

# Requirements for a Multimedia Ontology Framework

Carsten Saathoff, Simon Schenk, Steffen Staab  
ISWeb, University of Koblenz  
Koblenz, Germany  
{saathoff,schenk,staab}@uni-koblenz.de

January 27, 2006

## 1 Introduction

In our research group we use multimedia ontologies mainly for two purposes: description of multimedia content for further processing and analysis using multimedia reasoning techniques and annotation of multimedia content for sharing and retrieval of the content in the context of Semantic Desktop, P2P or PIM applications. Our publication list currently only contains the papers on the aceMedia Knowledge Infrastructure.

## 2 Requirements

We have collected a number of short-term and long-term requirements that we would like to find in an integrated multimedia ontology framework. We would also like to stress that we do not necessarily favor one single, all-in-one ontology, but would rather prefer a set of ontologies modeling different aspects of multimedia content in a way that is interoperable.

We see that multimedia is a diverse field covering numerous applications from traditional images to 5-dimensional videos and probably even more. And we also think, however, that a common upper model is desirable. Nonetheless we are convinced that starting with a set of specific ontologies, that are soon ready for adoption, is preferable. In fact, the number of people using ontologies for multimedia related tasks seems to be growing and therefore providing a concise and consistent framework early would lead to a larger number of early adopters and could thus deliver useful results on the deployment of the framework. Obviously this could serve as the basis for a common upper model.

## 2.1 Representation of content structure

Especially for multimedia reasoning, the structure of the multimedia content in terms of segmentations, key frames, shots and so on is utmost important. The structure is important to introduce domain knowledge and heuristics into the process of reasoning. This problem is already covered by the MPEG-7 MDS and ported to RDF or OWL in several proposals (aceMedia, Hunter, ...). We do not require a MPEG-7 based ontology, since we do not experience MPEG-7 as a widely adopted and supported standards when it comes to available tools. However, the basic means in MPEG-7 to describe content structure are very feasible and already used in our current research work.

Besides means for describing the content structure we also need to describe the relations of different structural elements. Currently we use spatial and topological relations, and for reasoning on videos we will start adding temporal relations. Therefore a module representing common and extractable relations based on some formally defined and consistent model (e.g. Alans interval calculus) would be another requirement.

## 2.2 Representation of low-level features

In order to map ontological concepts to parts of multimedia content we need to represent low-level features as parts of domain ontologies. Representation of low-level features should therefore form an distinct module in the framework. Again, MPEG-7 does solve this problem and transformations to RDF and OWL are available. However, we still do not require a MPEG-7 based representation.

## 2.3 Reasoning support

In principle it would be advantageous to have a framework that is represented in a logic for which reasoning methods exist. However, since it's still not known what reasoning approach will be useful for multimedia reasoning, or if there's one single approach at all, it could also be advantageous to provide means for meta-modeling or uncertainty within the framework. We think it would be desirable to develop a multimedia ontology framework, that provides different levels of complexity similar to family of OWL languages. Thus, a OWL-DL compatible subset can be used if DL reasoning is needed. Others requiring the representation of uncertainty could use another subset and develop custom methods for using the ontology.

In particular we would like the following features to be supported:

- support for DL-like reasoning
- rules
- uncertainty, fuzzyness
- meta modeling

## **2.4 Requirements for the framework**

As mentioned before we support the idea of an common upper ontology and common model for multimedia. But we also see that the number of requirements are largely diverse and maybe awkward to integrate without further experiences. We therefore propose to start with a smaller, easy to use and comprehensible set of ontologies that are aligned on the metadata level. We think that this would increase the number of people using the framework and would also favour the creation of tools and applications. In parallel work could start to develop a common upper ontology for multimedia and a concise model backing that ontology up.

## **2.5 Distinguish between information object and content**

We think that one must be able to differentiate between annotations describing the information object, i.e. the file, and the content of that information object. For instance, data like creation date and creator versus the semantic concepts depicted in an image like trees or the sky.

## **2.6 Multi-Modal and Multi-Tier Annotations**

Especially for videos multi-modality is important. In the case of multimedia reasoning also annotation provided by different analysis algorithms is important. A multimedia ontology framework should support annotations which such metadata in a way that different modalities or annotations of different sources can be distinguished and properly be linked for further processing. Especially annotations produced by different annotators (if human or not) might be inconsistent and mixing them might result in a invalid and useless annotation.

## **2.7 Retrieval of metadata**

The annotations of multimedia content can become rather complex, especially if they are used for reasoning purposes. Therefore we need and are currently working on means for structuring the metadata for easy retrieval. We think that a way to address sub-parts of the metadata directly, without the need to traverse the graph or use complex algorithms, is mandatory for successful adoption of an ontology framework for multimedia annotation and reasoning.

Currently we are working on a dynamic version of named-graphs, which will be especially useful for rich multimedia annotations, especially if several dimensions of annotation have to be covered, such as low-level features, structure, reasoning-oriented annotations, retrieval-oriented annotations and administrative annotations. Also provenance tracking and privacy related issues will be covered by our approach.

# **3 Summary**

To summarize our requirements, we list all important points in the following

- representation of content structure
- spatial, topological and temporal relations and concise model, possibly with support for reasoning on that level
- representation of low level features
- support for easy retrieval of metadata
- in principle, rather a set of interrelated but simple ontologies, that can be used independently, than one very complex but monolithic ontology
- distinguish information about annotated information object from information about content of annotated information object. (e.g. resolution of image vs. depicted person)
- link to upper ontology