A Video-oriented Knowledge Collection and Accumulation System with Associated Multimedia Resource Integration

Riho Nakano
Graduate School of Media and Governance
Keio University
Fujisawa, Kanagawa, Japan
nrihos@gmail.com

Yasushi Kiyoki
Faculty of Environment and Information Studies
Keio University
Fujisawa, Kanagawa, Japan
kiyoki@sfc.keio.ac.jp

Abstract

Knowledge-collection and accumulation are essential functions for realizing a knowledge-creation environment. Our system for an objective-video with a story on the Internet focuses on expert-knowledge collection and accumulation based video contents with multimedia, such as images, texts and audio data. The purpose of this system design is to generate "relationships of uploaded data" across heterogeneous categories in order to associate among uploaded contents in various categories related to specific topics included in objective-video, in a cross-disciplinary way. Our system creates two types of "relationships of uploaded data". One type is automatically generated according to similarities among the contents of uploaded data. The other relationship is created by experts once they are shown a list of uploaded content in other categories in response to a query. An important feature of this system is a data structure for expressing uploaded multimedia data, as knowledge-collection related to an objective-video and is called a "multi-dimensional knowledge model" (MDKM). Our proposed MDKM can determine the similarity or association between two sets of uploaded data by computing the relevance score between uploaded data in various dimensions that are their corresponding metadata, such as tag, color histograms, and harmony. By using this system, experts upload knowledge in categories related to the objective-video and accumulate their knowledge. Our system is applicable to E-learning systems, Internet video-sharing platforms, and participatory entertainment systems.

Keywords: Conceptual data modeling; collaborative knowledge building; knowledge accumulation; multimedia resource.

1 Introduction

A knowledge sharing system has recently been proposed as a novel knowledge-creation environment for collecting and sharing multimedia-information resources related to a main story included in a video (Dubnov and Kiyoki 2009). Knowledge sharing, collection and accumulation are significant functions for enabling experts to study a story from various aspects and viewpoints in multiple fields. Several research results are related to integrating knowledge collaboratively (Scardamalia 2002, Kekwaletswe and Bobela 2011). Most existing knowledge sharing systems and services implement collaborative knowledge construction in closed communities or categories, and restrict cross-disciplinary integration of knowledge from various fields. But, It is important to associate in a cross-disciplinary way knowledge related to a particular topic obtained from different fields.

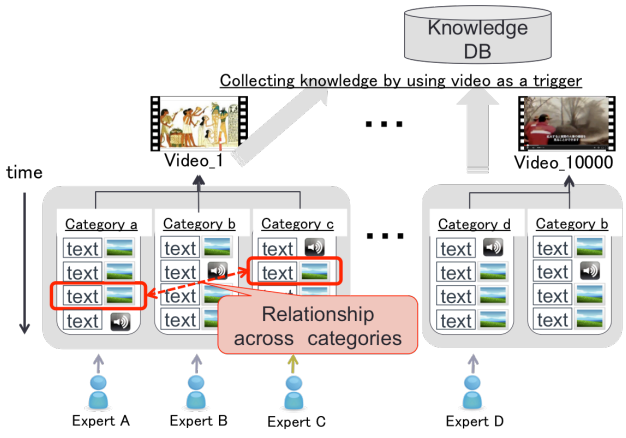

Figure 1: Concept of our knowledge collection and accumulation system

Toward this objective, we aim to generate and store the "relationships of uploaded data" between heterogeneous categories in order to associate among uploaded contents in various categories related to specific topics included in objective-video, in a cross-disciplinary way.

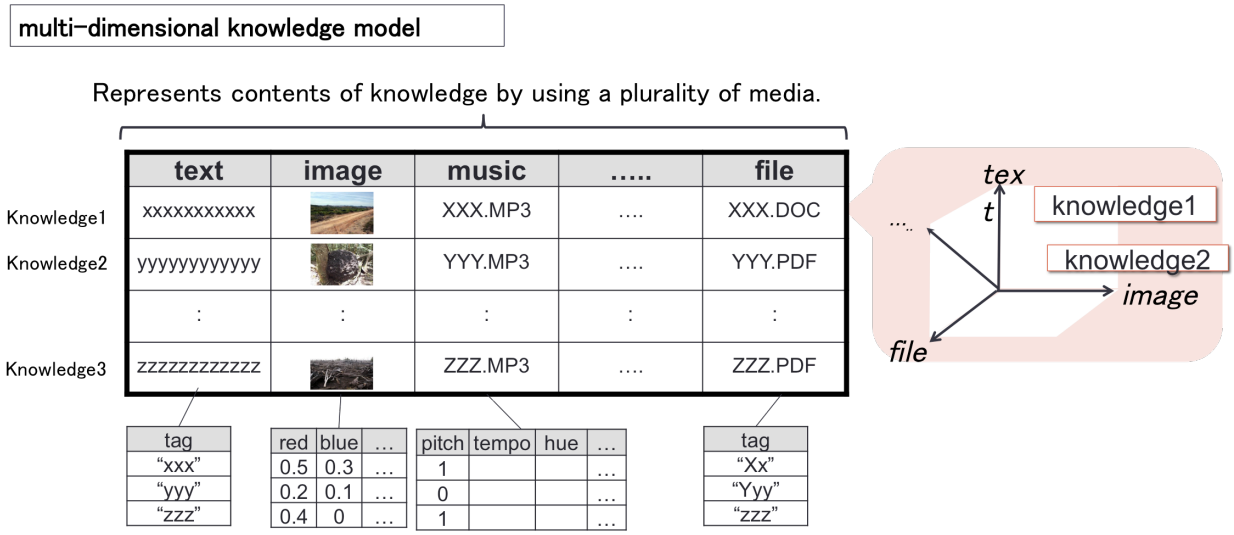

Figure 2: Model of data structure in our system

Our system realizes multimedia knowledge-collection and accumulation along a story expressed in a video content. Experts participating to the knowledge-collection and accumulation upload multimedia information resources expressing their knowledge in images, texts and audio, while watching the video. Experts are hence able to express their knowledge regarding the video that they are watching through the medium of their preference. Various interpretations in a story included in a video promote also experts to upload knowledge in various categories. This is because a video contains images, sound, and motion. This system has two main functions. The first function collects and uploads knowledge about each scene in a video. The other function shows a list of uploaded knowledge in each of various categories. This list is used by experts to refer to uploaded content in the categories related to expert's subject and interest. By referring this list, experts can create "relationship of uploaded data" among heterogeneous categories. In order to display the list, our system selects uploaded data from all categories related to an expert's area, from all the categories according to expert's area of interest by computing similarities among uploaded contents, in the form of text, images, and audio, in a cross-sectional way. Our system realizes two types of "relationships of uploaded data". One is automatically generated according to similarities among the contents of uploaded data, such as image, text or music, and the other is a reference to other uploaded data by experts.

The key technology of our system is a data structure of uploaded data that expresses the content through a plurality of media. We call this a "multi-dimensional knowledge model" (MDKM) and this is shown in Figure 2. This data structure contains text content as well as images and audio. The feature of MDKM can extract the similarity or degree of association between different uploaded data, in various dimensions that contains text, image, audio content. The similarity and association are computed as the relevance score between their metadata of text, image, and music, such as tags, color histograms, and harmony. By using this system, each expert finds knowledge in several categories related to his/her main subject and interest, and adds his/her knowledge to those categories. Thus, the system automatically generates the "relationship of uploaded data" and obtains implicit "relationships of uploaded data" that can be found by only experts. This system is applicable to E-learning systems, Internet video sharing platforms, and participatory entertainment systems. In addition, as an intended use of knowledge collection in our system, we propose an encyclopedia to gain knowledge passively for readers, and a platform to find and solve implicit problems by gathering many experts in various study fields.

This paper is structured as follows. Section 2 describes our system implementation. Section 3 discusses several related studies. Section 4 shows an architectural overview of our system. Section 5 shows the prototype of our system. Finally, Section 6 concludes this paper.

2 Motivating Example

Our system enables a user watching a video to upload text and image content related to the video, and to view relevant uploaded content in several categories, as shown in Figure 3.

For example, researchers at a Japanese university are using our system in a collaborative project with an Indonesian university to address environmental pollution in Indonesia. For this project, the researchers needed information such as the state of the current problem, the background of environmental pollution in Indonesia, etc. Following this, they asked a few graduate students to upload relevant content while watching a video about environmental pollution in Indonesia. They chose one graduate student majoring in each of "agriculture," "biology," "topography," and the "culture of Indonesia," which were the categories of knowledge set by the researchers for this project.

The details of this example are as follows: Student A is majoring in "biology" and has knowledge regarding insects. He is watching a video through our system in order to contribute knowledge from a biological perspective.

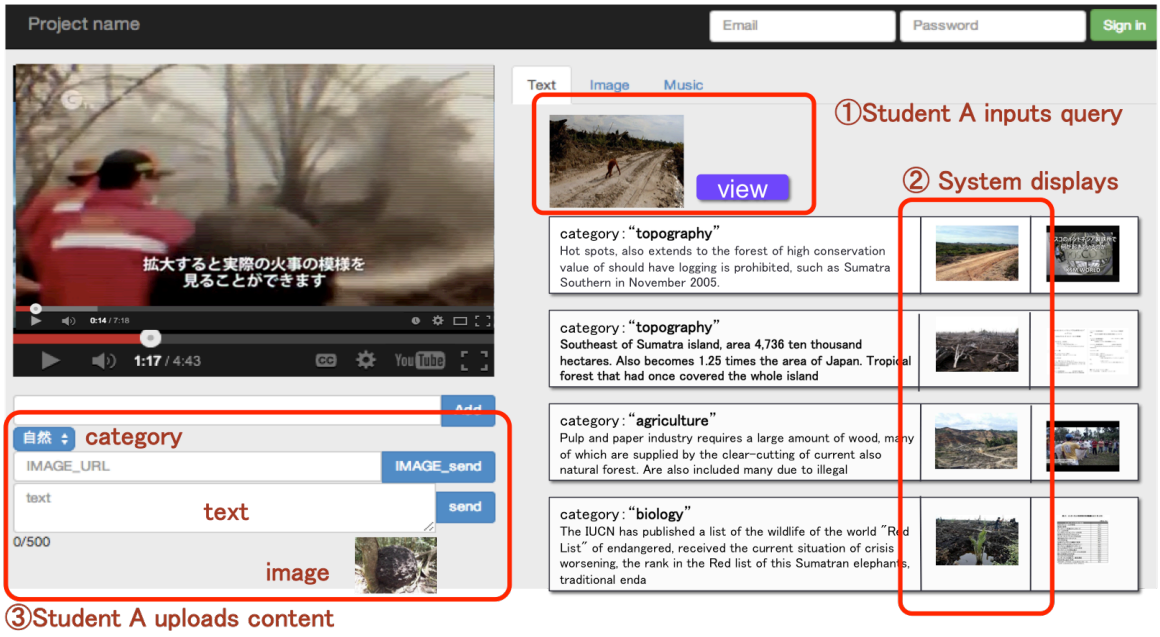

Figure 3: Screenshot of our system

At 0:10 in the video, the narrator says, "Forest fires frequently occur because of the rapid expansion of palm plantations in Indonesia." Student A submits a photograph of soil burned by forest fire in order to view a list of similar content from other categories, as shown in Figure 3-1. Student A is trying to add knowledge at his disposal to uploaded content in the other categories related to insects. Given the photograph, our system displays content containing photos that have similar colors — brown, grey, and black because it can extract similar photos with the inputted photo (Figure 3-2). Then, as the result, he finds the content, "For oil palm plantations, it is recommended that developers use fire." He knows that "Normally, forest is burned to exterminate termites. This is because termites cause serious damage to oil palm trees." As shown in Figure 3-3, Student A enters some text data, uploads a photo of a termite as image data, and selects "biology" as the relevant category from a drop-down list, as shown in Figure 3. He also enters a name for the relationship, "reason," that he has created between his uploaded content and the initial content obtained from his search.

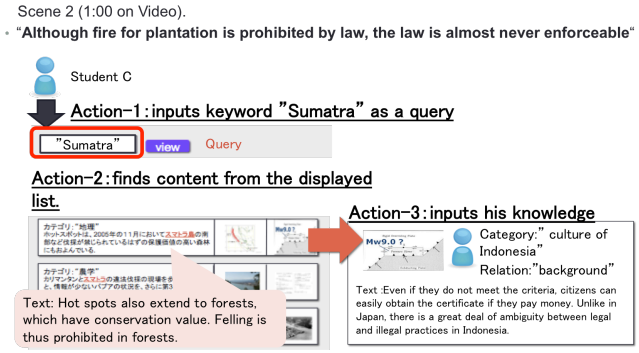

Figure 4: Student C's actions while using our system

We now describe another example of using the list generated by our system through a textual query. Student C watches the same video as Student A. She is knowledgeable about the culture of Indonesia. At 1:00, the narrator of the video says, "Although fire for plantation is prohibited by law, the law is almost never enforceable." Student C enters "Sumatra" as a query because she has lived in Sumatra (Figure 4-1). Our system promptly generates results containing the keyword. Student C chooses an item from the list (Figure 4-2), the content of which is: "Hot spots also extend to forests, which have conservation value. Felling is thus prohibited in forests." She knows that "Even if they do not meet the criteria, citizens can easily obtain the certificate if they pay money. Unlike in Japan, there is a great deal of ambiguity between legal and illegal practices in Indonesia." Student C then sets the content to upload as text data, submits a CSV file regarding corruption cases in Indonesia as optional data, and selects the "culture of Indonesia" as the category (Figure 4-3). She also names as "background" the relationship between her uploaded content and the video content.

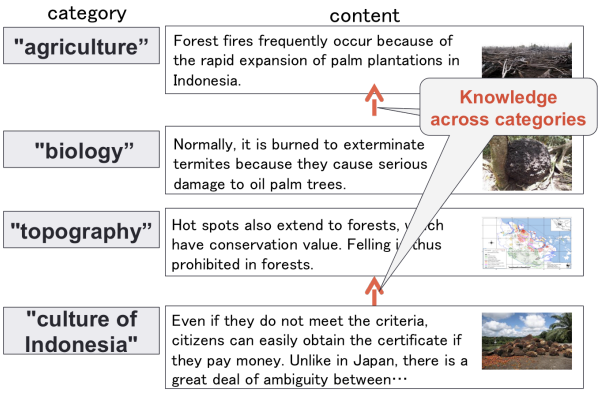

Figure 5: The results of knowledge building in our examples

The results of the two instances described above are shown in Figure 5. As we can see, the content uploaded by Student A in category "biology" is the cause of the content of category "agriculture." Similarly, the content uploaded by Student C in category "culture of Indonesia" provides the background of the content in "topography."

We now describe the operation of our system in the course of the above-mentioned queries. In order to display a list according to the query by Student A, the system calculates the inner product of the color histograms in the photo "burned soil by forest fire" as well as those of other photographs, and sorts them in descending order. When our database returns a set of corresponding data, the system displays these. It then collects the information submitted by Student A: the text "Normally, It is burned to exterminate termites …," the photograph "burned soil by forest fire," the time elapsed in the video, "0:10," the category "biology", the identifying information of the referenced video — "For oil palm plantations ..." — and the name of the relationship, "reason." Moreover, the system extracts the value of each color-based RGB as metadata of the photo submitted. It then inserts these recognized data items as well as the generated metadata into the database. In the case involving Student C, in order to display a list according to the keyword query, the system selects data tagged by "Sumatra" as well as those containing the term "Sumatra" in their text. It then carries out the same tasks as in the case involving Student A.

3 Related Work

In this section, we summarize research related to our proposed approach. We focus on collecting and sharing multimedia information related to a given video, including comments, annotations, and images from online resources. Our approach is related to data extraction methods proposed by Ballan et al. (2011) and Vallet, Cantador, and Jose (2010). The methods proposed by them focus on automatically extracting valuable information from online resources by tagging them or adding annotations to their metadata. For example, the system for automatic video annotation proposed by Ballanet et al. (2011) adds to the number of tags originally provided by users, temporarily localizes them, and associates the tags with articles by exploitig the collective knowledge embedded in tags as well as the online resources Wikipedia, YouTube and Flickr. Vallet et al. (2010) proposed a set of techniques to automatically obtain visual examples of additional queries from different external knowledge sources. Their proposed system provides high-quality visual examples for a content-based video retrieval system. These techniques effectively explain the content of a video by presenting content directly related to it, such as annotations.

Several methods have been proposed to collect information resources related to a given video. For example, the systems proposed by Bertini et al. (2012), Fagá, Motti, and Gonçalves (2010), and Godin, Neve, and Walle (2010) analyze information resources of external services, such as recommending videos and other applications. Fagá, Motti, and Gonçalves (2010) automatically created user profiles through the semantic analysis of annotations in order to discover new videos whose content matched their profile of interest. Bertini et al. proposed a method to benefit from several collaborative applications to integrate multimedia annotations into the end user's documents. These approaches focus on storing information resources as unstructured data, and their goal is not to share knowledge but to simply mine for valuable knowledge, trends, or user interests.

As a basic approach of our system, there are studies (Dubnov and Kiyoki 2009, Nanard and Nanard 2001) that allow users to upload information resources each other interactively. The "Opera of Meaning" proposed in (Dubnov and Kiyoki 2009) is a system for distributed, collaborative, and interactive viewing where the associations of different media elements are created. This system provides semantic and impression-based search performed by the public during the performance in the context of a main story. Nanard and Nanard (2001) proposed a community management mechanism to allow users to share knowledge in order to improve video indexing. These approaches are effective for collecting knowledge in detail and regarding a variety of topics.

Our system differs from these existing methods in expressing knowledge collected across categories by storing relationships between the content of a given video and associated knowledge contributed by experts. It enables viewers to create new collections of knowledge by assuming an interdisciplinary perspective.

4 System Architecture

Figure 6 shows the architectural overview of our proposed system. The main features of the system are that it (1) returns a list of uploaded data according to the user's query, and (2) inserts uploaded data into multiple tables.

4.1 Data Structure

The data structure used in our system consists of three data elements — hub collection and multimedia tables, a metadata table, and a relationship table — which are explained in detail as follows.

4.1.1 Hub Data

A hub collection is a set of sequences that consist of data items that help identify uploaded data. These data connect the other two tables in the system. We define "hub collection" (C) as a data structure determined based on a unit of data used to identify uploaded data (H).

$$C := \langle H, H_2, \ldots, H_i \rangle$$

H_i, which is the *i*-th hub data item, is defined by the following equation:

$$H := \{id, t, video_{id}, category\},$$

where H is a tuple consisting of the id (e.g., number, letter/strong, alphanumeric), timestamp information denoted by t and the id of a reproduced video, and a category of contents. The timestamp information t for uploaded data represents a playback time point in the relevant video stream when the hub data H_i flows into the system. The category is a keyword selected by the user.

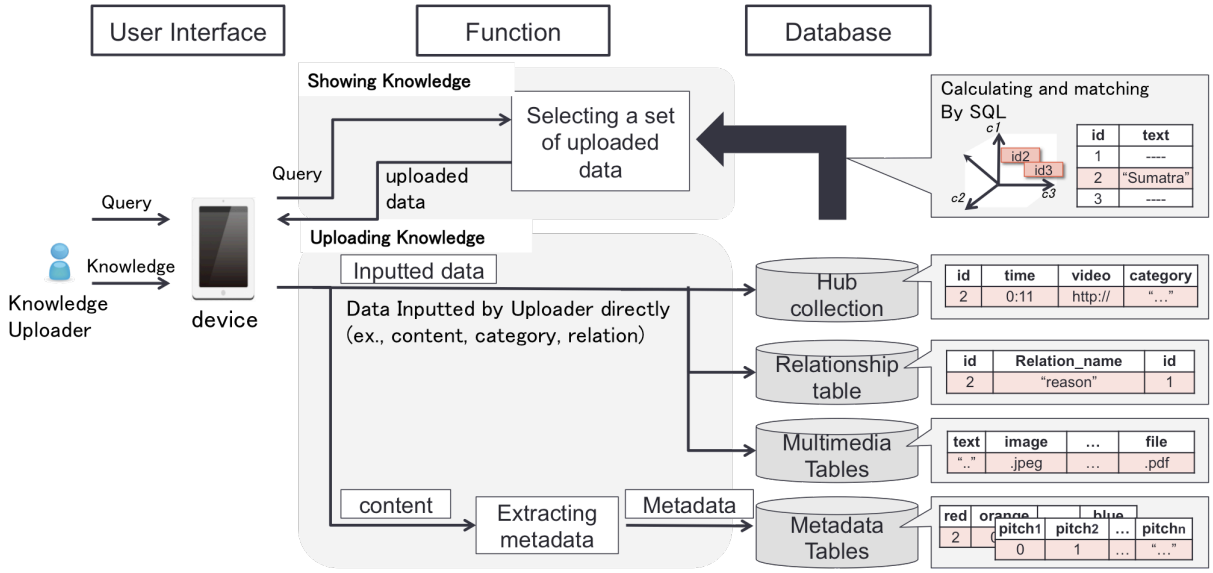

Figure 6: Architecture of our video-oriented knowledge collection and accumulation system

4.1.2 Multimedia Tables

Multimedia tables indicate various contents of knowledge (e.g., text, image, audio, video, etc.) and their corresponding metadata (e.g., tag, color histograms, pitch, tempo, etc.). These metadata are extracted using existing methods. This table is created by each type of media. The multimedia tables are modeled as follows:

$$\mathrm{M} := \{id, content\}$$

where *id* is the id of H_i and "content" is the relevant text, URL, or file (e.g., jpeg, mpeg, mp3, wav, doc, pdf, etc.).

4.1.3 Metadata Tables

The variable *f* expresses knowledge in order to compute similarity between sets of content. It is a metadata set consisting of attributes of metadata (e.g., "fish," "fire forest," red, blue, pitch etc.) and these value pairs. These tables can be created dynamically according to the type of media data, e.g., audio data can contain multiple metadata (pitch, tempo, etc.). This is modeled as follows:

$$\mathrm{f} := \{\langle k_1, v_1\rangle, \langle k_2, v_2\rangle, \ldots, \langle k_z, v_z\rangle\}$$

where $k[1....z]$ is an attribute of metadata and corresponds to $v[1....z]$.

4.1.3 Relationship Table

The relationship table links content from different categories or in the same category. It indicates the combination of different sets of content designated by the uploader. The relationship table is modeled as follows:

$$\mathrm{R} := \{r1_{id}, relationship_{name}, r2_{id}\}$$

where $r1_{id}$ and $r2_{id}$ are ids of the selected H_i and the other *H* and $relationship_{name}$ expresses the nature of the relation, e.g., cause and effect, explanation, etc.

4.2 Functions

The proposed system provides two main functions: extracting metadata from contents, and selecting knowledge from the knowledge collection.

4.2.1 Extracting Metadata from a Knowledge Collection

The system provides a function to generate the metadata of contents that is then inserted into the metadata table. The contents include images, audio, and video. The function is as follows:

$$Function_{extraction}(content) \Rightarrow \{\langle k_1, v_1\rangle, \langle k_2, v_2\rangle, \ldots, \langle k_z, v_z\rangle\}$$

4.2.2 Selecting Knowledge Set

The system selects a set *H* containing content as well as metadata corresponding to a query from C. It computes the relevance score of the two sets of content in question or manipulates string pattern-matching metadata of knowledge. This function is expressed as follows:

$$Function_{selection}(query) \Rightarrow \{H_1, H_2, \ldots, H_i\}$$

where the query is written in a programming language designed to manage data stored in a relational database management system, e.g., SQL. The calculated scores are ordered according to the values of metadata, and the sorted results are listed..

5. Implementation

In this section, we describe a prototype implementation of our proposed system. Figure 7 illustrates a prototype system architecture of the implementation of our system. Our prototype provides two options to the user or knowledge uploader: 1) Upload text or image data file while watching a video, and 2) dynamically display a list of uploaded content given a query by the knowledge uploader. Figure 6 shows the detailed architecture of our prototype. On the server side, the system communicates with a database to insert and select updated data.

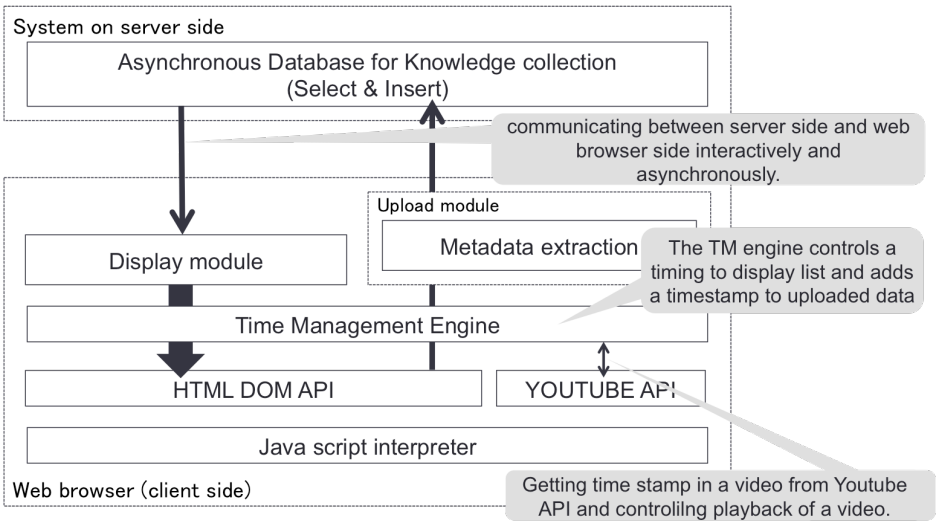

Figure 7: Prototype system architecture of the implementation of our system

On the client side, the system formats uploaded data and arranges these uploaded data to the end user.

The client-side module consists of the following four types of components: display modules, an uploading module, a metadata extraction module, and a timeline-management engine. The metadata extraction module calculates the value of each color and tempo of an image or audio, respectively, by using existing application programming interfaces (APIs) for image and music processing. Following this, the uploading module sends the raw data, the relevant category, and the extracted values to the sever. As shown in Figure 6, the timeline-management engine then obtains a time stamp for the video from YouTube's API, adds a time stamp to the uploaded data, and collates the data using the given time stamp. When the display module receives data from the sever side, it arranges them by utilizing HTML5 and the jQuery library. The engine also controls video playback.

The server-side program inserts the uploaded data into the database on receiving transmissions from the client. The data is then multicasted to other clients watching the same video. Furthermore, this server-side program selects a list of data from the database using SQL. Here, the system specifies the conditions under which data can be selected through tag matching, point directive character string, or value computation. The system then sends the selected data to the client side.

The most important feature of this prototype is its ability to instantly process and multicast a large amount of uploaded data to a large number of users. This is because our system utilizes technologies that reflect state changes based on edits made by users in real time in order to collect knowledge, as shown in Figure 6. This underlying technology relies on the socket.io library. The server-side components are run on a Node.js server, which is a server-side web technology that executes JavaScript.

6. Conclusion and Future Work

This paper proposed a knowledge-collection and accumulation system for videos with generating and storing a "relationship of uploaded data" in heterogeneous categories. Our system offers two functions: collecting knowledge regarding a scene in a video as a topic according to category, and compiling and displaying a list of related uploaded knowledge in other categories. The list is effective for experts to obtain opportunities to refer to uploaded data in the different categories from their main subjects and interests.

In future research, we plan to perform experimental studies to demonstrate the effectiveness of our system for computing and discovering associations between the uploaded data in heterogeneous categories. We also plan to extend the range implementation of our system by enabling it to process a greater variety of multimedia resources, such as audio data.

7. Acknowledgements

This work is partially funded by the Tokyo Foundation and the Mori research funding at Keio University SFC. In addition, we would like to thank Editage (www.editage.jp) for English language editing.

8. References

Dubnov, S. and Kiyoki, Y. (2009): Opera of Meaning: film and music performance with semantic associative search. *Proc. 2009 Conference on Information Modelling and Knowledge Bases* **20**, Maribor, Slovenia:384-391, IOS Press.

Scardamalia, M. (2002): Collective Cognitive Responsibility for the Advancement of Knowledge. *Liberal Education in a Knowledge Society*:67-98.

Kekwaletswe, R. and Bobela, T. (2011): Activity analysis of a knowledge management system: Adoption and usage case study. *Proc. SAICSIT '11,* New York, USA :287-289, ACM Press.

Ballan L., Bertini, M. and Bimbo, A.D. (2011): Enriching and localizing semantic tags in Internet videos. *Proc. 19th ACM International Conference on Multimedia*

(MM '11), Scottsdale, AZ, USA:1541-1544, ACM Press.

Vallet, D., Cantador, I. and Jose, J. (2010): Exploiting external knowledge to improve video retrieval. *Proc. International Conference on Multimedia Information Retrieval (MIR '10),* Philadelphia, PA, USA:101-110, ACM Press.

Bertini, M., Bimbo, A.D and Ferracani, (2012): A social network for video annotation and discovery based on semantic profiling. *Proc. 21st International Conference Companion on World Wide Web*, Lyon, France: 317-320, ACM Press.

Fagá, R., Motti, V. and Gonçalves, R. (2010): A social approach to authoring media annotations. *Proc. 10th ACM Symposium on Document Engineering,* Manchester, United Kingdom: 17-26, ACM Press.

Godin, F. Neve, W.D. and Van de Walle, R. (2010): Towards fusion of collective knowledge and audio-visual content features for annotating broadcast video. *Proc. 3rd International ACM Conference on Multimedia Retrieval,* Dallas, TX, USA: 329-332, ACM Press.

Nanard, N. and Nanard, J. (2001): Cumulating and sharing end users knowledge to improve video indexing in a video digital library. *Proc. 1st ACM/IEEE-CS Joint Conference on Digital Libraries,* Roanoke, VA, USA: 282-289, ACM Press.