

# Ontology- Reasoning for Web based Video Retrieval

G.Ayyapan

Associate Professor, Dept. of IT, Bharath University, Chennai, TN, India

**ABSTRACT:** In this paper presents two interactive systems for video search and browsing; one is a web application based on the Rich Internet Application paradigm, designed to obtain the levels of responsiveness and interactivity typical of a desk-top application, while the other exploits multi-touch devices to implement a multiuser collaborative application. Both systems use the same ontology-based video search engine, that is capable of expanding user queries through ontology reasoning, and let users to search for specific video segments that contain a semantic concept or to browse the content of video collections, when it's too difficult to express a specific query.

**KEYWORDS:** Ontology reasoning, multi-touch, multi user application

## 1. INTRODUCTION

Video search engines are the result of advancements in many different research areas: audio-visual feature extraction and description, machine learning techniques, as well as visualization, interaction and user interface design. The current video search engines are based on lexicons of semantic concepts and perform keyword- based queries.

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, while the web- based tools allow only limited user interaction

## II. THE SYSTEM

In this paper present an integrated system composed by *i*) a video search engine that allows semantic retrieval by content for different domains (possibly modelled with different ontologies) with query expansion and ontology reasoning; *ii*) web-based interfaces for interactive query com-position, archive browsing, annotation and visualization; *iii*) a multi- touch tangible interface featuring a collaborative natural interaction application.

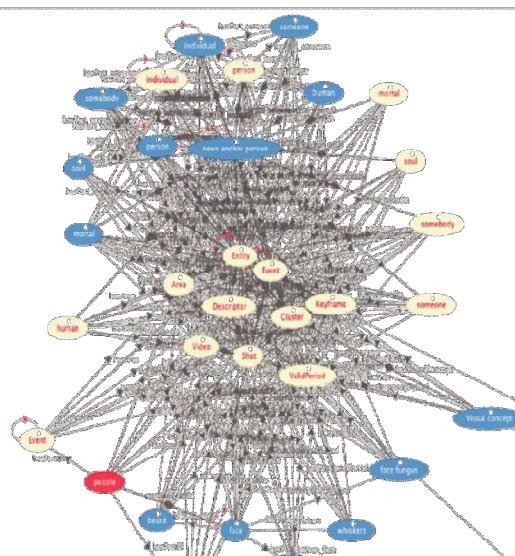


Figure 1. Automatically created ontology structure: view of the relations between some concepts related to “people”.



in Fig. 1. The ontology is modeled following the Dynamic Pictorially Enriched Ontology model, 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, using MPEG-7 descriptors (e.g. Color Layout and Edge Histogram) or other domain specific visual descriptors. 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.

As an example consider the query “*Find shots with vehicles*”: the concept specializations expansion through inference over the ontology structure permits to retrieve the shots annotated with *vehicle*, and also those annotated with the concept’s specializations (e.g. *trucks*, *cars*, etc.). In particular, WordNet query expansion, using synonyms, is required when using free-text queries, since it is not desirable to force the user to formulate a query selecting only the terms from a predefined lexicon.

**The web-based user interface** The web-based Sirio search system<sup>3</sup>, based on the Rich Internet Application paradigm (RIA), does not require any software installation, is extremely responsive

The GUI interface allows also to inspect and use 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.

For each result of the query it is shown the first frame of the video clip. 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,

This interface is based on some graphical elements typical of web 2.0 interfaces, such as the tag cloud. The user starts selecting concepts from a “tag cloud”, then navigates the ontology that describes the video domain, shown as a graph with different types of relations, and inspects the video clips that contain the instances of the concepts used as annotations. Users can select a concept from the ontology graph to build a query in the advanced search interface at any moment.

**The tangible user interface** MediaPick is a system that allows semantic search and organization of multimedia contents via multi-touch interaction. It has an advanced user centered interaction design, developed following specific usability principles for search activities, which allows users collaboration on a tabletop about specific topics that can be explored, thanks to the use of ontologies, from general to specific concepts. Users can browse the ontology structure in order to select concepts and start the video retrieval process. Afterwards they can inspect the results returned by the Orione video search engine and organize them according to their specific purposes. Some interviews to potential end-users have been conducted with the archivists of RAI, the Italian public broadcaster, in order to study their workflow and collect suggestions and feedbacks; at present RAI journalists and archivists can search the corporate digital libraries through a web-based system.

This provides a simple keyword based search on textual descriptions of the archived videos. Sometimes these descriptions are not very detailed or very relevant to the video content, thus making the document difficult to find. The cognitive load required for an effective use of the system often makes the journalists delegate their search activities to the archivists that could be not familiar with the specific topic and therefore could hardly choose the right search keyword. The goal of the MediaPick design is to provide to the broadcast editorial staff an intuitive and collaborative interface to search, visualize and organize video results archived in huge digital libraries with a natural interaction approach.

The user interface adopts some common visualization principles derived from the discipline of Information Visualization and is equipped with a set of interaction functionalities designed to improve the usability of the system for the end-users. The GUI consists of a concepts view, to select one or more keywords from an ontology structure and use them to query the digital library, and a results view, which shows the videos returned from the database, so that the user can navigate and organize the extracted contents

The concepts view consists of two different interactive elements: the ontology graph, to explore the concepts and their relations, and the controller module, to save the selected concepts and switch to the results view. The user chooses the concepts as query from the ontology graph. Each node of the graph consists of a concept and a set of relations. The concept can be selected and then saved into the controller, while a relation can be triggered to list the related concepts, which can be expanded and selected again; then the cycle repeats. The related concepts are only shown when a precise relation is triggered, in order to minimize the number of visual elements present at the same

time in the interface.

Each video element has three different states: idle, playback and in-formation. In the idle state the video is represented with a keyframe and a label visualizing the concept used for the query. During the playback state the video starts playing from the frame in which the selected concept was annotated. A longer touch of the video element activates the information state, that shows a panel with some metadata (related concepts, quality, duration, etc. over the video

At the bottom of the results list there are all the concepts related to the video results. By selecting one or more of these concepts, the video clips returned are filtered in order to improve the information retrieval process. The user can select any video element from the results list and drag it out-side. This action can be repeated for other videos, returned by the same or other queries. Videos placed out of the list can be moved along the screen, resized or played. A group of videos can be created by collecting two or more video elements in order to define a subset of results. Each group can be manipulated as a single element through a contextual menu: it can be expanded to show a list of its elements or released in order to ungroup the videos.

### III. ARCHITECTURE

The system backend and the search engine are 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). Videos are streamed using the RTMP video streaming protocol. The search engine is developed in Java and supports multiple ontologies and ontology reasoning services. Ontology structure and concept instances serialization have 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 with the Google-like interface. 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 and Medi-aPick, that are developed within IM3I and euTV projects. Audio-visual concepts are automatically annotated using either IM3I and euTV automatic annotation engines. The search results are produced in RSS 2.0 XML format, with paging, so that they can be used as feeds by any RSS reader tool and it is possible to subscribe to a specific search. Both the web-based interface and the multi-touch interface have been developed in Flex+Flash, according to the Rich Inter-net Application paradigm.

**Multitouch** MediaPick exploits a multi-touch technology chosen among various approaches experimented in our lab since 2004 [2]. Our solution uses an infrared LED array as an overlay built on top of an LCD standard screen capable of a full-HD resolution). The multi-touch overlay can detect fingers and objects on its surface and sends information about touches using the TUIO protocol at a rate of 50 packets per second. MediaPick architecture is composed by an input manager layer that communicates through the server socket with the gesture framework and core logic. The latter is responsible of the connection to the web services and media server, as well as the rendering of the GUI elements on the screen .

The input management module is driven by the TUIO dispatcher: this component is in charge of receiving and dispatching the TUIO messages sent by the multi-touch overlay to the gesture framework through the server socket. This module is able to manage the events sent by the input manager, translating them into commands for the gesture framework and core logic.

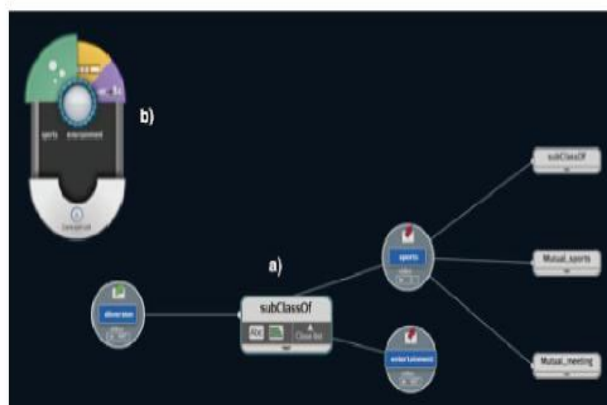


Fig:MediaPick System: a)Ontology Graph b)Controller Module

The logic behind the multi-touch interfaces needs a dictionary of gestures which users are allowed to perform. It is



possible to see each digital object on the surface like an active touchable area; for each active area a set of gestures is defined for the interaction, thus it is useful to link each touch with the active area in which it is enclosed. For this reason each active area has its own set of touches and allows the gesture recognition through the interpretation of their associated behavior. All the user interface actions mentioned above are triggered by natural gestures shown in the Tab. 1.

Gesture	Actions
	- Select concept
	- Trigger the controller module

### REFERENCES

[1] R. S. Amant and C. G. Healey. Usability guidelines for in-teractive search in direct manipulation systems. In *Proc. of the International Joint Conference on Artificial Intelligence*, volume 2, pages 1179–1184, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[2] S. Baraldi, A. Bimbo, and L. Landucci. Natural interaction on tabletops. *Multimedia Tools and Applications (MTAP)*, 38:385–405, July 2008.

[3] M. Bertini, A. Del Bimbo, G. Serra, C. Torniai, R. Cuc-chiara, C. Grana, and R. Vezzani. Dynamic pictorially en-riched ontologies for digital video libraries. *IEEE MultiMe-dia*, 16(2):42–51, Apr/Jun 2009.

Single tap	Long pressure touch
switch	
- Select video group contextual menu options	
- Play video element	
- Trigger video information	
state	
- Show video group contextual menu	
Drag	- Move video element
- Resize video	
Two-fingers drag	
- Group videos	

### IV.CONCLUSION

In this paper presents two semantic video search systems based on web and multi-touch multi-user interfaces. Future work will deal with further development of the interfaces, especially considering the new HTML5 technologies, extensive testing of the tangible user interface and thorough comparison of the two systems.