.

ACKNOWLEDGEMENTS

The author wishes to extend her thanks to the staff of the Network and Systems Solution (NSS) team at the Nortel Networks Wollongong Laboratory, Australia.

MobileOffice Scheduler: Another Mobile Office Solution

Latifah Mohamed, Ida Zuraida Zakaria
Telekom Research & Development Sdn. Bhd.,
Selangor, Malaysia
Email : [1]latifah, [2]ida@tmrnd.com.my

Abstract

This paper presents issues and challenges in implementing mobile office solution. Generally, the issues can be divided into two main categories: technical and non-technical. We then present MobileOffice Scheduler system - another mobile office solution. MobileOffice Scheduler system was designed as an SMS/MMS-based application, mainly to overcome the dependency on network availability as faced by WAP-based application. The system consists of two main modules, (i) MobileOffice Portal and (ii) MobileOffice Scheduler client. The prototype of the system, with limited functionalities has been successfully developed, while the real system is expected to be deployed by middle of year 2006. We hope the implementation of this system could help in organizing manager's appointment schedules more systematically and thus expand the boundaries of connectivity among employees of an organization.

Keywords:

Mobile office, scheduler, portal, SMS, throwaway prototyping

INTRODUCTION

More than three decades ago, people are only able to fantasize about having access to data in their offices when they are at home or travelling. However in current technological world, having access to mobile computing environment is considered crucial for today's business. For example, a portable computer and mobile phone enable mobile staff to be more productive, spending more time with customers, clients or suppliers and proportionally less time in the office. The goal of "anytime, anywhere, anyhow for anybody" is fast becoming a trend that should be implemented by business organization in order to strive in the business world today. By defining, developing and deploying a mobility solution, it can greatly enhance the productivity of the mobile workforce and provide competitive advantage if it is developed and deployed within the context of the business requirement.

Mobile and wireless devices are enabling organizations to conduct business more effectively (Fui-Hoon Nah, et al., 2005). Mobile applications can be used to support e-commerce with customers and suppliers, and to conduct e-business within across organizational boundaries.

This paper presents an overview of mobile office as well as the issues and challenges that need to be addressed by the developer in implementing mobile office solutions. The paper then will look briefly at the MobileOffice Scheduler that has been developed by TM R&D Sdn. Bhd.

MOBILE OFFICE

To support an increasingly mobile workforce, many organizations are deploying mobile office solution to extend network access to employees that are away from their desks. These employees could do their job by either roaming within the office, tele-networking from home or even travelling on the road. But these mobile office solutions do not replace the wired network infrastructure, but rather extend it to areas where access to the wired network is out of reach or limited.

Mobile Office solutions meet companies' expectations by offering access to key applications through different terminals. Mobile Office allows corporate clients to be accessible anywhere and at anytime to send, receive, forward and reply to e-mails, including the ability to view attachments and forward them to a fax machine. Other examples of applications that are being deployed on top of mobile office solution are intranet, instant messaging, corporate directory and Short Message Service (SMS) and Multimedia Message Service (MMS) services. This greater flexibility in data services enhances employees' communications while they are on the move.

One successful mobile office example is the one launched by Telstra. Back in year 2000, Telstra released a package of WAP applications designed to keep mobile workers in touch with the office (Aldred, 2000). The service, which was meant for WAP-enabled mobile phones, provides access to MS® Outlook and Exchange Server information including e-mail, calendar, contacts, tasks and notes. Telstra's first corporate wireless initiative is enabling up to 1500 Telstra employees with Mobile Office for internal use.

ISSUES AND CHALLENGES

To have a fully mobile wireless office application, there are needs for thorough understanding of its users, rigorous design process and meticulous attention to the usability of both devices and applications. These challenges can be categorized into two main categories, technical and non-technical challenges.

Technical Challenges

Technical challenges may be further divided into other sub-categories; mobility issue, bandwidth management issue and device issue.

Mobility of clients in this environment results in constantly changing topology of the systems, hence calling for mobility of resources. Thus, there is a need for location management that deals with mobility of clients and configuration management that deals with mobility of resources. The question now is how to locate users and how to manage the resources related to the users efficiently. There has been very little work to date on comparing different locating and addressing schemes. This problem involves several dimensions, making it difficult to be solved. Basically the fundamental trade-off in this location management is between searching and informing. Generally, the less informed the sender is, the more search cost incurred. Thus, mobility also introduces the cost of search to the global cost analysis. Another problem where mobility issue is concerned is how to provide adequate response to a user who wanders a substantial distance from their information service servers. Thus, the system will have to adapt to changing spatial distribution of clients by having dynamic data replication and services.

The second issue in technical challenge is the limited wireless bandwidth resource that will have to be carefully managed. Providing a smooth interaction of mobile users with information servers over narrow bandwidth is one of many key challenges that need to be addressed. Severe bandwidth constraint will limit the number of simultaneous mobile users and the nature of application that can be run. The concept of bandwidth management plays a crucial role in design of future wireless information services, where mobile clients will query and update remote companies database located on fixed network over a wireless channel. There should be a way to maximize the client's query capacity by efficiently allocate and better utilization of the wireless bandwidth. This bandwidth limitation will eventually define new cost measures for accessing and updating data and consequently may favour new solutions in mobile wireless environment.

The third sub-category in this technical challenge concerns with the hardware device itself. The amount of work that one can perform while moving is fundamentally constraint by the limited energy supplied by the device's battery. Unfortunately, despite considerable effort to prolong the battery lifetime of mobile devices, no effective solution for energy management has been found yet. In order to minimize energy spent by clients, the system should have to minimize the number of transmission and the total time client spent listening to the communication channel. Also to save energy, clients may avoid transmissions and waking up from doze mode only when necessary. Flinn & Satyanarayanam (2004) proposed a method called energy-aware adaptation that dynamically balances energy conservation and application quality. Using this method, energy usage can be reduced without affecting the perceived quality of the system. Another limitation of mobile device is the screen size. As reported by EContent Magazine (Daily, 2005), the most obvious limitation of mobile devices is screen size. With the average cell phone screen not measuring more than two inches square and those of PDAs only marginally larger, documents are inevitably formatted for a different viewing area. Designing a menu-driven application, with the use of nested menus and the use of simple forms to display the information may help mobile developers to solve the limited screen size issues. However, for applications that need to enable spreadsheet, the screen form factor rarely coincides with the much larger design of most spreadsheets or documents, leaving mobile developers a dilemma in overcoming this shortcoming. Another part of the trouble with current mobile device screens, though, is the even-sided nature of their displays.

All these three issues gave several impacts to the quality of the services of mobile wireless computing. Mobility issue for example has changed the assumption that data is situated in a fixed host. Resources now need to be dynamically configured in response to the mobility of the mobile clients. This involves dynamic replication of data as well as services. Querying wireless data also creates new problem for mobile wireless computing. There are questions what is the best to represent and query the fast-changing, update intensive data and how this information could be delivered to the distributed clients. The next section will briefly discuss the non-technical challenges that mobile wireless computing is facing.

Non-technical Challenges

Non-technical challenges in mobile wireless computing can be in terms of social challenges and usability challenges. Social challenges of mobile experiences include personalization, comfort, user's acceptance, adoption and privacy concern. Being accessible at any location and at any time creates concern about privacy issue among potential mobile users. Some users may not like having the idea of people know what their location at any time of the day is. The ability to monitor of client's whereabouts creates an opportunity for fraud and impersonation of mobile users. To protect privacy, it is necessary to develop sophisticated software for specification and enforcement of personal profiles of users. Having mobile users that are scattered in many locations, there will also be an authentication and security problems. Thus, there should be distributed services that are able to support authentication across administration domains which can be used to provide authentication and facilitate secure communication.

There are several usability issues of mobile wireless computing that are important to application designers and developers. These include understanding the nature of and differences among applications that are suitable for mobile office context and those are not. If unsuitable applications are ported on the mobile device, it will prevent most users from using the application effectively in mobile settings environment. As to date, there are very few benchmarks for applications which are fully mobile wireless application and users are likely to be tolerant of their imperfection. The developers should aim to preserve the much look and feel of the original application that they want to port into mobile settings. If the same application differs significantly in the settings, the user-perceived usability and satisfaction of the users may suffer.

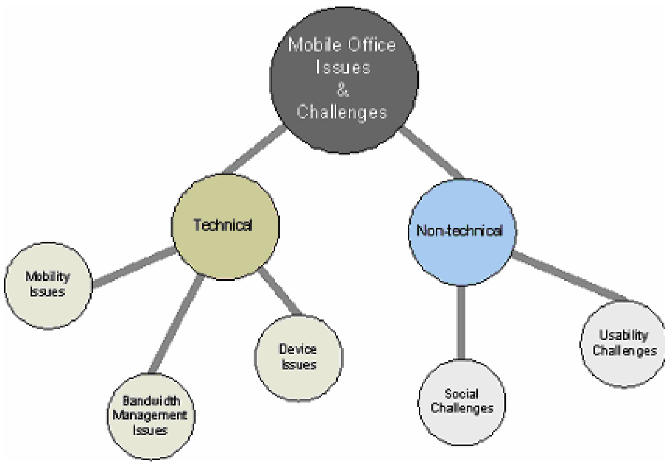


Figure 1: Issues and Challenges of Mobile Office

Issues and challenges of mobile wireless office applications could be summarized as illustrated by Figure 1 above.

MOBILEOFFICE© SCHEDULER

MobileOffice scheduler is a solution from TM Research & Development (TMR&D) towards realizing mobile office. The purpose of this system is to reduce the workload of a secretary in reminding his/her manager about his/her appointment schedule. It is like putting the manager's appointment schedule in his/her mobile device, without a need for the secretary to manually key in the information in his/her manager's devices to ensure the mobility of a manager. Hence, should an amendment required, secretary does not need to get back to his/her manager's devices to make the changes. Attachment of spreadsheets, document or presentation files is an advanced features expected to be developed soon.

Generally, the system consists of two modules, (i) MobileOffice portal and (ii) MobileOffice scheduler client. MobileOffice portal is a web-based portal system which is used by the information feeder regarding appointment schedule for a superior. In our case, we expect this system to be utilized by personal assistant or secretary. Once he/she input the appointment schedule for his/her manager, he/she may configure the system setting to locate when the schedules would be send to the manager. Differ from mobile office solution offered by Telstra many years back; our mobile office solution does not employ WAP as its main technology. We aim to reduce the dependency on the availability of the network. Another motivation is, we would like to reduce the needs for the manager to spend time accessing to the network to get his/her schedule when he/she could get the schedules ready for them without doing anything—just check to his/her MobileOffice scheduler application! For that reasons, we make use of SMS and MMS technology. The appointment schedule data is sent via SMS message to the manager's mobile device based on the configured setting. Once the manager's device got the SMS, it will be automatically processed by the MobileOffice scheduler client application which resides in the device. An alert message could also be configured if the manager wishes to be alerted. The alert time is based on his/her own preference, set via MobileOffice portal application.

Developing an SMS/MMS-based mobile office application is relevant since it is a convenient and low-cost mobile communication technology. Furthermore, text messaging is experiencing very rapid growth. Mobile Data Association (MDA), who commenced its official, ongoing, monitoring of SMS statistics since 11 August 1999 reported that since those early days, text messaging figures have gathered momentum year on year and July 2005 marks another major text messaging milestone (*Mobile Data Association*, 2005). Since then, there were more than 100 billion text messages in the UK. Text messaging growth for SMS in the UK can be seen in Figure 2 below.

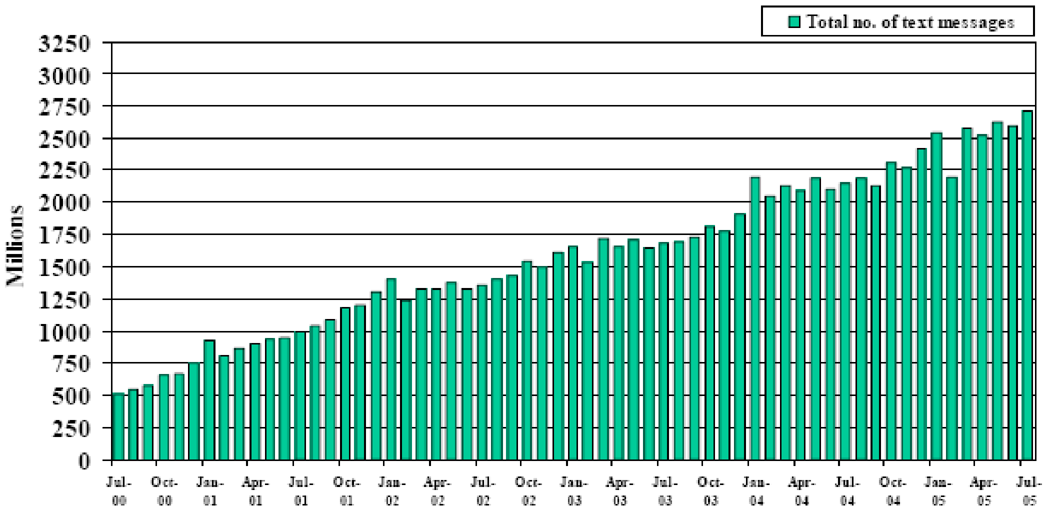


Figure 2: Text Messaging Growth (SMS) UK GSM Network Operator Totals (July 2000 – July 2005) (Source: *Mobile Data Association*, 2005)

Malaysia has also seen a phenomenal growth of SMS usage of late, with the total of SMS hitting 447 million per month in Quarter 1, 2003 (*mobileExec*, 2003). The total SMS sent in 2002 was 3, 605.9 million and in Quarter 1 2003, the total was 1, 433 million, with average 15.92 million SMS sent per day. Table 1 below shows Malaysia's mobile subscriber base and monthly SMS volumes in 2003.

Table 1: Malaysia's Mobile Subscriber Base & Monthly SMS Volume in 2003 (Source: mobileExec Sdn. Bhd., 2003)

Mobile Operators	Subscriber Base	Est. Monthly SMS Volume
Celcom	3.400 million	220 million
TMTOUCH	0.700 million	40 million
DiGi	1.950 million	70 million
Maxis	3.400 million	200 million
TOTAL	**9.45 million**	**530 million**

RESEARCH METHODOLOGY

We use throwaway prototyping methodology in building this system in order to reduce requirements risk. The prototype is developed from an initial specification, delivered for experiment then discarded. As the concept of throwaway prototype methodology, this prototype is not considered as a final system, since some system characteristics may have been left out. These include the file attachment and the alert functions for MobileOffice scheduler client application.

The process flow of throwaway prototyping is shown in Figure 3.

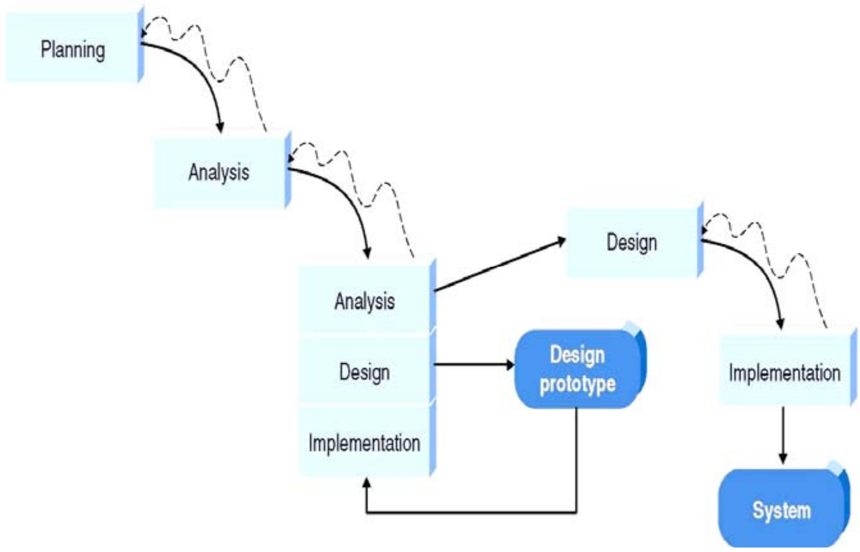


Figure 3: Throwaway Prototyping (Dennis, et al., 2004, p.13)

The throwaway prototyping-based methodologies have a relatively thorough analysis phase that is used to gather information and to develop ideas of the system concept. This type of prototyping is done right after the analysis phase of the software development lifecycle (SDLC). As throwaway-prototyping is a category of rapid application development (RAD), which is a combination of the changed SDLC phases, it is still possible to go backward from each phase, as shown by dotted lines in Figure 3 above. In our case, we will only throw away the MobileOffice scheduler client application, while MobileOffice portal is considered as a prototype and therefore will not be discarded. The details of MobileOffice scheduler client and MobileOffice portal is discovered in the next section.

SYSTEM DESIGN

MobileOffice portal was designed and developed using Active Server Pages (ASP), and we use VBScript for the scripting language. Since this application is meant for the information feeder, we divided the information feeder into three categories, namely, (i) personal assistant/secretary (ii) system administrator (iii) manager. These three categories carry different privileges and they are identified by different login ID. Personal assistant has the property of entering the appointment schedules for his/her manager and configure the setting of the scheduler SMS message. System administrator is responsible for services related to the system user, including creating, deleting and editing personal assistant-manager data. Manager as well could access to this system and he/she has similar roles and privileges as the personal assistant does.

The prototype MobileOffice scheduler client was developed using Java™ 2 Platform, Micro Edition (J2ME). The Mobile Information Device Profile (MIDP) platform, combined with Connected Limited Device Configuration (CLDC) is used for the Java™ runtime environment. The MIDP provides a mechanism for MIDP applications to persistently store data across multiple invocations. This persistent storage mechanism can be viewed as a simple record-oriented database model and is called the record management system (RMS). We use RMS to provide mechanism through which MIDlets (MIDP applications) can persistently store appointment schedule data and retrieve it later. In general, RMS provides the same functionality to J2ME application as what JDBC provides to J2SE applications i.e. persistent storage facility.

A few snapshots of the system are illustrated in the following figures. Figure 4 below shows the Appointment Manager page of MobileOffice portal application. The following figure shows MobileOffice scheduler client, which resides in the mobile device. Once MobileOffice scheduler icon is selected, it will display the list of appointment. Clicking on the topic will show the details of the appointment. Messages with attachment will be sent using Multimedia Messaging Service (MMS), whereby, the third image in Figure 5 shows the illustration. Currently, attachment scheduler message is not available in the prototype system.

Figure 4: Appointment Manager - Add New Appointment

Figure 5: Smart Client Scheduler client snapshots

CURRENT WORK

At this moment, we manage to finish developing the MobileOffice portal, and a prototype MobileOffice scheduler client. As stated in the previous topic, the scheduler does not have the attachment function yet, thus we can only view the list of appointments and observe the details of selected appointment. Since we are targeting for pen-based Symbian UIQ reference design version 3.0 types of mobile device's user interface, we plan to develop the real system using Symbian SDK Quartz 6.1. We also plan to port the application for Nokia Series 60 version soon in order to cater a bigger range of users. For that reason, we will discard the prototype MobileOffice scheduler client, which was developed using J2ME technology. The alert function for MobileOffice scheduler client has not been developed in the prototype version.

We expect the development of the system to be completed by middle of year 2006, at least for UIQ reference design.

CONCLUSION

To design an application that will be used in mobile environment, one must consider many factors in terms of technical and non-technical aspects. Questions like who will use the application, how it will be deployed, and what the interface should be like and application scalability should be the main factors that the developers are supposed to look into. The MobileOffice Scheduler has tried to answer these questions and it is a hope that this application can become another potential solution in mobile office solution.

REFERENCES

100 billion messages sent by UK's texting generation as texting hits all time high. (2005, August 23). *Mobile Data Association*. Retrieved September 1, 2005. From http://www.mda-mobiledata.org/MDA/documents/July05PressRelease.pdf

Aldred, J. (2000, Nov. 3). *WAP delivers mobile office*. Retrieved 7 Apr. 2005, from http://www.internetnews.com/bus-news/article.php/502641

Broadbeam Corporation. (2004). *Experience and expertise in mobility: Why does it matter?* Retrieved March, 7, 2005, from http://www.broadbeam.com/pdfs/broadbeam_experience_and_expertise_11252004.pdf

Churchill, E. F. & Munro, A. J. (2001). Work/place: Mobile technologies and arenas of activity. *ACM SIGGROUP Bulletin*. 22(3), 3-9. Retrieved July, 27, 2005 from ACM database.

Cisco System. (2003). *Implementing a Cisco mobile office solution: Implementation roadmap.* Retrieved March, 7, 2005 from http://www.cisco.com/application/pdf/en/us/guest/netsol/ns314/c714/cdccont_0900aecd80087a05.pdf

Daily, G. (2005, July, 27). Size matters: Optimizing mobile document delivery. *EContent: Digital Strategies & Resources*. Retrieved September, 2, 2005 from http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=8272

de Jode, M. (2002). *Porting applications from the Nokia Series 60 to the UIQ reference design*. Retrieved February, 3, 2005 from http://www.symbian.com/developer/techlib/papers/series60-UIQ/Series60-UIQ.pdf

Dennis, A., Wixom, B. H., & Tegarden, D. (2004). *Systems analysis and design with UML version 2.0: An object-oriented approach* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Flinn J., and Satyanarayanam M. (2004). Managing Battery Lifetime with Energy-Aware Adaptation. *ACM Transactions on Computer Systems*, 22, 137-179. Retrieved August 1, 2005 from ACM database.

Fui-Hoon Nah, F, Siau, K & Sheng, H. (2005). The value of mobile applications: A utility company study. *Communications of the ACM,* 48, 85-90. Retrieved August, 1, 2005 from ACM database.

Gorlenko, L., and Merrick, R. (2003). No wires attached: Usability challenges in the connected mobile world. *IBM Systems Journal*, 42, 639-651. Retrieved August 1, 2005 from ACM database.

Gupta, R. K. (2003, May, 12). *Intelligent appliances and J2ME's RMS.* Retrieved September 3, 2005 from http://java.ittoolbox.com/documents/document.asp?i=2755

Haaker, T., Faber, E. & Bouwman, H. (2004). Mobile services and technology track: Balancing strategic interests and technological requirements for mobile services. *Proceedings of the 6th international conference on Electronic commerce ICEC '04* (pp. 609-618). Retrieved August 5, 2005 from ACM database.

Imielinski T. & Badrinath B.R. (1994). Wireless computing. *Communications of the ACM*, 37(10), 18-28. Retrieved August 1, 2005 from ACM database.

Java 2 Platform, Micro Edition (J2ME). (n.d.). Retrieved September 3, 2005 from http://java.sun.com/j2me/index.jsp

Knudsen, J. (2003). *What's new in MIDP 2.0*. Retrieved September, 3, 2005 from http://developers.sun.com/techtopics/mobility/midp/articles/midp

mobileExec Sdn. Bhd. (2003). Enterprise Mobility Solutions. Retrieved February 2, 2005 from http://www.mobileexec.net/s_download/index.htm

Ortiz C. E., Giguere, E. (2001). *Mobile Information Device Profile for Java 2 Micro Edition (J2ME): Professional developer's guide*. New York: John Wiley & Sons.

Page, C. (2005). Designing for the mobile device: experiences, challenges, and methods: Mobile research strategies for a global market. *Communications of the ACM*, 48(7), 42-48. Retrieved August 5, 2005 from ACM database.

Sun Microsystem. (2001). MIDP APIs for wireless applications: A brief tour for software developers. Retrieved April 7, 2005 from http://java.sun.com/products/midp/midp-wirelessapps-wp.pdf

A mobile knowledge management expert system for automatically conducting a digital business

Dr. Wu Wen,
Lunghwa University of Science and Technology,
Taoyuan, Taiwan, R.O.C.
Email: wenwu@mail.lhu.edu.tw

MSc. Yuh-Huey Chen
Northern Taiwan Institute of Science and Technology,
Taipei, Taiwan, R.O.C.
Email: yhchen@mail.ntist.edu.tw

MSc. Hong-Shin Pao
Lunghwa University of Science and Technology,
Taoyuan, Taiwan, R.O.C.
Email: wo.ow@msa.hinet.net

Abstract

This paper presents a mobile knowledge management expert system by using multi-agent technology for automatically managing and controlling a digital company. The architecture consists of a user interface, an assignment control agent, a knowledge reasoning agent, a knowledge base, a database, and an inference engine. The inference engine selects appropriate rules from the knowledge base and uses facts in the database to reason suitable solutions which match the minimal price or better effectiveness via a foreword chaining reasoning method. Furthermore, to provide instant management and control, we have developed an automatic agent to manually or periodically monitor managerial problems using knowledge rules. In addition, the system makes use of ASP combined with JAVA to generate the functions so that it is platform independent. Furthermore, in order to allow users to easily manage the rule base and knowledge base, we have designed a user transparency approach to modify all vital rules, if necessary, without revising the program source codes. Users can modify, add or drop their latest knowledge rules from the system. When new knowledge is introduced, the system is able to automatically evaluate accurate actions or alternatives for conducting a digital firm.

Keywords:
Multiple agents, Knowledge bases, Inference engine, M-commerce, Rule reasoning

1. INTRODUCTION

The rapid growth of smart phones, personal digital assistants, and wireless networks during the last decade laid the foundation for mobile commerce (m-commerce). Some modern cities in the world are planning to build a wireless network to cover the whole city area so that citizens can use a notebook or personal digital assistant to access information from the Internet via a wireless network. Taipei and Kaohsiung, Republic of China, have already started the program and it will finish by the end of 2005. The supported protocols of standard are 802.11a, 802.11b, and 802.11g. The bandwidth of speed can reach 11-54 Mbps. Students and citizens in schools, administrative offices, and main streets are able to access the wireless network. Particularly, more and more digital learning and content programs subsidized by National Science Council, R.O.C., are executing and planning.

This research is motivated by several reasons. First, globalization and information technology have altered business management and competitive styles. Nowadays, many companies need to manage and control their organizations in a global marketplace. Most of them require facing global competition. Sometimes, they might need to organize a global work team or adopt a global delivery system. Conventionally, a global company will ask their subordinate organizations scattered across different countries to report their working performance or progress via the Internet. Next, they often require holding a meeting once per month or per season to gather all representatives from their oversea branches for discussing all the problems they face and give some instructions to ask their branches to follow. It looks very effective to employ Internet-based technology for transferring information but the procedure still takes tedious time and costs a lot if there is no Internet-based information system. Second, many businesses have already built enterprise systems, customer relationship systems, or supply chain management systems such that all the activities of the companies are under control. However, although these systems generate a large amount of important information, they lack the application of integration from management point of view. Finally, even in a local business, a top level manager always holds a meeting per week or per month in order to communicate or give instructions to his subordinators for running the company. In the meeting, there is nothing new except reporting the working progress or performance done in the past week or month.

Actually, all data can be found in the database produced by ERP, SCM, CRM, or other information systems. Why not to build an expert system to integrate all critical information and automatically inferring a suitable action or suggestion for the decision maker?

Accurate business performance measurement is critical to understanding business success and failure. Moreover, the increased attention to business performance evaluation by managers, consultants and academics reflect much pressure that more and more companies dedicate to improve their performance. Even if the assessment of the performance of a firm is a challenge and composite task, many organizations still do lots of effort for enhancing effectiveness for survival. Traditionally, through e-commerce, manufacturers are able to sell their products and offer services directly to customers bypassing intermediaries such as distributors and retailers. Erasing intermediaries in the distribution channel can reduce transaction costs. Similarly, if we make use of an automatic multi-agent system to help professional managers to trace and control a digital firm, top-level managers can get instant decision-making information via a wireless network. This system can save tedious time and huge money for gathering and analyzing all data, which are collected and integrated from various databases.

Normally, there are many views and thoughts in the evaluating of enterprise performance. Milee (1978) and Snow (1994) asserted that improved business performance needs an organizational structure, information systems, and management style related to a specific-firm strategy. Ford and Schellenberg (1982) summarized three basic frameworks frequently used to form business performance. The goal approach seeks a definition based on explicit goals or goals that can be implied from the behaviour of business's members. The system resource approach is to assess business performance according to the key internal and external factors on which the company depends for survival. The constituency approach provides a wide range of 'constituencies' with business performance assessment to satisfy constituent needs. Generally, two common categories of performance are used to measure: financial and non-financial performance. Venkatraman and Ramanujam (1986) stated that there were non-financial performance indicators (operational performance) besides financial indicators in the evaluation of performance. For examples, market share, product quality, R&D capability, manufacture efficiency, and customer satisfaction etc. Moreover, Murphy et al. (1996) surveyed 124 articles and classified as financial and operational performance. Then, performance measures were split into 8 dimensions: efficiency, growth, profit, size, liquidity, success/failure, market share, and leverage. Each dimension has its own measures and a total of 71 different measures of performance were collected. Finally they summed the total number of articles in which each dimension and measure were considered. They found that efficiency, growth, size, and profit were the widely used dimensions to evaluate performance. In the survey, 60 percent of their studies is only used one or two dimensions. No study in the research uses more than five of the eight dimensions. Some studies show that 19 percent of 52 articles adopt one measure and 71 percent use four or fewer measures. Furthermore, other pioneering research related to business performance has been found in Refs. (Dess & Robinson, 1984;Dwyer, 2003;Gilley et al., 2004; Hoque, 2004; Maiga & Jacobs, 2004; Murphy et al.,1996; Omero et al., 2005; Yamin et al., 1999) and we can roughly classified it into two domains as the following (see Table 1):

1. Financial

The financial measures primarily contain efficiency (e.g., return on investment, return on equity, return on assets, return on net worth, etc.), growth (e.g., change in sales), profit (e.g., return on sales, net profit margin, gross profit margin, net profit level, etc.), liquidity (e.g., sales level, cash flow level, current ratio, quick ratio, etc.), leverage, and others. Other measures include long-term capital/fixed assets ratio, circulating ratio in the debt payment aspect, receivable account velocity, duration of account receivable, stock velocity, fixed assets velocity and total assets velocity in the operating capability aspect, sales growth, profitability (e.g., return on investment, return on sales, return on equity, etc.), earnings per share, overall financial performance, and so forth. The enterprise performance can be analyzed and evaluated by financial data for the reference of decision makers.

2. Non-financial

Non-financial measures mainly contain innovation and stakeholder performance. The measures of innovation performance are composed of R&D outlays, product innovations, and process innovations. The stakeholder performance involves employment growth/stability, employee morale, customer relations, and supplier relations. Certainly, some others will also include in the non-financial performance such as new product introduction, product quality, marketing effectiveness, and manufacturing value-added.

Table 1. Measures for financial and non-financial performance.

Financial	Financial Performance	Return on Assets
		Return on Sales
		Overall Financial Performance
Non-financial	Innovation Performance	R&D Outlays
		Process Innovations
		Product Innovations
	Stakeholder Performance	Employment Growth/Stability
		Employee Morale
		Customer Relations
		Supplier Relation

The rest of pioneering research about non-financial performance is giving as the following. Yamin et al. (1999) devoted themselves to the relationships among generic strategies, competitive advantage and organizational performances under different situations. They suggested that there were significant differences when different generic strategies and types were adopted. Using a path analytical model, Hoque (2004) surveyed and discussed the impact on performance of two factors: strategy and environmental uncertainty from 52 samples of manufacturing. The results showed that management's strategy choice was positively related to performance remarkably. But there was no evidence to prove the relationship between environmental uncertainty and performance. Maiga (2004) examined the relationship between enterprise's benchmarking and enterprise performance. He found three elements in the benchmarking which affected the enterprise performance positively, including prior experience with benchmarking, the commitment of the organization to benchmarking, and internal preliminary competence analysis. They are helpful to the improvement of enterprise performance. Also, many of researches focused on the impacts of strategies and competition predominance on enterprise performance. Lebas (1995) paid attention to the relationship between performance management and measurement, and found that they were closely related. The result of performance measurement could be referenced by performance management, and performance management supported performance measurement correspondingly. Common criteria of evaluating enterprise performance include (1) employment creation, (2) social goods, (3) security of employment for the firm's personnel, (4) providing a satisfying return to corporate headquarters, (5) innovativeness in processes and product, (6) customer satisfaction, (7) growth of market share, (8) environmental contribution, and (9) technological leading edge. The enterprise performance covers a wide range of topics. Many scholars also state that the outsourcing affects the enterprise performance. Murphy et al. (1996) asserted that accurate performance evaluation was a key to the success for enterprises. His research can be summarized into two parts: 1) more than 60% of the researches evaluate the enterprise performance by one or two aspects and generally there is no appropriate choice. 2) They found that if the performance of an aspect increased, the performance of another was going to decrease. For example, if the enterprise wanted to broaden it business, a fairly amount of capitals were needed. Therefore, the enterprise's efficiency and cash flow would drop. In the research by Gilley et al. (2004), they found by statistics that the outsourcing of training and compensation might affect the enterprise performance.

About multi-agent technology, lots of work has been done on this research area. To provide better value-added services to buyers and sellers, Kong et al. (2004) developed an application of E-commerce in construction material procurement. They made use of mobile agents and web services to let information sharing between E-commerce systems. Through an E-Union server, communication and information between E-commerce systems can be facilitated. Primarily, a client registered in E-Union first sends a HTTP/HTML request, which stores material searching criteria, to the E-Union server. The application server in the E-Union server then explains its meaning and passes to the SQL server for finding E-Union members that have the specified type of material. Next, it sends a SOAP request to adequate sites. Upon receiving the SOAP request, each appropriate site sends a SQLXML request to the database server and gets the response in a SOAP message to the application server. The E-Union application server receives all the responses from all other appropriate sites and integrates all the responses and then sends them back to the client in HTTP/HTML format. Shakshuki et al. (2000) proposed a 3-tier multi-agent framework for assisting different users to locate, retrieve, and integrate information from the WWW. The system is comprised of a user agent, a broke agent, and a resource agent. The user agent allows the user to interact with the system and fulfill users and other user agent requests. It accepts user's topic of interest and the associated constraints. The broke agent plays an intermediary between the user agent and the resource agent. Using KQML, the broke agent identifies and matches agents based on their goals as well as interest. The resource agent accepts a query structure similar to that of the user agent. After identifying the information resources, downloading the relevant documents, the resource agent parses the document, extracts the topics and their qualities. Lee et al. (2001) designed a multi-agent system for internet information retrieval services. In order to get effective and correct results, they designed the existing retrieval agent with effective distribution and multi-processing. The integrated system consists of many agents with different functions. Each agent includes a facilitator and some small agents. The facilitator acts as communicator among related agent groups and manages its own agent group. Many other multi-agent techniques have been found in Refs. (Berry and Linoff, 1997; Bigus and Bigus, 1998; Liang and Huang, 2000)

Before we introduce knowledge management, an important area related to the applications of decision support systems shall be touched in the following. Wen et al. (2005) proposed a decision support system based on an integrated knowledge base for acquisition and merger. It not only provides merger processes, major problems and related regulations practically or procedurally, but also gives rational suggestions according to related regulations. Finally, it can suggest the user how to deal with uncertain growth rate and current evaluations. Zopunidis (1997) also developed a knowledge-based decision support system for financial management. This system integrated the technologies of decision support systems and professional systems, and was used to tackle past and current problems occurred frequently. Besides, decision support systems are also employed in the HR planning and decision, financial management and marketing decision makings. Other pioneering research to discuss DSS and knowledge management techniques can be found in Refs. (Li, 2000； Bharati & Chaudhury, 2004；Deshmukh, 1997；Eriksson, 1996;Liao, 2001；Masod & Soo, 2002；Matsatsinis et al., 1997; Mohanty &).

In this paper, we propose a multi-agent framework for measuring and managing business performance for the electronics industry in Taiwan. The remainder of the paper is organized as follows. Section 2 presents the multi-agent expert system for conducting an electronics enterprise. Then the functions and processes of the database and knowledge base, knowledge presentation and rule-based reasoning are explained respectively in Subsections 2.1, 2.2, and 2.3. Section 3 demonstrates the system's functions and evaluates its benefits through system implementation. Finally, in Section 4, some suggestions for future directions are also given in this paper.

2. THE MULTI-AGENT EXPERT SYSTEM FOR CONDUCTING AN ELECTRONICS ENTERPRISE

We have developed a multi-agent expert system intended for automatically monitor and control a company for proving instant suggestions to top level managers. The system integrates a smart phone or personal digital assistant with a build-in wireless card, a wireless network, which supports 802.11a, 802.11b, and 802.11g, and a high speed server, as shown in Figure 1. The server must have a database managed by DBMS, a knowledge base to store all known expert rules, a reasoning agent for knowledge inference, and a control agent for trigging and connecting with the reasoning agent.

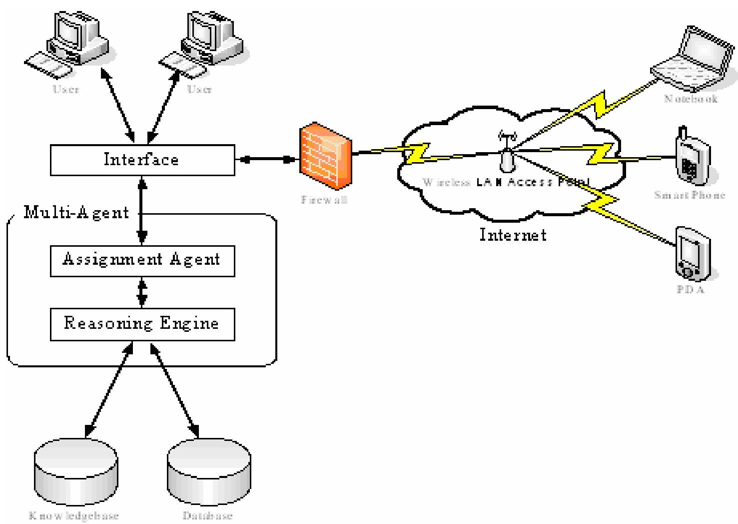


Figure 1. A structure of the mobile knowledge management expert system for managing and controlling enterprise performance.

2.1The database

Database technology acts as an important role in making organizations' information resources available on the Internet. A more common definition of a database is a collection of data designed to serve many applications efficiently by storing only one location and minimizing redundant data; whereas a database management system (DBMS) is the software that allows an information system to centralize data, manage them efficiently, and provide easy retrieval. Currently, most DBMSs provide heterogeneous database retrieval (e.g., access different databases such as Oracle, Sybase, Informix, SQL Server, etc.) by using an ODBC/JDBC interface. Relational databases management systems are the most popular development tools for building business information systems because their superior integration, distribution, and easy-of-use. The database consists of six tables, including Password, Production, Inventory, Finance, Sales, and Personnel. The main function of the Password table is to verify users' identification. Thus, the Password table is composed of four attributes: *id, password, unit*, and *management-level*. The Production table provides diverse descriptions and functions for production management and planning. The attributes of Production are *production name, production quantity, Produce-date, delivery-date,* etc. Inventory provides the number of stored products, parts, or materials. The attributes of the Inventory table consist of item-*no, item-name, quantity, cost, purchase-date, and valid-date*. The attributes of the Finance table are mainly item-name, cost, price, quantity, *year,* and *use-level.* The Personnel table is used to store personnel data including the attributes such as *code, name, employed-date, title, award, address, telephone,* and so on. Finally, the Sales table contains the key items needed to maintain the sales performance. The attributes of the table include *code, item-name, sales-date, delivery-date, quantity, delivery-address*, contact-tel, etc.

2.2 The knowledge base

Generally, a knowledge base involves two distinct bases. One is rule bases and the other is case bases. In this paper, the knowledge base only contains a rule base. In the rule base, the specialized domain knowledge is presented in the form of *if <antecedent clauses> then <consequent clauses>* statements. If the antecedent clauses are true, then the consequent clauses are true. The rules in the rule base are triggered if their antecedents match. Here, knowledge engineers must elicit knowledge rules based on professional experts (e.g., senior managers or top-level managers, etc.) experiences and judgments on routine working processes. All rules must be coded in IF-THEN-ELSE statements for execution on computers and stored in the knowledge base. The inference engine needs to select a suitable rule from the rule base. Then, the engine gets the facts from the database to infer a conclusion. Upon receiving the known facts from the database, the engine needs to infer and explain the situation and trigger actions or alternatives for decision makers. Furthermore, in order to enable users to easily manage the rule base, and knowledge management, we have designed a user transparency approach that can be used to modify all vital information, if necessary, without revising the program source codes. Users can modify, add or drop into the system their latest knowledge. Once the new knowledge has been input, the system is able to automatically diagnose all vital situations of which a company may face by using a controlling and tracing agent.

2.3 Knowledge presentation and rule-based reasoning

Knowledge presentation plays an important role in knowledge reasoning. A well designed knowledge presentation will affect the performance of an information system. In the past decades, many schemes of knowledge-based expert systems, for example production rules, predicate logic, semantic networks, frames, scripts, and decision trees, have been developed. However, production rules are most commonly used in knowledge presentation. Generally, knowledge-based expert (or decision support) systems must have the following characteristics:

1. Representational adequacy: this is the ability to professionally describe all knowledge and to compact with all knowledge in a knowledge base.
2. Inferential adequacy: it can inference new rules from some given rules and easily build a new structure.
3. Inferential efficiency: this is the ability to efficiently reason, quickly execute and get conclusions.
4. Acquisitioned efficiency: knowledge should effectively be accessed.

As we stated in Section 2.2, rules are presented as a set of *if <antecedent clauses> then <consequent clauses>* rules. Basically, there are two types of reasoning: backward chaining and forward chaining. Forward chaining is a data-driven approach; it starts from a basic idea and then tries to draw conclusions. It first checks the IF part of IF-THEN rules to find out whether the antecedent clauses of the rule is matched. As each rule is tested, the program can inference one or more conclusions. Conversely, backward chaining is a goal-driven method. It starts with a goal to be verified as either true or false. It then examines all the THEN parts of IF-THEN rules. A rule that contains this goal in its consequent clause will be checked to know whether its antecedent clause is true. If fails, the program searches for another rule whose conclusion is the same as that of the first one. This process continues until all possibilities have been tested. Our approach makes use of forward chaining.

Figure 2 shows an example of hierarchal structure for rule presentation. For convenient maintenance and effective inference, we organize the knowledge for business manufacturing management into 5 groups: production, inventory, sales, finance, and personnel. In order to have more space to accumulate the appending rules, the starting number begins from 100, 200, 300, 400, and 500, respectively. To gather correct and effective professional knowledge, 120 rules are collected by interviewing experts and senior managers, who work for Taiwan main electronics manufacturing companies. To clarify the forward chaining procedures, several example rules in the knowledge base for business management are shown below:

Rule 100: if Condition="production" then goes to Rule 101
Rule 101: if Quantity_per_month < Month_goal then Situation= "Warning"
Rule 102: if Accumulated_quantity < 30% Total_goal then Situation = "Ask production manager to explain the reason"
Rule 103: if Defector_rate□ Average_defector_rate then Method="Ask production manager to report"
Rule 104: if Reason="Bad skills" then Method="Improve training"
Rule 105: if Reason="Low capacity" then Method="Overtime"
Rule 106: if Situation="Warning" then Method="Keep monitor"
Rule 200: if Condition="Inventory" then go to Rule 201
Rule 201: if Condition="Stop production because of leak of materials" then Action="Emergency purchase" and show "Warning"
Rule 202: if Inventory_Level > Average_Inventory_Level then Method="Reduce Inventory_Level"
Rule 203: if Method="Reduction_Inventory" then Action="30% off discount to the parts"

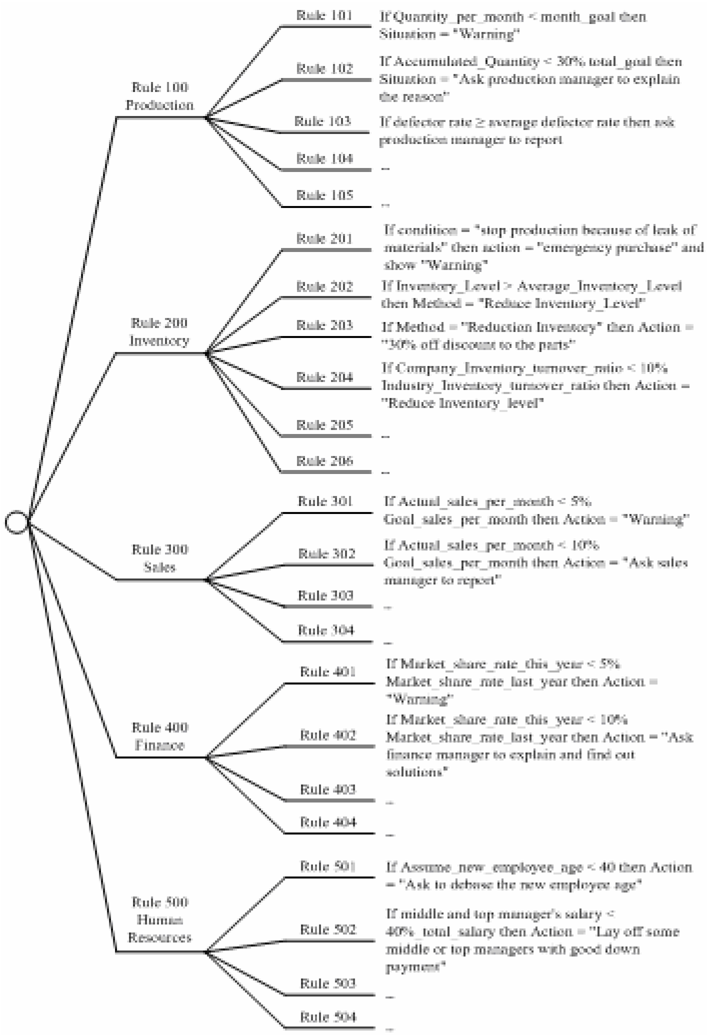


Figure 2. A hierarchal structure of management rules.

Now here is an example involving a production decision by using the forward chaining procedures. In forward chaining, we start with known facts and acquire new facts using rules that have known facts on their IF sides. When the condition in the IF part is satisfied (is true), then the rule will be fired and the conclusion in the Then part will be added to the assertion base. Assume that condition is equal to production and the goal of this month is to produce 150 notebooks. The production quantity of this month is 120 notebooks.

Step 1: Because condition is equal to production, according to Rule 100, the inference engine points to Rule 101.

Step 2: Next, the inference engine selects rule 101 and gets the facts from the database to reason.

Step 3: Now, Rule 101 has been fired since the value of quantity is less than the goal for this month.

Step 4: Using Rule 106, then the inference engine keeps to monitor the progress of production because situation is equal to warning,

During the reasoning period, when the middle or top manager makes some adjustments to or reduces company's goal, the system can use the forward chaining method shown above to automatically infer again without revising the program. Therefore, the mechanism is able to easily and quickly respond the changes in the rule base. To easily manage and effectively trace various business activities, the system allows users to make use of automatic and semi-automatic reasoning, which needs managers to choose the reasoning rules.

3. SYSTEM IMPLEMENTATION

This section describes the implementation of the mobile knowledge management expert system (MKMES). The ASP.net language is used to implement the system. However, there are no limitations for the kinds of expert systems which can be embedded in the agents. All agents are coded in Java and can provide a connection with relational databases via Microsoft's

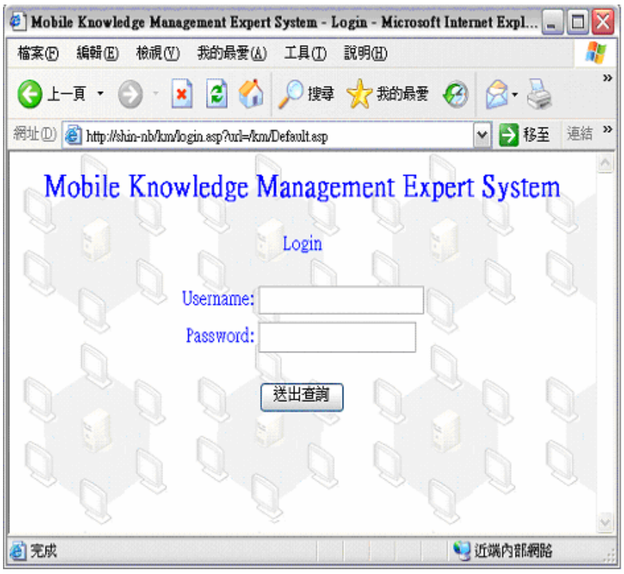


Figure 3. The login screen of the MKMES system for notebooks. Figure 4. The login screen of the MKMES system for smart phones.

Open Database Connectivity interfaces (ODBC). Because users may use a smart phone, personal digital assistant, or notebook to access the expert system, we have developed 3 types of Web pages to suitable different devices (see Figure3-Figure 5). Users do not need to worry about what device they use. The system is running on a server machine in the backend. Via a wireless network, a PDA in the frontend is able to automatically detect what device the user uses and directly link a suitable format to show the screen for the user. However, to simplify the contents of this paper, we only present all figures in a PDA form.

3.1 System functions and security

To assure security, we design a login menu to require users to fill in their Username and Password as shown in Figure 5. After submitting, the server in the backend needs to check whether the username and password are legal to login the system. Without permission, the user will be denied by the system. Besides this mechanism, a wireless access point (WAP) also allows the system administrator to set more than one secure key. Through the double check mechanisms, the MKMES system can act as a safe guard to protect it from hackers. The functions of the MKMES consist of knowledge reasoning, financial measure query, knowledge update, and system administration (see Figure 6).

3.2 System financial queries and automatic reasoning results

The Knowledge Reasoning icon can execute periodically automatic or semi-automatic inference by using an inference engine. During the inference period, the engine requires to select adequate rules in the rule base (i.e., knowledge base) and facts in the database to reason a conclusion as shown in Figure 7. Consequently, according to the conclusion of the reasoning, the system can provide some suggestions or actions to decision makers. Next, when users click the Financial Measure Query icon, they are able to inspect all vital financial data (see Figure 8). It is important to note that all rules in the rule base can be updated, deleted, or added through the Knowledge Update icon. Furthermore, in order to understand company's position in the electronics industry, the financial data of 257 Taiwan electronics firms are collected and analyzed. These data will be compared with the company's individual performance during the reasoning. Finally, the system administrator can use the System Administration icon to maintain various system functions including privileges, security, and financial measures. In this paper, knowledge reasoning has two major ways. The first one is an automatic reasoning and the second

Figure 5. The login screen of the MKMES system for PDAs. Figure 6. The menu of the mobile knowledge management expert system.

one is a semi-automatic reasoning. The automatic reasoning mechanism makes use of an automatic agent to access computer system time to commence a reasoning process; the system administrator can set a time interval to automatically infer knowledge. For the semi-automatic reasoning mechanism, a decision maker is able to select which knowledge group he would like to reason.

Figure 9 presents a modified rule accessing from the rule base. Any rule in the rule base can be modified, deleted, or added without affecting the program. Finally, managers can get the vital suggestions or actions via the system automatic or semi-automatic reasoning as shown in Figure 10.

Figure 7. A screen menu for knowledge reasoning. Figure 8. Results for querying financial measures.

Figure 9. The screen for knowledge update.

Figure 10. Results for automatic reasoning

4. CONCLUSIONS AND FUTURE WORK

Mobile knowledge management expert systems can help managers fast and effectively make decisions based on automatic or semi-automatic reasoning and various vital financial measure queries.

This paper clearly illustrates a framework for mobile knowledge management expert systems. Then we introduce knowledge presentation and rule-based reasoning, and finally system implementation is given to prove the possibility and effectiveness of the framework. It is worth to mention that we have designed a user transparency approach to modify all vital rules, if necessary, without revising the program source codes. Thus, the system allows users to easily manage the rule base (i.e., the knowledge base). Users can modify, add or drop their latest knowledge rules in the knowledge base. When new knowledge is introduced, the system is able to automatically reason accurate actions or alternatives for conducting a digital firm. Finally, the MKMES system has been implemented to inspect its functions. For the future work, we list the following two directions:

1. Build a machine learning mechanism to enhance system knowledge reasoning under incomplete data.
2. Adopt a neural network model to do some financial measure forecasting to make more easily manage and
control the future.

REFERENCES

Berry, M. J. A. & Linoff, G., 1997. Data Mining Technique for Marketing, Sale, and Customer Support, Wiley Co.mputer

Bharati, P., & Chaudhury, A., 2004. An empirical investigation of decision-making satisfaction in web-based decision support systems, Decision Support Systems, 37: 187-197.

Bigus, J.P., Bigus, J., 1998. Constructing Intelligent Agents with Java, WILEY.

Dess, G. G., & Robinson, R. b., 1984. Measuring Organizational Performance in the Absence of Objective Measures: The Case of the Privately-held Firm and Conglomerate Business Unit, Strategic Management Jouranl, 5: 265-273.

Dwyer, S., Richard, O. C., & Chadwick K., 2003. Genderdiversity in management and firm performance: the influence of growth orientation and organizational culture, Journal of Business Research, 56: 1009-1019.

Eriksson, H., (1996). Expert systems as knowledge servers. IEEE, Expert Systems and the Web, 14-18.

Ford, J.D. and Schellenberg. 1982. Conceptual issues of linkage in the assessment of organizational performance, Academy of Management Review, Jan., 49-58.

Gilley, K. M., Greer, C. R., & Rasheed, A. A., 2004. Human resource outsourcing and organizational performanc in manufacturing firms, Journal of Business Research, 57: 232-240.

Hoque, Z., 2004. A contingency model of the association between strategy, environmental uncertainty and performance measurement: impact on organizational performance, International Business Review, 13: 485-502.

Kong, S.C.W., Li, H., Hung, T.P.L., Shi, J.W.Z., Castro-Lacouture, D., & Skibniewski, (2004). Enabling information sharing between e-commerce systems for construction material procurement, Automation in Construction, 13, 261-276.

Lebas, M. J., 1995. Performance measurement and performance management, Internatinal Journal of Production Economics, 41: 23-35.

Lee, B.H., Park, M.S., & Kwon, H.C., 2001. Design and Implementation of Multi-Agent Environment for Internet Information Retrieval Service, IEEE, 318-321.

Li, S., 2000. The development of a hybrid intelligent system for developing marketing strategy, Expert Systems with Applications, 27: 395-409.

Liao, S. H., 2001. A knowledge-based architecture for implementing military geographical intelligence system on Intranet, Expert Systems with Applications, 20: 313-324.

Liang, T.P. & Huang, J.S., (2000). A framework for applying intelligent agents to support electronic trading. Decision Support Systems, 28, 305-317.

Maiga, A. S. & Jacobs, F. A., 2004. The Association Between Benchmarking and Organizational Performance: An Empirical Investigation, Managerial Finance, 30: 13-33.

Masod, S. H., & Soo, A., 2002. A rule based expert system for rapid prototyping system selection, Robotics and Computer Integrated Manufacturing, 18: 267-274.

Matsatsinis, N. F., Doumpos, M., & Zopounidis, 1997. Knowledge Acquisition and Representation for Expert Systems in the Field of Financial Analysis, Expert Systems with Applications, 12: 247-262.

Miles, R.E. & Snow, C.C., 1978. Organizational strategy, structure and process. New York: McGraw Hil.l

Miles, R.E. & Snow, C.C., 1994. Fit, failure and the hall of frame. New York: Free Press.

Mohanty, R. P., & Deshmukh, S. G., 1997. Evolution of an expert system for human resource planning in a petroleum company, Production Economics, 51: 251-261.

Murphy, G.B., Trailer, J.W., & Hill, R.C., 1996. Measuring performance in entrepreneurship research, Journal of Business Research, 36:15-23.
Omero, M., D'Ambrosio, L., Pesenti, R., & Ukovich, W., 2005. Multiple-attribute expert system based on fuzzy logic for performance assessment, European Journal of Operational Reseach, 160: 710-725.
Shakshuki, E., Ghenniwa, H. and Kamel, M., 2000. A Multi-Agent System Architecture for Information Gathering, IEEE, 732-736.
Venkatraman, N. & Ramanujam,, 1986. Measurement of business performancein strategy research：A comparison of approach, Academy of ManagementReview, 4: 801-814.
Wen, W., Wang, W. K., & Wang, T. H., 2005. A Hybrid Knowledge-based Decision Support System for Enterprise Mergers and Acquisitions, Expert Systems with Applications, 28: 569-582.
Yamin, S., Gunasekaran, A., & Mavondo, F. T., 1999. Relationship between generic strategies, competitive adavantage and organizational performance: an empirical analysis, Technovation, 19: 507-518.
Zopounidis, C., Doumpos, M., & Matsatsinis, N. F., 1997. On the use of knowledge-based decision support systems in financial management: A survey, Decision Support Systems, 20: 25

Developing a Brand Performance Measure for Internet Brands

Dr. Hsiu-Yuan Tsao
Ming Hsin University of Science and Technology
Hsin-Chu, Taiwan
Email: jodytsao@gmail.com

Prof. B. Ramaseshan
Curtin University of Technology
Perth, Australia
ramaseshanb@cbs.curtin.edu.au

Prof. Koong H.-C. Lin
Tainan National University of the Arts
Tainan, Taiwan
koong@mail.tnnua.edu.tw

Abstract

By using the Dirichlet model and brand performance measures, we have developed eBPMs (electronic brand performance measures) to assess Internet brand performance and to demonstrate how to use existing cyber panel data from web traffic measures to obtain the observed eBPMs. We present several potential applications of eBPMs for researchers and marketing practitioners. eBPMs may be capable of functioning as systemic web traffic measures to evaluate the market value of a web site, brand performance, and the effectiveness of web marketing activities, incorporating financial information, competitors' information, and online marketing activities.

Keywords:

Dirichlet model, Brand performance measures, Web traffic measures, eBPMs, Cyber panel data

INTRODUCTION

In an offline marketplace, brand performance is evaluated by measures such as how many customers buy the brand, how often they do so, and how much they buy other brands; these are known as "brand performance measures" (BPMs), an aggregate measure of brand loyalty, and are routinely reported by consumer panel operators (Ehrenberg et al., 2004). Over the past 20 years, empirical consumer behavior research using BPMs based on well-established Dirichlet models of repeat buying behavior has broadened to analyze brand repeat buying of fast-moving consumer goods (FMCG) in retail markets. The Dirichlet models have been utilized to benchmark loyalty patterns and to predict repeat choices (Bhattacharya, 1997; Danaher et al., 2003; Fader and Schmittlein, 1993; Stern, 2004). The approach to brand equity can provide a snapshot of a consumer's attitude towards a brand from a psychological perspective, and the data can be easily gathered at the individual level (Keller, 1993). However, the approach to brand performance is outcome oriented, and the data are collected from the aggregate measures such as brand purchase frequency, penetration, market share, and across-brand data.

In an online marketplace, some traditional consumer panel operators routinely compile web traffic measures using data from cyber consumer panels. This involves the presentation of research based on the existing web traffic measures—such as "the numbers of unique visitors" and "visits per visitor", and "visit duration"—to explore marketing insights and online consumer behavior (Bucklin et al., 2003; Park and Fader, 2004; Moe and Fader, 2004). However, research using web traffic measures to evaluate the performance of web sites from a marketing perspective is still rare; this is particularly the case with respect to content providers that offer a variety of functional services to attract targeted consumers for repeat visits to increase profits or to conduct brand building (Ilfeld and Winer, 2002).

The objective of this study was to: (1) develop eBPMs (electronic brand performance measures) to evaluate brand performance online; (2) demonstrate how to use existing cyber panel data from web traffic measures to obtain the observed eBPMs; and (3) propose research and industrial applications of eBMPs.

THE DIRICHLET MODEL AND BRAND PERFORMANCE MEASURES

The negative binomial distribution-Dirichlet ("NBD-Dirichlet" or, in the short form, "Dirichlet") is a descriptive model of buyers' repeat buying behavior. The Dirichlet model is an empirical generalization that describes certain patterns and aspects of aggregate behavior of buyers of competitive brands. The model uses a particular stochastic mathematical distribution of buyers' purchase probabilities—such as the Poisson, gamma, multinomial and beta binomial—to predict repeat buying behavior with respect to variables such as the timing of each purchase, the percentage of heavy buyers, and choice of brand (Goodhardt et al., 1984). Dirichlet modeling is increasingly used to provide norms against which brand performance can be interpreted (Ehrenberg et al., 2000). Marketing practitioners evaluate the performance of a brand on the basis of how many customers buy the brand, how often they do so, and how much they buy other brands. Well-known consumer panel operators, such as ACNielsen and IRI, routinely report BPMs to their clients. Ehrenberg et al. (2004) use aggregate data of loyalty—such as penetration, purchase frequency, and market share—to accurately predict repeat purchasing behavior and to interpret the BPMs. Table 1 sets out the definitions of BPMs used by that study.

Measures	BPMs (Ehrenberg *et al.,* 2004)
Brand market share (m)	number of purchases of a brand as a percentage of number of purchases of its category
Brand penetration (b)	percentage of potential buyers who have bought the brand at least once
Category penetration (B)	percentage of actual buyers who have bought the category at least once
Category purchase frequency (W)	number of purchases of the category per buyer
Brand purchases per buyer (w)	number of purchases of the brand per buyer
Category purchases per brand buyer (c)	number of purchases of any brands in the category for per brand buyer
SCR (share of category requirements) (w/c)	for those brand buyers, the percentage of purchases of the brand as a percentage of category purchases.

Table 1. Definitions of BMPs

ELECTRONIC BRAND PERFORMANCE MEASURES

Although the performance of a brand can be evaluated by how many customers buy the brand, how often, and how much they buy other brands in the offline marketplace (Ehrenberg et al., 2004), there is little research on to how to assess Internet brand performance. However, most research has recognized that web traffic plays a crucial role in the profitability and performance of web sites for those content providers (Riel et al., 2001; Drèze and Zufryden, 2004; Ilfeld and Winer, 2002; Serrano-Cinca et al., 2005; Wilson, 2004). Therefore, relying on the data obtained from web traffic measures, we argue that it is reasonable to evaluate the performance of a online brand on the basis of how many web surfers visit the web, how often they do so, and how much they visit other sites. In addition, BPMs had been adopted by marketing practitioners and researchers to assess the performance of a brand. Hence, this study evaluated Internet brand performance by using a number of loyalty-related measures based on Dirichlet and BPMs. We transformed these loyalty-related measures of BPMs into eBPMs, using "visits" in eBPMs instead of "purchases/buying" in BPMs, and "web surfer" in eBPMs instead of "buyer" in BPMs (Table 2).

Measures	eBPMs (defined by this study)
Brand market share (m^e)	number of visits to the brand site as a percentage of the number of visits to category sites
Brand penetration (b^e)	number of web surfers who have visited the brand site as a percentage of the number of web surfers who have visited category sites
Category penetration (B^e)	number of web surfers who have visited category sites as a percentage of all universe web surfers
Category purchase frequency (W^e)	number of visits to category sites
Brand purchases per buyer (w^e)	number of visits to the brand site
Category purchases per brand buyer (c^e)	number of visits to any brand site in the same category by per brand site surfer

SCR(share of category requirements) (w^e/c^e)	the number of visits to the brand site as a percentage of the number of visits to any brand site in the same category sites by per brand site surfer

*Superscript ^{e} indicates that the measures are of eBPMs rather than of the corresponding BPMs

Table 2. Definitions of eBPMs

CYBER PANEL DATA OF WEB TRAFFIC MEASURES

Consumer panel operators typically report the BPMs in Table 1 to their retail clients, whereas the cyber panel data operators typically measure web traffic. Cyber panel operators recruit large enough samples to reflect the population and require their participants to install software to record accurately their browsing behavior. Table 3 summarizes the web traffic measures currently used by large cyber panel operators. Research involving web traffic analysis has tended to use web traffic measures to evaluate the effectiveness of online advertising and marketing communication (Dewan et al., 2002; Chatterjee et al., 2003; Drèze and Zufryden, 2004) and the modeling of online consumer behavior (Moe and Fader, 2004; Park and Fader, 2004). Some researchers have also used web traffic measures to evaluate the market value of the Internet brand (Demers and Lev, 2001; Keating et al., 2003; Rajgopal et al., 2003).

Measures	Definition	Operator
Unique audience(UA)	number of unique surfers who have visited the site during the reported time period	MM/NR
Active reach(AR)	percentage of active surfers who visited the site during the reported time period	NR
Universal reach(UR)	estimated percentage of all surfers who have visited the web site during the reported time period	MM/NR
Visit duration(VD)	number of viewing hours by unique surfers who have visited the site during the reported time period	MM/NR
Visits(VS)	number of visits to the site per unique visitor during the reported time period	NR

*NR: NetRatings; MM:Media Metrix

Table 3. Web traffic measures

OBTAINING OBSERVED EBMPS FROM CYBER PANEL DATA

One of the main purposes of this study is to demonstrate how to use the existing cyber panel data of Web traffic to obtain eBMPs. Therefore, based on the definitions of eBPMs in Table 2, we adopt the Web traffic measures in Table 3 to propose a formula for calculating brand market share from two eBMPs.

$$\text{Brand market share } (m^e) = \frac{\text{Number of visits to the brand site}}{\text{Number of visits to category sites}} = \frac{VS_b}{VS_c} \quad (1)$$

The brand market share of a category (me) is obtained by dividing the total number of visits to the brand site by the total number of visits to category sites during the reported period. For the total visits to the brand site (we use the variable of VSb here) 1, the existing web traffic measure "the number of visits to the web property per unique visitor during the reported time period" (we use the variable VS in Table 3) is used. However, the total number of visits to category sites depends on the context of the research for which category of web sites (e.g., portals, content/community, or e-tail, etc.). Thus, it is calculated by summing the VSb for all brands in a specific category (we use the variable VSc here). Accordingly, the brand market share (me) is obtained by dividing VSb by VSc.

$$\text{Brand penetration } (b^e) = \frac{\text{The number of web surfers who have visited the brand site}}{\text{The number of web surfers who have visited category sites}} = \frac{UA_b}{UA_c}$$

[1] While the superscript $_b$ of VS_b indicates the brand level, the $_c$ of VS_c indicates the category level variable.

(2)

Brand penetration (be) is calculated by dividing the number of web surfers who have visited the brand site at least once by the total number of web surfers who have visited category sites at least once. Therefore, the number of web surfers who have visited the brand site can be obtained by the web traffic measure "unique audience" (UAb). As for the total number of web surfers who have visited category sites, it can be calculated by summing the UAb for all brands in the category (we use the variable UAc here). Therefore, brand penetration (be) is obtained by dividing UAb by UAc.

Category penetration (B^e) =

$$\frac{\text{The number of web surfers who have visited the web category}}{\text{The number of the universe of web surfers}} = UR_c$$

(3)

The category penetration (Be) is calculated by dividing the total number of web surfers who have visited category sites by the total number of all universe web surfers. The existing measure "universal reach" (we use the variable UR in Table 3) is used to estimate the percentage of all surfers who have visited a brand site during the reported period. Therefore, the measure of category penetration (Be) can be obtained by summing UR for all brands in the same category (we use the variable URc here).

Brand purchases per buyer (w^e) = number of visits to the brand site = VS_b. (4)

The brand purchases per buyer (we) is the average number of visits per brand site surfer. Thus, it can be obtained by the existing measure of "visits" (we use the variable VS in Table 3).

Category purchase frequency (W^e) = number of visits per category of web surfer = VS_c (5)

The category purchase frequency (We) is the average number of visits to category sites per category surfer. The most straightforward way to obtain the category purchase frequency (We) is to average the VSb for all brands in a category (we use the variable VSc here).

Category purchases per brand surfer (c^e)

= number of visits to any brand site in the same category per brand site surfer

= VS_{bc}.

(6)

The measure of category purchases per brand buyer (ce) represents aggregate data of average number of visits to any brand site in the same category per brand site surfer. It requires to identify the subset of surfers who visited each brand site and who also visited other sites in the same category sites. Thus, no existing web traffic measure can be used. However, the measure can be obtained by two steps: first, identify surfers for each brand who visited the brand site at least once and also visit other brand sites in the same category; second, calculate the average number of visits for those brand site surfer in step one (we use the variable VSbc here).

SCR^e (share of category requirements)

$$= \frac{\text{The number of visits to the brand site}}{\text{The number of visits any brand site in the same catgeory}}$$

$$= \frac{w^e}{c^e}$$

$$= \frac{VS_b}{VS_{bc}}.$$

(7)

Based on the definition of SCRe, it is "the number of visits to the brand site" (We) over the number of visits to any brand site in the same category sites by per brand site surfer (i.e., the category purchases per brand surfer (ce)). Therefore, the SCRe can be obtained directly by dividing VSb by VSbc.

POTENTIAL APPLICATIONS OF EBPMS

The eBPMs compiled from existing cyber panel data provide the theoretical basis for evaluating brand performance of Internet brands. There are several potential applications of eBMPs.

First, some researchers have attempted to incorporate measures of web traffic as non-financial variables, to explore the link between web traffic and the share prices of Internet brands (Demers and Lev, 2001; Keating et al., 2003; Rajgopal et al., 2003). Others have modeled online consumer behavior by relying on web traffic measures to evaluate the effectiveness of online advertising and marketing activities (Chatterjee et al., 2003; Moe and Fader, 2004; Wilson, 2004). However, previous studies that adopt different web traffic measures—e.g., reach, visit duration, and stickiness—might lead to different results. In addition, the theoretical basis of marketing requires further investigation for existing web traffic measures. Hence, eBPMs, based on the robust theory of the Dirichlet model and BPMs, can provide a valuable foundation for further research examining the link between web traffic and market value, the effectiveness of marketing activities, and online consumer behavior. In addition, findings based on eBPMs are readily comparable.

Second, the measures of eBPMs are obtained from across-site rather than within-site data. The measures thus provide a tool for evaluating the performances among competitive Internet brands. Hence, eBPMs have the potential to function as standard web traffic measures, routinely collected by cyber panel operators in the online marketing research industry.

Third, in this study, we did not explore the question of whether the assumption of the Dirichlet model about buyers' purchasing patterns—stochastic distribution—applies to a surfer's likelihood of repeat visits. However, research such as Park and Fader (2004) provides some support for the theoretical basis for the suitability of the stochastic theory to model repeat browsing behavior. Hence, we suggest that researchers can calculate the theoretical prediction of eBPMs by Dirichlet model analysis and compare the obtained eBPMs by cyber panel data to assess the capability of the Dirichlet model to predict surfers' patterns of repeat visits (Goodhardt et al., 1984). Ehrenberg et al. (2004) suggests ways of interpreting the results obtained from BPMs versus theoretical predictions of BPMs.

In summary, we suggest that eBPMs can act as systemic web traffic measures to evaluate the market value of a web site, brand performance, and the effectiveness of web marketing activities, incorporating financial information, competitors' information, and online marketing activities.

REFERENCES

Bucklin, R.E. and Sismeiro, C. (2003). A model of web site browsing behavior estimated on clickstream data. Journal of Marketing Research, 40(3), 249–267.

Bhattacharya, C.B. (1997). Is your brand's loyalty too much, too little, or just right?: Explaining deviations in loyalty from the Dirichlet norm. International Journal Research in Marketing, 14(5), 421–435.

Chatterjee, P., Hoffman, D.K., and Novak, T.P. (2003). Modeling the clickstream: implications for web-based advertising efforts. Marketing Science, 22(4), 520–541.

Danaher, P.J., Wilson, I.W., and Davis, R.A. (2003). A comparison of online and offline consumer brand loyalty. Marketing Science, 22(4), 461–476.

Demers, E. and Lev, B. (2001). A rude awakening: internet shakeout in 2000. Review of Accounting Studies, 6,(2–3), 331–359.

Dewan, R.M, Freimer, M.J., and Zhang, J. (2002). Management and valuation of advertisement-supported Web sites. Journal of Management Information Systems, 19(3), 87–98.

Drèze, X. and Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. Journal of Interactive Marketing, 18 (1), pp20–37.

Ehrenberg , A.S.C., Barnard, N.R. and Sharp, B. (2000). "Decision models or descriptive models?. International Journal of Research in Marketing, 17(2–3), 147–158.

Ehrenberg , A.S.C., Uncles, M.D., and Goodhardt, G.J. (2004). Understanding brand performance measures: using Dirichlet benchmarks. Journal of Business Research, 57(12), 1307–1326.

Fader, P.S. and Schmittlein, D.C. (1993). Excess behavioral loyalty for high-share brands: deviations from the Dirichlet model for repeat purchasing. Journal of Marketing Research, 30(4), 478–493.

Goodhardt, G.J., Ehrenberg, A.S.C and Chatfield, C. (1984). The Dirichlet: a comprehensive model of buying behaviour. Journal of the Royal Statistical Society A, Vol. 147, 621–655.

Ilfeld, J.S. and Winer, R. (2002). Generating website traffic. Journal of Advertising Research, 42(5), 49–61.

Keating, E.K, Lys, T.Z., and Magee, R.P. (2003). Internet downturn: finding valuation factors in spring 2000. Journal of Accounting & Economics, 34(1–3), 189–236.

Keller, K.L. (1993). Conceptualizing, measuring, and *managing* customer-based brand equity. Journal of Marketing, 57(1), 1–22.
Moe, W.W., and Fader, P.S. (2004). Capturing evolving visit behavior in clickstream data. Journal of Interactive Marketing, 18(1), 5–19.
Park, Y.H. and Fader, P.S. (2004). Modeling browsing behavior at multiple websites. Marketing Science, 23(3), 80–303.
Rajgopal, S., Venkatachalam, M., and Kotha, S. (2003). The value relevance of network advantages: the case of e-commerce firms. Journal of Accounting Research, 41(1), 135–162.
Riel, A.C.R. van, Liljander, V., and Jurriëns, P. (2001). Exploring consumer evaluations of e-services: a portal sit. International Journal of Service Industry Management, 12(4), 359–377.
Serrano-Cinca, C., Fuertes-Callen, Y., and Mar-Molinero, C. (2005). Measuring DEA efficiency in Internet companies. Decision Support Systems, 38(4), 557–573.
Stern, P. (2004). The relationship between customer loyalty and purchase incidence. Marketing Letters, 15(1), 5–19.
Wilson, R.D. (2004). Using web traffic analysis for customer acquisition and retention programs in marketing. Services Marketing Quarterly, 26(2), 1–22.

Effectiveness of Web-based Learning – An Evaluation

NaSrullah Memon
Department of Computer Systems and Software Engineering
Mehran University of Engineering and Technology Jamshoro, Sindh, Pakistan
Email: nm@acm.org

Abdul Qadeer Khan Rajput
Department of Computer Systems and Software Engineering
Mehran University of Engineering and Technology Jamshoro, Sindh, Pakistan
Email: vc@muet.edu.pk

Bhawani Shankar Chowdhry
Department of Electronics and Telecommunication Engineering
Mehran University of Engineering and Technology Jamshoro Sindh, Pakistan
Email: c.bhawani@ieee.org

Abstract

In this case study an attempt is made to assess the effectiveness of supplementary web-based materials and activities in computer programming courses using statistical methods. We have collected data on 66 students (of the department of Computer Systems and Software Engineering Mehran University of Engineering and Technology Jamshoro) from three classes: 01CS, 01SW, and 2KCS that describe various characteristics, use of web-based learning resources, and performance on graded quizzes and examinations. We use this data to evaluate the effectiveness of the web-based learning material.

Keywords:
Web-based learning, Distance education, Discussion/Chat Forums

INTRODUCTION

The use of web-based learning is increasingly common in many disciplines in higher education. Appropriate materials are available on-line from libraries, class discussions are held via e-mail, discussion/chat forums; textbook publishers provide WWW sites for their products, etc. Software developers are making such programs available to colleges and universities. Although these materials are generally used as supplements in traditional lecture settings, they also serve as a substitute for class meetings in the rapidly growing area of distance education.

Much more is still to be known about the effectiveness of these web-based supplements to face-to-face learning/teaching. How intensively will students utilize on-line course materials? Does access to on-line course materials increase comprehension and retention? Despite the paucity of answers to these and similar questions, the rush to make on-line technology an important component of higher education continues.

This study assesses the effectiveness of on-line materials in two courses of programming. These courses, "Introduction to C++ Programming (ICP)" and Object Oriented Programming (OOP) using Java, are being offered as traditional face-to-face courses for several years by the teachers with no web-based component. In this case, supplemental web-based components have been added to existing courses without a complete overhaul of the design and pedagogical approach, taught at MUET, Jamshoro.

LITERATURE REVIEW

As the use of computers and web-based learning has grown over the last decade, so has interest in assessing the effectiveness of these tools and methods. Due to expensiveness and unavailability of Internet, web-based learning uses a relatively small set of highly adaptable tools like synchronous and asynchronous computer mediated communications and hyperlinked content. Research on the effectiveness of such techniques can be applicable to many academic disciplines and settings. Consequently, our review of this literature is selective rather than comprehensive.

Agarwal and Day (1989) examined the effect of web-based instructional techniques on measures of student outcomes, as measured by course grades and results on the Test of Understanding College of Economics (TUCE). This study employed a control group that did not have access to web-based materials and an experimental group that did, both taught by the same instructor using the same text, tests, and instructional style. The experimental group made use of E-mail and discussion lists for communication and the WWW for information retrieval and access. In this study, those students with access to web-based instruction performed better than those who did not in the sense that their average score on the TUCE was 1.15 points higher (statistically significant).

Navarro and Shoemaker (2000) found that students in the principles of macroeconomics course who had access to a set of web-based instructional material (a CD-ROM with course content, class-related bulletin boards and chat rooms, and e-mail) performed better than students who did not have access to this material by scoring significantly higher on an 11 question final exam composed of essay questions.

John Siegfried and William Walstad (1998) summarized the extensive literature in this area. Our research can be viewed as an extension of the methods and techniques used in these studies to the area of web-based learning.

Kearsley, Lynch, and Wizer (1995) Bruce, Peyton, and Batson (1993); Berge and Collins (1995); Harasim (1989, 1993); Hiltz (1994); Mason and Kaye (1989); Waggoner, (1992) have all evaluated the role of online learning. The general theme of these studies is that student satisfaction is increased, there is greater interaction between students and between students and instructors, and critical thinking and problem-solving skills are frequently reported as improved. Moreover, grade point average and other measures of student achievement are as high as or higher under online teaching than in traditional classes. Makrakis, et al. (1998) used regression analysis to assess the effectiveness of a hypermedia system and courseware in computer science instruction. They found the design and presentation of instructional material and students' on-line interaction with teachers to be important explanatory variables.

EVALUATION FRAMEWORK

We employ a straightforward evaluation strategy. We want to evaluate the intensity of utilization of web-based learning, which affects the outcomes in introductory programming courses. Author # 1 began by referring the students to web-based learning materials (http://www.jquest.com, http://www.brainbench.com, http://javaonline.4t.com, www.javaranch.com, www.javaworld.com, www.jcertify.com, www.javaprepare.com, www.programmersheaven.com, www.devx.com, etc.) including interactive exercises and computer-graded quizzes, and made the access of web browsing available at library/labs to the students.

The materials were essentially supplemental; they did not take the place of any classroom-based activity. Instead, the web-based material and activities were intended to increase the interaction of the students with the material and the teachers, as well as the interaction between students, beyond the level of interaction found in a typical lecture course with no web-based material or activity.

Step 1. We collected random data from three random sample classes mentioned above. The selection of students was made keeping in view the demographic characters of general student population.

Step 2. We prepared the frequency distributions of the sample student data on class standing, hours spent by students on working, primary Internet access and commute time (Table-1) in order to depict the general trend of student sample.

Step 3. The frequency distribution is prepared on number of attempts on practice quizzes and high scores obtained by the students (Table-2) in order to accept or reject the null hypothesis of no relationship between the high scores and attempts per quiz and between the use of web based learning resources and performance in the final exams.

Step 4. The frequency distribution is prepared on number of students visiting a website lying within a range of times, i.e. starting from infrequent (0-100) to high frequency range (2001-2500) and the number of students participating in discussion / chat forums starting from less frequent (0-10) to more frequent participation i.e. (51-60) as shown in Table-3.

We use students' scores on various quizzes/tests and the final exam as evaluation instruments. The final exams had similar formats. Our investigation focuses on determining how much of the variation in the quiz and examination scores can be explained by variation in the students' use of the web-based material and activities after controlling for other observable factors that might affect those scores. Our prior belief was that the more a student utilized the web-based material and activities, the better that student would perform on the exam. Recognizing that web-based learning represents complex systems with many inter-related components, we included a number of different measures of student use of the web-based material.

DATA DESCRIPTION

The data were collected from 01CS, 01SW, and 2KCS classes at MUET during the academic year 2001-2002. Two courses were ICP and OOP, taught by first author.

The students in each class had been engaged to visit to course-related material on the above-mentioned web sites. The web-based material included course related content (including supplemental readings), practice quizzes, hyperlinks to the latest and updated course-related material on the Internet, involvement in discussion/chat forums for synchronous discussion, Net Meetings, and e-mail etc.

66 students in three classes, 38 in the ICP sections and 28 in the OOP courses showed interest. 06 students term backed (due to short of attendance) leaving 60 students that successfully completed the courses through the final exam. Prior to the first day in class, there was no indication that web-based material would be used in the classes.

DEMOGRAPHIC DATA

Demographic data on the students were collected by using statistical surveys. 14 students (21%) did not complete these surveys, and the following statistics are based on data from the 52 students who completed the surveys. The sample comprised of 77% male and 33% female. 69% reported being involved in extra curricular activities. Additional sample information is presented in Table 1.

Table 1 Demographic Data

Class Standing	**%**	**Primary Internet Access**	**%**
01CS	31	Home/Hostel	61
01SW	42	Library/Lab	35
2KCS	27	Other	4
Hours Spent Working	**%**	**Commute Time**	**%**
Did Not Work	18	Lived On Campus	63
< 10 hours per week	43	< 10 minute drive	8
10-20 hours per week	15	10-20 minute drive	12
20-30 hours per week	12	20-30 minute drive	10
> 30 hours per week	12	< 30 minute drive	8

INTERNET USE

How intensely do students utilize web-based learning resources? Answering this question is an important step in evaluating the effectiveness of these resources. If students are reluctant to use web-based resources, then no matter how effective these materials are at enhancing comprehension, they will ultimately have little value.

Simply making supplemental material available to students does not guarantee that students will utilize these materials. Copies of past exams and solutions to problems made available at department's seminar library and central library of MUET are often neglected by students; may be because the total cost of accessing them (including time, and copying costs) exceeded the expected benefit; where as computer-assisted learning material have a lower cost of access, which will lead to increased use and, consequently, comprehension and mastery of the material.

The learning material used in this study can be grouped into two general categories: material that enhances the student's interaction with the course material (which includes the practice quizzes and the supplemental web-based content) and the material that enhances the student's interaction with other students and the teacher, primarily through the involvement in discussion/chat forums (Net Meetings and, to a lesser extent, through e-mail). We examine each in turn.

UTILIZATION OF PRACTICE QUIZES

Students had been engaged to practice quizzes (composed of multiple choices, true-false) available at the mentioned sites. Each quiz consisted of a small set of questions drawn randomly from a large pool of potential questions. Each quiz could be taken up as many times as the user wishes, and the pool of potential questions is large enough that the probability of drawing the same question in multiple quizzes is small. The quizzes are also graded by the computer as soon as the quiz was submitted, and students could immediately see their score and the correct answer to each question.

The 01CS and 01SW classes (38 students) used five quizzes and appeared in the final exam, and 2KCS class (28 students) used five quizzes. Because each quiz was taken up to five times, there were a total of 1,840 (38 x 5 x 6 + 28 x 5 x 5) student quiz-opportunities for the 66 students. There were 1,195 actual student quiz-attempts, a 65% utilization rate by the students, suggesting that the students made considerable use of the practice quizzes. Table 2 shows the frequency distributions for utilization of the practice quizzes and the scores on the practice quizzes. Second row of 1st panel shows that in 29 instances a student took a particular practice quiz only once, despite the potential for taking that quiz for additional times; this represented 8% of the practice quiz attempts in the sample. A majority of the students who attempted any given practice quiz took that quiz the maximum number of times allowed (5), suggests that students perceived some benefit from multiple attempts at the quizzes.

Table 2. Frequency Distributions - Practice Quizzes

# of Attempts (x)	**Frequency f**	**x f**	**%**	**High Score**	**Frequency**	**%**
0	65	0	18	0-59	19	6
1	29	29	8	60-69	13	4
2	34	68	9	70-79	29	10
3	32	96	9	80-89	64	21
4	38	152	10	90-99	75	25
5	170	850	46	100	103	34
Total	368	1195			303	

The right panel of Table 2 shows that frequency distribution of the highest score on a quiz for 303 instances where a student took a quiz one or more times. In 80% of these cases the high score (80 to 89%) or (90 to 100%) was achieved. The modal percent of the possible points is 100%, which occurred in 103 cases. In other words, in 103 cases out of 303 observations, the student taking a quiz answered every question correctly. One possible explanation for this high frequency is that the quizzes were relatively short (5-10 questions each). Another possible explanation is the benefits students derived from taking the quizzes multiple times.

A natural question is to examine the possibility of a statistical relationship between the high score on a quiz and the number of times that quiz was attempted. Since quizzes were taken outside of class, one could interpret more attempts as greater student effort or more time spent studying for the course. A statistically significant relationship between these variables would be the evidence that some sort of learning took place when students exerted more effort by taking a quiz multiple times. If these variables were statistically independent, then no learning took place and the performance is unrelated to outside effort. The Pearson χ^2 statistic for this sample is 41.25, which has a P-value of 0.002. The null hypothesis of no relationship between high score and attempts per quiz is rejected, suggesting the presence of some relationship between these variables. All scores below 69 were placed in the same category for this test in order to obtain enough cells with a predicted value of more than 5% to make the χ^2 test valid. A likelihood-ratio χ^2 test similarly suggested a relationship between the variables.

UTILIZATION OF ASYNCHRONOUS COMMUNICATION TOOLS

The on-line resources are used to increase student interaction with the material by providing web-based content or to increase student interaction with other students. The latter category includes discussion/chat forums, Net Meetings and e-mail. Some measures of student use of these resources are summarized in Table 3.

The variable "Visits" on Table 3 is the total number of the course material visited by each student over the course of the semester. This variable reflects general student use of the course material.

The frequency distribution of "Visits" suggests that there was relatively little variation in the students' access of the on-line material. The total visits for a majority of the students' falls in the 101-500 range. A small group of about 10% of the students either utilized, or surfed through this material much more frequently (i.e. more than 1500 times).

Table 3. Frequency Distributions - Communication Tools

Visits	**Observed**	**%**	**No of Participations in Discussion/Chat forums**	**Observed**	**%**
0-100	4	6.1	0-10	34	51.5
101-500	36	54.6	11-20	8	12.1
501-1000	11	16.7	21-30	7	10.6
1001-1500	7	10.6	31-40	10	15.1
1501-2000	7	10.6	41-50	3	4.6
2001-2527	1	1.5	51-60	4	6.1
Total	66	100.0		66	100.0

The general pattern observed is that majority of students participate relatively infrequently (the modal number of participations made in the sample is 5) but a smaller but important group of students (about 38% in the next three groups) participated considerably more often.

We have decided not to use survey data on student attitudes about web-based material in this case study. Several factors affected this choice. Such data on student's attitudes about web-based material are frequently analyzed in the distance education literature and often find that students feel that web-based material is useful and beneficial. See, for example, Agarwal and Day (1998); Kearsley, Lynch, and Wizer (1995); Bruce, Peyton, and Batson (1993); Berge and Collins (1995); Harasim (1989, 1993); Hiltz (1994); Mason and Kaye (1989); Waggoner, (1992), and Rajput (2003). We chose to examine the relationship between use of web-based material and performance in this paper, and feel that, if done correctly, this analysis can increase our understanding of the appropriate role for these materials.

CONCLUSIONS AND SUGGESTIONS

We set out to describe and analyze students' use of web-based materials in programming courses viz. ICP and OOP. Students in the three classes were engaged to visit an array of web-based material, including content, computer graded quizzes that could be taken multiple times, and synchronous and asynchronous communications tools. The students made extensive use of the on-line quizzes, completing about 2/3 of the total available quizzes. The distribution of the total number of "visits" of the web sites suggests that students did not ignore the web-based content, but that a majority of the students probably did not return to this material multiple times. A majority of students were somewhat reluctant to make participation discussion/ chat forums although a small but significant fraction participated frequently.

Student utilization of the web-based material was significant. Our analysis suggests that on-line practice quizzes can be an effective tool. Taking multiple quizzes on a topic increased the high score on that topic significantly and there is some evidence that more attempts at the practice quizzes was positively correlated with the student's score on the final exam. Participation in discussion/chat forums also appears to be positively correlated with performance.

Participation in discussion/chat forums was a good predictor of performance. Since correlation between participation and performance reflects learning, therefore teachers using discussion/chat forums should focus on developing interesting discussion topics and designing discussion exercises that draw more students into the on-line discussion.

Finally, we note that these conclusions and observations are based on a relatively small sample of students. More data collection and analysis needs to be done in this area before definitive conclusions can be reached. We view this as an ongoing research project, and plan to continue to collect data. We strongly encourage other faculty who use web-based material in their classes to take the time to undertake similar studies.

REFERENCES

Agarwal, R. & Day, A. E. (1998). The impact of the internet on economic education. *Journal of Economic Education*, 29 (2), 99-110.

Berge, Z. L. & Collins, M. P. (1995). *Computer-Mediated Communication and the Online Classroom*, Creskill, NJ: Hampton Press.

Bruce, B., Peyton, J. K. & Baston, T. (1993). *Network-based Classrooms: Promises and Realities*, Cambridge, UK: Cambridge University Press.

Harasim, L. (1989). *Online Education*, New York, N.Y.: Praeger.

Harasim, L. (1993). *Global Networks*, Cambridge, MA: MIT Press.

Hiltz, R. (1994). *Virtual Classrooms*, Norwood, NJ: Ablex.

Kearsley, G., Lynch, W. & Wizer, D. (1995). The effectiveness and impact of online learning in graduate education. *Educational Technology*, 35, 37-42.

Makrakis, V., Retalis, S., Koutoumanos, A., Papaspyrou, N. & Skordalakis, M. (1998). Evaluating the effectiveness of an ODL hypermedia system and courseware at the National Technical University of Athens: A case study. *Journal for Universal Computer Science*, 4 (3), 1-15.

Mason, R. & Kaye, A. (1989). *Mindweave: Communications, Computers, And Distance Education*, New York, N.Y.: Pergamon Press.

Navarro, P. & Shoemaker, J. (2000). *Economics in the cyberspace: A comparison study*, Unpublished manuscript.

Siegfried, J. J. & Walstad, W. B. (1998). Research on teaching college economics. In Siegfried, J. J. & Walstad, W. B. (Eds.) *Teaching Undergraduate Economics*, Boston, MA: Irwin McGraw-Hill, 141-166.

Waggoner, M. D. (1992*). Empowering Networks: Computer Conferencing in Education*, Englewood Cliffs, NJ: Educational Technology.

AQK Rajput, FK Shaikh, and AQ Ansari, e-Learning: A Virtual Environment for Cooperative Distance Learning and Affordable Online Education*, MUET Res. J of Eng & Tech, Vol 22, No. 1, Jan 2003,* 57-64.

Strategy Analysis of Mobile Number Portability for Mobile Operators in the Use of Game Theory

Dr. Le-Pong Chin
Institute of Business Administration of Shih Chien University
Department of Information Management of Shih Chien
chin@mail.usc.edu.tw

Su-Wen Chuang
Institute of Business Administration of Shih Chien University

Abstract

Mobile number portability (MNP) allows the consumer to change to a new mobile phone service whilst keeping the same mobile phone number therefore the consumer can avoid the problems associated with changing of phone numbers when customers wish to change their operator so as to reduce the cost and inconvenience which is also called the transformation barrier. This article first demonstrates the monopoly market by mutual cooperation among the operators being unachievable through studying in the use of the concept of game theory and the CARTEL union monopoly model. This paper uses Game Theory to analyze the best course of action for the future market being cooperating for better survival but will fail due the possibility of distrust between each other or cheating. Therefore the competition among operators is inevitable. Our main contribution is to deduct possibility of the booming for operator to take opportunity by effectively utilization of the welfare strategy from starting point of MNP to grand huge market share so as to get the advantage of phenomenon of network externality.

Keywords:
MNP network externality CARTEL modeling AHP analysis Game theory

(9) INTRODUCTION

MNP allows the consumer to change to a new mobile phone service whilst keeping the same mobile phone number therefore the consumer can avoid the problems associated with changing phone numbers. At present MNP has been adopted by Hong Kong, Singapore, England, US, Canada, Australia, Sweden and so on. After achieving the first milestone of high penetration in the telecommunication industry the traffic saturate the infrastructure. Subscribers, sometimes, have to suffer the poor quality of service, for example high blocking rate, poor coverage, and high tariff, and wish to change to another service operator. According to the old regulation, subscribers are forced to change number when they decide to change operator, those subscribers have to suffer the inconvenience and, sometimes, pay additional expense. This phenomenon is called transformation barrier and reduce the motivation of users to change operator. The MNP policy encourages subscribers to get what they want and avoid the inconvenience of changing number. At the same time, operators have to propose their best quality policy and try their best to persuade the high quality of service with a reasonable tariff which is also beneficial in mobile industry development. Those activities are inevitable to increase the service cost and reduce the tariff income for operators. It's straightforward for operators to monopoly the market by mutual cooperation. If so, the regulation becomes ineffective in bringing beneficence to subscribers. This is obvious to analyze the cooperation and competition in the use of Game Theory. The CARTEL union monopoly model is one of suitable way to do this.

The second section surveys the necessary literatures. The CARTEL model is introduced in third section, and the analysis is conduct in section four. Finally, in section 5 the conclusion is made.

LITERATURE REVIEW

MOBILE SERVICE

Mobile communication becomes popular in coming era of m-commerce. More and more people feel uncomfortable and inconvenient without bringing their mobile phones. In Taiwan, the penetration rate increases rapidly when the first private mobile operator launches its service in 1995 and quantity of issued number is more than the population of Taiwan. Lots of persons have more than one mobile. After achieving the first milestone of high penetration in the telecommunication industry the traffic saturate the infrastructure. Subscribers, sometimes, have to suffer the poor quality of service, for example high blocking rate, poor coverage, and high tariff and wish to change service operator. According to the old regulation, subscribers are forced to change number when they decide to change operator and the inconvenience happens. This phenomenon is called transformation barrier and reduce the motivation of users to change operator. The MNP policy encourages subscribers to get what they want and avoid the inconvenience of changing number. At the same time, operators have to propose their best quality policy and try their best to persuade the high quality of service with a reasonable tariff. At present MNP has been adopted by Hong Kong, Singapore, England, US, Canada, Australia, Sweden and so on. This policy is proven to be beneficial to all subscribers. So as to have at present MNP has been adopted by Hong Kong, Singapore, England, US, Canada, Australia, Sweden and so on. This policy is proven to be beneficial to all subscribers and has beneficence in the development of mobile industry.

CARTEL MODEL

As a starting point, the centralized cartel model is used to increase the profits of all parties within the cartel. A strategy will be set while the most extreme model will allow the setting of price, productivity and profit for all parties in the cartel. However the centralized cartel model may not set the price and productivity at the same level where profit has been maximized. As shown by the following diagram and according to Friednam, due to companies engaging in deceit causes the centralized cartel model to deteriorate.

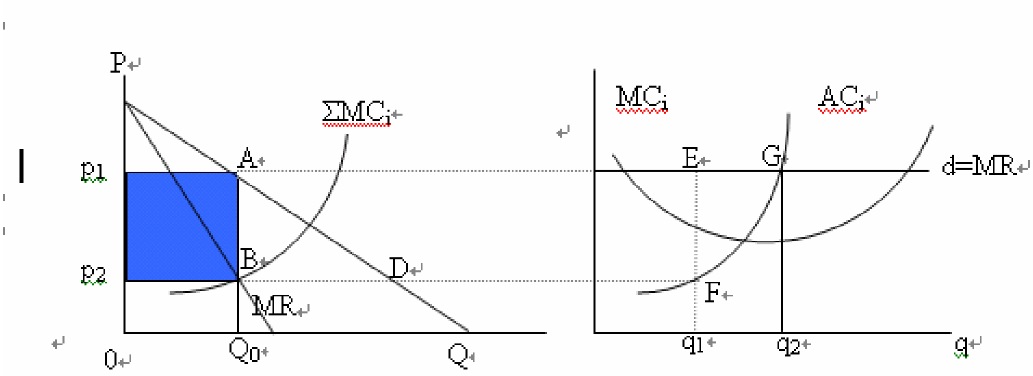
Diagram of Cartel Model

Plotted in Fig.1

P = Price Q = Quantity D = Demand

MR = Marginal Revenue AC = Average Cost MC = Marginal Cost

A monopolized industry sets price at p=p1. When producers unite in an attempt to quash the monopoly, they set their output to q=q1 ABP1P2. But when the producer produces out of self-interest at q=q2, profit becomes EGF, therefore all other producers will match the increment in production, causing the cartel to collapse.



Market producer

Fig. 1 the Cartel union model

The formation of a cartel which is some time to be charged as monopoly is usually not allowed. The characteristics of cartel can be described as follows: when prices are high and quantity is low the production can not be kept under the lowest cost. Generally speaking, the fewer members in a cartel, the easier for the parties to maintain agreement and keep the cartel going. When the market is booming, each producer may obtain a profit and the cartel is quite easy to maintain; If the economy is not booming, producers incurs a loss and thus in an attempt to strive for self-preservation, all parties attempt to deceive each other by secretly reducing prices to increase sales volume. Eventually, the cartel breaks down.

Telecommunication Network **Externality** Benefits

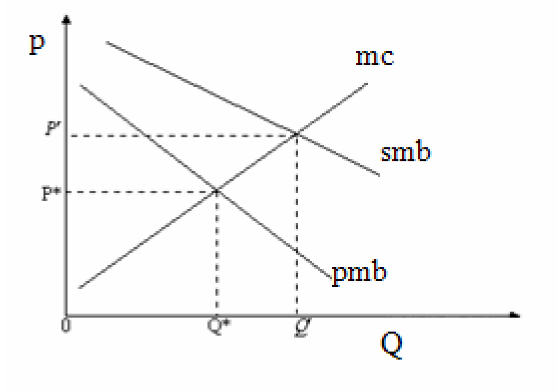


Fig.2 network externality l effect in market balance

From the cost point of view for a mobile operator, the cost for the calling party and the called party within in the same operator (PLMN) is lower than the call between two different operators (PLMN). Therefore, operators provide the discount for their subscribers when they call to the one who subscribe the same operator. This pricing policy encourage mobile user to select the operator which has the largest subscription.
Therefore the mobile phone has a network effect. Showing that a network externality effect is present in Fig.2. Because this network externality effect is existed, the system causes a domino effect, resulting in a "large is permanently large, small is getting smaller" competition situation. Therefore if there is an external market of competition, the domino effect results that the largest or the only operator will finally monopolize all the markets. When the network externality effect begins, the producer marginal cost will decrease because of the increase in consumers, whilst the marginal revenue will increase. In the figure 2, the vertical distance between the market profit line and the personal marginal revenue line represents the external nature of the network which is the extra advantage that consumers get when joining the network. For the individual user, the price they are willing to pay is P*, the quantity is Q*; For the entire market, the price is P', the quantity is Q', the market price is higher than the individual price, namely P' >P*, two disparities are precisely the external values which the network use producers.

GAME THEORY

Started in the 1980s, Game Theory gradually became an important analysis tool in the economics domain. The Game Theory fundamentally involves the studies of a group of stakeholders such as that under the forecast of possible choices, determines the best decision-making for the biggest beneficial gain, thus striving for survival within the competition (Wu Homao, the summer treasure,2,002) Tucker in 1950 to propose the renowned Prisoner Dilemma the Game Theory, satisfies the theory of Nash equilibrium.

Hypothesis:

According to Tucker in 1950 to propose the renowned Prisoner Dilemma, if police accuses A to accuse B each other

(1) A and B are cooperation to <u>deny</u> crime, A and B are guiltless, they won (5, 5)

(2)if A is to accuse B, B will went to jail one year, they won (6, -1); or B is to accuse A, A will went to jail one yea, they won (-1, 6)

(3) A accuses B each other, they won (0, 0)

prisoner to face a difficult choice		B	
		<u>deny</u> crime	accuse
A	<u>deny</u> crime	(5, 5)	(-1, 6)
	accuse	(6, -1)	(0, 0)

Fig. 3 Penalty diagram of Prisoner Dilemma

Research method

View point of enterprise strategy

(1)、combined monopolize – Cartel modeling

After the implementation of the MNP service, the consumer no longer needs many phone numbers, therefore D moves towards D '. MR also moves along with D towards MR ', therefore under the Cartel union monopoly model, producers may decide to cheat for their own self-interest. They do this by increasing quantity to q2; therefore the producers enter a Cartel union monopoly model, ending in failure state.

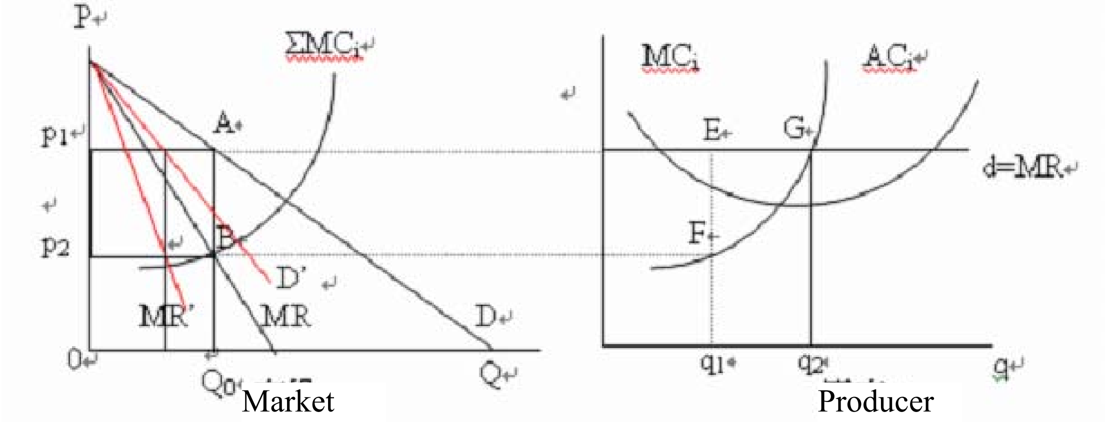


Fig.4 The Cartel after NMP

In this analysis model, by using the Cartel model from Economics theory as well as the network model, attempts to determine the status of competition after the implementation of the MNP service and find what strategy would be the best one. The reason that demand (D) decreases after the implementation of MNP service is that consumers no longer need to change their phone number when changing service providers, therefore the amount of phone numbers would decrease. Hence consumer demand for phone numbers will actually decrease after the implementation of the MNP service.

(2)、Game theory analysis

Starting from the above results, we know that a strategy of cooperation leads to failure. Even though the union monopoly is the best strategy, cheating existed potentially to leads to the final collapse of the cartel. The famous Prisoner's Dilemma demonstrates this phenomenon. Therefore in the following diagram no matter which producer first decides to cheat, in the end (competition, competition) is Nash equilibrium.

Producer B is another Telecommunication Company		cooperate	competition
	cooperate	$(\frac{1}{2}, \frac{1}{2})$	(1, 0)
	competition	(0, 1)	$(\frac{1}{2}, \frac{1}{2})$

Supposition: Producer A is Chinese Telecommunication and producer B is another telecommunication company. Explanation is described as follows: because the two companies are in cooperation mode, the market share is equally divided between A and B. However, due to the cartel model analysis, after MNP is implemented, the tendency of subscriber reduction entice operators impose cheating. Once when company begin to cheat to gain market share, the other company follows. The next competition between the two companies begin again.

V、Conclusion

Our main contribution is to explain the monopoly by cooperation among mobile operator is not feasible from the Cartel model of game theory. Similar to another example, Prisoner's Dilemma, the cooperation among the operators can grand maximum benefit but the explanation of Cartel model, the cooperation will collapse. We conclude that when MNP policy begins, all parties are mutually competing with each other as the most advantageous strategy.

The future studies are proposed as follows

Competition model for multiple parties can be studied.

Dynamic game theory can be conduct for future study.

REFERENCES

1▯ Pierre Martineau▯ Developing the Corporate Image▯ N.Y Charles Scribner's Sons▯ 1960▯ PP. 3-13 ▯

2▯ Robert M. Worcester▯ Consumer Marketing Research Handbook▯ N.Y▯ McGraw-Hill Book Co.▯ 1972▯ PP. 507▯

3▯ Rohlfs, J.▯ "A Theory of Interdependent Demand for a Communication Service▯ " Bell Journal of Economics and Management Science▯ 5 . 1974▯ PP. 16-37▯

4▯ http://corporate.fetnet.net/▯

5▯ http://www.cht.com.tw/▯

6▯ http://www.tcc.net.tw/about/▯

7▯ http://www.kgt.com.tw/▯

8▯ http://www.ulycom.com.tw/technology/communication_detail.asp?id=90▯

The Effects of Mobile Web Services on Reengineering Business Processes of a Cluster

Abbass Ghanbary
University of Western Sydney
ghanbary@cit.uws.edu.au

Abstract

This paper discusses the doctoral level research undertaken by the author in regards to the effects of mobile technology and web services on the business processes of a cluster or group of organizations The main focus of this research is to investigate the impact of mobile, Internet, computer and web services technologies (Emerging Technologies) in bringing various organizations together in a cluster and the subsequent effect of these technologies on their business processes. This research will also investigate the factors influencing the use of mobile technologies by personal users and concentrate on identifying the advantages and limitations of the aforementioned emerging technologies in terms of their impact on the way businesses are organized.

Keywords:

Mobile technologies, web services, organization, personal mobile phone usage.

INTRODUCTION

This paper outlines a doctoral level research on the effects of emerging technologies on re-engineering business processes of a cluster. Furthermore, this research also investigates the factors that impact the use of mobile technologies in the business processes. As a part of this research, two questionnaires are designed to investigate the usage of mobile devices by individuals and organizations. The foundation of this study is based on the behaviour of the people who live and work in an environment inundated with wireless communication devices. The accompanying surveys, discussed later in this paper, will also help this researcher in distinguishing the limitations and the applications of mobile technologies in the context of the personal usage. This research will also investigate the organizational expectancies when mobile technologies are involved.

As per Lacohee, et al, (2003) advances in technology has undoubtedly played a role in the rapid and unprecedented take up and wide spread availability of mobile communications. Perry, et al, (2001) state that the rapid and accelerated move towards the mobile technologies has provided the individuals the ability to work away from the office and be on the move. Thus we can surmise that the primary areas of impact of mobility include: general mobile usage, mobile banking, mobile shopping and the use of mobile devices at work.

Normal day to day activities becomes more related to the way we communicate with each other. Faster transfer of data, transaction and communication independent of location and time with identifying the functionality of mobile devices has become the mainstay of all challenged business processes. The mobile web service with the aid of web engineering will make it possible for customer to have access to the product and services. When the demand is identified the availability of the supply will be easily analysed and the sales or services will take place.

As per Unhelkar (2005) the reason that the service model is so attractive is its ability to incorporate standards and open protocols for calling services and transmitting data. Web services make software functionality available over the Internet so that programs can request a service running on another server (a web service) and use that program's response in web sites, WAP services, or other applications. The possibilities are endless.

This transformation of businesses has been evolutionary. At the beginning of the Internet age, with the aid of its communications capabilities, businesses were transferred to e-business and now the technology is making it possible to run our daily activities disregard of location and time boundaries. Ranjbar (2001) states that many technological products that were first considered by people as luxuries were later consumed as necessary items in daily life. The IT products are of no exception. If the computer was an item of luxury in the 1980s, it has already established itself as an essential component of most contemporary homes. Like electricity, water and gas, information seems a necessary commodity for running the everyday home life in a modern society. As one may open the tap for water and press a switch for light, one can also switch on the computer and obtain information and numerous IT-based services. Ghanbary (2003) has described the Internet as the most powerful tool that bring information to our homes and offices like water and electricity that come by power lines and pipes. Now, mobile technology is allowing the information to be accessible anywhere and anytime with the aid of advanced mobile gadgetry.

The users get interested in technology only when it provides them with the ability to achieve their product/services from the relevant supplier. As per Sun (2003) the user context is the user sensible information related to consumer behaviour, including their locations, surroundings and timings of the available information. Therefore, the data gathered during this research will be statistically analysed based on the given demographic and the personal experiences of the users to construct a theory on how mobile technology could advance its capabilities to create better value to individuals and their personal processes

To achieve our aim of research, this paper is presented in the following sections: a) abstract b) keywords c) introduction d) hypothesis and research questions e) questionnaires development f) methodology in gathering the data g) proposed gathering and analyses of the data and h) conclusions and future directions.

HYPOTHESIS AND RESEARCH QUESTIONS

Thus mobile technologies along with Web Services would form a formidable front for tomorrow's technology as far as the Internet is concerned the major areas of investigation would be:

- Technological issues such as modelling of mobile software applications.
- Emerging mobile technology in terms of bandwidth, hand held units, continuity of transactions on the move.
- The issues and challenges in incorporating Web Service in businesses with the aid of mobile technologies.
- The subsequent issues and challenges of Business Process Reengineering.
- Mobile applications that would facilitate business entities to m-enable their business processes.
- The methods to transit already e-transformed businesses into collaborative m-enablement in order to fully capture the opportunities.
- Methods to transform the m-enabled business entities smoothly in order to capture the full potential true mobility.
- Transformation of other related areas such as Customer Relationship Management systems into the m-enabled environment.

This study is based on the following hypotheses and questions:

H1: Mobile technology has influenced the way people adapt to their banking, shopping and work related activities
Arising from the above hypothesis, there are several research questions that has to be answered.

H2 : M-transformation of business processes enhance the delivery of service in business.

- Which stage does the organisation stand in terms of m-transformation?
- How could we apply the m-transition road map at this stage?
- What is the best possible method to apply m-transition road map? Is it organisation wide, Department wise or business unit wise?
- What are business processes that could be transformed in the unit already identified?
- What are the changes happening in the industry where the organisation belongs to in terms of this unit?
- What is the expected impact after mobile technology is introduced to selected business units in the organisation?

- What would be the direct impact on the customers once the new technology is fully implemented?
- Would there be any anticipated problems during the changeover?
- Is the continuity of transactions on the move a major concern of the mobile users?
- Do the users of mobile technology decide on the methods to transit already e-transformed businesses into collaborative m-enablement in order to fully capture the mobile opportunities?
- Do the limitations of mobile technology such as security, privacy and integrity concerns, the major drawbacks for the mobile users to fully rely on this technology for their daily activities such as banking, shopping and working?

We strongly believe there are still other questions existing which we propose to explore after the statistical analyses of the data arising from the questionnaire. The development of the questionnaire is described in the following section.

QUESTIONNAIRES DEVELOPMENT

The surveys will be validated and tested by collecting the sample data and placing the result in the spread sheet already designed in Microsoft Excel to test and verify our final result. The descriptive collected data will be transferred into the numeric values. The descriptive data will also be statistically analysed and they will be tested against the hypothesis of this research.

The questionnaire filled by personal users has been categorised into five different sections:

- Section A: General mobile usage
- Section B: Mobile Banking
- Section C: Mobile Shopping
- Section D: Mobile at work
- Section E: Demographics

Section A General mobile usage: This section classifies the participants who use mobile devices and whether they are connected to Internet using the personal device. The reasons for their purchase of a mobile device are also queried together with their perceived advantage and disadvantage of the mobile Internet.

Section B Mobile Banking: The questions asked in this section of the questionnaire give us the opportunity to identify the usage of mobile devices by the individuals in the mobile banking domain. The participants are asked to identify the advantages and disadvantages of using mobile devices in their banking activities. This will probe in to the sort of transactions that people conduct on their mobile devices. We also query on the transactions that people would want to conduct through the mobile devices. Any problems/difficulties/complaints in the use of mobile banking will also be queried. Finally, the participants will be asked to list improvements they would like to see in their mobile banking experience.

Section C Mobile Shopping: In this section the participants are asked about the usage of their mobile devices to shop on line. The participants are requested to list the perceived advantages and disadvantages of mobile shopping. They are also queried on the sort of products/services that is bought using their mobile devices and the improvements they would require to make their mobile shopping experience more satisfactory. The problems/difficulties/complaints they have in mobile shopping and the sort of products/services they would not buy using their mobile devices will also be probed. The participants will also be queried on the setup applications and facilities they would like to see, which would comfort and attract them into the mobile shopping environment.

Section D Mobile at Work: The questions on this section are concentrated on how mobile devices could improve the daily processes of persons at work. The participants will be queried, why they need mobile devices to do their given tasks and the advantages and disadvantages, using their mobile devices at work environment. Problems/difficulties/complaints experienced by using mobile devices at work will also be queried. The preference of the participants, whether to take given tasks in the office or anytime, anywhere using mobile technology will also be probed.

Section E Demographics: Participants are asked to give details about their age, level of education, level of income, occupation and their gender. The reason for this section is to identify the type of people and how the usage of the mobile device is influenced by their demographics.

And the questionnaire filled by the organizations has been categorised into five different sections:

- Section A: Your organization
- Section B: Mobile technology information
- Section C: Mobile technology management
- Section D: Mobile business processes
- Section E: Mobile web services

Section A Your organization: This section of the questionnaire is related to the size and the nature of the trade of the organization.

Section B Mobile technology information: The questions in this section are related to the use of mobile devices in the organization.

Section C Mobile technology management: The mobile technology management questions are related to the concepts of development and managing mobile networks as well as related to the general issues.

Section D Mobile business processes: The Mobile Technology Process Issues are used to see the process issues that are faced by organisations when adopting mobile technology.

Section E Mobile web services: The Mobile Web Services are used in collaboration of the business processes of different organisation while adapting to the new technology.

The sample of the participants would be urban based users. The demographics would be varied due to the sensitivity of the data which will differ as per the demographic factors of the samples. For the public questionnaire the categories of age and the level of education would be very important factors when investigating the usage of mobiles. The gender of the user also could have a bearing to certain usage. And for the organization's questionnaire the importance of the organizations size and the nature of the trade play the major factor while analysing the data. All these areas are tested in the questionnaires which would be analysed to see whether such demographic factors have any impact on the mobile usage.

METHODOLOGY IN GATHERING THE DATA

The research was firstly initiated by literature review to understand the nature of the existing problems in the mobile environment. It was realized that there is hardly any relevant study on the personal usage of mobile technology therefore the literature review was mainly concentrated on:

- The issues and challenges of electronic and mobile commerce.
- The issues and challenges of electronic and mobile business.
- The fundamentals of the data communication and its related protocols.
- The fundamentals of the mobile data communication and its related protocols.
- The fundamentals of web engineering and Web Services.
- The advantages and disadvantages of the Mobile Technologies.

All the data will be collected by a questionnaire while the participants remain anonymous. The data will be analysed and processed by statistical methods for deduction and testing of the hypothesis.

By careful analysis of the different research methodologies, the action research is also is considered as a research strategy for the continuation of this study. Due to nature of PhD research, it is required to analyze and identify the characteristics of the emerging technologies in the real world. It has been decided to continue with the action research after the completion of the data collection for the surveys.

PROPOSED GATHERING AND ANALYSES OF DATA

The available options in collecting the data are as follows:

Questionnaire and Personal Interview: This method was considered the most suitable for the project and was adopted.

Telephone Interview: This is also very suitable method for gathering the required data for the organization's questionnaire.

For the statistical analyses of the data, two separate spread sheets are already created using Microsoft Excel. The results will be exported to Microsoft Excel spreadsheets where detailed graphical representations would be constructed and statistical values would be calculated. Also, it has been decided to change all the descriptive answers to the numeric values and place them on Microsoft Excel. For the purpose of this research, the graphical presentation of the result and the usage of different statistical methods to test against the hypothesis are crucial.

CONCLUSION & FUTURE DIRECTION

This research will delve into Web Services, the impact of mobile technologies including mobile gadgets, wireless networks, relevant operating system as well as the sociological aspects of personal mobile usage including user profiling (segmentation), user analysis, variations in usage and so on.

With the event of mobile technologies, emerging as a leading technology in the business organisations, new knowledge has to be created in the lines of business processes. The process owners of various business processes would have to be empowered in order to realise the full potential of emerging technology. This study aims to bring new business models based collaborative Web Services m-enabled business organisations**.**

REFERENCES

Ghanbary, A. (2003). *" Effects of Computers on Family and Leisure Time"*. B. Applied Science, Honour Thesis. Australia, Sydney: University of Western Sydney.

Lacohee, H., Wakeford, N., Pearson, I. (2003). *"A social history of the mobile telephone with a view of its future",* BT Technology Journal, Vol. 21(3),July 2003..

Lientz, P. B., & Rea, P. K. (2001). *"Start Right in E-Business. A Step by Step Guide to Successful E-Business Implementation "*. Academy Press. ISBN: 0-12-449977-5(pb)

Perry,M.,O'Hara,K.,Sellen,A.,Brown,B. and Harper R.(2001), *"Dealing with Mobility: Understanding Access Anytime, Anywhere"*, ACM Transactions on Computer Human Interaction, Vol. 8(4),pgs 323-347, December 2001

Ranjbar, M. (2001), *"Contemporary Computing".* Sydney: University of Western Sydney.

Sun, J. (2003). *"Information Requirement Elicitation in Mobile commerce",* Communications of the ACM, Vol. 46(12), December 2003..

Unhelkar, B. (2005). *"Web Services Extending BPR to Industrial Process Reengineering"* Proceedings of Information Resource Management Association, International Conference. San Diego, California. May 2005.

The effects of m-transformation of business processes in the delivery of service to customers

Dinesh Arunatileka
PhD Research Student, University of Western Sydney

Abstract

The paper discusses the m-transformation of business processes of an organization. M-transformation subject the business processes to change in a systematic way in order to facilitate mobile technologies and reap the advantages of new technology to the organization. There are two models proposed, one to identify the current stage of an organization with regard to mobile technology and the other to take the organization through the transformation, systematically.

Keywords:

M-transformation, transformation of business processes, m-transition, Scale of four stages in an organisation in m-transformation

INTRODUCTION

The focus of this paper is to describe a framework to m-transform business processes of an organization in order to enhance the service delivery to customers. There are two models proposed by the author in order to identify and transform the business processes. It is important to use a systematic approach when transforming an organization to embrace mobile technology. The paper discusses such a systematic approach in order to fully utilise the advantages offered by the mobile technologies.

The Internet's widespread popularity and the ubiquity of the World Wide Web have revolutionised the face of business (Chung, Lin, & Shim, 2005). The digital revolution has changed the way business is being done to such an extent that place is becoming space and time is no more a barrier for business transactions. The existence of eMarket places makes many buyers and sellers, big or small, to compete in the same online space for transactions and collaborations (S. Arunatileka & Arunatileka, 2003). Thus the business organizations have become more customer-centric in order to retain their customers. The service organizations especially have used technology in servicing customers to be more customer centric and being pro active in gaining, retaining and growth of customers. Zureik & Mowshowitz (2005) state that the rapid growth of e-commerce attests to the advantages of using internet facilities to identify, reach and persuade an audience to join an action. The first phase of such rapid change brought upon by the internet culminated in electronic commerce. M-commerce is riding on e-commerce but adds the very valuable element of mobility into the business transactions. The transition from e-commerce to m-commerce is described in the next section after the description of the paper structure.

The paper is structured as follows:
Introduction is followed by a chapter named e-commerce to m-commerce where e-transformation in general is discussed with a view to introduce m-transformation. This is followed by the proposed m-transformation methodology chapter where the proposed models by the author are discussed. Then comes the research goals in general followed by the specific hypothesis and questions concluded with the future work.

E-COMMERCE TO M-COMMERCE

There is substantial research material regarding the electronic transformation of businesses in order to reap the opportunities and advantages offered by e-business. There have been several methodologies already proposed by such researchers. Sawhney & Zabin (2001) have stated that the transformation of a business organisation in to e-business is not a one-step process but a journey. They have proposed seven steps in this journey namely: E-vision (broadening the view), E-volution (Climbing the ladder), E-strategy (Playing with LEGOs), E-syncronization (Breaking the boundaries), E-Infrastructure (Opening the hood), E-Capitalization (Placing the winning bets) and E-Organization (Rallying the people). The best advantages of mobile technologies could be obtained by the organization if the etransformation has already happened and the electronic systems in the organization is already in place. M-transformation could be identified as an extension to e-business which also adds the element of mobility, to fully extend the flexibilities of the organization. However, service organizations may want to enable mobile facilities to service customers due to the increased pressures of the knowledgeable customers and further cultivate a more customer centric business culture, even though the e-transformation process has not yet fully matured in those organizations. Therefore our discussion in this paper is based upon organisations using mobile technology irrespective of the fact, whether the organizations are fully e-transformed or not.

S. Arunatileka & Arunatileka (2003) proposes another model for e-transformation named Seven E's in e-transformation. The seven stages are: environment analysis, e-business goals/strategies, e-readiness, e-transformation road map, e-transformation methodology, e-systems, evolution (change Management). In this model, a very high emphasis is given to the environmental factors of the organizations in order to fine tune the transformation. The e-transformation road map which is part of this model was initially proposed by (Ginige, Murugesan, & Kazanis, 2001). The m-transformation road map proposed in the paper is based on the e-transformation road map due to the similarities of the scenarios. The next chapter describes the m-transformation methodology.

PROPOSED M-TRANSFORMATION METHODOLOGY

Mobile technologies have facilitated the customer centric business culture in a substantial way. Mobile technologies, riding on the back of traditional internet, ensure that information is available to its users independent of a physical location. This is a major wave after the internet, as users have a high expectation of mobile applications – particularly being able to access their applications from anywhere and at any time. Mobile technologies and the resulting mobile applications have brought about unique phenomena in personal, social and business domain that is unparalleled (D. Arunatileka & Unhelkar, 2003). Banks, as primary institutions of service oriented business, have increasingly leaned towards e-commerce based operations. Emerging mobile technologies offer "anytime anywhere" type of banking that results in better customer orientation and provide personalization of services to the customer (D. Arunatileka, 2006).

The proposed methodology for m-transformation comprises two models to identify and m-transform the business processes of an organization. Model (1) is about the level of mobile technology maturity in the organisation under consideration. Model (1) is known as the Scale of four stages in m-transformation in an organization. This model will categorise the organizations in one of four stages with regards the organisation's usage of mobile technologies. The proposed four stage scale of m-transformation in organisations has been arrived by the author after careful consideration of earlier research in this area, notably by (Kalakota & Robinson, 2002). Kalakota and Robinson (2002) have depicted the market evolution of m-business in five stages. They are: System Integration, Business Integration, Web enabled commerce, Web enabled transformation and Multi-device multi web transformation (M-business). The Kalakota and Robinson model depicts a roadmap towards M-business. However, that model misses out on the rigours of identifying and analysing where the organisation is currently standing in terms of m-transformation. This is partly due to the timing of the research since this road map was introduced in 2002 when m-business was at very early stages. (Lan & Unhelkar, 2005) have also considered a Globalization transition road map (GET) for organizational transitions.
Figure 1 depicts the Scale of four stage in m-transformation. The model identifies the stage of an organisation with regard to its current usage of mobile technologies.

Model (1)

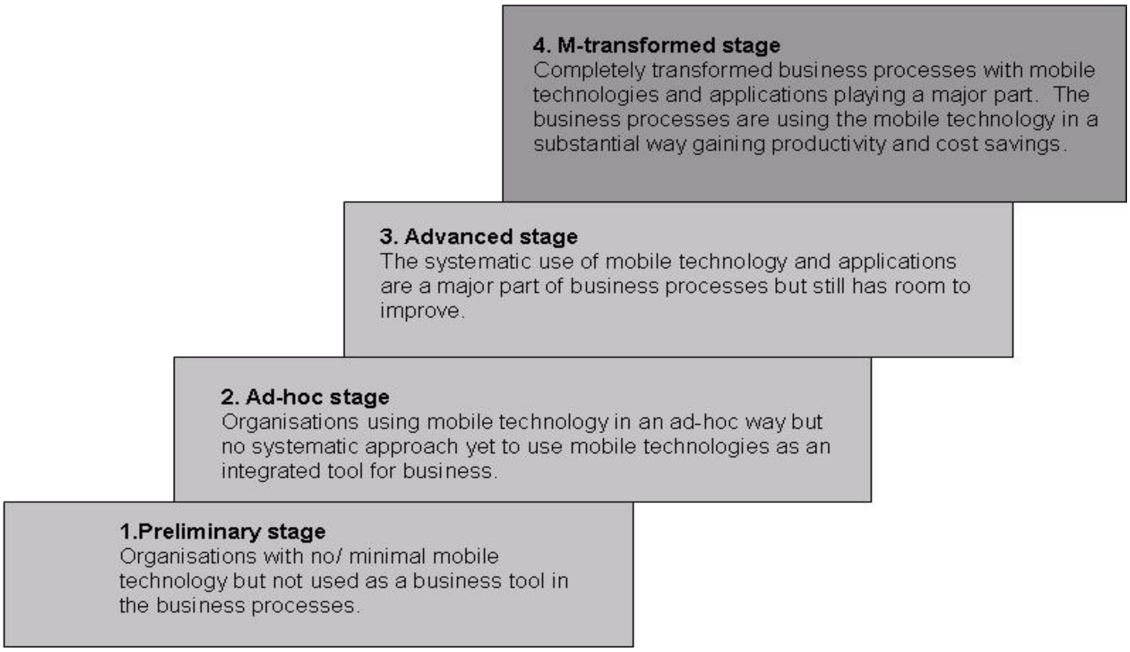


Figure 1 – Scale of four stages in an organisation in m-transformation

Once the stage at which the organisation is currently standing, with regard to m-transformation is identified, then the organization could look at the business processes that needs transformation in order to use mobile technologies. Model (2) proposes a model to identify and action the m-transformation process thereafter.

Figure 2 depicts the mobile transition roadmap, a generic model to transform a business organisation into a mobile enabled organisation. This roadmap idea has been adopted from the electronic transformation roadmap (Ginige et al., 2001) and a case study carried out in a leading bank in Australia studying the business processes and functionalities of a business unit.

Model (2)

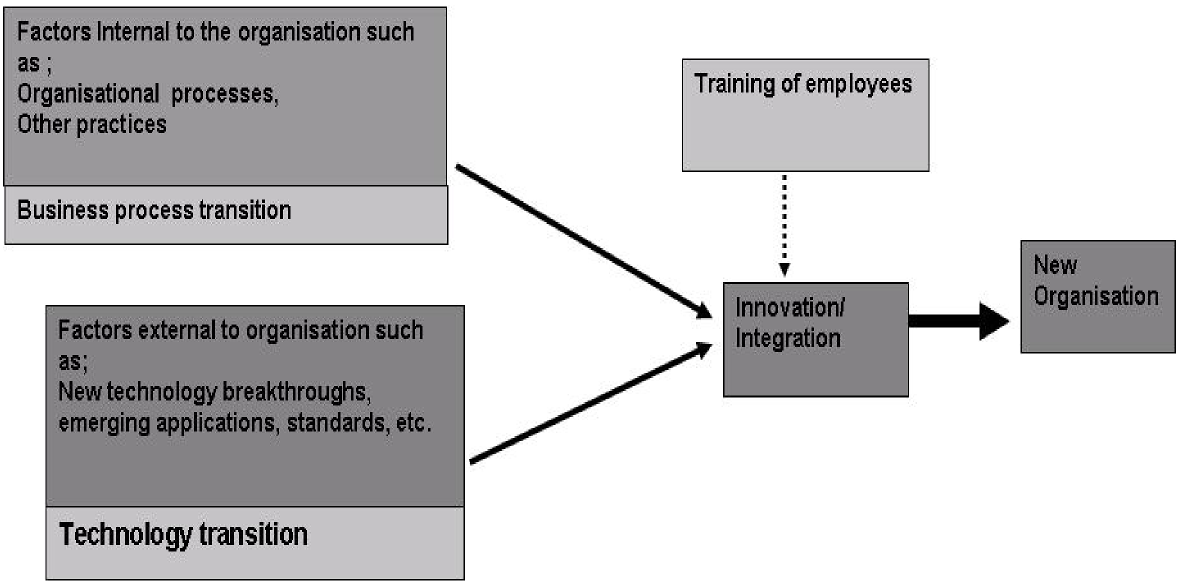


Figure 2: Mobile transition roadmap (Adapted from Ginige et. al, 2001)

The actual change in the organization would be a more holistic change. This change would be looked at in terms of different perspectives namely : The technology perspective, the methodology perspective and the social perspective. The model deals with all three of these perspectives when m-transforming the business processes of an organization.

RESEARCH GOALS

The next research goal will be to verify the validity of these models in order to apply them in large scale to business organizations. A survey has been designed and awaiting clearance for this verification. Once the survey forms are collected and analysed, the models could be verified so that they could be directly applied to business organizations as generic models to identify and transform there business processes. The survey will be based on the following hypothesis and the research questions.

RESEARCH QUESTIONS AND HYPOTHESES

The main research question under investigation is:

“Does m-transformation of business processes enhance the competitive delivery of service in business organizations?”

In order to probe into this research question we use the hypothesis;

H1 : M-transformation of business processes enhance the delivery of service in business organizations.

To establish the above hypotheses, there are several leading research questions to be answered. The following questions were considered when a banking case study was done in order to identify the m-transition process (Arunatileka, 2006).

- Which stage does the organisation stand in terms of m-transformation?
- How could we apply the m-transition road map at this stage?
- What is the best possible method to apply m-transition road map? Is it organisation wide, Department wide or business unit wide?
- What are business processes that could be transformed in the units already identified?
- What are the changes happening in the industry where the organisation belongs to in terms of this unit?
- What is the expected impact, after mobile technology is introduced to selected business units in the organisation?
- What would be the direct impact on the customers once the new technology is fully implemented?
- Would there be any anticipated problems during the changeover?
- What are the cultural issues that should be considered?

By considering these questions the mobile transition road map is applied by identifying the internal business processes and the external enablers in order to systematically transform the organisation.

CONCLUSION AND FUTURE WORK

The immediate action in the future would be to get the survey across to a sizeable sample of organizations and collect their data. Then the organizations could be categorized as per their current level of mobile technology. M-transformation road map could then be applied to individual organizations in order to systematically change the relevant business processes. The research questions discussed in the previous chapter is used to run this process smoothly. The road map will enable the organization to follow a smoother transformation which would have considered the technology, methodology and sociology perspectives of the organization in a full scale.

REFERENCES

Arunatileka, D. (2006). Applying Mobile Technologies to Banking Business Processes. In B. Unhelkar (Ed.), *Mobile Business : Technological, Methodological and Social Perspectives*. New York: Idea Group Publishing.

Arunatileka, D., & Unhelkar, B. (2003). *Mobile Technologies, providing new possibilities in Customer Relationship Management.* Paper presented at the 5th International Information Technology Conference, Colombo, Sri Lanka.

Arunatileka, S., & Arunatileka, D. (2003, December, 2003). *eTRANSFORMATION AS A STRATEGIC TOOL FOR SMEs IN DEVELOPING NATIONS.* Paper presented at the 1st International Conference on E-Governance, New Delhi, India.

Chung, J.-Y., Lin, K.-J., & Shim, S. S. Y. (2005). Beyond Electronic Commerce. *Computer, 38*(8), 92-93.

Ginige, A., Murugesan, S., & Kazanis, P. (2001). A Roadmap for Successfully Transforming SMEs in to E-Businesses. *Cutter IT, 14*.

Kalakota, R., & Robinson, M. (2002). *M-Business : The Race to Mobility*. New York, USA: McGraw Hill Professional.

Lan, Y., & Unhelkar, B. (2005). *Global Enterprise Transitions : Managing the Process*. USA: IDEA Group Publishing.

Sawhney, M., & Zabin, J. (2001). *The Seven Steps to Nirvana*: McGraw Hill.

Zureik, E., & Mowshowitz, A. (2005). Consumer power in the digital society. *Communications of the ACM, Volume 48, Number 10*, Pages 46-51.

The Organisation of Information

Eric Rayner
University of Wollongong
Wollongong, Australia
Email: epr01@uow.edu.au

Abstract

Existing literature does not provide a formal model of information with demonstrable utility with respect to the organisation, retrieval and presentation of information. In order to achieve this, we define a number of key concepts, including information space, knowledge libraries and restricted set space. We present a formal model of information based on restricted set space—a more general variant of metric space. We compare and contrast this model with the relational model of data. This model will form the basis of knowledge libraries that offer significant advantages over existing applications.

Keywords:

Information space, knowledge library, *N*-space, set space, metric space, Minkowski equation, Hausdorff metric.

INTRODUCTION

After discussing the current state of knowledge and identifying a gap in the literature, we outline our approach to modeling information. Next, we present the essentials of a new model of information, which we compare and contrast with the relational model of data. We conclude with a discussion of our past, present and future research.

CURRENT STATE OF KNOWLEDGE

There exist a variety of methods to generate reliable information, including: the scientific method; simulation and logical inference.

'Information theory' (Shannon, 1948) has found numerous applications within information science. However because it only discusses the transmission of information, it is a small piece of the information puzzle.

Information retrieval models such as the probabilistic model (Robertson & Jones, 1976) and the vector space model (Salton, 1971; Salton & Lesk, 1968) are really just models of the process of information retrieval when documents are indexed by keywords. Existing information retrieval systems suffer from:

- information islands—which occur when the index terms known only to those who already have access to the information;
- info-gluts—which occur when the index terms are ambiguous;
- poor effective indexing resolution—which occur when the information desired by the user can not be well described by the retrieval language available.

Classification schemes such as Dewey Decimal Classification (Dewey, 1876; OCLC, 2005), Universal Decimal Classification (Otlet & Fontaine, 1907; Stevens, 2005), Bibliographic classification (Bliss, 1950-1953; Curwen, 2004) and Colon classification (Ranganathan, 1962), provide schemes to arrange documents on library shelves according to hierarchical subdivisions of the 'universe of knowledge'. These subdivisions frequently have cultural, rather than universal, justifications.

Formal concept analysis (Priss, 2003) is a data analysis technique that organises a set of 'objects' with 'attributes' into a lattice of 'formal concepts'. This lattice is of limited practical use as a classification structure, making formal concept analysis, in this respect, more a theoretical curiosity than a practical tool.

Models of data such as the relational model (Codd, 1970) give low level models of information without addressing the question of what information is.

The notion of 'memes' (Dawkins, 1976) gives insight into the nature of information and how it may evolve. This view of information suggests we can model the spread of information using existing epidemiological models.

Table 1 overleaf lists the perspectives of information discussed here and relates each perspective with its unit of information.

Perspective	**Unit of Information**
Information Theory	Bits
Information Retrieval	Documents
Library Classification	Papers, books
Formal Concept Analysis	Objects with attributes
Data Models	Atomic values or data (numbers, letters, words)
Evolutionary	Memes

Table 1. Perspectives of information and the units of information used.

None of these approaches attempt to provide a formal model of information with demonstrable utility with respect to the organisation, retrieval, and presentation of information.

OUR APPROACH

We treat information as a network—the nodes and edges of which are also information. This treatment fits in well with constructs such as the WWW, and its inspiration, the memex (Bush, 1945). It also fits in closely with what we know about the structure of the brain.

Our challenge is not to simply construct a model of information, but to construct a model that has demonstrable utility in respect to organising, relating, retrieving and presenting information. Our approach to this problem is to use a spatial metaphor.

Our aim is to model *information space*, by which we mean a system (analogous to physical space) for organising, relating, retrieving and presenting information. Information space:

- defines relations between the *information elements* it contains and
- describes how users may navigate the space and select information elements.

Models of information based on information space will provide the necessary theoretical underpinning for a new type of system that satisfies needs not well served by existing systems.

Definition: A *knowledge library* (KL) is a system for the creation, storage, modification, querying, and display of sets of information elements.

- KLs inform users of the content (i.e. the elements) of the library and the relations that exist between elements in the library.
- KLs allow users to select a (possibly empty) subset of the library containing elements that satisfy arbitrary requirements specified by the user.

Some examples of potential knowledge library sets are: academic papers, company reports, product fact sheets, letters, algorithms, chess games, web sites, and encyclopaedia entries.

Hypothesis: A formal model of information, based on information space, can be used as the basis for a knowledge library that facilitates the organisation, retrieval and presentation of information—offering significant advantages over existing applications.

MODELLING INFORMATION SPACE

We develop a model of information from a particular type of information space, which we call *N-space. N-space* is an *N*-dimensional information space where each point can be completely described in terms of *N* coordinates—one coordinate from each dimension. A region in *N*-space is a set of points. Information elements can be attached to *N*-space by associating them with regions. *N*-space provides a navigable space where relations between information elements can be discussed in terms of distance and direction.

Our formal model of *N*-space is based on a variant of metric space.

Definition: A *metric space* is a pair $\langle \mathrm{A}, d \rangle$. The *metric* $d : \mathrm{A} \times \mathrm{A} \to \Re$ is a function where, for each $x, y, z \in \mathrm{A}$:

- if $x \neq y$, $d(x, y) > 0$ (strict positiveness);
- $d(x, y) = d(y, x)$ (symmetry);
- $d(x, x) = 0$ (reflexivity);
- $d(x, y) \leq d(x, z) + d(z, y)$ (the triangle inequality).

Satisfying the triangle inequality guarantees the important spatial characteristic that the direct distance between two points is never greater than the total distance along a path that includes one or more intermediate points.

The problem of searching in metric spaces has been explored in (Chavez, Navarro, Baeza-Yates, & Marroquin, 2001).

In order to model *N*-space we must generalise metric space in a number of ways.

- Because we need to find distances between information elements that are attached to more than one point in the space, we require distance measures between pairs of *sets* of points, as well as just pairs of points.
- Because we may wish to model one-way relations, we need to: relax the symmetry constraint, allowing $d(x, y) \neq d(y, x)$ for $y \neq x$.
- Because extra properties of an information element that satisfies our requirements are irrelevant, we extend reflexivity to subsets and weaken the strict positiveness requirement.
- Finally, because we are now dealing with distances between sets of points, rather than just points, we need to generalise the triangle inequality so that it provides a sensible constraint for distances between sets.

The triangle inequality often does not hold for distances between pairs of sets of points when this distance is given in terms of a minimum distance between points in each set. Fortunately, because the *span* of the intermediate set is taken into account, the generalised triangle inequality, given in the definition below, is more easily satisfied.

Restricted set space, defined below, generalises metric space in these ways.

Definition: A *restricted set space* is a pair $\langle \mathrm{A}, d \rangle$. Each element of the set A is itself a set. The *restricted set distance* $d : \mathrm{A} \times \mathrm{A} \to \Re$ is a partial function for which, for each $x, y, z \in \mathrm{A}$:

- if $x \subseteq y$, $d(x, y) = 0$ (subset reflexivity);
- if $x \not\subseteq y$, $d(x, y) > 0$ (restricted strict positiveness);
- $d(x, y) \leq d(x, z) + s(z) + d(z, y)$ (the generalised triangle inequality).

Here the function $s : \mathrm{A} \to \Re$. For each $z \in \mathrm{A}$, $s(z) = \max\{d(a, b) | a, b \in z\}$. We call $s(z)$ the *d span* of z.

In this paper ϕ is the empty set. Note that, if $x \in \mathrm{A}$, $\phi \subseteq x$ and so $d(\phi, x) = 0$ and for $x \neq \phi$, $d(x, \phi) > 0$. Also, $x \subseteq x$ so $d(x, x) = 0$ (reflexivity). Furthermore, if z is a singleton set then, due to reflexivity, $s(z) = 0$ and the generalised triangle inequality can be written $d(x, y) \leq d(x, z) + d(z, y)$ (the triangle inequality). Finally, for the triangle inequality to be satisfied, it must be the case that, for each $x \in \mathrm{A}$, $d(x, \phi) \geq \max\{d(x, y) | y \in \mathrm{A}\}$.

We use restricted set space as the basis for models of N-space. We may be able to make use of an ordering of points in A to simplify construction of the function d for a dimension $\langle \mathrm{A}, d \rangle$ of N-space.

Definition: Given a binary relation R on a set A, a *path* is a sequence, with n>1, $x_1, x_2, ..., x_n$ such that $x_1Rx_2R...Rx_n$. Given such a path we say that x_1 is *before* x_n in R and that x_n is *after* x_1 in R. Furthermore, given a function $d : \mathrm{A} \times \mathrm{A} \to \Re$, the d length of the path is $\sum_{i=1}^{n-1} d(x_i, x_{i+1})$. In any set of paths, a d shortest path is a path with the least d length.

Definition: A *networked space* is a triple $\langle \mathrm{A}, d, R \rangle$. Each element of the set A is itself a set, $d : \mathrm{A} \times \mathrm{A} \to \Re$, R is a binary relation and, for each $x, y \in \mathrm{A}$:

- if $x \subseteq y$, $d(x, y)$= 0;
- if xRy, 0 < $d(x, y)$ < the d length of any path $xR...Ry \neq xRy$;
- if $x \not\subseteq y$ and $\neg xRy$ and x is before y in R, $d(x, y)$ = the d length of a d shortest path $xR...Ry$;

Networked spaces based on lists, trees, or graphs are suitable for forming information space dimensions.

We give the central definition of our formal model of information space below. In this definition, the function G_p^N generalises the Minkowski metric

$$L_p(\langle y_1,...,y_N \rangle, \langle z_1,...,z_N \rangle) = \left(\sum_{i=1}^{N} | y_i - z_i |^p \right)^{1/p}$$

in two ways. First, G_p^N is based on the *distance*, rather than the difference, between y_i and z_i. Second, G_p^N allows us to assign a weight w_i to each dimension. Both G_p^N and L_p generalise Euclidian distance which, for G_p^N, is obtained when each w_i = 1, each $d(y_i, z_i) = |y_i - z_i|$, and p = 2.

Definition: An Np-networked space, is a pair $\langle \mathrm{A}, G_p^N \rangle$ where:

- $\mathrm{A} = \{\langle a_1,...,a_N \rangle \mid a_1 \in \mathrm{A}_1,...,a_N \in \mathrm{A}_N\}$;
- for each $1 \leq i \leq N$: $\langle \mathrm{A}_i, d_i \rangle$ is a networked space, $w_i \geq 0$ and $y_i, z_i \in \mathrm{A}_i$.
- $p \geq 1$;
- $G_p^N : \mathrm{A} \times \mathrm{A} \to \Re$;
- $G_p^N(\langle y_1,...,y_N \rangle, \langle z_1,...,z_N \rangle) = \left(\sum_{i=1}^{N} w_i \cdot d_i(y_i, z_i)^p \right)^{1/p}$.

All that remains, to cover the essentials, is to show how we can determine the distance between pairs of sets of points, rather than just pairs of points. In order to do this we generalise the distance metric. In other words, given an Np-space $\langle \mathrm{A}, d \rangle$, where A is a set of N-tuples, called points, we generate a restricted set space $\langle \mathrm{P}, e \rangle$, where P is the power set of A.

One likely definition for the function e is, for each $X, Y \in$ P, $e_1(X, Y) = \max\{\min\{d(x, y) | y \in Y\} | x \in X\}$. Note that $e_2(X, Y) = \max\{e_1(X, Y), e_1(Y, X)\}$, a symmetric version of e_1, is the Hausdorff metric. One problem with this metric is that a single discrepancy in otherwise identical large sets X, Y, may result in a large value for $e_2(X, Y)$ (Dubuisson & Jain, 1994).

If $X = \{x_1, x_2, \ldots, x_n\}$ then $e_3(X,Y) = \sum_{i=1}^{n} \min\{d(x_i, y) \mid y \in Y\}/n$, proposed in (Dubuisson & Jain, 1994) as part of a technique to compare digital images, is another distance function that is designed to overcome this problem. This function has a number of properties that make it quite suited to our purposes. While e_3 is not symmetric and does not satisfy the triangle inequality and so is not a metric, it can be shown that, because it satisfies the generalised triangle inequality and other requirements, e_3 can be used as a restricted set distance function.

Similarly, it can be shown that $e_4(X,Y) = \min\{d(x,y) \mid x \in X, y \in Y\}$, which also has some useful properties, can also be used as a restricted set distance function.

N-SPACE AND THE RELATIONAL MODEL OF DATA

As models of information, *N*-space has a number of similarities to the relational model of data. For example:

- *N*-space is an extension of the concept of *n*-ary relations;
- points in *N*-space are *n*-tuples;
- dimensions of *N*-space correspond to domains (in the relational model);
- dimensions that are *N*-spaces correspond to nonsimple domains (in the relational model);
- describing and selecting information elements belonging to regions in *N*-space corresponds to symmetric exploitation (in the relational model);
- the concept of subspaces of *N*-space is equivalent to projection (in the relational model).

The most significant differences between *N*-space and the relational model are:

- the relational model has no equivalent to distances between points in *N*-space;
- the relational model has no equivalent to networked dimensions in *N*-space;
- while points in *N*-space are merely *associated* with information elements, *n*-tuples *are* the data in the relational model.

This last point underlines what is really the main difference between *N*-space and the relational model of data. The relational model models the information—in the form of data—itself, while *N*-space is content to classify and relate information. Of course the information elements in an *N*-space may themselves be *N*-spaces. It may even be feasible to present *all* information in *N*-space form.

COMPLETED AND FUTURE INFORMATION SPACE RESEARCH

Completed research

We have completed a thorough and extensive review of the relevant literature. This process, in itself, was instrumental in identifying some of the shortcomings of existing information retrieval and organisation technology. We have identified a gap in the literature and formulated our hypothesis.

We have developed a formal model of information, including a number of definitions, theorems and proofs. As part of the process of developing this model we have proposed a number of generalisations to important mathematical ideas such as: the triangle inequality; metric space; the Minkowski metric; and the Hausdorff metric. These generalisations may have applications in other areas.

We have designed a computer interface to *N*-space and built an experimental, small scale, desktop application. As a result of this experiment we are confident our research will be successful.

We have presented various aspects of our research at:

- the University of Wollongong School of Information Technology and Computer Science colloquium series;
- the University of Wollongong Pure Mathematics seminar series;
- the University of Wollongong higher degree research student conference and
- the IEEE Tencon 2005 regional conference.

We have published (Rayner, 2005; Rayner, Piper, & Bunder, 2005) and have further papers in the final stages of preparation.

Research to be completed in the near future

We are working on a number of projects that will be completed in the near future (i.e. before PhD completion). This includes:

- further research papers;
- an application programming interface for *N*-space knowledge libraries;
- extending the formal model to cover uncertainty.

Ongoing research

One of the main reasons for the success of information retrieval based on keyword indexing is the ease with which documents can be automatically indexed. We believe that techniques can be developed to allow automatic indexing of documents in keyword ontology dimensions in *N*-space. By disambiguating keywords by context in documents and dimensions, this technique should solve a number of problems inherent in the existing information retrieval methodology.

Practical *N*-space knowledge library applications will necessarily include various levels of security. Because *N*-space knowledge libraries organise, relate, retrieve and present information, *N*-space security models may provide generic information security solutions.

N-space provides new possibilities in the way we interact with information. In particular, graphical representations of *N*-space can be developed that will provide users with rapid overviews of large complex collections of information, visually bringing to life metaphors such as:

- information density;
- holes or gaps in information and
- frontiers of knowledge.

As a novel and powerful information model, we believe *N*-space can offer potent alternatives to traditional paradigms in teaching, learning, research and commerce. In particular, *N*-space could help us to determine, and appreciate, the value of information.

REFERENCES

Bliss, H. E. (1950-1953). *A bibliographic classification. Vols. 1-4*. New York: H. W. Wilson.

Bush, V. (1945). *As We May Think*, from http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush-all.shtml

Chavez, E., Navarro, G., Baeza-Yates, R., & Marroquin, J. L. (2001). Searching in Metric Spaces. *ACM Computing Surveys, 33*(3), 273-321.

Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM, 13*(6), 377-387.

Curwen, T. (2004). *Bliss Classification Association*, 2005, from http://www.sid.cam.ac.uk/bca/bcahome.htm

Dawkins, R. (1976). *The selfish Gene*. Oxford: Oxford University Press.

Dewey, M. (1876). *A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library*. Amherst.

Dubuisson, M.-P., & Jain, A. K. (1994). *A Modified Hausdorff Distance for Object Matching.* Paper presented at the International Conference on Pattern Recognition, Jerusalem, Israel.

OCLC. (2005). *OCLC - a worldwide library cooperative: Dewey services*, from http://www.oclc.org/dewey/

Otlet, P., & Fontaine, H. L. (1907). *Manuel du Repertoire Bibliographique Universel*. Brussels: Institut International de Bibliographie.

Priss, U. (2003). *A Formal Concept Analysis Homepage*, 2005, from http://www.upriss.org.uk/fca/fca.html

Ranganathan, S. R. (1962). *Elements of library classification*. Bombay: Asia publishing house.

Rayner, E. (2005). Searching for the Science in Information Science. *Rhizome: A cross-disiplinary postgraduate journal, 1*, 215-228.

Rayner, E., Piper, I., & Bunder, M. (2005). *Introducing N-Tree Space: A Classification System for Knowledge Library Applications.* Paper presented at the IEEE TENCON, Melbourne.

Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Sciences, 27*(3), 129--146.

Salton, G. (1971). *The SMART Retrieval System--Experiments in Automatic Document Processing*. Englewood Cliffs, N.J.: Prentice Hall.

Salton, G., & Lesk, M. E. (1968). Computer Evaluation of Indexing and Text Processing. *Journal of the ACM, 5*(1), 8-36.
Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*, 379-423, 623-656.
Stevens, A. (2005). *UDC Consortium*. Retrieved 17 October, 2005, from http://www.udcc.org/

Critical Success Factors for the Implementation of Business Intelligence System in Engineering Asset Management Organisations

William Yeoh
School of Computer and Information Science,
University of South Australia
william.yeoh@unisa.edu.au

Abstract
Today's asset management organisations (AMOs) thrive to turbulence. Increased competition has forced them to continuously drive more productivity out of their assets, while enormous information has been generated from their disparate information systems for managerial works. In response to these growing pressures, AMOs are urged to improve their business execution and decision support through the implementation of business intelligence (BI) system - which is a complex undertaking, involves an immense expenditure and a variety of stakeholders over a period of time. Moreover, BI is a relatively new area that has primarily been driven by IT industry; it appears there has been little academic research both theoretically and empirically in BI systems implementation issues. Therefore, this research seeks to contribute to the understanding of critical success factors (CSFs) affecting successful BI system implementation. It aims to investigate BI system implementation process, and identifies the CSFs influence on BI system implementation, particularly in AMOs. Furthermore, the research will explore the inter-relationships between these key factors, and broader contextual and process issues. As this study is of exploratory nature, there is a need moving in a nonlinear research path, and modify the hypothesised CSFs based on real-life practice. Hence this research adopts qualitative positivist approach and case study method.

Keywords:
Critical Success Factors, Business Intelligence System, Asset Management

INTRODUCTION

Today's asset management organisations (AMOs) have been generating more information than ever before. Many business segments have been computerised within asset management and in a variety of ways; meanwhile increased competition has forced them to continuously drive more productivity out of their assets. For example, maintenance management, registry management, operation management, reliability assessment, work management, risk management, inventory control and contract management. According to Gartner Research (Steenstrup 2004), asset intensive industries spend more than half of their operation budget and hence, management attention in managing assets lifecycle information.

Although, these business automation initiatives are geared at providing enhanced value and efficiency to the business; nevertheless, issues like lack of interoperable data architecture, disparate asset managing and administrative systems, and a complementing organisational issues result in the inability of organisational stakeholders to fully utilise information for optimum advantage of the business. In other words, they are suffering from the inability to properly and appropriately exploit the critical managerial information from their BI systems. Therefore, in order to provide reliable information to the overall business execution and decision support, the importance of BI system cannot be taken too lightly.

However, the configuration and implementation of BI system is a complex undertaking, always involves an immense expenditure and various stakeholders including project leader, data architects, system administrators, business analysts, external consultants, BI vendor and key BI users of all management levels over a period of time. Moreover, project managers often tend to focus on the financial and technical aspects, and hence neglect the non-technical perspectives such as organisational issues. In order to overwrite this biased perspective, this study applied a typical approach used to assure and define the success of an information system implementation - critical success factors approach (as used by Esteves & Pastor, 2000; Holland & Light, 1999; Parr et al, 1999; Somers & Nelson, 2001; Williams & Ramaprasad, 1996).

As BI is essentially driven by business and IT industry, very little research both theoretically and empirically on the BI system implementation issue has been done. Hence, this research seeks to contribute to the understanding of CSFs affecting successful BI system implementation. It aims to investigate BI system implementation process, and identifies the key factors influence on BI system implementation, particularly in AMOs. Furthermore, the research will explore the inter-relationships between these key factors, and broader contextual and process issues.

RESEARCH QUESTIONS

The absence of stated key factors for the successful implementation of BI system has led to growing frustration among BI practitioners and users. The BI system implementation process is complex, usually requiring simultaneous attention to a wide variety of human, organisational, budget, technical, project management, cultural and external variables. Therefore, this research attempts to resolve the following overarching questions:

- How the 'success' of a BI system implementation can be determined?
- What are the most critical factors that influence the success of BI system implementation in AMOs?
- What are the inter-relationship between these critical success factors, and broader contextual and process issues?
- Which recommendations and guidelines can be given to BI practitioners and users to achieve successful BI system implementation?

PRELIMINARY LITERATURE REVIEW

This section of the research proposal is an overview of the body of work in the research context and outlines the relevance to the thesis. The following figure 1 shows the different key areas that should be considered in order to answer the research questions.

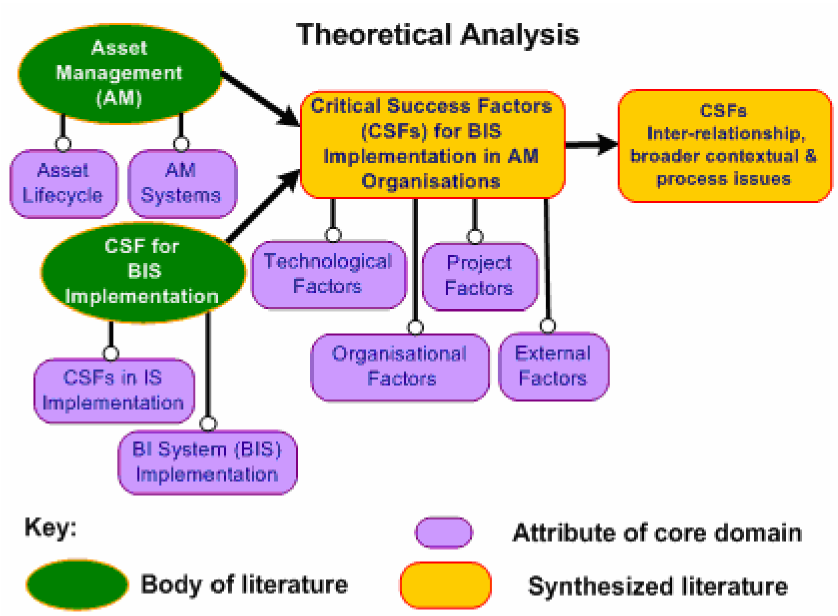


Figure 13: Overview of Key Research Area

ASSET MANAGEMENT

Recently, prudent and effective management of engineering assets is an issue that has gained significant momentum in Australia. Increased competition has forced AMOs to continuously drive more productivity out of their industrial assets. In other words, asset management is moving to the forefront contributing to an organisation's business and financial objectives. Moreover, there is a large degree of inter-dependency among these asset networks; a malfunction within a network may undermine the performance of other networks (IIMM, 2002). Yet little systematic analysis has been done to obtain reliable information about the performance of out-of-sight assets to support business decisions. For example, the water supply scare crisis in Sydney and South Australia were obvious failure evidences to deliver adequate services.

Asset Management (AM) is a broad term and has multi-facets of definitions. The nature of definition really depends on individual business considers what to be an asset. However, a common definition of AM is "the coordinated activities and practices through which an organisation optimally manages its assets and their associated performance, risks and expenditures over their lifecycle for the purpose of achieving its organisational strategic plan" (BSI, 2004). In other words, AM aims to meet a required level of service in the most cost-effective way through its entire lifecycle initiating from assets needs definition, planning, acquiring, installing, operation, maintenance to retiring stage. Effective AM optimises asset utilisation, increases output, maximises availability, and lengthens asset lifespan, while simultaneously maintaining lowest long-term cost rather than short-term savings in order to reap the maximum value.

Besides that, Eerens (2003) indicates that it is impossible to optimise asset lifecycle profits if the various business functions do not synchronise or align with each other. This requires integration and synchronisation of management plans and practices at all levels from asset needs definition to asset retiring stage. In order to achieve that, Bever (2000) suggest that AMOs should focus on two main arenas. First, they need to treat assets as core business enabler and manage assets from a strategic perspective rather than seeing them as secondary supporter to business functions, and second, there should be a focused approach on asset management and its interrelationships with other processes and systems. As a result, the asset operation, maintenance, human resource, supply chain, financial and other business functions are considered inter-dependent and mutually supportive towards the achievement of the overall enterprise goals.

Besides that, Haider and Koronios (2003) point out the intricacies of effective asset management. As they noted, there are silos of operational and administrative systems work in isolation, which not only control and monitor the assets operation and maintenance but also provide administrative support throughout the asset lifecycle. In practice, data and thus information are collected both manually and electronically, in an assortment of formats, and stored in various customised and off the shelf legacy systems that are deficient in interoperability. Most importantly, there has always been a degree of difficulty to which critical business information can be accessed appropriately, due to a lack of interoperable data architecture, disparate asset lifecycle management and administrative functions, and the complementing organisational and people issues. In other words, the methodologies employed to develop these systems are to increase productivity alone, and very much functional oriented.

After all, the process of asset management is sophisticated because it is a complex and integrated function that requires substantial information to be collected from many different parts of the organization. This stored information will be extracted, transformed and loaded to data warehouse or data mart for further data mining and business analysis purposes. The asset management engineering and planning process uses this information to provide enhanced decision support in the planning and scheduling of asset maintenance, rehabilitation, and replacement activities. Thus the effective and efficient extraction, analysis and provision of quality information are critical to improve the bottom line of AMOs.

BUSINESS INTELLIGENCE (BI) SYSTEM

BI systems, as Negash and Gray (2003) explain, are the “combination of data gathering, data storage, and knowledge management with analytical tools to present actionable information to decision makers”. With a well-planned BI system, it can transform the large, evolving, disparate data silos that have previously been a ubiquitous part of asset management into crucial business information through appropriate data mining and analysis process. Furthermore, information gap between the collection and analysis process can be identified and improved, while overlapping effort can be greatly reduced.

BI systems come as standardised software packages from vendors like Business Objects, Cognos, IBM, Informatica, Information Builders, Oracle and SAS, allowing customers to adapt them to their specific requirements, rather than having to develop in-house complex software applications. In fact, it has been a preferred method for replacing proprietary decision support systems such as executive information system and corporate performance management system. The power of the standard BI system solution lies in its ability to present the business a new and thorough insight across the entire range of its functions. In particular, critical information from many different sources can be integrated into a coherent body for strategic planning and decision making.

With the increasing advancement in data warehouse, web architecture, storage capability, data mining and analytical techniques, demand for BI applications continues to increase at an ever more explosive rate as the vendors continue to report substantial profits (Chen, 2002; Soejarto ,2003; Whiting, 2003). A Gartner survey found that more than 95 percent of the surveyed enterprises would support the use of BI systems, particularly for group-based decision making (Buytendijk, 2002). Gartner Research (Buytendijk et al., 2004) further forecasts that BI market will continue to grow and flourish through 2009. At the meantime, IDC Research expects that the BI market will enjoy a compound annual growth rate of 23%, and report US$3.3 billion dollars turnover in the Asia Pacific region alone (Giang, 2002).

This finding is echoed by Darrow's (2003) prediction that the worldwide market will be reaching US$12 billion by 2006.

However, although BI has been vigorously pushed by IT industry, it is a relatively new area and hence there has been limited research activity undertaken towards BI systems implementation issues. Moreover, the whole project planning and actual implementation of BI system always involves an immense amount of expenditure and a variety of stakeholders, including project leader, data warehouse architects, system administrators, business analysts, external consultants and key BI users spanning over a large period of time. Therefore an extensive academic research to explore the issues of CSFs and their inter-relationship for the successful implementation of BI system is of great importance. It can serve as primary guidelines for those who are considering, planning or evaluating the potential risks involved in adopting BI systems.

CRITICAL SUCCESS FACTORS (CSFS) FOR SYSTEMS IMPLEMENTATION

Pinto and Slevin (1987) defined CSFs as a set of function of $S = f(\chi_1, \chi_2, \ldots \chi_n)$ where S is a project success and χ_i is the i-th CSF. As this research aims to identify CSFs for a successful BI system implementation, thus the primary task is to define each χ_i. Moreover, Rockart (1979), who was the seminal author that applied CSF approach in information system research, further described CSFs as "the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization. If results in these areas are not adequate, the organisation's effort for the period will be less than desired". CSFs are vital areas, as Rockart (1984) observed, that top executive should pay constant attention for the project to flourish. In other words, CSFs are high-level management considerations, as distinct from a detailed set of project deliverable specifications. Hence, a set of CSFs identified for a large-scale project effort, like BI system implementation, is fundamentally different from the set of interlinked detailed tasks which must be accomplished satisfactorily to ensure a successful completion (Dobbins, 2000). As a result, the success of BI system implementation may not be fully ensured by the accomplishment of the required CSFs, but failure to accomplish the CSFs successfully will be a major deterrent to success.

As CSFs are simple to comprehend, document and monitor, they have been widely applied in information systems analysis and research (Fitzgerald & Carroll, 2003). Despite limited literature on BI system implementation issues, however, there are quite a large number of literature on the success issues and the necessary factors within system evaluation research and ERP implementations (such as Allen et al., 2002; Bancroft et al., 1998; Bingi et al., 1999; DeLone & McLean., 1992, 2003; Esteves & Pastor, 2000; Holland & Light, 1999; Koh et al., 2000; Parr et al., 1999; Pinto & Slevin, 1987; Somers & Nelson, 2001; Zhang et al., 2002). Although these literature are oriented towards general information system as well as ERP implementations; nonetheless, they form a theoretical foundation for this research to infer candidate key factors for successful BI system implementation.

Among all, Esteves and Pastor (2000) developed a unified model of CSFs for ERP system implementation by integrating relevant published works. They mapped these CSFs in a matrix model composing of four perspectives: organisational, technological, strategic and tactical. Their work is well-recognised by Somers and Nelson (2001) who also developed a CSFs model consisting of twenty two factors essential for ERP system implementation based on the responses from 86 organisations. Moreover, Parr et al (1999) have identified 10 necessary factors for successful ERP system implementations through the studies of 42 implementation projects and 10 in-depth interviews with ERP experts. Coincidentally, it is noted that many of Parr et al's findings are consistent with the studies of Bancroft et al (1998), whom list of factors was derived from three case studies and in-depth interviews with 20 practitioners.

Besides that, Pinto and Slevin (1987) argued that in order to ensure successful project implementations, project managers must equipped with both strategic and tactical capabilities in managing projects. In line with this, they developed a strategic-tactical framework that comprises ten CSFs serving as guidelines to project managers. As for strategic perspective, Holland and Light (1999) emphasised the importance for a project mission, sustained management support and a formal schedule outlining individual responsibility in the project implementation. There is no doubt that these issues are most critical at the outset of a project. Tactical aspects, on the other hand, play a more significant role as the implementation phase moves closer. They comprises open communication among stakeholders, balanced team with necessary technical and business knowledge, and underlying technological infrastructure.

However, a few researchers focus their lens on people issues. Among all, user satisfaction is most widely recognised as a criterion of successful system implementation (Benard & Satir, 1993; Chen, 2000; DeLone & McLean, 1992, 2003; Gelderman, 1998; Parikh & Fazlollahi, 2002). In particular, DeLone and McLean (2003) noted that user satisfaction is a success factor depending on the quality of information, system and service. Following this, user acceptance is deemed as another key requirement in the successful implementation of information systems. Numerous research have been undertaken to study the critical determinants of user acceptance (for example, Davis et al., 1989; Leung & Wong, 1997; Lowry, 2002; Wright & Papke-Shields, 2002). Moreover, according to a recent online survey of 544 IT professionals who are active in the use BI system, they have voted that user acceptance is the most important prerequisite leading to flourishing implementation (Information Builders, 2005).

On the other hand, there are substantial studies focusing primarily on organisational context too. In this context, support from top management is considered as the most crucial success factors (Bingi et al., 1999; Esteves & Pastor, 2000; Koh et al., 2000; Lawrence & Low, 1993; Parr et al., 1999; Steinbart & Nath, 1992; Somers & Nelson, 2001). In particular, Bingi et al. (1999) claimed that the successful implementation completely relies on the strong, persistent commitment of the top management. Besides that, Beath (1991) has undergone extensive study on the importance of a “champion” role in the implementation of an information system. He indicates that the champion, the project manager, plays the most pivotal role in ensuring the success of a system implementation. Furthermore, a balanced multifunctional team is well regarded as a key factor for many systems and ERP implementations success. This team consists of stakeholders with various skills from an assortment of backgrounds (Willcocks & Sykes, 2000). Mumford (1995, p. 192) indicates that the chosen team member must be fully “empowered” to achieve greatest “user involvement”. His assertion is echoed by Levin et al (1998), who believes that the empowerment of project team is of great importance to assure successful implementation.

Another critical factor is open communication among the stakeholders. In modeling CSFs for ERP implementation, Holland and Light (1999) noted that open communication plays a critical role to ensure successful implementation. Sarker and Lee (2000) confirmed through their case study, systems implementation like ERP can only be successful if there is open and honest communication among the stakeholders. Apart from that, Willcocks and Sykes (2000) indicate the importance of vendor support for successful ERP implementation. They maintained that the relationship between the software buyer and vendor should be strategic in nature. In their studies, Zhang et al. (2002) further classified this variable into three dimensions: service response time; qualified vendor with knowledge in both business process and IT; and active participation of vendor throughout the whole implementation process. However, little has taken into account the technological factors. Only Zhang et al. (2002) considered that the suitability of software and hardware has an impact on ERP implementation success.

RESEARCH CONTRIBUTION

As the study of BI system is a new area that has been primarily driven by business and IT industry; the academic research that directly reflects on CSFs for BI system implementation is very limited. Hence this proposed research will make both theoretical and practical contributions to the field of BI system and asset management in the following ways:

1) This research will make a contribution to the body of knowledge of BI system implementation and asset management by identifying the CSFs as well as their broader contextual and processes issues, and verifying their interrelationships that are essential for ensuring successful BI system implementation, particularly in asset management organisations.

2) This research will provide detailed recommendations that could enable BI project stakeholders not only to identify, but also to concentrate on the CSFs, especially in the planning of BI initiative. Such outcome will help them to improve the effectiveness and efficiency of their implementation activities, by obtaining a better understanding of possible obstacles that hinders the successful BI system implementation, and the organisation is more likely to have a more reliable BI system after its implementation.

3) This study will also present a detailed analysis of inter-relationships between those CSFs for further inquiry into the possible emergence of implications and series of effects that lead to project success, or failure. This is particularly important for project managers to predict the degree of success in the implementation of BI system through the control and monitoring of these CSFs.

RESEARCH LIMITATIONS

Due to research commitments, the scope of this research is primarily focussed on engineering asset management organisations in Australia, particularly within the participating engineering organisations of CRC for Integrated Engineering Asset Management (CIEAM). However, with the analysed results at the end of this research project, further studies in other industries or geographical regions are possible and indeed strongly encouraged. Also, the progress of this study depends on the motivation of relevant BI stakeholders; therefore it is important to explain to them on the envisaged benefits of this research effort and to get their highest degree support and cooperation.

RESEARCH METHODOLOGY AND APPROACH

This research follows a qualitative positivist approach, whereby the objective is to identify CSFs that lead to successful BI system implementation in AMOs. Qualitative approach is cyclical and towards a nonlinear direction (Neuman, 2000); and circularity enables the researcher to permanently reflect on the entire research process and on particular steps when needs be (Flick, 1998). It hence allows for exploration of nature of elements and thereby uncovers new elements through their interrelationships (Bussen & Myers, 1996). This research is hypothetic and assumes that there are critical factors impacting the BI system implementation success, particularly in assets managing business. It puts a methodical structure in the already existing knowledge for literature as well as practical world. Hence qualitative approach is best suited to this research as it provides further insight and understanding to the research body, especially for complex phenomena.

Within qualitative research, positivist research approach has been chosen to enhance the predictive understanding of phenomena, as the positivist focuses on testing of theory and deducing causal relationships that form the basis of the generalised knowledge. Most importantly, hypothesis and questions are proposed in advance to undergo empirical testing with carefully controlled and repeatable conditions. Through devising particular measures which detect the dimensions that are of interest to the researcher, the understanding of phenomena is essentially the modelling of the problem situation and its measurement (Orlikowski & Baroudi, 1991). In such situation, the researcher is an impartial observer, who carries out objective evaluation and makes predictions on processes, without forming any subjective judgement (Orlikowski & Baroudi, 1991). In line with this, the approach comprises evidence of formal propositions, hypothesis testing, quantifiable measure of variables, and the drawing of inferences about a phenomenon. This paradigm forms the base of assumption for external validity and the degree to which outcomes can be generalised (Hirschheim, 1992). The issue at hand is to explore how CSFs could enhance the degree of success for BI system implementation within AMOs; as it includes complexity of asset management and its associated influence on BI system development. Therefore, in the context of this research, it is a matter of forming theoretical CSFs, and their empirical deduction of generalised CSFs in the implementation of BI system for AMOs.

Among several methodologies available for qualitative positivist research, case study method is chosen to study empirically the contemporary phenomenon in its real-life context, particularly, when the boundaries between phenomenon and context are still at their early, formative stages (Yin, 1994). Case studies allow for a deeper investigation into a phenomenon by capturing detailed “reality” and thus reveal the highly complex, and unquantifiable events associated with that phenomenon. More importantly, it enables the study of many variables simultaneously without the need to determine them beforehand (Galliers, 1992). In addition, Yin (1994) asserts that it is especially useful in uncovering the ‘why’ and ‘how’ questions associated with a phenomenon. Walsham (1993) states that “the validity of an extrapolation from an individual case or cases depends on the plausibility and cogency of the logical reasoning used in describing the results from the case and in drawing conclusions from them”. Thus case study approach provides better explanations on the examined phenomenon as it allows for a rich description (Miles & Huberman, 1994), which would otherwise be lost in other quantitative designs. Moreover, in-depth explanation about the interrelationships between these CSFs, and broader contextual and process issues can be provided through multiple case studies. In exploratory research like this, as Benbasat et al (1987) and Cavaye (1996) recognised, case studies are deemed most suitable for building theory and refining concepts for further investigation. This research method is also consistent with other studies on the system implementation issues (such as Moynihan, 1993; Shanks et al., 1997; Watson et al., 1998).

In short, since the academic research on CSFs for successful BI system implementation in AMOs is limited, there is a critical need to examine the issues of CSFs by moving in a nonlinear research path, and modify the hypothesised CSFs based on real-life practice. A nonlinear research path can make successive passes through steps, sometimes moving backward and sideways before moving on (Neuman, 2000). Therefore, qualitative positivist approach and case study method seem most appropriate for this research.

CASE STUDY DESIGN

The purpose of the case study is to investigate the perceptions of AMOs stakeholders on the CSFs for BI system implementation, and to determine the empirical validity of the proposed CSFs concepts, leading to the identification of CSFs for successful BI system development. Essentially, the first case study will serve as pilot case study, whereby the case study protocol will be tested for validity and reliability. Based on the findings of pilot case study, necessary refinement will be carried out on both the questionnaire and proposed CSFs model. Subsequently, multiple case studies will be conducted to further investigate and verify the key factors for BI system implementation. Esteves (2004) recognised that the different organisational contexts offer valuable insights in the management of CSFs as well as implications resulted from the patterns of communality between those case studies. This methodology is also supported by Herriot and Firestone (1983), that the evidence gathered from multiple cases is considered more compelling and hence a more robust result can be produced. Numerous researchers have also applied case study method in their studies for critical factors (such as Akkermans & van Helden, 2002; Allen et al., 2002; Murray & Coffin, 2001; Sarker & Lee, 2000).

DATA COLLECTION

During investigation the following methods will be used for data collection.

a. Business reports,
b. Archival records,
c. Survey,
d. Direct observation, and
e. Interviews.

The documentation sources include annual reports, business documents such as position description, policy manuals, organisational structure charts, training documents, and archival records like BI project documents and presentation material for the implementation of BI project. Yin (1994) asserts that these documents can be used for the corroboration and augmentation of other evidence sources.

To facilitate the interview, a set of structured questionnaires will be developed, whereby interviewees can also propose their own factors which they deem important. The interviewees consist of relevant BI system stakeholders including BI system project leader, project members, system administrators, top management, key users and project consultants. Moreover, Personal Construct Psychology (PCP) will be used to elicit additional key factors during the interview. PCP is a constructivist system of psychology developed by George Kelly in 1950s and later refined by others (Boose & Bradshaw, 1987). It concerns about individual and a group psychological process, which suggests that experts function as anticipatory systems - they develop conceptual models so as to better understand and make predictions concerning their immediate world. By drawing distinction between elements, as Gaines and Shaw (1997) noted, the formation of theoretical constructs is based on every person's way of seeing relationships between objects. Moreover, with a solid theoretical foundation for knowledge acquisition and practical approach to knowledge modelling research, PCP is well-accepted to model the cognitive processes of human experts, hence it was considered appropriate to elicit the CSFs beliefs from those who are expert in BI system implementation. Also, it supports creative processes, and encourages divergence through mutation, constraint relaxation and such mechanisms (Ford & Bradshaw, 1993). In the last 40 years, PCP theory has been widely applied in various research areas including artificial intelligence, education, management, human computer interaction and knowledge acquisition research, it is thus considered appropriate for this study too.

PROJECT TIMETABLE

Aim	**Year 1**		**Year 2**		**Year 3**	
	1-6 months	7-12 months	1-6 months	7-12 months	1-6 months	7-12 months
Research Proposal	*Proposal submitted					
Ethics Proposal						
Literature Review		Draft	Review			Rewrite

Research Methods		Draft				Rewrite
Conference Paper		* 1 paper under review				
Pilot Case Study						
Refinement of Instruments						
Data Collection						
Data Analysis						
Journal Article						
Write Up Analysis						
Write Up Thesis						

Current stage. Next stage will be the development of structured questionnaire for interviews.

REFERENCES

Akkermans, H., & van Helden, K. (2002). Vicious and virtuous cycles in ERP implementation: a case study of interrelations between critical success factors. *European Journal of Information System*, (11), 35-46.

Allen, D., Kern, T., & Havenhand, M. (2002). ERP Critical Success Factors: an exploration of the contextual factors in public sector institutions. *Proceedings of the 35th Hawaii International Conference on System Sciences*, Hawaii: HICSS.

Bancroft, N., Seip, H., & Sprengel, A. (1998). *Implementing SAP R/3.(2nd ed.).* Greenwich: Manning Publications.

Beath, C. (1991). Supporting the Information Champion. *MIS Quarterly*, September, 355-371.

Benard, R., & Satir, A. (1993). User satisfaction with EISs: meeting the needs of executive users. *Information Systems Management*, 10(4), 21-29.

Benbasat, I., Goldstein, D., & Mead, M. (1987). The Case Study Research Strategy in Studies of Information Systems. *MIS Quarterly (11)*, 369-386.

Bever, K. (2000). Understanding Plant Asset Management Systems. *Maintenance Technology,* July/August, 20-25.

Bingi, P., Sharma, M., & Godla, J. (1999). Critical Issues Affecting an ERP Implementation. *Information Systems Management*, 7-14.

Bussen, W., & Myers, M. (1996). Executive information system failure: A New Zealand case study. *Journal of Information Technology, (12),* 145-153.

Boose, J., & Bradshaw, J. (1987). Expertise Transfer and Complex Problems: Using AQUINAS as a knowledge Acquisition Workbench for knowledge-based Systems. *International Journal of Man-Machine Studies,* (26), 3-28.

Cavaye, A. (1996). Case study research: A multi-faceted research approach for IS. *Information Systems Journal,* 227-242.

Buytendijk, F. (2002). *Business Intelligence Networks: The Future of BI.* Gartner Research.

Buytendijk, F., Dresner, H., Linden, A. (2004). *Hype Cycle for Business Intelligence*. Gartner Research.

British Standards Institution (BSI). (2004). *PAS 55-1 Asset management part 1: Specification for the optimised management of physical infrastructure assets*. London: British Standards Institution.

Chen, E. (2002). *The 2003 IT Budget Pinch: Belts Get Pulled Even Tighter.* Retrieved September 22, 2005, from eWeek website: http://www.eweek.com/article2/0,3959,547545,00.asp

Chen, L., Soliman, K., & Frolick, M. (2000). Measuring user satisfaction with data warehouses: an exploratory study. *Information and Management*, 37(3),103-110.

Darrow, B. (2003, Feb 3). Making the Right Choice - Solution providers are evaluating a plethora of options as they puzzle over the future of business intelligence. *Computer Reseller News.*

Davis, F., Bagozzi, R., & Warshaw, P. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Model. *Management Science*, August, 982-1003.

DeLone, W., & McLean, E. (1992). The Quest for the Dependent Variable. *Information System Research*, 60-95.

DeLone, W., & McLean, E. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19, 9-30.

Dobbins, J. (2000). *On a Generalized CSF Process Model for Critical Success Factors Identification and Analysis*, PhD Thesis, George Washington University.

Eerens, E. (2003). *Business driven asset management for industrial & infrastructure assets*. Victoria, Australia: Le Clochard.

Esteves, J. (2004). *Definition and Analysis of Critical Success Factors for ERP Implementation Projects,* PhD Thesis, Technical University of Catalonia.

Esteves, J., & Pastor, J. (2000). Towards The Unification Of Critical Success Factors For ERP Implementations, *Proceedings of 10th Annual Business Information Technology (BIT) Conference*: Manchester.

Fitzgerald, L., & Carroll, J. (2003). The Role of Governance in ERP System Implementation. *Proceedings of the 14th Australasian Conference on Information System:* Perth.

Flick, U. (1998). *An introduction to qualitative research*. Thousand Oaks, CA: Sage.

Ford, K., & Bradshaw, J. (1993). Knowledge Acquisition As Modeling. *International Journal of Intelligent Systems,* 8(1).

Gaines, B., & Shaw, M. (1997). Knowledge acquisition, modeling and inference through the World Wide Web. *International Journal of Human-Computer Studies, 46(6),* 729-759.

Galliers, R. (1992). Choosing information systems research approaches. In *Information systems research: Issues, methods and practical guidelines,* London: Blackwell Scientific,144-162.

Gelderman, M. (1998). The relation between user satisfaction, usage of information systems and performance, *Information and Management*, 34(1), 11-18.

Giang, R. (2002). *Asia/Pacific Business Intelligence Solutions Market Spending Dynamics*. IDC Research Group.

Haider, A., & Koronios, A. (2003). Managing Engineering Assets: A Knowledge Based Approach through Information Quality, *Proceedings of 2003 International Business Information Management Conference*, Cairo:IBIMA.

Herriot, R., &Firestone, W. (1983). Multisite Qualitative Policy Research: Optimizing Description and Generalizability. *Educational Researcher*, 12(3), 14-19.

Hirschheim, R. (1992) Information Systems Epistemology: An Historical Perspective. In *Information Systems Research: Issues, Methods and Practical Guidelines,* R. Galliers (Ed.), Oxford: Blackwell Scientific Publications, 28-60.

Holland, C., & Light, B., (1999). A framework for understanding success and failure in Enterprise Resource Planning System Implementation. *The Proceedings of 7th European Conference on Information Systems,* Copenhagen.

International Infrastructure Management Manual (IIMM) - Australia New Zealand Ed. (2002). Thames: National Asset Management Steering Group.

Information Builders. (2005). Business Intelligence Tools: An Objective Comparison. Retrieved September 10, 2005, from Information Builder website: http://www.informationbuilders.com/products/whitepapers /pdf/BI_Tools_WP.pdf

Koh, C., Soh, C., & Markus, M. (2000). A Process Theory Approach to Analysing ERP Implementation and Impact: The Case of Revel Asia. *Journal of Information Technology Cases and Application*, 4-23.

Lawrence, M., & Low, G. (1993). Exploring Individual User Satisfaction Within User-led Development. *MIS Quarterly, (2),* 195-203.

Levin, R., Mateyaschuk, J., & Stein, T. (1998) Faster ERP Rollouts. *Information Week,* July.

Leung, H., & Wong, P. (1997). A study of user acceptance tests. *Software Quality Journal*, 6(2), 137-149.

Lin, S., Gao, J., & Koronios, A. (2004). Data Quality Issues in Asset Management. *Proceedings of Asia-Pacific Conference on System Integrity and Maintenance*, New Delhi: ACSIM

Lowry, G. (2002). Modelling user acceptance of building management systems. *Automation in Construction*, 11(6), 695-705.

Miles, M., & Huberman, A. (1994). *Qualitative Data Analysis: An Expended Source Book*, Thousand Oaks, CA: Sage.

Moynihan, G. (1993). An executive information system: Planning for post-implementation at NASA. *Journal of Systems Managemen, 44* (3), 8-31.

Murray, M., & Coffin, G. (2001). A Case Study Analysis Of Factors For Success In ERP System Implementations. *Proceedings of Seventh Americas Conference on Information Systems,* Boston.

Mumford, E., (1995). Creative Chaos or Constructive Change: Business Process Reengineering versus Socio-Technical Design. *In Examining Business Process Reengineering*: *Current Perspectives and Research Directions,* London: Kogan Page.

Nandhakumar, J. (1996) Design for Success? Critical Success Factors in Executive Information System Development. *European Journal of Information Systems, (5),* 62-72.

Neuman, W. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*, 4th Ed, MA: Allyn & Bacon.

Negash, S., & Gray, P. (2003). Business Intelligence. Paper presented at 9th American Conference on Information Systems, US. Retrieved September 9, 2005, from http://www.terry.uga.edu/~jaronson/DSS-Readings/GrayAMCIS2003AugBITutorial.pdf

Orlikowski, W., & Baroudi, J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research,* (2), 1-28.

Parikh, M., & Fazlollahi, B. (2002). Analyzing User Satisfaction with Decisional Guidance. *Proceedings of Annual Meeting of the Decision Sciences Institute*, 128-133.

Parr, A., Shanks, G., & Darke, P. (1999). Identification of Necessary Factors for Successful Implementation of ERP Systems. *New Information Technologies in Organizational Processes: Field Studies and Theoretical Reflections on the Future of Work*, Kluwer academic publishers, chapter 8, 99-119.

Pinto, J., & Slevin, D., (1987). Critical factors in Successful Project Implementation. *IEEE Transactions on Engineering Management*, 34(1), 22-27.

Rockart, J. (1979). Chief Executives Define Their Own Data Needs. *Harvard Business Review,* 57(2), 81-93.

Rockart, J. (1984). Engaging Top Management in Information Technology. *Sloan Management Review*, (57), 3-16.

Sarker, S., & Lee, A. (2000). Using a case study to test the role of three key social enabales in ERP implementation. *Proceedings of the 21st International Conference on Information Systems,* Brisbane.

Shanks, G, O'Donnell, P & Arnott, D 1997, 'Data warehousing: A preliminary field study', *Proceedings of the 8th Australasian Conference on Information Systems,* Sydney: ACS.

Soejarto, A. (2003). *Tough Times Call for Business Intelligence Services, an Indisputable Area of Growth*, Retrived September 8, 2005, from website: http://vb.channelsupersearch.com/news/var/40682.asp

Somers, T., & Nelson, K. (2001). The Impact of Critical Success Factors across the Stages of Enterprise Resource Planning Implementations. *Proceedings of the 34th Hawaii International Conference on System Sciences*, Hawaii.

Steenstrup, K. (2004). *Asset-Intensive ERP II and EAM/CMMS MQ Criteria 1Q03*. Gartner Research.

Steinbart, P., & Nath, R. (1992). Problems and Issues in the Management of International Data Communications Network: The Experiences of American Companies. *MIS Quarterly,* 16(1), 55-76.

Walsham, G. (1993). *Interpreting Information Systems in Organizations.* Chichester: Wiley.

Watson, H., Haines, H., & Loiacono, E. (1998). The approval of data warehousing projects: Findings from ten case studies. *Journal of Data Warehousing,* 3(3), 29-37.

Whiting, R. (2003). Look Within-Business-Intelligence Tools have a New Mission: Evaluating All Aspects of a Company's Business. *Information Week***,** 32-33.

Willcocks, L., & Sykes, R. (2000). The Role of the CIO and IT Function in ERP. *Communications of the ACM,* 43(4).

Williams, J., & Ramaprasad, A. (1996). A taxonomy of critical success factors. *European Journal of Information Systems,*(5), 250-260.

Wright, K., & Papke-Shields, K. (2002). Education, training, and user acceptance in the domain of enterprise resource planning systems. *33rd Annual Meeting of the Decision Sciences Institute*, 6-11.

Yin, P. (1994). *Case Study Research: Design and Method.* (2nd ed.), Thousand Oaks, CA: Sage.

Zhang, L., Lee, M., Zhang, Z., & Banerjee, P. (2002). Critical Success Factors of Enterprise Resource Planning Systems Implementation Success in China. *Proceedings of the 36th Hawaii International Conference on System Sciences,* Hawaii.

PANEL PROPOSAL

Global Smartsourcing: Key Opportunities and Challenges

Panel Chair:

Dr. Mahesh S. Raisinghani
Associate Professor
TWU School of Management, USA
Email: mraisinghani@mail.twu.edu

IMB 2006

Sydney, Australia

IMB 2006 Tutorial and Panel Chairs:

Yogesh Deshpande, University of Western Sydney, Australia

y.deshpande@uws.edu.au, Ph. +61 2 4620 3233, Fax +61 2 4620 3075

San Murugesan, Southern Cross University, Australia

smurugesan@scu.edu.au, Ph. +61 2 6659 3676, Fax +61 2 6659 3650

Global Smartsourcing: Key Opportunities and Challenges

Premise:

The greatest advance in the 21st century will come from the ability to fuel innovation and reduce the cost of production by moving work to where the workers are. This is the foundation of the recent move to outsourcing. However, it is only the start of a much deeper revolution called smartsourcing. Smartsourcing looks at how organizations will need to partner with service providers to integrate competencies in order to achieve substantial innovations across the full spectrum of their business.

This panel seeks to extend the thought process started by this phenomenon to consider the meta-analysis of existing literature in global information systems, international business, strategic management and related fields that can help guide future research efforts and inform practice. Contrary to the way the practice of research is often taught and written, the activity of research is clearly a social process, not merely a rationally contrived act. Researchers are urged to go beyond models of national culture and adopt a more dynamic view of culture (e.g., an alternate theory-based view of culture via social identity theory (SIT), which serves as grounding for cultural research programs) that sees culture as contested, temporal and emergent (Myers and Tan, 2001; Straub et al., 2001).

This panel will seek practical advise on the following questions:

1. **How can research in global smartsourcing be made more useful?**
2. **What have we learned from existing theory and frameworks in global smartsourcing?**
3. **How can a meta-analysis be applied in doing useful research?**
4. **What are the critical issues in doing global smartsourcing research that contributes to theory and practice?**
5. **What strategies are necessary for research on global smartsourcing that is potentially useful for both theory and practice?**

Panel Goals:

This panel will extend the research and thought process on the following topics:

- Global smartsourcing Research Frameworks
- Studies that validate and/or extend existing global smartsourcing frameworks
- Conceptualizing and developing new models in global smartsourcing
- Evolution of research in global smartsourcing
- Classification schemes, typologies involving issues in global smartsourcing

This panel will provide a revealing insight into how to synthesize and apply these ideas to theory and practice. Our intention is to provide a constructive discussion on conducting effective global IS research by involving people from four levels:

a) Editors and senior researchers from the global IS area

b) Comments, observations from researchers who have conducted doctoral dissertations relating to global IS

c) Researchers from various parts of the world conducting “unique” or “experimental” research in the global IS area

d) Future researchers in global IS and issues facing such researchers

Procedure:

I am in the process of finalizing the list of panelists and discussants. An informal discussion via a bulletin board will begin to take place by November 2005. This will allow the various individuals to contribute to the panel discussion at the conference without having to waste valuable time on clarification of terms, concepts, and questions.

Panel Outputs:

In the short term it will be required to identify possible domains for this initiative. Also in the long term we expect to see higher level of collaboration, research continuity, accountability, among researchers across the globe.

Panel Proposal/Participation Form

Those interested in contributing to this panel are requested to use the form below to notify the panel chair. Given below is the outline for the panel discussion. If your expertise fits into any of the categories, we would appreciate your input. Please use the attached form to indicate which topic(s) you would like to contribute to. It will be our effort to accommodate contributions from the attendees in the best possible manner. It is our goal to allow sufficient time for discussion and audience participation.

Please fill up the form below and attach a brief c.v or link to your c.v.

How would you like to contribute?

Name and affiliation:____

____

Check all that apply:

Panelist_____________

Discussant___________

Topic(s) you would like to contribute to:

1.________________________________

2.________________________________

3._________________________________

Have you attached your c.v./URL to your C.V.?_________

Special Equipment Needs

____

Please let one of the panel chairs know ASAP, if you would like to contribute to this exciting panel.

Panel Chair:

Dr. Mahesh S. Raisinghani
Associate Professor
TWU School of Management, USA
Email: mraisinghani@mail.twu.edu

AIMB 2006 Conference
Sydney, Australia
Tutorial Session

Business Process Engineering in a Collaborative Boundaryless Environment

PRESENTED BY:
BHUVAN UNHELKAR, PHD, FACS
BHUVAN@CIT.UWS.EDU.AU
(WWW.UNHELKAR.COM)
UNIVERSITY OF WESTERN SYDNEY

14, Feb, 2006, Tuesday

1. Tutorial on "Business Process Engineering in a Collaborative Boundaryless Environment"

Description of the Tutorial

Increasingly, with rapid advances in communications technologies, organizations are finding themselves in a situation where they are constantly interacting with other organizations through their software applications. For example, web-services based software applications are able to transgress the traditional software boundaries (such as operating systems) and interact with software applications from the "other" or external organizations. As a result, it is now possible for the end-user (ie the Customer) to have a single unified view of their business requirements – although such requirements might be satisfied by a collaboration of numerous businesses.

Due to such comprehensive "meshing" of numerous business processes to satisfy the needs of a customer, it has now become essential to consider "cross-organizational" business processes. This is different to the earlier attempts at re-engineering of business processes that focussed solely internal to a business. Today, information and communications technology has made it imperative to consider business processes in the context of all other business partners – supplier and clients as well as government agencies.

This tutorial is focused on considering these uniquely important issues of engineering business processes in the context of collaborative known and unknown business. Collaboration of numerous businesses to satisfy single client demand leads to a relatively "boundaryless" environment as far as the client is concerned. Organizations can respond through this need by modelling and implementing their business processes as singular boundaryless processes – as considered in this tutorial.

This tutorial will also consider a Hospital Management System as an example of engineering of business processes in a global environment. This tutorial is primarily based on a book jointly authored by the presenter titled "Global Enterprise Transitions" (Lan and Unhelkar, 2005, IDEAS publishing group, Hershey, PA, USA).

Topic wise Schedule

This tutorial is a 1-day tutorial for SEPG2002 – although with relevant additional material it has also been presented in two-day format. The three sessions are a rough description, and there will be a half-hour break in between, or spread between sessions.

	Session-wise topics			
Ses sion	*Timing*	*Topic*		*Comments*
1	9:00 AM (1:30 hour)	**Business Process Engineering (in Global Context)**		Why Engineer Business Processes Business Usage of the Internet – land based and Mobile Collaborative commerce and the need to reengineer beyond traditional organizational boundaries.
	10:30	**Coffee Break**		
2	11 AM (1:30 hour)	**A Hospital Management System Working Example (possible hands-on modelling)**		Major issues and challenges in BPR in a boundaryless environment Case study based on Hospital Management System A hands-on attempt to reengineer processes

Audience

1. Managers and senior managers – both business and technical
2. Business Analyst / Requirements modeller involved in analysing and documenting business processes (for both businesses as well as software systems)
3. Researchers and academics, who might be interested in using this material in advancing their modelling work

References

Lan, Y. and Unhelkar, B., *Global Enterprise Transitions"* IDEAS publishing group, Hershey, PA, USA, 2005

2. The Presenter

Profile of Dr Bhuvan Unhelkar

Bhuvan Unhelkar (BE, MDBA, MSc, PhD; FACS) is a 9 times author and has 24 years of strategic as well as hands-on professional experience in Information and Communication Technology. He earned his Doctorate in the area of "object orientation" from the University of Technology, Sydney. In his academic role at the University of Western Sydney, he leads the Mobile Internet Research and Applications Group (MIRAG) and has extensively presented and published research papers and case studies. Founding Principle of MethodScience.com, he also has notable past consulting and training expertise in software engineering (modelling, processes and quality), enterprise globalization, web services and mobile technologies. He is a sought-after orator, a Fellow of the Australian Computer Society, a Rotarian and a previous TiE Mentor. Please see www.unhelkar.com for greater details:

Administrative

The entire tutorial will be delivered from a lap top machine using Powerpoint 2000. Suitable Datashow and projection equipment will be required.

Whiteboard and markers for discussion sessions

Level: Intermediate

Participants will get a softcopy of the presentation slides. Relevant papers and book details can be emailed by the presenter.

End of tutorial Proposal

Full Proceedings of the 2nd International Conference on Information Management and Business (IMB2006) Sydney, Australia

www.aimb.org

13 - 16 Feb, 2006

The 2nd Annual IMB Conference 2006, is being held in Sydney, Australia. This conference will provide excellent opportunity for academics and practitioners to present their work and define the future of information management and business applications. The conference organizers are committed to the success of this event and envisage its significant influence on the business information systems and management community internationally.

These proceedings contain research intense reviewed papers, invited keynote talks, tutorials & panel discussions. Topics covered in these proceedings include:

* AI and Fuzzy Logic for Business Applications
* Business Computing
* Business Intelligence and Data Mining
* Business Modelling
* CRM, ERP, & SCM
* Digital Media Technologies and Applications
* eCommerce and mBusiness
* eTransformation
* Financial Information Systems
* Global Information Systems Management
* Health Information Management
* Information Systems and Technologies
* Knowledge Management and Ontology

Sponsorship & support of the Australian Computer Society, University of Western Sydney & the AeIMS Research Group therein is gratefully acknowledged.

School of Computing and Mathematics,
University of Western Sydney, Sydney, Australia
ISBN: 1 74108 122 X