



Ontology-based Video Retrieval in a Semantic- based Learning Environment

Antonella Carbonaro

Department of Computer Science

University of Bologna

carbonar@csr.unibo.it

Abstract

The paper presents an ontological approach for enabling semantic-aware information retrieval and browsing framework facilitating the user access to its preferred contents. Through the ontologies the system will express key entities and relationships describing learning material in a formal machine-processable representation. An ontology-based knowledge representation could be used for content analysis and concept recognition, for reasoning processes and for enabling user-friendly and intelligent multimedia content search and retrieval.

1 Introduction

Vast amounts of multimedia information including video are becoming ubiquitous as a result of advances in multimedia computing technologies and high-speed networks. Video is rapidly becoming the most popular media, due to its high information and entertainment power.

The main challenge is to index information retained in video in order to make them searchable and thus (re-) usable. This requires the multimedia content to be annotated, which can either be done manually or automatically. In the first case, the process is extremely work- and thus cost-intensive; in the second case, it is necessary to apply content-analysis algorithms that automatically extract descriptions from the multimedia data. The aim is to create a concise description of the multimedia content features, that is, its metadata. Metadata descriptions may vary considerably in terms of comprehensiveness, granularity, abstraction level, etc. depending on the application domain, the tools used and the effort made for creating the descriptions.

Multimedia annotation systems need standard output, which must be compliant with other tools for browsing or indexing. MPEG-7 [ISO/IEC] standard was defined for this purpose. It represents an elaborate standard in which a number of fields, ranging from low level encoding scheme descriptors to high level content descriptors, are merged to be useful for describing a video or part of it.

In text-based applications, it is often sufficient to annotate only the generic properties of the document (such as title and creator) and to perform keyword search, based on full-text information retrieval approaches. For non-textual resources, however, full text search is only an option if there are sufficient associated textual information (for example, Google's image search based on surrounding text in HTML or video retrieval based on subtitles, closed captions or audio transcripts). In many other cases content descriptions are inevitable. Since content descriptions are not often about the entire document (e.g. a specific shot in a film or a specific region in a picture), it is necessary to implement shot detection and keyframe extraction procedures to deal with video files.

An important step towards efficient manipulation and retrieval of visual media is semantic information representation [Calic et al. 2005], [Bloehdorn et al., 2004]. In the digital library community a flat list of attribute/value pairs is often assumed to be available. In the Semantic Web community, annotations are often assumed to be an instance of an ontology. Through the ontologies the system will express key entities and relationships describing video in a formal machine-processable representation. An ontology-based knowledge representation could be used for content analysis and object recognition, for reasoning processes and for enabling user-friendly and intelligent multimedia

content search and retrieval.

In [Tjondronegoro and Spink, 2008] they report findings from a study examining the state of multimedia search functionality on major general and specialized Web search engines. They investigated 102 Web search engines to examine the degree of multimedia searching functionality offered by major Web search engines and to compare the functionalities of each Web search engine which is significant for the development of more effective multimedia IR systems. Their findings show that despite the growing level of interest in multimedia Web search, most major Web search engines currently offer limited multimedia search functionality. Keywords are still used as the only mean of multimedia retrieval. For search formulation, ontology-based classification can help users in redesigning their query if it is too specific. For example, instead of looking for ‘‘aloe vera’’, users can be suggested to search on ‘‘green plants’’. Moreover, a unified indexing on keywords and semantic summaries will enable search engines to support users in finding related topics.

The aim of this paper is to present our video retrieval and browsing framework based on both collaborative and semantic approaches. The collaborative approach is exploited both in retrieving task (to cover recommendation and resource sharing tasks) and in semantic coverage of the involved domain. The semantic approach is exploited introducing an ontology space covering domain knowledge and resource models based on word sense representation. Also the ontology level exploits system collaborative aspect. We show how the semantic technologies can enhance the traditional e-learning keyword approaches facilitating the user retrieval and browsing by adding semantic information in the resource and user profiles.

Applications that could benefit from semantic video representation are manifold, from education and training to medical, from entertainment to system analysis and evaluation, etc. For example, home entertainment systems (management of personal multimedia collections, including manipulation of content, home video editing, searching, etc.) need a mechanism to interpret human’s queries, and retrieve the closest match. However, this search outcome may result very unsatisfactory due to the blurred link between the low-level measured features and the human semantic queries. This discrepancy between the way video data is coded digitally and the way it is experienced by a human user is called the semantic gap, [Smeulders et al., 2000]. Differently, in education, semantic annotations of video recording of lectures distributed over the Internet can be used to augment the material by providing explanations, references or examples, that can be used for efficiently access, find and review material in a student personal manner [Carbonaro and Ferrini, 2005], [Carbonaro, 2005]. Moreover, in television, semantic annotation of programmes, for example news, could produce electronic programme guides, which would allow the user to

view details of forthcoming programmes in terms of entities referred to in particular broadcasts [Dowman et al., 2005].

2 Personalized Video Retrieval Framework

Traditional approaches to personalization include both content-based and user-based techniques [Dai and Mobasher, 2004]. If, on one hand, a content-based approach allows to define and maintain an accurate user profile (for example, the user may provides the system with a list of keywords reflecting hir/her initial interests and the profiles could be stored in form of weighted keyword vectors and updated on the basis of explicit relevance feedback), which is particularly valuable whenever a user encounters new content, on the other hand it has the limitation of concerning only the significant features describing the content of an item. Differently, in a user-based approach, resources are processed according to the rating of other users of the system with similar interests. Since there is no analysis of the item content, these information management techniques can deal with any kind of item, being not just limited to textual content. In such a way, users can receive items with content that is different from that one received in the past. On the other hand, since a user-based technique works well if several users evaluate each one of them, new items cannot be handled until some users have taken the time to evaluate them and new users cannot receive references until the system has acquired some information about the new user in order to make personalized predictions. These limitations often refer to as the sparsity and start-up problems [Melville et al., 2002]. By adopting a hybrid approach, a personalization system is able to effectively filter relevant resources from a wide heterogeneous environment like the Web, taking advantage of common interests of the users and also maintaining the benefits provided by content analysis.

A hybrid approach maintains another drawback: the difficulty to capture semantic knowledge of the application domain, i.e. concepts, relationships among different concepts, inherent properties associated with the concepts, axioms or other rules, etc.

Semantic-based approach to retrieving relevant material can be useful to address issues like trying to determine the type or the quality of the information suggested from a personalized learning environment. In this context, standard keyword search has a very limited effectiveness. For example, it cannot filter for the type of information (tutorial, applet or demo, review questions, etc.), the level of information (aimed to secondary school students, graduate students, etc.), the prerequisites for understanding information, or the quality of information. Some examples of semantic-based e-learning

systems can be found in Mendes and Sacks [Mendes and Sacks, 2004] and in Lytras and Naeve [Lytras and Naeve, 2005].

The aim of this paper is to present our personalized learning retrieval framework based on both collaborative and semantic approaches. The collaborative approach is exploited both in retrieving task (to cover recommendation and resource sharing tasks) and in semantic coverage of the involved domain. The semantic approach is exploited introducing an ontology space covering domain knowledge and resource models based on word sense representation. Also the ontology level exploits system collaborative aspect.

The Scout-V module assists authors in annotating video sequences. Each shot belonging to the video sequence can be annotated on the base of underlying ontologies. These descriptions are labelled for each shot and are stored as MPEG-7 descriptions in the output XML file. Scout-V can also save, open, and retrieve MPEG-7 files in order to display the annotations for corresponding video sequences. The Scout-V main page shows all the videos that should be elaborated performing shot detection, editing or removing. Given the segmentation of video content into video shots, the second step is to define the semantic lexicon to label the shots. A video shot can fundamentally be described by using five basic classes: agents, objects, places, times and events. These five types of lexicon define the initial vocabulary for our video content; they correspond to the SemanticBase MPEG-7 tags. We have also defined attributes to describe class characteristics. Each attribute corresponds to a specified MPEG-7 tag used in storing phase. By using the defined vocabulary for static agents, key objects, places, times and events, the lexicon is imported into Scout-V for describing and labelling each video shot. The shots are labelled for their content with respect to the selected lexicon. Note that the lexicon definitions are database and application specific, and can be easily modified and imported into the annotation tool.

Scout-V annotation tool is divided into three graphical sections: the Scene Matching frame in which are shown the algorithms that can be used to obtain video annotation recommendations (Block Truncation Coding, edge histogram, colour histogram), the Ontology Visualization frame, providing interactivity to assist authors of the annotation tool and the Video Presentation frame with the key frame image display and the frame characteristics. The Ontology Editor module allows to modify the ontology tree creating and populating all the necessary classes and instances. The aim of the instance creation phase is to effectively represent the domain knowledge, so as to achieve a better precision in the annotation task. Annotations are then stored and used by recommendation procedure to help users finding similar frames which have been annotated also by other users.

Once the scene as a whole has been annotated, the system produces a MPEG-7 file. The system comprises automatic shot detection and scene matching modules to obtain video annotation recommendations in a collaborative framework. By the shot detection method the video can be automatically segmented into shots. A shot is a contiguous sequence of video frames which have been recorded from a single camera operation [Grana et al., 2005]. The method is based on the detection of shot transitions (hard cuts, dissolves, and fades). One or more keyframes are extracted from the obtained shots set in dependence on the visual content dynamics. Several experiments tested the effectiveness of both the shot detection module and the frame matching module on the annotation process. More detailed description can be found in [Carbonaro and Ferrini, 2007].

The video retrieving framework is shown in Fig. 1. We introduce an example ontology from the travel domain; it could be published on fixed URI's as OWL files. The ontology would define concepts such as ActivityProvider to link an Activity with a ContactAddress. There could be a set of subtypes of activities such as BungeeJumping or IceClimbing, and these could be categorized into types like AdventureActivity. Based on the rich expressiveness of OWL, it is furthermore possible to define classes by their logical characteristics. For example, a class BackpackersDestination could be defined as a destination that offers budget accommodation and some adventure activities. These defined classes allow reasoners to automatically classify existing domain objects into matching categories.

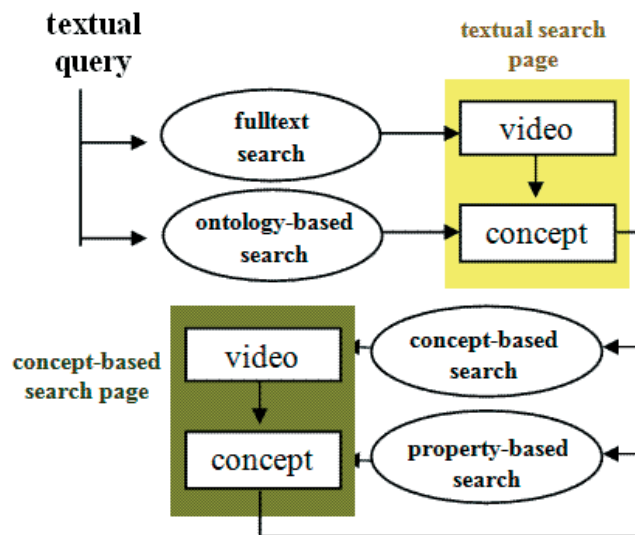


Fig. 1 Video retrieval functioning scheme

Figure 2 shows the implemented ontology. The retrieving process starts with a keyword-based search. We use database textual fields to retrieve video that correspond to user query; these fields are concerning to both general video features like title, keywords and description and shot elements (see Figure 3). On the left-hand side of the screen we show the set of concepts extracted from the ontology that the user can use to perform a semantic-based video search. These concepts are ontology instances related both to performed query and to retrieved video. We can consider the retrieving function as an interactive transformation of the starting query. The user can iteratively performs concept-based search choosing a concept from the set of relevant extracted from the ontology using ontology properties that link individuals. Since visualized concepts are OWL individuals they are showed using their name in the ontology. Figure 4 shows the ontology-based search results using the keyword “snow”: The related concepts are relative to winter sports, to those corresponding to the unique retrieved video and to the concepts child of returned class in the ontology.

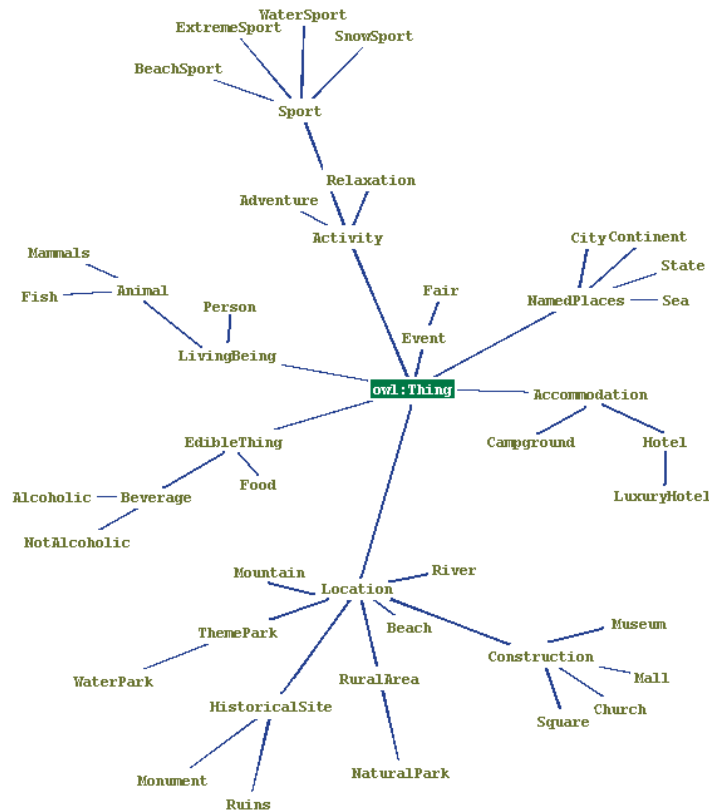



Fig. 2 Developed ontology

Risultati della ricerca per la stringa "water"

Concetti associati alla ricerca:


- Colorado
- surf
- ColoradoRiver
- Italy
- WildWadi
- OceanBeach
- PacificOcean
- DubaiCity
- SanDiego
- rafting



✓ ▶

Wild Wadi Water Park

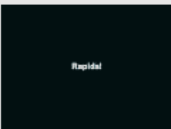
Keywords: dubai water surf
Descrizione: Surf inside a water park in Dubai



✓ ▶

Canyoning

Keywords: canyon water sport
Descrizione: Me and my friends, canyoning in central italy



✓ ▶

Rafting on Colorado River

Keywords: rafting sports water grandcanyon
Descrizione: Colorado river descent in our summer holiday. We travel through Grand Canyon!

Fig. 3 Simple query results

Risultati della ricerca per la stringa "snow"

Concetti correlati alla ricerca:

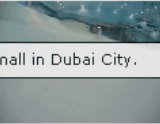
- snowboard
- ski

Concetti associati ai video:

- MallOfTheEmirates
- DubaiCity
- ski

Classi correlate alla ricerca:

- SnowSport
- ski
- snowboard



✓ ▶

Skiing in Dubai

Keywords: ski snow mall
Descrizione: Skiing inside a mall in Dubai, with real snow!

Fig. 4 Ontology-based search mode

3 Considerations

In this paper we have presented a methodology for semantic video content retrieving. The system comprises automatic shot detection and scene matching modules to obtain video annotation recommendations in a collaborative framework. Several experiments tested the effectiveness of both the shot detection module and the frame matching module on the annotation process. The ontology-based retrieving framework offers a valuable multimedia search functionality. Future works will include the study and the implementation of an ontology layer able to maintain several existing ontologies the user knows. This approach could allow to compare the knowledge of any user without having a single consensual ontology.

BIBLIOGRAPHY

- Bloehdorn S., Petridis K., Simou N., Tzouvaras V., Avrithis Y., Handschuh S., Kompatsiaris Y., Staab S., Strintzis M. G. (2004), *Knowledge Representation for Semantic Multimedia Content Analysis and Reasoning*, Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology.
- Calic J., Campbell N., Dasiopoulou S., Kompatsiaris Y. (2005), *A Survey on Multimodal Video Representation for Semantic Retrieval*, the Third International Conference on Computer as a tool (Eurocon 2005), IEEE.
- Carbonaro A., Ferrini R. (2007), *Ontology-based Video Annotation in Multimedia Entertainment*, Proc. 3rd IEEE International Workshop on Networking Issues in Multimedia Entertainment (NIME'07) - 2007 4th IEEE Communications and Networking Conference (CCNC 2007), Las Vegas (USA), IEEE Communications Society.
- Carbonaro A. (2005), *Defining personalized learning views of relevant learning objects in a collaborative bookmark management system*, Web-Based Intelligent e-Learning Systems: Technologies and Applications" Idea Group Inc.
- Carbonaro A., Ferrini R. (2005), *Considering semantic abilities to improve a Web-Based Distance Learning System*, ACM International Workshop on Combining Intelligent and Adaptive Hypermedia Methods/Techniques in Web-based Education Systems.
- Dai H., Mobasher B. (2004) *Integrating semantic knowledge with web usage mining for personalization*, *Web Mining: Applications and Techniques*, A. Scime (Ed.), Hershey: Idea Group Publishing, 276-306.
- Dowman M., Tablan V., Cunningham H., Popov B. (2005), *Web-assisted annotation, semantic indexing and search of television and radio news*, Proceedings of the 14th international conference on World Wide Web, 225 – 234.

- Grana C., Tardini G., Cucchiara R. (2005), *MPEG-7 Compliant Shot Detection in Sport Videos*, Proceedings of IEEE International Symposium on Multimedia (ISM2005), Irvine, California, USA, 395-402.
- ISO/IEC (2002), *Overview of the MPEG-7 Standard (version 8)*, ISO/IEC JTC1/SC29/WG11/N4980, Klagenfurt, July 2002.
- Lytras M. D., Naeve A. Eds. (2005), *Intelligent Learning Infrastructure for Knowledge Intensive Organizations*, Information Science Publishing, London.
- Mendes M. E. S., Sacks L.(2004), *Dynamic Knowledge Representation for e-Learning Applications*, In: M. Nikravesh, L.A. Zadeh, B. Azvin and R. Yager (Eds.), *Enhancing the Power of the Internet - Studies in Fuzziness and Soft Computing*, Springer, vol. 139, 255-278.
- Smeulders A. W. M., Worring M., Santini S., Gupta A., Jain R. (2000), *Content based image retrievals at the end of the early years*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, issue 12, 1349-1380.
- Tjondronegoro D., Spink A., (2008), *Web search engine multimedia functionality*, InformationProcessing and Management 44 (2008) 340–357.