SocialSkip: Pragmatic Understanding within Web Video

Konstantinos Chorianopoulos Ioannis Leftheriotis Chrysoula Gkonela
Ionian University
Department of Informatics
7 Tsirigoti Sq, 49100 Corfu, Greece
+30 26610 87707
choko@ionio.gr

ABSTRACT
In this paper, we present a system that facilitates the analysis of user activity within a web video. Previous research in user-based techniques has assumed an extra effort from the users, such as video replies, comments, tags, and annotations. We have developed and evaluated the SocialSkip system, which improves sense making of web videos by visualizing the simplest form of user interactions with video, such as pause, and seek. In contrast to previous stand-alone implementations, the SocialSkip system employs a web-video player and cloud-based resources (application logic, database, content). The system was validated with two user studies, which provided several hundreds of user interactions with five types of web video (sports, comedy, lecture, documentary, how-to). We found that seeking activity within web video is reversely proportional to how interesting the video is. Moreover, we suggest that if the users are actively seeking for information within a video (e.g., lecture, how-to), then the user activity graph could be associated with the semantics of the video. Finally, SocialSkip provides an open architecture for collaborative contributions to the analysis of the user activity data, in a fashion similar to the TRECVID workshop series.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems-*Video*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces-*Web-based interaction*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing-*Abstracting methods*; H.3.5 [**Information Storage and Retrieval**]: On-line Information Services-*Web-based services*

General Terms
Human Factors.

Keywords
Web Video, Pragmatics, User-Based, cloud computing.

1. INTRODUCTION
The Web has become a very popular medium for sharing and watching many types of user or commercially generated video content [3]. Users can also perform other important tasks, such as commenting videos, replying with other videos, or just expressing their likeness. Despite the increasing number of web videos, there is limited support for understating the content of those videos, besides the text provided by the users (e.g., title, tags, comments, micro-blogs).

Content semantics is an important concept that facilitates the retrieval of multimedia information, such as video. In previous research, there are two broad categories of approaches: 1) Content-based methods extract meaning by analyzing the video itself (e.g., scene change, sound, closed captioning), and 2) User-based methods extract meaning by analysis the user activity on the video (e.g., comments, annotations). In this work, we are concerned with the latter. Previous user-based works have focused on the meaning of the comments and other metadata information of the video. Although there are various methods that collect and manipulate this information [9][10], the majority of them is considered burdensome for the users. Moreover, the percentage of users leaving a comment is too small in comparison to the number of viewers.

Previous efforts have introduced several applications in order to evaluate methods for browsing video content. We have identified the following types of systems: 1) simulation of broadcast, 2) rapid prototypes, and 3) extension of media players. Kim et al [7] built an application on top of a simulated broadcast system in the lab. SmartSkip [6] was developed with multimedia authoring tools and facilitates browsing through thumbnails and video content. Li et al [8] employed the Microsoft Windows Media Player to develop a browser with video navigation features. Crockford et al [5] designed a system as a wrapper around an ActiveX control of Windows Media Player. Lastly, researchers and practitioners have been cooperating for more than a decade on a large-scale video library and tools for analyzing the content of video. The TRECVID[1] workshop series provides a standard-set of videos, tools, and benchmarks, which facilitate the incremental improvement of sense making for videos. In summary, the majority of previous systems operates only locally and depends on video clips that have to be edited and maintained on the respective computer.

2. SOCIALSKIP SYSTEM
Researchers have already realized that the viewer is not the end of the video production-distribution-consumption chain, but could be regarded as node in an edit-share-control process [2]. Therefore, user interactions with the video might add value to the content.

EuroITV2011, June 29th-July 1st 2011, Lisbon, PORTUGAL.

[1] TREC Video Retrieval Evaluation: http://trecvid.nist.gov/

Then, the cumulative user interactions could be leveraged for the benefit of future viewers. In this way, we created SocialSkip: a system that facilitates the analysis of simple user interactions on web videos. In the following sub-sections we are describing the main properties of the system.

We have developed an experimental system based on internet and cloud-based technologies (figure 1). Google App Engine[2] has enabled programmers to develop a web application, maintain it and administrate the traffic and data storage online. In order to implement the SocialSkip system, we used Eclipse (Java). The SocialSkip web video player is based on the YouTube API[3] (application programming interface).

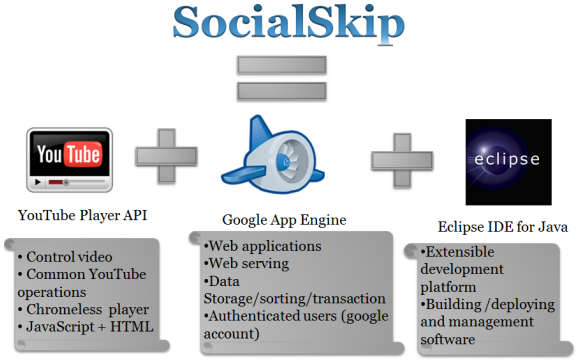

Figure 1: SocialSkip is based on web technologies

The development tools have been seamlessly integrated into a flexible architecture (figure 2). In particular, the YouTube API provides a chromeless user interface (without any controls). Moreover, we used HTML to create the buttons and JavaScript to implement their functions. The YouTube API exposes some important events such as stop or pause. Moreover, it provides methods for controlling the timing of the video. Alternative video APIs could be also used as long as they allow developers to control the current state of the video. In particular, an HTML 5 video player could substitute the YouTube API implementation in a future release, as soon as HTML 5 is supported efficiently from browsers, in several client devices (e.g., mobile, tablet, desktop computers).

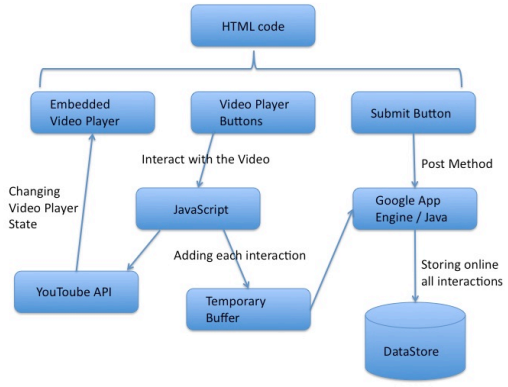

Figure 2: SocialSkip software architecture

[2] http://code.google.com/appengine/

[3] http://code.google.com/apis/youtube/overview.html

Users of SocialSkip should have a Google account in order to sign in and watch the uploaded videos. Thus, users' interactions are recorded and stored alongside with their Gmail addresses. When a user visits the SocialSkip website there are the following elements on the screen: 1) the web video window, and 2) the video buttons, and 3) a submit button. Pushing one of the player buttons has two effects. Firstly, the video player acts accordingly to the button pressed. Simultaneously, we add the specific interaction in a local buffer. When the user pushes the submit button, the interactions are stored in the Google DataStore. Finally, an additional tool was used in the development process of SocialSkip. Questionnaires were employed next to the main SocialSkip video-player and Google forms allowed their efficient integration. In the next section, we are presenting the results from the two versions of the player that we have employed in usability tests and experiments.

3. EXPLORATORY USABILITY TEST

In the first version of SocialSkip player (Figure 3), we aimed to assess the technical feasibility of the above architecture, as well as to get early user feedback.

In addition to the standard buttons (Play, Pause, Stop), the Rewind button shows the video on the reverse order and the Forward button shows the video on a faster pace. Overall, the SocialSkip player provides the main functionality of a typical VCR device, because the users are familiar with them [5][8]. Moreover, we added the GoBack button, which jumps thirty seconds backwards the video.

In the exploratory usability tests, we employed three different videos. The first one was the popular lecture "The last lecture" by Randy Pausch, the second one was a sport video "Australia Open 2010 Final Federer vs Murray" and the third one was a segment of the TV Series «Big Bang Theory». We selected 9 participants and we instructed them to watch the videos in their own pace and interact in any way they wanted with the player. There was no time restriction to complete this task. Furthermore, after the completion of this procedure, we interviewed the users, in order to explore their preferences.

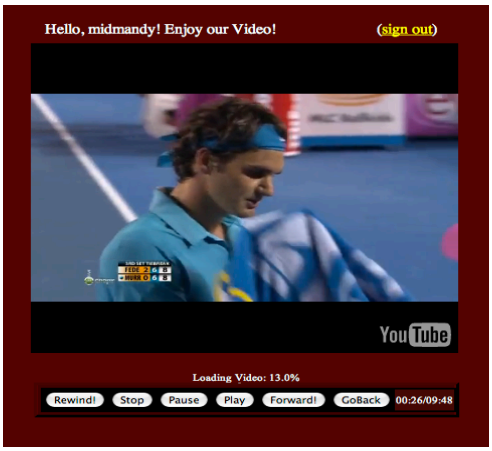

Figure 3 First SocialSkip player had several buttons

We found that the level of user activity on the video depends on how interesting it is. According to the users the most interesting of the three videos was «The Last Lecture» by Randy Pausch, although it was the one with the fewest interactions. On the other hand, the sports video «Australia Open 2010 Final Federer vs Murray» we observed the greatest number of interactions, and

users commented that it was an indifferent video. This finding is in agreement with previous research [9]; it seems that the more attractive a video is, the fewer user interactions are taking place. Finally, we found that user activity was spread over the available buttons, which made difficult to associate video semantics with user activity. According to those findings we have decided to focus on a few buttons and to set-up a more controlled experimental procedure.

4. UNDERSTANDING PRAGMATICS

In the second version of the SocialSkip player, we aimed to visualize activity data from the users, as well as to establish a flexible experimental procedure that can be replicated and validated by other researchers.

The second version of SocialSkip (Figure 4) employs fewer buttons, in order to be simpler to associate user actions with video semantics. Firstly we have merged pause and play on a single button and secondly we have modified the forward and backward buttons to Goforward and Gobackward. The first one goes backward 30 seconds and its main purpose is to replay the last viewed seconds of the video, while the Goforward button jumps forward 30 seconds and its main purpose is to skip insignificant video segments. The thirty-second step is a popular seek window in previous research and commercial work due to the fact that it is the average duration of commercials. Furthermore we have observed replay functions and buttons in mobile devices such as Apple's iPhone and Safari QuickTime video players, which has the default time of 30 seconds as a replay.

We did not use a random seek timeline because it would be difficult to analyze users' interactions. Li et al [8] observed that when seek thumb is used heavily, users had to make many attempts to find the desirable section of the video and thus caused significant delays. Drucker et al [6] and Li et al [8] tested different levels of speed for the functions of forward and rewind, too. User could make the choice of speed and locate more quickly the segment he wanted. For example, there have been commercial systems such as ReplayTV and TiVo that provide the ability to replay segments, or to jump forward in different speeds. Next to the player's button the current time of the video is shown followed by the total time of the video in seconds. Although we did not have a seek bar, we suggest that the data collected from the fixed skip could simulate the use of random seek, because any random seek activity can be modeled as a factor of fixed skipping actions (e.g., a random seek of 180 seconds is equal to 6 skips of 30 seconds).

Figure 4: The 2nd version of the SocialSkip Player is focused on skipping buttons, and questionnaire functionality.

In this case study, we selected three videos (lecture, how-to, documentary) that are as much visually unstructured as possible, because content-based approaches have already been successful with those videos that have visually structured scene changes.

In order to experimentally replicate user activity we added an electronic questionnaire that corresponds to a few segments of the video (Figure 4). According to Yu et al [11] there are segments of a video clip that are commonly interesting to most users, and users might browse the respective parts of the video clip in searching for answers to some interesting questions. In other words, it is expected that in a future field study, when enough user data is available, user behavior will exhibit similar patterns even if they are not explicitly asked to answer questions. The experiment took place in a lab with internet connection, general purpose computers and headphones. Twenty-five users spent approximately ten minutes to watch a video with all buttons muted, so they could not skip or pause. Next, there was a time restriction of five minutes, in order to motivate the users to actively browse through the video and answer the questions that corresponded to a few key-frames. We informed the users that the purpose of the study was to measure their performance in finding the answers to the questions within time constraints.

In order to understand video pragmatics, we visualized the user activity data with a simple user heuristic. Firstly, we considered that every video is associated with an array of k cells, where k is the duration of the video in seconds. Next, we modified the value of each cell by 2, depending on the type of interaction. For each Play, Pause and GoBackward, we increased the value. We decreased the value for each GoForward. In this way we have created the following activity graph (lecture video) that assist the understanding of video content, based on the pragmatics (user video browsing actions) rather than the content itself (Figure 5). Shamma et al [10] has created user activity graphs for visualizing the activity of re-mixing video segments, but has not considered simple user actions, such as pause and seek. In this research work, we focused on a flexible system and the respective experimental procedure for capturing the pragmatics of web videos. In further research, we plan to elaborate and compare alternative methods of analyzing user activity data.

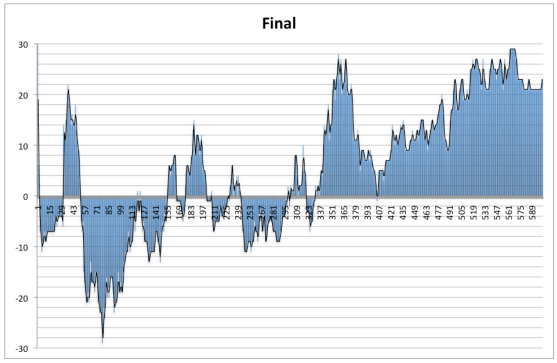

Figure 5 The user activity graph provides a comprehensive visualization of cumulative user interactions

5. CONCLUSIONS

Previous research has developed several applications, in order to evaluate video semantics. Nevertheless, the majority of previous systems runs locally and needs special modification on the software, and on the video clips. Another important procedural parameter is that subjects must be at a specific place where the experiment is conducted and those systems do not scale to large number of subjects without major modifications. Besides stand-

alone applications, there are web-based systems. For example, Shamma et al [10] employed a web-based application where the user can watch, browse, select and annotate video material. In table 1, we present the benefits of web-based implementations compared to the stand-alone applications. SocialSkip is a web application and has several advantages in contrast to stand-alone applications. For example, users do not have to go through an installation process, they just have to visit the link and if there is an updated version they just have to refresh the page.

Table 1. Stand-alone versus Web applications

Stand-alone applications	**Web applications**
Static DB	Large and growing video DB
Off-line user testing	On-line user testing
Open source dev	On-line and open source dev
No need for internet	Need for broadband, start-up time
Difficult to share data	Open to compare data-sets

Besides the flexible programming model, the SocialSkip system provides an open data model. One of the main advantages is that it can use a large number of videos available in the YouTube library. In further work, it could also connect to other growing (lecture and how-to) video libraries, such as Vimeo, and Khan academy. In this way, the materials of our experimental procedure become available to other researchers, who can validate and/or augment the research findings and the understanding of the respective videos. Finally, we provide open access to both source code and the growing data-set of user interactions, which might facilitate further implementations, as well as alternative user activity graphs.

The system design and the user activity analysis have several advantages, as well as some limitations. Firstly, users are not highly involved in the interaction and data-recording process, which is seamless and based on simple user actions, such as skip. Although the user activity approach might not be suitable for videos that are considered very entertaining [9], the pragmatics should be suitable for a growing number of videos on the web contributed by schools and hobbyists, such as lecture and how-to (e.g., cooking, DIY) videos.

Finally, the SocialSkip system has the potential to support remote usability tests, because test-subjects might be at any place, as long as there is broadband internet. Moreover, we suggest that user-based content analysis has the benefits of continuously adapting to evolving users' preferences, as well as providing additional opportunities for the personalization of content. For example, researchers might be able to apply several personalization techniques, such as collaborative filtering, to the user activity data. In this way, video pragmatics is emerging as a new playing field for enriching the user experience. In summary, our motivation for the SocialSkip system has been the realization that we can learn more about a content item not by studying it, but by studying the way people use it.

6. ACKNOWLEDGMENTS

We are thankful to the participants of the user study. The work reported in this paper has been partly supported by project CULT (http://cult.di.ionio.gr). CULT (MC-ERG-2008-230894) is a Marie Curie project of the European Commission (EC) under the 7th Framework Program (FP7).

7. REFERENCES

[1] Fabricio Benevenuto, Tiago Rodrigues, Virgilio Almeida, Jussara Almeida, and Keith Ross. 2009. Video interactions in online video social networks. ACM Trans. Multimedia Comput. Commun. Appl. 5, 4, Article 30 (November 2009), 25 pages.

[2] P. Cesar and K. Chorianopoulos, "The Evolution of TV Systems, Content, and Users Toward Interactivity.," *Foundations and Trends in Human–Computer Interaction*, p. 2(4): 279–374.

[3] S. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., And Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," *Proceedings of the 7th ACM SIGCOMM Conference on internet Measurement*, San Diego, California, USA: ACM, New York, NY, 2007, pp. 1-14.

[4] L. Chen, G. Chen, C. Xu, J. March, and S. Benford, "EmoPlayer: A media player for video clips with affective annotations," *Interacting with Computers*, 2008, pp. 17-28.

[5] C. Crockford and H. Agius, "An empirical investigation into user navigation of digital video using the VCR-like control set," *International Journal of Human-Computer Studies*, 2006, pp. 340-355.

[6] S. Drucker, A. Glatzer, and S.D. Mar, C, "SmartSkip: consumer level browsing and skipping of digital video content," *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves*, Minneapolis, Minnesota, USA: 2002, pp. 219-226.

[7] J. KIM, H. KIM, and K. PARK, "Towards optimal navigation through video content on interactive TV," *Interacting with Computers*, vol. 18, Jul. 2006, pp. 723-746.

[8] F.C. Li, A. Gupta, E. Sanocki, L.-wei He, and Y. Rui, "Browsing digital video," *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, vol. 2, 2000, pp. 169-176.

[9] C. Pinhanez, C.-marie Karat, J. Karat, J. Vergo, and R. Arora, "" Less Clicking , More Watching ": An Option for Entertainment on the Web ? CHI2001," *Communication*, 2001, pp. 3-4.

[10] D.A. Shamma, R. Shaw, P.L. Shafton, and Y. Liu, "Watch what I watch: using community activity to understand content," *Proceedings of the international workshop on Workshop on multimedia information retrieval - MIR '07*, N. ACM, New York, ed., Augsburg, Bavaria, Germany: ACM Press, 2007, p. 275.

[11] B. Yu, W.-Y. Ma, K. Nahrstedt, and H.-J. Zhang, "Video summarization based on user log enhanced link analysis," Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA '03, New York, New York, USA: ACM Press, 2003, p. 382.