

Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles

Raphaël Troncy¹,

Werner Bailer², Michael Hausenblas², Philip Hofmair², and Rudolf Schlatte²

¹ CWI Amsterdam, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
`raphael.troncy@cwi.nl`

² JOANNEUM RESEARCH Forschungsgesellschaft mbH, Institute of Information
Systems and Information Management, Steyrergasse 17, 8010 Graz, Austria
`firstName.lastName@joanneum.at`

Abstract. MPEG-7 can be used to create complex and comprehensive metadata descriptions of multimedia content. Since MPEG-7 is defined in terms of an XML schema, the semantics of its elements have no formal grounding. In addition, certain features can be described in multiple ways. MPEG-7 profiles are subsets of the standard that apply to specific application areas, which can be used to reduce this syntactic variability, but they still lack formal semantics. In this paper, we propose an approach for expressing semantics explicitly by formalizing the semantic constraints of a profile using ontologies and rules, thus enabling interoperability and automatic use for MPEG-7 based applications. We demonstrate the feasibility of the approach by implementing a validation service for a subset of the semantic constraints of the Detailed Audiovisual Profile (DAVP).

1 Introduction

The amount of multimedia data being produced, processed and consumed is growing, as is the number of applications dealing with multimedia content. In many of these applications, metadata descriptions of the content are important. MPEG-7 [13], formally named Multimedia Content Description Interface, is designed as a standard for representing such descriptions in a broad range of applications. In order to cover diverse requirements scenarios [16], many *descriptors* and *descriptions schemes*, as well as the relationships between them, have been defined. The descriptors and description schemes are together referred to as *description tools*, and a description is a particular instantiation of these. There are description tools for diverse types of annotations on different semantic levels, ranging from very low-level features, such as visual (e.g. texture, camera motion) or audio (e.g. melody), to more abstract descriptions. The flexibility of MPEG-7 is very much based on structuring tools, which allow descriptions to be associated with arbitrary multimedia segments or regions, at any level of granularity, using different levels of abstraction.

The downside of the breadth targeted by MPEG-7 is its complexity and its fuzziness [1,19,21]. For example, very different syntactic variations may be used in multimedia descriptions with the same semantics, while remaining valid MPEG-7 descriptions. Given that the standard does not provide a formal semantics for these descriptions, this syntax variability causes serious interoperability issues for multimedia processing and exchange, for example on the Web. To reduce this syntax variability, MPEG-7 has introduced the notion of *profiles* that constrain the way the multimedia descriptions should be represented for particular applications. However, these additional constraints, such as the MPEG-7 description tools, can only be represented in XML Schema, which allows only very limited control over the semantics of the descriptions [9,14,18]. 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 this paper, we propose a method to formalize the semantic constraints of an MPEG-7 profile and a semantic validation service using the formalization. In contrast to other work [4,9,18,20], we do not intend to completely map the MPEG-7 description tools onto an OWL ontology [15,17], but rather to use Semantic Web technologies to represent those constraints defined in natural language in the standard that cannot be expressed using XML Schema.

The paper is organized as follows. In the next section, we present a scenario motivating our work and justify the need for a partial formalization of MPEG-7. In section 3, we briefly introduce the notion of MPEG-7 profiles and we describe one of them as an example illustrating our work. In section 4, we detail how the MPEG-7 profiles can be formalized using Semantic Web technologies, building first an OWL ontology and rules capturing the semantic constraints, and developing then some tools converting from the XML-based MPEG-7 descriptions to RDF triples. In section 5, we illustrate the use of this work by describing an implemented Web-based application providing a semantic validation service for MPEG-7 multimedia descriptions. Finally, we give our conclusions and outline future work in section 6.

2 Motivating Scenario

The MPEG-7 description tools (Descriptors and Description Schemes) are represented using the *Description Definition Language* (DDL), the core part of the standard. Having collected high-level requirements [16] for the language, several DDL proposals have been proposed and evaluated [11]. A decision was taken to adopt the XML Schema recommendation [22] as the most appropriate schema language¹. The MPEG-7 XML Schema defines numerous elements, types, and rules for their valid combinations. The standard, however, allows the specification of different descriptions with equivalent semantics. This raises interoperability problems when exchanging MPEG-7 descriptions, where different applications may use the standard differently.

¹ Several extensions (array and matrix datatypes) have been added in order to satisfy specific MPEG-7 requirements.

The following simple MPEG-7 description example illustrates the problem. The example is a format for reference data from the TREC Video Retrieval Evaluation, where the goal is to describe the shot structure of a video and the key frames representing each shot. The sample shown in Figure 1 is part of a description that validates against the MPEG-7 XML Schema. However, without additional knowledge about how MPEG-7 has been used, one cannot grasp the semantics of the elements in the description. For instance, the `VideoSegment` is used to represent at the same time the whole video content, the shots and the key frames.

```

<VideoSegment id="TRECVID2005_1">                                <!-- whole video -->
  <MediaLocator>
    <MediaUri>20041116_110000_CCTV4_NEWS3_CHN.mpg</MediaUri>
  </MediaLocator>
  [...]
  <TemporalDecomposition gap="false" overlap="false">
    <VideoSegment id="shot1_1">                                <!-- shot -->
      <MediaTime>
        <MediaTimePoint>T00:00:00:0F30000</MediaTimePoint>
        <MediaDuration>PT00H00M03S26116N30000F</MediaDuration>
      </MediaTime>
      <TemporalDecomposition>
        <VideoSegment id="shot1_1_RKF">                       <!-- key frame -->
          <MediaTime>
            <MediaTimePoint>T00:00:01:27057F30000</MediaTimePoint>
          </MediaTime>
        </VideoSegment>
      </TemporalDecomposition>
    </VideoSegment>
    [...]
  </TemporalDecomposition>
</VideoSegment>

```

Fig. 1. A valid MPEG-7 shot structure description from the TRECVID evaluation

More generally, the problem comes from the modeling foundations of the languages used. XML Schema allows the derivation of new element types from existing ones, either by restriction or by extension. However, even if this derivation mechanism is reminiscent of the inheritance of object-oriented programming languages, it is rather a reuse of the content model defining the super type. The purpose of XML Schema is to define and to reuse the definitions of types and elements, that is, the set of syntactic constraints that define their content model, whereas well-defined semantics is the basis of the Semantic Web technologies. Coming back to our example, a proper OWL modeling of the `VideoSegment` should come with the definition of three different concepts rather than a single element type.

To overcome this interoperability issue, MPEG-7 has introduced the notion of profiles that constrain the use of subsets of the language, suitable for certain application domains (see section 3). For example, one can convert the previous MPEG-7 description into another one, conforming to the Detailed Audiovisual Profile (DAVP) [1], which specifies a number of semantic constraints in textual form. The conversion can be done by a straightforward XSLT transform that adjusts the structure of the description and adds the additional elements and attributes that are required by the DAVP schema. The result corresponding to Figure 1 is partly shown in Figure 2.

```

<VideoSegment id="TRECVID2005_1">
  <StructuralUnit
    href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.programme"/>
  [...]
  <TemporalDecomposition gap="false" overlap="false"
    criteria="visual shots">
    <VideoSegment xsi:type="ShotType" id="shot1_1">
      <StructuralUnit
        href="urn:x-MPEG-7-davp:cs:StructuralUnitCS:2005:vis.shot"/>
      <MediaTime>
        <MediaTimePoint>T00:00:00:0F30000</MediaTimePoint>
        <MediaDuration>PT00H00M03S26116N30000F</MediaDuration>
      </MediaTime>
      <TemporalDecomposition criteria="key frames">
        <VideoSegment id="shot1_1_RKF">
          <StructuralUnit
            href="urn:x-MPEG-7-davp:cs:StructuralUnitCS:2005:vis.keyframe"/>
          <MediaTime>
            <MediaTimePoint>T00:00:01:27057F30000</MediaTimePoint>
          </MediaTime>
        </VideoSegment>
      </TemporalDecomposition>
    </VideoSegment>
  [...]
</TemporalDecomposition>
</VideoSegment>

```

Fig. 2. A valid MPEG-7 shot structure description conforming to the DAVP profile

Additional elements include the `StructuralUnit` element on the segments that specify its semantic type (shot or key frame) and the `criteria` attribute of the decompositions. With these additional elements and attributes, it is then possible to distinguish between the different types of elements for which `VideoSegment` has been used. This enrichment of the XML schema allows more explicit expression of the structure of the description. The semantics of the DAVP constraints cannot, however, be formalized using XML Schema and thus cannot be checked for full semantic conformance. For example, the DAVP XML Schema requires the

StructuralUnit elements and criteria attributes to be present. But there is no way to check if their values are set correctly, for example, whether a key frame of a shot is a part of its temporal decomposition, or whether a key frame is not further decomposed into subparts. One can still put a **VideoSegment** element at any place allowed by the XML Schema and assign it the **StructuralUnit** value for key frame, or decompose a key frame further. The resulting description would still conform to the schema, although it would violate the non-formally represented semantic constraints of DAVP. Similarly, both the MPEG-7 and the DAVP XML schemas are able to specify whether a temporal decomposition of a video segment contains some *gap*, or whether the sub-segments *overlap*, by stating a true/false value for the corresponding XML attributes. It is not possible, however, to validate whether these values actually correspond to the time-coded temporal decomposition for a given description using only XML Schema conformance tools.

To summarize this discussion, we observe that the differences between the TRECVID MPEG-7 format and the DAVP MPEG-7 format prevent the exchange of the descriptions across applications, although the descriptions contain the same information. Worse, because of the lack of a formal definition of their semantics, it is not possible to automatically define mappings between the corresponding elements of the two descriptions. We thus propose a general method for formalizing the semantic constraints of any MPEG-7 profile.

3 MPEG-7 Profiling

Profiles have been proposed as a means of reducing the complexity of MPEG-7 descriptions. As specified in the standard, the definition of a profile consists of three parts, namely: *i) description tool selection*, i.e. the definition of the subset of description tools to be included in the profile, *ii) description tool constraints*, i.e. definition of constraints on the description tools such as restrictions on the cardinality of elements or on the use of attributes, and *iii) semantic constraints* that further describe the use of the description tools in the context of the profile. The first two parts are represented using the MPEG-7 DDL and result in a specific and more constrained XML Schema. The third part is currently expressed in natural language which prevents any automated process from efficiently tackling the complexity and interoperability problems associated with the use of MPEG-7 profiling for describing multimedia content.

3.1 Semantic Constraints

The semantic constraints in a profile can be compared to the prose descriptions specifying the semantics of any MPEG-7 descriptor or description scheme: both are written in natural language. Because of the generic nature of many description tools and the envisioned broad application area of MPEG-7, these descriptions do not contain many constraints with respect to their use. A profile is tailored to fit the needs of a much narrower application area, and thus the intended use of the description tools is better defined. The semantic constraints

of the profile definition render the use of its description tools more precisely and thus avoid ambiguities in the descriptions.

The more generic and flexible a description tool is, the greater the need for semantic constraints. This is especially true for the structuring tools that allow the building of arbitrary hierarchical structures describing the media, spatial or temporal decompositions of the content into any kind of regions or segments. A profile can further specify the allowed structural decompositions for a given document type, genre or application. For example, a general `VideoSegment` descriptor describes some temporal segment of a video. In contrast, a `VideoSegment` used in the description structure of a profile might have a much more well-defined semantics. Hence, it can be refined in `Shot`, `KeyFrame` and `Transition` if the targeted application is to describe an automatic shot structure detection, or in `Report`, `Interview` and `IndoorStudio` in the context of the description of a weekly sports magazine broadcast on a TV channel [18].

3.2 Existing Profiles

Several profiles have been under consideration for standardization. Among them, the Simple Metadata Profile (SMP) allows the global definition of the general metadata of multimedia content of simple collections. The User Description Profile (UDP) consists of tools for describing user preferences and usage history in order to personalize multimedia content delivery. The Core Description Profile (CDP) contains the tools for the description of relationships between multimedia content, media information, creation information, usage information and semantic information. However, none of these adopted profiles include the visual and audio description tools that are useful, for example, for representing the results of feature extraction or automatic analysis of audiovisual content. Nor do they include the semantic constraints of the selected description tools that clarify how they should be used in the profile.

To overcome this flaw, a profile for the detailed description of single audiovisual content entities called Detailed Audiovisual Profile (DAVP)² has been proposed in [1]. This profile includes many of the MDS tools as well as those for audio and visual features description. Furthermore, it has been designed for improved support of interoperability between systems using MPEG-7, by avoiding possible syntax ambiguities and clarifying the use of the description tools in the profile. The semantic constraints thus play a crucial role in the DAVP definition since they define and constrain the use of the MPEG-7 description tools in the context of the profile. Because of the lack of formal semantics of XML Schema, these constraints are only described in textual form in the profile definition.

In summary, profiles are tailored towards specific application areas. They come with a schema that selects and further constrains a subset of description tools of the standard. Nevertheless, this does not solve the interoperability problem since the existing profiles are defined using the same informal notation as MPEG-7.

² <http://mpeg-7.joanneum.at>

4 Defining Semantics of MPEG-7 Descriptions

We detail in this section our approach to define the missing semantics layer on top of existing MPEG-7 profiles (section 4.1). We show that this task amounts to defining an ontology capturing the semantics of a given profile (section 4.2) as well as some additional rules to fully represent all its semantic constraints (section 4.3).

4.1 Overview of the Proposed Approach

MPEG-7 profiles are able to restrict the syntax variability of MPEG-7 descriptions (section 2). However, they cannot represent and thus check the correct use of the description tools according to their informal yet intended semantics. We therefore propose the following layered approach to validate *semantically* the conformance of MPEG-7 descriptions to a given profile:

Well-formedness. The MPEG-7 description is a well-formed XML document.

Validity. The description document validates against the full MPEG-7 XML Schema and the MPEG-7 profile it claims to conform to.

Semantics Consistency. The description is consistent with the ontology and rules describing the semantic constraints expressed informally in both the MPEG-7 standard and the profile definition.

We propose to use Semantic Web languages to express these constraints, and later inference tools to check the semantic consistency of the descriptions. The whole task can then be carried out with an appropriate combination of the following languages [10]:

- Using XML Schema to define largely structural constraints, i.e. what types are allowed and how they may be combined;
- Using OWL to formally capture the intended semantics of particular description tools;
- Using XSLT to convert from the original XML descriptions to RDF statements to assert the class-membership of particular description tools due to the presence of certain properties;
- Using additional rules to express relationships between structurally different but semantically equivalent constructs.

Therefore, achieving interoperability for MPEG-7 descriptions requires an ontology formally describing the profile (in terms of the concepts used) the MPEG-7 description purports to adhere to, and a way to convert automatically from the descriptions, the instances of the concepts modeled in this ontology. As an example, we show in section 4.2 how this method can be realized for the particular DAVP profile. The OWL expressivity being not enough for capturing all the semantics constraints, we give in section 4.3 some example rules to finalize the semantics formalization.

4.2 Building a Profile Ontology

Several attempts have been made to map the MPEG-7 description tools onto an OWL ontology. An automatic mapping from XML Schema to OWL covering the whole standard has been proposed [4,20]. However, without re-engineering work, the resulting ontology is unable to capture the intended semantics, not represented in the MPEG-7 schema, of the description tools. Other attempts have manually modeled an MPEG-7 ontology. However, the result is either restricted to the upper level elements and types of MPEG-7 [9], or adapted to a very specific use of the standard in a particular application [18]. Furthermore, the additional semantic constraints specified in the profile definitions cannot be taken into account.

Given this experience, we do not intend to completely map the MPEG-7 description tools onto an OWL ontology. We observe that there is also no need to model the complete profile definitions in the ontology. We therefore only capture the semantic constraints that cannot be handled by pure XML Schema tools [2]. For example, for the DAVP profile, we model a DAVP ontology containing the concepts and properties necessary to represent some structural restrictions using a container-contained pattern (**hasParent**). In this ontology, both **GlobalTransition** and **Shot** classes are defined as a subclass of a **VideoSegment** with a particular value for the **hasStructuralUnit** property. The class **KeyFrame** is also defined as a **VideoSegment** but one that cannot be further decomposed temporally and that must be contained in a **Shot**. Figure 3 shows part of the resulting DAVP ontology for later describing a shot/keyframe structure compliant with the DAVP profile.

One of the problems underlined in section 2 was that a single description tool (e.g. **VideoSegment**) can be used in different contexts with a different semantics. It is now possible to model these different concepts in the ontology and to automatically instantiate them from the XML descriptions using the additional DAVP information. The functional role of a **VideoSegment** can be inferred by an OWL reasoner given the value of its **hasStructuralUnit** property.

Finally, it has to be noted that some constraints implicitly defined in MPEG-7 are not formalized in the ontology. We detail in the next section how logical rules can be used to capture these additional constraints.

4.3 Using Additional Rules for Expressing All Semantic Constraints

The limitations of XML Schema, especially with respect to MPEG-7, have been discussed earlier in this paper. OWL itself has some serious limitations regarding the composition of properties [6,8]. The often cited example is the obvious relationship between the composition of the **isParentOf** and **isSiblingOf** properties in conjunction with the class **Male** that yields the **isUncleOf** property:

$$\text{Male}(x) \sqcap \text{isSibling}(x,y) \sqcap \text{isParentOf}(y,z) \models \text{isUncleOf}(x,z)$$

Using rules in combination with Description Logics has been investigated for quite a long time, thus a range of proposals exists [3,5,12]. In the Semantic

Web stack, it is expected that a rule language will complement the ontology layer. For example, the Semantic Web Rules Language (SWRL) [7] is a Horn clause rules extension to OWL for describing relationships between a composite property and other properties. Using SWRL would be a desirable approach, but was not feasible in practice since at the time of writing, no native SWRL-rules engine exists. The Jena Framework³ is an example of an available integrated system combining an OWL reasoner with rules, and supporting both forward and backward reasoning. The rules are expressed in the N3 format, and the support of built-ins (validation, non-monocity) and RDF/OWL reasoning facilities are good arguments for its practical usage at the moment.

```

Class(a:Segment) Class(a:VideoSegment partial
  a:Segment
  restriction(a:hasStructuralUnit cardinality(1)))
Class(a:VideoProgrammeSegment complete
  a:VideoSegment
  restriction(a:hasStructuralUnit value (
    mpeg7:urn_x-MPEG-7-davp_cs_StructuralUnitCS_2005_vis.programme)))
Class(a:GlobalTransition complete
  a:VideoSegment
  restriction(a:hasParent allValuesFrom(a:VideoProgrammeSegment))
  restriction(a:hasStructuralUnit value (
    mpeg7:urn_x-MPEG-7-davp_cs_StructuralUnitCS_2005_vis.transition)))
Class(a:Shot complete
  a:VideoSegment
  restriction(a:hasKeyframe minCardinality(1))
  restriction(a:hasParent allValuesFrom(a:VideoProgrammeSegment))
  restriction(a:hasStructuralUnit value (
    mpeg7:urn_x-MPEG-7-davp_cs_StructuralUnitCS_2005_vis.shot))))
Class(a:Keyframe complete
  a:VideoSegment
  restriction(a:hasParent allValuesFrom(a:Shot))
  restriction(a:hasTemporalDecomposition maxCardinality(0))
  restriction(a:hasStructuralUnit value (
    mpeg7:urn_x-MPEG-7-davp_cs_StructuralUnitCS_2005_vis.keyframe))))

```

Fig. 3. Excerpt of the OWL ontology corresponding to the DAVP profile using an abstract syntax

Figure 4 shows part of the rule set used in our system to check for semantically invalid constructs in an MPEG-7 DAVP description. The first rule states that a `KeyFrame` without the required `StructuralUnit` descriptor will raise an error, while the second rule states that a `KeyFrame` must be contained in a `Shot`. When an invalid instance is found, it is annotated with a `hasError` property that can later be queried with SPARQL.

³ <http://jena.sourceforge.net/inference/> and
<http://jena.hpl.hp.com/juc2006/proceedings.html>

```

# KeyFrame must have the appropriate structural unit
[keyframe_no_SU_rule:
  (?k a mpeg7:KeyFrame),
  noValue(?k mpeg7:hasStructuralUnit
    mpeg7:urn_x-mpeg-7-davp_cs_StructuralUnitCS_2005_vis.keyframe)
  ->
  (?k mpeg7:hasError mpeg7:DAVPViolationError)]

# KeyFrame must be contained in a Shot
[missplaced_keyframe_rule:
  (?k a mpeg7:KeyFrame),
  (?k mpeg7:hasParent ?s),
  noValue(?s a mpeg7:Shot)
  ->
  (?p mpeg7:hasError mpeg7:DAVPViolationError)]

```

Fig. 4. Example rules expressed in the Jena rule language

In the same way, the Jena rule language is used to represent the semantic constraints implicitly defined in MPEG-7 and that are not yet formalized in the ontology. For example, a rule can check whether the start/end time of the segments involved in a temporal decomposition are compatible with the start/end time range of their encompassing segment. It has to be noted that the use of rules comes potentially with serious scalability problems, since the languages mentioned above are known to be undecidable. However, during our experiments, we did not enter in the bad reasoning cases and the overall impact of these particular rules from the complexity point of view should be further explored in the future.

5 Implementation and Results

To illustrate the methodology detailed above, we have developed a Web-based application that provides a semantic validation service for any DAVP descriptions. The running application is available at <http://mpeg-7.joanneum.at/Validator/>.

Figure 5 depicts the different components of the application.

- A first XSLT component takes an MPEG-7 document and pre-processes it to ensure that the input is compliant with respect to the DAVP schema (structural and XML-Schema-based syntactic validation);
- A second XSLT component takes the output of the first component and converts it into RDF-triples according to the DAVP OWL ontology;
- An inference service, such as the Jena Framework, takes as inputs the RDF-triples obtained previously and the additional logical rules modeling the semantic constraints of the DAVP profile in order to produce an inferred graph;

- Finally, a set of predefined SPARQL-queries can be executed on the resulting inferred graph in order to check the semantic validity of the MPEG-7 description.

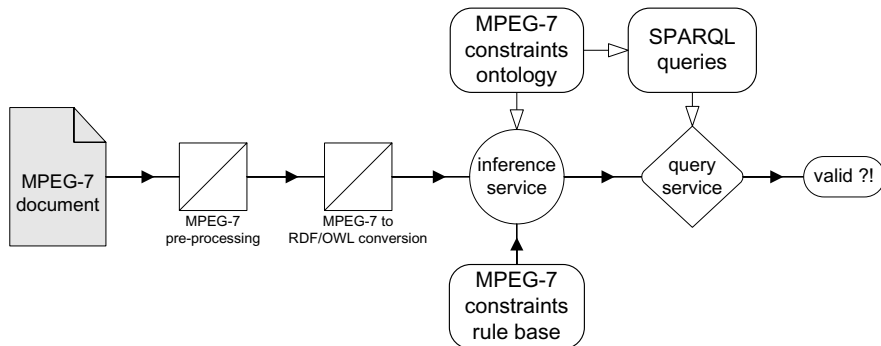


Fig. 5. Components and dataflow of the DAVP semantic validation application

Coming back to the scenario discussed in the section 2, we have pointed out that the problems caused by the unclear definitions of MPEG-7 description tools can be overcome by formalizing the semantic constraints defined in the profile. Hence, given an MPEG-7 description syntactically valid with respect to the DAVP schema, we want to validate semantically the description against the profile using the environment described above. We have shown in the previous sections how the semantic constraints related to the `Shot` and `KeyFrame` have been formalized. The top-level structure of the DAVP schema (the `Mpeg7` element and its children) has also been defined in the ontology⁴.

According to the layered approach advocated in the section 4.1, the first step is to check the wellformedness and the syntactic validity of a description with respect to the DAVP schema. An XML processor such as Xerces is used in our application for this task. The XML description is then converted into an RDF graph that contains the instances of the classes defined in the DAVP ontology. An OWL reasoner such as RacerPro⁵ or Pellet⁶ is then used to check the consistency of the graph. Finally, the RDF representation of the description is processed by the Jena rule engine, and the resulting RDF graph is checked with a SPARQL query for any incorrect annotations produced by the inference engine. Figure 6 shows the user interface of our semantic validation web-based application.

Let's assume that an MPEG-7 document given as input to the semantic validation service has already been found to be valid against the DAVP XML Schema.

⁴ The complete DAVP ontology as well as the rules modeled are available at <http://mpeg-7.joanneum.at/Validator>.

⁵ <http://www.racer-systems.com/>

⁶ <http://www.mindswap.org/2003/pellet/>

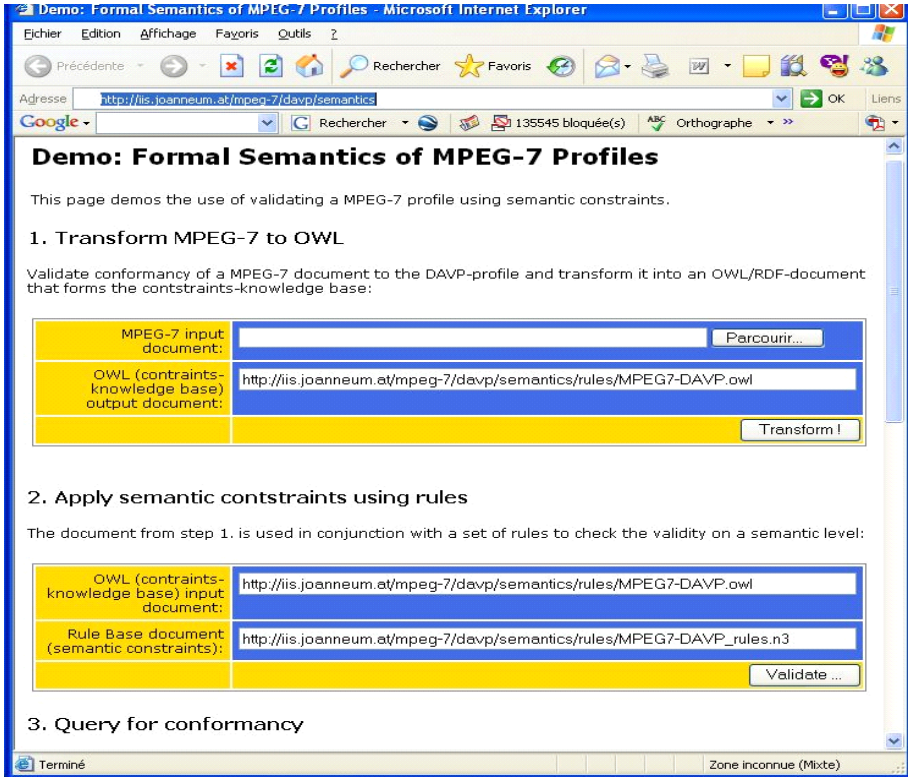


Fig. 6. Web-based application providing a semantic validation service for MPEG-7 descriptions

The following examples describe possible violations of the semantic constraints defined in the profile, while still being valid against the XML Schema:

- The OWL reasoner is able to detect “local” violations of the DAVP constraints such as a bad use of the `StructuralUnit` element on shots and key frames, or a key frame that is not contained in a shot description.
- An OWL reasoner cannot detect the missing `StructuralUnit` elements because of the open world assumption underlying to the OWL language. This error can be however detected by the Jena rule engine, thus motivating our choice to use additional rules for representing all semantic constraints of a profile.

Overall, this web-based application provides a semantic validation service for MPEG-7 descriptions. The times needed to perform the checks depends on the size and the complexity of the descriptions: it is quite fast (a few seconds) for some real MPEG-7 descriptions such as the ones produced by the TRECvid evaluation.

6 Conclusion and Future Work

In this paper, we have proposed an approach to overcome the interoperability problems that result from the lack of formal semantics of the MPEG-7 description tools by formalizing the semantic constraints of a profile. The approach is based on the definition of profiles, which are not just subsets of the MPEG-7 standard, but that also define a set of semantic constraints that specify the use of the description tools in a particular context. We have proposed to build an ontology that includes the concepts being described in the profile and to formalize the semantic constraints of the profile in terms of these concepts using description logic and rules. It was not our intention to completely represent the profile in the ontology, but only those constraints that cannot be validated using pure XML Schema tools. We have demonstrated the feasibility of this approach by modeling a subset of the DAVP profile, and we have set up a Web-based application that provides a semantic validation service for checking the conformance of any MPEG-7 descriptions against the semantic constraints of the profile.

As stated above, our approach requires the definition of semantic constraints on the description tools. One could argue that it is thus possible to apply this approach to the complete MPEG-7 standard, as there are many such constraints expressed in natural language in the standard. While this might work well for some of the description tools that have already a well-defined semantics, it is not the case for those generic Descriptors such as `VideoSegment`, which are the most problematic with respect to the interoperability problems discussed in Section 2. This means that a key prerequisite to make the proposed approach efficient is the definition of profiles that include sufficient semantic constraints to prevent ambiguities.

A next step in our work is to finalize the formalization of the semantic constraints for the whole DAVP profile. This task may require a stricter and more precise formulation of some of these constraints. There is of course always a trade-off between flexibility and strictness with respect to description tool semantics. If we require the semantic constraints to be very strict, this might prevent the use of any structures in the description not foreseen in the profile definition, even if they are used as extension and do not interfere with the structures defined in the profile. Thus it could be an option to introduce different levels of conformance to profile semantics, which one could name "semantic levels" of a profile.

Representing formally the semantic constraints of the MPEG-7 description tools is not only useful for semantically validating the descriptions as discussed in Section 5, but also for establishing mappings between profiles and heterogeneous MPEG-7 descriptions. Actually, the greatest potential with semantic definitions of MPEG-7 profiles is in the ability to use these descriptions to relate the content to other audiovisual streams/segments described using alternative MPEG-7 profiles or other domain ontologies. Current multimedia applications on the Web, such as multimedia search engines, need to index multimedia metadata from heterogeneous sources. Formalizing the semantics of the profiles used for representing this metadata allows the auto mapping of the descriptions, guess the

semantics of each audiovisual segments, etc. In the future, we plan to investigate further how the approach presented in this paper can be used in this particular use case.

Acknowledgments

The authors would like to thank the following colleagues at JOANNEUM RESEARCH (Georg Thallinger, Herwig Zeiner and Peter Schallauer) and at CWI (Lynda Hardman, Jacco van Ossensbruggen) for their feedback on the prototype and earlier versions of this paper. The research leading to this paper was partially supported by the European Commission under contract FP6-027026, "Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space", and under contract IST-2-511316, "IP-RACINE: Integrated Project - Research Area CINE".

References

1. Werner Bailer and Peter Schallauer. The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 based Systems. In *12th International MultiMedia Modelling Conference (MMM'06)*, pages 217–224, Beijing, China, 2006.
2. Werner Bailer, Peter Schallauer, Michael Hausenblas, and Georg Thallinger. MPEG-7 Based Description Infrastructure for an Audiovisual Content Analysis and Retrieval System. In *Proceedings of SPIE - Storage and Retrieval Methods and Applications for Multimedia*, pages 284–295, San Jose, California, USA, 2005.
3. Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Andrea Schaerf. ALlog: Integrating Datalog and Description Logics. *Journal of Intelligent Information Systems*, 10(3):227–252, 1998.
4. Roberto Garcia and Oscar Celma. Semantic Integration and Retrieval of Multimedia Metadata. In *5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot'05)*, Galway, Ireland, 2005.
5. Benjamin Grosf, Ian Horrocks, Raphael Volz, and Stefan Decker. Description Logic Programs: Combining Logic Programs with Description Logics. In *12th International World Wide Web Conference (WWW'03)*, Budapest, Hungary, 2003.
6. Ian Horrocks, Peter F. Patel-Schneider, Sean Bechhofer, and Dmitry Tsarkov. OWL rules: A proposal and prototype implementation. *Journal of Web Semantics*, 3(1):23–40, 2005.
7. Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosf, and Mike Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Technical report, World Wide Web Consortium, May 2004. <http://www.w3.org/Submission/SWRL/>.
8. Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen. From *SHIQ* and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics*, 1(1):7–26, 2003.
9. 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.

10. Jane Hunter and Carl Lagoze. Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles. In *10th International World Wide Web Conference (WWW'01)*, pages 457–466, Hong Kong, 2001.
11. Jane Hunter and Frank Nack. An overview of the MPEG-7 Description Definition Language (DDL) proposals. *Signal Processing: Image Communication*, 16(1-2):271–293, 2000.
12. Alon Y. Levy and Marie-Christine Rousset. Combining horn rules and description logics in carin. *Artificial Intelligence*, 104(1-2):165–209, 1998.
13. MPEG-7. Multimedia Content Description Interface. Standard No. ISO/IEC n° 15938, 2001.
14. 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.
15. OWL. Web Ontology Language Reference. W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/owl-ref/>.
16. Fernando Pereira. MPEG-7 Requirements Document V.16. ISO/IEC JTC1/SC29/WG11/N4510. Pattaya, Thailand, December 2001.
17. RDF. Ressource Description Framework Primer. W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-primer/>.
18. Raphaël Troncy. Integrating Structure and Semantics into Audio-visual Documents. In *2nd International Semantic Web Conference (ISWC'03)*, pages 566–581, Sanibel Island, Florida, USA, 2003.
19. Raphaël Troncy and Jean Carriève. A Reduced Yet Extensible Audio-Visual Description Language: How to Escape From the MPEG-7 Bottleneck. In *4th ACM Symposium on Document Engineering (DocEng'04)*, Milwaukee, Wisconsin, USA, 2004.
20. Chrisa Tsinaraki, Panagiotis Polydoros, and Stavros Christodoulakis. Interoperability support for Ontology-based Video Retrieval Applications. In *3rd International Conference on Image and Video Retrieval (CIVR'04)*, Dublin, Ireland, 2004.
21. 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.
22. XML Schema. W3C Recommendation, 2 May 2001. <http://www.w3.org/XML/Schema>.