Video Search: New Challenges in the Pervasive Digital Video Era

Katerina Pastra and Stelios Piperidis

Institute for Language and Speech Processing,

Artemidos 6 and Epidavrou, 151-25, Maroussi, Greece,

email `kpastra,spip@ilsp.gr`

Abstract

The explosion of multimedia digital content and the development of technologies that go beyond traditional broadcast and TV have rendered *access* to such content important for all end-users of these technologies. While originally developed for providing access to multimedia digital libraries, video search technologies assume now a more demanding role. In this paper, we attempt to shed light onto this new role of video search technologies, looking at the rapid developments in the related market, the lessons learned from state of art video search prototypes developed mainly in the digital libraries context and the new technological challenges that have risen. We focus on one of the latter, i.e., the development of cross-media decision mechanisms, drawing examples from REVEAL THIS, an FP6 project on the retrieval of video and language for the home user. We argue, that *efficient* video search holds a key to the usability of the new "pervasive digital video" technologies and that it should involve cross-media decision mechanisms.

Keywords: Interactive TV, Video Search, Video Retrieval, Pervasive Digital Video, Digital Libraries,

First presented at the 4th European Interactive TV Conference EuroITV 2006, extended and revised for JVRB

REVEAL THIS

1 Introduction

The proliferation of digital multimedia content and the subsequent growth of the number of digital video libraries have boosted research on the development of video retrieval systems. Video search technology has traditionally been conceived as a way to enable efficient access to *large video data collections*, with automatic indexing and cataloguing of this data being an essential derivative of its development. The quest for efficient video retrieval mechanisms is still ongoing, with a number of different unimodal and multimodal techniques being implemented and evaluated (cf. TRECVID competitions, [HC04]).

However, the emergence of technologies crossing the boundaries between traditional broadcast and the Internet, and between traditional television and computers broadens the scope of developing video search functionalities. Does this, new, reinforced role of video search render it indispensable for end-users of the new technologies? Within the rapidly changing multimedia-processing context, does video search become an even more challenging task in terms of the retrieval performance required for achieving high usability of the new technologies it is embedded in?

In this paper, we look into the role of video search in the light of the new "convergent" technologies and the technological challenges that are subsequently posed on its development. In order to do so, we present the market status and trends in video search for the new technologies, as well as the search mechanisms used within commercial and research prototypes. We discuss the effects of the new role subsumed by video search and focus mainly on the use of cross-media decision mechanisms for dealing with such effects. Last,

we present REVEAL THIS, a research project which attempts to implement cross-media mechanisms for increasing the usability of both its pull and push video access scenarios.

2 Market Status and Trends in Video Search

The formation of regulations for the new technologies that enable traditional broadcast and the internet to converge (e.g. Internet Protocol TV (IPTV), Peer-to-Peer (P2P) networks, Mobile TV) proves that these technologies -with the ever growing popularity- have become something more than a trend or optimistic prospect ([Car05, Bro05]). They already are a new reality, in which:

- Traditional TV sets can be extended with intelligent digital video recorders (DVRs), set-top boxes with PC-like functionalities, or can even communicate with personal computers for displaying streamed digital media through gaming consoles enabling an enhanced, interactive TV experience [SBH+05].

- TV viewing goes beyond traditional TV sets, in mobiles and portable digital media players (i-pods) enabling on-the-move TV watching [Cha05].

- File-swapping networks and headline syndication technology facilitates the exchange of not only professional/copyrighted TV programmes but also of consumer-generated ones (e.g. podcasts, video blogs, etc.) [Bor05, Jaf05].

In opening new markets, suggesting new business models, and indicating new content distribution channels, all these technologies boost the availability of digital video content and motivate digitisation. In this light, the question of easy and efficient access to video content that was once posed in relation to digital video collections raises again more demanding than ever.

In this section, we will look into the role of video search within this new context. This will shed light on the implications this new role has on the development of video search mechanisms and will point towards possible aspects that need to be taken into consideration for developing highly usable video access services. We present the interested parties in the market of the new TV/video technologies and their stance towards video search functionalities.

2.1 The market players

Electronics manufacturers, software companies, telecommunication giants, cable companies, broadcasters, content owners all have an interest in the new market that is being created and which extends beyond professional users to everyday laymen, to home users. As expected, the convergence of internet and TV has not only resulted in the convergence of the corresponding business sectors, but has also created new ones, which are interested in providing end-to-end services, i.e., aggregation of video (and other) digital content and distribution of this content to interested users through a push (data routing according to a user profile) or pull (data search and retrieval) model.

The business sectors which are being actively involved in providing an enhanced TV experience (i.e. IPTV, i-TV, mobile TV etc.) could be classified in the following categories according to their main business activities[1]:

a *Content owners:* production companies (broadcasters who also produce their own content e.g. BBC are included in this category too)

b *TV service providers:* satellite and cable companies (e.g. Comcast, BeTV etc.) and TV broadcasters (e.g. CNN, RTBF etc.)

c *P2P service networks:* networks that allow for file-swapping e.g. BitTorrent, e-donkey etc.

d *Electronics manufacturers:* manufacturers of DVRs and set-top boxes (e.g. TiVo, Akimbo, Scientific Atlanta etc.), portable digital media players (e.g. Apple), mobiles (e.g. Siemens).

e *Computer networking companies* (e.g. Cisco), *Internet Service Providers* (e.g. AOL) and *Phone companies* (e.g. BellSouth, Verizon etc.)

f *Internet Protocol TV software developers* (e.g. Microsoft, Myrio, Virage)

g *Content service providers:* content monitoring companies which provide push and/or pull services (e.g. TVEyes, BlinkX)

---

[1]The categories are not mutually exclusive; on the contrary, there are organisations with a wide range of activities, which span more than one category.

h *Web content aggregators:* companies that aggregate digital media (text, audio, video) or links to these media and present them online to a user upon request/search e.g. Google and Yahoo

i *Content re-packaging companies:* companies that acquire content e.g. sports videos/TV programmes and re-package it for meeting various user needs e.g. interactive viewing of a car race, where the user can choose his/her preferable viewing angle(s) (e.g. Nascar)

All these market players have their own interest in the new technologies and therefore attempt to extend the services they provide by either ***enriching in-house developments*** or by ***making alliances*** with each other. Figure 1 illustrates the relative position of each business sector in the chain from content owners to end-users of enhanced-TV. The chain reveals the dependencies between the different market sectors for achieving an end-to-end service; these dependencies justify the increasing number of alliances, partnerships and mergings among key players in these sectors and the tendency of some businesses to extend their own development activities in order to avoid such collaborations. The dependencies illustrated in figure 1 are transitive, namely, if X is dependent on Y, and Y is dependent on Z, then X is dependent on Z too. Dependency, in this case, denotes a need for a company X to collaborate with a company Y that belongs to another business sector (i.e. provides different services or/and products) or a need for company X to compensate for the Y services/products on its own, by extending its own services/products.

In particular, content owners are essential in allowing digitisation/monitoring/capture of and access to TV/video content; all parties interested in allowing users to watch proprietary content need agreements with them. They are themselves interested in taking advantage of the new content distribution channels that are opened. TV service providers and phone companies (cellular and other telecommunication giants) are also competing for entering the IPTV market; for the former this is a natural extension to their current services for which they already have a successful business-model, while the latter consider this a new opportunity [Rea05a] for extending their internet service provision services or/and allowing mobile-TV.

On the other end, electronics manufacturers want to make the best out of the functionalities of the devices they develop. TiVo, for example, has extended the features of its DVRs, allowing the user to download movies from the internet, buy products, search local movie theatre listings etc. [Shi06]. Computer networking companies such as Cisco are also extending their reach to the consumer networking market and IPTV for set-top boxes, in particular, through a number of company acquisitions [Rea05b]. In most cases, in-house IPTV software comes part-and-parcel with the products of the electronics manufacturers (i.e. set-top boxes, DVRs etc.). When not, alliances or acquisitions of businesses take place, with IPTV software developers playing a key-role; for example, Siemens has acquired Myrio in its attempt to enter the IPTV market [Rea05a]. Deals between more than one business sectors are also growing; for example, Microsoft, BT and Virgin are collaborating for providing mobile TV services, Microsoft contributing all software needed for packaging and viewing TV on mobiles, BT contributing a new network for mobile TV and Virgin making use of the network and providing the service in a new mobile phone [Rea06]. However, if IPTV software is what makes enhanced-TV-experience possible, video search software seems to be what makes it usable [Del05].

2.2 Trends in providing video search services

The new business sectors that have emerged (g to i above) rely heavily on video/digital media search technology. This is the main asset and the source of competition for both content service providers and web content aggregators. For example, TVEyes e-mails links to digital media from a number of sources to a user according to his profile and/or his queries, it allows a user to search both archived and real time broadcast content, it has extended its search functionalities to podcasts and blogs [Pog05] and has even included a desktop search functionality to its engine. One of its main competitors, Blinkx, allows also for an integrated search providing results from a user's desktop and the web, including TV content, blogs, images, P2P content and for the creation of a customized TV channel based on topics/keyword searches provided by the user. Automatic continuous streaming video is played whenever the user logs in and can be downloaded to a PC [Mil05]. The search tool is proactive, in the sense that it looks for information related to what the user is working on in other programs, without waiting for the user to initiate a search [Reg06].

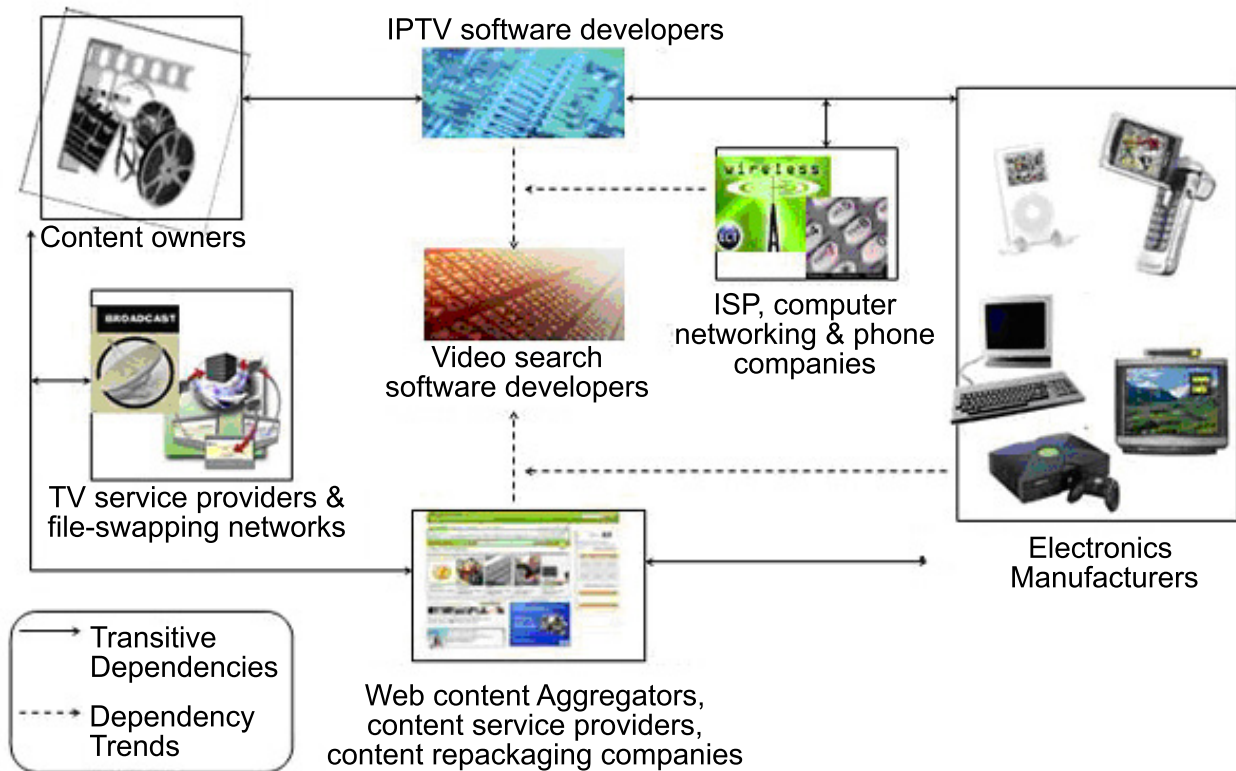On their turn, some of the big search engines have

Figure 1: Dependencies & dependency trends among different business sectors

extended their searches to video too (cf. Yahoo video and Google video - [Ols05]), while others, such as AOL, have acquired video search software companies (e.g. SingingFish and Truveo) for accommodating these needs [OM05]. What is more important to note though, is that other business sectors involved in the enhanced-TV-experience market have realised the importance of video search technology too and have proceeded in partnerships with video search market players. TiVo, for example, has partnered with Yahoo: from any yahoo TV episode page found, the user can click on a "record to my TiVo box button" and schedule the recording through the yahoo portal [Sin05]. Microsoft has also been developing its own video search software, which could possibly be used by telecom operators who already make use of its IPTV software [LaM05].

Furthermore, electronics manufacturers make alliances with content aggregators (e.g. Motorola and Google deal, Intel and Google etc.), and so do phone companies (e.g. Vodafone and Google deal) in order to include video search in the services they provide. Partnerships between content owners and content service providers are another sign of the growing need of all business sectors in the TV/video market to manage and distribute content efficiently to their end-users. For example, the Press Association, BBC and TVEyes partnership in developing the "Politics Today" portal, in which users can search for and watch audio and video files from e.g. the British parliament along with related news articles reveals a tendency of content owners for going beyond traditional archival indexing and retrieval of their collections for their own professional use to providing such access to laymen [Rel05].

Figure 1 illustrates these direct and indirect dependencies (dependency trends) of some business sectors on the use of video search technology.

2.3 Employed video search mechanisms

While most implementation details are kept aside, it seems that the video search mechanisms currently in the market have one or more of the following characteristics:

- They make use of *metadata* provided for each video/TV programme by the content owners

or/and the broadcaster (e.g. information also available in electronic programme guides). Microsoft, AOL and Yahoo rely exclusively on such data; the latter has also developed a headline syndication based protocol for facilitating publishers in adding metatags to media files (Media RSS) [Ols05].

- They are *text-based* and actually run on closed captions (e.g. Google) or/and on automatically generated transcriptions of the audio (e.g. TVEyes). BlinkX, on its turn, searches for keywords in the audio track by transforming a keyword into a concatenation of phonemes and matching the latter with the audio track.

- They apply mostly to *English* video files.

- They are mainly keyword-based searches with semantic expansion being quite restricted (e.g. Google identifies concepts associated to the textual keywords)

- The video retrieved is either the entirety of a programme (or a part of its structure e.g. a film chapter) or a short segment consisting of a few seconds before and after the keyword matched [Inc05].

Will such video search mechanisms suffice for efficient access to the increasing number of digital video content? The above-mentioned video search services are still in their infancy, with beta versions of the engines having been released in early summer of 2005; therefore, no usability evaluation of these services is currently available. Still, some problems with such mechanisms are evident:

- Metadata creation is time and effort consumming and in many cases video files are accompanied by limited or no such information. Especially in the case of consumer-generated video content, metadata cannot be taken for granted for retrieving a video file

- Closed captions are not always available for TV programmes (not available at all for other types of video data), while automatic transcriptions are not very robust for being the sole source of retrieval; especially in cases when the recording conditions are noisy (e.g. field interviews), automatic speech recognition (ASR) degrades considerably [Hau05].

- There is an abundance of non-English video/TV programmes of interest to users, which currently cannot be retrieved, unless English metadata are associated to them.

- Going beyond keyword matching to intelligent processing through query expansion and semantic search strategies is important not only for achieving better precision and recall in retrieval, but mainly for making the best out of queries formed by laymen (i.e. uninitiated everyday users of the video search services).

- Users are expected to lose interest, if they have to watch long video files until they (probably) reach the segment that is directly related to their query. On the other hand, sufficiently long segments are necessary for users being properly informed regarding their query.

The quest for more accurate and effective video search mechanisms has been undertaken within research prototypes. We turn to them for suggestions on developing more efficient video search mechanisms.

3 Video Search in Research Prototypes

In this section, we first present the lessons learned from state of art prototypes that have been developed and evaluated within a "digital collection" access scenario; in doing so, we explore whether they go beyond the above-mentioned video search drawbacks and if so, to what extent their suggestions could be implemented in the new enhanced-TV context. We discuss how differences in the context of use affect the development of video search mechanisms and turn to suggestions on effective video search for the new technologies by recently completed and on-going research projects.

3.1 Video search mechanisms in digital collection scenarios

The large digital library initiatives in mid-nineties boosted research on the development of video analysis technologies and in particular of video indexing and retrieval. Projects such as Informedia [Hau05] and Fischlar [Sme05] explored a variety of unimodal and/or

multimodal mechanisms for intelligent indexing, retrieval, summarization and visualization of mainly broadcast news files.

In 2001, the need to compare the different video retrieval approaches/prototypes led to the organisation of a Video track within the Text Retrieval and Evaluation Conference (TRECVID), a competition that is being organised every year since [HC04, Sme05]. In this evaluation campaign, competing systems are given a test video collection (mainly broadcast news in English), a number of statements expressing an information need (topic/query) and a common shot boundary reference (retrieval unit) and they are asked to return a ranked list of at most a thousand shots which best satisfy the need. The users who form the queries are trained analysts and the queries themselves may consist of a mix of visual descriptions and semantic information (e.g. "find shots zooming on the US Capitol dome") [Sme05].

Competing systems implemented a variety of indexing and retrieval mechanisms, which were able to tackle lack of metadata and closed captions, degraded ASR output and strict keyword matching. It was generally found that [HC04]:

- Systems that performed medium-specific analysis (e.g. automatic speech recognition, face detection, image analysis etc.) and which subsequently fused the medium-specific individual retrieval scores for each shot were found to perform slightly better than systems which relied on e.g. text-based only retrieval mechanisms.

- Learning the best linear weights for fusing such multimedia information according to the type of the query was found to be helpful for retrieval.

- Text query enhancement through lexical resources or associated information found in web documents retrieved by a search engine (cf. also work in the PRESTOSPACE project, [DTCP05]), relevance feedback mechanisms [Jon04] and use of high-level concepts associated with low-level features, all seemed to contribute in more accurate video retrieval [Sme05].

These findings point to *slight* benefits from the implementation of multimedia approaches for video indexing and retrieval, ones that make use of both visual and linguistic analyses of video content. Still, they fail to convince on the *necessity* of going beyond text-based retrieval and actually the query-type based fusion they suggest opens a number of research questions that remain to be addressed [Hau05].

The TRECVID evaluation setup itself (and the participating systems) is more focused on a digital video collection application scenario and remains detached from the "pervasive digital video" reality. The users of its video search scenario are not the everyday laymen/novice computer users reached by the new enhanced-TV services. The queries/information need statements used are not relevant to the new context of convergent technologies and the search unit is a keyframe (representative of a shot retrieved for the user) rather than a video segment that will be presented to the user. A structured video collection with no variety of genre/domain (mostly news broadcasts) is presupposed. The development of query-type dependent multimedia mechanisms for video retrieval relies on the availability of a "query and corresponding retrieved videos" corpus for training a system; given that there is a non-finite type of queries a user may submit in the enhanced-TV context, this is an unrealistic approach for the corresponding video search services.

All these must be taken into consideration when drawing conclusions regarding the direction the development of video search mechanisms in the new context should take. It is no coincidence that stronger suggestions on the need for multimedia approaches to video search, come from research projects that implement applications for leisure and entertainment within the "digital home" and/or for the "mobile user". Projects such as UP-TV [TPKC05], BUSMAN [XVD+04] and AceMedia [BPS+05, PKS+04] are characteristic cases; the projects consider the association of low-level image features and higher-level concepts important for video indexing and retrieval and develop tools and formalisms for the corresponding semi-automatic annotation. Such an association goes beyond query-type dependent learning for multimedia fusion and actually takes the TRECVID findings a step further. In what follows, we look into the parameters that effect the development of video search technologies for use in the enhanced-TV context and which have led researchers to focusing on the image-language association need.

3.2 Technological challenges in the new video search context

In going from digital video libraries to the pervasive digital video reality a number of video access aspects have been extended affecting the development of video search mechanisms:

From advanced computer *users* to laymen

While digital video library users were advanced computer users, the enhanced-TV context reaches everyday people, uninitiated computer users. This affects not only the type of queries (information need) formed, but also their quality and accuracy, the expectations/demands on video retrieval accuracy, the length of the retrieved unit, the domain and language of the retrieved data.

From structured *data* collections to pervasive video data

Digital video collections consist of data of a specific genre and domain (or of finite number thereof), with expected content and structure and of professional quality. The availability of any type of video in the "convergent technologies" reality entails that video access mechanisms should deal with not only professionally made videos, but also consumer-generated ones of probably lower quality, and more noise. Furthermore, abundance of video means non-finite variety in domains and languages, language genres, pronunciations, visual imagery and even cues revealing programme structure. The variety of sources of data implies that associated metadata will not always be of the same granularity or even of the same type and in many cases will not even be available.

From static to dynamic *search*

Within the enhanced-TV context, video search becomes dynamic in two senses; first of all, not only archived data but also real-time broadcast data needs to be searched, and actually the timely delivery of such data to the user can become a competitive advantage for a content service provider. This imposes time-constraints to a search mechanism. On the other hand, the search model itself goes beyond traditional search triggered by a user through an interface to pro-active, personalised search for a user according to his/her known profile and interests. Effectiveness in indicating "interesting" data for the user becomes even more crucial and personalization is rendered central for the functionality of the search mechanism.

Based on the above, one could argue that the ideal video search mechanism for use by e.g. a content service provider, should be *language, domain and genre independent*, should be able to analyse any medium-specific part of the video (i.e. audio, image, subtitles etc.) *robustly*, to *structure* the video (i.e. identify meaningful, comprehensive stories/segments in it covering a specific topic for presentation to the user), to take user preferences into consideration and work in *real time*, with the highest *precision*. However, technology is far from achieving this. As mentioned earlier, recent research projects that aim at developing video access prototypes for applications within the enhanced-TV context, focus on the multimedia nature of videos for dealing with this challenging task, and in particular on the association of semantically equivalent medium-specific pieces of information extracted from a video.

In the next section, we present one such project, REVEAL THIS, which takes things a step even further, suggesting that one should go beyond this association to crossing media for performing efficient video indexing.

4 The REVEAL THIS Suggestion: cross-media decision mechanisms

The REVEAL THIS project (REtrieval of VidEo And Language for The Home user in a Information Society) is an EC funded FP6 project which aims at developing software for efficient access to multimedia content[2]. The envisaged use scenario of the technology that is being developed in the project is illustrated in figure 2.

Web, TV and /or Radio content monitored by the service provider is fed into the REVEAL THIS prototype, it is being analysed, indexed, categorized and summarized and stored in an archive. This data can be searched and/or can be pushed to a user according to his/her interests. Novice and advanced computer users are targeted; the former will be using mainly a simple mobile phone interface where information will be pushed to, while advanced computer users will use a web interface for searching for information. Cross-lingual retrieval is enabled (i.e. the user can issue a query in one language and retrieve data in another), while the summaries presented to the user are not only translated to the user's language but are also personalized according to a user's interest or specific query. English and Greek for EU politics, news and travel-related data are handled by the prototype.

Figure 3 illustrates the system architecture; input

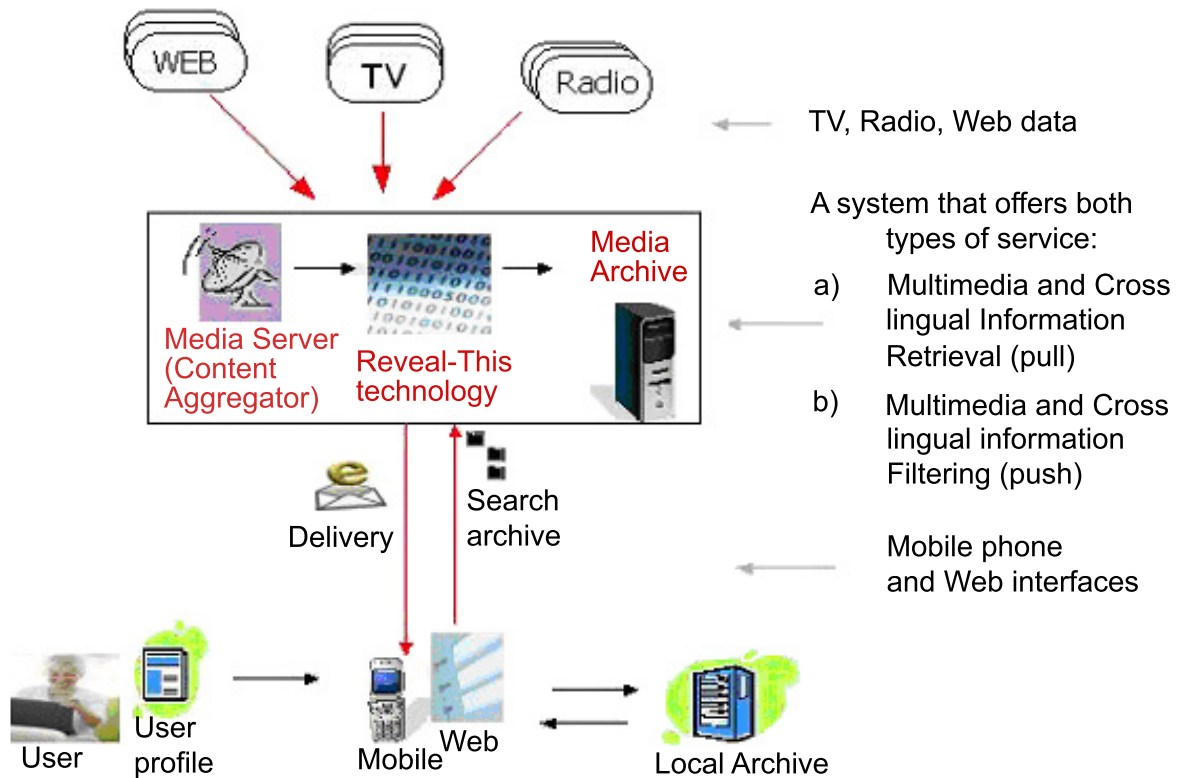[2]For more details cf. http://www.reveal-this.org and [PP05].

Figure 2: The REVEAL THIS use scenario

data is analyzed by a number of different medium-specific processors (i.e. speech processing component, image analysis component, etc.) each of which provides its own interpretation of the data contributing to the "bag" of time-aligned (in the case of audio and video files) medium-specific pieces of information (e.g. speaker turns, named entities, facts, image categories, face ids, text categories etc.). These pieces of information are available for the more complex modules, which are responsible for indexing, categorization and summarization.

These more complex modules are viewed as *decision mechanisms* that should take advantage of the semantic interaction between the different medium-specific pieces of information available for a video segment. The latter is the result of a similarly viewed segmentation mechanism, which stands as the processing (indexing- categorization- summarization-) unit for these modules. These units/segments are semantically coherent video segments covering a particular topic. The REVEAL THIS attempts to develop cross-media decision mechanisms for video segmentation, video indexing, video categorization and multimedia summarization are still ongoing. What are these mech-

anisms supposed to decide on though?

Crossing-media for achieving a task has normally taken the form of finding semantic equivalences between different media both of which express the same information (e.g. a TV news programme and a corresponding web news article) [BKW99]. REVEAL THIS suggests mechanisms for crossing-media *within* the same multimedia document (i.e. video), where one needs to go beyond the notion of semantic equivalence to other semantic interaction relations.

Cross-media decision mechanisms go a step further from medium-specific feature fusion suggested in the literature (cf. section 3.1) and attempt to decide whether two medium-specific representations:

a are associated (i.e. there is a semantic equivalence in their meaning) - what multimedia integration does [PW04] (e.g. the image of the European Parliament building and the corresponding textual label),

b are not associated but they complement each other, they collaborate in forming a message (e.g. the image of Josep Borrell and the audio of his
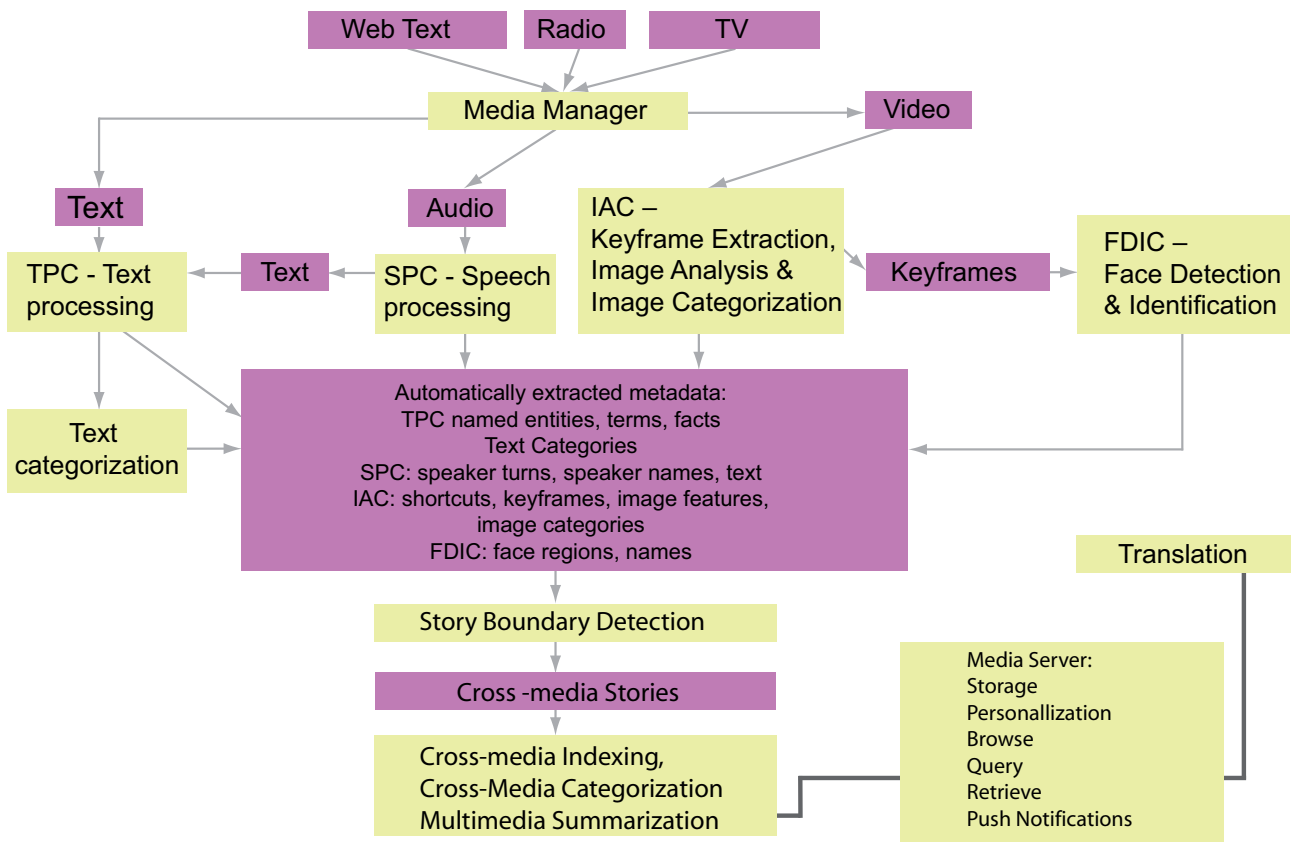
Figure 3: The REVEAL THIS system architecture

speech on the European Commission stance on the war in Iraq) , or

c are not associated, they do not collaborate in forming a message, they are actually *semantically incompatible* (contradicting), due to e.g. errors in automatic medium-specific analysis/interpretation

Taking as an example the case of a cross-media indexing mechanism for a video retrieval system, one could say that once the mechanism takes as input the medium-specific analysis of a video segment (e.g. the output of the image analysis component, the face detection component, the speech processing component and the natural language understanding component), it has to decide which medium-specific pieces of information (medium-specific interpretations of the video) are more representative of the content of the segment for indexing.

In case (a), the association indicates a highly significant piece of information, which could be used as an indexing term. In case (b), a conjunction of different medium-specific pieces of information forms the

indexing term. In case (c), the indexing mechanism will have to choose (a) piece(s) of information from a specific medium that is to be trusted as more reliable for expressing the video segment content. This choice could be domain- and/or genre-specific, it could rely on each medium processor's confidence scores etc. The third case is particularly important in view of the technological challenges mentioned in section 3.2.

One could go on elaborating the notion of cross-media decision mechanisms addressing also the methods that could be used for implementing such mechanisms. However, this goes beyond the scope of this paper, which aims at suggesting these mechanisms as a direction in video search research, which seems more appropriate in the new "pervasive digital video reality".

5 Conclusion

In this paper, we looked into the growing market of the convergent-technologies that go beyond traditional broadcast and TV and in particular the effects on and

from video search services in this context: market trends have shown that video search technology is indispensable for those involved in the product chain from TV/video content to the end-user. We further presented the characteristics of the video search services provided in this market and the ones of the more advanced video search techniques developed in state of the art prototypes. The latter point to the fact that one should take advantage of the different medium-specific pieces of information that can be extracted from a video; however, they fail to convince on the necessity for doing so which could be due to their restricted development and evaluation within a digital libraries scenario.

We, therefore, discussed the effects on the development of video search mechanisms when shifting from the digital libraries application scenario (for which most of these systems were developed) to the pervasive digital video context; this shift was found to render the need for taking advantage of the multimedia nature of videos more demanding and to make the suggestion of crossing-media for the task worth investigating. Thus, we presented the notion of cross-media decision mechanisms as suggested by the RE-VEAL THIS project, a project that attempts to implement such mechanisms for the home user.

We have, therefore, concluded that efficient video search is *indispensable* for highly usable technologies that go beyond traditional broadcast and TV and that *cross-media decision mechanisms* may hold the key for its development.

The rapid developments in the new "pervasive digital video" market require that we step back from core research and development in video analysis and retrieval, in order to look at the new role video search technologies have come to play and the subsequent technological challenges that have risen, to learn from the last decade of developing video search mechanisms within the digital libraries context and re-direct research accordingly; otherwise, there is a risk of developing technologies that are detached from the actual needs of the target users, detached from the new reality that is rapidly being formed around us. In this paper, we attempted to take this step back and get an overview of the video search related situation.

6 Acknowledgements

References

[BKW99] Susanne Boll, Wolfgang Klas, and Jochen Wandel, *A Cross-Media Adaptation Strategy for Multimedia Presentations*, Proceedings of the seventh ACM international conference on Multimedia (Part 1) (New York, NY, USA), ACM Press, 1999, ISBN 1-58113-151-8, pp. 37–46.

[Bor05] John Borland, *File-swap TV comes into focus*, http://www.news.com, August 8th, 2005, 2005, Last visited July 31st, 2006.

[BPS+05] Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Vassilis Tzouvaras, Yannis S. Avrithis, Siegfried Handschuh, Ioannis Kompatsiaris, Steffen Staab, and Michael G. Strintzis, *Semantic Annotation of Images and Videos for Multimedia Analysis*, Proceedings of the Second European Semantic Web Conference, ESWC 2005, Lecture Notes in Computer Science, Vol. 3532, Springer, 2005, ISBN 3-540-26124-9, pp. 592–607.

[Bro05] Anne Broache, *Digital TV changeover suggested for 2009*, http://www.news.com, July 12th, 2005, 2005, Last visited July 31st, 2006.

[Car05] Doreen Carvajal, *New eu rules for television: content without frontiers?*, The International Herald Tribune, November 28, 2005 Monday, 2005.

[Cha05] B. Charney, *Free tv for your mobile*, http://www.news.com, May 17th, 2005, 2005.

[Del05] Kevin J. Delaney, *Tv's future may be web search engines that hunt for video*, Wall Street Journal, December 16, 2004, Thursday, Section B; Page 1, Column 5, 2005.

[DTCP05] Mike Dowman, Valentin Tablan, Hamish Cunningham, and Borislav Popov, *Content Augmentation for Mixed-Mode News Broadcasts*, Proceedings of the 3rd European Conference on Interactive TV: User Centred ITV Systems, Programmes and Applications, 2005.

[Hau05] Alexander G. Hauptmann, *Lessons for the Future from a Decade of Informedia Video Analysis Research*, Image and Video Retrieval, 4th International Conference, CIVR 2005, Lecture Notes in Computer Science, Volume 3568, 2005, ISBN 978-3-540-27858-0, pp. 1–10.

[HC04] Alexander G. Hauptmann and Michael G. Christel, *Successful Approaches in the TREC Video Retrieval Evaluations*, Proceedings of the 12th annual ACM international conference on Multimedia, SESSION: Brave new topics – session 3: the effect of benchmarking on advances in semantic video, 2004, ISBN 1-58113-893-8, pp. 668–675.

[Inc05] StreamSage Inc., *Rich Media Indexing Technology Approaches*, White Paper, March 2005, 2005.

[Jaf05] Joshua Jaffe, *Online travel-guides: the next generation*, http://www.news.com, September 28, 2005, 2005.

[Jon04] Gareth J. F. Jones, *Adaptive Systems for Multimedia Information Retrieval*, Adaptive Multimedia Retrieval, First International Workshop, AMR 2003, Lecture Notes in Computer Science, Volume 3094, 2004, ISBN 978-3-540-22163-0, pp. 1–18.

[LaM05] Martin LaMonica, *Microsoft signs on Alcatel for IPTV*, http://www.news.com, February 22nd, 2005, 2005, Last visited July 31st, 2006.

[Mil05] Elinor Mills, , http://www.news.com, October 3, 2005, 2005, Last visited July 31st, 2006.

[Ols05] Stefanie Olsen, *Yahoo, Google turn up volume on video search battle*, http://www.news.com, May 4, 2005, 2005, Last visited July 31st, 2006.

[OM05] Stefanie Olsen and Elinor Mills, *AOL launches video search service*, http://www.news.com, June 30, 2005, 2005, Last visited July 31st, 2006.

[PKS⁺04] Kosmas Petridis, Ioannis Kompatsiaris, Michael G. Strintzis, Stephan Bloehdorn, Siegfried Handschuh, Steffen Staab, Nikos Simou, Vassilis Tzouvaras, and Yannis S. Avrithis, *Knowledge Representation for Semantic Multimedia Content Analysis and Reasoning*, Knowledge-Based Media Analysis for Self-Adaptive and Agile Multi-Media, Proceedings of the European Workshop for the Integration of Knwoledge, Semantics and Digital Media Technology, EWIMT 2004, 2004, ISBN 0-902-23810-8.

[Pog05] D. Pogue, *A very cool podcast service*, The New York Times, July 27th, 2005, 2005.

[PP05] Stelios Piperidis and Harris Papageorgiou, *REVEAL THIS: retrieval of multimedia multilingual content for the home user in an information society*, The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, EWIMT 2005, 2005, ISBN 0-86341-595-4, pp. 461–465.

[PW04] Katerina Pastra and Yorick Wilks, *Vision-Language Integration in AI: A Reality Check*, Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, 2004, ISBN 1-58603-452-9, pp. 937–944.

[Rea05a] Marguerite Reardon, *Siemens tackles Microsoft IPTV dominance*,

http://www.news.com, June 13th, 2005, 2005, Last visited July 31st, 2006.

[Rea05b] Marguerite Reardon, *Tech firms focus on TV*, http://www.news.com, November 23rd, 2005, 2005, Last visited July 31st, 2006.

[Rea06] Marguerite Reardon, *Microsoft, BT, Virgin team on mobile TV*, http://www.news.com, February, 14th 2006, 2006, Last visited July 31st, 2006.

[Reg06] Keith Regan, *Revamped Blinkx Program to Deliver Automatic Searches*, E-Commerce Times, July 3rd, 2006, 2006, Last visited July 31st, 2006.

[Rel05] PA Press Release, *Pa business website set to revolutionize parliamentary monitoring*, The Press Association, November 21st, 2005, 2005.

[SBH⁺05] Richard Shim, John Borland, Eva Hansen, Stefanie Olsen, and B. Schuler, *MeTV: Finally you are in control*, http://www.news.com, April 2005, 2005, Last visited July 31st, 2006.

[Shi06] Richard Shim, *Tivo beefs up patent portfolio*, http://www.news.com, April 6th, 2005, 2006, Last visited July 31st, 2006.

[Sin05] Michael Singer, *Yahoo, TiVo to connect services*, http://www.news.com, November 6th 2005, 2005, Last visited July 31st, 2006.

[Sme05] Alan F. Smeaton, *TRECVID: 5 Years of Benchmarking Video Retrieval Operations*, MUSCLE/ImageCLEF Workshop 2005, 2005.

[TPKC05] Chrisa Tsinaraki, Panagiotis Polydoros, Fotis Kazasis, and Stavros Christodoulakis, *Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content*, Multimedia Tools and Applications **26** (2005), 299–325, ISSN 1380-7501.

[XVD⁺04] Li-Qun Xu, Paulo Villegas, Mónica Díez, Ebroul Izquierdo, Stephan Herrmann, Vincent Bottreau, Ivan Damnjanovic, and Damien Papworth, *A User-Centred System for End-to-End Secure Multimedia Content Delivery: From Content Annotation to Consumer Consumption*, International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, Volume 3115, 2004, ISBN 3-540-22539-0, pp. 656–664.