A Multimedia Adaptation Framework based on Semantic Web Technology

Dietmar Jannach[1] and Klaus Leopold[2]

[1]Department of Business Informatics and Application Systems
[2]Department of Information Technology
University Klagenfurt
9020 Klagenfurt, Austria
{dietmar.jannach, klaus.leopold}@uni-klu.ac.at

Abstract. Intelligent, server-side *adaptation of multimedia resources* is becoming increasingly important and challenging for two reasons. First, the market continuously brings up new mobile end-user devices to which the content has to be adapted as these devices support different display formats and operate on various types of networks. On the other hand, with the help of metadata annotations which are now available in the MPEG-7 and MPEG-21 standard, advanced forms of resource adaptations *on the content level* become possible. As none of the existing multimedia transformation tools and libraries can support all these different forms of basic and advanced adaptation operations, an intelligent multimedia adaptation server has to integrate such external tools and algorithms and perform an adequate sequence of adaptation operations on the original resource before sending it to the client.

In this paper we present the results of the *ISO/IEC MPEG Core Experiment* on using Semantic Web Services technology as a tool for declaratively describing the semantics of adaptation services and constructing multi-step adaptation sequences in an open and extensible multimedia adaptation framework. We show how the semantics of adaptation operations can be captured in the form of input, output, precondition, and effects, how the problem of finding adequate adaptation sequences can be viewed as an AI planning problem, and finally, how the existing MPEG standards are technically integrated into the service descriptions and how they serve as the shared ontology of the domain. Our approach both introduces declarative, AI-based technology into to the involved multimedia communities and on the other hand broadens the application scope of Semantic Web Service technology in the area of general semantic service descriptions and automated program construction.

Introduction

Currently, a lot of different end user devices are connected to the Internet. One can find desktop PCs, notebooks, workstations, set-top boxes, TV sets, but also mobile devices such as PDAs, cell phones, and hand held devices using Internet services. All those devices feature different capabilities in terms of computational power, memory size, display size, or network capabilities. Displaying regular web content (HTML) on these devices is a task of manageable complexity. However, rendering multimedia content in such an environment remains challenging, because the content itself is heterogeneous in terms of encoding. For instance, a video can be encoded in different formats such as MPEG-1, -2, -4, H264, or WMV, using different encoder settings such as spatial and temporal resolution, color depth, or bit rate. As a consequence, the end user devices are in general not able to display all kind of multimedia data. Therefore, a lot of research work went towards *Universal Multimedia Access* [1], where any user (respectively device) can consume any multimedia content, anytime and anywhere. To enable UMA, the multimedia content has to be adapted before it is sent to the client so that it fits the consumer's terminal capabilities and usage environment.

MPEG-21 is an emerging ISO/IEC standard that aims at addressing these new challenges by defining a normative open framework for multimedia delivery and consumption involving all parties in the delivery and consumption chain. A major part of these standardization efforts deals with the definition of an interoperable framework for *Digital Item Adaptation* (DIA) [2]. Currently, the scope of MPEG-21 standards with respect to adaptation comprises mechanisms (normative description schemes) enabling interoperability by describing the adaptation problem, like the terminal capabilities, the user preferences, or the format of the resource itself. At the same time, new standards like MPEG-7 [3] have become available, allowing us to enrich media content with semantic content annotations, which in turn facilitates new forms of multimedia experience, like search on specific topics or semantics-based content selection and filtering. The implementation of an engine that actually performs the required adaptation steps to the multimedia content is left to tool vendors and is beyond the scope of standardization.

We argue, that it does not seem realistic that one single software tool will be able to perform all required adaptation steps for the various user preferences, terminal capabilities, network characteristics, or even for the diverse set of coding formats. Furthermore, an adaptation engine has to be open and extensible, such that no changes in the general mechanism are required when new forms of adaptation are possible as the standards evolve or new tools become available.

In this paper, we present a knowledge based multimedia adaptation framework which is able to automatically construct adaptation plans for multimedia resources to fulfill terminal constraints. In our approach, we incorporate third-party adaptation tools to meet adaptation goals and rely on standardized MPEG metadata as well as *Semantic Web Service* [4] technologies for capturing the semantics of these adaptation tools.

A Knowledge based Multimedia Adaptation Framework

In order to intelligently adapt multimedia resources according to the client's needs, a server-side adaptation engine has to support various media types like audio, video, and image with respect to different adaptation classes, like for instance, media transcoding, transmoding, or even content-based adaptation. As no single software piece exists that can handle the huge variety of multimedia adaptation methods, an intelligent adaptation engine has to be open and extensible in order to integrate already existent adaptation tools. Furthermore, standards emerge and additional coding formats and forms of adaptation will become possible. Consequently, the mechanism for computing the required adaptation operations should be based on declarative specifications in order to be robust and extensible with respect to such additions.

The KOMMA project [7] carried out in official *ISO/IEC Core Experiments* [9, 10] addresses these new requirements in a novel, *knowledge-based* multimedia adaptation framework. The framework consists of two major parts, the *adaptation decision taking engine* and the *adaptation engine* (Figure 1). The adaptation decision taking engine is responsible for finding a suitable sequence of transformation steps – the *adaptation plan* – that can be applied on a multimedia resource. The adaptation plan is then forwarded to the adaptation engine which performs the actual transformation steps on the multimedia resource. In parallel, also the accompanying (MPEG-7) content descriptions are transformed adequately.

We view the finding of a suitable transformation sequence as a typical state space planning problem (see e.g. [8]), where actions are applied on an initial state to reach a goal state. In our domain, the *start state* corresponds to the original multimedia resource which is specified by means of MPEG-7 metadata descriptions. The *goal state* is an adapted version of the multimedia resource which suits the client's needs and preferences which are expressed using MPEG-21 documents. *Actions* are the adaptation methods that can be applied on the multimedia content and which are specified in terms of *inputs*, *outputs*, *preconditions*, and *effects* (IOPE).

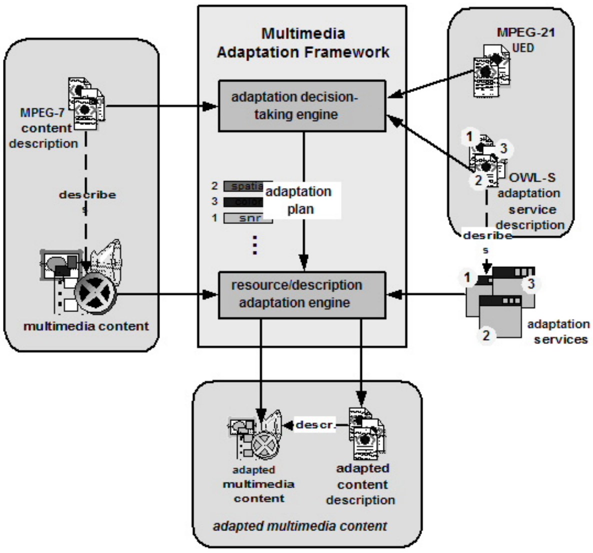

Figure 1: The KOMMA multimedia adaptation framework

The following simplified example shows the description of the start state and the goal state of a multimedia adaptation problem where an image has to be spatially scaled and the color has to be removed. Note that for the sake of readability we use an informal notation rather than the internal XML representation.

The start state – which is the MPEG-7 description of the existing resource – can be described as follows:

jpegImage(http://path/to /image.yuv), width(640), height(480), color(true).

The goal state – which is the MPEG-21 UED – can be described as follows:

jpegImage(file://path/to/image.yuv), horizontal(320), vertical(240, color(false)

The following example shows the description of the spatial scaling operation for images based on the IOPE approach.[1]

Operation: *spatialScale*

Input:	*imageIn, oldWidth, oldHeight, newWidth, newHeight*
Output:	*imageOut*
Preconditions:	*jpegImage(imageIn), width(oldWidth), height(oldHeight)*
Effects:	*jpegImage(imageOut), width(newWidth), height(newHeight), horizontal(newWidth), vertical(newHeight)*

The adaptation framework features standard read and write operations used to read images from input sources and write images to output sources. The computed adaptation plan of the adaptation decision taking engine may look like as follows:

1. *read(http://path/to/image.yuv, outImage1)*
2. *spatialScale(outImage1, 640, 480, 320, 240, outImage2)*
3. *greyscale(outImage2,outImage3)*
4. *write(outImage3, file://path/to/output/image.yuv)*

By using the IOPE approach for modeling the functionality of adaptation services, our engine remains flexible, since the core planner operates on arbitrary symbols, such that new adaptation services can be easily added without changing the implementation. Moreover, IOPE-style descriptions have shown to be are expressive enough for a wide range of problem domains. Another interesting aspect in our problem domain is that the level of detail of the functional descriptions of available adaptation services can vary. In the example given, each action is an atomic, single-step picture transformation. This fine-granular specification is reasonable in cases when, e.g., open-source transformation software can be used in the adaptation engine. In this scenario, the adaptation chain and the execution plan is composed of API calls to a local media processing library. On the other hand, as newer standards for semantic content annotation like MPEG-7 are increasingly established, specialized software companies provide advanced adaptation functionality as service for their clients. With the approach described in this paper, however, the potential distributed nature of the individual services is transparent for the engine.

To summarize, the framework is presented in this paper is capable of computing and executing multi-step adaptation sequences based on semantic descriptions of the available transformation operations. We use OWL-S for representing inputs, outputs, preconditions, and effects and a Prolog-based planning engine that interprets these descriptions and produces adequate adaptation plans. For interoperability reasons, we view the existing MPEG standards as the shared domain ontology, which for instance defines which terms can be used in the IOPE descriptions. In the subsequent sections, we will show how this integration of the involved technologies (MPEG/OWL-S) was implemented on the technical level.

[1] We omit the similar grey-scaling specification for sake of brevity.

MPEG Standards as Domain Ontology

The MPEG-7 and MPEG-21 DIA standards form a precise way of how multimedia resources and usage environments are described by using XML Schema technology. Document 1 shows a fragment of how a video resource is annotated in MPEG-7; in particular it specifies the resolution of an image, which we used in the example.

```
<Mpeg7>
  <Description xsi:type="ContentEntityType">
    <MultimediaContent xsi:type="VideoType">
      <Video>
        <MediaInformation id="news1_media">
          <!-- MediaIdentification ... -->
          <MediaProfile>
            <MediaFormat>
              <!-- Content, Medium, FileFormat, Filesize ... -->
              <VisualCoding>
                <Format href=" urn:mpeg:mpeg7:cs:FileFormatCS:2001:1"
                     colorDomain="color">
                  <Name xml:lang="en">
                     JPEG image</Name>
                </Format>
                <Pixel aspectRatio="0.75" bitsPer="8"/>
                <Frame width="640" height="480" rate="25"/>
              </VisualCoding>
            </MediaFormat>
          </MediaProfile>
        </MediaInformation>
      </Video>
    </MultimediaContent>
  </Description>
</Mpeg7>
```

Document 1. Fragment of an MPEG-7 video description

Although not explicitly mentioned as such, the definitions in the standards implicitly form a precise ontology for the multimedia domain. The dimensions and corresponding syntactical structures in which a resource can be annotated are strictly defined. The intended semantics of the terms is specified in natural language. For the examples above, the *colorDomain* attribute describes the color domain of the video with a frame size of 640×480 pixels indicated by the *width* and *height* attributes. The resolution and color capabilities of the rendering device are described by the *horizontal* and *vertical* attributes and the *colorCapable* attribute respectively.

Note that we deliberately do not propose to develop an explicit (OWL-based) ontology of the multimedia domain. We argue that the acceptance of such an ontology would be limited among service providers because it would lead to the fact that the same things have to be expressed and maintained both in the MPEG XML standard documents and in OWL. Thus, we rather aim at viewing the existing standards as the domain ontology and smoothly integrating the semantic service descriptions on the technical level with the help of standard XML-technology.

Nonetheless, the creation of an additional ontology as described in [8] for improving the quality and effectiveness of the adaptation plans, would be an interesting extension of our framework. With the help of such additional domain knowledge, we could for instance model domain rules and heuristics that help us to optimize the transformation sequence by e.g., reducing the image size before doing other operations.

Semantic Multimedia Adaptation Services

Within the ISO/IEC Core Experiments, we showed how adaptation services can be described and executed with Semantic Web Service technology by using OWL-S [5] together with SWRL [6] to specify adaptation services. OWL-S provides a computer-interpretable description of services and the means by which they can be accessed. It comprises mechanisms to describe the functionality of a service in terms of a description of the transformation that is caused when a service is invoked. In particular, with OWL-S one can declaratively describe the *inputs*, *outputs*, *preconditions*, and *effects* (IOPE) of a service. The structuring of the OWL-S description is motivated by the need to provide three essential types of knowledge about a service: profile, process, and grounding. The *service profile* specifies what the service requires from the user or agent and tells us what it provides for them. The *process model* specifies how the service works. One can consider the profile as the interface for the service, whereas the process represents its internal structure. Finally, the *service grounding* specifies the details of how an agent can access the service. In this paper we only present the process model because it is the most crucial part of the adaptation service description. SWRL provides an XML format for describing logical facts and rules and can be used within OWL-S for specifying the logic of preconditions and effects.

Document 2 shows fractions of an OWL-S description for the spatial scale service of our example. It basically consists of an atomic process called *SpatialScale* and its properties. First, the input parameters are defined by mapping each parameter to an MPEG term in order to specify its meaning (i.e., the semantics). For instance, the parameter *imageIn* is mapped to the MPEG-7 term *urn:mpeg:mpeg7:cs:FileFormatCS:2001:1* which means that *imageIn* is a JPEG image. The adaptation decision taking engine would interpret this parameter mapping as precondition for the planning problem because it tells us that the parameter imageIn *has to be* a JPEG image. The effects of the spatial scale service are represented as a list of SWRL atoms. In our example, the parameter *newWidth* is mapped to both the MPEG-7 *width* term *urn:mpeg:mpeg21:01-DIA-MediaInformationCS-NS:81* and to the MPEG-21 *horizontal* term *urn:mpeg:mpeg21:2003:01-DIA-AdaptationQoSCS-NS:6.5.9*. The mapping to the MPEG-21 term tells the adaptation decision taking engine that the parameter has to meet the MPEG-21 terminal constraint. Furthermore, the mapping to the MPEG-7 *width* term signals that the original MPEG-7 descriptor is no longer valid and has to be updated according to the MPEG-21 horizontal term.

At the moment, only a small fraction of the SWRL language capabilities is used in our framework. Basically, we are using literals (see Document 2) for describing the service semantics in terms of IOPE which we then automatically transform into the internal Prolog-based representation that is used by the planner. Our experiments show that the expressiveness of this restricted language is sufficient to describe important classes of possible transformation services. Nonetheless, we are currently working on extending the expressiveness of the language to cover more of the available SWRL constructs, which will however require the implementation of an own SWRL engine. Still, the usage of SWRL even for simple literal-based statements has the advantage that it already conforms to the evolving standards and is embedded in the OWL-S framework.

```xml
<?xml version="1.0" encoding="ISO-8859-1"?> ...
<rdf:RDF> ...
  <owl:Ontology rdf:about="">
    <owl:versionInfo>$Id: Spatial Scaler V 1.301, KL</owl:versionInfo> …
  </owl:Ontology>   <!--Process-->
  <process:AtomicProcess rdf:ID="SpatialScale">
    <!-- The parameters of the process -->
    <process:hasInput>
      <process:Input rdf:ID="imageIn">
        <process:parameterType rdf:datatype="&xsd;#anyURI">
          <!-- JPEG image -->
           urn:mpeg:mpeg7:cs:FileFormatCS:2001:1
        </process:parameterType>
      </process:Input>
    </process:hasInput>
    <process:hasInput>
      <process:Input rdf:ID="oldWidth">
        <process:parameterType rdf:datatype="&xsd;#anyURI">
          <!-- reference to MPEG-7 width -->
          urn:mpeg:mpeg21:01-DIA-MediaInformationCS-NS:81
        </process:parameterType>
      </process:Input>
    </process:hasInput>
    ... <!-- more input parameters here --> ...
    <process:hasOutput>
      <process:Output rdf:ID="imageOut">
        <process:parameterType rdf:datatype="&xsd;#anyURI">
          <!-- JPEG image -->
            urn:mpeg:mpeg7:cs:FileFormatCS:2001:1
        </process:parameterType>
      </process:Output>
    </process:hasOutput>
    <process:hasResult>
      <process:Result rdf:ID="eff_resolution">
        <rdfs:comment> Effects are changed MPEG-7 width and height
              and meet MPEG-21 hor. and vert.
        </rdfs:comment>
        <process:hasEffect>
          <expr:SWRL-Expression>
            <expr:expressionBody rdf:parseType="Literal">
              <swrlx:AtomList>
                <rdf:first>
                  <swrlx:classAtom> <!-- MPEG-7 width -->
                    <owlx:Class
                      owlx:name="urn:mpeg:mpeg21:01-DIA-MediaInformationCS-NS:81"/>
                    <owlx:Individual owlx:name="#newWidth"/>
                  </swrlx:classAtom>
                </rdf:first>
                <rdf:rest>
                  <swrlx:AtomList>
                    <rdf:first>
                      <swrlx:classAtom> <!-- MPEG-21 horizontal -->
                        <owlx:Class
               owlx:name="urn:mpeg:mpeg21:2003:01-DIA-AdaptationQoSCS-NS:6.5.9.1"/>
                        <owlx:Individual owlx:name="#newWidth"/>
                      </swrlx:classAtom>
                    </rdf:first>... <!-- more effects here --> ...
                  </swrlx:AtomList>
                </rdf:rest>
              </swrlx:AtomList>
            </expr:expressionBody>
          </expr:SWRL-Expression>
        </process:hasEffect>
      </process:Result>
    </process:hasResult>
  </process:AtomicProcess>
```

Document 2. OWL-S process description for a spatial scale adaptation service

Concluding remarks

Within the *Core Experiment* we developed a prototype system implementing the described combination of MPEG-metadata, Semantic Web Service technology, and AI-planning. This prototype system is described in more detail in [10] and the experiments indicate that the presented approach is practicable also for more complex adaptation problems. Currently the work is under review in the current MPEG-21 DIA standardization process.

Overall, we view our work as a first step towards bringing two different research areas together that both aim at adding more *semantics* to the contents (and services) available on the Internet, as we introduce knowledge-based technology to the MPEG community and on the other hand recognized another promising application area for Semantic Web Services technology in a domain where a shared domain ontology forms a solid basis for interoperation and automated program construction.

References

1. F. Pereira and I. Burnett, “Universal Multimedia Experiences for Tomorrow”, IEEE Signal Processing Magazine, pp 63 – 73, March 2003
2. A. Vetro and C. Timmerer, “Overview of the Digital Item Adaptation Standard”, IEEE Transactions on Multimedia, Special Issue on MPEG-21, 2004
3. J. M. Martinez, R. Koenen, and F. Pereira, “MPEG-7 - The Generic Multimedia Content Description Standard, Part 1”, IEEE MultiMedia 9(2):78–87, April–June 2002.
4. S. McIlraith, T.C. Son, and H. Zeng, “Semantic Web Services”, IEEE Intelligent Systems, Special Issue on the Semantic Web, 16(2):46–53, 2001.
5. D. Martin et al, “OWL-S: Semantic Markup for Web Services”, available under: http://www.daml.org/services/owl-s/1.1/overview/
6. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission 21 May 2004, I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B.Grosof, M. Dean, http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/.
7. D. Jannach, K. Leopold, Christian Timmerer, and H. Hellwagner, “A knowledge-based framework for multimedia adaptation.”, *to appear in:* Applied Intelligence, Springer, 2005.
8. I. Bratko, *Prolog Programming for Artificial Intelligence*, 3rd Edt., Addison-Wesley 2000.
9. C. Timmerer, K. Leopold, D. Jannach, H. Hellwagner, Mariam Kimiaei, “Report of CE on Conversion Parameters”, ISO/IEC JTC 1/SC 29/WG 11 M11262, Palma De Mallorca, Spain, October 2004
10. K. Leopold, C. Timmerer, “CE Report on Conversions”, ISO/IEC JTC 1/SC 29/WG 11 M11742, Hong Kong, China, January 2005
11. P. Soetens and M. De Geyter, "Applying domain knowledge to multi-step media adaptation based on Semantic Web Services", Proceedings of Workshop on Image Analysis for Multimedia Interactive Systems (WIAMIS), Montreux, April, 2005.