Capturing MPEG-7 Semantics

Stamatia Dasiopoulou[1], Vassilis Tzouvaras[2], Ioannis Kompatsiaris[1] and Michael G. Strintzis[1]

[1] Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
[2] Image, Video and Multimedia Systems Laboratory, National Technical University of Athens, Greece

Abstract. The ambiguities due to the purely syntactical nature of MPEG-7 have hindered its widespread application as they lead to serious interoperability issues in sharing and managing multimedia metadata. Acknowledging these limitations, a number of initiatives have been reported towards attaching formal semantics to the MPEG-7 specifications. In this paper we examine the rationale on which the relevant approaches build, and building on the experiences gained we present the approach followed in the BOEMIE project[3].

1 Introduction

Multimedia content is omnipresent on the Web, rendering the availability of interoperable semantic content descriptions a key factor towards realising applications of practical interest. The recent literature includes a number of efforts towards extracting well-defined descriptions, utilising the formal and exchangeable semantics provided by ontologies and the Semantic Web. However, descriptions of multimedia content come intertwined with media related aspects, for which a unified representation is also required to ensure truly interoperable multimedia metadata.

The MPEG-7 standard [5] constitutes the greatest effort towards a common framework to multimedia description. However, despite providing a wide coverage of the aspects of interest, MPEG-7 leads to a number of ambiguities that hinder the interoperability and sharing of the resulting descriptions, due to its XML Schema implementation lack of explicit semantics and the allowed flexibility in using the provided description tools.

To overcome the resulting ambiguities and align with the Semantic Web initial for machine understandable metadata, a number of efforts towards adding formal semantics to MPEG-7 through its ontological representation have been reported. Different approaches have been followed, ranging from those targeting direct translations pertaining to the MPEG-7 specifications, to those focusing more on the need for unambiguous, well-defined semantics of the corresponding descriptions. In the following we examine the existing approaches

[3] http://www.boemie.org/

and discuss on the respective design rationales and issues raised. Based on such study, we present the Multimedia Content Ontology (MCO) developed within the BOEMIE project for the purpose of capturing content structure and decomposition semantics in a rigid way.

The remainder of the paper is organised as follows. Section 2, briefly overviews on MPEG-7 and highlights the inherent interoperability issues, while Section 3 goes through the existing MPEG-7 ontologies, focusing on the different engineering principles followed. Section 4, presents the proposed MCO ontology, examining the objectives, engineering methodology and attained value. Finally, Section 5, concludes the paper and discusses future perspectives.

2 MPEG-7 Interoperability Issues

The MPEG-7 standard, known as ISO/ICE 15938, constitutes the greatest effort towards a common framework to multimedia description. Formally named "Multimedia Content Description Interface", it was developed by the Moving Pictured Expert Group, a working group of ISO/IEC. It aims to provide a rich set of standardised tools for the description of multimedia content and additionally support some degree of interpretation of the information's meaning, enabling thus smooth sharing and communication of multimedia metadata across applications and their efficient management, e.g. in terms of search and retrieval.

MPEG-7 consists of four main parts: the Description Definition Language (DDL), i.e. the XML-based language building blocks for the MPEG-7 metadata Schema, the Visual and Audio parts that include the description tools for visual and audio content respectively, and the Multimedia Description Schemes (MDSs), the comprise the set of Description Tools (Descriptors and Description Schemes) dealing with generic as well as multimedia entities. Generic entities are features, which are used in audio and visual descriptions and therefore "generic" to all media (e.g. vector, time, etc.). Apart from these, more complex Description Tools have been defined that can be grouped into five categories according to their functionality: i) content description, which address the representation of perceivable information, ii) content management, which includes information about media features, creation and usage of audiovisual content, iii) content organisation, iv) navigation and access, which refers to summaries specifications and variations of content, and v) user interaction that addresses user preferences and usage history pertaining to the consumption of multimedia content. Consequently, using MPEG-7 one can construct descriptions referring to the media itself, the content conveyed, management and organisation aspects.

MPEG-7 is implemented in the form of XML Schemas that define how well-formed and valid content descriptions can be constructed. However, the intended semantics of the provided description tools are available only in the form of natural language documentation accompanying the standard's specifications, thereby leaving proper use of the description tools in the responsibility of each application/system. For example, the conceptual difference between *StillRegion* and *VideoSegment* is not, and cannot be, reflected in the corresponding XML

schemas. Further ambiguities result from the flexibility adopted in the MPEG-7 specifications: the provided tools and descriptions are not associated with unique semantics. For example, the MovingRegion DS can be used to describe an arbitrary set of pixels from an arbitrary sequence of video frames, one pixel of a video frame, a video frame or even the full video sequence. This is contrary to the human cognition that perceives each of the aforementioned as conceptually different notions. Obviously such flexibility leads to many problems in terms of interoperability of the produced descriptions, and particularly with respect to the preservation of intended semantics.

In addition, MPEG-7 aiming to provide a generic framework for multimedia content description rather than committing to specific application aspects provides significant flexibility in the usage of the provided description tools. For example, describing an image of Zidane scoring can be done using among others keywords or free text, and can be done either at image or image region level, with all possible combinations conveying exactly the same meaning. However, in this paper we focus on media rather than content aspects of multimedia metadata; consequently issues related to the utilisation, alignment and coupling of domain ontologies with multimedia ones are out of scope.

As direct result of the aforementioned, different applications may produce perfectly valid MPEG-7 metadata that are however non-interoperable to each other as they implicitly conform to different conceptualisations. This practically means, that in order to access and retrieve such metadata in a uniform and integrated way, appropriate mappings and customised queries translations (e.g., expansions) need to be pre-defined.

3 State of the Art MPEG-7 Ontologies

Motivated by the advances in content annotation brought by, within the Semantic Web content, research in multimedia analysis and annotations shifted towards the exploration of Semantic Web technologies, especially for expressing the extracted content descriptions. However, as multimedia come in two layers, soon the need to formalise semantics at media level as well became apparent. As described in the following, a number of approaches have been reported targeting (partial) translations of MPEG-7 into an ontology. However, the standard's normative semantics force each author to decide individually how to interpret the semantics from the syntax and description provided, inevitably leading to different methodologies for mapping the provided XML Schema to RDFS/OWL models.

Chronologically, the first initiative to make MPEG-7 semantics explicit was taken by Hunter [3]. The RDF Schema language was proposed to formalize the Multimedia Description Scheme and the Descriptors included in the corresponding Visual and Audio parts, while later the proposed ontology was ported to OWL [4]. The translation follows the standards specifications and preserves the intended flexibility of usage, while making explicit the semantics related to the implied hierarchy by formalizing the "is-a" relations through subclass relations.

As a result, the different segment types are treated as both segments (due to the subclass relation with respect to the Segment class) and multimedia content items (due to the subclass relation with respect to the MultimediaContent class). Such an approach, simplifies on one hand addresses issues with respect to part-whole semantics, e.g. a query for an image depicting Zidane would return images containing a still region depicting Zidane, but retains the issues related to semantic ambiguities, e.g. it is still not possible to differentiate conceptually a moving region from a frame or a video segment.

Two RDFS MPEG-7 based multimedia ontologies have been developed within the aceMedia project [1] in order to address the representation of the multimedia content structure and of visual description tools. The use of RDFS restricts the semantics captured to subclass and domain/range relations, however, the defined content and segment concepts are kept distinct, contrary to the approach of [4]. Furthermore, additional concepts that represent notions common to humans, such as the Frame concept, have been introduced to represent content decomposition aspects (and semantics) not present in MPEG-7.

Another effort towards an MPEG-7 based multimedia ontology has been reported within the context of the SMART Web project [6]. The developed ontology focuses on the Content Description and Content Management DSs. Two disjoint top level concepts, namely the MultimediaContent and the Segment, subsume the different content and segment types. A set of properties representing the decomposition tools specified in MPEG-7 enables the implementation of the intrinsic recursive nature of multimedia content decomposition. Although in this approach, axioms have been used to make explicit parts of the MPEG-7 intended semantics, ambiguities are present due to the fact that in many cases the defined concepts and properties semantics lie in linguistic terms used.

Contrary to the aforementioned efforts that target partial translations of MPEG-7 in a manual fashion, an initiative towards a more systematic approach to translate the complete MPEG-7 to OWL has been reported in [2]. The proposed approach is based on a generic XML Schema to OWL mapping designed to ensure that the intended semantics are preserved. The resulting MPEG-7 ontology has been validated in different ways, one of which involved its comparison against [4], which showed their semantic equivalence. However, as the resulting ontology is an OWL Full ontology, issues emerge with respect to the complexity and decidability of the applicable inferences. Another related activity refers to the work conducted in the context of the DS-MIRF framework [8], where a methodology has been presented for translating the complete MPEG-7 MDS (including content metadata, filtering metadata etc.) into OWL. The main characteristic of this approach is that the resulting MPEG-7 ontology is used as a core ontology in order to integrate domain specific annotations.

Similar to the efforts of porting MPEG-7 to an ontology, Troncy et. al. proposed to represent both the Schema and the semantics of DAVP using Semantic Web technologies [7]. An ontology has been developed to capture the DAVP Schema semantics while additional rules have been defined to capture additional

constraints. Although contributing towards less ambiguous descriptions semantics and usage, the proposed approach .

4 The BOEMIE approach

From the aforementioned, it is clear that two core issues with respect to formalising MPEG-7 semantics include: modelling unambiguously conceptually distinct entities, and capturing the logical relations semantics of the considered entities. Building on the experiences gained from the existing literature, two ontologies have been implemented to address structural and audiovisual descriptors within the BOEMIE project. In the following we exemplify the proposed approach for the so called Multimedia Content Ontology (MCO), the ontology that addresses the semantics of content structure and decomposition.

Taking into account the intertwined in multimedia content media and content layers, multimedia ontologies are involved in analysis, annotation and retrieval. As such, an ontology addressing multimedia content structural aspects needs to provide the means to represent the following types of knowledge:

- The different types of media content considered, e.g. images, captioned images, web pages, video, etc.
- The semantics of decomposition of the corresponding media types into their constituent parts according to the level of the produced annotations, and
- Logical relations among the different media types, e.g. a web page may consist of a text extract, two images, and an audio sample.

Based on the identified MPEG-7 interoperability issues and the solutions proposed in the existing literature, three main challenges have been identified with respect to reaching a well-defined conceptualisation, providing cleaner semantics:

- MPEG-7 allows different Segment DS types to be used for representing the same multimedia entity. For example, to represent an image one can use both the StillRegion DS and the VideoSegment DS.
- The provided Segment Description Tools do not have unique semantics in themselves either. For example, the MovingRegion DS can be used to describe an arbitrary set of pixels from an arbitrary sequence of video frames, one pixel of one video frame, a full video frame or even the full video sequence.
- The above ambiguities extend to the decomposition relations among the different multimedia content items and types of multimedia segments. For example, a video can be temporally decomposed into segments of both the VideoSegment type and of the StillRegion type, while StillRegion type segments may result as well from a spatial decomposition

4.1 Multimedia Content Ontology Engineering

The MCO engineering starts with the identification of the different types of multimedia content addressed and the introduction of respective concepts. The

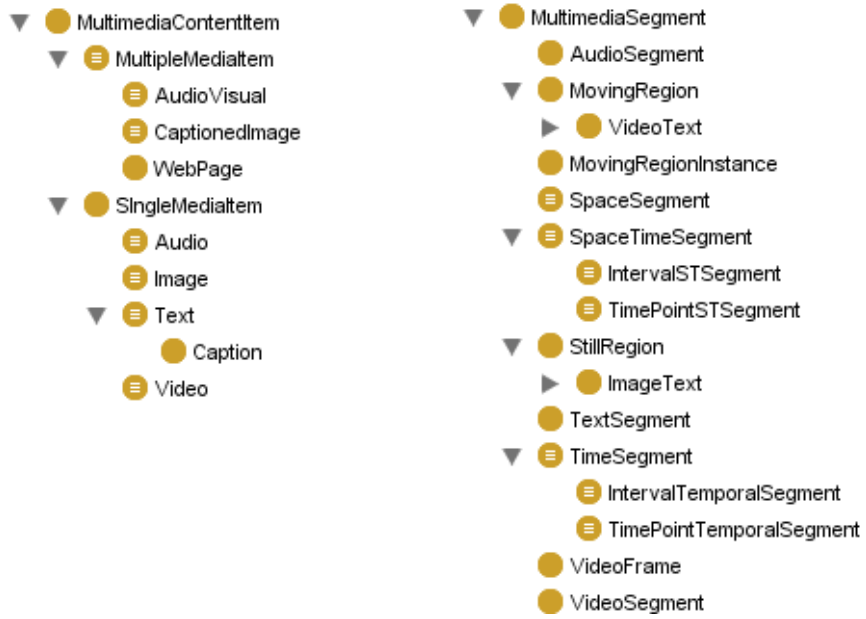

Fig. 1. Multimedia content item types and segment types hierarchies.

concepts of SingleMediaItem and MultipleMediaItem have been introduced to discriminate between multimedia content items consisting of a single media item and composite ones that contain multiple different media content items (e.g., a captioned image is be defined as a type of composite content that includes an image and an associated text). Figure 1 illustrates the respective content item hierarchy. Axioms have been included to formally capture the definitions of the composite items based on the different types of content they may include, while for the single media content items respective axioms have been introduced only this time based on their decomposition properties.

The class MultimediaSegment has been introduced to represent the different types of constituent segments into which the different single media content items can be decomposed. The different types of multimedia segments have been further grouped with respect to the decomposition dimension, namely spatial, where only position information is required to identify the desired segment, temporal, where information with respect to the time point or interval is required to define the corresponding segment, and spatiotemporal, where both position and time related information is required, thus leading to the following hierarchy. As result the classes: *SpaceSegment*, *TimeSegment* and *SpaceTimeSegment* have been defined, while appropriate axioms have been introduced to define the kind of decompositions applicable to each of the three multimedia segment types, as shown in Table 1. The applicable decompositions are a direct result (semanti-

cally) of the aforementioned definitions of what constitutes a spatial, temporal and spatiotemporal segment respectively.

$$\text{SpaceSegment} \equiv \exists\ \text{hasSegmentLocator.SpaceLocator}$$
$$\text{SpaceSegment} \equiv \forall\ \text{hasSegmentLocator.SpaceLocator}$$
$$\text{SpaceSegment} \sqsubseteq \forall\ \text{hasSegmentDecomposition.SpaceSegment}$$
$$\text{TimeSegment} \equiv \exists\ \text{hasSegmentLocator.TimeLocator}$$
$$\text{TimeSegment} \equiv \forall\ \text{hasSegmentLocator.TimeLocator}$$
$$\text{TimeSegment} \sqsubseteq \forall\ \text{hasSegmentDecomposition.(TimeSegment} \sqcup \text{SpaceTimeSegment)}$$
$$\text{SpaceTimeSegment} \equiv \exists\ \text{hasSegmentLocator.SpaceLocator}$$
$$\text{SpaceTimeSegment} \equiv \exists\ \text{hasSegmentLocator.TimeeLocator}$$
$$\text{SpaceTimeSegment} \sqsubseteq \forall\ \text{hasSegmentDecomposition.SpaceTimeDecomposition}$$

Table 1. Segment types definitions based on decomposition dimension.

Given the types of single media items addressed, corresponding classes of the different decomposition segments have been included, basing the definitions on restrictions with respect to the corresponding segment locator and decomposition properties. In Figure 1, the corresponding segment type hierarchy is depicted, while in Table 2, indicative segment type definitions are provided.

$$\text{StillRegion} \sqsubseteq \forall\ \text{hasSegmentSpaceDecomposition StillRegion}$$
$$\text{StillRegion} \sqsubseteq \exists\ \text{hasSegmentSpaceDecomposition StillRegion}$$
$$\text{StillRegion} \sqsubseteq \forall\ \text{hasSegmentLocator.SpaceLocator}$$
$$\text{StillRegion} \sqsubseteq \exists\ \text{hasSegmentLocator.SpaceLocator}$$
$$\text{MovingRegion} \sqsubseteq \forall\ \text{hasSegmentSpaceTimeDecomposition MovingRegion}$$
$$\text{MovingRegion} \sqsubseteq \exists\ \text{hasSegmentTimeDecomposition MovingRegion}$$
$$\text{MovingRegion} \sqsubseteq \forall\ \text{hasSegmentLocator.SpaceLocator}$$
$$\text{MovingRegion} \sqsubseteq \exists\ \text{hasSegmentLocator.TimeLocator}$$

Table 2. Illustrative segment types definitions.

The concept SegmentLocator has been introduced to represent the different types of locators one can use to identify a particular segment (spatial, such as a visual mask, and temporal, such as a time interval).

A number of additional properties, not directly related to the structure of multimedia content but necessary within the BOEMIE analysis and interpretation, have been defined. These include the association of a content item with the URL providing the physical location of the file, the association of a segment/content item to the domain concepts it depicts and its respective ABox (as a URL including the respective domain-specific instances).

The aforementioned definitions, in combination with the disjointness axiom defined between the MultimediaContentItem and MultimediaSegment classes, enforce distinct semantics to the different segment and decomposition types, overcoming the respective ambiguities present in relevant MPEG-7 specifications

part and in parts of the existing MPEG-7 ontologies. Thus, cleaner semantics are achieved, while, and most importantly, well-defined inference services can be applied to ensure the semantic coherency of the produced annotations. The corresponding ontology files can be accessed at $http://www.boemie.org/$.

4.2 Instantiation Example

In this subsection, we provide an annotation example based on the developed Multimedia Content Ontology. Assuming a captioned image depicting Feovana Svetlanova's pole vault trial, the assertions utilising the proposed MCO are shown in Table 3, where the $N3$ notation has been used for readability purposes, and *aeo* corresponds to the athletics domain ontology that provides the domain specific vocabulary.

$$
\begin{gathered}
mco : Image(Image1) \\
mco : Caption(Caption1) \\
mco : CaptionedImage(CaptionedImage1) \\
mco : contains(CaptionedImage1, Image1) \\
mco : contains(CaptionedImage1, Caption1) \\
mco : hasURL(CaptionedImage1, \text{“}http://..pole_vault1.jpg\text{”}) \\
mco : StillRegion(StillRegion1) \\
mco : hasMediaDecomposition(Image1, StillRegion1) \\
mco : Mask(Mask1) \\
mco : hasSegmentLocator(StillRegion1, Mask1) \\
hasURL(Mask1, \text{“}http://..mask-url..\text{”}) \\
mco : hasMediaDecomposition(Image1, StillRegion2) \\
mco : StillRegion(StillRegion2) \\
mco : hasSegmentLocator(StillRegion2, Contour2) \\
mco : Contour(Contour2) hasURL(Contour2, \text{“}http://..contour-url..\text{”}) \\
mco : depicts(StillRegion1, PersonFace1) \\
aeo : PersonFace(PersonFace1) \\
mco : depicts(StillRegion2, HorizontalBar1) \\
aeo : HorizontalBar(HorizontalBar1) \\
mco : TextSegment(TextSegment1) \\
mco : hasMediaDecomposition(Caption1, TextSegment1) \\
mco : TokenLocator(TokenLocator1) \\
mco : hasSegmentLocator(TextSegment1, TokenLocator1) \\
mco : StartCharacter(StartCharacter1) \\
mco : EndCharacter(EndCharacter1) \\
mco : hasOffsetValue(StartCharacter1, \text{“}0\text{”}) \\
mco : hasOffsetValue(EndCharacter1, \text{“}17\text{”}) \\
mco : depicts(TextSegment1, PersonName1) \\
aeo : PersonName(PersonName1) \\
aeo : hasPersonNameValue(PersonName1, \text{”}FeovanaSvetlanova\text{”})
\end{gathered}
$$

Table 3. Example annotation utilising the MCO of an image depicting a high jump attempt of an athlete.

5 Conclusions and Perspectives

MPEG-7 constitutes the greatest effort towards standardised multimedia content descriptions, comprising a rich set of broad coverage tools. Aiming to serve as a common framework, it builds on a generic schema, allowing for great flexibility in its usage, a feature partially accountable for the confronted interoperability issues, the other main reason being the lack of formal semantics that would render such descriptions unambiguous. On the other hand, the Semantic Web provides the languages and means to express, exchange and process the semantics of information.

The aforementioned study reveals two core issues with respect to formalizing MPEG-7 intended semantics: modelling unambiguously conceptually distinct entities, and capturing the logical relations semantics of the considered entities. The former is a prerequisite to ensure interoperability among the different existing MPEG-7 ontologies, and to allow the definition of respective mappings. The latter forms a key enabling factor for truly utilizing the expressivity and inference services provided by ontologies. The significance of providing interoperable formal semantics is further emphasized by initiatives such as the Multimedia Semantics (MMSEM) Incubator Group[4], a standardization activity of the World Wide Web Consortium (W3C) that aims at providing a common framework for achieving semantic interoperability and the Common Multimedia Ontology Framework[5].

The presented Multimedia Content Ontology although not providing an immediate solution towards interoperable multimedia metadata, constitutes a valuable contribution towards a cleaner conceptualisation of media related aspects that allows the utilisation of ontology reasoners in order to assess the semantic consistency of the produced annotations. Ensuring precise media semantics is of significant importance for the realisation of a multimedia enabled Web, as it strongly relates to the feasibility of applications involving semantic handling of multimedia content, as for example multimedia presentation generation applications, hypermedia ontology-frameworks, and of course all services involving semantic-based annotation and retrieval of multimedia content.

References

1. Bloehdorn, S. et. al. : Semantic Annotation of Images and Videos for Multimedia Analysis. Proc. ESWC, Heraklion, Crete, Greece, May 29 - June 592-607, 2005.
2. Garcia, R., and Celma, O.: Semantic Integration and Retrieval of Multimedia Metadata. Proc. ISWC, Galway, Ireland, Nov. 6-10, 2005.
3. Hunter, J.: Adding Multimedia to the Semantic Web: Building an MPEG-7 Ontology. Proc. 1st Semantic Web Working Symposium, Stanford University, California, USA, July 30 - August 1, 2001, pp. 261-283.
4. Hunter, J., Drennan, J., Little, S.: Realizing the Hydrogen Economy through Semantic Web Technologies. IEEE Intelligent Systems, (1): January 2004, pp. 40-47.
5. Martinez, J.M. (editor). : Overview of the MPEG-7 Standard (v4.0) ISO/MPEG N3752.
6. Oberle, D. et.al. : DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology). Tech. Report, AIFB, University of Karlsruhe. July 2006.

[4] $http://www.w3.org/2005/Incubator/mmsem/$

[5] $http://www.acemedia.org/aceMedia/reference/multimedia_ontology/$

7. Troncy, R. et. al.: Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles. Proc. SAMT, Athens, Greece, December 6-8, 2006, pp. 41-55.
8. Tsinaraki, C. et. al.: Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content. Multimedia Tools Appli. 26(3), pp. 299-325, 2005.