Appl Intell (2006) 24: 109–125
DOI 10.1007/s10489-006-6933-0

A knowledge-based framework for multimedia adaptation

**Dietmar Jannach · Klaus Leopold ·
Christian Timmerer · Hermann Hellwagner**

Abstract Personalized delivery of multimedia content over the Internet opens new business perspectives for future multimedia applications and thus plays an important role in the ongoing MPEG-7 and MPEG-21 multimedia standardization efforts. Based on these standards, next-generation multimedia services will be able to automatically prepare the digital content before delivery according to the client's device capabilities, the network conditions, or even the user's content preferences. However, these services will have to deal with a variety of different end user devices, media formats, as well as with additional metadata when adapting the original media resources. In parallel, an increasing number of commercial or open-source media transformation tools will be available, capable of exploiting such descriptive metadata or dealing with new media formats; thus it is not realistic that a single tool will support all possible transformations.

In this paper, we present a novel, fully knowledge-based approach for building such multimedia adaptation services, addressing the above mentioned issues of openness, extensibility, and concordance with existing and upcoming standards. In our approach, the original media is transformed in multiple adaptation steps performed by an extensible set of external tools, where the construction of adequate adaptation sequences is solved in an Artificial Intelligence *planning process*. The interoperability issue is addressed by exploiting standardized *Semantic Web Services* technology. This technology allows us to express tool capabilities and execution semantics in a declarative and well-defined form. In this context, existing multimedia standards serve as a shared domain ontology.

The presented approach was implemented and successfully evaluated in an official *ISO/IEC MPEG (Moving Picture Experts Group) Core Experiment* and is currently under further evaluation by the standardization body.

D. Jannach (✉) · K. Leopold · C. Timmerer · H. Hellwagner
Universität Klagenfurt, A-9020 Austria
e-mail: {Dietmar.Jannach, Klaus.Leopold, Christian.Timmerer, Hermann.Hellwagner}@uni-klu.ac.at

Keywords Multimedia applications · Knowledge-based systems · Semantic web services · Planning

1. Introduction

The importance of multimedia content provisioning over the Internet has been steadily increasing over the last years. The most prominent application domains can, for instance, be found in the entertainment sector and in the areas of distance education and online training. Quite naturally, the field will continuously evolve as bandwidth limitations will be decreasing and new (mobile) end user devices will open new opportunities for multimedia-based applications. In a traditional multimedia content provisioning architecture, the content, e.g., a video, is stored on one or more servers or network nodes and streamed to the client in a certain data format and encoding. However, the variety of the new end user devices, user preferences, and other application-specific requirements like mobility show that future multimedia environments must be more flexible and adapt the content to the client's needs when transferring it over the network [5].

Figure 1 sketches this general trend towards increased heterogeneity in multimedia content delivery over the Internet, where the end-user devices significantly differ both w.r.t. their features and capabilities (i.e., which content can be consumed in which form and quality) as well as in the underlying types of networks they are using (like LAN, WLAN, GPRS, UMTS, etc.).

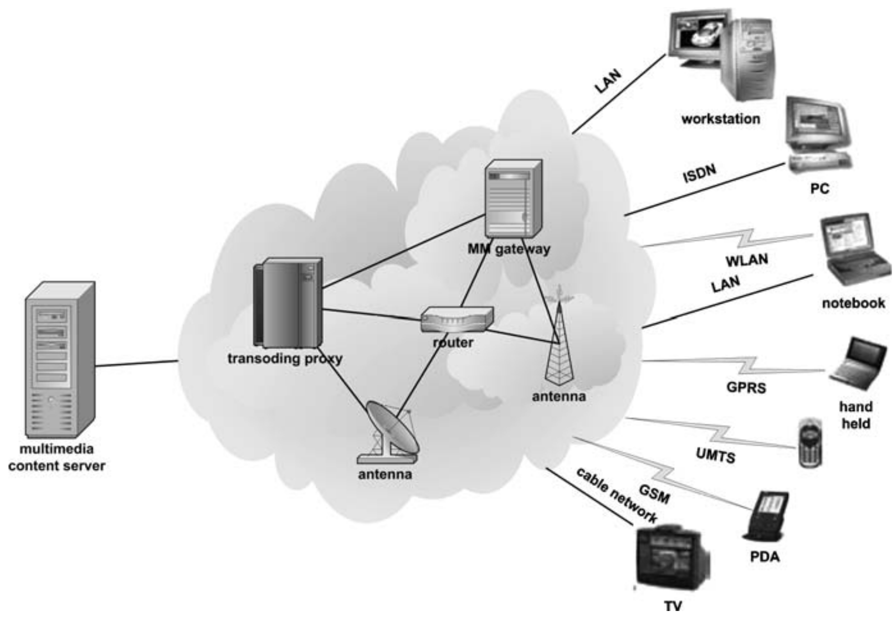

Fig. 1 Multimedia provisioning over the Internet

The following typical scenario shall illustrate why intelligent server-side adaptation of resources is crucial in such a heterogeneous environment. Typically, high quality video streams are stored at multimedia servers in a given format. In a situation, however, where the server receives a request from a mobile user that uses a PDA (Personal Digital Assistant) and a low-bandwidth wireless connection, it is of course not reasonable to transfer the video in high quality or resolution, respectively, and leave the adaptation to the client. First, transferring the large video file or "fat" stream would cause higher network traffic while at the same time increasing the user's expenses when following a pay-per-byte accounting model. In addition, client-side transformation would consume a large portion of the limited processing power of the handheld device.

Thus, an intelligent network node like a video server or proxy will transform the video into an appropriate format before sending or forwarding it to the client. The needed transformation steps typically include scaling (e.g., spatial adaptation of the presentation size, quality reduction, or grey-scaling), transcoding (e.g., MPEG-2 to MPEG-1), or modality conversion (e.g., video-to-image or image-to-text) techniques.

Note that besides such basic transformations on the format or encoding level, future forms of multimedia consumption will require even more complex types of adaptation. A typical example is related to the *personalization* of the content, e.g., based on MPEG-7 metadata [22]. This will, for instance, include semantics-based selection or removal of scenes from a movie or the selection of certain channels from a live broadcast.

Another example can be found in the context of *session mobility*, where active user sessions need to be maintained while being switched between terminals. In such a scenario, the adaptation strategy of the server has to be adjusted even during a session, when the user, for instance, wants to transfer the current video stream from a PC to a PDA.

These new challenges are being addressed by the ISO/IEC standardization body in the emerging MPEG-21 standard [9], which aims at defining a normative open framework for interoperable multimedia applications involving all parties in the multimedia delivery and consumption chain. *Digital Item Adaptation* [18, 36] is a major part of this standardization effort and provides mechanisms (normative description schemes) that enable describing the adaptation problem, like the terminal capabilities, the user preferences, or quality of service (QoS) information pertaining to the media resources. The implementation of the engine that actually performs the required adaptation steps is left to tool vendors and is thus beyond the scope of standardization.

There is, however, a specific requirement in that context that is only marginally addressed in the current standard, i.e., the multimedia framework "*... shall support a convenient and efficient way to concatenate multiple adaptation steps*" [7]. Therefore, in order to enable efficient treatment and automatic construction of such transformation sequences by an adaptation engine, we argue that mechanisms are needed to express the *semantics* of the individual adaptation steps. Note that it does not seem realistic that a 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. Thus, we need a mechanism that is, on the one hand, expressive enough for describing such adaptations steps and is, on the other hand, independent from their actual implementation. Then, interoperability among tools of different vendors would become possible. In addition, an open and extensible approach has to be chosen, such that changes in the general mechanism are not required when new forms of adaptation emerge or new tools become available.

In this paper, we present a novel, fully *knowledge-based* framework for addressing the above-mentioned challenges. Our approach adds declarative knowledge representation, AI-based planning, and interoperable XML-based services to the MPEG-21 Digital Item Adaptation framework. The main features of this framework and the contributions of this work can be summarized as follows.

"Planning" adaptation sequences. We view the problem of determining and executing adequate adaptation sequences as a classical state-space planning problem (see, e.g., [8]). Starting with the description of a given multimedia resource, the goal is to apply a set of transformation operations on the resource such that the goal state is reached, in which the media is finally converted into the format which maximizes the user's experience for the given environment. Thus, a knowledge-based approach will allow us to build adaptation "services" that are capable of dealing with arbitrary sets of transformation operations, media descriptions, and user requirements.

Interoperable adaptation "services". In order to address the challenges of openness and interoperability with existing and future tools and standards, we argue that *Semantic Web Services* (see, e.g., [24]) can serve as the key enabling technology in that context. The main challenge in the described problem setting is to find a common way (among tool vendors and content providers), for the content itself, the adaptation tools, and the *semantics* of the execution of transformation operations to be described, which is a pre-requisite for the automated construction of meaningful adaptation sequences. As will be described later in this paper, the new W3C-standardized[1] languages from the field of Semantic Web Services like OWL-S[2] exactly serve this purpose: they aim at providing a generic and well-defined mechanism for describing the semantics of method execution - in their case for Web services - in terms of inputs, outputs, preconditions, and effects.

Integration into MPEG-21. Besides a common language for knowledge-representation, also a shared *domain ontology* is needed, such that these descriptions are compatible, which is currently a major challenge in most Semantic Web Service applications. In our application domain of multimedia provisioning, however, such a shared definition of terms and concepts already exists in terms of the existing MPEG-7 and MPEG-21 multimedia standards.

The remainder of this paper is organized as follows. In Section 2 we present an example problem, discuss the proposed planning approach, and give an architectural overview of an experimental prototype implementation. Details of the planner component and planning technology are discussed in Section 3. In Section 4 we show how OWL-S can serve as a knowledge representation mechanism for capturing the semantics of external transformation tools and algorithms and how existing MPEG standards serve as a shared domain ontology. By discussing the results of an official *ISO/IEC MPEG Core Experiment*, we show in Section 5 how our work has been incorporated into the existing MPEG-21 Multimedia Framework enabling interoperability among different applications or vendors. In Section 6 we discuss related work in the area. Finally, Section 7 concludes and gives an outlook on future extensions.

2. Architectural overview and example problem

Figure 2 shows an overview of the architecture of an intelligent multimedia adaptation server. We use the following MPEG terminology for describing the individual components: a client requests a certain content (*multimedia resources*) from a multimedia server and also provides a declarative description of its *usage environment*, i.e., device capabilities, network conditions, and user preferences. This description is contained in a Usage Environment Description (UED) whose specification is part of the MPEG-21 Digital Item Adaptation (DIA) standard [18, 36]. Together with the multimedia resources, the media server stores descriptions of the content and the available adaptation tools. The *content descriptions* are expressed in terms of MPEG-7 [22] media information metadata and contain information about, e.g., media encoding, color, or spatial resolution.

All these different types of metadata serve as the input for the *adaptation decision-taking engine* which produces a sequence of appropriate transformation steps, i.e., an adaptation plan. The adaptation plan is then forwarded to the *adaptation engine* which executes the required transformation steps on the original resource according to the adaptation plan and, thus, produces the adapted multimedia resource (see Fig. 2).

As one major novelty, the process of constructing the adaptation plan is in our framework modeled as a classical state-space planning problem (see, e.g., [8, 13]), where the description of the existing multimedia resource is the *start state*, the functionality provided by existing media transformation

[1] World Wide Web Consortium, see http://www.w3c.org

[2] OWL-based Web Service Ontology, see, e.g., http://www.daml.org/services/owl-s/

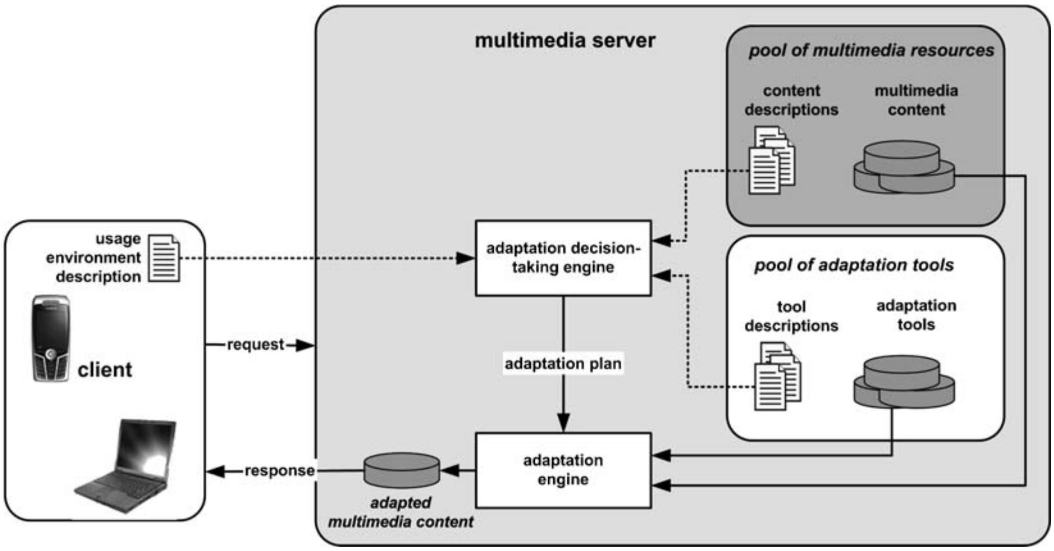

Fig. 2 Architecture overview

tools corresponds to world-altering *actions*, and the usage environment properties can be mapped to the *goal state*.

The following simple example illustrates our approach and sketches how transformation steps, i.e., the possible actions of a plan, can be described in terms of inputs, outputs, preconditions, and effects (IOPE) and how a suitable adaptation plan for a given problem is constructed. For sake of better readability and comprehensible semantics, we will use a simplified logic notation instead of the XML representation used by MPEG. The actual XML representation for spatial scaling will be shown in Section 5. In the example, an *MPEG-4 Advanced Simple Profile* encoded color video with a certain resolution (640 × 480) is stored on a server. The existing resource can thus be described by the following logical facts[3]:

coding_format(*mpeg_*4*_ASP*).
color_domain(*true*).
frame_size(640, 480).

When a client (e.g., some sort of PDA) sends a request for the video, it also transmits a description of its terminal's capabilities (MPEG-1 decoding, no color, small display size).[4]

decoding_capabilities(*mpeg_*1).
color_capabilities(*false*).
display_size(320, 240).

Let us assume that the adaptation steps *grey scaling* and *spatial scaling* are available—among others, such as encoding and decoding—in an existing image transformation toolkit (e.g., software library) and Application Programming Interface (API). Note that the adaptation operations as described in the following, are performed on a single image of a video. The decomposition of a video into frames is done by another tool executed by the adaptation engine. The inputs, outputs, preconditions, and effects can thus be described as shown in Fig. 3.[5]

A corresponding plan can be computed by a standard state-space planner and is also listed in Fig. 3, where "fb1" to "fb5" ("frame buffer") are the unified variables.

Note that when multiple steps are required, a handle to the partially adapted content has to be forwarded from one adaptation tool to the next. The appropriate forwarding of (pointers to) the partially adapted resource is also part of the plan, e.g., the output of step 1 of the plan (*decode*) is contained in variable "fb2," which is then used as an input in the next step (*spatialscale*). During plan execution—which involves the invocation of (external) tools—this information is then automatically evaluated to correctly parameterize the transformation tools.

In general, using the IOPE approach and a logic language for modeling the functionality of adaptation services has two main advantages. First, the general approach is simple to comprehend and has a defined, commonly agreed semantic such that the automated composition of action sequences is possible. Second, the approach is flexible, since the core planner operates on arbitrary symbols, such that new types of predicates or new actions can be easily added when they are available without changing the implementation.

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,

[3] These descriptions are contained in MPEG-7 metatdata.

[4] These usage environment descriptions are expressed using MPEG-21 metadata formats.

[5] In fact there is actually no means in current MPEG standards to express this sort of knowledge.

Grey scaling (greyscale)

Input:	image
Output:	greyImage
Preconditions:	yuvImage(image), color(true).
Effects:	yuvImage(greyImage), color(false).

Spatial scaling (spatialscale)

Input:	image, x, y, newx, newy
Output:	scaledImage
Preconditions:	yuvImage(image), width(x), height(y)
Effects:	yuvImage(scaledImage), width(newx), height(newy), horizontal(newx),vertical(newy).

1:	decode(fb1,mpeg_4_ASP,fb2)
2:	spatialscale(fb2,640,480,320,240,fb3)
3:	greyscale(fb3,fb4)
4:	encode(fb4,mpeg_1,fb5)

Fig. 3 An example transformation plan

each action is an atomic, single-step picture transformation. This fine-granular specification is reasonable in cases when, e.g., open-source transcoding software like FFMPEG or ImageMagick[6] should be used in the adaptation engine. In such a scenario, the adaptation chain and the execution plan is composed of API calls to a local media processing library, showing that existing libraries can easily be integrated. On the other hand, as newer standards for semantic content annotation like MPEG-7 are increasingly established in commercial environments, specialized software companies will be able to provide such advanced adaptation functionality as (pay-)services for their clients. With the approach described in this paper, however, the potential distributed nature of the individual services is transparent for the adaptation engine.

In our specific problem setting, the IOPE modeling approach can also serve as a means for addressing further domain-specific challenges: First, as sketched in the example, frame width and height are used for describing the size of the existing resource; the client's terminal capabilities, however, are described in terms of the display size, i.e., vertical and horizontal resolution as will be shown later in the MPEG XML fragments. Such a situation is quite common in the problem domain, as for instance different multimedia standards are involved. However, we can address this problem as described in the example, i.e., the ontological mapping between those terms or other (mathematical) relations can be explicitly modeled using preconditions and effects of the actions.

As a second side aspect, adaptation is not necessarily limited to the transformation of the media resource itself, but the same mechanism can be utilized to transform the accompanying (semantic) content descriptions, if this is needed by the client. In other words, in the adaptation process also the metadata may be updated correctly.

3. The planner component—implementation and evaluation

3.1. Implementation

In order to demonstrate the general feasibility of our approach, we have implemented a prototype system that supports all the required steps in the process in basic form. This includes parsing and analysis of the available MPEG metadata documents, plan construction and execution, and an editor for semantically annotating the available transformation tools.

At the core of this prototype system developed in the *ISO/IEC MPEG Core Experiment*, we use a light-weight Prolog realization of a standard means-ends planner with goal regression (see, e.g., [8]). Upon a client's request for a multimedia resource, the first task of the surrounding framework (implemented in Java) is to set up the problem space for the planner. The set of available transformation tools that correspond to *planning actions* remains stable over a

[6] http://ffmpeg.sourceforge.net, http://www.imagemagick.org

longer period of time and can be pre-loaded once upon server startup. Therefore, the main task is to parse and evaluate the MPEG-compliant metadata containing the resource descriptions as well as the usage environment that accompany the request. The information contained in those descriptions is then converted into Prolog facts that represent the start and goal states of the problem. We use a straightforward translation scheme that uniquely maps the metadata elements to Prolog predicates by exploiting the XML tree structure. An example of such metadata documents can be found in Section 4.

The actual invocation of existing external tools like the open source transcoding tools FFMPEG or ImageMagick can be performed in our current framework in three different forms. The most interoperable form is based on a *Web Service* interface (see later sections) which represents a programming-language independent invocation format. In addition, our framework is able to invoke locally available tools that implement a defined Java interface. When integrating such tools, we use appropriate wrapper classes that are dynamically loaded by the adaptation server at run-time. Finally, a dynamic invocation mechanism for fast C/C++ implementations with some restrictions has been implemented (see [20]).

In summary, the main advantage of the declarative planning approach lies in the fact that the reasoning engine is able to work with arbitrary sets of predicate symbols. Consequently, the introduction of new types of predicates, i.e., usage environment or content descriptions, does not require modifications in the planning algorithm. Therefore, the approach is robust against typical changes in the problem domain when, e.g., new media formats, codecs, or standardized metadata become available over time.

As a matter of fact, the employed search algorithm is provably sound and complete and will find a transformation plan if such a plan exists. Although we cannot automatically check the correctness of the *semantic* descriptions of transformation tools in general, we can at least check whether the set of predicate symbols is valid, i.e., that the used predicate symbols correspond to terms that are defined in the domain ontology formed by the MPEG standards. Nonetheless, in the error case when undefined symbols are used in an action description, the planner will produce no plans that contain such actions and rather try to reach the goal state with the help of other transformation actions.

3.2. Evaluation

The major aspects to be considered in the evaluation of the presented system are *running-times for plan construction* and the *quality of the produced plans*. We deliberately did not consider the time needed for performing a single step of the actual low-level adaptation of the resources themselves, as this is done by external, highly optimized tools which is not within the scope of our work. As such, it is not relevant for our purposes, whether we evaluate the system with simple operations only (like grey-scaling or spatial-scaling as in our examples) or complex, time-consuming transformations like decoding and re-encoding of whole video streams.

If we look at the plan construction process, the complexity of state-space planning problems is well-understood and has been discussed in detail, e.g., in [10]. In the context of our work, the actual complexity of the search problem therefore mainly depends on the length of the required plans and on the number of available actions. With respect to the first issue, the experiments we made so far suggest that the typical *length of the required transformation sequences* remains rather short, i.e., it generally does not exceed five to seven steps. In many cases, complex transformation operations (like content selection) are typically done by a single external tool and thus appear only as a single step in the adaptation plan.

On the other hand, we expect that also the number of available actions (i.e., external transformation operations) will be limited in concrete instantiations of the system, as we do not assume that alternative tools will be embedded for performing a certain type of resource modification. As a result, in our tests that involved around twenty different transformation actions, plans of such a length could be computed in a few milliseconds for all cases where a plan exists. Furthermore, the non-optimized planner implementation needs less than one second to determine that no such plan exists. Note that in general there are no hard real-time restrictions for plan generation in our application scenario because once a transformation plan is computed it is valid for a long time during video streaming. The costly operations are typically the transformation itself or the establishment of streaming sessions, in particular when network conditions change or session mobility is involved, i.e., when during resource consumption a session is migrated to another end user device.

The evaluation of further examples dealing with other real-world problems is part of our ongoing work and will be done in particular in the context of the envisioned embedding of our framework in a distributed multimedia system including content authoring, retrieval, delivery, and personalization.

Nonetheless, at the current state we can already mention the following possible performance improvements for our implementation, which were beyond the scope of our first prototype.

Planning time. First, there have been significant advances in the field of AI-based planning over the last years which made the state-space planning approach suitable for large real-world problems (see, e.g., [4, 6]. Consequently, the running times for plan construction in our framework can be strongly reduced without conceptual changes solely by the usage of more elaborated planning algorithms. The integration of such a high-performance planning engine and the usage of the de-facto standard PDDL-2 (Planning Domain Description Language) [13] for internally representing the planning problems is thus part of our current work. Other performance improvements can be reached more simply with the means of, e.g., a *plan cache*: once a plan is computed, we can store this plan together with the client's preferences and the media content description and access the plan immediately once a similar request is received without searching for a new plan at all. Even more, in cases where only some parameters like the network conditions change, we do not necessarily have to compute a new adaptation plan from scratch but can rather try to update an existing plan with the new parameters.

Quality of plans. The general challenge with respect to plan quality is that there may be several possible plans to reach the same goal. While we consider the selection of alternative tools for the same transformation step as a minor problem, the effectiveness of the produced plans can heavily depend on the *order of the transformation steps*. As an example, for performance reasons it is reasonable to reduce the video's size before the color is reduced, although the inverse order would also result in the same desired video format.

In the current implementation, we tackle this problem by introducing search heuristics into the knowledge base. Based on this additional information, we can steer the search strategy of the underlying planner which exploits the annotated priorities of each action every time a plan step has to be selected. While this approach helps us to improve the plan quality in many cases—in particular for standard operations—the resulting plans may still be sub-optimal with respect to execution efficiency, as there is no global optimization. Therefore we are currently investigating how we can introduce general concepts for expressing such search and optimization strategies as well as heuristics into our framework. In particular, we see opportunities in re-using and extending the already existing mechanisms for providing *adaptation hints* in the context of Adaptation QoS [18] (see Section 6). The principal advantage of this approach is that it is already part of the current MPEG-21 standard; the expressiveness of the provided mechanisms is however limited and could be insufficient for our purposes.

Another interesting option we are evaluating is to use an additional *ontology* with a high expressiveness for modeling domain rules and heuristics for improving plan quality. In that context, we also want to be able to cope with situations where no plan can be found that *exactly* produces the desired format. In these cases, we need mechanisms for defining how *acceptable* solutions for the client look like. The applicability of an approach based on domain ontologies for a similar multimedia adaptation problem was, for instance, already demonstrated in [30]. The interoperability with the MPEG-21 Digital Item Adaptation standard, however, is still an open issue.

3.3. Integration with other systems

The work presented in this paper aims at providing a core framework for building intelligent and flexible adaptation engines that can be used within larger multimedia systems. Currently, we are participating in the CoCOMA project carried out within the EU-funded DELOS Network of Excellence,[7] which focuses on the integration of *content-based multimedia retrieval* in digital libraries, and on the delivery and consumption of the retrieved multimedia data. The project aims at providing users of digital library systems with intelligent and personalized retrieval capabilities for large media collections.

The objective of this project is therefore also to develop an end-to-end system that integrates the various technologies and frameworks that are needed for personalized content delivery like, for instance, *metadata extraction* [12] or multimedia *authoring* [3]. As such, this project will serve as an appropriate testbed for further evaluation of the applicability and interoperability of our approach.

4. Semantic service descriptions

At the current stage of the standardization process, the MPEG-21 framework does not yet provide normative mechanisms for describing *the semantics* of available transformation tools, which is a major pre-requisite for intelligent construction of multi-step adaptation sequences. Within this section, we will now discuss the specific requirements for such a knowledge representation mechanism and show how *Semantic Web Service technology* can serve as a suitable means for capturing the semantics of such transformation tools and operations.

In the field of knowledge-based planning, PDDL-2 (Planning Domain Description Language) [13] is nowadays well-established as the standard knowledge representation language. It has been shown that the expressiveness of this language is sufficient to cover a wide range of application and problem domains. Nonetheless, besides expressiveness considerations, there are specific aspects in our problem setting to be taken into account:

- At the moment, there is limited awareness of PDDL-2 outside the Artificial Intelligence and planning communities, which can potentially hamper the acceptance of such

[7] http://www.delos.info/

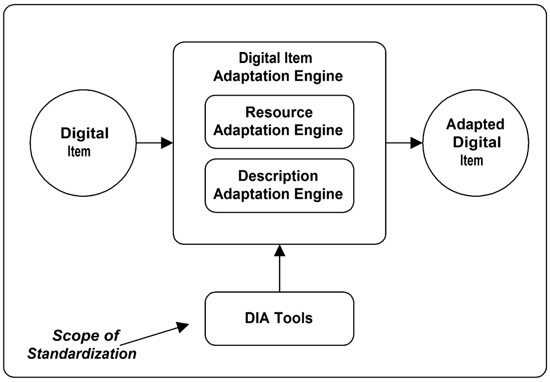

Fig. 4 Overview and scope of the MPEG-21 DIA standard.

a language. In addition, the logic-based notation does not really comply with the current trends in open, Web-based systems, where XML-based technology is now state of the art.

- Even more importantly, the chosen representation mechanism for describing tool semantics has to be combined and interoperate with the representation format that is used in the existing XML-based ISO/IEC standards MPEG-7 and MPEG-21. This particularly holds since we want to directly use these standards as a shared domain ontology, i.e., we want to guarantee that only a defined set of terms and symbols are used for describing tool execution semantics.
- Finally, such a language still needs to have precise and well-defined semantics such that the automatic composition of meaningful, multi-step transformation sequences is possible.

With respect to the problem of interoperability with the existing multimedia standards, a closer look at these standards is needed. At the current state of standardization of *Digital Item Adaptation* (DIA) within the MPEG-21 framework, the normative part is mainly limited to the tools that have to be used for describing, e.g., the user's preferences, his/her usage environment or QoS information on the media resources. These descriptions are provided in terms of XML documents. The actual engines that perform the adaptation of the content are non-normative and the implementation is left to vendors of adaptation components (see Fig. 4).

However, this current DIA approach does not account for cross-system reuse and interoperability of existing adaptation tools as required in our approach. If a provider of an online adaptation service wants to integrate an existing transformation tool or algorithm, the semantics of executing a tool's method must be hand-coded into the logic of the internal, non-standardized adaptation engine. As a consequence, we have the problem that the adaptation server's core has to be changed each time a new tool is available or the standards are extended. In addition, there are no guarantees that the manually engineered tool semantics corresponds to the real semantics encoded in the tool.

Therefore, the scope of standardization has to be extended to include such standardized mechanisms for capturing the semantics of transformation operations as specified in [33]. In the near future, such descriptions could be even provided by the tool and algorithm providers themselves, if we think of highly specialized content provisioning services like personalized live-streams as a new business model. However, there is still agreement that the internal implementation of adaptation engines in the sense of MPEG-21 should not be within the scope of the standards, e.g., whether a planning engine is used for construction of the transformation sequences or not.

In following subsections, we will give a short introduction to the ideas of Semantic Web Services and basic terminology before we show how this knowledge representation scheme can serve our purposes and can be smoothly integrated with the domain ontology formed by the MPEG standards.

4.1. Semantic web services

The idea of the Semantic Web [2] is to have resources available on the Web that can be retrieved by content and not by keyword [32] and can also be automatically accessed by software applications or agents. Services in that context are specific "resources" where the invocation of such a service typically results in a world-altering action, like the sale of a book. In order to facilitate these new forms of using the Web, the available resources have to be semantically annotated, which leads to the development of new markup languages. The Web Ontology Language (OWL) [23, 32] is the most prominent approach having the status of a W3C recommendation. Languages like OWL can be both used for creating ontologies for a domain as well as for the semantic description of particular Web sites.

If we look at service-type resources, the envisioned automation in the Semantic Web means that we need annotation mechanisms such that software agents can, e.g., search for appropriate services, invoke them with the required parameters, or even compose a complex transaction from several distributed services. A typical example from the domain of travel planning is described in [24], where a complete travel arrangement requires the invocation of multiple services like making a hotel reservation or booking a flight. These requirements are currently addressed in OWL-based Web Service Ontology (OWL-S) [32], which constitutes a general ontology for the domain of services.

Automatic service composition and interoperation are the most important aspects with respect to our problem domain of intelligent multimedia adaptation. (Other OWL-S features related to organizational aspects, e.g., service registration and publication, are not discussed in this paper.) In order to

Fig. 5 Fragment of an MPEG-7 video description.

```
<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:VisualCodingFormatCS:2001:3.3.1"
                                colorDomain="color">
                  <Name xml:lang="en">
                        MPEG-4 Visual Advanced Simple Profile @ Level 0</Name>
                </Format>
                <Frame width="640" height="480"/>
              </VisualCoding>
            </MediaFormat>
          </MediaProfile>
        </MediaInformation>
      </Video>
    </MultimediaContent>
  </Description>
</Mpeg7>
```

support these tasks, OWL-S 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. As already mentioned, this particular *IOPE* approach is also widely used for planning tasks in Artificial Intelligence.

Currently, OWL-S 1.1 does not yet specify the rule language to be used for describing conditions or outputs of actions. Different candidates like Semantic Web Rule Language (SWRL)[8] are currently evaluated and will soon be part of OWL-S, because such a common language with defined semantics is crucial for interoperability. However, for our problem domain we feel that in typical cases no unusual requirements on the expressiveness of the language should arise that are not within the scope of the current proposals.

In our opinion, one of the most important aspects that can hamper the usage of Semantic Web Services is the problem of having a shared domain ontology. If we think again of the well-known travel planning problem, automatic service composition can only be done if all participants not only use the same terms like "hotel reservation", but also associate the same meaning with that term. Semantic markup languages like OWL (or its predecessor DAML+OIL[9]) only provide a common infrastructure for defining, integrating, and reasoning about such ontologies, but cannot solve the problem of agreeing on a domain-specific ontology in a community. In fact, even for well-understood domains like business-to-business electronic commerce or electronic supply-chain management, several competing pseudo-standards for XML-based data exchange emerged.[10] In our domain, however, such a common understanding can be reached by interpreting the existing MPEG standards as the common basis to be used for describing the functionality and semantics of the adaptation services.

4.2. Multimedia standards as domain ontology

Multimedia standards like MPEG-7 or MPEG-21 precisely specify the way multimedia resources and usage environments (e.g., terminal and network capabilities) can be annotated with additional information. XML Schema[11] technology is extensively used in these standards as the specification language for the shared vocabulary. Specific media resources are annotated with document instances that correspond to that standardized schema. Figure 5 shows an example of how a video resource is annotated in MPEG-7; in particular it describes the parameters color domain, encoding format, and the frame size, which we used in the planning example of Section 2. The MPEG-21 DIA description of the end device used in the planning example is shown in Fig. 6; in particular it describes the terminal's resolution as well as its decoding and color capabilities.

Although not explicitly mentioned as such, the definitions in the relevant MPEG-7 and MPEG-21 standards parts ([15]

8 http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/

9 http://www.daml.org

10 http://www.ebxml.org/

11 http://www.w3.org/XML/Schema

```
<DIA>
  <Description xsi:type="UsageEnvironmentType">
    <UsageEnvironmentProperty xsi:type="TerminalsType">
      <Terminal>
        <TerminalCapability xsi:type="DisplaysType">
          <Display>
            <DisplayCapability
                    xsi:type="DisplayCapabilityType"
                    activeDisplay="true" colorCapable="false">
              <Mode>
                <Resolution horizontal="320" vertical="240"
                    activeResolution="true"/>
                <Resolution horizontal="160" vertical="120"
                    activeResolution="false"/>
                <Resolution horizontal="80" vertical="60"
                    activeResolution="false"/>
              </Mode>
            </DisplayCapability>
          </Display>
        </TerminalCapability>
      </Terminal>
    </UsageEnvironmentProperty>
  </Description>
</DIA>
```

Fig. 6 MPEG-21 DIA description of a terminal with certain display capabilities.

and [18, 36]) 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. In the examples above, the *colorDomain* attribute describes the color domain of the video, and the frame size of 640×480 pixels is 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. Please note that such a rendering device could support multiple resolutions; the resolution mode currently used is indicated by the Boolean *activeResolution* attribute.

For our multi-step adaptation problem, the representation mechanism for the resource descriptions (MPEG-7) and the client preferences (MPEG-21) is therefore already given. At the moment, the language to be used for describing the semantics of the transformation services is not within the scope of OWL-S. Quite obviously, the concepts that are used in preconditions and effects of the services have to correspond to those that are used in the relevant MPEG standards. In Section 5, we will show, by discussing the results of a *ISO/IEC MPEG Core Experiment*, how these requirements can be met by integrating the different technologies and representation mechanisms on the technical level.

Because the rule language is not yet standardized in OWL-S, we have chosen the current version of SWRL for evaluation purposes in our prototypical implementation. We describe inputs, outputs, preconditions, and effects in SWRL notation which is then automatically transformed to an internal, logic-based format which can be exploited by a Prolog-based planning engine. Our experiments show that the complexity of the conditions and expressions that we need for describing a transformation service is typically very limited and the usage of simple facts is sufficient in most cases. The same also holds for a potential extension of the approach for the problem of Digital Rights Management (DRM)[12]: if some transformations, like the extraction of the audio stream from a movie, require particular rights, such information can be easily incorporated in our approach and encoded in the preconditions of the transformation functions.

Overall, we are optimistic that the rule language that is developed for OWL-S will be expressive enough for our problem domain and that compatibility of our prototype can be reached by transforming the descriptions on the syntactical level. Note that a full re-implementation of the MPEG standards with the representation means of OWL-S is in principle possible and would allow for a consistent usage of OWL technology throughout. As a side effect, such an approach would in our opinion also improve the comprehensibility and clarity of the standards. A first step in that direction can be the construction of an OWL-based ontology for the domain, which can serve as accompanying material that can be used for constructing the semantics descriptions for adaptation services.

5. Integrating the MPEG-21 multimedia framework and semantic web technology

In this section we present how the MPEG-21 Multimedia Framework and Semantic Web technology can be integrated following our knowledge-based adaptation decision-taking approach. Therefore, we present how our approach has been implemented in the course of an official ISO/IEC MPEG Core Experiment (CE), and how it has been integrated into the multimedia framework utilizing OWL-S and SWRL, in particular, into the so-called *Conversions and Permissions* amendment [33] of MPEG-21 Digital Item Adaptation (DIA) [18, 36]. Before going into the details and results of this effort, we give a brief overview of the related MPEG-21 goals, concepts, and parts.

In recent years, researchers within the multimedia community have developed technologies and standards to address some of the issues described in the previous sections, in support of what is generally referred to as Universal Multimedia Access (UMA) [34]. In order to create the big picture of how these different technologies and standards fit together, the Moving Picture Experts Group (MPEG)[13] decided to standardize the MPEG-21 Multimedia Framework

[12] Open Mobile Alliance (http://www.openmobilealliance.org) and Digital Media Project (http://www.chiariglione.org/project)

[13] http://www.chiariglione.org/mpeg

[9] with the ultimate goal to support users during the exchange, access, consumption, trade, or other manipulation of so-called Digital Items in an efficient, transparent, and interoperable way. A Digital Item (DI) is defined as a structured digital object with a standard representation, identification, and metadata. It acts as the fundamental unit of distribution and transaction within this framework. In practice, a Digital Item is a structured XML-based container format of one or more multimedia resources and associated metadata.

The MPEG-21 parts relevant for our work are briefly introduced in the following. MPEG-21 DIA [18] aims to achieve interoperable transparent access to (distributed) advanced multimedia content by shielding users from network and terminal installation, management, and implementation issues. In order to reach this goal, the adaptation of Digital Items is required. Furthermore, *Amendment 1* to DIA [33] entitled *Conversions and Permissions* enables finer-grained control over the changes that can occur when playing, modifying, or adapting Digital Items and their resources. One major part of this amendment comprises *conversions* in terms of adaptation operations a terminal is capable of, as well as adaptation operations suggested by certain users within the multimedia delivery chain in case DI adaptation is required. In our work, however, we focus on the former type of conversions, i.e., the *conversion capabilities*. For the different types of conversions we use the MPEG-21 Rights Data Dictionary (RDD) [17] which provides a non-exhaustive but comprehensive set of multimedia related terms and definitions. Additionally, such RDD terms can be further refined and specialized through pre-defined extensibility mechanisms. Note that according to the basic guideline of MPEG-21, i.e., to fill the gap between existing standards and to specify only the minimum that is required for the sake of interoperability, this amendment provides only the basic framework for describing such conversions. As such, the actual description of specific conversion operations, e.g., spatial scaling of images, is not a normative part of this amendment but can be easily incorporated due to its open and extensible design. Fortunately, the Semantic Web technologies like OWL-S and SWRL provide mechanisms for describing such specific conversion operations which can make use of terms as specified in existing MPEG standards, i.e., our *shared domain ontology*. Therefore, the integration has been carried out in both directions:

1. **OWL-S → MPEG-21.** Specific conversion descriptions can be included in MPEG-21 descriptions, e.g., as terminal capabilities as shown in Fig. 7.
2. **MPEG-21 → OWL-S.** MPEG-21 syntax and semantics are referenced from within OWL-S descriptions, e.g., file format, display resolution, or the actual type of conversion as shown in Figs. 7 and 8, respectively.

Figure 7 shows a conversion description describing an adaptation operation (spatial scaling of an image) which the adaptation engine is capable of performing, as an MPEG-21 *terminal capability* and as a *service profile* of OWL-S.

The semantics of this description is as follows:

- The DIA description describes the *conversion* capabilities of a terminal.
- In this case, OWL-S is used for describing the actual conversion capabilities which is possible due to the extensible design of the MPEG-21 DIA schema.
- MPEG-21 Rights Data Dictionary (RDD) serves as the shared domain ontology providing a specialized term for spatial scaling of rectangular images. As such, interoperability is guaranteed. Note that the actual RDD term is referenced through the `profileHierarchy:SpatialScaleRectangularImage` tag where its prefix will be resolved to `urn:mpeg:mpeg21:2003:01-RDD-NS`, the standardized MPEG-21 RDD universal resource name (URN). For better human readability, the full reference to the RDD term is included within the `profile:serviceName` tag.
- For this kind of spatial scaling operation, an image as well as the resulting width and height as inputs are required which results in the adapted version of the input image as output. The current width and height of the input image are obtained from its associated MPEG-7 description.

The actual service model, i.e., how it works, is referenced from within the `owl:Ontology` tag and is excerpted in Fig. 8, i.e., only preconditions and effects are shown:

- MPEG-7 and MPEG-21 terms are referenced from within the `owlx:Class` and `owlx:Individual` tags.
- According to the preconditions only bitmap images are accepted.
- Furthermore, it shows how the display's horizontal and vertical resolutions are mapped onto the image's width and height by referencing terms from MPEG-7 and MPEG-21.
- Finally, the output format of the image is defined which is the same as the input format, i.e., bitmap.

More details are given in [19].

So far we have concentrated on how MPEG-21 and Semantic Web technology can be integrated. Figure 9 illustrates the big picture, i.e., the architecture, of an MPEG-21 DIA compliant multi-step capable knowledge-based adaptation engine. The user requests a multimedia resource which she/he wants to consume with a maximum user experience. Together with the *context Digital Item* including the usage environment description (e.g., color capabilities, resolution, or decoding capabilities of the rendering device as shown in Fig. 6) it

```xml
<dia:DIA>
  <dia:Description xsi:type="TerminalsType">
    <dia:Terminal>
      <dia:TerminalCapability xsi:type="ConversionCapabilitiesType">
        <dia:ConversionDescription xsi:type="ConversionDescriptionXMLType">
          <rdf:RDF>
            <owl:Ontology>
              <!-- version info and comments ... -->
              <owl:imports rdf:resource="&profile;"/>
              <owl:imports rdf:resource="&process;"/>
              <owl:imports rdf:resource="&rescale_process;"/>
            </owl:Ontology>
            <profileHierarchy:SpatialScaleRectangularImage rdf:ID="SpatialScaler">
              <!-- reference to the process model specification -->
              <profile:has_process rdf:resource="&rescale_process;#rescale"/>
              <profile:serviceName>
                urn:mpeg:mpeg21:2003:01-RDD-NS:SpatialScaleRectangularImage
              </profile:serviceName>
              <profile:textDescription>This service...</profile:textDescription>
              <profile:contactInformation>...</profile:contactInformation>
              <!-- Descriptions of IOPEs -->
              <profile:hasInput rdf:resource="&rescale_process;#imageIn"/>
              <profile:hasInput rdf:resource="&rescale_process;#width"/>
              <profile:hasInput rdf:resource="&rescale_process;#height"/>
              <profile:hasOutput rdf:resource="&rescale_process;#imageOut"/>
            </profileHierarchy:SpatialScaleRectangularImage>
          </rdf:RDF>
        </dia:ConversionDescription>
        <!-- further conversion descriptions, e.g., grayscaling, SNR scaling, etc. -->
      </dia:TerminalCapability>
    </dia:Terminal>
  </dia:Description>
</dia:DIA>
```

MPEG-21 terminal capability

reference to RDD term from within OWL-S description

Fig. 7 MPEG-21 DIA description of a terminal capable of performing spatial scaling of rectangular images

forms one input to the adaptation engine. The requested resource and its associated metadata (e.g., MPEG-7 metadata describing the characteristics of the multimedia content such as color domain, resolution, or coding format as shown in Fig. 5), which is referred to as the *content Digital Item*, constitutes the other input to the adaptation engine. The Digital Item Adaptation engine itself consists of two major parts which are the adaptation decision-taking engine and the resource/description adaptation engine. Additionally, a list of conversion programs described by MPEG-21 terminal capabilities utilizing OWL-S and SWRL is maintained by the adaptation engine. Please note that these conversion programs are possibly distributed over several physical adaptation nodes from different tool vendors. The *adaptation decision-taking engine* computes an *adaptation plan* based on the *MPEG-7 content description* of the original resource, the *MPEG-21 UED*, and the *conversion descriptions* of the available *conversion programs*. The adaptation plan is then forwarded to the *resource/description adaptation engine* which applies the individual conversion programs according to the adaptation plan on the original multimedia content. As a result, an *adapted content Digital Item* comprising the *adapted multimedia content* and the *adapted content description* is produced and forwarded to the user where it is consumed according to the current usage environment.

By using standardized description formats as a shared domain ontology we have shown how our knowledge-based multimedia framework enables interoperability among different users within the multimedia delivery chain, e.g., consumers, producers, and network operators. As such, the framework can serve as a basis for deploying and distributing independent adaptation services in open markets and environments.

6. Related work

Due to the integrative nature of our approach, there are quite many different research fields that are related to our work. We will therefore only focus on recent, representative examples in the different areas and structure our discussion along the following main lines: multimedia adaptation systems and low-level media adaptation, Semantic reference to MPEG-7 term from within OWL-S description

```
<!-- Preconditions: the input image format has to be bmp -->
<process:hasPrecondition>
  <expr:SWRL-Condition>
    <expr:expressionBody>
      <swrlx:Atom>
        <owlx:Class owlx:name="urn:mpeg:mpeg7:cs:FileFormatCS:2001:11"/>
        <owlx:Individual owlx:name="#imageIn"/>
      </swrlx:Atom>
    </expr:expressionBody>
  </expr:SWRL-Condition>
</process:hasPrecondition>
<!-- Effects are that the output image's width and height correspond
     the MPEG-21 constraints horizontal and vertical -->
<process:hasEffect>
  <expr:SWRL-Condition>
    <expr:expressionBody>
      <!-- display's horizontal resolution is the same as image width -->
      <swrlx:classAtom>
        <!-- reference to the MPEG-7 description: frame width -->
        <owlx:Class owlx:name="urn:mpeg:mpeg21:01-DIA-MediaInformationCS-NS:81"/>
        <!-- reference to the MPEG-21 description: horizontal resolution -->
        <owlx:Individual owlx:name="urn:mpeg:mpeg21:01-DIA-AdaptationQoS-NS:6.5.9.1"/>
      </swrlx:classAtom>
      <!-- display's vertical resolution is the same as image height -->
      <swrlx:Atom>
        <!-- reference to the MPEG-7 description: frame height -->
        <owlx:Class owlx:name="rn:mpeg:mpeg21:01-DIA-MediaInformationCS-NS:82"/>
        <!-- reference to the MPEG-21 description: vertical resolution -->
        <owlx:Individual owlx:name="urn:mpeg:mpeg21:01-DIA-AdaptationQoS-NS:6.5.9.2"/>
      </swrlx:Atom>
      <!-- Return value -->
      <swrlx:Atom>
        <owlx:Class owlx:name="urn:mpeg:mpeg7:cs:FileFormatCS:2001:11"/>
        <owlx:Individual owlx:name="#imageOut"/>
      </swrlx:Atom>
    </expr:expressionBody>
  </expr:SWRL-Condition>
</process:hasEffect>
```

reference to MPEG-21 term from within OWL-S description

mapping between MPEG-7 media information and MPEG-21 terminal capabilities

Fig. 8 Preconditions and effects of the spatial scaling operation.

Web (Services) technology, and ontologies in multimedia applications.

6.1. Multimedia adaptation systems

Server-side adaptation of resources is well established in the area of distributed multimedia systems, mostly in order to cope with the real-time requirements of continuous media. For an overview, refer to [14], for instance.

While most research in that area for a long time focused on proprietary low-level bitstream adaptation mechanisms, it was only recently that work was reported which exploits the new metadata possibilities of the emerging MPEG standards; examples are the *ViTooKi*[14] *Video Tool Kit* project [5, 28] or the work described in [31].

Comparable to our work, *ViTooKi* provides an architecture for an adaptive proxy for audio/visual streams capable of adapting multimedia resources according to terminal capabilities (display size and color capabilities) and network characteristics (available bandwidth). For the adaptation process, the concept of *adaptor chains* is introduced which enables the concatenation of several adaptation steps [28]. Such adaptor chains are dynamically instantiated at runtime according to the usage environment. In that work, the adaptation process is mainly steered by the MPEG-7 *VariationSet Description Scheme* which can be used to provide hints as to when to apply an adaptation and which algorithm to use.

In contrast to our knowledge-based approach, the construction process for adequate adaptation chains in *ViTooKi* is not automated in the sense that these chains are intelligently assembled from semantically-enriched tool descriptions. Rather, they rely on a set of pre-defined, manually engineered adaptation chains, which are looked up dynamically at run-time. When the adaptation chain is executed, the corresponding external tools like those contained in the FFMPEG library are parameterized and executed. Again, the set of such external tools and algorithms is manually engineered and the existing chains possibly have to be revised

[14] http://vitooki.sourceforge.net/

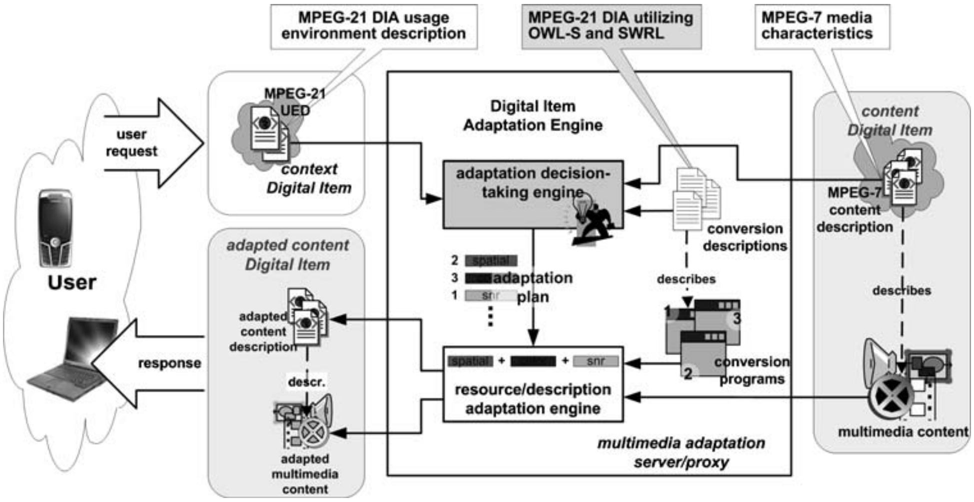

Fig. 9 MPEG-21 DIA compliant multi-step capable knowledge-based adaptation engine using OWL-S and SWRL.

when a new tool is introduced to the framework, which is not necessary in our approach.

Currently, we are working on a detailed analysis of how our approach can be integrated into the architecture and infrastructure provided within the *ViTooKi* project, in particular since this framework implements additional features like intelligent video caching mechanisms.

6.2. Low-level media adaptation

Bitstream Syntax Description (BSD) is a low-level adaptation mechanism which is also covered by the current MPEG-21 DIA standard. Basically, BSD is an additional metadata layer that can be used to describe the high-level structure of a media bitstream. A BSD can be used to perform manipulations directly on the bitstream without requiring costly decoding or encoding operations. BSD is designed to be used together with *Adaptation QoS* [18, 36], another tool defined within the MPEG-21 framework which enables users (primarily content producers) to describe quality of service (QoS) trade-offs. Adaptation QoS specifies the relationship between environmental (e.g., terminal and network) constraints, media quality, and feasible adaptation operations. For example, a constraint could be the limitation of the network throughput on the path to the client. Adaptation of a video in this case could be accomplished by performing operations like, e.g., frame dropping, coefficient dropping, or wavelet-based reduction, which would lead to changes in the quality of the multimedia stream. Consequently, Adaptation QoS is useful to hint adaptation nodes to meet given constraints with regard to the quality of the adaptation result.

We see our work as more general and on a higher layer than BSD and Adaptation QoS. BSD could be one adaptation step in the presented architecture, providing a fast adaptation mechanism for compressed media streams. Furthermore, Adaptation QoS could be integrated into our approach, providing a basis for deriving a heuristic used by the planner to find, e.g., the best adaptation steps with regard to quality.

Similarly to BSD-based media stream adaptation, *transcoding* techniques and systems as covered, e.g., in [35] for videos, could be employed as building blocks (individual adaptation steps) in our higher-level architecture.

6.3. Semantic web services

The extension of the ideas of the Semantic Web to the *services* domain is still in its early phases, and the standardization process for the required markup languages within the *World Wide Web Consortium* (W3C) is an ongoing progress. On the research side, most efforts are currently spent on mechanisms and algorithms for intelligent service discovery, composition, and execution, but there are only very few reports on *real-world* applications that already exploit these new possibilities.

One of the currently few examples is described in the Industrial Track of the 2004 International Semantic Web Conference[15]: Lopez et al. [11] report on an application from the financial domain, where an intelligent *Notification Agent* manages alerts when critical financial situations arise by automatically discovering and selecting customer notification services. Comparable to our approach, they base parts of their framework on a commonly-agreed ontology (i.e., standard) in the domain, in their case the IFX financial standard.[16] In contrast to our approach, they use *Description Logics* [32] as the language for describing services and a proprietary mechanism to describe and execute composite, multi-step services.

[15] http://iswc2004.semanticweb.org/

[16] http://www.ifxforum.org/.

Another Semantic Web Services application, although from the academic environment, is reported in [29]. The *my*Campus environment aims at extending the nowadays relatively simple infotainment services available for *mobile* users by providing higher levels of automation and the development of services that respect the *context* in which users operate. The main similarity to our approach lies in the fact that the authors aim at personalized delivery of (multimedia) content for different usage environments, both with respect to hardware as well as software; in addition, they also use AI-based planning technology for the construction of composite services. In contrast to our work, however, they rely on proprietary, domain-specific ontologies for service annotations whereas we use the commonly-agreed multimedia standards as shared domain ontology.

6.4. Ontologies in multimedia applications

Up to now, Semantic Web technology was not broadly used in the field of current multimedia-related research. Still, there are a few reports (like [1, 25, 26] or [21]) that aim at exploiting the advantages of joining current MPEG technology for describing multimedia content with the shared knowledge representation provided by Semantic Web technology.

In [21], for instance, an approach towards automated, knowledge-assisted content analysis for multimedia resources is described. This work shows how domain specific ontologies can be used for narrowing the semantic gap that arises when concepts and objects should be automatically extracted without human intervention from image or spatio-temporal features. Semantic Web technology is used as a representation mechanism for modeling additional domain-specific knowledge with the side intention that general, domain-independent, knowledge representation and reasoning mechanisms from arbitrary other domains can be re-used.

The *SCULPTEUR* project described in [1] is an approach where ontologies are serving as integrating technology for advanced concept, metadata, and content browsing and retrieval in the domain of digital museum information. In this work, a domain ontology and a semantic layer with references to the actual objects of the digital collection like 2-D or 3-D images and models were developed. Semantic Web technology was mainly used for the purpose of system integration, as the existing information like textual metadata on items is in many cases stored in separate collections and legacy systems.

Finally, another project where Semantic Web technology is applied in the multimedia domain is *OREL* (Ontology-based Rights Expression Language) [26]. OREL is a rights expression language designed on top of the OWL Web Ontology Language. With the help of an OREL ontology, users and applications can handle digital rights and permissions at the semantic level. For instance, it is possible to express that "Alice is granted to play the audio clip *sample.mp3* for five times free of charge before Christmas 2004". However, in fact this ontology implements a subset of the MPEG-21 Rights Expression Language (REL) [16, 37] and MPEG-21 Rights Data Dictionary (RDD) [17, 27] and it is still open how these ontologies interoperate and whether they are fully compatible.

All of the described approaches use Semantic Web technology basically as a tool for further semantic annotations and domain-specific ontologies that go beyond the capabilities of MPEG-7, that was designed with the goal of highly accurate content description and retrieval. In our work, we use Semantic Web technologies at a different level as we aim at transforming the contents themselves and assume that content annotations are expressed in terms of MPEG-compliant metadata.

7. Conclusions

In this paper, we have presented a novel, fully knowledge-based approach for intelligent multimedia adaptation, which is required in the context of the rapid evolution of new possibilities of multimedia consumption over the Internet.

The contributions of this paper can be summarized as follows. First, we have shown how well-established, domain-independent reasoning techniques from the field of Artificial Intelligence can be applied to solve the real-world problem of dynamic, semantics-based composition of transformation steps for multimedia content adaptation. The declarative approach allows us to build an extensible framework that is capable of flexibly integrating and re-using arbitrary external multimedia adaptation components and that is thus robust against changes and extensions in this very rapidly evolving application domain.

As such, our approach on the one hand broadens the application scope of AI technology and, on the other hand, for the first time introduces knowledge-based technology into emerging ISO/IEC MPEG multimedia standards. Moreover, our framework establishes a new important link between the up-to-then rather isolated standardization worlds of ISO/IEC MPEG and W3C.

Our work also showed that the mechanisms of Semantic Web Service technology have already reached a state such that they can be used for describing and solving non-trivial service composition problems. Although in our approach the transformation services are not necessarily distributed over the Web, we also see our work as a contribution to the field of Semantic Web Services. Our work both describes a realistic application problem and test-bed for these techniques, but also emphasizes the fact that the existence of a shared, commonly-agreed domain ontology is a fundamental prerequisite for the success of Semantic Web Services.

The proposed framework was successfully evaluated in an official *ISO/IEC MPEG Core Experiment* within the MPEG-21 standardization process, to which some of the authors have been actively contributing for several years. Our future work mainly includes the continuation of the evaluation process, the development of a stable reference implementation, and finally the further propagation and integration of our work into the MPEG-21 multimedia standards and into a digital library system for personalized, content-based multimedia retrieval and delivery.

References

1. Addis M, Boniface M, Goodall S, Grimwood P, Kim S, Lewis P, Martinez K, Stevenson A (2003) SCULPTEUR: Towards A new paradigm for multimedia museum information handling. In: Proceedings of the International Semantic Web Conference 2003, FL, USA, Lecture Notes in Computer Science, 2870, pp 582–596
2. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. Scientific American 284:34–43
3. Bertino E, Ferrari E (1994) Stolf M, MPGS: An interactive tool for the specification and generation of multimedia presentations. IEEE Transactions on Knowledge and Data Engineering 12(1): 102–125
4. Blum A, Furst M (1997) Fast planning through planning graph analysis. Artificial Intelligence 90:281–300
5. Boeszoermeny L, Hellwagner H, Kosch H, Libsie M, Podlipnig S (2003) Metadata driven adaptation in the ADMITS project. EURASIP Signal Processing: Image Communication, Special Issue on Multimedia Adaptation 18:749–766
6. Bonet B, Geffner H (2003) Faster heuristic search algorithms for planning with uncertainty and full feedback. In: Int. Joint Conference on Artificial Intelligence (IJCAI-2003). Acapulco, Mexico, pp 1233–1238
7. Bormans J (2003) (ed.) MPEG-21 requirements v 1.5 ISO/IEC JTC 1/SC 29/WG 11 N5873, Trondheim, Norway
8. Bratko I (2000) Prolog programming for artificial intelligence, 3rd edn., Addison-Wesley
9. Burnett I, de Walle RV, Hill K, Bormans J, Pereira F (2003) MPEG-21: Goals and achievements. IEEE MultiMedia Magazine 10(6):60–70
10. Bylander T (1994) The computational complexity of propositional STRIPS planning. Artificial Intelligence 69(1/2): 165–204
11. Cobo JML, Losada S, Corcho Ó, Benjamins VR, Niño M, Contreras J (2004) SWS for financial overdrawn alerting. In: International Semantic Web Conference, pp 782–796
12. Eidenberger H, Breiteneder C (2003) VizIR—A framework for visual information retrieval. Journal of Visual Languages and Computing 14:443–469.
13. Fox M, Long D (2003) PDDL2.1: An extension to PDDL for expressing temporal planning domains. Journal on Artificial Intelligence Research 20:61–124
14. Gecsei J (1997) Adaptation in distributed multimedia systems. IEEE MultiMedia 4(2):58–66
15. ISO/IEC 15938-5:2003, Information Technology—Multimedia Content Description Interface—Part 5: Multimedia Description Schemes, 2003
16. ISO/IEC 21000-5:2003, Information Technology—Multimedia Framework—Part 5: Rights Expression Language, 2003
17. ISO/IEC 21000-6:2003, Information Technology—Multimedia Framework—Part 6: Rights Data Dictionary, 2003
18. ISO/IEC 21000-7:2004, Information Technology—Multimedia Framework—Part 7: Digital Item Adaptation, 2004
19. Jannach D, Leopold K (2005) A multimedia adaptation framework based on semantic web technology. In: Workshop on Multimedia and the Semantic Web, held in conjunction with 2nd European Semantic Web Conference
20. Jannach D, Leopold K, Hellwagner H (2004) An extensible framework for knowledge-based multimedia adaptation. In: Proceedings of 14th Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE), Lecture Notes in Artificial Intelligence 2070, pp 434–444
21. Kompatsiaris I, Mezaris V, Strintzis MG (2004) Multimedia content indexing and retrieval using an object ontology. In: G. Stamou (ed.) Multimedia content and semantic web—methods, standards and tools, New York, USA
22. Martinez JM, Koenen R, Pereira F (2002) MPEG-7—The generic multimedia content description standard, Part 1. IEEE MultiMedia, 9(2):78–87
23. McGuiness DL, van Harmelen F (2004) (eds.) OWL Web ontology language overview. Technical report. http://www.w3.org/TR/owl-features/.
24. McIlraith S, Son T, Zeng H (2001) Semantic web services. IEEE Intelligent Systems, Special Issue on the Semantic Web 16: 46–53
25. Mezaris V, Kompatsiaris I, Strintzis MG (2004) Region-based image retrieval using an object ontology and relevance feedback. EURASIP Journal on Applied Signal Processing, Special Issue on Object-Based and Semantic Image and Video Analysis pp 886–901
26. Qu Y, Zhang X, Li H (2004) OREL: An ontology-based rights expression language. In: Proceedings of the 2004 ACM World Wide Web Conference (WWW'2004) Alternate Track Papers and Posters pp 324–425
27. Rust G, Barlas C (2005) The MPEG-21 rights data dictionary. IEEE Transactions on Multimedia 7(2)
28. Schojer P, Boeszoermenyi L, Hellwagner H, Penz B, Podlipnig S (2003) Architecture of a quality based intelligent proxy (QBIX) for MPEG-4 videos. In: Proceedings of the 2003 ACM World Wide Web Conference (WWW'2003). Budapest, Hungary, pp 394–402
29. Sheshagiri M, Sadeh N, Gandon F (2004) Using semantic web services for context-aware mobile applications. In: Proceedings of MobiSys2004 Workshop on Context Awareness. Boston, USA, pp 782–796
30. Soetens P, Geyter MD (2005) Applying domain knowledge to multi-step media adaptation based on semantic web services. In: Proceedings of Workshop on Image Analysis for Multi-media Interactive Systems (WIAMIS) Montreux, Switzerland
31. Steiger O, Sanjun DM, Ebrahimi T (2003) MPEG-based personalized content delivery. In: Proc. of the IEEE Int. Conf. on Image Processing, ICIP 2003, Barcelona, Spain, pp 14–16
32. The OWL Services Coalition (2004) OWL-S: Semantic markup for Web Services. Technical report. http://www.daml.org/services/owl-s/1.0/.
33. Timmerer C, DeMartini T, Barlas C (2005) (eds.) Information technology—multimedia framework (MPEG-21)—Part 7: Digital Item Adaptation, AMENDMENT 1: DIA Conversions and Permissions

34. Vetro A, Christopoulos C, Ebrahami T (2003) (eds.) IEEE Signal Processing Magazine, Special Issue on Universal Multimedia Access, 20
35. Vetro A, Christopoulos C, Sun H (2003) Video transcoding architectures and techniques: An overview. IEEE Signal Processing Magazine 20(2):18–29
36. Vetro A, Timmerer C (2005) Digital item adaptation: Overview of standardization and research activities. IEEE Transactions on Multimedia 7(2)
37. Wang X, DeMartini T, Wragg B, Paramasivam M (2005) The MPEG-21 rights expression language. IEEE Transactions on Multimedia 7(2)