

# Sirio, Orione and Pan: an Integrated Web System for Ontology-based Video Search and Annotation

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/vim>  
SUBMITTED to ACM MULTIMEDIA 2010 DEMO PROGRAM

## ABSTRACT

In this technical demonstration we show an integrated web system for video search and annotation based on ontologies. The system is composed by three components: the Orione<sup>1</sup> ontology-based search engine, the Sirio<sup>2</sup> search interface, and the Pan<sup>3</sup> web-based video annotation tool. The system is currently being developed within the EU IM3I<sup>4</sup> project. The goal of the system is to provide an integrated environment for video annotation and retrieval of videos, for both technical and non-technical users. In fact, the search engine 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. The annotation tool can be used to create ground-truth annotations to train automatic annotations systems, or to complement the results of automatic annotation, e.g. adding geolocalized information.

## 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

<sup>1</sup>Orione was a heroic and mythical hunter, first cited in Homer's *Odyssey*.

<sup>2</sup>Sirio was the hound of Orione. It was a dog so swift that no prey could escape it.

<sup>3</sup>A Greek divinity, his name means "all".

<sup>4</sup><http://www.im3i.eu>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

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

Copyright 2010 ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

## 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. Automatic video annotation systems are based on large sets of concept classifiers [4], typically based on supervised machine learning techniques such as SVMs; therefore there is need to easily create ground-truth annotations of videos, indicating the objects and events of interest. The current video search engines are based on lexicons of semantic concepts and perform keyword-based queries [3]. 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, 4]. 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.

Within the EU VidiVideo project has been conducted a survey formulated following the guidelines of IEEE Standard 830-1984 Guide to Software Requirements Specifications, to gather user requirements for a video search engine. 40 qualified users from 10 different countries participated; half of the users were part of the scientific and cultural heritage sector and half from the broadcasting and entertainment industry. 73.5% consider a web-based interface "mandatory" and 20.5% "desirable"; this requirement has resulted as fundamental also during the collection of user requirements done within the IM3I project. Users also requested the possibility to formulate complex composite queries and to expand queries based on ontology reasoning. They also requested to have an interface able to show concepts and their relations. Regarding the current annotation practices the vast majority uses controlled lexicons and ontologies. All these requirements have been taken into account in the design of the integrated system.

Regarding manual annotation tools, these have become popular in several mainstream video sharing sites such as YouTube and Viddler, because they extend the popular idea of image tagging applying it to videos. This information is managed in a collaborative web 2.0 style, since annotations can be inserted not only by content owners (video profes-

sionals, archivists, etc.), but also by end users, providing an enhanced knowledge representation of multimedia content.

In this demonstration we present an integrated system composed by *i*) 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; *ii*) a web-based system for manual annotation of videos, developed with the aim of creating, collaboratively, ground truth annotations that can be used for training and evaluating automatic video annotation systems. Annotations can be exported to MPEG-7 and OWL ontologies and are directly integrated with the ontology used by the search engine.

The search engine 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 search engine.* The Sirio system<sup>5</sup> is composed by three different interfaces shown in Fig. 1: a GUI to build composite queries that may include Boolean and temporal operators and visual examples, a natural language interface for simpler queries with boolean/temporal operators, a free-text interface for Google-like searches. The GUI interface allows to inspect a local view of the ontology graph, when building queries, to better understand how one concept is related to the others and thus suggesting to the users possible changes of the composition of the query. 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 vehicles*”: the concept specializations expansion through ontology structure permits to retrieve not only the shots annotated with *vehicle*, but also those annotated with its specializations (*cars*, *trucks*, 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 Orione 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*) as shown, partially, in Fig. 3. The ontology is modelled following the Dynamic Pictorially Enriched Ontology model [1], 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 queries created in each interface are translated by the search

<sup>5</sup><http://deckard.micc.unifi.it/im3i/>

engine into SPARQL, the W3C standard ontology query language.

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. Ontology structure and concept instances serialization has been designed so that inference can be executed simultaneously on multiple ontologies, without slowing the retrieval; this design allows to avoid the need of selecting a specific ontology when creating a query. The engine has also been designed to fit into a service oriented architecture, so that it can be incorporated into the customizable search systems, other than Sirio, that are developed within IM3I project. Audio-visual concepts are automatically annotated using either the VidiVideo annotation engine [4] and the IM3I annotation engine. 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.

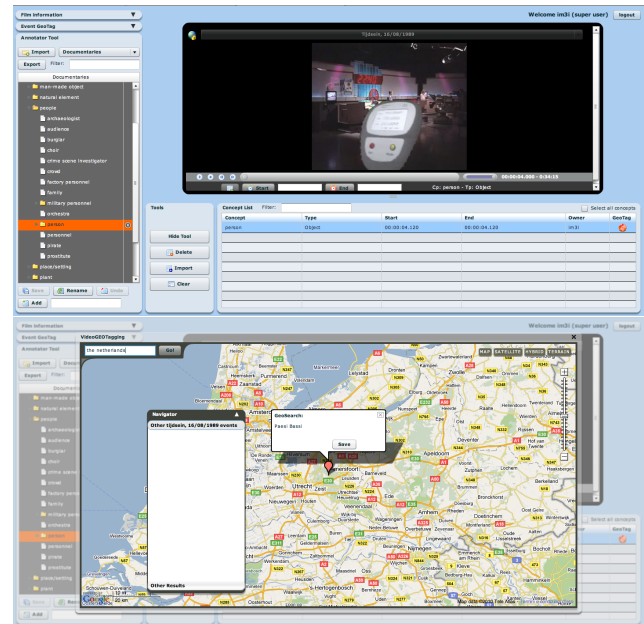


Figure 2: The video annotator tool.

*The annotation tool.* The video annotation system allows to browse videos, select concepts from an ontology structure, insert or modify annotations and download previously saved annotations (Fig. 2). The annotation tool can work with multiple ontologies, to accommodate different video domains. Ontologies can be imported and exported using the MPEG-7 or OWL ontology, or created from scratch by users. The system administrator can set the ontologies as modifiable by end-users. Annotations, managed by the Orione search engine, can be exported to both MPEG-7 and OWL

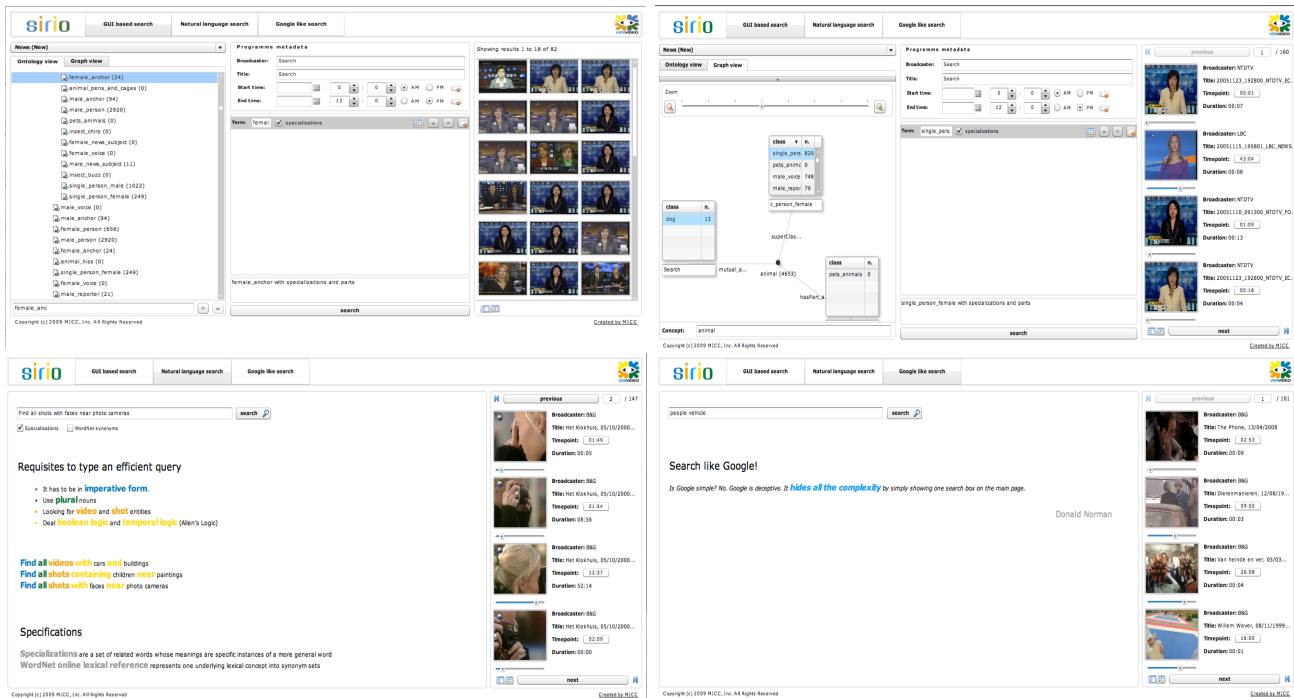


Figure 1: Search interfaces: GUI query builder (with ontology tree and ontology graph); natural language search; Google-like search.

ontology formats.

The system is implemented according to the Rich Internet Application paradigm: all the user interfaces of the system are written in Adobe Flex and Action Script 3, and run in a Flash virtual machine inside a web browser. RIAs offer many advantages to the user experience, because they 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. Both high levels of interaction and collaboration (obtainable by a web application) and robustness in multimedia environments (typical of a desktop application) are guaranteed for the users. With this solution installation is not required, since the application is updated on the server, and run anywhere regardless of what operating system is used. All the modules of the system are connected using HTTP POST, XML and SOAP web services. This allows to integrate the system within a web-services architecture such as that of the IM3I project.

### 3. USABILITY TESTS

The video search engine (Sirio+Orione) has been tested at the end of the VidiVideo project to evaluate the usability of the system, with a set of field trials. A group of 12 people coming from broadcasting, media industry and cultural heritage institutions, in Italy and in the Netherlands, have tested the system on-line (running on the UniFi-MICC servers), performing a set of pre-defined queries and interacting with the system. The methodology used follows the practices defined in the ISO 9241 standard, and gathered:

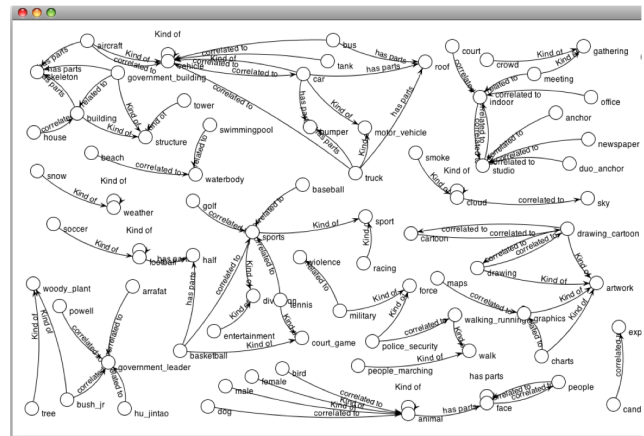


Figure 3: An example of the relations between the DOME ontology concepts.

*i)* observational notes taken during test sessions by monitors, *ii)* verbal feedback noted by test monitor and *iii)* an online survey completed after the tests by all the users. The observational notes and verbal feedbacks of the users have been analyzed to understand the more critical parts of the system and, during the development of the IM3I project, have been used to redesign the functionalities of the Orione search engine and of the Sirio interface. Fig. 4 summarizes the results of the tests. The overall experience is very positive and the system proved to be easy to be used, despite the objective difficulty of interacting with a complex system for which the testers received only a very limited training. The type of interaction that proved to be more suitable for the

majority of the users is the GUI based interface, followed by the Google-like. The NLP interface received the lower marks when users were asked to rate its usefulness for their work, but still it was considered easy to be used. This is due probably to the fact that this type of interface is not commonly used, so that the average user has not enough familiarity with it.

#### 4. 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. We also demonstrate the use of the annotation tool to integrate the results obtained by automatic annotation systems, and the addition of geolocalized annotations. Differently from other web based tools the proposed system allows to annotate the presence of objects and events in frame accurate time intervals, instead of annotating a few keyframes or having to rely on offline video segmentation. The possibility for the annotators to extend the proposed ontologies allow expert users to better represent the video domains, as the videos are inspected. The import and export capabilities using MPEG-7 standard format allow interoperability of the system.

#### Acknowledgments.

This work was partially supported by the EU IST Vidi-Video project ([www.vidivideo.info](http://www.vidivideo.info) - contract FP6-045547) and EU IST IM3I project (<http://www.im3i.eu/> - contract FP7-222267). The authors thank Thomas M. Alisi and Nicola Martorana for their help in software development.

#### 5. REFERENCES

- [1] M. Bertini, A. Del Bimbo, G. Serra, C. Torniai, R. Cucchiara, C. Grana, and R. Vezzani. Dynamic pictorially enriched ontologies for digital video libraries. *IEEE MultiMedia*, 16(2):42–51, Apr/Jun 2009.
- [2] A. Natsev, J. 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.
- [3] A. 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.
- [4] C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. R. R. Uijlings, M. van Liempt, M. Bugalho, I. Trancoso, F. Yan, M. A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J.-M. Geusebroek, T. Gevers, M. Worring, D. C. Koelma, and A. W. M. Smeulders. The MediaMill TRECVID 2009 semantic video search engine. In *Proceedings of the 7th TRECVID Workshop*, Gaithersburg, USA, November 2009.

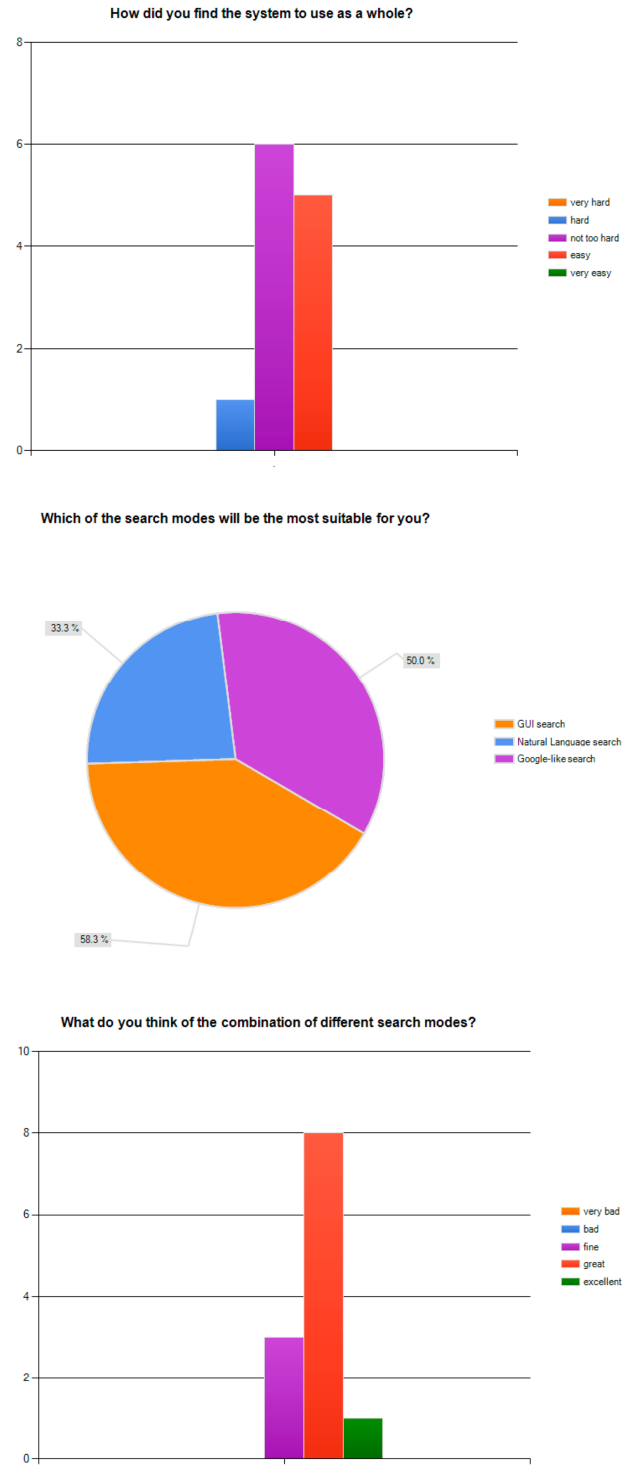


Figure 4: Overview of usability tests: overall usability of the system, preferred search modalities, usability of the combination of search modalities.