*Collaboration and the Knowledge Economy: Issues, Applications, Case Studies*
*Paul Cunningham and Miriam Cunningham (Eds)*
*IOS Press, 2008 Amsterdam*
*ISBN 978–1–58603–924-0*

# Ontological Multimedia Information Management System

Hilal TARAKCI[1], Nihan Kesim CICEKLI[2]
[1]*MilSOFT Software Technologies, METU, Technopolis, Ankara, 06531, Turkey*
*Email: hilaltarakci@yahoo.com*
[2]*Department of Computer Engineering, METU, Ankara, 06531, Turkey*
*Tel: +90 312 2105582, Fax: + 90 312 2105544, Email: nihan@ceng.metu.edu.tr*

**Abstract:** In order to manage the content of multimedia data, the content must be annotated. Although any user-defined annotation is acceptable, it is better if many systems use the same annotation format. MPEG-7 is a widely accepted standard for multimedia content annotation. In MPEG-7, semantically identical metadata can be represented in multiple ways due to lack of precise semantics in its XML-based syntax. This unfortunately prevents metadata interoperability. To overcome this, MPEG-7 standard is translated into an ontology. In our work, we use an MPEG-7 ontology on top and wrap the given user-defined ontologies with MPEG-7 ontology, thus building MPEG-7 based ontologies automatically. Our proposed system is an ontological multimedia information management framework due to its modular architecture, ease of integrating with user-defined ontologies naturally and automatic harmonization of MPEG-7 ontology and domain-specific ontologies.

**Keywords:** semantic querying of video content, multimedia content annotation, mpeg-7, mpeg-7 ontology, mpeg-7 based ontology

## 1. Introduction

Nowadays, using computer technology is the most common way to socialize. People share their special or common multimedia data on youtube, facebook and similar web sites. Besides, everybody has a personal digital library of photos, videos etc. and has experienced the annoyance of looking for a specific video scene inside a huge amount of data without the help of an intelligent multimedia data management system.

Many projects have been developed for the purpose of managing multimedia data with respect to its content. Among these, we can list AceMedia [1], K-Space[2], BilVideo [3], Informedia [4], VideoQ[5]. In this paper, we are concerned with semantic annotation of multimedia data, especially videos. We will summarize the state of the art and then present the difference of our proposed system. In order to manage huge amount of multimedia data, the content must be annotated. The way in which the content is annotated depends on the annotation environment of the multimedia information management system. Although any user-defined annotation is acceptable, it is better if many systems use the same annotation format. In other words, standardizing the metadata of the content is much better than each system using its own defined annotation format. The widely accepted content annotation standard is Multimedia Content Description Standard known as MPEG-7[6]. MPEG-7 is an ISO/IEC standard and developed by MPEG (Moving Picture Experts Group). Furthermore, MPEG-7 uses XML as the language of choice for the textual representation of content description. In MPEG-7, semantically identical metadata can be represented in many different ways due to lack of precise semantics in XML-based syntax. This unfortunately prevents metadata interoperability. In order to overcome the interoperability issues, efforts have been spent to translate MPEG-7 standard into an ontology and to enable its integration

with other ontologies through appropriate frameworks, thus enhancing interoperability. There exist four OWL/RDF proposals of MPEG-7. These are Jane Hunter's MPEG-7/ABC ontology [7], Tsinaraki's MPEG-7/Tsinaraki ontology [8], Garcia and Celma's Rhizomik model[9] and Arndt's COMM[10]. In this paper, basics of MPEG-7 ontologies are mentioned and our choice of MPEG-7 ontology is presented. In our work, we use an MPEG-7 ontology on top and wrap the given user-defined ontologies with MPEG-7 ontology, thus building MPEG-7 based ontologies automatically. Prior to wrapping the user-defined ontology, we let the user to select concepts that are going to be used in annotation. On the annotation and querying interface, we let the user to annotate or query the wrapped concepts with their attributes. Our proposed system is an ontological multimedia information management framework due to its modular architecture, ease of integrating with user-defined ontologies naturally and automatic harmonization of MPEG-7 ontology and domain-specific ontologies, which does not include automated or semi-automated annotation but enables integration of any such module. In the paper, these concepts are explained and some user interface screenshots are presented to give a flavour of the usage of the system. The ontological multimedia information framework can be easily used in specific domains naturally. Moreover, the system can easily be modified according to domain-specific requirements due to its modular architecture.

Since our proposed system is based on MPEG-7 ontology, when a mapping between MPEG-7 and another multimedia content description is available, the system can be easily expanded to welcome new annotations. A domain-specific integration of the framework can be easily produced.

The rest of the paper is organized as follows. A brief summary of related projects is given in Section 2, emphasizing the difference between our framework and the existing work. Section 3 reviews MPEG-7 standard and MPEG-7 based ontologies. Our ontological video model is presented in Section 4. Section 5 summarizes the implementation of the proposed multimedia information management system. Section 6 concludes the paper with some comments about future work.

## 2. Related Work

People want to search and find digital content according to its semantics. In order to achieve this, there should be knowledge about the content. This knowledge comes from the metadata of the content. Metadata can be on different levels of abstraction. On the lowest syntactical level there are basic visual features of content like shape, size, texture, color and movement of a camera or an object in a scene. On a higher level these physical features are interpreted to derive semantic information. This includes taxonomies (e.g. genre), organizational information (e.g. scenes for supporting indexing) and basic descriptions (e.g. identification of objects involved in a scene, roles, etc.). Another type of semantic information is the description of the content as annotations in natural language. As the abstraction level of the metadata increases, the management and querying power of the system increases.

Many projects exist that have been developed on the management of multimedia data with respect to its content [1,2,3,4,5].

AceMedia[1] aims to automate annotation process at all levels and ease content creation, search, access, consumption and re-use.

K-Space [2] focuses on narrowing the semantic gap between content descriptors, which may be computed automatically, and the richness and subjectivity of semantics in high-level human interpretations of audiovisual media.

BilVideo provides an integrated support for queries on spatio-temporal, semantic and low-level features (color, shape, and texture) on video data [3]. BilVideo is an application-independent system. In other words, the system can easily be tailored for the specific

requirements of such applications with the help of the definition of external predicates supported by the system's query language without much effort.

The aim of the Informedia project is to achieve machine understanding of video and film media, in terms of search, retrieval, visualization and summarization [4]. Informedia provides full-content search and retrieval of TV and radio news and documentary broadcasts.

VideoQ[5] is a Web based video search system, where the user queries the system using animated sketches that is defined as a sketch where the user can assign motion to any part of the scene. VideoQ adopts client-server architecture. The client is a java applet that is loaded to a web browser. The user sketches a query scene as a collection of objects with different attributes including motion, spatio-temporal ordering, shape, color and texture.

Another research activity focuses on creating a framework for the automatic annotation of videos in soccer domain and the semantic retrieval of soccer videos based on high-level concepts [11]. The Multimedia Ontologies Annotator is the framework that allows users to import basic ontology schemas, generate the multimedia ontology and annotate videos according to the given ontology. Besides, the system performs complex queries in order to retrieve videos containing specific visual concepts and high-level linguistic concepts.

In our project, we do not focus on automating or semi-automating annotation process itself as most of the above projects do. We focus on providing the user with an ontological video management system framework enabling him to make use of his ontologies in video annotation and querying without any extra knowledge. Amongst the mentioned projects, our proposed system is closest to the Multimedia Ontologies Annotator[11]. Moreover, we aim to import existing MPEG-7 xml files to support backward compatibility to existing systems. Our system is a general framework that delegates the annotation and querying power from the system to the user's hands by allowing him to configure the system with his user-defined ontologies, in opposition to most of the above systems.

## 3. MPEG-7 Standard and MPEG-7 Ontology

It is preferable that all systems use the same format for annotating the multimedia content. In other words, standardizing the metadata of the content is much better than each system using its own annotation format. To visualize the benefit of multimedia data content standardization, assume you annotated your video in a system that is using the standard format. Then, you can query your annotated video in any system that is using the same standard in annotation, since the systems are talking the same language. In real world, there is such a standard named as MPEG-7[6].

MPEG-7 is an ISO/IEC standard for descriptions of multimedia content. It can be classified into the group of standardised description schemes, however in contrast with other standardised description schemes, it has not been developed in a restricted application domain but it has been intended to be applicable to a wide range of application domains.

The goal of the MPEG-7 standard is to allow interoperable searching, indexing, filtering, and access of audiovisual (AV) content by enabling interoperability among devices and applications that deal with AV content description. MPEG-7 specifies the description of features related to the AV content as well as information related to the management of AV content. The scope of the standard is to define the representation of the description that is, the syntax and the semantics of the structures used to create MPEG-7 descriptions. The MPEG-7 does this by attaching complex semantics to the content.

In MPEG-7, semantically identical metadata can be represented in multiple ways due to lack of precise semantics in XML-based syntax. This unfortunately prevents metadata interoperability. To overcome these interoperability issues, efforts have been spent to translate MPEG-7 standard into an ontology and to enable its integration with other ontologies through appropriate frameworks, thus enhancing interoperability. There exist

four OWL/RDF proposals of MPEG-7 [13]. In the approach proposed by Jane Hunter [7], ABC ontology is used as the core ontology and it provides attachment points for integrating MPEG-7 and domain specific ontologies. Technically, the mpeg7: MultimediaContent class is defined as a subclass of the abc:Manifestation class, while the corresponding domain ontologies are assumed to be appropriately attached to corresponding ABC classes. Complexity of Hunter's MPEG-7 ontology is OWL-Full.

In Tsinaraki's MPEG-7 ontology [8], the semantic part of MPEG-7 is translated into an ontology that serves as the core ontology for the attachment of domain specific ontologies, in order to achieve MPEG-7 compliant domain specific annotations. Its complexity is OWL-DL.

Garcia and Celma's Rhizomik model [9] is fully automatic translation of the whole standard. Therefore, it is not limited to description schema and has an OWL-DL complexity.

Core Ontology for Multimedia, which is abbreviated as COMM[10] is re-engineering of MPEG-7 using DOLCE design patterns. COMM is an OWL DL ontology.

A brief comparison for existing four MPEG-7 ontologies is given in Table 1 [13].

| | **Hunter** | **DS-MIRF** | **Rhizomik** | **COMM** |
|---|---|---|---|---|
| **Foundations** | ABC | None | None | DOLCE |
| **Complexity** | OWL-Full | OWL-DL | OWL-DL | OWL-DL |
| **Coverage** | MDS+Visual | MDS+CS | All | MDS+Visual |
| **Applications** | Digital Libraries | Digital Libraries | Digital Right | MM Analysis |

*Table 1: Comparison of four MPEG-7 ontologies*

In our proposed system, we want to accept existing MPEG-7 annotated multimedia metadata xml files, and convert them to the system supported MPEG7 based ontology. Therefore, we chose to use an MPEG-7 ontology that fully covers the existing MPEG-7 standard, which directs us to use Rhizomik Model.

## 4. Ontological Video Model

In order to achieve efficient querying and retrieval of multimedia data, the data should be stored in the multimedia database in such a way that queries could be answered in a reasonable time. This can be carried out by using a powerful video modelling technique. In Advanced Video Information System (AVIS) [14], video is divided into frame sequences to which activities, events and objects are associated. These associations are modelled by using special data structures. These data structures are frame segment tree, which is abbreviated as FST, and arrays that contain activities, events and objects. The model has been extended in order to support spatio-temporal query types [16]. In our current system we focus on providing the necessary infrastructure for a general framework, which makes use of ontologies in annotation and spatio-temporal querying of videos.

MPEG-7 ontologies try to find an elegant way to integrate MPEG-7 ontology with domain-specific ontologies. Our goal on the other hand, is to automate the integration of MPEG-7 ontology and domain-specific ontologies in an acceptable way. In order to automate the integration of MPEG-7 and user-define ontologies, we propose a video model ontology specification inspired by AVIS. The resulting video specification is the glue between MPEG-7 ontology and other user-defined ontologies. The specification is composed of the following rules:

1. There is a concept class, which will be the superclass of the classes in the user-defined ontology that are to be used in annotation and querying.
2. There is a temporal holder class, which is the subclass of the mpeg7: VideoSegmentType in Rhizomik MPEG-7 ontology.
3. There is a hasAppearanceOf property whose domain is video segment class and range is video concept class.
4. There is an isAppearedOn property, which is inverse property of hasApperanceOf property.

The third and fourth items of the specification are inspired by AVIS FST tree. In our system, we implemented this specification and use it to stick MPEG-7 ontology and domain-specific ontologies together.

## 5. Proposed System: Ontological Multimedia Information Management System

Two approaches exist in binding MPEG-7 and domain-specific ontologies. In the first approach, which is Hunter's MPEG-7/ABC approach, there is a core ontology that provides attachment points for MPEG-7 ontology and other domain-specific ontologies. In the second approach, which is MPEG-7/Tsinaraki approach, MPEG-7 is the core ontology and it provides attachment points for other ontologies to bind.

In our work, we first examined Hunter's MPEG-7 ontology, however we had technical problems due to its incompleteness and therefore we head for another MPEG-7 ontology. We decided to use Rhizaomik MPEG-7 ontology because of its one-to-one correspondence with MPEG-7 schema. In addition, we adopted Tsinaraki's approach, which uses MPEG-7 as the core ontology. However, we do not leave the users with the complex details of binding their domain-specific ontologies to attachment points provided by MPEG-7 ontology. Instead, we define a standard way of integrating MPEG-7 ontology and other user-defined ontologies and automate the harmonization process.

As stated above, in our system we use Rhizomik MPEG-7 ontology on top and wrap the given user-defined ontologies with MPEG-7 ontology, thus building MPEG-7 based ontologies automatically. Prior to wrapping the user-defined ontology by MPEG-7 ontology, we let the user select concepts that are going to be used in annotation and querying. The concepts that are going to be used in annotation and querying are wrapped by MPEG-7 ontology. The user interface for this process is shown in Figure 1.

The screenshot in Figure 1 is taken from Ontology Management module of our system. In this module, the user enters his ontology to the system by specifying a model name for the imported ontology and by giving the system the path of the file of the imported ontology. When the user clicks IMPORT button, the given model is imported by the system. Afterwards, the system loads the nontrivial concepts in Concept Selection pane, allowing the user to select which concepts in this ontology are subject to be used in annotation and querying. When the user clicks EMBED button, the system integrates the ontology, actually the selected concepts, with the MPEG-7 ontology.

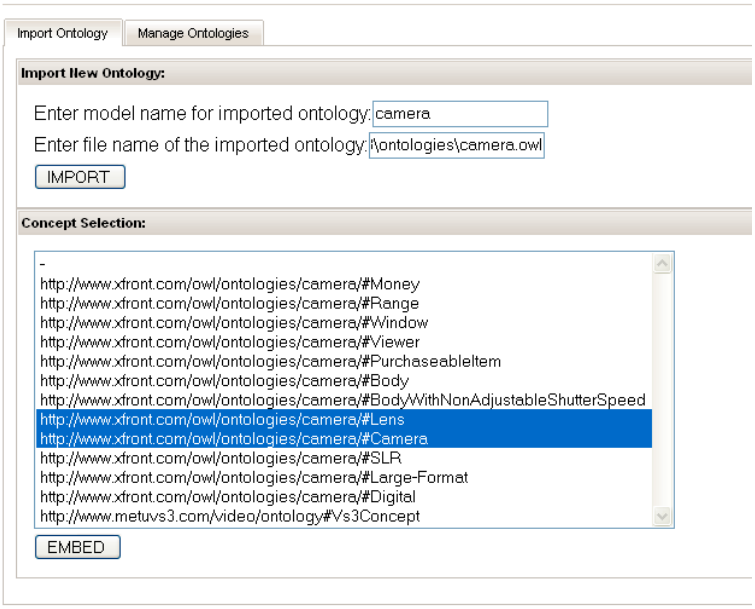


*Figure 1: Screenshot for binding MPEG-7 and user-defined ontology*

The screenshot in Figure 2 is taken from New Video Metadata page of the system. In this page, the user annotates his videos by using the concepts from his ontology that is entered to the system via Ontology Management page. In this page, the user can specify the interval to annotate, via JMF applet by clicking “Click to see JMF Applet” link in Video Properties pane. Moreover, the user can automatically select region of an object by using JMF applet. Meanwhile, region selection is meaningful for spatio-temporal queries and it is provided by the system infrastructure in order to support spatio-temporal querying. The user selects the model in Model Selection pane to load concepts that are going to be used in annotation. The embedded concepts of the selected model are loaded in the listbox in Concept Selection pane. Then the user selects a concept. Following concept selection, the individuals belonging to that concept are loaded to individual combo-box and attributes for the selected concept are loaded in listbox in Attribute Selection pane. The user can annotate the retrieved attributes in Attribute Selection pane.

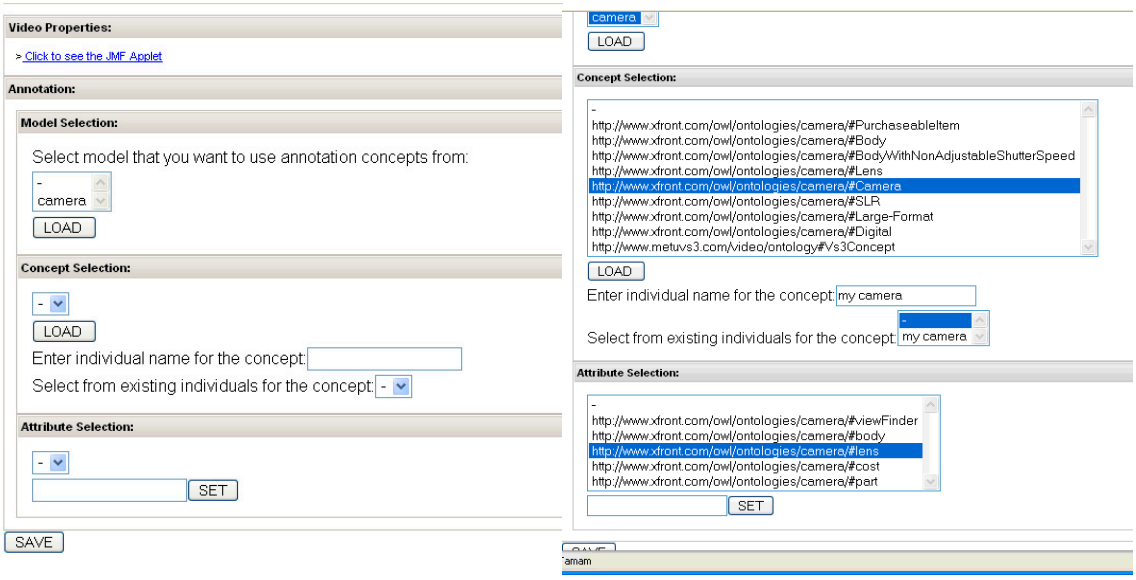


*Figure 2: Annotation screenshot: before and after loading concepts*

In the implementation, we have used JSF and facelets in front side with myfaces and richfaces components; Java programming language in server side; Jena as ontology API and MySQL as database.

## 6. Conclusion and Future Work

The ontological multimedia information framework can be easily used in some specific domain naturally. For example, a user who wants to use the system in the soccer domain, can achieve this just by feeding the program with a soccer-domain ontology. Moreover, the system can easily be modified according to domain-specific requirements due to its modular architecture. For example, if the user wants to specialize the interface according to a specific domain, it is possible since the implementation is as modular as possible.

In our proposed work, we focused on binding MPEG-7 ontology and domain-specific ontologies in a standard and automated way, the general annotation and querying features, importing existing MPEG-7 xml files into system. We do not consider automating the annotation process itself. As a future work, automatic or semi-automatic annotation feature may be added to the system. The architecture of the system is designed to allow this improvement. Furthermore, a domain-specific multimedia content management system can be put on top of the proposed system.

Our proposed system is an ontological multimedia information management framework with a modular architecture, ease of integrating with user-defined ontologies naturally and automatic harmonization of MPEG-7 ontology and domain-specific ontologies. The system makes use of existing user-defined ontologies in multimedia content annotation. Therefore, the system delegates the querying power of the multimedia management system from the system to the user-defined ontology.

The system opens a door for many future works due to its general structure. One of the aims of this work is to give a starting point to developers who want to develop a multimedia information management system in a specific domain. With this system in their hand, they do not have to worry about MPEG-7 details, ontology API details and interface details

## Acknowledgements

This work is partially supported by The Scientific and Technical Council of Turkey Grant ‘‘TUBITAK EEEAG-107E234”.

## References

1. S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V.Tzouvaras, Y. Avrithis, S. Handschuh, Y. Kompatsiaris, S. Staab and M. G. Strintzis, "Semantic Annotation of Images and Videos for Multimedia", in Proc. of 2nd European Semantic Web Conference, (ESWC '05),Heraklion, Greece, May 29 - June 1,2005.
2. E. Spyrou, G. Koumoulos, Y. Avrithis et al., "K-Space at TRECVid 2006", 4th TRECVID Workshop, Gaithersburg, USA, November 2006.
3. E. Şaykol, Web-based user interface for query specification in a video database system, M.S. thesis, Dept. of Computer Engineering, Bilkent University, Ankara, Turkey, Sept. 2001.
4. M. Christel, T. Kanade, M. Mauldin, R. Reddy, M. Sirbu, S. Stevens, and H. Wactlar, "Informedia Digital Video Library," Communications of the ACM, Vol. 38, No. 4, 1995, pp. 57 - 58.
5. Shih-Fu Chang ; Chen, W. ; Sundaram, H. VideoQ: a fully automated video retrieval system using motion sketches [1998-01-01]
6. http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
7. J. Hunter, "Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology", International Semantic Web Working Symposium (SWWS), Stanford, July 30 - August 1, 2001
8. C. Tsinaraki, P. Polydoros and S. Christodoulakis. Interoperability support for Ontology-based Video Retrieval Applications. In Proc. of 3rd International Conference on Image and Video Retrieval (CIVR 2004), Dublin, Ireland, 21-23 July 2004.

9. R. Garcia and O. Celma. Semantic Integration and Retrieval of Multimedia Metadata . In Proc. of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005), Galway, Ireland, 7 November 2005.
10. Richard Arndt, Raphaël Troncy, Steffen Staab, Lynda Hardman, Miroslav Vacura: COMM: Designing a Well-Founded Multimedia Ontology for the Web. ISWC/ASWC 2007: 30-43
11. http://www.ercim.org/publication/Ercim_News/enw66/del_bimbo.html
12. http://www.w3.org/2005/Incubator/mmsem/XGR-mpeg7/
13. MPEG-7 based Multimedia Ontologies: Interoperability Support or Interoperability Issue?,Troncy R., Celma O., Little S., García R., Tsinaraki C.,Multimedia Annotation and Retrieval enabled by Shared Ontologies Workshop, MARESO'07.
14. S. Adali, K.S. Candan, S. Chen, K. Erol, V.S. Subrahmanian, The advanced video information system: data structures and query processing, Multimedia Systems 4 (1996) 172–186.
15. Arslan, U., Donderler, M.E., Saykol, E., Ulusoy, Ö., Gudukbay, U., A Semi-Automatic Semantic Annotation Tool for Video Databases, In SOFSEM 2002, Workshop on Multimedia Semantics, Milovy, Czech Republic, November 2002.
16. Koprulu M., Çiçekli, N.K., Yazici, A., Spatio-temporal Querying in Video Databases, Information Sciences 160, 2004, Elsevier Science, pp. 131-152, 2004.,