vESP: Enriching Enterprise Document Search Results with Aligned Video Summarization

Pål Halvorsen[1,2], Dag Johansen[3], Bjørn Olstad[4], Tomas Kupka[1], Sverre Tennøe[4]
[1]University of Oslo, Norway [2]Simula Research Laboratory, Norway
[3]University of Tromsø, Norway [4]Microsoft

ABSTRACT

In this demo[1], we present a video-enabled enterprise search platform (vESP), an application prototype that enhances a widely deployed commercial enterprise search engine with video streaming. In large enterprises, there exists a lot of information in form of presentations with corresponding video. Using our enhancements, a user can select and combine slides from different presentations generating a new slide deck dynamically and the corresponding video clips are concatenated and presented vis-a-vis the slides on-the-fly. The prototype is evaluated using a data set from Microsoft, and our initial user surveys indicate that the opportunity to enrich the search results with corresponding video is embraced by potential users.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video

General Terms

Design, Experimentation, Performance

Keywords

Video search, content-based composition, personalization

1. INTRODUCTION

Popular search engines still have room for improvements with regard to how video is supported. In particular, indexing of videos is still difficult since automatic video analysis have limitations with regard to precision and resource requirements. Also, the volume and size of videos create problems, in particular if the user is interested in specific events within larger videos, not the whole videos. A search result pointing to a list of, for instance, 90 minutes videos enforces the user to explicitly download whole videos, before manually parsing them for the potentially relevant event within. This is a resource demanding, tedious and time-consuming task.

[1]**A 4-minute video of the** vESP **demo is available at http://home.ifi.uio.no/paalh/DAVVI-vESP.mp4**

MM'10, October 25–29, 2010, Firenze, Italy.

We have previously demonstrated how efficient search across large video archives can be efficiently supported. Our DAVVI system [1] provides a drill-down search interface to voluminous soccer video archives, where selected events are extracted, arbitrary concatenated from several sources into one continuous video playout and smoothly streamed to the end user. This system, however, assumes that third-parties publish relevant textual data that can be crawled by a search engine, indexed and aligned temporally to the corresponding video events. This assumption is fair, but is also specific to the sports application domain.

We have been interested in investigating a broader application area, but where as much as possible of existing components can be used. For instance, large enterprises with tens of thousands of employees typically provide multimedia archives for internal problem solving, educational and sales supporting purposes. Our application domain chosen is therefore for enterprise archives containing multimedia data, and we use data available internally for Microsoft employees. Also, we did not want to create a brand new user interface or application, but rather investigate how to better integrate video into existing search offerings – like the FAST enterprise search platform (ESP). Current solutions tend to treat video separately through specific video search services, not as an integral part mixing highly related multimedia data. Hence, we have enhanced and integrated video in the Microsoft enterprise search engine released this spring.

2. vESP

FAST ESP is already a scalable, high-end enterprise search engine commercially used world-wide. In the next version, labeled FS14, improved contextual meta-data aggregation, more configurable relevancy models and support for in-page document browsing give the users better and more relevant search results. In this context, our aim is to further enrich search results with corresponding video data.

An example scenario is given by the large information collection from Microsoft giving searchable information about the Windows7 operating system. This data set often consists of Powerpoint presentations and a corresponding video that is recorded from the original presentation. Currently, FS14 is able to return a search result where the Powerpoint slides can be browsed within the search page, and alternatively another link (url) to the corresponding video, i.e., the user must manually browse each set of slides and seek in the video to find the interesting parts. The main idea of vESP is to align the video with the Powerpoints and then allow the user to select the interesting slides from possibly sev-

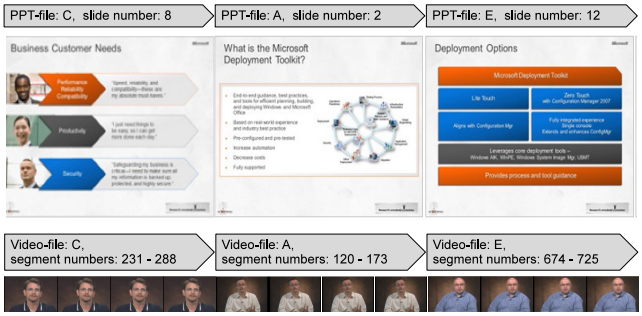

Figure 1: Playlist playout

eral presentations into a new, personalized slide deck. The relevant parts of corresponding videos will subsequently be extracted and placed into a playlist as depicted in figure 1, and both the slides and the aligned videos are played as one continuous presentation.

One of the main challenges in systems like vESP is to find metadata which can be used for search and annotation of the video. In our scenario, we have a set of presentations and corresponding videos. In our demonstration, we have indexed the slides and the slide transcripts. The transcripts can be obtained using speech-to-text tools, or as separate text documents which were available in our data set. So far, the transcripts only add indexable text as metadata, but we are investigating how to use the corresponding timing information in the audio stream to make a fine-granular index, i.e., also including smaller explanations within a slide. The data set used to demonstrate the functionality already contains a slide transition table which is used to map video segments to slides, i.e., the video interval corresponding to as particular $slide_i$ is given by the its start- and end-times. However, we are currently investigating different approaches of automating this process by detecting a slide change in a video as also researched in [2]. Nevertheless, when such functionality is available, it will also create an incentive to produce more metadata, e.g., using the timing functionality already existing in Powerpoint to have the slide transition times recorded.

vESP uses our adaptive segmented HTTP streaming technology [1] similar to Move Networks, Microsoft's Smooth Streaming and Apple's HTTP Live streaming to deliver the video content. Each video is partitioned into 2-second segments, each coded as an independent video in multiple qualities for dynamic video quality adaption in real-time and for arbitrary concatenation of segments into a content-based, personalized video matching the selected slides. Similar to torrent technology, the segments are uploaded to servers in the Internet and retrieved in playout order for display using HTTP GET requests.

Figure 2 shows the prototype interface. At the top, one can submit traditional search queries. The in-page document preview is available below each search result, and the composition of a slide deck and playout of the video is located on the right. Moreover, to test the enhanced functionality of the enterprise search platform using this interface, we have performed a user study using a 7-point balanced keying Likert scale (1-useless, 2-strongly dislike, 3-dislike, 4-neutral, 5-like, 6-strongly like and 7-wow). 33 assessors, both technical and non-technical persons, tested the system

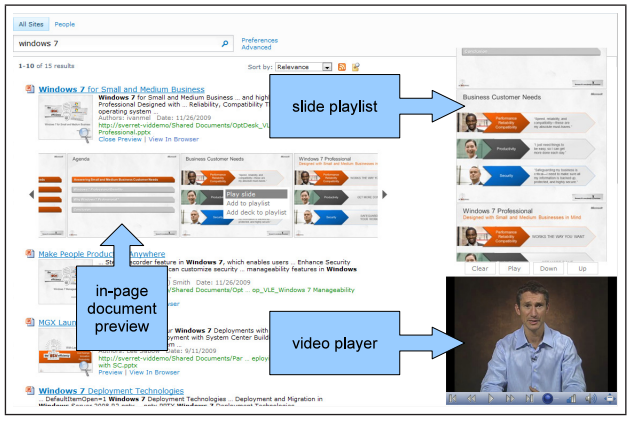

Figure 2: FS14 with document preview and integrated video functionality

System	avg	max	min	stdev
Plain	3.87	5	3	0.64
Document preview	5.36	6	4	0.65
vESP	5.60	7	3	1.06

Table 1: User study

with respect to the impression of the search page, and the results are presented in table 1. We did observe a small difference between the technical and non-technical persons, where non-technical persons were slightly more reluctant (many of them thought that they *must* use the video functionality all the time, and not as the intended option that *could* be used). Nevertheless, vESP received the best average score, also compared to the in-page document preview, and was the only system in total given the highest score ("wow") – 7 out of 33 persons gave a top score, some even from the group of technical people.

3. DEMO

In this demo, we present vESP. We illustrate how video clips are extracted according to slide boundaries and how to dynamically generate a new presentation across multiple search results. Furthermore, we demonstrate the smooth playout of the on-the-fly, multi-source generated video and its synchronized slide presentation. Finally, we experimentally compare vESP to the traditional ESP functionality and the extended in-page document preview over the same data set.

4. REFERENCES

[1] Johansen, D., Johansen, H., Aarflot, T., Hurley, J., Kvalnes, Å., Gurrin, C., Sav, S., Olstad, B., Aaberg, E., Endestad, T., Riiser, H., Griwodz, C., and Halvorsen, P. DAVVI: A prototype for the next generation multimedia entertainment platform. In *Proceedings of the ACM International Multimedia Conference (ACM MM)* (Oct. 2009), pp. 989–990.

[2] Ngo, C.-W., Pong, T.-C., and Huang, T. Detection of slide transition for topic indexing. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)* (2002), pp. 533–536.