# Monetising User Generated Content Using Data Mining Techniques

**Yu-Hsn Liu[1], Yongli Ren[2], Robert Dew[1]**

[1]School of Information Technology, Deakin University
221 Burwood Highway, Victoria 3125, Australia
{yuhsnliu, rad}@deakin.edu.au

[2]School of Information Engineering, Zhengzhou University
Zhengzhou 450052, China
yonglitom@gmail.com

**Abstract**

Social media systems such as YouTube are gaining phenomenal popularity. As they face increasing pressure and difficulties monetising the large amount of user-generated content, there are intense interests in technologies capable of delivering revenue to the owners. In this paper, we propose to use data mining techniques to help companies increase their revenue stream. Our approach differs principally in the underlying monetisation model and hence, the algorithms and data utilised. Our new model assumes both consumer and commercial content being entirely user-generated. We first present an algorithm to demonstrate one of possible monetisation technique that could be used in social media systems such as YouTube. A large volume of real-data harvested from YouTube will also be discussed and made available for the community to potentially kick start research in this direction.

*Keywords:* YouTube, User-Generated Content, Monetisation, Web Mining, Data Mining, Business Intelligence.

## 1 Introduction

Three years ago, most of the content published by the media exists as a linear stream coming from a single information source such as a TV channel, the radio, or the newspaper. The 'consumer' as the name suggests is largely responsible for the consumption of information. Their role in publishing or their influence in the content is minimal in most cases.

As communication technologies improve significantly in speed, capacities and forms, we are seeing the emergence of a new media. Characterised largely by 'user contribution', 'sharing', 'decentralisation' and being 'free', these social media systems are gaining phenomenal popularity and success on the Internet. FaceBook, MySpace, YouTube, Wikipedia, and other Web 2.0 sites are overtaking traditional media and to a certain extent, creating transparency levels never seen before.

Just Australia alone, the significance and impacts are clear. In the last two years, many traditional media reported poor earnings results (Cartman, Australian Media 2007, Australian Associated Press 2007, Becker & Posner 2009), and the fire sales of traditional media (Ali Moore 2009, Australian Associated Press 2009) only appear to confirm the bearish outlook of these businesses. As more users turn towards a new paradigm of content consumption, where they are also publishers on a collaborative and free platform, the traditional approach of monetising published content needs to be relooked.

Most social media systems operate without boundaries and are unconstrained by geographical locations, language and time differences. Consequently, they have a subscriber base many times larger than most traditional media in existence. YouTube for example generates more than 100 million views a day and receives more than 65,000 video uploads in 24 hours (Cha, Kwak, Rodriguez, Ahn & Moon 2007). This level of consumption and content creation delivered YouTube the video publishing power that traditional media is incapable of matching. Yet, the success of these systems is also the very reason for their poor financial position as their exponential growth result in significant costs that is matched by a disproportional income stream.

YouTube for example is reportedly losing millions of dollars every day (Fritz 2009, Silversmith 2009, Hartley 2009) because viewers get the content for free. From a business perspective, there is no way YouTube could charge viewers a fee no matter how small that is. Similarly, it is not possible to charge the content publishers who are users themselves. YouTube's mantra of keeping content free soon became the reason of its current success and also the threat of its future failure. Therefore, there is an increasing pressure among such content providers to monetise their businesses (Steffens 2009, Dignan, Diaz & Nusca 2009, Bogatin 2009) before investors pull out.

The idea of monetising content is not new. TV channels, radios and newspapers all publish content for 'free' (or a small fee) in return for 'eyeballs' that could be sold to businesses in the form of advertisements. Social media systems simply adopt this model in hope of achieving the same outcome. *The Age* for example inserts commercial footages on the electronic version of their news. This approach annoys a large number of readers as they find the content intrusive, consumes their bandwidth (which they have to pay for), and are most of the time, untargeted.

If these social media systems are to continue operations, a new monetisation model and appropriate mechanisms are needed. In this paper, we propose a new monetisation model based on user-generated content and user meta-data. This model excludes businesses from the direct involvement of the content users consume. Instead, they would identify user-generated content to push commercial messages on their behalf. To achieve this, we believe data mining technologies would be the best candidate. However, existing algorithms will need to be redesigned to utilise the new model so as to bring about significant

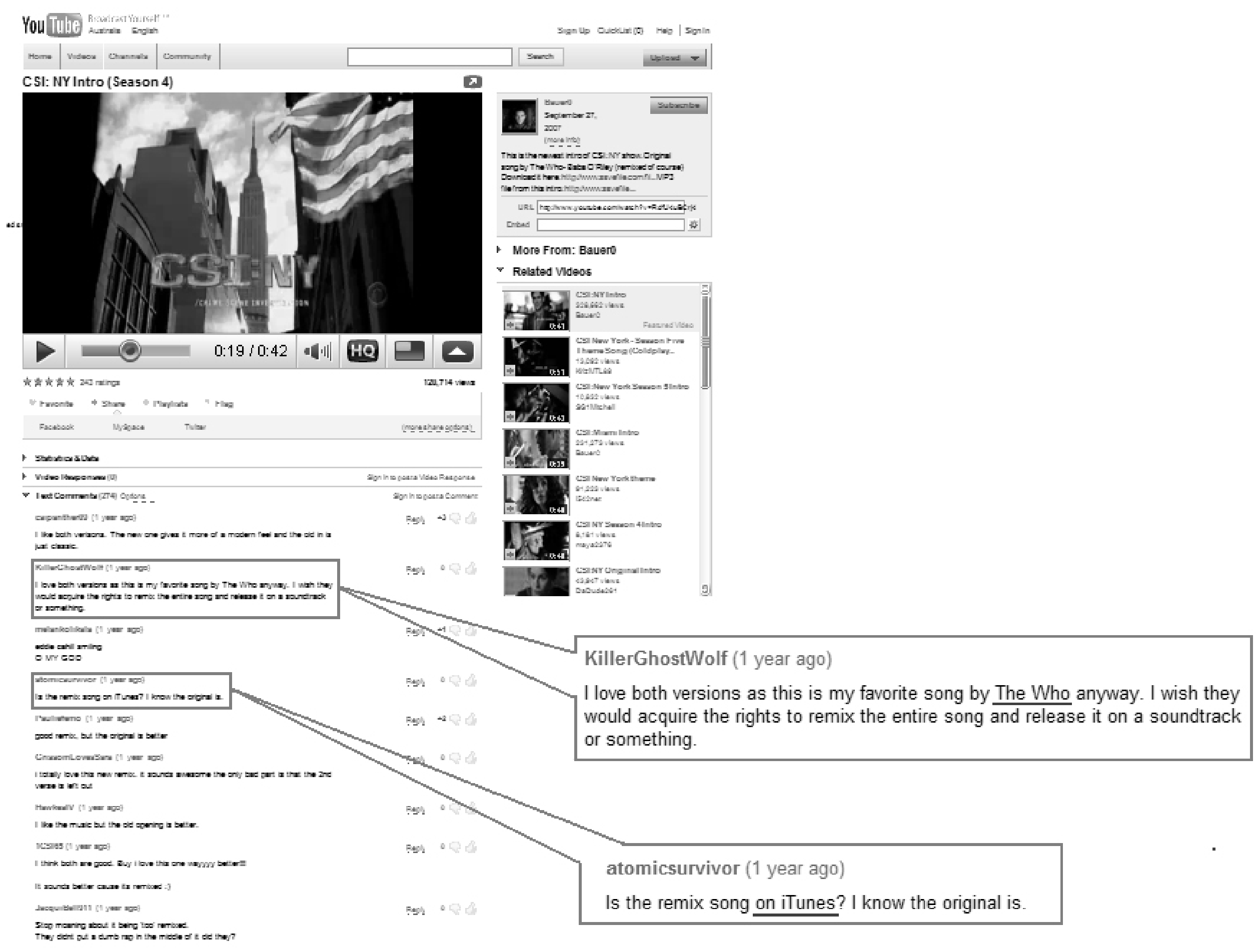


Figure 1: One of the many *CSI: New York* videos posted on YouTube. We will discuss our monetisation model based on this real example. Source: `http://www.youtube.com/watch?v=RdfU4uBCrj4`. Notice that there isn't any advertisements, which means one monetisation opportunity lost. While it's possible to include an advertisement, it is likely to be untargetted for TV programme like *CSI: New York*. Nevertheless, we could exploit what's in the comments, since the comments are a result of watching the video.

increase in the revenue stream.

Among the many social media systems, we will focus on systems similar to YouTube, where the underlying content is user-generated and that user-generated meta-data (e.g., video profiles and video comments) are available. On this particular model, we make the following contributions:

- We propose a novel monetisation model, where both the consumer and commercial content (as well as meta data) are entirely user-generated.
- Using the suggested monetisation model, we proposed a possible monetisation scenario and an algorithm aimed at increasing the revenue streams of content providers.
- As a consequent of undertaking the above research work, we will also contribute our real-world data sets harvested from YouTube so as to provide a platform for other researchers to explore this new direction.

The remaining sections of this paper is organised as follows. In the next section, we discuss further details of our monetisation model. Specifically, we will demonstrate the potential of our proposal with an example. In Section 3, we suggest a monetisation algorithm to realise our proposed scenario. In Section 4, we present our preliminary results before presenting our conclusions in Section 5.

## 2 A New Monetisation Model

The novelty of our approach lie in the observation that advertisements do not necessarily deliver the same level of impact on users of social media systems than messages delivered by a user within their community. With the bulk of social media users in the age of 20 to 30 years old, they made up a significant group whose beliefs are radically different. According to (McCrindle 2009), this group value collaboration, sharing and the freedom of opinion. As a result, they tend to be more receptive to peer opinions rather than commercial messages.

The 'Gen-Y' group of users aside, chances are that many individuals have seek peer opinions (or are influenced by peer comments) when it comes to making a decision about a product or service. Therefore, the significance of peer opinions cannot be undermined. While it was common to manage user opinions to minimise the level of negativity of an organisation's business, user opinions are now leveraged to improve on the positivity of an organisation's product or service. This is commonly seen in the virtual world demonstrated by two key technologies: collaborative filtering (Herlocker, Konstan, Terveen & Riedl 2004, Herlocker, Konstan, & Riedl 2000) popularised by `Amazon.com` and viral marketing (Domingos 2005, Leskovec, Adamic & Huberman 2005).

While social media systems have capitalised on these characteristics to fuel their growth, monetisa-

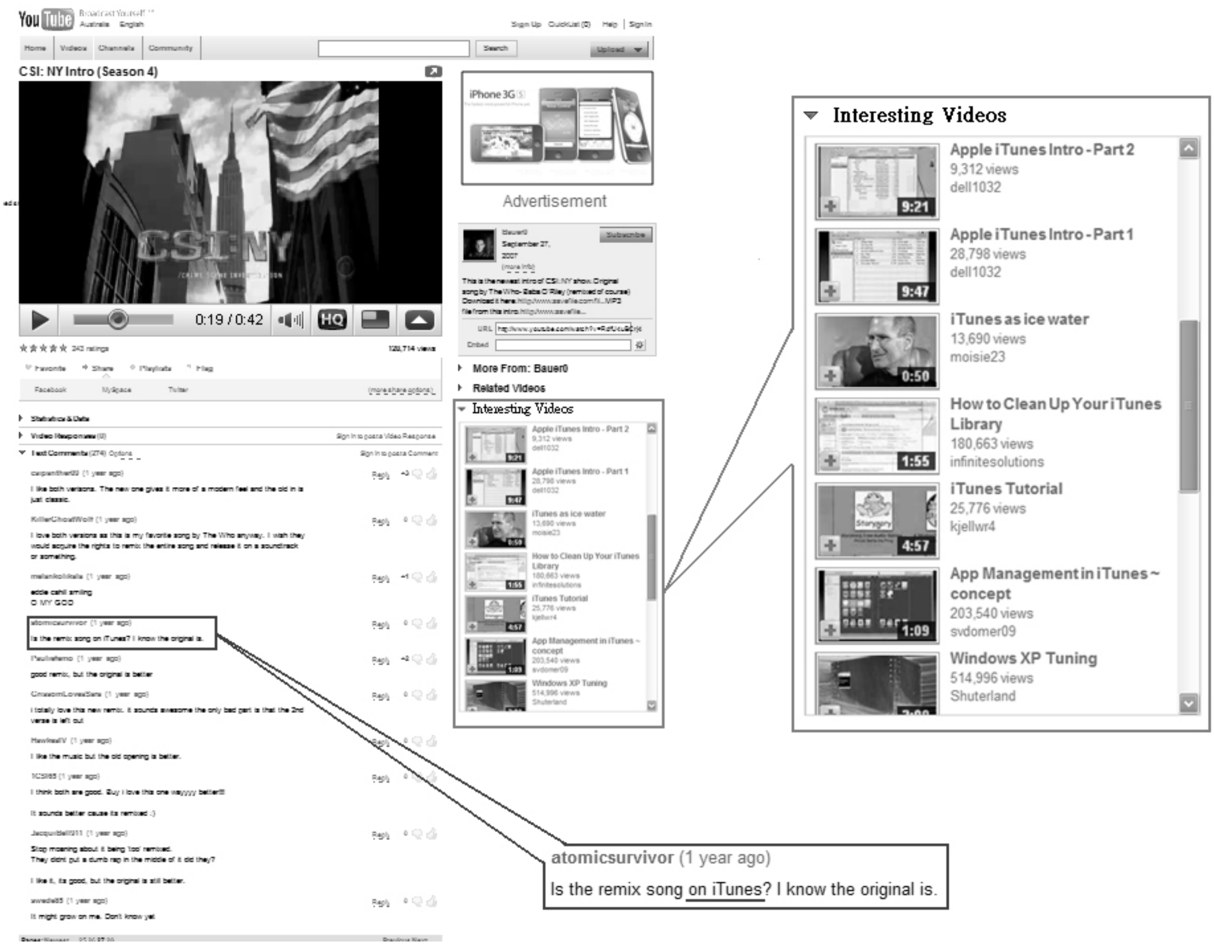


Figure 2: This is a mock up screen showing the enhancements possible, and realising the scenario we discussed in Section 2. In terms of the interface change, this is almost undetectable from the user's perspective. For the media system however, this change together with the monetisation mechanism that works in the background will deliver the revenue stream it needs.

tion methods haven't. Ideally, monetisation methods should exploit the user characteristics the same way the success of social media systems have. More importantly unless an improved revenue stream is achieved, the future of these systems is unclear. Given the large amount of user videos created for the products they use, it is clear that there is a large group of users who would like to offer their views in addition to the commercial messages. Regardless of whether those views are negative or positive, their videos offer a more balanced evaluation for the next potential customer. This led us to consider the possibility of creating a monetisation mechanism, where the commercial messages could be entirely user-generated. To appreciate the virtue of this model, let us consider a possible scenario.

Our example uses YouTube and will assume the readers know how the system operates, and have seen some user comments before. Many of us would have watched one of our favourite TV programmes on YouTube either because you missed the telecast, or you want to watch a certain part again. Let's say you want to watch *CSI: New York* on YouTube as Figure 1 shows. On a system like YouTube not only do you get the video, you would notice a few existing features such as comments posted by other users, a list of related videos found on the right, and also an advertisement along with other features. In this case, there isn't any advertisement shown. This means one monetisation opportunity lost. Even if an advertisement is shown, chances are the advertisement isn't related to *CSI: New York* and would appear random to most users. At first thought, one may argue that a TV programme like *CSI: New York* do not contain material for target marketing. If they do, TV advertisements would not be the way they are now.

We counter argue that this is not true. In fact, the level of target marketing is increasingly apparent even in conventional TV programmes. *MasterChef, Australia* for example has more food related advertisers than a programme like *Big Brother.* On YouTube, the fact that it has all the user generated comments provide a fertile ground for target advertisements to be taken to a new level. Potentially, what conventional TV programmes cannot do, e.g., target marketing on a generic programme like *CSI: New York*, YouTube could with all the other meta information.

To see how this is possible, let us consider a real life example using a CSI video from YouTube (Figure 1 contains the URL for this video). As mentioned, the user comments is one of the differences between watching the same programme on TV and YouTube. If we are to advertise without intruding the viewer and if our intention is to keep the advertisements targeted, one way is to monitor the comments of the video being watched.

In our example, one of the users commented on the theme song of *CSI: New York.* This sparked a number of related posts with one user eventually mentioning *iTunes* and another mentioning the band who sung the theme song – *The Who.* As a result of these comments, there is a probability that a viewer is influ-

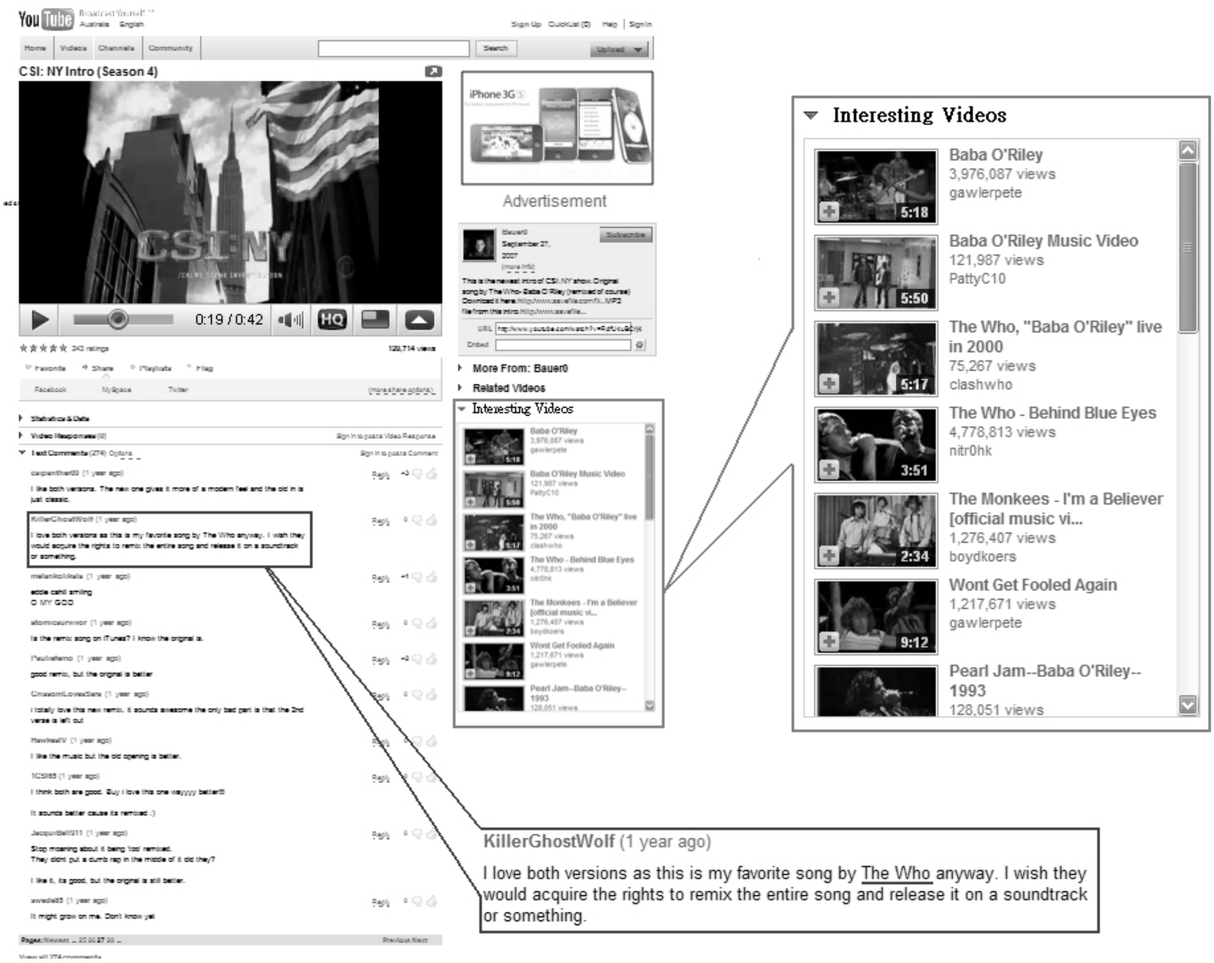


Figure 3: This is another mock up screen showing the effects of our proposed monetisation mechanism. In this case, the mention of both *iTunes* and *The Who* band could be utilised by displaying an advertisement from Apple, and also populating the *Interesting Videos* section with the band videos. Clicking on the band videos may or may not result in direct monetisation. A more ideal system (which we have yet to address in our algorithm) is to direct a click closer towards monetisation. In this case, clicking on the *The Who* videos may not results in direct monetisation but it would maintain the eyeball and eventually perhaps, lead to buying an album by the band.

enced to find out more about *The Who*, or even go online to purchase the song from *iTunes*. Clearly, if the appropriate monetisation mechanisms are in place, an *iTunes* advertisement could be used in place of a random one. This would deliver significant improvements in the click-through. Even better, we could also introduce user-created videos on *iTunes* and videos on *The Who* band along with the CSI video. Figure 2 shows a mock up of the existing screen in Figure 1 to help our readers visualise our monetisation scenario. Notice the *iTunes* advertisement from Apple instead of no advertisement (or a random one) because *iTunes* was detected in the user comments. Additionally, notice a new section called *Interesting Videos* been added after *Related Videos*. While *Related Videos* contains a list of CSI videos, *Interesting Videos* are actually videos that are retrieved from keywords such as *iTunes* and *The Who*, which when clicked could potentially lead to monetisation opportunities if those videos were sponsored by the commercial entities.

From the technical perspective, our interest is in how we could populate the *Interesting Video* section. This is the section where its videos, when clicked, will lead to monetisation. Therefore, an underlying requirement for videos listed in this section is that they must have a monetary value attached. This monetary value could be a payment from the advertiser who selects a user-created content as its agent for commercial messages about a product, or a video that would increase the likelihood of a user clicking on an advertiser's commercial message.

Our example also illustrated the possibility of having more than one keyword in the comments that could lead to multiple sets of videos being candidates of monetisation. In Figure 2, we show how the mention of *iTunes* could lead to targeted advertisements and also a list of *Interesting Videos* about *iTunes*. Another possible scenario from the other keyword is the mention of *The Who* band – see Figure 3. If the commercial owners of the band wants to increase publicity, they could pick some of the user produced *The Who* videos to include in the *Interesting Video* section. Consequently, these videos become monetisable for the social media system but rised another technical challenge – in the presence of multiple candidate keywords (and thus multiple sets of *Interesting Videos* for monetisation), which video should we decide upon?

## 3 Monetisation Algorithm

Now that we have discussed the possible monetisation scenario, we turn our attention to the discussion of a possible monetisation algorithm.

Let $U = \{u_1, u_2, \ldots, u_i\}$ be the set of all users (or

**Algorithm 1** CreateInterestingList(user $u$, video $v$)

1: Let $C(u, v) =$GetComments$(u, v)$
2: **for all** $c \in C(u, v)$ **do**
3: $\quad c' \leftarrow$ StemWord$(c)$
4: $\quad$ **for all** $w \in c'$ **do**
5: $\quad\quad \mathcal{C}_w = \{\phi\}$
6: $\quad\quad$ **if** $w \in \mathcal{M}$ **then**
7: $\quad\quad\quad \mathcal{C}_w = \mathcal{C}_w \cup \{v_t | w \subseteq v_t.\texttt{MonetisationKeyword}\}$
8: $\quad\quad$ **end if**
9: $\quad$ **end for**
10: **end for**
11: Let $\mathcal{R} = \{\phi\}$
12: **for all** $w \in \{c'_1, c'_2, \ldots, c'_j\}$ **do**
13: $\quad$ Sort $\mathcal{C}_w$ such that $\mathcal{C}_w = \{\mathcal{U}(v_{t1}) \geq \mathcal{U}(v_{t2}) \geq \ldots \geq \mathcal{U}(v_{t\ell})\}$
14: $\quad$ Let $\mathcal{C}'_w = \{$top $n$th elements of $\mathcal{C}_w\}$
15: $\quad \mathcal{R} = \mathcal{R} \cup \{\mathcal{C}'_w\}$
16: **end for**
17: Sort $\mathcal{R}$ such that $\mathcal{R} = \{\sum_{i=1}^{|\mathcal{C}'_{w1}|} \mathcal{U}(v_{ti}) \geq \sum_{i=1}^{|\mathcal{C}'_{w2}|} \mathcal{U}(v_{ti}) \geq \ldots \geq \sum_{i=1}^{|\mathcal{C}'_{wj}|} \mathcal{U}(v_{ti})\}$
18: **return** $\mathcal{C}'_{w1}$ in sorted $\mathcal{R}$

user accounts) in the social media system. For a social media system like YouTube, a user $u_j$ can upload a number of videos. We denote the videos uploaded by $u_j$ as $V(u_j) = \{v_1, v_2, \ldots, v_k\}$. For a given video $v$ from a user $u$, a set of comments made about $v$ by other users of the system are available. We will denote this as $C(u, v) = \{c_1, c_2, \ldots, c_\ell\}$.

For a user $u$ watching a video $v$, the objective of our monetisation algorithm is to find a set of videos $\mathcal{T} = \{v_1, v_2, \ldots, v_m\}$ such that for every $v_t \in \mathcal{T}$, $v_t$'s monetisation keyword contains one or more word terms $w_1, w_2, \ldots$ found in $C(u, v)$. For convenience, we will use an object-oriented notation when referring to a user, video or comment property. Therefore, a video $v_t$ will have a monetisation keyword denoted as $v_t.\texttt{MonetisationKeyword}$.

Since the choice of a video can be highly subjective from user to user, it will be difficult to guarantee a click-through. In other words, we may have selected $v_t$ as the video that leads to monetisation but the user may not necessarily click on it. Therefore, given a set of candidate videos $\mathcal{C}$ (or $\mathcal{T}$), we want to rank (or rate) every $v_t \in \mathcal{C}$ so that we can pick the best video in $\mathcal{C}$ for monetisation. To do this, we define a utility measure $\mathcal{U}(v_t \in \mathcal{C})$ to quantify the probability (or likelihood) of $v_t$ delivering a click-through (Joachims, Granka & Pan 2005, Jansen 2009, Zhao, Liu, Bhowmick & Ma 2006). Clearly, there are many ways one could define utility and is likely to vary even within the same system for users on different geographical location. For discussion sake, we define the utility $\mathcal{U}$ of a video $v_t \in \mathcal{C}$ as

$$\mathcal{U}(v_t) = f_1(v_t.\texttt{ViewCount}) + f_2(v_t.\texttt{Rating}) + \ldots \quad (1)$$

where $f_1$ and $f_2$ are functions that would output a normalised value based on the property of $v_t$, in this case `ViewCount` and `Rating`, so that a consistent score could be obtained for each video. Finally, a word term $w$ belongs to $\mathcal{M}$, the collection of monetisable keywords if and only if $w$ is a keyword tagged to a video $v_t \in \mathcal{T}$. In other words, given $w \in \mathcal{M}$, we have a set of videos $\{v_t | w \subseteq v_t.\texttt{MonetisationKeyword}\}$. Given these definitions, we can now present our monetisation algorithm as shown in Algorithm 1.

For a given user $u$ and video $v$, the algorithm will retrieve all the comments $C(u, v)$ associated with $v$ using `GetComments()` as shown in Line 1. For each comment $c$, we will first stem the words so as to make matching easier. While in theory this maybe sufficient, our preliminary analysis of the raw data highlights a number of technical challenges. Of the 3,480,580 comments investigated, there are large number of short-form words, e.g., 'dun' for 'don't' or 'b4' for 'before', that stemming will not adequately address. At the same time, a lot of symbols need to be removed including ones like '`xoxo`', or '`:)`', which has no direct bearing on monetisation.

Once the entire line of comment has been stemmed (and 'cleaned' of short forms and symbols), we will cycle through each word term $w$ in Line 4. If $w \in \mathcal{M}$ holds, then we will add all the videos $\{v_t | w \subseteq v_t.\texttt{MonetisationKeyword}\}$ to $\mathcal{C}_w$ making the videos our candidates for monetisation for $w$ as shown from Lines 5 to 9. Once all the word terms are processed, we sort the videos in each $\mathcal{C}_w$ for each $w$ by their utility $\mathcal{U}$ keeping only the highest $n^{th}$ videos from $\mathcal{C}_w$ in $\mathcal{C}'_w$. This is then added to $\mathcal{R}$ representing the all the candidate videos $\mathcal{C}'_{wj}$ under consideration. We then find the sum of utility for all videos associated with a given word term $w$ as represented in Line 17 – $\sum_{i=1}^{|\mathcal{C}'_{wj}|} \mathcal{U}(v_{ti})\}$. This is then sorted with the collection of videos having the most utility being selected, i.e., $\mathcal{C}'_{w1}$ in sorted $\mathcal{R}$ (Line 18).

At this point in time, we have not considered the speed issue in favor of an easy way to quickly test our idea. Hence, the algorithm's focus is on the logic of the recommendation rather than the practicality of its implementation within a system such as YouTube. In addition, data cleaning regarding the short-forms and symbols were omitted, which we admit could affect the accuracy of our matching a word term to $\mathcal{M}$. We also hand-picked word terms that are easily recognisable as keywords for monetisation by counting and looking for word terms referring to a product as the basis for establishing $\mathcal{M}$ and $\mathcal{T}$ in advanced. All these considered, we wish to highlight to the readers the preliminary nature of our model and we concede the need for further investigation before we could reliably report the impacts of this model. On a significantly small sample and controlled test environment however, the results look very promising.

Finally, we conclude this section with an interesting note for the curious reader. The 'iTunes' word term in our example was mentioned 1,117 times from over 3.4 million comments spread across 623,730 videos and 65,645 users. While these numbers appear to be large, we remind our users that YouTube records 65,000 video uploads a day. With 623,730 videos (equivalent to looking at 10 days of video uploads) we are really looking at the tip of an iceberg

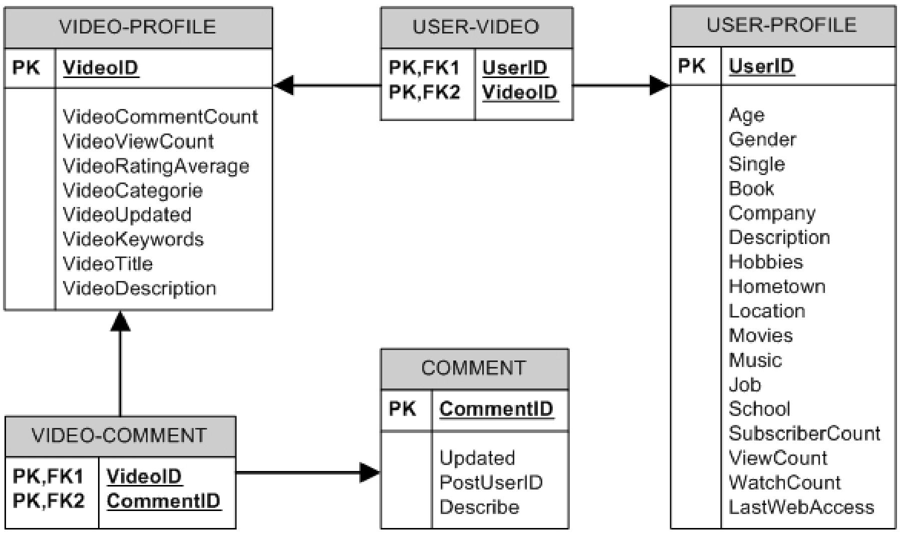


Figure 4: ER-diagram depicting the relationship and structure of our harvested YouTube data. The data can be obtained from our research Webpage at `http://www.deakin.edu.au/~yuhsnliu/youtube`.

and therefore, should not recklessly dismiss its significance.

## 4 Data Sets

We next present the data we harvested from YouTube. Figure 4 shows the ER diagram depicting the way we structured the information we harvested. From this ER-diagram, we were able to load our data files onto a database system to produce joins of a flat file so that different types of analysis could be carried out. Over the lifetime of this project, we endeavor to release weekly updates of what we harvested from our crawler. As we use SQL Server and wrote our initial code in `C#` and `.NET`, instructions on how to load the files onto SQL Server, how to create joins, and how to export the joins as a flat file for analysis are available along with our data sets. Also, the description of each fields and the ER-diagram can be found on our Website. The URL of the Webpage is given in Figure 4.

As a condition of use, the data sets are meant for research and non-profit purposes only. Any form of commercial use should be consulted with the authors. We also kindly request that proper acknowledgement is made when using the data sets downloaded from our Website. Further information on the citation details can be found on the Webpage at the URL given in Figure 4.

## 5 Conclusions

Social media systems with rich video content are emerging rapidly in recent years. As collaborative access and sharing of information becomes the ‘norm’, it becomes vital that businesses utilise these systems and incorporate social media technologies in their operations. *The Age* for example may be a news publisher but incorporating social media technologies on their Website allows them to deliver content through a new dimension. In doing so, it is important that monetisation tools are available so that *The Age* can continue to deliver the cutting edge experience to their readers via a sustainable business model.

In this paper, we contribute to the above in three ways. First, we propose to use user-generated content throughout our monetisation process. We argue the effectiveness of such an approach based on the user characteristics of these social media systems. Second, we propose a monetisation algorithm based on our monetisation model to realise the scenario discussed in Section 2. While our results are preliminary, we are confident that with further refinement to the algorithm and possibly the development of a prototype, it will be possible to demonstrate its impact in the near future. Finally, we will make available the large volume of user-generated content we harvested on YouTube to the research community. With access to this real-world data sets, not only will we help advance existing data mining research but also, we may potentially spark of new research ideas in data mining - particularly, in the area of using data mining to achieve monetisation of user-generated content.

## References

Cartman/Australian Media. Packer Sells Nine to CVC. http://www.australian-media.com.au/-news/4217/packer-sells-nine-to-cvc/, last accessed on July 31, 2009.

Australian Associated Press. Ten Tipped to Post Loss as Ads Vanish. http://business.theage.com.au/business/ten-tipped-to-post-loss-as-ads-vanish-20090401-9jgv.html, last accessed on July 31, 2009.

Ali Moore. PBL Considers Further Media Sell-off. http://www.abc.net.au/lateline/business/items-/200705/s1935762.htm, last accessed on July 31, 2009.

Australian Associated Press. PBL Sell-off to Ignite Gaming Expansion. http://www.theage.com.au/news/Business/PBL-selloff-to-ignite-gaming-expansion/2006/10/18/1160850987522.html, last access on July 31, 2009.

Gary Becker and Richard Posner. The Future of Newspaper, http://www.becker-posner-blog.com/archives/2009/06/the_future_of_n.html, last accessed on June 23, 2009.

Ben Fritz. YouTube Loosing Less Money than Thought, Happy Hollywood Doesn't Know

It. Los Angela Times, June 17, 2009 - http://latimesblogs.latimes.com/entertainment-newsbuzz/2009/06/report-youtube-losing-less-money-than-thought-and-happy-that-hollywood-doesnt-know-it.html, last accessed on August 1, 2009.

David Silversmith. Google Loosing Up to $1.65 million a Day on YouTube. Internet Evolution, April 14, 2009 - http://www.internetevolution.com/author.asp?-section_id=715&doc_id=175123, last accessed on August 1, 2009.

Emma Hartley. YouTube is losing money hand over fist, says Credit Suisse. As is Twitter. Telegraph, UK, May 6, 2009 - http://blogs.telegraph.co.uk/news/emmahartley/-9721127/YouTube_is_losing_money_hand_over_fist-_says_Credit_Suisse_As_is_Twitter/, last accessed on August 1, 2009.

Miriam Steffens. Murdoch Looks for New Ways to Monetise MySpace Traffic. The Age, Australia, February 9, 2009 - http://www.theage.com.au/news/technology/biz-tech/murdoch-looks-for-new-ways-to-monetise-myspace-traffic/2009/02/08/1234027900076.html, last accessed on August 1, 2009.

Larry Dignan, Sam Diaz, and Andrew Nusca. Monetising Social Media: Still an Uphill Climb. ZDNet, May 8, 2008 - http://blogs.zdnet.com/BTL/?p=8764, last accessed on August 1, 2009.

Donna Bogatin. Web 2.0 Hype: Popularity Without Profits, ZDNet, November 2, 2006 - http://blogs.zdnet.com/micro-markets/?p=620&tag=rbxccnbzd1, last accessed on August 1, 2009.

Mark McCrindle. Understanding Generation Y. Government of South Australia, Department of Education and Children's Services, April 2009.

Thorsten Joachims, Laura Granka, and Bing Pan. Accurately Interpreting Click-through Data as Implicit Feedback, Proc. ACM SIGIR Int. Conf. on Research and Development in Information Retrieval, Salvador, Brazil, August 15, 2005.

Bernard Jansen. Investigating Customer Click-through Behaviour with Integrated Sponsored and Non-Sponsored Results. Int. Journal of Marketing and Advertising, 5(1), p. 74 - 94, 2009.

Qiankun Zhao, Tie-Yan Liu, Sourav Bhowmick, and Wei-Ying Ma. Event Detection from Evolution of Click-through Data. Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Philadelphia, USA, August 2006.

Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. Proc. Int. Conf. Internet Measurements, San Diego, California, USA, October 2007.

Jonathan Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems, 22(1), pp. 5 - 53, 2004.

Jonathan Herlocker, Joseph Konstan, and John Riedl. Explaining Collaborative Filtering Recommendations. Proc. ACM Int. Conf. on Computer Supported Cooperative Work, Philadelphia, Pennsylvania, USA December 2000.

Pedro Domingos. Mining Social Networks for Viral Marketing. IEEE Intelligent Systems, 20(1), 80-82, 2005.

Jure Leskovec, Lada A. Adamic, Bernardo A. Huberman. The Dynamics of Viral Marketing. Proc. ACM Int. Conf. on Electronic Commerce, pp. 228 - 237, Ann Arbor, Michigan, USA, 2006.