Director of Editorial Content: Kristin Klinger
Senior Managing Editor: Jennifer Neidig
Managing Editor: Jamie Snavely
Assistant Managing Editor: Carole Coulson
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com



Library of Congress Cataloging-in-Publication Data

Encyclopedia of multimedia technology and networking / Margherita Pagani, editor. -- 2nd ed.
p. cm.
Includes bibliographical references and index.
Summary: "This publication offers a research compendium of human knowledge related to the emerging multimedia digital metamarket"--Provided by publisher.
ISBN 978-1-60566-014-1 (hardcover) -- ISBN 978-1-60566-015-8 (ebook)
1. Multimedia communications--Encyclopedias. I. Pagani, Margherita, 1971-
TK5105.15.E46 2009
621.38203--dc22
2008030766

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

# Ontology and Multimedia

**Roberto Poli**
*University of Trento, Italy*

**Achilles Kameas**
*Hellenic Open University, Italy*

**Lambrini Seremeti**
*Hellenic Open University, Italy*

## INTRODUCTION

Audiovisual resources in the form of still pictures, graphical, 3D models, audio, speech, and video play an increasing pervasive role in our lives, and there will be a growing need to manage all these multimedia objects. This is a task of increasing importance for users who need to archive, organize, and search their multimedia collections in an appropriate fashion.

To cope with this situation, much effort has been put in developing standards both for multimedia data (natural and synthetic (e.g., photography, face animation), continuous and static (e.g., video, image)) and for data describing multimedia content (metadata). The aim is to describe open multimedia frameworks and achieve a reasonable and interoperable use of multimedia data in a distributed environment.

## BACKGROUND

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 bridge 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" (Smeulders, Worring, Santini, Gupta, & Jain, 2000).

Because of the high cost and subjectivity associated with human-generated metadata, a large number of research initiatives are focusing on technologies to enable automatic classification and segmentation of digital resources. Many consortia are working on a number of projects in order to define multimedia metadata standards, which are being developed in order to describe multimedia contents in many different domains and to support sharing, exchanging, and interoperability across different networks. They are distinguished in Salvetti, Pieri, & Di Bono, 2004):

- *Standardised description schemes* that are directly related to the representation of multimedia content for a specific domain (like METS, MPEG-7).
- *Standardised metadata frameworks* that consider the possibility of integrating more metadata standards mapped on different application domains, providing rich metadata models for media descriptions together with languages allowing one to define other description schemes for arbitrary domains (like PICS, RDF, MPEG-21).

For example, the vision of MPEG-21 is to define a multimedia framework to enable augmented and transparent use of multimedia resources across a wide range of networks and devices used by different communities. The intent is that this framework will cover the entire multimedia content delivery chain, including creation, production, delivery, personalization, presentation, and trade.

The development of metadata standards will increase the value of multimedia data, which are used by various applications. Nevertheless, there are disadvantages in current metadata representation schemes (Smith & Schirling, 2006). Some of them are cost, unreliability, subjectivity, lack of authentication, and interoperability with respect to syntax, semantics, vocabularies, and languages (Salvetti et al., 2004).

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 toward 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 (Staab & Studer, 2004).

Professional groups increasingly are building metadata vocabularies (or ontologies). A number of research and standards groups are working on the development of common conceptual models (or upper ontologies) to facilitate interoperability between metadata vocabularies and the integration of information from different domains.

## MAIN FOCUS OF THE ARTICLE

### Multimedia Ontologies

Ontologies have applications in many areas, including natural language translation, medicine, standardization of product knowledge, electronic commerce, and geographic information systems, among others. Many of these applications use or will use multimedia data in the immediate future, making the creation of multimedia ontologies a crucial component (Alejandro & Smith, 2003).

It is well known that the word "ontology" generates a lot of controversy in discussions about Artificial Intelligence, although it has a long history, in which it refers to the categorical framing of what is (Poli, 2001, 2002, 2007; Poli & Simons, 1996). Briefly it can be claimed that ontology deals with what can be rationally understood, at least partially. According to this interpretation, science in all of its branches is the most successful and powerful ally of ontology.

One may say that there are material things, plants and animals, as well as the products of the talents and activities of animals and humans in the world. This first almost trivial list already indicates that the world comprises not only things, animate or inanimate, but also activities and processes and the products that derive from them. It is likewise difficult to deny that there are thoughts, sensations, and decisions, and the entire spectrum of mental activities, just as one is compelled to admit that there are laws and rules, languages, societies, and customs. We can set about organizing this list of objects by saying that there are independent items that may be real (mountains, flowers, animals, and tables), or ideal (sets, propositions, values), and dependent items which in turn may be real (colors, kisses, handshakes, and falls) or ideal (formal properties and relations). All these are in various respects items that are.

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.

Four different levels of information are represented in multimedia ontologies (Euzenat et al., 2004):

- Signal information
- Featural information
- Symbolic information
- Semantic information

Multimedia ontologies can be of two types:

- Media-specific ontologies have taxonomies of different media types and describe properties of different media. For example, video may include properties to identify the length of the clip and scene breaks.
- Content-specific ontologies describe the subject of the resource, such as the setting or participants. Because such ontologies are not specific to the media, they could be reused by other documents that deal with the same domain.

Multimedia ontologies are used for different goals by different applications, including the following (Alejandro & Smith, 2003):

- **Content visualization:** They can be used to create tables of content and used for browsing.
- **Content indexing:** They can be used to improve indexing consistency in manual annotation systems (Schreiber et al., 2001) (e.g., use the term apartment instead of flat), or in the propagation of labels in automatic indexing systems (e.g., a face detected implies a person was detected).
- **Knowledge sharing:** Annotated multimedia collections can be more easily shared if they use a common conceptual representation.

- **Learning:** Collections annotated by different individuals using common ontologies lead to annotation consistency which is of extreme importance in applying approaches based on learning techniques that use annotated collections for training.
- **Reasoning:** Information not explicit in the data may be obtained automatically with the help of ontology.

## Creating a Multimedia Ontology

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 characterizes 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 (essentially the ontology is built in one cycle). Alternatively, one can develop a separate ontology for different media and create a link between nodes for every cross reference.

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; and(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 semiautomatic ontology construction techniques, which aim at facilitating each of the steps above.

In the data-driven approach, for example, a semiautomatic construction of a multimedia ontology uses a video collection, and models the concepts, their properties, and their relationships. For each of the videos one applies the following steps: (1) automatic scene cut detection; (2) automatic speech recognition; (3) parsing of metadata; and (4) automatic concept analysis. Preprocessing consists of steps one through three, after which an ontology can be constructed based on textual content alone. Step four forms the basis for manually adding multimedia components to the ontology (Smith, Tseng, & Jaimes, 2003).

When building multimedia ontologies, one should address the requirements listed below:

1. Multimedia ontologies should be designed to serve one or more of these purposes, accurately and adequately:
   - *Annotation* (e.g., summarization of multimedia content)
   - *Analysis* (e.g., ontology driven semantic analysis of multimedia content, etc.)
   - *Retrieval* (e.g., context-bases 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, etc.)
2. 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 (e.g., semantic concepts depicted in an image).

3. 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.*

Ontologically speaking, possibly the most striking aspect of semantic visual information is its multilayer structure. To provide but a feeling of the multilayered stratification of visual object, one may consider the case of a portrait, where one may distinguish the following levels:

- The three-dimensional space in which the subject of the portrait and some elements of the setting appear;
- The movement of the subject's apparent corporeality;
- The subject's character;
- His or her individual idea, or the idea that the person portrayed has of him/herself; and
- The symbolic, or the universal content manifested by the portrait (Hartmann, 1950; Ingarden, 1962; Poli, 1998).

A *hierarchical structure of multimedia segments* is thus needed in order to capture all possible types of media decompositions and relations. The regions that correspond to semantic objects need to be described in terms of their *location* within the multimedia content.

Ontology should be capable of capturing the low-level descriptor information (in order to semantically express this kind of characteristics associated with a concept, several different audiovisual attributes (e.g., color, shape, texture, motion, localization, etc.) need to be represented, depending on the concept) and to allow for basic and complex data types. Hence, alignment with standards (MPEG-7, TV-Anytime, MPEG-21) is a safe approach in order for the ontology to be of practical interest.

4. Ontology should be rich enough to describe the spatiotemporal relationships between the entities depicted.
   - *Spatial* relations are needed to describe how segments are placed and relate to each other in 2D space (e.g., right and above).
   - *Topological* spatial relations are needed to describe how the spatial boundaries of the segments relate (e.g., touches and overlaps).
   - *Temporal* relationships among segments or events, providing information about the sequence in time, need to be represented, especially when the multimedia object is of type video.
5. 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, and so forth, should be represented in ontology, because *provenance information* provides rather important metadata of the multimedia document. A multimedia ontology framework should *support annotations* produced by different annotators (humans or not), thus *enabling their linking and further processing*.
6. Due to the large number and wide variety of Web access devices, it has become quite essential for a Web service to *know the required device capabilities* in order to support playback of a media object and other characteristics of the delivery context that influence their presentation.
7. Multimedia ontologies need to be *open* in order to be widely-adopted by multiple and heterogeneous communities.

## Existing Multimedia Ontologies

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. Existing multimedia ontologies are divided in groups, according to their domain of application or their framework (Eleftherohorinou, Zervaki, Gounaris, Papastsiaris, & Hobson, 2006). Table 1 shows a list of existing multimedia ontologies, grouped according to their domain or framework; the groups of ontologies are:

- **Content structure ontologies:** They focus on the description of multimedia content structure.
- **Specific domain ontologies**: they 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

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

| Content Structure | Specific Domain | Multimedia Upper Ontology | Multimedia Core Ontology |
|---|---|---|---|
| aceMedia Framework | Medical Image Domain | SWeMPs | CIDOC CRM |
| AIM@SHAPE | NM2 | ZyX | ABC |
| Music Information | MEPCO | Salero | D OLCE |
| INA | ImageStore | MPEG-7 | WordNet |
| | Soccer Domain | Semantic User Preference | |
| | Formula 1 Domain | X3D | |
| | | Modality | |

specifications of domain independent concepts and relations based on formal principles derived from philosophy, mathematics, linguistics, and psychology.

An increasing number of multimedia ontologies are being used in various projects for multimedia processing; examples are given below.

The main goal of **MEPCO** (Kienast, Zeiner, Hofmair, Schlatte, Thallinger, Burger et al., 2006) is the cross-relation of media campaigns over the media TV, press, and Internet and furthermore the ambitious goal to cross link media campaigns over different countries. What makes a media campaign unique from others is not completely straightforward; however, there are rules that a human uses to determine whether a media campaign is new. These heuristic rules will be formally encoded as to describe media campaigns in a generic way. The MediaCampaign Ontology (MEPCO) will be based on the upper-level ontology PROTON and will be aligned to media-related metadata standards, such as NewsML and News Codes from IPTC.

**ACEMedia** 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 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.

The **ZyX** ontology provides an ontological description of an abstract multimedia presentation model and is based on the ZyX model by Boll, Klas, and Westermann (2000). The ZyX model describes complete or fragments of multimedia documents by the means of a tree. The nodes of the tree are called presentation elements. Each presentation element has got a binding point associated with it. Such a binding point 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.

## FUTURE TRENDS

Development of multimedia ontologies is still quite an empirical process. Due to that reason research communities involved with ontology-driven analysis, while sharing the same knowledge do not share a mutual consensus. Harmonization approaches, therefore, need to be followed by all these ontology-driven based systems and applications, in order to pave the way toward the development of an integrated knowledge infrastructure. Achieving ontology harmonization may be difficult and complex in practice. The most significant factors are the efficient modularisation of ontologies, the easy linking to other ontologies, and the specification of a minimum set of ontologies to be used for arbitrary applications.

## CONCLUSION

Multimedia ontologies enable the inclusion and exchange of multimedia content through a common understanding of the multimedia content description and semantic information. They model the domain of multimedia data, in terms of low-level features and media structure descriptions, thus increasing sharing capabilities of multimedia objects. Additionally, such an ontology could also be used as a translator, to integrate multiple, heterogeneous data sources. With the aid of multimedia ontologies the vision of querying and retrieving multimedia content from distributed databases has started to become more feasible.

## REFERENCES

Alejandro, J., & Smith, J.R. (2003). Semi-automatic, datadriven construction of multimedia ontologies. In *Proceedings of the IEEE International Conference On Multimedia and Expo,* Baltimore.

Boll, S., Klas, W., & Westermann, U. (2000). Multimedia document formats: Sealed fate or setting out for new shores? *Multimedia-Tools and Applications, 11*(3), 267-279.

Eleftherohorinou, H., Zervaki, V., Gounaris, A., Papastathis, V., Kompatsiaris, Y., & Hobson P. (2006). *Towards a common multimedia ontology framework*.

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

Hartmann, N. (1950). *Aesthetik*, de Gruyter, Berlin.

Ingarden, R. (1962). *Untersuchungen zur Ontologie der Kunst: Musikwerk. Bild*, Architektur. Film, Niemayer, Tübingen 1962.

Kienast, G., Zeiner, H., Hofmair, P., Schlatte, R., Thallinger, G., Burger, T. et al. (2006). *Representation techniques for multimedia objects*. SALERO Deliverable 3.1.1. IST FP6-027122.

Kosch, H. (2004). *Distributed multimedia database technologies supported by MPEG-7 and MPEG-21*. Boca Raton: CRC Press.

Poli, R. (2001). The basic problem of the theory of levels of reality. *Axiomathes, 12*(3-4), 261-283.

Poli, R. (2002). Ontological methodology. *International Journal of Human-Computer Studies, 56*, 639-664.

Poli, R. (2007). Ontology. The categorical stance. In R. Poli, J. Seibt, & J. Symons (Eds.), *Theory and applications of ontology (provisional title). Vol 1: The philosophical stance.* Dordrecht: Springer-Verlag.

Poli, R., & Simons, P. (Eds.). (1996). *Formal ontology*. Dordrecht: Kluwer.

Salvetti, O., Pieri, G., & Di Bono, M.G. (2004). WP9: *A review of data and metadata standards and techniques for representation of multimedia content*. MUSCLE. Network of Excellence FP6-5077-52.

Schreiber, A.T., et al. (2001). Ontology-based photo annotation. *IEEE Intelligent Systems*.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell*, *22,* 1349-1380.

Smith, J.R., & Schirling, P. (2006). Metadata standards roundup. *IEEE Multimedia, 13*, 84-88.

Smith, J.R., Tseng, B.L., & Jaimes, A. (2003, July 24-25). Modal keywords, ontologies, and reasoning for video understanding. In *Proceedings of the International Conference On Image and Video Retrieval*, Urbana, IL, USA.

Staab, S., & Studer, R. (2004). Handbook on ontologies. *International handbooks on information systems.* Heidelberg: Springer-Verlag.

## KEY TERMS

**Artificial Intelligence:** AI is a branch of computer science that deals with intelligent behaviour, learning, and adaptation in machines.

**Content Structure Ontologies:** Ontologies that focus on the description of multimedia content structure. They should be capable of capturing the low-level descriptor information, represent several different audiovisual attributes (e.g., color, shape, texture, motion, localization, etc.) depending on the concept, and allow for basic and complex data types.

**Metadata:** They are "data about other data;" they are data segments that describe structural, behavioural, or functional aspects of other data segments. Multimedia 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.

**Multimedia Core Foundational Ontologies:** They are conceptualizations that contain specifications of domain independent concepts and relations based on formal principles derived from philosophy, mathematics, linguistics, and psychology. They are used as a starting point for the construction of new ontologies or as a bridge between existing ontologies.

**Multimedia Ontology**: In a multimedia ontology concepts might be represented by multimedia entities (images, graphics, video, audio, segments, etc.) or terms. A multimedia ontology is a model of multimedia data, especially of sounds, still images and videos, in terms of low-level features and media structure. Multimedia ontologies enable the inclusion and exchange of multimedia content through a common understanding of the multimedia content description and semantic information.

**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.

**Ontology:** An ontology is a formal, explicit specification of a domain. It deals with what can be rationally understood, at least partially. Typically, an ontology consists of concepts, concept properties, and relationships between concepts. In a typical ontology, concepts are represented by terms.

**Specific Domain Ontologies:** They have been created to serve a particular domain; they consist of terms that represent concepts particular to that domain using constructs of content structure.