Web-based semantic browsing of video collections using multimedia ontologies

Marco Bertini, Gianpaolo D'Amico, Andrea Ferracani,
Marco Meoni and Giuseppe Serra
Media Integration and Communication Center, University of Florence, Italy
{bertini, damico, ferracani, meoni, serra}@dsi.unifi.it
http://www.micc.unifi.it/
SUBMITTED to ACM MULTIMEDIA 2010 DEMO PROGRAM

ABSTRACT

In this technical demonstration we present a novel web-based tool that allows a user friendly semantic browsing of video collections, based on ontologies, concepts, concept relations and concept clouds. The system is developed as a Rich Internet Application (RIA) to achieve a fast responsiveness and ease of use that can not be obtained by other web application paradigms, and uses streaming to access and inspect the videos. Users can also use the tool to browse the content of social and media sharing sites like YouTube, Flickr and Twitter, accessing these external resources through the ontologies used in the system. The tool has won the second prize in the Adobe YouGC[1] contest, in the RIA category.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Algorithms, Experimentation

Keywords

Video retrieval, browsing, ontologies, web services

1. INTRODUCTION

Currently, the most common approach to access and inspect a video collection is by using a video search engine. Typically such systems are based on lexicons of semantic concepts, presented as lists or trees, and let to perform keyword-based queries [1]. These systems are generally desktop applications or have simple web interfaces that show the results of the query as a ranked list of keyframes [2,3]. Video browsing tools are developed aiming more at summarization of video content, as in [4] where different visual features are used to provide an overview of the content of a single video, or aiming at the suggestion of new query terms, as in [5]. In other approaches, e.g. [6], the content of a video collection is clustered according to some visual features, and users browse the clusters to inspect the various instances of a concept. Similarly to the interfaces of video search engines, also these browsing tools are desktop based applications or more rarely form-based web applications, that have relatively limited user interaction and simple presentation of results as lists and tables. Finally, all these systems are designed to work only on a single repository of videos, missing the opportunities to exploit the large amount of multimedia data now available on the web from multimedia sharing sites like Youtube and Flickr.

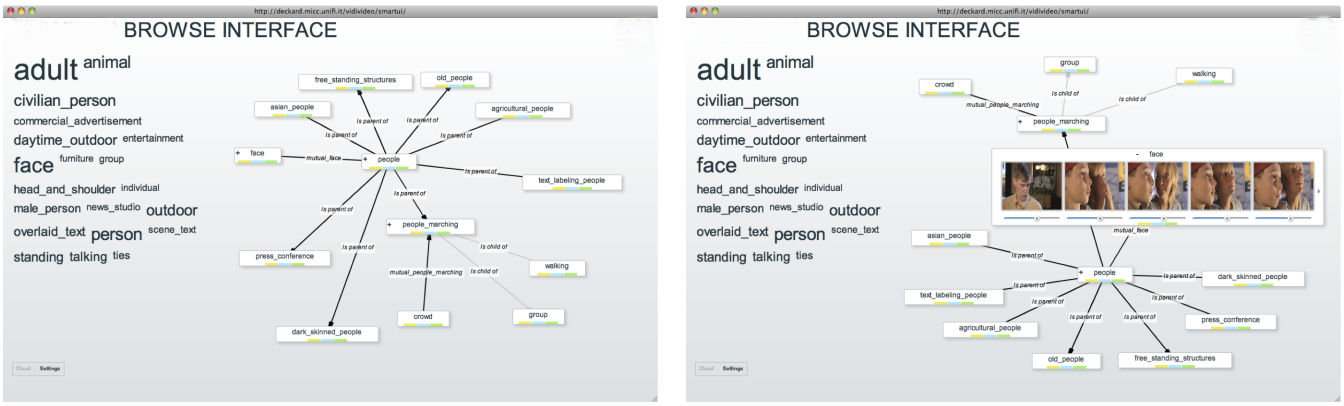
In this demonstration we present a web video browsing system that allows semantic access to video collections of different domains (e.g. broadcast news and cultural heritage documentaries), with advanced visualization techniques derived from the field of *Information Visualization* [7], with the goal of making large and complex content more accessible and usable to the end-users. The user interface was designed in order to optimize comprehension of the structure of the ontology used to model a domain, and to integrate diverse information sources within the same presentation. This objective is achieved using graph representation [8, 9], that maximizes data comprehension and relations analysis. The system uses also concept clouds to summarize the content of a collection, a form of data presentation that has now become extremely familiar to web users. Finally our web system, using the Rich Internet Application paradigm (RIA), does not require any installation and provides a responsive user interface.

2. THE SYSTEM

The tool provides means to explore archives of different video domains, inspecting the relations between the concepts of the ontology and providing direct access to the video instances of these concepts. The interface aims at bringing some graphical elements typical of web 2.0 interfaces, such as the tag cloud, to the exploration of video archives. The user starts selecting concepts from a "tag cloud", than inspects the ontology that describes the video domain, shown as a graph with different types of relations, and inspects the instances of the concepts that are annotated (see Fig. 1*a*).

[1]http://www.adobeyougc.com/

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



(a) (b)

Figure 1: Browsing interface: *a)* a view of part of the ontology (concepts related to "people"); *b)* inspection of some instances of the "face" concept.

Since the presentation of the full graph of the ontology would not be efficient only the concepts that are nearby (in terms of number of relations) the selected one are shown. Users can select a threshold of this distance and also different automatic layout algorithms for the for the spatial positioning of the nodes of the graph, to better understand the concepts relations. The "tag cloud" shows the most frequent concepts in the video collection.

Results of the browsing are shown in the interface in a compact form, to show both ontology structure and video clips. For each video clip that contains an instance of a concept is shown the thumbnail in a mini video player, that let users immediately watch the video (Fig. 1*b*). These thumbnails and videos are obtained from the video streaming server, and for each of them also programme metadata (such as title, broadcaster, etc.) are provided. Users can play the video sequence and, if interested, zoom in each result to watch it in a larger player that provides more details on the video metadata as shown in Fig. 2. Below every element of the ontology we show buttons to access the related contents of the concept from different sources, differing from each other by color. The objective is the enrichment of information with associated material in order to provide the user with a larger number of information sources to improve information completeness and increase user's knowledge. Our system uses the publicly available APIs of YouTube as source of video contents, Flickr for images and pictures and Twitter for social and real-time content. An example of this functionality is shown in Fig. 3. The interface was designed to let users interact with the presentation of the browsing results, allowing drag & drop of the concepts to correct errors of automatic layout positioning, and delving deep into the concepts related to a certain element by double clicking on the represented concept. The graph is animated with smooth transitions between different visualizations [10].

A back-end search engine analyzes the structure of the ontology used to annotate the videos, translating the verbose ontology representation into a more compact and manageable XML representation, that can be easily parsed and processed by the user interface that runs within a browser plugin. The ontology has been created automatically from a flat lexicon, using WordNet to create concept relations (*is_a*, *is_part_of* and *has_part*). The *co-occurrence* relation has been obtained analysing the ground-truth annotations of the TRECVid 2005 training set. The ontology is modelled following the Dynamic Pictorially Enriched Ontology model [11], that includes both concepts and visual concept prototypes. These prototypes represent the different visual modalities in which a concept can manifest; they can be selected by the users to perform query by example. Concepts, concepts relations, video annotations and visual concept prototypes are defined using the standard Web Ontology Language (OWL) so that the ontology can be easily reused and shared. The back-end search engine uses SPARQL, the W3C standard ontology query language, to create simplified views of the ontology structure.

The system frontend is based on the Rich Internet Application paradigm, using a client side Flash virtual machine which can execute instructions on the client computer. RIAs can avoid the usually slow and synchronous loop for user interactions, typical of web based environments that use only the HTML widgets available to standard browsers. This allows to implement a visual querying mechanism that exhibits a look and feel approaching that of a desktop environment, with the fast response that is expected by users. Another advantage of this solution regards the deployment of the application, since installation is not required, because the application is updated only on the server; moreover it can run anywhere, regardless of what operating system is used, provided that a browser with Flash plugin is available. The user interface is written in the Action Script 3.0 programming language, using Adobe Flex. The graphical representation of the graph is made using the open source visual analytics framework Birdeye Ravis [12].

The system backend is currently based on open source tools (i.e. Apache Tomcat and Red 5 video streaming server) or freely available commercial tools (Adobe Media Server has a free developer edition). The RTMP video streaming protocol is used. The search engine that provides access to the ontology structure and concepts instances is developed in Java and supports multiple ontologies and ontology reasoning services. Audio-visual concepts are automatically annotated using the VidiVideo annotation engine [2] or the

automatic annotation tools of the IM3I project[2]. To deal with limitations in the number of streaming connections to the streaming server while maintaining a fast interface response, a caching strategy has been adopted. All the modules of the system are connected using HTTP POST, XML and SOAP web services.

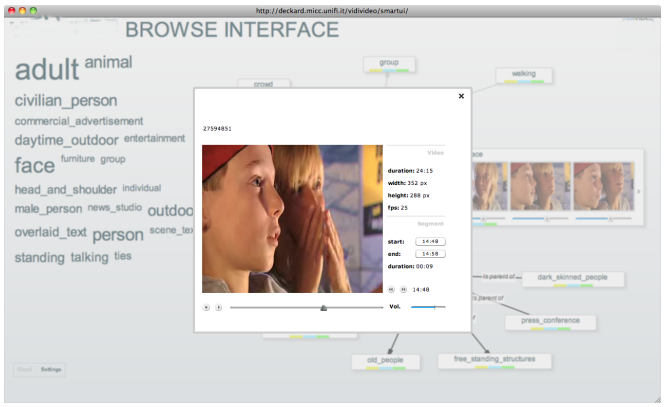


Figure 2: Large streaming video player: the user can expand the mini video players to better inspect each instance of the ontology concepts and analyze the video metadata. The video player shows the position of the concept within the whole video.

The system ranked second in the Adobe YouGC contest, in the Rich Internet Application category.

3. DEMONSTRATION

We demonstrate the browsing functionalities of the system in different video domains: broadcast news and cultural heritage documentaries. We show how to navigate the video collections using the ontology, with its concepts and concepts relations, and with the concept clouds. We demonstrate also how the browsing can be expanded from the video collections to include related material from other sources; the same ontology used for video browsing is used also to access videos on YouTube, images on Flickr and tweets on Twitter.

Acknowledgments.

This work was partially supported by the EU IST VidiVideo project (www.vidivideo.info - contract FP6-045547) and EU IST IM3I project (http://www.im3i.eu/ - contract FP7-222267). The authors thank Nicola Martorana for his help in software development.

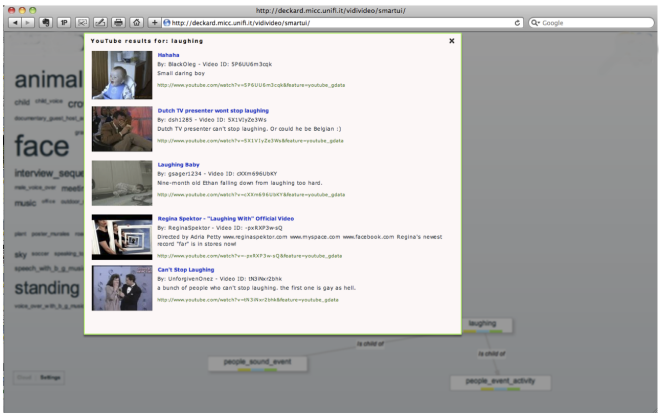
4. REFERENCES

[1] A.F. Smeaton, P. Over, and W. Kraaij. High-level feature detection from video in TRECVid: a 5-year retrospective of achievements. *Multimedia Content Analysis, Theory and Applications*, pages 151–174, 2009.

[2] Cees G. M. Snoek, Koen E. A. van de Sande, Ork de Rooij, Bouke Huurnink, Jasper R. R. Uijlings, Michiel van Liempt, Miguel Bugalho, Isabel Trancoso, Fei Yan, Muhammad A. Tahir, Krystian Mikolajczyk, Josef Kittler, Maarten de Rijke, Jan-Mark Geusebroek, Theo Gevers, Marcel Worring, Dennis C. Koelma, and Arnold W. M. Smeulders. The MediaMill TRECVID 2009 semantic video search engine. In *Proceedings of the 7th TRECVID Workshop*, Gaithersburg, USA, November 2009.

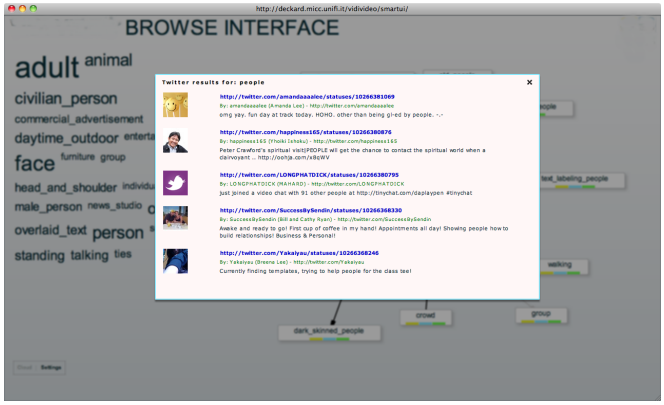
[3] A. Natsev, J.R. Smith, J. Tešić, L. Xie, R. Yan, W. Jiang, and M. Merler. IBM Research TRECVID-2008 video retrieval system. In *Proceedings of the 6th TRECVID Workshop*, 2008.

[2]http://www.im3i.eu



(a)

(b)



(c)

Figure 3: Searching related material in external collections: the user can extend the browsing to other repositories or social sites, seeing the instances of an ontology concept in *a)* YouTube, *b)* Flickr and *c)* Twitter.

[4] Klaus Schoeffmann and Laszlo Boeszoermenyi. Video browsing using interactive navigation summaries. In *CBMI '09: Proceedings of the 2009 Seventh International Workshop on Content-Based Multimedia Indexing*, pages 243–248, Washington, DC, USA, 2009. IEEE Computer Society.

[5] Thierry Urruty, Frank Hopfgartner, David Hannah, Desmond Elliott, and Joemon M. Jose. Supporting aspect-based video browsing: analysis of a user study. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8, New York, NY, USA, 2009. ACM.

[6] W. Bailer, W. Weiss, G. Kienast, G. Thallinger, and W. Haas. A video browsing tool for content management in postproduction. *International Journal of Digital Multimedia Broadcasting*, 2010.

[7] Stuart K. Card, Jock Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, January 1999.

[8] E. Di Giacomo, W. Didimo, L. Grilli, and G. Liotta. Graph visualization techniques for web clustering engines. *Transactions on Visualization and Computer Graphics*, 13(2):294–304, 2007.

[9] Ivan Herman, Guy Melançon, and M. Scott Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000.

[10] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of Visual Languages & Computing*, 6(2):183–210, 1995.

[11] Marco Bertini, Alberto Del Bimbo, Giuseppe Serra, Carlo Torniai, Rita Cucchiara, Costantino Grana, and Roberto Vezzani. Dynamic pictorially enriched ontologies for digital video libraries. *IEEE MultiMedia*, 16(2):42–51, Apr/Jun 2009.

[12] Birdeye information visualization and visual analytics library, http://code.google.com/p/birdeye/.