

# Semantics of Temporal Media Content Descriptions

Martin Höffernig, Michael Hausenblas, Werner Bailer

(JOANNEUM RESEARCH Forschungsgesellschaft mbH

Institute of Information Systems and Information Management

Steyrergasse 17, 8010 Graz, Austria

{firstname.lastname}@joanneum.at)

**Abstract:** The temporal structure of multimedia content is an important aspect of the description of time-based media and needed in many applications. Expressive content description languages, such as MPEG-7, provide tools for describing the temporal decomposition of content into segments. Although the semantics of temporal decomposition are apparent, the validation of the semantics of the temporal decompositions (e.g. temporal extent of child segments, gaps, overlaps) is not possible on the syntactic level. We propose therefore to model the semantics of temporal decompositions using ontologies and rules. As a proof of concept we apply the formalisation to a validation use case, implemented as a Web application.

**Key Words:** Metadata, Ontology, Semantic Web, MPEG-7, Temporal Decomposition, Time Interval, Validation

**Category:** H.5.1

## 1 Introduction

The description of multimedia content is of growing importance in a number of applications dealing with multimedia content creation, processing and archiving. Media content descriptions can be on a global scope (i.e. describing only metadata related to a complete media item, such as title and production information) or related to spatial, temporal and spatiotemporal segments of the content. An important aspect of a detailed content description of time-based media is the description of the temporal structure of the content, i.e. its decomposition into formal and logical units, such as e.g. shots, scenes or speech segments. There are several use cases where semantic annotations on these segments are relevant.

Imagine a YouTube<sup>1</sup>-like service, which provides summaries of the videos in the database in order to facilitate browsing. To produce such summarisation clips automatically, annotations on temporal segments of the source video are used in order to determine the relevant snippets that are put into the summary. This requires combining semantic descriptions of the content and the temporal structure of the source material.

Now imagine that we use the videos we have found with the help of the summaries to edit new content, like in classic post-production or in a Web 2.0

---

<sup>1</sup> <http://www.youtube.com>

application such as jumpcut<sup>2</sup>. Instead of just getting the final video as output, it would be great to also get a metadata description of the output. This requires metadata editing, i.e. automatically applying the edit decisions taken on the audiovisual material to the related metadata. There are annotations for each of the segments in each of the source contents and the edit operations creates a new segmentation. The task is to identify which metadata from the source applies to which segments of the target content, and whether there are potentially conflicting descriptions from the two source contents.

Last but not least, one can think of the automatic, semantic validation of temporal decomposition. Systems may produce descriptions of media assets that conform to a certain standard, such as MPEG-7, on a syntactic level—but how about the semantics? We aim at answering this question in this paper by formalising the semantics of temporal decompositions of media content descriptions.

### 1.1 Existing Work

The description of the temporal structure of the content is one of the most important aspects of a detailed content description of time-based media, and in particular a strength of MPEG-7 over other multimedia content description standards. The flexibility of MPEG-7 is based on allowing descriptions to be associated with arbitrary multimedia segments or regions, at any granularity, using different levels of abstraction. The downside of the breadth targeted by MPEG-7 is its complexity and its fuzziness [Bailer and Schallauer 2006, Ossenbruggen et al. 2004]. For example, very different syntactic variations may be used in multimedia descriptions with the same intended semantics, while remaining valid MPEG-7 descriptions. To reduce this syntax variability, MPEG-7 has introduced the notion of *profiles* that constrain the way multimedia descriptions should be represented for particular applications. Profiles are therefore a way of reducing the complexity of MPEG-7 (i.e. only a subset of the whole standard can be used) and of solving some interoperability issues (i.e. English guidelines are provided on how the descriptors should be used and combined). However, these additional constraints are only represented with XML Schema<sup>3</sup>, and, for most of them, cannot be automatically checked for consistency by XML processing tools. In other words, profiles provide only very limited control over the semantics of the MPEG-7 descriptions [Hunter 2001, Nack et al. 2005]. Because of this lack of formal semantics, the resulting interoperability problems prevent an effective use of MPEG-7 as a language for describing multimedia.

In [Troncy et al. 2006] the authors present an approach to formalise a subset of the semantic constraints of the Detailed Audiovisual Profile (DAVP)<sup>4</sup>.

---

<sup>2</sup> <http://www.jumpcut.com>

<sup>3</sup> <http://www.w3.org/XML/Schema>

<sup>4</sup> <http://mpeg-7.joanneum.at>

The formalisation of the semantic constraints can be used to automatically validate *semantically* the conformance of MPEG-7 descriptions to a given profile [Troncy et al. 2007].

In this work we do not focus on the semantics of such a temporal segment in terms of the type of unit it represents (e.g. shot, scene), as this is already modeled in the ontology described in [Troncy et al. 2006]. We concentrate on the semantics of the temporal segmentation [Allen and Ferguson 1994]. A temporal decomposition of a segment is a container for a set of segments, thus defining parent-child relations between the segment to be decomposed and the segments in the set. The temporal extent of a segment is specified by its time point and duration elements, which are pattern-restricted strings in MPEG-7. In addition, attributes of the temporal decomposition specify, whether overlaps of segments or gaps between them are allowed.

## 1.2 Problem Formulation

The semantics of the temporal decomposition are clearly defined. However, due to the limitations of XML Schema, documents containing one of the following two violations of temporal decomposition semantics are still valid w.r.t. to the profile schema:

**Invalid parent-child segment relation** A temporal decomposition of a segment into subsegments is only meaningful if the time range filled by each of the subsegments is at most the time range of the segment being decomposed, i.e. a part of a temporal segment cannot start before or end after its parent segment.

**Gap and overlap** A temporal decomposition can be qualified whether the subsegments in the decomposition overlap or have gaps between them. These properties are specified with the **gap** and **overlap** attributes of the decomposition that have a **true/false** value. There is, however, no mechanism to check whether the actual time description of the segments conforms to the value of the attribute or not.

An example of a temporal decomposition of a segment is shown in Figure 1. Segment  $S_1$  is decomposed into Segments  $S_2$ ,  $S_3$  and  $S_4$ . For example  $S_1$  has start point  $t_1$  and a duration  $d_1$ . This temporal decomposition contains three gaps (between  $t_1$  and  $t_2$ , between  $t_3$  and  $t_4$ , and between  $t_7$  and  $t_8$ ) and one overlap between  $t_5$  and  $t_6$ .

Our approach is to model the semantics of temporal decompositions using Semantic Web languages to formalise the semantics, and later inference tools to check the semantic consistency of the segments. Section 2 describes the approach we are proposing and Section 3 its implementation and integration into the

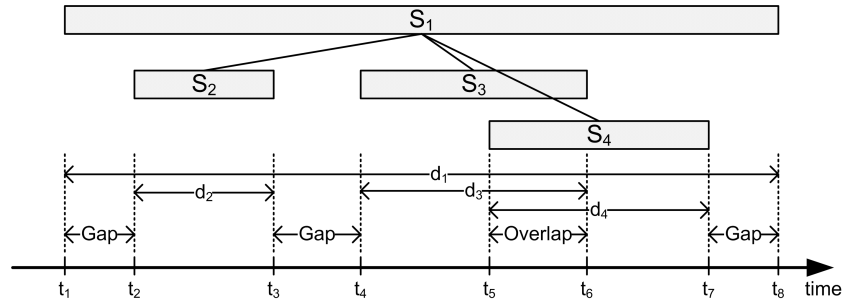


Figure 1: Temporal decomposition of segment  $S_1$  into three segments ( $S_2$ ,  $S_3$ ,  $S_4$ ) with gaps and overlap.

validation service. In Section 4 we conclude the discussion and outline future research.

## 2 Formal Representation of Temporal Media Descriptions

An ontology is used for the formal representation of temporal segments. A temporal segment is described by a start point and a duration. The ontology contains classes and properties for describing the temporal behavior of a temporal segment and the relations between these temporal segments. Hence the ontology models (i) the time interval of a temporal segment with start point and duration, (ii) the parent-child relation between temporal segments and (iii) the temporal decomposition attributes of temporal segments (overlap and gap).

### 2.1 An Ontology for Temporal Segments

Several classes and properties are needed to model the required relationships:

**Class Segment** This is the main class in the ontology. Every temporal segment is an instance of class Segment. Every instance of this class has exactly one `hasStartPoint` relation and exactly one `hasDuration` relation.

**Class ParentSegment** This class describes all temporal segments that are decomposed into further temporal segments (using `hasChild`, `hasAssertedGap` and `hasAssertedOverlap`. This class is a subclass of class Segment.

The exemplary temporal decomposition in Figure 1 is partially represented as an ontology in Figure 2<sup>5</sup>.

```

prefix tsmd: http://mpeg-7.joanneum.at/semantics/temporal#
5 prefix ex: http://mpeg-7.joanneum.at/semantics/example#
prefix rdfs: http://www.w3.org/2000/01/rdf-schema#
prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

```

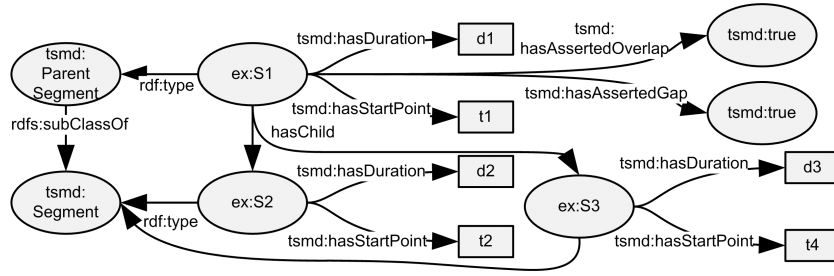


Figure 2: Excerpt of the ontology for describing the temporal decomposition shown in Figure 1.

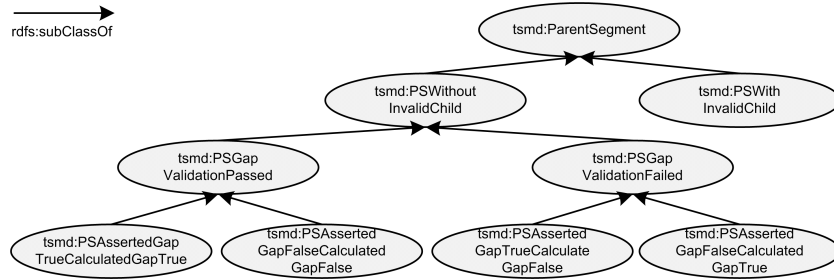


Figure 3: Validation hierarchy (classes for parent-child relation- and gap- verification)

### 3 Validation of Temporal Decompositions

The purpose of the validation process is to find invalid parent-child segment relations and to verify the asserted gap and overlap relation of a parent segment.

The presented ontology is capable of representing a temporal decomposition of a segment. For the validation purpose the classes depicted in Figure 3 are relevant. Classes needed for overlap verification are not depicted for simplicity.

#### 3.1 Validating Temporal Decompositions using Rules

Rules are used to produce new statements about a temporal segment. The Jena rules syntax<sup>6</sup> is used for defining the rules. First the rule `calculate_end_point` computes the value of the property `hasEndPoint`, which represents the end point of a temporal segment. Additional rules for calculating the property values

<sup>6</sup> <http://jena.sourceforge.net/inference/index.html#rules>

```

[parent_has_invalid_child_true:
 (?parent rdf:type tsm:ParentSegment),
 noValue(?parent tsm:hasInvalidChild tsm:true),
 (?parent tsm:hasChild ?child),
 (?parent tsm:hasStartPoint ?parent_sp),
 (?parent tsm:hasEndPoint ?parent_ep),
 (?child tsm:hasStartPoint ?child_sp),
 (?child tsm:hasEndPoint ?child_ep),
 parentHasInvalidChild(?parent_sp, ?parent_ep, ?child_sp, ?child_ep)
 ->
 (?parent tsm:hasInvalidChild tsm:true)]

[parent_has_invalid_child_false:
 ?parent tsm:hasInvalidChild tsm:false
 <-
 (?parent rdf:type tsm:ParentSegment),
 noValue(?parent tsm:hasInvalidChild tsm:true)]

```

**Figure 4:** Rules for computing `hasInvalidChild` property value.

(true or false) of the properties `hasInvalidChild`, `hasCalculatedGap` and `hasCalculatedOverlap` are defined. These properties are needed to define the classes for validation (see Figure 3) in the ontology. The rules for computing the value of the property `hasInvalidChild` are depicted in Figure 4. For this purpose, the forward rule `parent_has_invalid_child_true` and the backward rule `parent_has_invalid_child_false` are needed. New statements generated by the forward rule are the input for the backward rule. If all terms of the forward rule pass, a new statement with property `hasInvalidChild` and value `true` is generated for an instance of class `ParentSegment`. If no invalid parent-child relation has been found the backward rule produces a new statement with property `hasInvalidChild` and value `false`.

### 3.2 Implementation

As shown in Figure 5 the following steps are performed to semantically validate the temporal decomposition of a segment:

1. Classify parent segments (using a reasoner)
2. Calculate invalid children, gaps and overlaps respectively (using rules)
3. Classify the validation results (using a reasoner)

The approach is integrated into the VAMP service<sup>7</sup> for semantic validation of MPEG-7 descriptions [Troncy et al. 2007] and enables it to validate the description of the temporal media structure in terms of semantics.

<sup>7</sup> <http://vamp.joanneum.at>

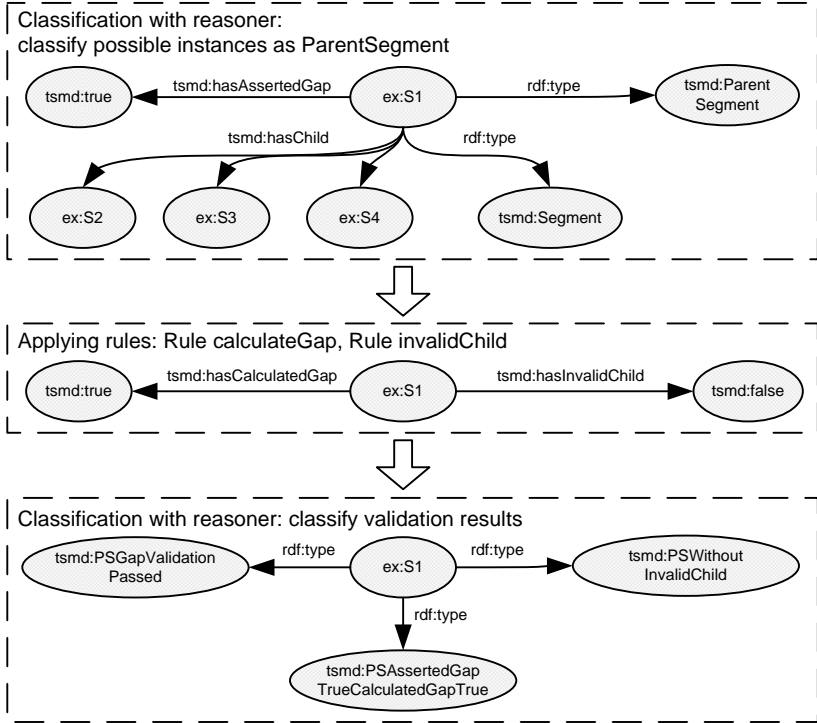


Figure 5: Validation steps for parent-child relation and gap

## 4 Conclusions and Future Work

We have proposed a formalisation of temporal decompositions in multimedia content description. Rules are subsequently used to check the validity of a the temporal description. The chosen approach is generic, as it models general properties of the description of temporal multimedia content structure, independent of the actual description format or standard. It can thus be applied to the validation of descriptions using other standards that support the concept of temporal segmentation, such as MXF DMS-1.

Another application are for the proposed model is that of search and retrieval applications, that allow querying temporal segments of the content. The metadata describing the content are often related to different segments, so that only subsegments are relevant to the query. Our approach can be used to determine the segment that shall be returned from the segments to which to the metadata description of the content relates.

A possible enhancement for the validation of temporal segmentation is to

more precisely report the violation, i.e. the segments and the elements/attributes that caused it. This would improve the quality of the error messages that the user receives and thus the usability of the service.

## Acknowledgments

The authors would like to thank Raphaël Troncy, Philip Hofmair and Rudolf Schlatte for their contributions. The research herein was partially supported by the European Commission under the contracts FP6-027026, "Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space", IST-2-511316, "IP-RACINE: Integrated Project - Research Area CINE", and IST-004124, "NM2: New Millennium – New Media".

## References

- [Allen and Ferguson 1994] James F. Allen and George Ferguson. Events and actions in interval temporal logic. *Journal of Logic and Computation, Special Issue on Actions and Processes*, 4(5):531–579, Oct. 1994.
- [Bailer and Schallauer 2006] Werner Bailer and Peter Schallauer. The Detailed Audio-visual Profile: Enabling Interoperability between MPEG-7 based Systems. In *12<sup>th</sup> International MultiMedia Modelling Conference (MMM'06)*, pages 217–224, Beijing, China, 2006.
- [Hunter 2001] Jane Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *First International Semantic Web Working Symposium (SWWS'01)*, Stanford, California, USA, 2001.
- [Nack et al. 2005] Frank Nack, Jacco van Ossenbruggen, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). *IEEE Multimedia*, 12(1), 2005.
- [Ossenbruggen et al. 2004] Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). *IEEE Multimedia*, 11(4), 2004.
- [Troncy et al. 2007] Raphaël Troncy, Werner Bailer, Michael Hausenblas, and Martin Höffernig. VAMP: Semantic Validation for MPEG-7 Profile Descriptions. Technical Report INS-E0705, (CWI) Centrum voor Wiskunde en Informatica, 2007.
- [Troncy et al. 2006] Raphaël Troncy, Werner Bailer, Michael Hausenblas, Philip Hofmair, and Rudolf Schlatte. Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles. In *1<sup>st</sup> International Conference on Semantics And digital Media Technology*, pages 41–55, Athens, Greece, 2006.