Towards an Ontology for MPEG-7 Semantic Descriptions

Md. Anisur Rahman, M. Anwar Hossain, Iluju Kiringa and Abdulmotaleb El Saddik
Multimedia Communication Research Laboratory (MCRLab)
School of Information Technology and Engineering (SITE)
University of Ottawa, Canada
{anis, anwar}@mcrlab.uottawa.ca, kiringa@site.uottawa.ca, abed@mcrlab.uottawa.ca

Abstract

Multimedia resources may be described using several metadata standards, MPEG-7 being the most comprehensive among those standards. MPEG-7 provides different tools to describe any multimedia content. Semantic Descriptor Scheme is one of them, which is used to describe the semantics of the content in terms of Events, Objects, Concepts, Places, Time and Abstraction. Since there is no hard and fast rule regarding how the Semantic Descriptor Scheme should be used, different semantic descriptions may result for a single multimedia object. These semantic descriptions may be unified in order to enhance the overall knowledge about the associated multimedia object. We proposed an ontology in this paper that semantically represents the structure of MPEG-7 Semantic DS and primarily acts as a resource of such unification. The knowledge representation provided by this ontology can be used to develop tools that perform automatic multimedia reasoning from different existing semantic descriptions, which are narrated with other domain specific ontology.

1. Introduction

In the recent years most of the media are reproduced as digital media in different forms such as image, audio, video and multimedia in general. As World Wide Web carries out the distribution of the media, availability of digital media has hugely increased. Fine-grained computerized support is a must to manage that enormous amount of digital media. A lot of efforts have been invested on automatic segmentation or structuring of multimedia content. However, very little development was made on machine-generation of semantic descriptions of audiovisual information. Knowledge-based techniques based on Web ontologies can be a solution to overcome this problem. Ontologies can provide automated integration and meaningful retrieval of multimedia – both content and metadata from different sources.

The term ontology is defined as a specification of a conceptualization [17]. That is, an ontology is a description of the concepts and relationships that can exist, like a formal specification of a program, providing a shared and common understanding of a domain that can be communicated between people and computer systems. Each multimedia object evokes one or more concepts, as any word of a vocabulary does. Considering this Massimiliano et. al. [13] informally defined a Multimedia Ontology as a mean for specifying the knowledge of the world through multimedia objects and representing the organization of multimedia documents in a structured way such that users and applications can process the descriptions with reference to a common understanding.

There are several standards in the multimedia domain to describe multimedia objects. MPEG-7 [1, 2] is the most comprehensive among all. MPEG-7 uses several XML-based descriptors and description schemes. This approach limits the reasoning power of MPEG-7 that would otherwise allow automatic deduction of facts from multimedia descriptions. However, if an ontology can be built for MPEG-7, it can serve both as data model and definition of the semantics of MPEG-7 terms and relation between them. Developing the ontology may also be helpful for resolving any probable inconsistencies, ambiguities or duplication among the MPEG-7 descriptor schemes and descriptors. Class and property hierarchies, which could have been taken from the data model if it was provided with the specification of MPEG-7, can now be derived from the English-text semantic descriptions.

The key components of MPEG-7 semantic descriptions are semantic entities such as objects and events, attributes of these entities such as labels and

properties, and finally, relations of these entities such these are very powerful and useful as a description tool, they can be a potential candidate to develop smart applications capable of dealing with multimedia at high level.

Several researches [8, 11, 13] attempted to provide ontology for MPEG-7 as a whole or in part. However, as MPEG-7 is too large and generic, it is very difficult to represent all its features in greater details in a single ontology. Therefore, they used a core subset of the MPEG-7 specification together with a top-down approach to generate the ontology. However, to the best of our knowledge, an ontology to cover the Semantic Descriptor part of MPEG-7 remains uncovered, which is presented in this paper. We name the ontology as Semantic Content Description Ontology (SCDO). This ontology may be used for the unification of different MPEG-7 semantic descriptions of multimedia contents. Later we describe a framework that utilizes SCDO in combination with domain-specific ontologies that are used in different descriptions.

The remaining outline of the paper is as follows. Section 2 states some related works. Section 3 provides an overview of the MPEG-7 description tools. Section 4 discusses the MPEG-7 Semantic Description tool in details. In Section 5 we give a motivating example. The design and implementation of SCDO is given in section 6. In Section 7 we present a framework that uses SCDO to facilitate unification and retrieval of facts from different multimedia descriptions. We conclude and present the future work in Section 8.

2. Related work

Idea of developing ontology for MPEG-7 descriptors is not new. Most of the works in this direction was motivated by the work of Jane Hunter [6, 7, 8]. The ontology developed by Hunter mainly focuses on low level descriptors, such as segment, color, regionLocator etc. But his work didn't cover the content dimension and corresponding annotations, such as person, vehicle or an event. The subsequent works in this area either contributed very little or totally ignored that part. In our work we mainly give emphasis to Semantic Descriptor part of MPEG-7.

In [9] a methodology is given for the interoperability of OWL with the MPEG-7 Multimedia Description Scheme (MDS) so that domain ontologies described in OWL can be integrated with the MPEG-7 MDS metadata. Unlike ours, this paper mainly focuses on video descriptions.

The authors in [11] developed an ontology that represents the structure of the MPEG-7 visual part to as an object being the participant of an event [4]. As enable machines to generate and understand visual descriptions which can be used for multimedia reasoning. This work lacks any architecture or framework for using the developed ontology.

An automatic mapping of the MPEG-7 standard to OWL is given in [14]. It is based on a generic XML Schema to OWL mapping. The authors linked three different music schemas to retrieve related information from instances of all the metadata sources. This linking was done through ontology that maps MPEG-7 vocabulary to MusicBarinz, Simac and Music vocabularies. However, the MPEG-7 vocabulary used for such linking does not cover the Semantic Content Description tool in general.

The tools standardized by MPEG-7 for describing the semantics of multimedia are presented in [3]. The authors discuss the key components of MPEG-7 semantic descriptions - semantic entities such as objects and events; attributes of these entities such as labels and properties; and relations of these entities such as an object being the agent of an event. They used English-Text rather than any formal model to express the semantics. This work remains as our guideline for formalizing the entities, attributes and relations by creating an ontology for them.

In [15] the authors discuss a tool for semantic annotation and search in a collection of art images. They used multiple existing ontologies to support this process, including the Art and Architecture Thesaurus, WordNet, ULAN and Iconclass. The authors used MPEG-7 ontology as a basis and developed an upper ontology by reusing several domain specific ontologies. Although they tried to capture some semantic aspect while describing an image, their semantic description does not use the MPEG-7 vocabulary.

A framework was presented in [16] for building an ontology to provide semantic interpretations in image contents. The authors developed an ontology for representations of multimedia contents in the domain of natural scenes and used it on top of MPEG-7 ontology to enable fast and efficient image query and retrieval. The MPEG-7 ontology on which their framework is based doesn't include the Semantic Content descriptor tools.

3. MPEG-7 description tools

Descriptors (Ds) and Description Schemes (DSs) representing features of multimedia, and more complex structures grouping these Ds and DSs are the components of The MPEG-7 framework.

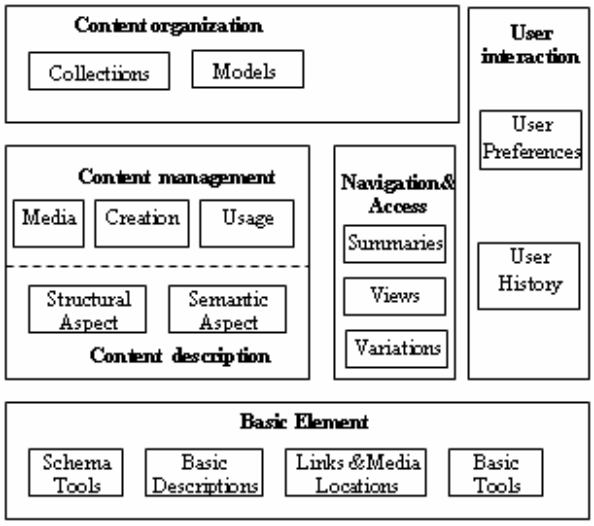

Figure 1. MPEG-7 description tools [1]

MPEG-7 standard comprises of different parts which provides a set of multimedia description tools to generate descriptions. The basic structure of MPEG-7 tools is depicted in Figure 1 and a brief description of the tools is given below.

- The basic elements include data types, vectors, matrices, and constructs for linking media files, localizing pieces of content, places, time, persons, individuals, groups, organizations, and other textual annotations.
- The Content Description tools represent perceptible information, including structural aspects (structure description tools), audio and visual features, and conceptual aspects (semantic description tools).
- The Navigation and Access tools facilitates browsing and retrieval of audio-visual content by defining summaries, partitions and decompositions, and variations of the audio-visual material.
- The Content Organization tools are used for organizing and modeling collections of audio-visual content and descriptions.
- The User Interaction tools describe usage history and user preferences related to the consumption of the multimedia object

Among the tools, Semantic DS is used for describing the semantics of multimedia, and is the target of our ontology. In the following section we discuss the Semantic DS tool in details.

4. Semantic description tool

The MPEG-7 Semantic DS tool describes the semantics of multimedia content in terms of events. Events can be perceived as occasions upon which something happens. Time, objects, people and places can populate such occasions. These entities can have properties and states and are interrelated among themselves. In MPEG-7 the participants, background, context, and all the other information that makes up a single narrative are referred to as a "narrative world" [4]. The components of the semantic descriptions are entities in narrative worlds, their attributes, and their relations. A multimedia object may have multiple narrative worlds, or vice versa.

An example of a semantic description of an image is illustrated in Figure 2. In this example, two persons, an event, a place, a time and a concept depicted or symbolized in the image are described together with several relationships among them.

A narrative world in MPEG-7 is represented using the Semantic DS, which is described by a number of semantic entities and of graphs of their relationships. The Semantic DS is derived from the SemanticBag DS, which is an abstract type representing any kind of collection of semantic entities and their relationships. Some specialized SemanticBase DSs are the Object, AgentObject, Event, SemanticPlace, SemanticTime, SemanticState and Concept DSs. These are the entities that collectively describe the narrative world such as an object, agent and event; the where and when of things; a parametric entity, and a concept, respectively.

The Object and Event DSs describe perceivable semantic entities (objects and events) that can exist or take place in time and space in narrative worlds. The Object DS and Event DS are recursive to describe the subdivision of objects and events into other objects and events. The AgentObject DS extends from the Object DS to describe an object that acts as a person, a group of persons or an organization in a narrative world. The SemanticPlace and SemanticTime DSs describe a location and time in a narrative world, respectively. The SemanticState DS describes the state or parametric attributes of a semantic entity at a given place and/or time in the narrative world. The Concept DS describes concepts as collections of one or more properties.

MPEG-7 has defined some standard semantic relations. However, descriptions of non-normative relations are also allowed. Normative semantic relations may describe how several semantic entities

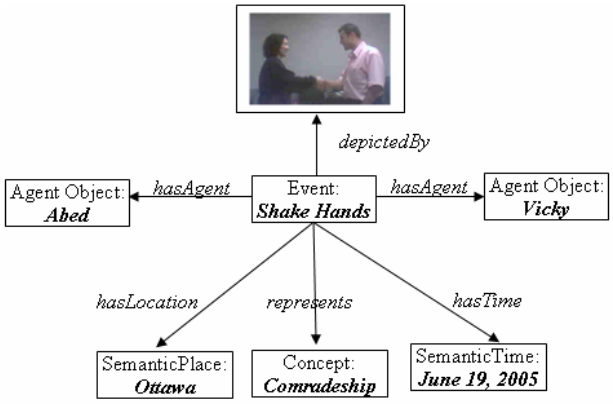

Figure 2. Semantic description of an image

relate in a narrative (e.g., agent, patient, and accompanier). For example, semantic relations may describe relations between objects and/or events that draw upon typical linguistic relations such as an object being the agent of an event.

5. Motivating example

MPEG-7 formalism lacks the semantics and reasoning support in many ways. For example, the search of any implicit fact cannot be deduced from MPEG-7 documents as its description is based on XML Schema. To overcome such limitations several attempts to create MPEG-7 ontology have been made [6, 11, 12, 15]. However, those researches did not focus on MPEG-7 Semantic Descriptor tools, which is very important in terms of unifying several semantic descriptions of multimedia objects. By the following example we describe a scenario that shows a potential problem that may arise due to the absence of Semantic Descriptor ontology.

Let us consider the scenario shown in Figure 2. In that scenario, "Abed" is "Shaking Hands" with "Vicky". A possible MPEG-7 description of the event in that scenario is given in Figure 4 and a simple tree representation of a selected part (Event, Object, and their relationships) is shown in Figure 3.a. Let us now consider another scenario where "Abed" is "Shaking Hands" with "Vicky" in a different occasion. The tree representation of this event is shown in Figure 3.b.

Without the presence of Semantic Descriptor ontology, there is no way to deduce that the event descriptions of the two scenarios are semantically same. Such an ontology may describe that AgentObject is a sub-class of Object and hasAgent is an inverse property of agentOf. Given these relationships, the Object instances "Abed" and "Vicky" in Figure 3.a

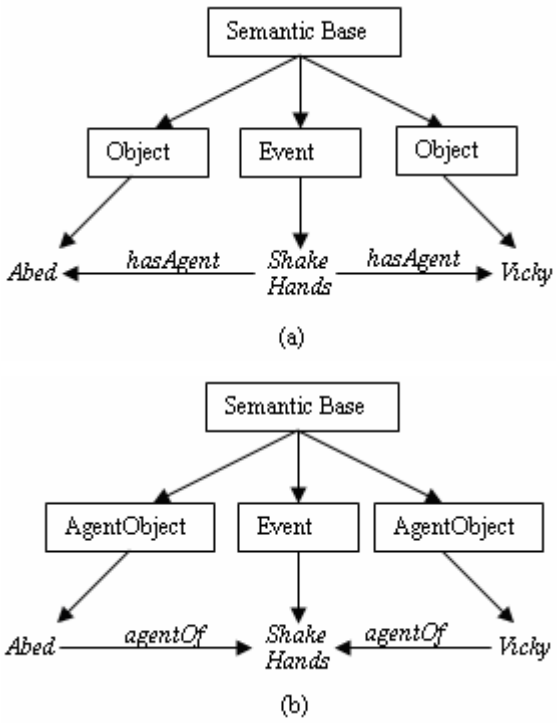

Figure 3. Semantic description of two images in tree structure

fall into the same class of AgentObject in Figure 3.b. We were motivated for creating an Ontology for retrieving such implicit information and thus unify different descriptions.

6. Design of SCDO

MPEG-7 standardizes the core technologies allowing description of audiovisual data content in multimedia environments. This effort was achieved by standardizing Descriptors, Descriptor Schemes and Description Definition Language [11]. Descriptors are the representations of features that define the syntax and the semantics of each feature representation.

Description Schemes on the other hand specify the structure and the semantics of the relationships between their components. These components can be Descriptors or Description Schemes. Description definition language allows the definition of Descriptors and Descriptor schemes. MPEG-7 makes this standardization into seven parts. Semantic Content Descriptor Ontology (SCDO) focuses on the part that specifies the conceptual aspect of the content. It specifies both the concepts inherent in semantic description features (e.g. event, object, concept etc.) and their interrelationships.

```
<Mpeg7>
 <Semantic>
  <Label><Name>Abed shakes hands with vicky </Name></Label>
  <SemanticBase xsi:type="EventType" id="EV1">
   <Label><Name>Shake hands</Name></Label>
   <Relation xsi:type="ObjectEventRelationType“  name="Agent“   target="#AO1"/>
   <Relation xsi:type="ObjectEventRelationType“ name="Agent“ target="#AO2"/>
   <Relation xsi:type="ConceptSemanticBaseRelationType“  name="hasPropertyOf" target="#C1"/>
   <SemanticPlace> <Label><Name>Ottawa</Name></Label> </SemanticPlace>
   <SemanticTime> <Label><Name>June 19, 2005</Name></Label> </SemanticTime>
  </SemanticBase>
  <SemanticBase xsi:type="ObjectType" id="AO1">
   <Label><Name>Abed</Name></Label>
  </SemanticBase>
  <SemanticBase xsi:type="ObjectType" id="AO2">
   <Label><Name>vicky</Name></Label>
  </SemanticBase>
  <SemanticBase xsi:type="ConceptType" id="C1">
   <Label><Name>Comradeship</Name></Label>
   <Property>Associate</Property> <Property>Friend</Property>
  </SemanticBase>
  </Semantic>
</Mpeg7>
```

Figure 4. An MPEG-7 Description describing the semantics of an image

SCDO is intended for giving a common language for sharing and reusing knowledge about phenomena in MPEG-7 Semantic Descriptors.

We followed inspirational approach for the design of SCDO. We were motivated to develop the ontology to deal with situations as stated in section 5. The concepts of SCDO were created to accommodate the representation and processing of descriptive, procedural, and/or reasoning knowledge within a system devised for supporting decision-making about the similarity/dissimilarity of different multimedia content descriptions.

6.1. SCDO Structure

This section describes the structure of the SCDO. In general, ontology consists of several components of which the most important are concepts, relations, attributes, instances, and axioms [18]. Concepts in general are abstract terms and are organized in taxonomies. These are the vital part of any ontology. In SCDO there are several concepts (classes) shown in Figure 5 and relations (properties) shown in Figure 6.

6.1.1. Classes in SCDO

The classes (e.g. SemanticBase, Object, Event etc.) depicted in Figure 5 represents the entities of MPEG-7 Semantic DS tool as mentioned in Section 4. As SemanticBase describes narrative world and semantic entities related to audio-visual content, it is the root class of our ontology. A number of specialized subclasses are derived from this class such as Object, AgentObject, Event, Concept, SemanticPlace, and SemanticTime. The AgentObject class is an extension of Object class. Person, Organization, GroupOfPeople and PersonalizedObjects are sub-classes of AgentObject. The Event class describes a perceivable or abstract event. The Concept class describes a semantic entity that cannot be described as a generalization or abstraction of a specific object, event, time, place, or state. The Location class is an extension of SemanticPlace class, which is further sub-classed as Country, State/Province, City and Zip/PostalCode. The SemanticTime class is extended with Year, Month, Day, Hour, Minute and Second classes.

6.1.2. Properties in SCDO

MPEG-7 standard describes several normative semantic relations [3] that act as properties in SCD ontology. These relations may be abstractly categorized into three groups: narrative relations, definitive relations and spatial/temporal relations. Narrative relations illustrate the connections among the entities in a narrative world.

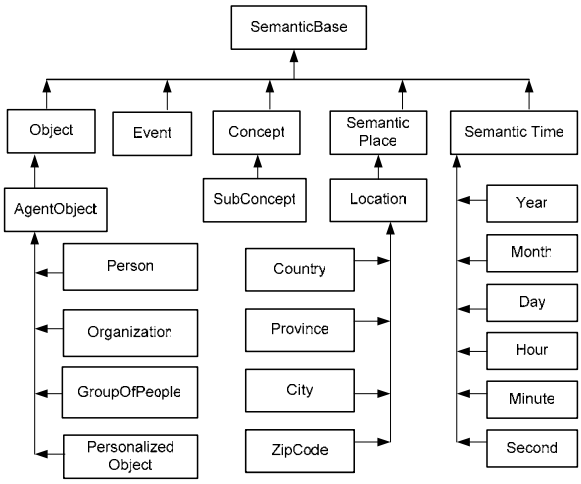

Figure 5. Class hierarchy of SCDO

Some example of narrative relations are hasAgent, agentOf, patient, patientOf etc. as listed in Table 1. Figure 6 shows an example where we use such relations. The property agentOf relates an individual of AgentObject class to an individual of Event class.

Therefore, the domain of the property agentOf would be AgentObject and its range would be Event. The property hasAgent is the inverse property of AgentOf (i.e. the domain of hasAgent is Event and its range is AgentObject).

Definitive relations are used to find the relationships among definitions of several semantic entities. These definitions define such relationships whether an entity is a combination of two or more entities, or is a generalization of another entity and so on.

6.2. Implementation of SCDO

The class hierarchy, attributes and properties of SCD ontology has been represented using OWL, the W3C's Ontology Language for the Web. The ontology properties described in Section 6.1.2 link an individual to an individual and are referred to as Object properties in OWL. The Protégé 2.0 [available at: http://smi.stanford.edu/projects/protege/] tool was used to create the SCD Ontology. As an ontology development tool Protégé provides considerable functionality both at the modeling and user-interface levels.

7. System architecture

The process of unifying different MPEG-7 semantic descriptions of multimedia objects through SCDO is highlighted in the conceptual architecture in Figure 7.

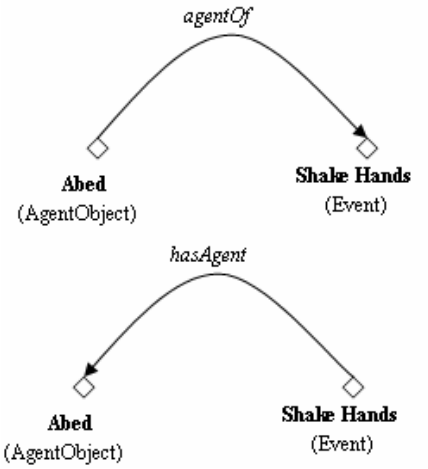

Figure 6. An example of properties in SCDO

The architecture is composed of several modules such as query interface, query processor, OWL reasoner, and MPEG-7 parsing engine. In addition to this the architecture uses several other resources including existing MPEG-7 descriptions of multimedia files and domain specific ontologies that are referenced in those descriptions. The unification process is handled by using our SCD ontology.

Table 1. Semantic relations in MPEG-7

Type	Normative relations
Narrative	agent, agentOf, patient, patientOf, experiencer, experiencerOf, stimulus, stimulusOf, causer, causerOf, goal, goalOf, beneficiary, beneficiaryOf, them, themOf, result, resultOf, instrument, instrumentOf, accompanier, accompanierOf, summarizes, summarizedBy, state, stateOf
Definitive	combination, specializes, generalizes, similar, opposite, exemplifies, exemplifiedBy, interchangeable, identifier, part, partOf, contrasts, property, propertyOf, user, userOf, component, componentOf, substance, substanceOf, entailment, entailmentOf, manner, mannerOf, influences, dependsOn, membershipFunction
Spatial/ Temporal	key, keyFor, annotes, annotatedBy, shows, appearsIn, reference, referenceOf, quality, qualityOf, location, locationOf, source, sourceOf, destination, destinationOf, path, pathOf, time, timeOf, depicts, depictedBy, represents, representedBy, context, contextFor, interprets, interpretatedBy

In the following we briefly describe the functionalities of the modules.

Query Interface: The user of the system interacts through this interface to pose semantic query and gets the result returned by the system.

Query Processor: This is the central module of the system. With the help of the OWL Reasoner module it determines which variation of structures and values of MPEG-7 semantic descriptions can be the target of a given query. It then uses the MPEG-7 Parsing Engine to get all MPEG-7 descriptions that matches some of the variants.

OWL Reasoner: The SCDO is implemented in OWL language. The domain-specific ontology is also assumed to be expressed in OWL. The OWL Reasoner uses those OWL documents to extract the matching concepts and properties to a given set of queries.

MPEG-7 Parsing Engine: MPEG-7 documents are written in XML Schema. This module is dedicated to search those documents to find a match.

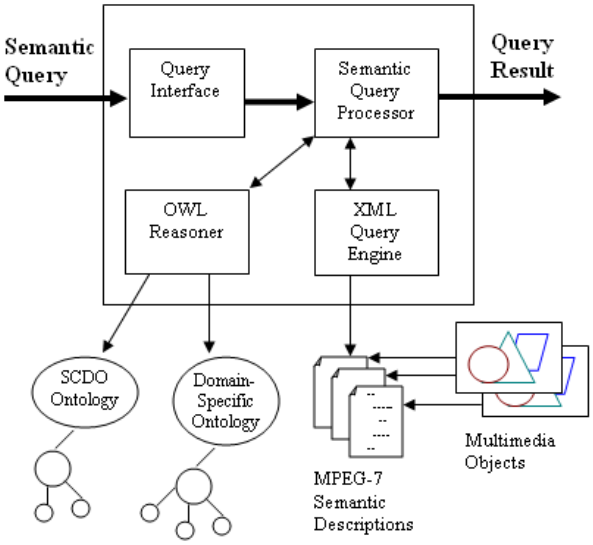

Figure7. System architecture

8. Conclusion

In this paper we presented the construction of an ontology that represents the structure of the MPEG-7 Semantic Content Description tool. The goal of this ontology is to extract facts by unifying different MPEG-7 semantic descriptions related to multimedia files. We also provided a framework that uses the ontology for such unification. The use of ontology enables machines to generate and understand semantic content descriptions, which can be leveraged for efficient multimedia reasoning. The knowledge representation provided by the ontology can be utilized to develop different tools targeting different multimedia domains such as multimedia content management, intelligent information retrieval, etc. One of the visible limitations of our approach is that we used SCDO in a test environment. However, as a future work we like to address this issue and plan to adopt the ontology in larger information retrieval domain such as web searches.

9. References

[1] M.M. Jose, *MPEG-7 Overview*, Technical Report, ISO/IEC JTCI/SC/29/WG11 N6828, 2004.

[2] S.F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 688-695, 2001.

[3] A.B. Benitez et al., Description of a Single Multimedia Document. Manjunath, B. S., Salembier, P., Sikora, T. (eds.). Introduction to MPEG-7 – Multimedia Content Description Interface, John Wiley & Sons Ltd. (2002), pp. 111–138.

[4] A.B. Benitez et al., "Semantic of multimedia in MPEG-7," *Proc. IEEE Conf. on Image Processing* (ICIP-2002), Rochester, New York, USA, September 2002.

[5] M. Horridge, H. Knublauch, and A. Rector, A Practical Guide To Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools. Edition 1.0, Technical Report, University Of Manchester, 2004.

[6] J. Hunter, "Enhancing the semantic interoperability of multimedia through a core ontology," *IEEE Trans. Circuits and Syst. Video Technology*, vol. 13, Jan. 2003, pp. 49-58.

[7] J. Hunter, "Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology," *International Semantic Web Working Symposium* (SWWS), Stanford, 2001.

[8] M. Doerr, J. Hunter, and C. Lagoze, "Towards a Core ontology for Information Integration," *Journal of Digital Information*, vol. 4, no. 1, April 2003.

[9] C. Tsinaraki, P. Polydoros, and S. Christodoulakis, "Interoperability support for Ontology-based Video Retrieval Applications," Proc. *CIVR*, 2004.

[10] S. Hammiche et al., "Semantic Retrieval of Multimedia Data**,"** *The Second ACM International Workshop on Multimedia Databases*, 2004.

[11] N. Simou et al., "A Visual Descriptor Ontology for Multimedia Reasoning," *Workshop on Image Analysis for Multimedia Interactive Services* (Montreux, Switzerland, 2005).

[12] H.H. Kim, S.S. Park, and W. Kim, “A Framework for the Integration of Multimedia Data,” *Journal of Object Technology*, vol. 4, no. 5, pp. 27-35, Jul.-Aug. 2005.

[13] M. Albanese, P. Mareska, and A. Picariello, “Towards a Multimedia Ontology System: an Approach Using TAO_XML,” *Technical Report, FIRB Project*, 2005.

[14] R. García and O. Celma, “Semantic Integration and Retrieval of Multimedia Metadata,” *Proc. of International Semantic Web Conference* (ISWC 05), 2005.

[15] G.L. Hollink, J. Schreiber, and B. Wielinga, “Semantic annotation of image collections,” *In Workshop on Knowledge Markup and Semantic Annotation, KCAP’03*, 2003. Available at http://www.cs.vu.nl/˜guus/ papers/ Hollink03c.pdf.

[16] S. Liu, L.T. Chia, and S. Chan, “Ontology for nature-scene image retrieval,” *In Proceedings OTM confederated international conferences*, Oct. 2004.

[17] R. Neches et al., “Enabling technology for knowledge sharing,” *AI Magazine*, vol. 12, no. 3, pp. 36-56, 1991.

[18] T.R. Gruber, “A translation approach to portable ontologies,” *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220, 1993.