

vESP: A Video-Enabled Enterprise Search Platform

Pål Halvorsen^{1,2}, Dag Johansen³, Bjørn Olstad⁴, Tomas Kupka¹, Sverre Tennøe⁴

¹University of Oslo, Norway ²Simula Research Laboratory, Norway

³University of Tromsø, Norway ⁴Microsoft

Abstract—In this paper, we present how to provide a novel and potentially disruptive multimedia service by modifying a widely deployed commercial enterprise search engine. The idea is to transparently integrate rich multimedia data with traditional textual-oriented query results. This includes that the search engine automatically discovers and extracts relevant scenes from a large knowledge repository of existing videos and produces a new, customized video of events matching the user query. To evaluate our prototype, we have performed experiments using a data set from a knowledge repository in Microsoft consisting of PowerPoint presentations with corresponding videos. Our initial results demonstrate that such integration can be implemented efficiently, and that potential users prefer to have the opportunity to enrich the search results with corresponding video.

Index Terms—Video enriched search results, user study

I. INTRODUCTION

Search engines have emerged as a fundamental interface for locating and accessing data, whether on the Internet or in large enterprises. The initial textual based search service has gradually been complemented with support for additional media types, often organized as specific media-type search services. For video, this includes Blinx, Google Video and Bing Video. YouTube, even if not always framed as a video search engine, clearly belongs in this category.

We are investigating a complementary approach to these popular video services where video data is not managed as a specific media type, but rather is better integrated in traditional search experiences. To investigate this more thoroughly, we have built vESP. This is a prototype deviation building on the latest version of the widely deployed enterprise search engine FAST ESP (Enterprise Search Platform) from Microsoft. This is technology currently deployed inside large enterprises, or behind the scenes of huge web portals. This includes, for instance, Verizon, New York Times, Financial Times, Disney, Elsevier and Comcast. As an indication of volume, current ESP installations have accumulated query volumes in the order of billions a day. Precisely resolving a textual query into a video or image is vital for many of these enterprises, in particular for e-commerce sites.

In general, user dialogues with video search engines are still strongly influenced by the iterative, minimalistic, and text-inspired approach used for more than a decade in leading search engines. For video, for instance, a search engine typically replies with a set of video thumbnails for further manual handling. This puts the user heavily in the knowledge discovery loop, in particular since video is voluminous in nature. Pro-

viding a reference to, for instance, a 90 minutes movie where the specific event searched for amounts to seconds within, is clearly a solution with room for improvement. Closely related is how to extract meaningful and accurate metadata to build a useful video search index. This problem has been gradually improved over the last decade for discrete media types, but a similar improvement for video event extraction is still, at best, ahead of us.

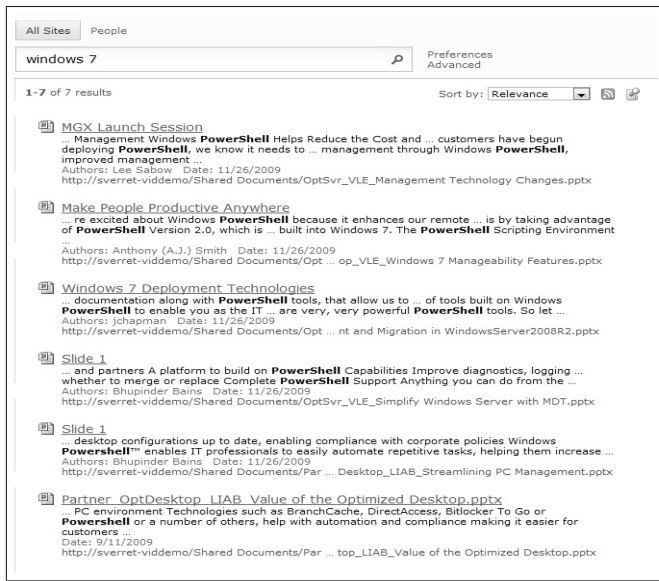
Here, we describe a prototype where we integrate rich multimedia data with traditional textual search results in a commercial enterprise search engine. The prototype is tested in the context of an enterprise knowledge database comprising presentations and corresponding videos. The results from a user evaluation indicate that potential users prefer the additional, but optional video-complemented search results compared to the plain text, document-browsing results.

The rest of this paper is organized as follows. The next section briefly describes some examples of existing systems. In section III, we present how we approach these and related problems with our vESP prototype. Then, in section IV, we detail the architecture, and section V presents a user evaluation of the extended multimedia functionality. Section VI gives a brief discussion, before we conclude in section VII.

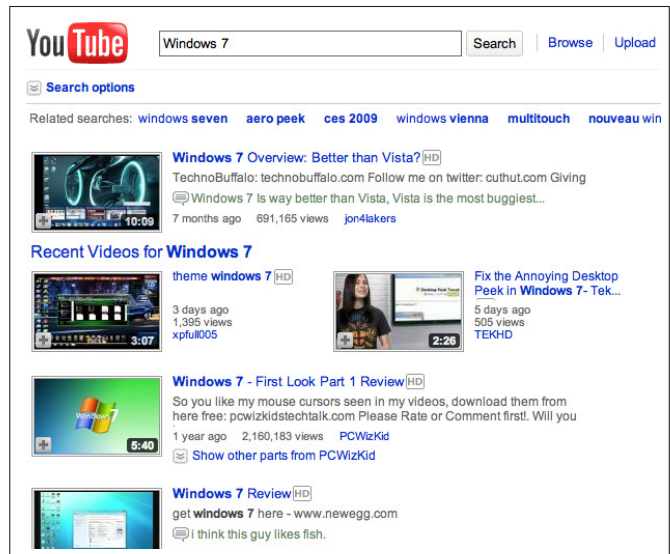
II. RELATED WORK AND EXISTING SYSTEMS

The FAST enterprise search platform (ESP) [1] is a highly efficient text-oriented search platform supporting a large range of document formats. It supports a lot of different configuration, e.g., result ranking using the concept of tunable rank profiles (individual ranking), and query tools like refinements, lemmatization, anti-phrasing, stemming, entity extraction and spell-checking. These are all of which are well known features of enterprise search engines. The search results are then typically returned in a Google-like style as shown in figure 1(a).

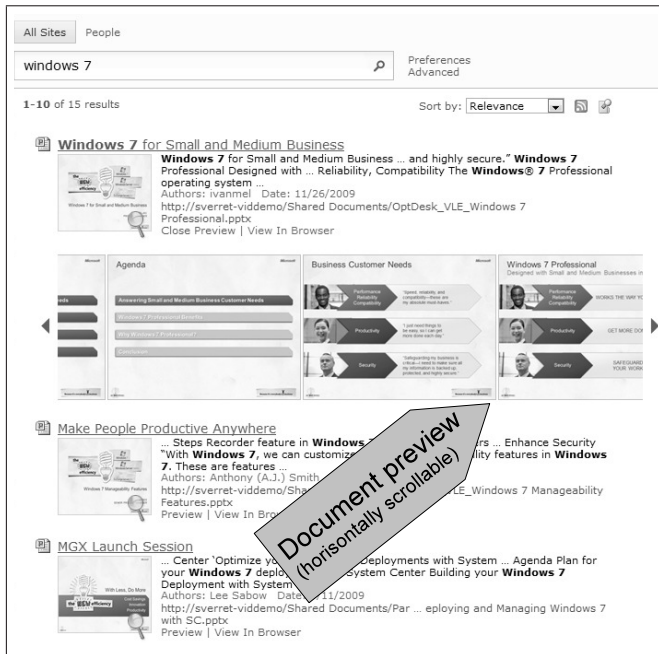
In the next release, labeled FS14, there is also support for so called “Best Bets” which allows administrators to map keywords to hyperlinks and keywords to arbitrary content, e.g., video, audio and images. This gives great possibilities in terms of enhancing the most popular queries with relevant, actionable content, all directly in the search results. Best Bets and Visual Best Bets are user context aware, meaning you can expose groups of users to different content, based on user properties such as location, work responsibilities, expertise, title, or any custom property defined. In FS14, there is also support for in-page document preview as shown in figure 1(b)



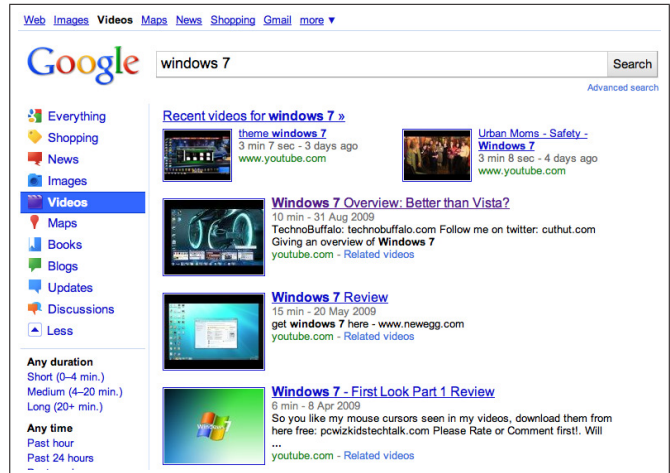
(a) Textual, Google-like



(a) YouTube [2]



(b) Document preview



(b) Google video [3]

Figure 1. FS14 without integrated video functionality

where a user may browse through the document without opening a separate application like Word or PowerPoint. However, currently there is no support for on-the-fly generation of customized presentations from the search results and there is no connection between documents and the corresponding video.

Video libraries with video search are currently identified by YouTube-like systems as shown in figure 2. You may search for a given topic like “Windows 7” where a list of links to videos about this topic is returned. The challenge however is that maybe only small parts of different videos are of interest,

Figure 2. Examples of search capability in existing systems

and the user must in this case manually adjust the payout interval and manually jump between the different videos when switching from one clip to another. Additionally, in these systems, there is no integration of the presentation document and the video.

With respect to tools for presenting slides and corresponding videos, there exist several approaches where slides and video presentation is synchronized. For example, the Altus vSearch [4] is the system that closest resembles our system with respect to this sub-functionality. As shown in figure 3, it combines enterprise video search with PowerPoint synchronization and scrolling transcripts into an accessible and searchable video archive. vSearch is also able to generate custom made presentations for later reuse. However, the functionality that is missing when searching for knowledge, is the ability to additionally present a customized video for a selected set of slides, i.e., without manually going through each presentation.

There exists many individual components that may be used

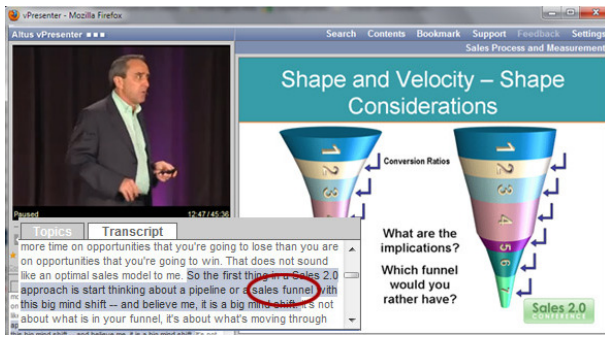


Figure 3. Altus vSearch [4]

in our scenario. However, the novelty of our system is the full integration of document search and customization with a corresponding video extraction, on-the-fly generation and synchronized presentation. Thus, in the vESP system, we have addressed this challenge by developing a running prototype integrating the next generation enterprise search engine with in-page document preview with an adaptive and dynamic video streaming system.

III. VESP PROPERTIES

The current FAST ESP (FS14) is a high-end scalable enterprise search engine already providing conversational search experiences. This interactive property gives users better and more relevant answers, in particular when combined with contextual meta-data aggregation, and configurable relevancy models, query processing and results processing. In this context, our approach has been to enrich this dialogue with directly relevant video data.

The typical vESP flow of operations is depicted in figure 4. In our example application using data from Microsoft, the knowledge database consists of presentations related to various technical topics. The presentation data include the PowerPoint slides, the speech transcripts and slide timing information (slide change time). In our scenario, this data is explicitly stored in the system, but in a general case, the speech transcript could be extracted using speech-to-text tools, and the slide transition time can often be obtained from automatic video analysis detecting the change in an area of the picture. Additionally, the video data is stored in a “video storage server”. Thus, using vESP, the user can, based on a textual Google-like search, select individual slides or complete presentations in a new on-the-fly generated presentation and optionally have the related video displayed along with the slide presentation.

A main property of vESP is thus that it attempts to align directly related video events to specific textual queries. This is obviously application-specific as there are numerous discrete elements represented in a search index without a directly related video event. However, one such domain is knowledge repositories of large enterprises which, in addition to traditional documents, may contain videos of oral presentations like lectures, conference talks and sales presentations. Slides and transcripts of such talks are often made public on the

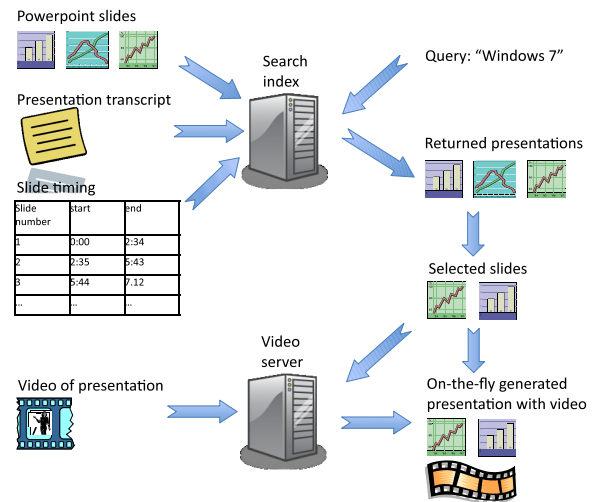


Figure 4. vESP operations: indexing, querying and data presentation

web, and now more recently commonly accompanied with a recorded video footage. Thus, a massive amount of (possibly distributed) information is available, i.e., efficient search is the enabler to find and retrieve the requested information.

Precision is important in this context. Aligning a general reference to slides and the related video is neither novel nor fine-grained enough to support accurate resolution of textual queries to the directly associated events in the video. As such, we have a second and more novel property in vESP in which fine-level events within a video is captured and represented with their metadata. An important part for such an index entry is its time offset into the related video. This is used to map an index keyword into a specific part of the larger video. For instance, the result returned when searching for “Windows 7” in vESP contains the traditional url and text snippet listing, but also PowerPoint and video event thumbnails. Since FS14 supports in-page browsing, the actual slides can be directly opened and viewed by clicking on the PowerPoint thumbnail. What vESP adds is that the click on a particular slide also activates a video player displaying the part of the overall video that talks about the content on the slide. This is easier said than done, but is illustrated in figures 1 and 5. The result page presented in figure 1(a) is the traditional media-poor. In figure 1(b), the result page has been enriched with a document preview, while figure 5 displays vESP with document preview and video. The point now is that only the parts of the video related to the query “Windows 7” can be played out.

A third and novel property of vESP is that it potentially summarizes multimedia content when serving individual users. Normally, a search engine acts as an extra level of indirection pointing users to remote and relevant data sources. The user then explicitly determines which data to manually download. A search engine cache violates this architecture, but still the cache content is a copy of some data once published on the web. vESP brings this to a new level by prefetching already published video data, converting it to smaller, fixed-size seg-

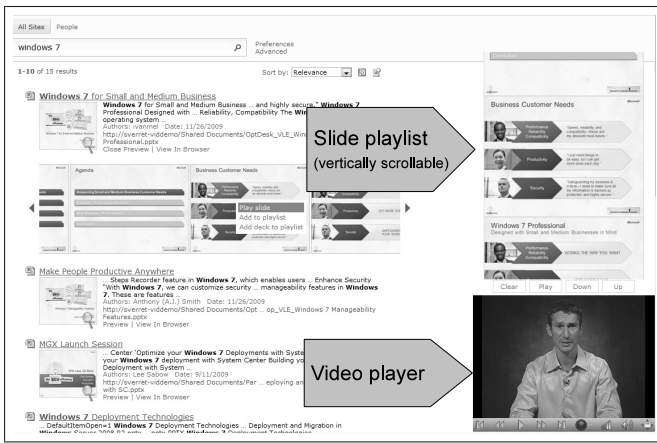


Figure 5. FS14 with document preview and integrated video functionality

ments, indexing it, and storing it in the enterprise infrastructure for efficient distribution. A traditional video scene is made up of a series of such fixed size segments, and we chose to fragment a video into a finer resolution than traditional scenes since there might be scenes with redundant content relative to a query. With this fine-grained representation, only the segments within a scene directly related to a query can potentially be used.

Each segment is stored as a small video, and several such segments can be efficiently integrated to produce a customized video. An entire scene will naturally be easily produced by concatenating the associated segments together. A more complex example is when a slide has been used multiple times, and VESP produces a video with content from totally different talks captured on different videos. Finally, and more the common case, is an example where a user selectively builds a playlist of slides from possibly different sets of slides (illustrated in figure 6). VESP produces on-the-fly the corresponding continuous video summarization with seamless transition between the source videos.

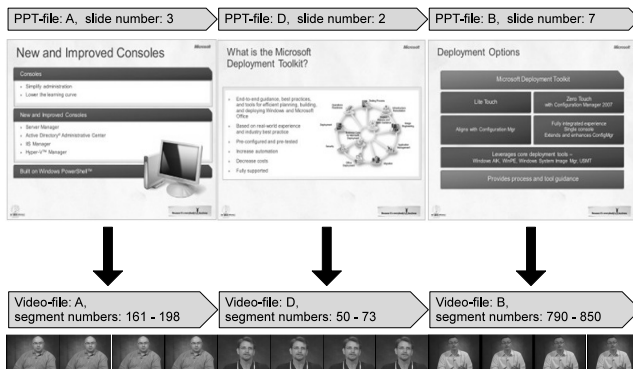


Figure 6. Playlist payout

IV. SYSTEM ARCHITECTURE

Search and video playing are the visible subsystems of VESP, but other important subsystems works under the hood

to provide an efficient search, video production, and seamless video streaming solution. We have previously demonstrated similar ideas in the soccer domain in an end-to-end prototype where extremely customized soccer videos can be produced in response to a textual query [5].

Almost the same architecture and several of the same subsystems are used in VESP, but for a completely different application area. Here, the use case is search and information extraction in an enterprise knowledge database. Users can search a company’s relevant content, retrieve and compose highly detailed multimedia presentations. This is a compelling application, given the vast amount of knowledge that exists in large, corporate organizations, but where it is virtually impossible to manually resolve detailed demand and precise supply of knowledge.

The architecture consists of two integrated subsystems, which are ① metadata extraction and search¹ and ② video segmentation and dissemination. An overview of the architecture and how the different subsystems interact are presented in figure 7. Notice that the services provided by each subsystem are integrated by the client-side software.

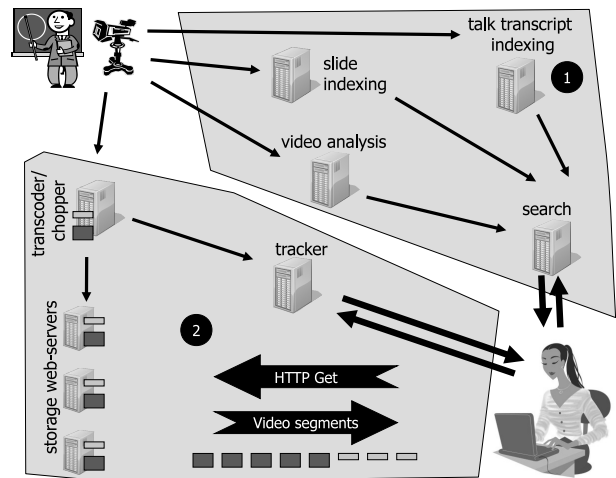


Figure 7. VESP Architecture

A. Metadata Extraction and Search

In order to be able to identify events in a video and enable search, we aimed for a fully automated metadata extraction solution. Not surprisingly for the soccer domain, we experienced that automatic video analysis tools were not precise enough to serve our needs. We achieved low-level event capturing with a precision in the order of about 80-85%, but important high-level events like, for instance, a missed goal or an exciting move could be missed.

We solved this problem by a semi-automated metadata extraction solution, where we crawled the web for reliable

¹In our previous work, metadata extraction and search were two separate subsystems, but they are here integrated into FS14. In the experiments described below, we did not use automatic video analysis. We do, however, have tools available, but this will then be a separate subsystem outside FS14.

and close to real-time sport publishing sites. Next, we aligned time-annotated discrete events indexed by the search engine and annotated with their time offsets, to the respective time offset in the raw video. The metadata aggregation problem was now reduced partly to a time synchronization issue.

The metadata aggregation approach for vESP is close to fully automated. First, we use transcript of talks, either published by the same publishers of the talk video and the associated slides, or obtained using speech-to-text tools. The resulting text and the slides themselves are indexed similar to normal documents in a traditional search engine. 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. A search for “Windows 7 Powershell” will then return an even more limited video interval containing only the part where “Windows 7 Powershell” is mentioned within the slide. This is a well known problem [6].

As with the soccer domain, the remaining problem is how to synchronize the index elements correctly with the offset in the video and length. That is, which sequence of video segments should a query, and thus a slide, be resolved and linked to. In our test scenario, we use real data already used internally by Microsoft, and it is not produced to fit our prototype.

The data set (see section III) 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 start- and end-times of $slide_i$ in the table. However, we are currently investigating different approaches of automating this process by detecting a slide change in a video [7]. Furthermore, as such functionality becomes available, this creates an incentive for people to produce more metadata. This can in this scenario for example be done using the already existing timing functionality in PowerPoint and have the slide transition times be recorded.

B. Video Segmentation and Dissemination

The user can either be involved in refining the video result from a textual query, or this process can be fully automated. In the former case, users can select one or more events initially suggested by vESP and compose those into a playlist. In the latter approach, vESP automatically generates a specified-length playlist based on the search result ranking.

Next, the playlist content is streamed to the user for playback as a content-based, personalized video seamlessly switching from one event to the next in the playlist with a reasonable degree of continuity when played out. To disseminate the multi-quality video streams, we use a segmented, torrent-like HTTP streaming approach. Such a solution has been adopted by, for instance, Move Networks [8], Microsoft’s Smooth Streaming [9] and Apple’s HTTP Live streaming [10]. This architecture has been proven very scalable [8], but we use our own container format to achieve even better performance [11].

The internal encoding of the different segments also takes into account heterogeneity issues like, for instance, varying user devices, buffer fill level, and network bandwidth. This

is supported by having each segment encoded in different quality representations and adapting to available resources in real-time by changing video quality when streaming. The size of each segment depends on target search granularity, delivery efficiency, coding efficiency, and video-quality adaption frequency. We conform to the de facto industry standard by using two-second video segments. A typical video event thus consists of several consecutive two-second segments.

Next, all the segments are distributed to web-servers either out in the Internet or the enterprise infrastructure. vESP finally retrieves segments using HTTP GET requests for each segment in a torrent-like manner each fed in-order into the video player for a smooth playout.

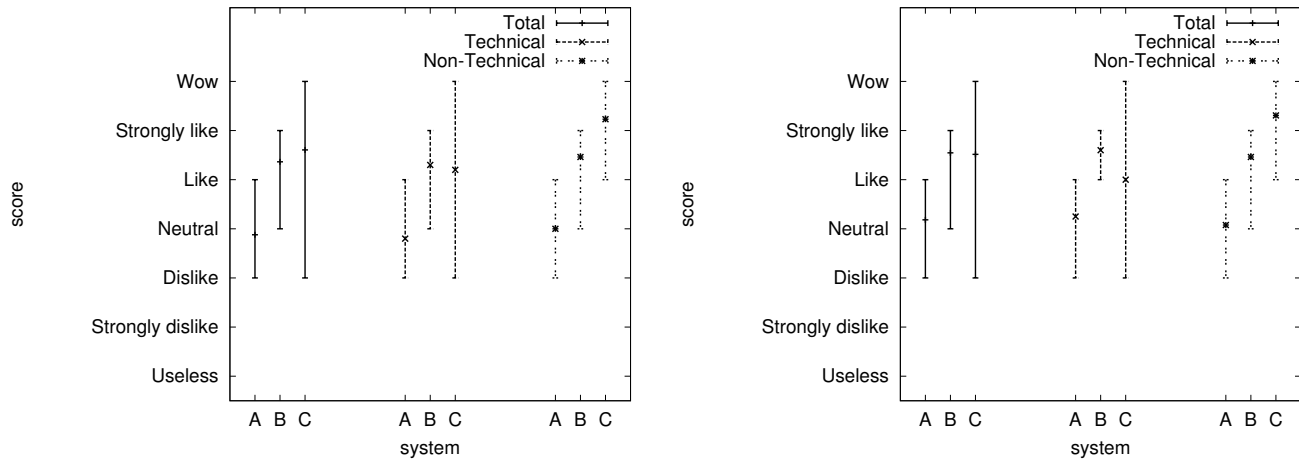
V. EXPERIMENTS AND RESULTS

The efficiency of the technical streaming solution and arbitrary composition of personalized videos have been shown in [5]. For instance, we demonstrated a very short startup latency, smooth/hiccup-free playout of on-the-fly composed videos and seamless video quality adaption to available resources. Our focus here is to investigate if the proposed integration of search and in-page document browsing with vESP adds value to the user experience. We have therefore performed a user evaluation where the users evaluate three different versions of the system. The first approach (figure 1(a)) presented search results using a textual, Google-like interface where returned documents must be opened using an external program like PowerPoint. The second approach (figure 1(b)) adds the possibility to browse slide thumbnails in-page, and vESP (figure 5) adds further support for presenting the corresponding video when clicking on a particular slide.

For the experiments, we used the described data set with technical presentations from Microsoft. The users then evaluated the features of the systems, and after the subjective tests, they answered a short questionnaire about their impression of the systems and how they valued the functionality. The order in which the assessors tested the different systems were changed to see if this affected the results. For each system, we asked what the assessors first impression of the system was and what the value of such a system was (if they would like to use it). Furthermore, we used a 7-point Likert scale with balanced keying to express user satisfaction as shown in table I². In total, we had 33 assessors, all familiar with and frequent users of search systems. Their job positions divide the group in two, i.e., technical (20) and non-technical (13) people.

The results are presented in figure 8 and table II. First, we did not observe any difference with respect to the order they tested the system. Then, we evaluated if the results are statistically significant (unlikely to have occurred by chance) using a non-parametric statistical Friedman test to analyze the

²The table shows the value evaluation scale, but the first impression scale is equal, however, with slightly adjusted comments. Note also that there is not necessarily a linear relation between the values and the keys, i.e., when we later have calculated averages, they do not directly map to a key, but the values are meant as indications of the user preferences.



(a) First impression of the search result page

(b) System value, willingness to use the system

Figure 8. User evaluation results with max/avg/min scores (A = plain, B = document preview and C = document preview + video)

Group of people	System	first impression				system value			
		avg	max	min	stdev	avg	max	min	stdev
Total	A - Plain	3.87	5	3	0.64	4.18	5	3	0.63
	B - Document preview	5.36	6	4	0.65	5.54	6	4	0.56
	C - vESP	5.60	7	3	1.06	5.51	7	3	1.30
Technical	A - Plain	3.8	5	3	0.61	4.25	5	3	0.63
	B - Document Preview	5.3	6	4	0.65	5.6	6	5	0.50
	C - vESP	5.2	7	3	1.05	5	7	3	1.33
Non-technical	A - Plain	4	5	3	0.70	4.07	5	3	0.64
	B - Document preview	5.46	6	4	0.66	5.46	6	4	0.66
	C - vESP	6.23	7	5	0.72	6.30	7	5	0.75

Table II
STATISTICS ABOUT THE USER STUDY

Value	Key	Description
1	Useless	The page is too complex and nobody would ever use it.
2	Strongly dislike	I did not like this search page. I would never use it.
3	Dislike	I do not really need this search page, but maybe my colleagues would use it.
4	Neutral	This search page could be useful sometimes. I would use it occasionally.
5	Like	This search page covers my needs. I would use it regularly.
6	Strongly like	This search page saves time and effort. I would recommend it to my colleagues.
7	Wow	This search page was fantastic! I would use it as my preferred way of viewing presentations.

Table I
THE USED BALANCED KEYING, 7-POINT LIKERT SCALE.

user data. Using a p-value (probability of the outcome) of 1%, regardless of grouped in total, technical or non-technical, is significant for both tests.

With respect to the user evaluations, we first grouped the whole assessor group together (labeled “total”). Then, we observe that the overall impression of the search result page

is improved when adding in-page document preview and again slightly more when adding the video features in vESP (figure 8(a)). For the system value test (figure 8(b)), the trend is almost the same, but the willingness is approximately the same to use the document preview system with and without video. In this respective, we observed that the degree of excitement for the added video feature value heavily depends on the background of the assessors and how they use search.

Non-technical people gave very positive feedback like “tremendous value” and “very impressive”, and they would really like to use the added video feature. For this group, the deviation between the different assessors also is quite small as seen by the standard deviation (similar to the two other systems). All scores are between “liked” (5) and “wow” (7).

In the group of technical people, they were slightly more reluctant. The scores varied more (larger gap between max and min, but also a larger standard deviation), i.e., between “disliked” (3) to “wow” (7). The average is approximately as for the system with document preview. Some liked it a lot. However, they also had comments like “watching a video may take too long time”, even though some added “for me as a developer”. A problem here seemed to be that some of the assessors did not see the feature as one that *could be used*,

but one to *use all the time*.

Nevertheless, in total, vESP was ranked higher than the other systems and was the only system given the top score “wow” – even in the group of technical people. 7 out of 33 in total gave vESP this highest score.

VI. DISCUSSION

The common search engine result case for a decade has been a list of text snippets and urls, and we conjecture that this is ripe for change. An important driver for this is likely the new generation of users familiar with short multimedia downloads. Hence, Google Video, Bing Video, Blinx and YouTube already provide a different approach with thumbnail oriented interfaces. Additionally, focus has not been on video the first decade of search engine developments because of the relative high volume of HTML pages published, and very importantly, that discrete media types lend themselves naturally to precise analysis. A rich and accurate search index provides the foundation for any acceptable search experience. Video metadata extraction is so far in its infancy, in particular when comparing with metadata extraction for discrete media types.

The relative amount of video being published is growing relatively fast, so the demand for better indexing is excelling. As such, we consider our semi-automated video metadata extraction solution as a practical intermediate step towards more general automated metadata extracting techniques.

The vESP prototype demonstrates how multimedia search results can be integrated more tightly. The main novelty in this solution is that it demonstrates a new principle for search engine technology. Prior to our work, search engines were primarily considered mediators guiding users where to locate remote data, and assigning some priority order on the data corpus evaluated. We introduce more computing to this data intensive critical path by converting a search engine into a multimedia production engine for extremely customized results.

Mixing multimedia results is far from novel. Interesting video summarization solutions like, for instance, storyboards and collages [12] are just two examples of novel approaches that are interesting to integrate into next generation of our vESP prototype. What we have implemented and demonstrated so far is mainly the back-end components that support the production and streaming properties. At the same time, we needed a first interface that demonstrates the integration property. This initial approach based on the traditional url and text snippet result pages is what potential users of vESP evaluated. What we evaluate as a start then is the user dialogue part of the system, which we consider has most potential for improvement.

A more thorough user evaluation of our vESP and DAVVI systems, both demonstrating search as a production mediator, is in progress. Anecdotic results from the soccer domain though far exceed our expectations. We also consider the results from the vESP evaluation as preliminary since we need more external users for a more objective evaluation. We

were fortunate to have access to a commercially top-of-the line search engine just before being released, and we even built our own extension prior to this release. Getting objective and external users to evaluate and compare these systems was not an option since the commercial system is in its public release as of this writing.

The traditional search user experience was clearly less regarded when compared with the two alternatives. This was not a surprising result, in particular for this user group. Our vESP when compared with the latest state-of-the art enterprise search engine gave more mixed results. For technology savvy engineers, the initial acceptance rate did not trigger any particular higher acceptance than the same search solution just supporting in-line browsing of PowerPoint slides. The engineers we used in our user satisfaction experiments were all involved in core development of this product, but not in the video extension work primarily done by external academics. Whether this has impact on the subjective user satisfaction is hard to determine. Our conjecture is that the normal work pattern of technology savvy engineers does not have video downloads in the critical path for knowledge sharing. This is a fact that even some of the academic authors of this paper admit confirms with their own user behavior while at work. Research papers and specific developer web sites are still more the rules, videos the exceptions.

However, the more sales oriented users in the experiment rated the vESP video solution better than the two subset solutions. One aspect that might favor the vESP video solution is that this type of personnel tends to use internal company produced videos in their professional workday on a more frequent basis.

VII. CONCLUSION

The rapid growth and popularity of videos published on the Internet also impact search technologies, both through new services being provided and how to structure old ones. We have studied aspects of the latter by integrating video data with existing discrete data results in an enterprise search engine. Initial experience with a concrete application case study using a data subset from a knowledge repository in Microsoft indicates that integration of video can be efficiently done for search, and that potential users really liked to have the opportunity to enrich the traditional search result with video data.

The planned deployment for vESP is first for internal use in a large international software corporation like Microsoft where videos already are used for internal knowledge sharing, but also to communicate to the outside world. We conjecture this trend will continue and probably accelerate, not only due to the sheer volume increase of videos published, but also impacted by new services emerging. The vESP prototype we have built provides such a new service. It not only responds to a search with further references to external resources, it actually pulls down, analyzes and produces specific multimedia results. This is customized both to the user query, but also to provide a

smooth user experience with a good degree of continuity when inspected.

The novelty in vESP is that video data is fully integrated and aligned with the semi-structured data in an enterprise search engine. A traditional textual query is resolved against a traditional search index, but also extracts related video segments from a multimedia storage. The net effect is that a user does not have to retrieve the individual media types and relevant results directly from their sources, but is provided with a high-quality (defined as precise sequence of relevant scenes) video composed by the search engine and streamed to the user.

The use case scenario where we evaluated vESP is for knowledge sharing among employees of Microsoft. The company, like many other large enterprises, already has a large multimedia database for internal knowledge management. Presentations, slides and textual transcripts add to the process of capturing intellectual capital in an enterprise, and vESP brings this even a step further by making it easier to extract and integrate related videos in this process.

ACKNOWLEDGEMENTS

The authors would like to thank Berit Herstad, Pengpeng Ni and Alexander Eichhorn for assistance with the user evaluation experiments.

This work has been performed in the context of the *iAD* (Information Access Disruptions) centre for Research-based Innovation, project number 174867. *iAD* is directed by Fast Search & Transfer (a Microsoft subsidiary) in collaboration with Accenture, Cornell University, University College Dublin, Dublin City University, BI Norwegian School of Management and the Norwegian universities in Tromsø (UiT), Trondheim (NTNU) and Oslo (UiO).

REFERENCES

- [1] "fast ESP," 2010. [Online]. Available: <http://www.microsoft.com/enterprisearch/en/us/Fast.aspx>
- [2] "YouTube," 2010. [Online]. Available: <http://youtube.com>
- [3] "Google Video," 2010. [Online]. Available: <http://video.google.com/>
- [4] "Altus vSearch," 2010. [Online]. Available: http://www.altuslearning.com/altus_vsearch.php
- [5] D. Johansen, H. Johansen, T. Aarflot, J. Hurley, Å. Kvalnes, C. Gurrin, S. Sav, B. Olstad, E. Aaberg, T. Endestad, H. Riiser, C. Griwodz, and P. Halvorsen, "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.
- [6] H. D. Wactlar, M. G. Christel, Y. Gong, and A. G. Hauptmann, "Lessons learned from building a terabyte digital video library," *Computer*, vol. 32, no. 2, pp. 66–73, 1999.
- [7] C.-W. Ngo, T.-C. Pong, and T. Huang, "Detection of slide transition for topic indexing," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2002, pp. 533–536.
- [8] "Move Networks," <http://www.movenetworks.com>, 2009. [Online]. Available: <http://www.movenetworks.com>
- [9] "SmoothHD," <http://www.smoothhd.com>, 2009. [Online]. Available: <http://www.smoothhd.com>
- [10] R. Pantos (ed), "HTTP Live Streaming," <http://www.ietf.org/internet-drafts/draft-pantos-http-live-streaming-00.txt>, 2009. [Online]. Available: <http://www.ietf.org/internet-drafts/draft-pantos-http-live-streaming-00.txt>
- [11] H. Riiser, P. Halvorsen, C. Griwodz, and D. Johansen, "Low overhead container format for adaptive streaming," in *Proceedings of the Multimedia Systems Conference (MMsys)*, Feb. 2010.
- [12] M. G. Christel, A. G. Hauptmann, H. D. Wactlar, and T. D. Ng, "Collages as dynamic summaries for news video," in *Proceedings of the ACM International Multimedia Conference (ACM MM)*, 2002, pp. 561–569.