Multimedia Ontologies

Lambrini Seremeti
Hellenic Open University &
Technological Educational Institute of Lefkada
11 Kapodistriou str, Neapoli
Lefkada, Greece
+30-26450-26160
seremeti@vodafone.gr

Achilles Kameas
Hellenic Open University
23 Sahtouri str
26222, Patras, Greece
+30-2610-367402
kameas@eap.gr

ABSTRACT
The development and maintenance of large multimedia databases has attracted much attention nowadays frm companies and organizations that held multimedia content (archives, broadcast companies, radio and TV channels, etc.). the goal is to bypass the ineffective and time-consuming process of manual searching and retrieval of multimedia content and use computers to make the content easy to be found and accessible to other parties. Thus, two critical points are identified in making the above goal a reality; effective representation as well as effective retrieval and exploitation of multimedia content. For accomplishing the above goal researchers have started to use ontologies in the field of multimedia in order to construct machine-understandable, descriptive versions of the multimedia content based on multimedia ontologies. In this paper, we present multimedia ontology characteristics, its construction and some existing multimedia ontologies.

Categories and Subject Descriptors
I.2.4 [Knowledge Representation Formalisms and Methods: Semantic networks; H.3.7 [Information Storage and Retrieval]: Digital Libraries

General Terms
Management

Keywords
Metadata, Ontology, Multimedia Ontology

1. INTRODUCTION
The amount of digital multimedia information accessible to the end-users is growing every day, not only in terms of consumption but also in terms of production. But if it is today easier and easier to acquire, process and distribute multimedia content, it should be equally easy to access the available information, because huge amounts of digital multimedia information is being generated, all over the world, every day. In fact, there is no point in making available multimedia information that can only be found by chance. Unfortunately, the more information becomes available, the harder it is to identify and find what you want, and the more difficult it becomes to manage the information. People looking for content are typically using text-based browsers with rather moderate retrieval performance. These text-based engines rely on human operators to manually describe the multimedia content with keywords and free annotations. This solution is increasingly unacceptable for two major reasons [19]:

- First, it is a costly process, and the cost increases quickly with the growing amount of content.
- Second, these descriptions are inherently subjective and their usage is often confined to the specific application domain for which the descriptions were created.

To enable multimedia content to be discovered and exploited by services, agents and applications, it needs to be described semantically. Although significant progress has been made in recent years on automatic segmentation or structuring of multimedia content and the recognition of low-level features within such content, the generation of descriptions of multimedia content is inherently problematic because of the volume and complexity of the data, its multidimensional nature and the potentially high subjectivity of human-generated descriptions.

Metadata are a representation of the administrative, descriptive, preservation, usage and technical characteristics associated with multimedia objects. They can be extracted manually or automatically from multimedia documents. This value-added information helps bridging the semantic gap, described as: "The lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation".

There are many ways to describe multimedia content using metadata, and, indeed, today many proprietary ways are already in use in various digital asset management systems. Such systems, however, do not allow a search across different repositories for a certain piece of content, and they do not facilitate content exchange between different databases using different systems. These are interoperability issues, and creating a standard is an appropriate way to address them.

The metadata standards increase the value of multimedia data which are used by various applications, such as: architecture, real estate, interior design; cultural services; digital libraries; e-commerce; education; home entertainment; investigation services; journalism; multimedia directory services; remote sensing; wireless and mobile environments. All these applications share the capability of handling multimedia information based on its semantic information.

The MPEG [23] family of standards address this kind of interoperability. For example, MPEG-7 defines the metadata, that is data about data, elements, structure, and relationships that are used to describe audiovisual objects including still pictures, graphics, 3D models, music, audio, speech, video, or multimedia collections, whereas MPEG-21 was developed to address the need for an overarching framework to ensure interoperability of digital multimedia objects [27].

Mobimedia'07, Month 8, 2007, Nafpaktos, Aitolokarnania, Greece.

It is necessary to have a common understanding of the semantic relationships between metadata terms from different domains. Representation and semantic annotation of multimedia content have been identified as an important step towards more efficient manipulation and retrieval of multimedia. In order to achieve semantic analysis of multimedia content, ontologies are essential to express semantics in a formal machine-processable representation [22].

In this paper, we shall introduce multimedia ontologies, as a candidate solution to the problem of representing heterogeneous metadata for multimedia content. In the next section we have included a brief introduction to ontologies, while in section 3 we shall present in detail the current state of the art in multimedia ontology construction tools and methodologies. The paper concludes with a discussion on future trends.

2. ONTOLOGIES

"Ontology" is the name given to a structure of knowledge, used as a means of knowledge sharing within a community. Gruber defined an ontology as "a formal, explicit specification of a shared conceptualization" [15]. A "conceptualization" refers to an abstract model of some phenomenon in the world, which identifies the relevant concepts of that phenomenon. "Explicit" means that the type of concepts used and the constraints on their use are explicitly defined. "Formal" refers to the fact that the ontology should be machine readable. "Shared" reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group.

Ontologies have been proposed to solve the problems that arise from using different terminologies to refer to the same concept, or using the same term to refer to different concepts. They consist of definitional aspects such as high-level schemas and assertional aspects, entities, attributes, interrelationships between entities, domain vocabulary and factual knowledge, all connected in a semantic manner [21]. They have generally been associated with logical inferencing and recently have begun to be applied to the Semantic Web [2]. Ontologies provide specific tools to organize and provide a useful description of heterogeneous content. For humans, ontologies enable better access to information and promote shared understanding. For computers, ontologies facilitate comprehension of information and more extensive processing. In summary, ontologies are about adding meaning to information in order to provide clear value, delivering "the right information to the right person at the right time and in the right way".

Ontologies can be classified, according to the issue of the conceptualization into [16]:

- **Representation ontologies or meta-ontologies**. They capture the representation primitives used to formalize knowledge in a given knowledge representation system.
- **General or upper-level ontologies**. They classify the different categories of entities existing in the world. Very general notions which are independent of a particular problem or domain are represented in these ontologies. Knowledge defined in this kind of ontologies is applicable across domains and includes vocabulary related to things, events, time, space.
- **Domain ontologies**. They are more specific ontologies. Knowledge represented in this kind of ontologies is specific to a particular domain. They provide vocabularies about concepts in a domain and their relationships, or about the theories governing the domain.
- **Application ontologies**. They describe knowledge pieces depending both on a particular domain and task.

Ontologies can effectively be used to perform semantic annotation of multimedia content. They can also be used for multimedia object retrieval.

3. MULTIMEDIA ONTOLOGIES

Ontologies are formal and explicit representations of domain knowledge, typically expressed with linguistic terms that include concepts, concept properties, and relationships between concepts. Although linguistic terms are appropriate to distinguish event and object categories, they are inadequate when they must describe specific patterns of events, or multimedia entities. To this end, high level concepts expressed through linguistic terms and pattern specifications represented through visual, or auditory concepts, should both be organized into new extended ontologies that couple linguistic terms with visual/audio information. These new extended, or multimedia enriched ontologies, are known as multimedia ontologies.

3.1 Multimedia ontology characteristics

Multimedia ontologies are necessary because the concepts and categories defined in a traditional ontology are not rich enough to fully describe the plethora of events that can occur in multimedia objects. A multimedia ontology, informally, is a means for specifying the knowledge of the world through multimedia documents, in a structured way, such that users and applications can process the descriptions with reference to a common understanding. They model the domain of multimedia data, especially the visualizations in still images and videos in terms of low-level features and media structure descriptions. Low-level features are machine-oriented and can be automatically extracted (e.g. MPEG-7 compliant descriptors), whereas high-level semantic concepts require manual annotation of the medium or are restricted to specific domain. Ontologies define a formal language for the structuring and storage of the high-level features, facilitate the mapping of low-level to high-level features and allow the definition of relationships between pieces of multimedia information. Structure and semantics are carefully modeled to be largely consistent with existing multimedia description standards like MPEG-7.

Multimedia ontologies should be able to represent the structure of a multimedia document itself, depending on the type of document and the relations between structural elements. Multimedia ontologies need to describe and represent knowledge for either one, or even more of the following top-level hierarchical types of multimedia documents: image, video, 3D graphics, audio, audiovisual, multimedia presentation. They also need to distinguish between annotations describing the information object and these concerning the multimedia document's content. Multimedia ontologies should be capable of capturing the low-level descriptor information and to allow for basic and complex data types. They should be rich enough to describe the spatiotemporal relationships between the entities depicted. Details regarding the description of the multimedia object itself, such as the creation date, the creator, the purpose it was created for, or, even its subsequent history, should be represented in a multimedia ontology, since provenance information provides rather important metadata of the multimedia document.

Multimedia ontologies represent four different levels of information: *signal* information, *featural* information, *symbolic* information and *semantic* information. In order for someone to be able to use all the levels of information, from the semantic to the raw audiovisual one, a proper alignment framework should be provided. For example, ALIMO [12] (Alignment of Multimedia Ontologies) is an ontology alignment system that pay special care to each one of the subparts of a multimedia ontology and the attributes with the special meaning and structure. Semantic descriptions will be aligned using methods hybrid alignment systems (terminological, structural, etc.). the signal description parts will be compared by using visual matching algorithms from the field of digital image and video processing. The feature description by examining the XML schema of the MPEG-7 visual part and at last the symbolic description by referring to the definitions of the concepts that those labels are instances of, and also by examining the datatypes of the attributes assigned to those instances.

Multimedia ontologies can be of two types [31]:

- **Media-specific ontologies**. They have taxonomies of different media types and describe properties of different media. For example, video may include properties to identify length of the clip and scene breaks. They describe the format of files and related information. For an image, ontological markup may include file format and file size, plus information about how the image was produced, such as the camera that took the photo or focal length.
- **Content-specific ontologies.** They describe the subject of the resource, such as the setting or participants. Since such ontologies are not specific to the media, they could be reused by other documents that deal with the same domain. Such reuse would enhance search that was simply looking for information on a particular subject, regardless of the format of the resource. They allow an author to describe what the media is about. For a photo, this could include the date and time it was taken, where it was taken, who and what is in the picture, and what is happening. For other media, like sound, attributes like lyrics, chord progressions, or historical information may also be relevant.

Multimedia ontologies are used for content visualization, content indexing, knowledge sharing, learning and reasoning. They should be designed to serve one or more of the following purposes:

- *annotation* (e.g. summarization of multimedia content),
- *analysis* (e.g. ontology driven semantic analysis of multimedia content),
- *retrieval* (e.g. context-based retrieval and recommendations),
- *reasoning* (e.g. application of reasoning techniques to multimedia content),
- *personalized filtering* (e.g. delivery of multimedia content according to user preferences),
- *meta-modeling* (e.g. ontologies used to model multimedia processes, procedures).

3.2 Multimedia ontology construction

The construction of multimedia ontologies is difficult, because different correct specifications of the same domain or collection are possible and many decisions have to be made, which depend on the domain, the purpose of the multimedia ontology, the complexity of content and structure that characterize the multimedia objects, and the user's knowledge.

One can build a multimedia ontology simultaneously for all media. For each concept, all media specific concepts are encoded into the nodes of the ontology simultaneously. Alternatively, one can develop a separate ontology for different media and create a link between nodes for every cross reference.

Multimedia ontology construction is usually a manual, iterative process consisting of at least three steps:

1. selection of concepts to be included in the ontology,
2. establishment of properties for the concepts and relationships between concepts in the ontology,
3. maintenance of the ontology.

The ontology can be constructed using a concept-driven or a data-driven approach. The concept-driven approach does not require any data: the ontology is built from general or domain specific knowledge. In the data-driven approach, the ontology is constructed primarily from data, but domain knowledge is also used in manually constructing it. In general, however, fully automatic construction of ontologies is not possible because automatically selecting relevant concepts and relationships is hard. An alternative is to use semi-automatic ontology construction techniques, which aim at facilitating each of the steps above.

One could mention several attempts for building multimedia ontologies. In [18] multimedia ontologies are constructed manually. Text information available in videos and visual features are extracted and manually assigned to concepts, properties, or relationships in the ontology. In [1] new methods for extracting semantic knowledge from annotated images are presented. Perceptual knowledge is discovered grouping images into clusters based on their visual and text features and semantic knowledge is extracted by disambiguating the meaning of words in annotations using WordNet and image clusters. In [24] a Visual Descriptors Ontology and a Multimedia Structure Ontology, based on MPEG-7 Visual Descriptors and MPEG-7 MDS respectively, are used together with a domain ontology in order to support content annotation. Visual prototypes instances are manually linked to the domain ontology. In [9] an approach to semantic video object detection is presented. Semantic concepts for a given domain are defined in an RDF(S) ontology together with qualitative attributes (e.g. color homogeneity), low-level features (e.g. model components distribution), object spatial relations and multimedia processing methods (e.g. color clustering) and rules in F-Logic are used for detection on video objects. In [3] pictorially enriched ontologies have been introduced for the purpose of automatic video annotation. Video clips of highlights are regarded as instances of concepts in the ontology and are directly linked to the corresponding concepts, clustered into subclasses according to their perceptual similarity. Visual concepts are defined as the centers of these clusters, such that each of them is assumed to represent a specific pattern in which the concept can manifest itself.

BOEMIE [28] (Bootsrapping Ontology Evolution with Multimedia Information Extraction) advocates a synergistic approach that combines multimedia extraction and ontology evolution in a bootstrapping process involving, on the one hand, the continuous extraction of semantic information from multimedia content in order to populate and enrich the ontologies and, on the other hand, the deployment of these ontologies to enhance the robustness of the extraction system.

MOM (Multimedia Ontology Manager) [4] is a complete system, which has been developed according to the principles and concepts of pictorially enriched ontologies. It supports dynamic creation and update of multimedia ontologies, provides facilities to automatically perform annotations and create extended text (and audio) commentaries of video sequences, and allows complex queries on video databases, based on the ontology itself.

OntoMedia [17] is an ontology-based multimedia information system. Its main goal is the management of large multimedia collections using semantic integration techniques for metadata by applying state-of-the-art ontology driven Semantic Web technology to the multimedia domain.

Annotations of multimedia documents typically have been pursued in two different directions. Either previous approaches have focused on low-level descriptors, such as dominant color, or they have focused on the content dimension and corresponding annotations, such as person or vehicle. A software environment to bridge between the two directions is M-OntoMat-Annotizer [20] that allows for linking low-level MPEG-7 visual descriptions to conventional Semantic Web ontologies and annotations. It is used in order to construct ontologies that include prototypical instances of high-level domain concepts together with a formal specification of corresponding visual descriptors. Thus, it allows formalizing the interrelationship of high-level and low-level multimedia concept descriptions allowing for new kinds of multimedia content analysis, reasoning and retrieval.

A multi-ontology based multimedia annotation model is presented in [11], which ensures effective utilization of multimedia assets by a variety of users. In this model, a domain independent multimedia ontology is integrated with multiple domain ontologies in an effort to provide multiple, domain-specific views of multimedia content. Thus, access to multimedia content can better address different users' information needs. Dong and Li developed a Multimedia Ontology based on MPEG-7 multimedia content description tools, proposed a strategy to integrate multiple domain ontologies, and designed a term extraction procedure to automatically extract domain-specific ontological terms from textual resources of multimedia data.

There are many ontology languages (like OWL or WSMO family) available with different expressiveness and reasoning capabilities. The main criteria for the selection of an ontology language are its knowledge representation mechanism and the inferencing/reasoning support needed by an application. The high complexity of multimedia modeling requires a representation language with high semantic expressiveness. This fact in combination with the compliance to W3C standards makes OWL the most appropriate language for multimedia knowledge representation.

An ontology designed for multimedia applications should enable integration of the conceptual and media spaces. M-OWL [14] is a new ontology language that supports this capability. It supports explicit definition of media properties for the concepts. The language has been defined as an extension of OWL, the standard ontology language for the Web. Ghosh et al also propose a new Bayesian Network based probabilistic reasoning framework, with M-OWL for semantic interpretation of multimedia data, and a new model for ontology integration, based on the similarity of the concepts in the media domain. The framework can be used to integrate several multimedia and traditional ontologies.

TAO_XML is another suitable multimedia ontology description language. TAO (TeleAction Object) [8] is a paradigm for representing multimedia objects based on the following two elements: a hypergraph that specifies the component objects and their structural relations, and a knowledge structure which describes the environment and the actions of the object. TAOs can be described using XML, thus opening the way towards the representation of multimedis ontologies.

3.3 Existing multimedia ontologies

Existing multimedia ontologies are classified, according to their domain of application, or their framework, in the following groups:

- **Content Structure Ontologies**, that focus on the description of multimedia content structure.
- **Specific Domain Ontologies,** that have been created to serve a particular domain.
- **Multimedia Upper Ontologies**. Upper level ontologies are intended for more general use and describe higher level concepts that can be refined by domain ontologies, in order to make multimedia-handling procedures more homogeneous.
- **Multimedia Core Foundational Ontologies**. The role of core ontologies is to serve as a starting point for the construction of new ontologies, to provide a reference point for comparisons among different ontological approaches and to serve as a bridge between existing ontologies. Core ontologies are typically conceptualizations that contain specifications of domain independent concepts and relations based on formal principles derived from philosophy, mathematics, linguistics and psychology.

Table 1. Existing multimedia ontologies, grouped by their domain or framework

Content Structure	Specific Domain	Multimedia Upper	Multimedia Core
AIM@SHAPE	ImageStore	ZyX	ABC
aceMedia	MEPCO	MPEG-7	CIDOC CRM

AIM@SHAPE ontology has been developed for representing, modeling and processing knowledge, which derives from digital shapes. A digital shape is any individual object having a visual appearance which exists in a two or higher dimensional space. Pictures, sketches, images, 3D objects, videos, and animations are examples of shapes. Digital shapes are used in many different contexts, including industrial design, biomedical applications, entertainment, environmental monitoring, cultural heritage.

MEPCO [25] is an ontology developed for cross relating and linking media campaigns. It is created as a domain-specific ontology that models partially the media domain and the advertising domain. It was designed as an extension of PROTON upper-level ontology and has a twofold objective: it serves to organize a specific DB where creatives are stored, and at the same time to aid the discovery, interlinking and tracking of media campaigns thus enhancing the other modules in the system architecture.

The ZyX model is an upper ontology developed for multimedia metamodeling. It provides an ontological description of an abstract multimedia presentation model and is based on the ZyX model by Boll/Klas [5]. It describes complete or fragments of multimedia documents by means of a tree, the nodes of which are called presentation elements. Each presentation element has got a binding point associated with it, which can be bound to one variable of another presentation element, thus creating the edges of the tree. The presentation elements are the generic elements of the model. They can represent atomic media elements (e.g. videos, images and text) or operator elements which combine presentation elements with certain semantics. There are operator elements that allow for temporal synchronization, definition of interaction, adaptation, and for the spatial, audible, and visible layout (the so-called projector elements) of the document.

The initial and ongoing goal of the *ABC* model [6] is threefold:

- To provide a conceptual basis for understanding and analyzing existing metadata ontologies and instances.
- To give guidance to communities beginning to examine and develop descriptive ontologies.
- To develop a conceptual basis for automated mapping amongst metadata ontologies.

As such, the ABC ontology is not intended as a metadata vocabulary per se, but as a basic model and ontology that provides the notional basis for developing domain, role, or community specific ontologies. In this spirit, the ABC model incorporates a number of basic entities and relationships common across other metadata ontologies including time and object modification, agency, places, concepts, and tangible objects. In particular, it has been designed to model physical, digital and analogue objects held in libraries, archives, museums and on the Internet. This includes objects of all media types, such as text, image, video, audio, web pages, and multimedia. Communities wishing to build their own metadata ontologies and models may then extend the ABC entities and relationships as needed.

The *CIDOC* Conceptual Reference Model (*CRM*) [29] provides definitions and a formal structure for describing the implicit and explicit concepts and relations used in cultural heritage documentation. It is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modeling. It is proposed to be considered as part of a core ontology for multimedia objects supporting all multimedia objects and especially those concerning cultural heritage items and events.

The *MPEG-7* ontology [30] is an ontology developed for supporting knowledge-based multimedia applications so as to capture all the MPEG-7 MDS.

ImageStore [7] is an ontology developed for BioImage Database. All user interactions with the database are mediated through this ontology, which defines the metadata descriptors used for the images. The ImageStore Ontology makes a clear distinction between several semantic “worlds”:

- The Real World, in which exist the objects and events that have been partially captured by various forms of media.
- The Media World, in which exist such representations of aspects of the real world.
- The Cognitive World, in which exist man’s ideas, interpretations and conceptualizations both about objects and events in the real world and about the portrayal of these objects and events in the media world.
- The Logical World, in which exist the formalizations (i.e. ontologies) of the conceptualizations of all three worlds and of their interconnections.

Specifically, the ImageStore Ontology specifies the metadata requirements for a wide range of information, including:

- Image acquisition (including information about who took the original micrograph, where, when, under what conditions, with what equipment, for what purpose, etc., and details of any image processing undertaken).
- The media object itself (image type, dynamic range, resolution, format, codec, etc.).
- The image denotation (i.e. brute facts about the subject that has been imaged, specimen preparation method, etc.).
- The image connotation (the interpretation, meaning purpose or significance of the imaged subject, its relevance to its creator and others, and its semantic relationship to other imaged subjects).

aceMedia [26] focuses on generating value and benefits to end users, content providers, network operators, and multimedia equipment manufactures, by introducing, developing and implementing a system based on an innovative concept of knowledge assisted, adaptive multimedia content management, addressing user needs. It extends and enriches ontologies to include low level audiovisual features, descriptors and behavioral models in order to support automatic annotation; a core ontology was described based on extensions of the DOLCE [13] (Descriptive Ontology for Linguistic and Cognitive Engineering) core ontology and the multimedia-specific infrastructure components, the Visual Descriptor Ontology, based on an RDFS representation of the MPEG-7 Visual Descriptors and the Multimedia Structure Ontology based on MPEG-7 MDS. Its main aims are the support of audiovisual content analysis and object/event recognition, the creation of knowledge beyond object and scene recognition through reasoning processes and enabling user-friendly and intelligent search and retrieval.

5. CONCLUSIONS & FUTURE TRENDS

Ontologies are a very promising technology for a variety of application areas. Some are still to come: intelligent environments (contextually-appropriate personalized information spaces), personal knowledge networking (Social Semantic Desktop), and business performance management (near-real-time semantic information integration of critical business performance indicators to improve the effectiveness of business operations and to enable business innovations).

More and more application scenarios depend on the integration of information from various kinds of resources that come in different formats and are characterized by different formalization levels. In a lot of large companies, for example, in the engineering domain, information can be typically found in text documents, e-mails, graphical engineering documents, images, videos, sensor data, and so on, that is, information is stored in so-called cross-media resources. Taking this situation into account, the next generation of semantic applications has to address various challenges in order to come up with appropriate solutions for: ontology learning and metadata integration, information integration, advanced multimedia ontology mapping [10].

Pervasive computing envisions a world in which computational devices are ubiquitous in the environment and are always connected to the network. In the pervasive computing vision, computers and other network devices will seamlessly integrate into the life of users, providing them with services and information in an "always on", context sensitive fashion. Semantic technology such as ontologies can make a significant contribution by supporting scalable interoperability and context reasoning in such systems.

With the use of mobile devices and current research on ubiquitous computing, the topic of context awareness is a major issue for future IT applications. Intelligent solutions are needed to exploit context information, for example, to cope with the fuzziness of context information and rapidly changing environments and unsteady information sources. Since multimedia ontologies encode a view of a given domain that is common to a set of individuals or groups in certain settings for specific purposes, the mechanisms to tailor multimedia ontologies to the need of a particular user in his working context are required. The efficient dealing with a user's context posts several research challenges: formal representation of context, context reasoning, context mapping.

Developing methods and tools that are able to meet challenges as consistency checking of multimedia ontologies, evolution of multimedia ontologies and metadata and reasoning, is an essential requirement to devise an ontology and metadata infrastructure that is powerful enough to support the realization of applications that are characterized by an open, decentralized, and ever changing environment. What is needed is a formal model of networked ontologies that supports their evolution and provides the basis for guaranteeing their (partial) consistency in case one of the networked ontologies is changing.

6. REFERENCES

[1] Benitez, A., and Chang S. Automatic multimedia knowledge discovery, summarization and evaluation. *IEEE Transactions on Multimedia, Submitted,* 2003.

[2] Berners-Lee, T., Hendler, J., and Lassila, O. *The Semantic Web*. Scientific American. May, 2001.

[3] Bertini, N., Cuchiara, R., Del Bimbo, A., and Torniai, C. Video annotation with pictorially enriched ontologies. In *Proceedings of IEEE Int'l Conference on Multimedia & Expo*, 2005.

[4] Bertini, M., Del Bimbo, A., Torniai, C., Cucchiara, R., and Grana, C. MOM: Multimedia Ontology Manager. A Framework for Automatic Annotation and Semantic Retrieval of Video Sequences. MM'06, (Santa Barbara, California, October 23-27), USA, 2006.

[5] Boll, S., Klas, W., and Westermann, U. *Multimedia Document Formats-Sealed Fate or Setting Out for New Shores?* In Multimedia-Tools and Applications. 11(3), 267-279, 2000.

[6] Brickley, D., Hunter, J., and Lagoze, C. *ABC: A Logical Model for Metadata Interoperability*. Harmony Project Working Paper, 1999.

[7] Catton, C., Sparks, S., and Shotton, D. *The ImageStore Ontology and the BioImage Database: Semantic Web Tools for Biological Research Images.* Bioimage, 2006.

[8] Chang, H., Chang, S., Hou, T., and Hsu, A. The management and the applications of Tele-Action Objects. *ACM Journal of Multimedia Systems*, 3(5-6):204-216, 1995.

[9] Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V., and Strintzis, M. Knowledge-assisted semantic video object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(10): 1210-1224, Oct, 2005.

[10] Davies, J., Studer, R., and Warren P. *Semantic Web Technologies: trends and research in ontology-based systems*. WILEY, 2006.

[11] Dong, A., and Li, H. Multi-ontology Based Multimedia Annotation for Domain-specific Information Retrieval. *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing.* Vol.2, 158-165, 2006.

[12] Euzenat et al. D2.2.3: *State of the art on ontology alignment*. KWEB EU-IST-2004-507482.

[13] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. Sweetening Ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Proceedings of the 13th International Conference on Knowledge Acquisition, Modeling and Management*, EKAW, Vol. 2473 of Lecture Notes in Computer Science, Siguenza, Spain, 2002.

[14] Ghosh, H., Chaudhury, S., Kashyap, K., and Maiti, B. *Ontologies. A Handbook of Principles, Concepts and Applications in Information Systems.* Springer US, 2007, 265-296.

[15] Gruber, T. A translation approach to portable ontologies. *Knowledge Acquisition* 5(2): 199-220, 1993.

[16] Guarino, N. *Formal Ontology and Information Systems*. IOS Press, 3-15, 1998.

[17] Husemann, B., and Vossen, G. *OntoMedia – Semantic Multimedia Metadata Integration and Organization*, 2005.

[18] Jaimes, A., and Smith, J. Semi-automatic, data-driven construction of multimedia ontologies. In *Proceedings of IEEE Int'l Conference on Multimedia & Expo*, 2003.

[19] Pereira, F., and Van de Walle, R. *Multimedia Content Description in MPEG-7 and MPEG-21*. In Multimedia Content and the Semantic Web. Ed. Stamou, G, and Kollias, S., WILEY, 2005.

[20] Petridis, K., Anastasopoulos, D., Saathoff, C., Timmermann, N., Kompatsiaris, I., and Staab, S. M-OntoMat-Annotizer: Image Annotation. Linking Ontologies and Multimedia Low-Level Features. *Engineering Applications of Semantic Web Session (SWEA) at the 10th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES)*, Bournemouth, U.K., 2006.

[21] Sheth, A. *Semantic Metadata for Enterprise Information Integration*. DM Review. July, 2003.

[22] Staab, S., and Studer, R. *Handbook on Ontologies*. International Handbooks on Information Systems, Springer Verlag, Heidelberg, 2004.

[23] Stamou, G., and Kollias, S. *Multimedia Content and the Semantic Web. Standards, Methods and Tools*. WILEY, 2005.

[24] Strintzis, J., Bloehdom, S., Handschuh, S., Staab, S., Simou, N., Tzouvatras, V., Petridis, K., Kompatsiaris, I., and Avrithis, Y. Knowledge representation for semantic multimedia content analysis and reasoning. In *European Workshop on the Integration of Knowledge, Semantics and Digital Media technology*, Nov. 2004.

[25] Todorova, P., and Popov, B. *MEPCO Design*. Delivarable 3.2, 2006.

[26] http://www.acemedia.org

[27] http:www.boemie.org

[28] http://cidoc.ics.forth.gr

[29] http://rhizomik.net/mpeg7ontos

[30] http://www.altova.com/products/semanticworks

[31] http://www.w3.org/TR/webonto_req