eduKEN: A Tool for Fine-grained Video Comment Collection and Analysis

Wai Gen Yee, Andrew Yates
Department of Computer Science
Illinois Institute of Technology
Chicago, IL 60616 USA
yee@iit.edu, ayates@iit.edu

Ophir Frieder
Department of Computer Science
Georgetown University
Washington, DC 20057 USA
ophir@cs.georgetown.edu

Armin Moehrle
TOCA, LLC
Chicago, IL 60647 USA
armin@toca.com

ABSTRACT
An increasing amount of Web information is in video format. Today's search technology allows videos to be found using graphical features and textual descriptions. However, the information gleaned from video features is coarse, while textual descriptions are often short and fail to capture the precise content of videos. We hypothesize that user comments contain supplemental information that effectively describes the content of a video. This information, once extracted, can be applied to a search engine index to improve video search accuracy.

A complete comment-based indexing system must encourage users to post comments and be able to analyze comment data. To support research in the relevant areas of interface design, text analysis, and information retrieval, we present eduKEN, a flexible Web-based tool that allows comment collection, analysis, and search. To support the collection of real-world comment data in an educational setting, eduKEN includes an additional user interface designed for classroom use.

Categories and Subject Descriptors
D.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process;* H.5.0 [**Information Interfaces and Presentation**]: General

General Terms
Design, Experimentation

Keywords
search, annotations

1. INTRODUCTION
The volume of video data on the Internet is exploding. YouTube is already the third-most popular Web site on the Internet. It had more than one hundred million site visits daily in 2009, a 13% increase from the year before [3]. The volume of available video data is increasing dramatically as well, with over ten hours of video uploaded to YouTube every minute [15].

The usefulness of this video data is severely compromised without an effective way of searching it. The search problem of most of these videos is that it consists of comparing a few query terms to a relatively small amount of descriptive metadata. In YouTube, video searches are based on string matching on three information fields created by the video's publisher: the title, the description, and the keywords. Results, consequently, are in "coarse" terms of entire videos.

We hypothesize that the comments that viewers make on the videos also contain information that enhance their description. Moreover, a comment can be associated with a particular segment of video. Such comments can describe to a very fine grain the contents of a video not captured by broad descriptions.

We propose, eduKEN, a system that allows users to view videos and post comments that are automatically and naturally associated with a time within the video (i.e., "timestamped" comments). This differs from YouTube and related systems, which allow users to post comments, but require them to explicitly timestamp them using textual tags (e.g., in the form *dd*:*dd*, where *d* is a digit).

If the information within a comment that describes a video can be identified and extracted, it can be added to a search engine, allowing unprecedented search accuracy and granularity. Such a system is particularly useful with longer videos, where short descriptions cannot hope to describe all of the content. For example, consider a recorded 90-minute college lecture. Several topics are covered over the course of the lecture, much of which cannot be directly discovered with textual search.

The characteristics of college lectures make them well-suited for evaluating the benefits of eduKEN. An alternative educational user interface with a lecture slide browser was designed for this purpose. (We are currently evaluating eduKEN's usability and educational impact in select college classrooms.)

To develop a usable system, research must be conducted in the following areas:

- Interface design – Users must have an intuitive interface that encourages commenting and captures necessary information;
- Text mining – The "information" contents of the comments must be reliably identified;
- Information retrieval – The extracted information must be integrated with other metadata in a search engine index.

Our system is designed to support such research.

Demo
Our demonstration of eduKEN exposes at a high level how its commenting features support data creation and location, improve user interaction, and potentially aid search. Users are able to browse and post comments while watching the videos. They are also able to identify relevant sections of video.

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04

We also demonstrate eduKEN's fine-grained search capabilities. As query responses, users have the option of retrieving either coarse results (entire videos), or fine-grained results (video segments). Specifically, fine-grained results include timestamps of the points in the video that match the query.

Finally, we demonstrate eduKEN's modular architecture, which makes easy the independent modification of components, including the user interface, the video sources, and the data stores.

2. RELATED WORK

Several popular commercial Web-based video sharing systems allow users to post comments. Recently, they have begun to support timestamp tags within comments as well by making them "clickable." Clicking on such a timestamp forwards the video to the point in time corresponding to the timestamp, providing context for the associated comment. However, we argue that requiring the user to explicitly insert timestamps into comments causes the feature to be underused. Newer systems, such as Viddler, allow users to post timestamped comments in a more natural way, suggesting the growing popularity of the feature. Viddler, however, does not support fine-grained search.

The Microsoft Research Annotation System (MRAS) also supports timestamped comments [1]. Despite critical acclaim, it could not sustain long-term interest, and therefore, few studies evaluated the effectiveness of timestamped comments in improving the user's video viewing experience. MRAS possibly lacked support because Web video was in its infancy when it was released (ca. 1999), and users were unaccustomed to its practice. The subsequent rise of video sharing sites with commenting features is strong evidence for this argument.

Videolyzer [4], a system for evaluating the information quality of videos, has a user interface and features that bear some similarity with that of eduKEN, including timestamped comments. However, it is expressly designed for the critiquing the information in a video through the use of a transcript.

Project Pad 2 [10] and VideoANT [13] are designed for educational use, but have specialized interfaces. Project Pad 2 focuses on the ability to add notes to time segments. VideoANT does not support user comments, but instead requires them to be supplied by the video's publisher. Both support streaming video but are not intended for regular classroom use.

In terms of interaction design, our work can borrow concepts from the Games With a Purpose (GWAP) project [12]. The idea behind GWAP is that some problems are too computationally complex for computers to feasibly solve alone, including image recognition. GWAP's solution is to use humans, motivated via games, to solve the problems. In our case, what we want to do is encourage users to post as many comments as possible, many of which are expected to contain descriptive information.

Comment analysis involves the extraction of user opinions about "aspects" of a real-world object, such as opinions of features of sale products [5][6][8]. Our desire to extract descriptive information from comments bears similarities to aspect identification, but is unique in that the "aspects" we search for come from qualitatively different bodies of text and are not constrained by a particular product. Aspect identification yields much of its accuracy from pre-defined syntactic and semantic constraints. We are also not concerned about opinions as they do not describe content, and therefore, cannot be directly applied to improving search accuracy.

A common method of video search is via feature analysis (e.g., color histograms) [11]. Although this is viable in some contexts, it is less natural in many other contexts, as users often express their searches using text. A likely more effective approach is to combine both feature-based and text-based search for retrieval.

Little work has been done expressly on the analysis of comments to improve the description of video [9], though work has been done on identifying interesting conversations within comments through analysis [2]. We expand on previous work by directly considering the information content within comments and discussing their potential for improving search accuracy.

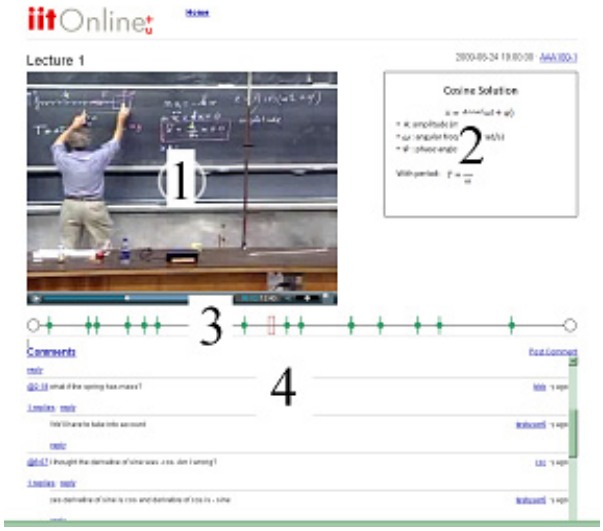

Figure 1. A Prototype Interface Design for the Classroom

3. SYSTEM DESCRIPTION

3.1 Basic Functionality

Basic eduKEN functionality includes video playback, comment browsing, comment posting, and search. Some of this functionality is illustrated in Figure 1: video playback (1) and comment browser (4). A slide browser (2) is included to support a classroom application. A timeline (3) indicates the points in the video that are commented.

By default, the comments display is synchronized with the video playback. Comments are sorted by timestamp, and as the video plays, the comments scroll vertically in the comment browser in synchrony with the video. Sorting can be based on the time when the comment was made, as well.

To post a comment, the user clicks a button and starts typing. Video playback pauses until the comment has been committed. By default, the comment is timestamped with that point in the video. eduKEN also supports un-timestamped comments, such as those found in typical video sharing Web sites, such as YouTube.

The user can issue text queries to search for videos or video segments. The return values of the queries are entire videos by default, but can also be video segments in the form (video, timestamp) pairs. Each result is ranked by a relevance score (e.g., TF-IDF) based on a comparison of the text to which the query is compared (e.g., the title of the video) and includes some of the text that matched the query.

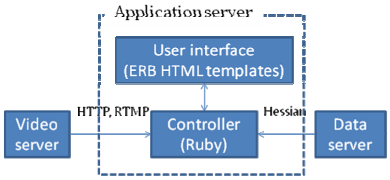

Figure 2. Application Server

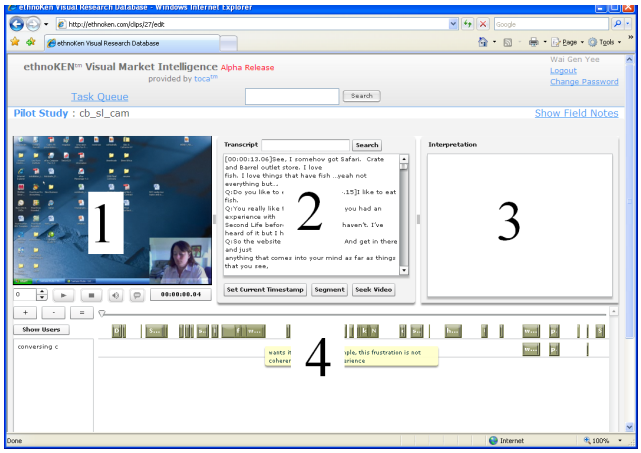

Figure 3. Alternate User Interface Design

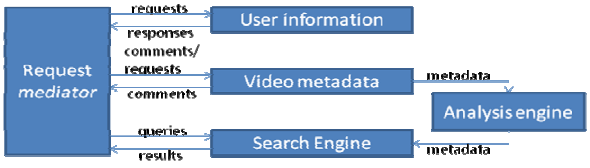

Figure 4. Data Server

3.2 Basic Architecture

The system consists of three main components: an application server, a video server, and a data server as shown in Figure 2. The application server consists of a Web-accessible user interface, which interacts with the video server and data server via a *controller* component. The video server serves the video content to the user. The data server stores user information and video metadata (including comments) and provides data analysis and search functionality.

3.3 Design Features and Implementation

eduKEN is designed to be modular, following a rough model-view-controller (MVC) paradigm. This allows the application, video, and data servers to be independently maintained as long as they obey their well-defined interaction protocol.

The application server is logically made up of two components: a user interface and a controller (in accordance with the MVC) as shown in Figure 2. The user interface can be modified independent of the controller.

We implemented the application server in Ruby using the Rails framework. The interface is implemented using ERB HTML templates, which provide a way to keep program logic in the controller and out of the interface, allowing the interface to be modified by changing only the templates.

In Figure 3, we show an alternate interface design with a video viewer (1), space for a transcript (2), space for typing comments (3), and horizontally scrolling comments (4). Note that Figure 3 is smaller than Figure 1 to conserve space.

The application server communicates with the video server via HTTP or RTMP. HTTP can be used to host the video on a Web server, while RTMP can be used to stream the video from a Flash streaming server. These protocols allow eduKEN to play from standard sources (e.g., Flash and MP4) via the Internet.

The application server communicates with the data server via a customizable protocol, implemented on top of *Hessian*. Hessian is a popular and simple way of implementing communications between Internet-connected servers.

Data management and analysis is handled by the data server. The data server stores user accounts and video metadata. It also hosts the video search engine. See Figure 4.

Many vital operations are done by the data server. First, a user is authenticated with the user data. Next, the user finds a video using the search engine. While a video is playing, requests are made to the data server for comments, which are buffered at the application server and displayed to the user as necessary. Any comment posted by the user is stored at the data server.

Finally, the data server is a natural place to conduct analyses on the video metadata. In our architecture, we analyze video metadata and feed the results into the search engine index.

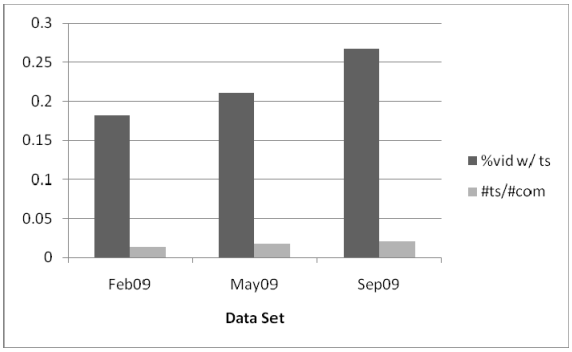

Figure 5. Percentage of Videos and Comments with Timestamps

4. RESULTS

We currently have two research foci. First, we are analyzing comment data for information useful to search. We have three hypotheses:

1. Information useful to search exists in user comments;
2. This information can be reliably extracted;
3. Comments with timestamps contain even more information useful to search.

To test these hypotheses, we present preliminary results of an analysis conducted on YouTube data using our analysis engine.

The data set we used was crawled from YouTube in three batches in 2009: once in February, once in May, and once in September. The videos were retrieved by issuing queries made up of random terms from the SCOWL English word list [7]. For each random query, we randomly picked one video from the result set. For each video, we recorded metadata including title, description, keywords, and comments. Our data set contains over 70,000 videos; a typical video has the characteristics:

- Title length (terms): 6
- Description length: 39
- Keyword length: 11
- Number of comments: 210
- Comment length: 18.

From these statistics, it is clear that the comments contain potentially much more information (210 * 18 = 3780 terms) than do the publisher-supplied title/description/keyword (6 + 39 + 11 = 56 terms) fields. Indeed, in a separate report, we show that including information from comments when indexing video data into a search engine can increase search accuracy by 15% [14].

Other results suggest the increasing importance of timestamps in comments. We identified over 290,000 timestamps in our videos. They occur in approximately 18,000 of our over 70,000 videos. However, their rate of occurrence is increasing. In February, 18% of videos had comments with at least one timestamp, while in September, this rate increased to 27%, as shown in Figure 5.

Based on our initial observations, the information found in timestamped comments is unique from the general description of the video and reliably describes the associated video segment. We are currently preparing these results for review.

Our other focus is to integrate eduKEN into the online education delivery at the Illinois Institute of Technology. A student survey recently revealed that most online students find that the current video player/discussion board paradigm lacks interactivity and does not address certain common problems with the system such as poor recording quality. eduKEN was designed to allow students to make comments contextualized within a lecture video, simulating the raising of a hand in a live classroom. Such functionality should address many of the student concerns.

Recent user studies indicate that 80% of students would use eduKEN over the current system if it were made available.

5. CONCLUSION

The increasing volume of video data on the Web necessitates effective ways to search for them. Publisher annotation is bound to be too sparse. At the same time, Web 2.0 technologies allow users to post comments on these videos.

We hypothesize that these comments contain information useful to search and that this information can be identified and incorporated into search engines to improve their accuracy.

A comment-based search system engenders challenges in the areas of interface design, text mining, and information retrieval. eduKEN is designed to be a flexible way of allowing independent research in these areas.

One initial finding is that YouTube comments contain a significant amount of useful information, improving search accuracy by 15%. Another finding is that students find that eduKEN is very likely to address a need that they have in online education – namely its lack of interactivity.

We are continuing our analysis of YouTube comments, particularly those containing timestamps. We are also currently conducting usability tests on eduKEN in select college classrooms.

6. REFERENCES

[1] D. Bargeron, A. Gupta, J. Grudin, and E. Sanocki. "Annotations for Streaming Video on the Web: System Design and Usage Studies." In *Proc. WWW*, 1999.

[2] M. Choudhury, H. Sundaram, A. John, and D. Seligmann. "What makes conversations interesting?: themes, participants and consequences of conversations in online social media." In *Proc. WWW*, 2009.

[3] comScore. "U.S. Online Video Viewing Surges 13 Percent in Record-Setting December." Press Release. Feb. 4, 2009.

[4] N. Diakopoulos, S. Goldenberg, and I. Essa. "Videolyzer: quality analysis of online informational video for bloggers and journalists." In *Proc. Conf. on Human Factors in Computing Systems* (CHI), 2009.

[5] X. Ding, B. Liu and P. S. Yu. "A Holistic Lexicon-Based Approach to Opinion Mining." In *Proc. ACM Conf. on Web Search and Data Mining* (WSDM), 2008.

[6] W. Jin, H. H. Ho, R. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction," In *Proc. ACM SIGKDD*, 2009.

[7] Kevin's Word List Page. Spell Checker Oriented Word List. Accessed February, 2009. http://wordlist.sourceforge.net/

[8] Y. Lu, C. Zhai and N. Sundaresan. "Rated Aspect Summarization of Short Comments", In *Proc. WWW*, 2009.

[9] M. Potthast. "Measuring the Descriptiveness of Web Comments." In *Proc. ACM SIGIR*, 2009.

[10] Project Pad 2. Accessed October, 2009. http://dewey.at.northwestern.edu/ppad2/documents/help/index.html

[11] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. "Content-based image retrieval at the end of the early years" IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000.

[12] L. von Ahn and L. Dabbish. "Labeling Images with a Computer Game." In *Proc. ACM Conf. on Human Factors in Computing Systems* (CHI), 2004.

[13] VideoANT. Accessed October, 2009. http://www.cehd.umn.edu/ats

[14] W. G. Yee, A. Yates, S. Liu, and O. Frieder. "Are Web User Comments Useful for Search?" In *Wkshp. Large Scale Dist. Sys. Inf. Ret.* (LSDSIR09), http://lsdsir09.isti.cnr.it, 2009.

[15] YouTube Fact Sheet. Accessed September, 2009. http://www.youtube.com/t/fact_sheet