Towards a Multimedia Ontology System: an Approach Using TAO_XML *

Massimiliano Albanese, Paolo Maresca, Antonio Picariello and Antonio Maria Rinaldi
Dipartimento di Informatica e Sistemistica
Università di Napoli "Federico II"
Napoli, Italy
E-mail: {malbanes,paomares,picus,amrinald}@unina.it

Abstract

Archiving, organizing, and searching multimedia data in an appropriate fashion is a task of increasing importance. The ontology theory may be appropriately extended in order to face with this challenging issue. In this paper we propose a novel multimedia ontology theory. We first describe the multimedia ontology concepts and then we adopt TAO_XML *as a suitable ontology description language. Eventually, we propose a general architecture for supporting creation and management of multimedia objects.*

1. Introduction

The rapid evolution of digital technology is producing a tremendous amounts of digital data such as images, music, movies, and other types of media. Thus, the management of multimedia objects is a task of increasing importance for users who need to archive, organize, and search their multimedia collections in an appropriate fashion.

At the present, users typically arrange their multimedia collections into file systems which provide poor naming mechanisms and hierarchical directory structures for organization and searching. In particular, this approach has the following drawbacks:

- the categorization depends on the used classification hierarchies;
- the logical organization strictly depends on the physical storage system;
- identification based on file names alone is often not globally consistent (e.g., duplicates are possible);
- the semantic content of multimedia objects is difficult to represent and manage.

Semantics of digital multimedia materials are very hard to capture either by manual or automatic way: these semantics may be viewed as the set of terms created or linked in the practice, which forms the multimedia ontology of the discourse. Till now, knowledge management has been accomplished by the use of *ontologies*, having their primary area of application in the field of knowledge engineering. An ontology is thus a *terminological abstraction of the real world* [3]. By the way, it's the author opinion that an ontology based representation of multimedia information may greatly improve multimedia systems, in order to structure content and support retrieval.

In this paper we propose a novel multimedia ontology theory. We first describe the multimedia ontology concepts and then we use TAO_XML as a suitable ontology description language. We also propose a general architecture for supporting creation and management of multimedia objects.

The remainder of the paper is organized as follows. Section 2 provides a motivating example that will be used throughout the entire paper, while section 3 discusses related works. Multimedia ontology concepts are introduced in section 4. Section 5 provides a background of the TAO paradigm and language, while section 6 describes the proposed multimedia ontology system. Eventually conclusions are reported in section 7.

2. Motivating Example

Providing content-based information is an important activity for a number of applications. In the world wide web domain, for example, search engines may be greatly enhanced if they can use a conceptualization of the managed data. By the way, we notice that no information about the multimedia content is considered at all, simply because the ontology definition does not allow any kind of *multimedia content information*.

*This work has been carried out partially under the financial support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) in the framework of the FIRB Project "Middleware for advanced services over large-scale, wired-wireless distributed systems (WEB-MINDS)"

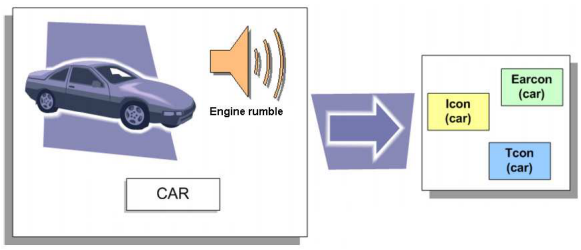

Figure 1. Motivating example

Let us consider, for example, figure 1. From a psychological point of view, humans associate the concept of *car* to either the picture of a car, or the word *car*, or the engine rumble.

Anyway, how to model and represent multimedia data in an ontology system is not a trivial task, especially for the lack of a model that can describe this kind of ontology. Our vision is that of providing a general framework which takes into account not only the textual information, but also the multimedia content of documents.

Let us consider a user who wants to retrieve all the information related to the famous modern painter *Pablo Picasso*. When she submits the query to the search engine, the use of an ontology may surely enhance the retrieval process, thus collecting all the information related to the painter, as the artistic field, the information about the subjects represented in his painting and so on.

In the following of the paper, we will define a model for multimedia ontologies and show a specific implementation using the TAO_XML environment. Figure 2 shows an example of using the TAO_XML representation for the concepts related to *Picasso* (see figure 4 for details).

3. Related Works

Multimedia data description and presentation is a hot topic in the research community. In the last years several systems have been presented to provide formal models and languages to address the issues related to the complex nature of multimedia objects. In [2] the authors present Flavor, a formal language for audio visual object representation. The system uses an innovative description, called Flavor, to generate C++ and Java code for describing, processing and producing bit streams according to a specific syntax. The system also provides a framework extension, called XFlavor, in order to offer XML features for media representation.

AROM [13] is an object based knowledge representation system which, together with V-STORM [7] supplies a general framework to manage and describe multimedia data. This paper presents an AROM knowledge base called AVS

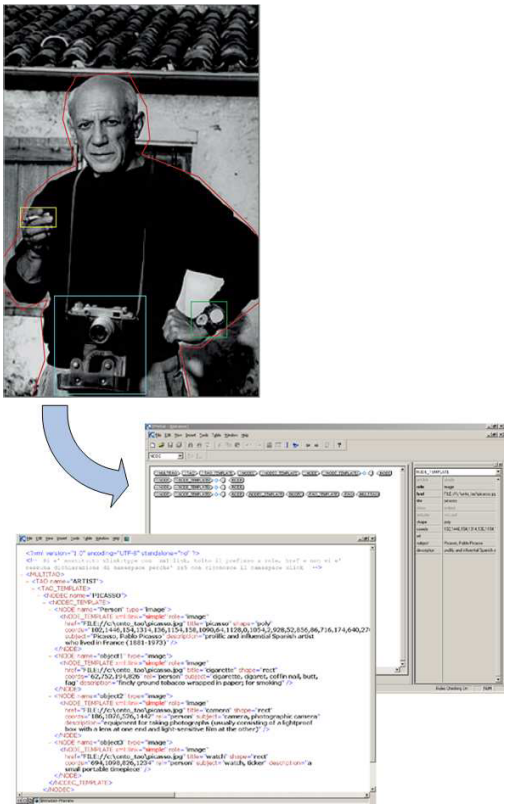

Figure 2. A portrait of Picasso and its representation using TAO_XML

[5] which represents a generic model for multimedia presentation, using the SMIL standard [14]. The AVS model is described through a graphical notation using UML. Authors of [12] propose a novel framework for describing and indexing multimedia data using a statistical approach. A4SM [11] is a framework for automatic and semi-automatic annotation of multimedia data. The authors use an RDF schema for describing relation based semantics. Different description frameworks as Dublin Core, XML, RDF and so on, are analyzed in [4], where the authors propose a Multimedia Description Framework (MDF), designed to give a unified view of multiple description schemas.

At the best of our knowledge there is no formal definition of what a multimedia ontology should be. In this paper we propose a formalization of a multimedia ontology and propose a language for describing its structure and content.

4. Multimedia Ontology Concepts

It is well known that the word “ontology” generates a lot of controversy in discussions about AI, although it has a long history in philosophy, in which it refers to the sub-

ject of existence. It is also often confused with epistemology, which is about knowledge and knowing. We adopt the Gruber idea [3], who argues that the term ontology means a specification of a conceptualization: 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.

In this paper we make a first attempt towards the definition of a Multimedia Ontology. The consideration that are at the basis of our idea is that each multimedia object evokes one or more concepts, as any word of a vocabulary does.

In this sense, we can informally define 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.

Example 1 *Consider the picture in figure 2: it recalls the concept of 'Pablo Picasso', the famous Spanish painter who depicted a lot of wonderful paintings among which the famous 'Guernica'. The same concept is evoked by the word 'Picasso'.*

We can draw the following considerations from the above example: the first step towards the definition of a Multimedia Ontology requires to extend the concepts of 'word' and 'dictionary'. Starting from this consideration let us now introduce some preliminary and fundamental definitions.

Definition 1 (MM-Alphabet) *A* MultiMedia Alphabet *is a finite set of* MM-Symbols*, where each MM-Symbol is an alphanumerical character, a pixel or an audio sample.*

$$\text{MM-Alphabet} = \{\text{MM-Symbol}\} \tag{1}$$

Two MM-Symbols are said to be homogeneous if they are of the same type.

Definition 2 (MM-Word) *Given an alphabet $\mathcal{A}$, a* Multi-Media Word *of length k over $\mathcal{A}$ is a composition of k homogeneous MM-Symbols from $\mathcal{A}$.*

$$\text{MM-Word}_{\mathcal{A}} = \langle s_1, ..., s_k \rangle,\ s_i \in \mathcal{A}\ \forall\, i \in [1, k] \tag{2}$$

A MultiMedia Word *is said to be* composite *if it can be decomposed into meaningful MultiMedia Words,* atomic *if it cannot be further decomposed.*

Definition 3 (MM-Dictionary) *Given an alphabet $\mathcal{A}$, a* MultiMedia Dictionary *over $\mathcal{A}$ is a set of MM-Words over the alphabet $\mathcal{A}$.*

$$\text{MM-Dictionary}_{\mathcal{A}} = \{\text{MM-Word}_{\mathcal{A}}\} \tag{3}$$

It's worth noticing that the concept of decomposition of MM-Words is different for different kinds of media. In the case of images a component MM-Word is a subregion of the whole picture; in the case of videos a component MM-Word is a subsequence of frames; in the case of text a component MM-Word is a word.

Example 2 *The picture in figure 2 is a composite MM-Word. Clearly the main subject is Picasso, but we can recognize several subregions: the camera, the cigarette and the watches. Each component MM-Word recalls a different concept and contributes to determine the overall meaning of the composite MM-Word.*

Definition 4 (MM-Document) *A* MultiMedia Document *is a composition of MM-Words through a set $\mathcal{R}$ of relations that represents the logical structure of the documents.*

$$\text{MM-Document}_{\mathcal{A}} = (\{\text{MM-Word}_{\mathcal{A}}\}, \mathcal{R}) \tag{4}$$

As a particular case, we notice that, if w is a MM-Word, then $(\{w\}, \emptyset)$ is still a MM-Document.

Definition 5 (Extended MM-Dictionary) *Given a Multi-Media Dictionary $\mathcal{W}$, an* Extended MultiMedia Dictionary *$\mathcal{D}$ over $\mathcal{W}$ is a set of MM-Documents composed of MM-Words in $\mathcal{W}$.*

In the following we will be using the terms *multimedia object* to refer to both MM-Words and MM-Documents.

Definition 6 (Concept) *We define a* Concept *C as a pair $(\mathbb{D}, \mathbb{R})$, where $\mathbb{D}$ is a domain and $\mathbb{R}$ a set of relations between the elements in $\mathbb{D}$. Let $\mathcal{C}$ denote the set of all concepts.*

The elements in $\mathbb{D}$ can be thought as elementary concepts that allow to define a more complex concept. Figure 3 shows an example of such a concept, derived from ConceptNet [6], a freely available commonsense knowledge base and natural-language-processing toolkit which supports many practical textual-reasoning tasks over real-world documents. The concept of *car* is defined through a few simpler concepts connected by several kinds of relations.

We can now define how to map object from the Extended Multimedia Dictionary into concepts.

Definition 7 (Mapping Function) *We define a* Mapping Function *ρ as a function that relates a multimedia object to a specific concept.*

$$\rho : d \in \mathcal{D} \rightarrow C \in \mathcal{C} \tag{5}$$

We say that a mapping function is complete iff

$$\forall d \in \mathcal{D}\ \rho(d) \neq null \tag{6}$$

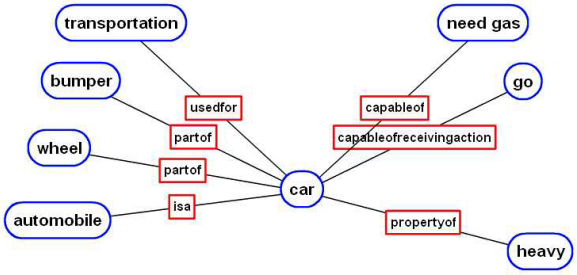

Figure 3. The concept of car from ConceptNet

We say that a mapping function is partial iff

$$\exists d \in \mathcal{D} \mid \rho(d) = null \tag{7}$$

Different mapping functions can assign different meanings to the same multimedia object. In fact each multimedia object, depending on the context, may represent different concepts. The inverse function of ρ returns all the objects that represent a given concept.

$$\rho^{-1} : C \in \mathcal{C} \rightarrow \{d_i\} \in \mathcal{D}^n \tag{8}$$

Given the definition of mapping function we can define the synonymy of multimedia objects.

Definition 8 (Synonymy) *Given a mapping function ρ and two MM-Documents $d_1, d_2 \in \mathcal{D}$, d_1 and d_2 are synonyms w.r.t. ρ iff*

$$\rho(d_1) = \rho(d_2) \tag{9}$$

We remark that the synonymy property holds between terms from the dictionary. In other words two object are synonyms w.r.t. a specific mapping function if they represent the same concept in the domain of that function.

Example 3 *Figure 3 shows an example of different MM-Words – a picture, a sound and a written word – that express the concept of car.*

We can finally give a formal definition of Multimedia Ontology.

Definition 9 (Multimedia Ontology) *We define a* Multimedia Ontology *MO as*

$$MO = (\mathcal{D}, \mathcal{C}, \mathcal{F}) \tag{10}$$

where $\mathcal{D}$ is an extend multimedia dictionary, $\mathcal{C}$ is a set of concepts and $\mathcal{F}$ a family of mapping functions.

Definition 10 (Domain Multimedia Ontology) *We define a* Domain Multimedia Ontology MO_{dom} *as*

$$MO_{dom} = (\mathcal{D}_{dom}, \mathcal{C}_{dom}, \rho_{dom}) \tag{11}$$

where ρ_{dom} is a partial mapping function, $\mathcal{C}_{dom} \subset \mathcal{C}$ is the codomain of ρ_{dom} and $\mathcal{D}_{dom}$ is the subset of all objects $d \in \mathcal{D}$ such that $\rho_{dom}(d)$ is not null.

Example 4 *Figure 2 shows a picture and its* TAO_XML *description. In this picture we notice the presence of several objects which are well represented through our extension of* TAO_XML *. Some elements have been added to the language in order to i) manage spatial and temporal information related, as an example, to subregion of an image or time intervals in an audio file; ii) semantic relations between objects.*

5. TeleAction Objects for Multimedia Ontologies

The second major contribution of this paper is the definition of a model for describing the structure and the content of multimedia ontologies using a suitable language. In this section we introduce the TAO_XML language and show how it can fit the multimedia ontology definitions. In particular we show how the ontology concepts defined so far can be mapped into the language elements.

TAO (TeleAction Object) [1] 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. In this section, we will introduce TAO and show how it can be described using XML, thus opening the way towards the the representation of Multimedia Ontologies.

5.1. Theoretical Background: TeleAction Objects

TeleAction Objects (TAOs) are multimedia objects with an associated hypergraph representing both the multimedia object and the knowledge structure. The knowledge structure allows the TAO to automatically react to certain events. A TAO can be divided into two parts: a hypergraph $\mathcal{G}(\mathcal{N}, \mathcal{L})$ and a knowledge $\mathcal{K}$, where $\mathcal{N}$ is a set of nodes and $\mathcal{L}$ is a set of links. There are two types of nodes: base and composite nodes. Each node represents a TAO and each link represents a relation among TAOs. There are the following link types: (i) attachment, (ii) annotation, (iii) reference, (iv) location and (v) synchronization. Base and composite nodes are called *bundled* when they are grouped, thus defining a single entity. With respect to the formal definitions introduced in section 4, base and composite TAO nodes respectively correspond to atomic and composite MM-Words, while a TAO corresponds to a MM-Document. A whole multimedia system is defined by the MULTITAO element, the root element of the TAO_XML document. A MULTITAO consists of one or more TAOs.

The knowledge structure $\mathcal{K}$ of a TAO is organized into four levels: System Knowledge, Environment Knowledge, Template Knowledge, and Private Knowledge. The knowledge is structured as an active index (IX), which is a set of index cells (IC) from an index cell base (ICB). The index cells define the reactions of the TAO to events filtered by the system. An index cell accepts input messages, performs some action, and sends output messages to a group of ICs. The messages sent will depend on the state of the IC and on the input messages. An IC may be seen as a kind of finite-state machine.

5.2. TAO_XML

The need to represent TAOs through an XML-based language has led to the introduction of TAO_XML [8, 9]. In general, multimedia systems may be viewed as consisting of a set of connected and interacting elementary TAOs. Each TAO is obtained by constructing an hypergraph whose nodes are attached to the index cells which provide the knowledge necessary to the system to react to external events. The hypergraph contains base and composite nodes which are connected via links that describe the relations between the components nodes of the TAO. TAO_XML links are classified as structural, temporal and spatial and correspond to the location, synchronization and annotation and reference TAO links. The attachment link is not used in TAO_XML since the attachment relation is implicitly described by the structure of the document.

5.3. TAO ontologies

We have already seen as the concepts introduced in section 4 can be strictly mapped into the elements of the TAO paradigm. The TAO_XML is thus highly suitable for describing multimedia ontologies. In particular we have extended the language to address some specific issues, adding elements useful to manage spatial and temporal information related, as an example, to subregions of an image or time intervals in an audio file. We have also introduced elements useful to manage semantics of the objects and semantic relations among them. To this aim we use WordNet [10] as a uniform way for representing concepts. Figure 4 shows an example of TAO_XML document, describing the picture of Picasso.

6. Building a Multimedia Ontology System: the architecture and the process

In this section we describe the architecture of the system that has been prototyped at the University of Napoli *Federico II* and currently in the experimentation phase.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!DOCTYPE MULTITAO (View Source for full doctype...)>
<MULTITAO>
  <TAO name="ARTIST">
    <TAO_TEMPLATE>
      <NODEC name="PICASSO">
        <NODEC_TEMPLATE>
          <NODE name="Person" type="image">
            <NODE_TEMPLATE title="Picasso" role="image" href="FILE://c:/tao/picasso.jpg".. />
          </NODE>
          <NODE name="object1" type="image">
            <NODE_TEMPLATE title="cigarette" role="image" href="FILE://c:/tao/picasso.jpg".. />
          </NODE>
          <NODE name="object2" type="image">
            <NODE_TEMPLATE title="camera" role="image" href="FILE://c:/tao/picasso.jpg".. />
          </NODE>
          <NODE name="object3" type="image">
            <NODE_TEMPLATE title="watch" role="image" href="FILE://c:/tao/picasso.jpg".. />
          </NODE>
        </NODEC_TEMPLATE>
      </NODEC>
    </TAO_TEMPLATE>
  </TAO>
</MULTITAO>
```

Figure 4. An example of TAO_XML document

Figure 5 shows the main processes provided by the system, and described in the following:

- *Reverse Document Production - RDP*: it is composed by an initial extraction phase - which takes care of the identification of the basic TAO components - and a second abstraction phase, in which we adopt a representation that is independent from any particular application. The resulting document contains all the basic components and their interconnections. A possible abstraction is a labelled tree as we can see in the example in figure 4. In the example, we want to associate a concept to a multimedia object. In order to do that, we perform multimedia dictionary association and concept association. Once both the associations have been performed, the document is stored into a repository and opportunely indexed using a relational $DBMS$ (Oracle 8i, in our case).
- *Direct Document Production - DDP*: the document extracted from the repository is reconstructed in all of its parts and transformed into its original form. In order to do this, it is necessary to have a subprocess analogous to the one described before.

The architecture has been realized by reusing some tools such as $XMetal$, or implementing new tools at our laboratories, such as transformation, abstraction and extraction tools.

An example of the TAO_XML result of the RDP phase is shown in figure 2. Actually, that is a complete document, which define the component objects and their role in the whole structure.

7. Discussion and Conclusions

The approach we have presented in this paper demonstrates the following advantages:

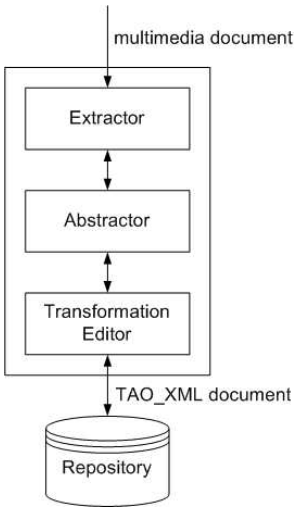

Figure 5. System architecture

- the first obvious advantage is the portability on different hardware/software platforms and the extreme interoperability; the end user, in order to interact with the document collection, only needs a simple browser (e.g. Internet Explorer, etc.);
- the second advantage is the fact that XML is already integrated with the most popular database management systems (e.g. Oracle 8i);
- third, XML is a widespread language in the Internet, thus the documents represented in this format are no longer limited to use in a single organization but may be distributed to other organizations as well;
- eventually, the possibility to share several objects coming from different databases and different organizations may help to build distributed ontologies.

References

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

[2] A. Eleftheriadis and D. Hong. Flavor: a formal language for audio-visual object representation. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 816–819. ACM Press, 2004.

[3] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, June 1993.

[4] M. J. Hu and Y. Jian. Multimedia description framework (MDF) for content description of audio/video documents. In *DL '99: Proceedings of the fourth ACM conference on Digital libraries*, pages 67–75. ACM Press, 1999.

[5] M. H. Ketfi A., Gensel J. An object-based knowledge representation approach for multimedia presentations. In *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, 2001.

[6] H. Liu and P. Singh. Commonsense reasoning in and over natural language. In *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES'2004)*, 2004.

[7] R. Lozano, M. E. Adiba, H. Martin, and F. Morcellin. An object dbms for multimedia presentations including video data. In *ECOOP '98: Workshop ion on Object-Oriented Technology*, pages 553–554. Springer-Verlag, 1998.

[8] P. Maresca, T. Arndt, and A. Guercio. Unifying distance learning resources: The metadata approach. *Journal of Computers*, 13(2):60–76, November 2001.

[9] C. Marmo and P. Maresca. TAO_XML. Master's thesis, University of Napoli, 1999.

[10] G. A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, November 1995.

[11] F. Nack and W. Putz. Designing annotation before it's needed. In *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, pages 251–260. ACM Press, 2001.

[12] A. P. Natsev, M. R. Naphade, and J. R. Smith. Semantic representation: search and mining of multimedia content. In *KDD '04: Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 641–646. ACM Press, 2004.

[13] M. Page, J. Gensel, C. Capponi, C. Bruley, P. Genoud, D. Ziebelin, D. Bardou, and V. Dupierris. A new approach in object-based knowledge representation: The AROM system. In *Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, pages 113–118. Springer-Verlag, 2001.

[14] W. Recommendation. Synchronized Multimedia Integration Language (SMIL 2.0) [2nd Edition]. http://www.w3.org/TR/2005/REC-SMIL2-20050107/, January 2005.