Sirio: an Ontology-based Web Search Engine for Videos

Thomas Alisi, Marco Bertini, Gianpaolo D'Amico, Alberto Del Bimbo,
Andrea Ferracani, Federico Pernici and Giuseppe Serra
Media Integration and Communication Center, University of Florence, Italy
{alisi, bertini, damico, delbimbo, ferracani, pernici, serra}@dsi.unifi.it
http://www.micc.unifi.it/vim

ABSTRACT

In this technical demonstration we show a web video search engine based on ontologies, the Sirio[1] system, that has been developed within the EU VidiVideo project. The goal of the system is to provide a search engine for videos for both technical and non-technical users. In fact, the system has different interfaces that permit different query modalities: free-text, natural language, graphical composition of concepts using boolean and temporal relations and query by visual example. In addition, the ontology structure is exploited to encode semantic relations between concepts permitting, for example, to expand queries to synonyms and concept specializations.

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, ontologies, web services

1. INTRODUCTION

Video search engines are the product of progress in many technologies: visual and audio analysis, machine learning techniques, as well as visualization and interaction. The current video search engines are based on lexicons of semantic concepts and 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

---

[1]Sirio was the hound of the mythical hunter Orion. It was a dog so swift that no prey could escape it.

of keyframes [2, 3]. These systems do not let users to perform composite queries that can include temporal relations between concepts and do not allow to look for concepts that are not in the lexicon. In addition, desktop applications require installation on the end-user computer and can not be used in a distributed environment.

In this demonstration we present the Sirio system, a web video search engine that allows semantic retrieval by content for different domains (broadcast news, surveillance, cultural heritage documentaries) with query interaction and visualization. The system permits different query modalities (free text, natural language, graphical composition of concepts using boolean and temporal relations and query by visual example) and visualizations, resulting in an advanced tool for retrieval and exploration of video archives for both technical and non-technical users. In addition the use of ontologies permits to exploit semantic relations between concepts through reasoning. 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 Sirio system[2] is composed by three different interfaces: a GUI to build composite queries that may include boolean/temporal operators and visual examples, a natural language interface for simpler queries with boolean/temporal operators, a free-text interface for Google-like searches. In all the interfaces it is possible to extend queries adding synonyms and concept specializations through ontology reasoning and the use of WordNet. Consider, for instance, a query "*Find shots with animal*": the concept specializations expansion through ontology structure permits to retrieve not only the shots annotated with *animal*, but also those annotated with its specializations (*dogs*, *cats*, *etc.*). In particular, WordNet query expansion, using synonyms, is required when using natural language and free-text queries, since it is not possible to force the user to formulate a query selecting terms from a lexicon, as is done using the GUI interface.

The search engine uses an ontology that has been created automatically from a flat lexicon, using WordNet to create concept relations (*is_a*, *is_part_of* and *has_part*). The ontology is modelled following the Dynamic Pictorially Enriched Ontology model [4], 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 exam-

---

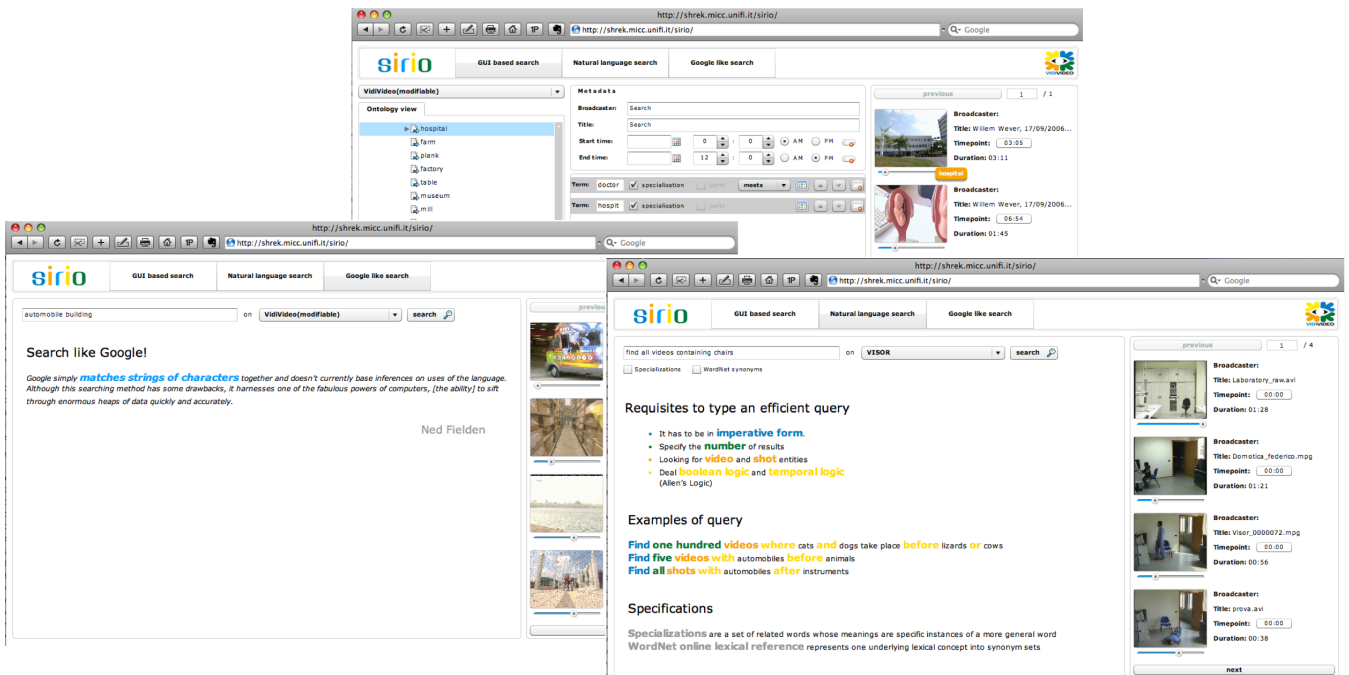[2]http://deckard.micc.unifi.it/sirio/

Figure 1: Search interfaces: natural language search; Google-like search; GUI query builder.

ple. 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 queries created in each interface are translated by the search engine into SPARQL, the W3C standard ontology query language.

The system 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 usual 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. With this solution the application installation is not required, since the system is updated on the server, and run anywhere regardless of what operating system is used.

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 is developed in Java and supports multiple ontologies and ontology reasoning services. Audio-visual concepts are automatically annotated using the VidiVideo annotation engine [2]. The search results are in RSS 2.0 XML format with paging, so that they can be treated as RSS feeds. Results of the query are shown in the interface and for each video clip of the result set is shown the first frame. These frames are obtained from the video streaming server, and are shown within a small video player. Users can then play the video sequence and, if interested, zoom in each result displaying it in a larger player, that provides more details on the video metadata and allows better video browsing. The user interface is written in Adobe Flex and Action Script 3. All the modules of the system are connected using HTTP POST, XML and SOAP web services.

3. DEMONSTRATION

We demonstrate the search modalities of the system in three different video domains: broadcast news, video surveillance and cultural heritage documentaries. We show how each interface is suitable for different users: the GUI interface allows to build composite queries that take into account also metadata, as required by professional archivists, the natural language interface allows to build simple queries with boolean and temporal relations between concepts, the free-text interface provides the popular Google-like search.

4. REFERENCES

[1] A. F. Smeaton, P. Over and W. Kraai. High-Level Feature Detection from Video in TRECVid: a 5-Year Retrospective of Achievements, Multimedia Content Analysis, Theory and Applications, 151–174, 2009, Springer Verlag.

[2] C. G. M. Snoek *et al.* The MediaMill TRECVID 2008 Semantic Video Search Engine, In Proceedings of the 6th TRECVID Workshop, 2008.

[3] A. Natsev, J. R. Smith, J. Tešić, L. Xie, R. Yan, W. Jiang, M. Merler IBM Research TRECVID-2008 Video Retrieval System, In Proceedings of the 6th TRECVID Workshop, 2008.

[4] M. Bertini, R. Cucchiara, A. Del Bimbo, C. Grana, G. Serra, C. Torniai and R. Vezzani. Dynamic Pictorially Enriched Ontologies for Video Digital Libraries. In *IEEE Multimedia*, to appear, 2009.