MPEG-7 Based Description Infrastructure for an Audiovisual Content Analysis and Retrieval System

Werner Bailer*, Peter Schallauer, Michael Hausenblas, Georg Thallinger

JOANNEUM RESEARCH, Institute of Information Systems and Information Management,
Steyrergasse 17, 8010 Graz, Austria

ABSTRACT

We present a case study of establishing a description infrastructure for an audiovisual content-analysis and retrieval system. The description infrastructure consists of an internal metadata model and access tool for using it. Based on an analysis of requirements, we have selected, out of a set of candidates, MPEG-7 as the basis of our metadata model.

The openness and generality of MPEG-7 allow using it in broad range of applications, but increase complexity and hinder interoperability. Profiling has been proposed as a solution, with the focus on selecting and constraining description tools. Semantic constraints are currently only described in textual form. Conformance in terms of semantics can thus not be evaluated automatically and mappings between different profiles can only be defined manually. As a solution, we propose an approach to formalize the semantic constraints of an MPEG-7 profile using a formal vocabulary expressed in OWL, which allows automated processing of semantic constraints.

We have defined the Detailed Audiovisual Profile as the profile to be used in our metadata model and we show how some of the semantic constraints of this profile can be formulated using ontologies. To work practically with the metadata model, we have implemented a MPEG-7 library and a client/server document access infrastructure.

Keywords: metadata, MPEG-7, content-analysis, profile, semantics, ontology, OWL

1. INTRODUCTION

A growing amount of audiovisual data are produced, processed and stored digitally. Many applications, for example those dealing with multimedia archival and media monitoring, are required to handle large amounts of digital audiovisual data. The main challenge is to index this data in order to make them searchable and thus (re-)usable. This requires the audiovisual content to be annotated, which can either be done manually, in an extremely work- and thus cost-intensive process, or by applying content-analysis algorithms that automatically extract descriptions of the audiovisual data. In both cases, the aim is to create metadata, which is a concise and compact description of the features of the audiovisual content. Metadata descriptions may vary considerably in terms of profundity, comprehensiveness, granularity, abstraction level, etc. depending on the application area, the tools used and the effort made for creating the description.

In this paper, we present a case study of establishing a metadata description infrastructure for an audiovisual content-analysis, documentation, search and retrieval system. The description infrastructure consists of the internal metadata model and the tools for accessing, modifying and storing metadata descriptions. The system, which manages video, audio and still images, consists of the following components:

- An ingesting component, which is used to import audiovisual data into the system and to perform and control automatic content-analysis tools, which extract a number of low- and mid-level metadata.
- A manual documentation component, which is used for textual descriptions and describing high-level semantic information, which cannot be extracted automatically.

* werner.bailer@joanneum.at; phone +43 316 876-1218; fax +43 316 876-1191; http://iis.joanneum.at

- A search component for query formulation and result presentation, which provides search options for both textual and content-based queries.
- A backend infrastructure providing storage and search functionalities.

The system shall enable an optimized annotation workflow and parallel operation of all the components.

The internal data model of comparable systems is usually proprietary. Until a few years ago, there were no standards available, that could be used as a basis for such a data model. A number of standards for metadata of audiovisual content are designed as exchange formats only an thus cannot be used for an internal data model (cf. Section 3.1) With the standardization of MPEG-7, the internal data model of some systems has been based on MPEG-7 or derived some concepts from it, especially in research systems (e.g. [3, 5]). We intended to follow a similar approach of basing the metadata model on an existing standard. However, as will be described in Section 4, some application specific adaptations are necessary.

The paper is organized as follows: We start by describing the requirements of such a system on the description infrastructure. Based on these requirements we define an internal metadata model. We have chosen to base our metadata model on MPEG-7 and we describe the rationale for this choice. One of the consequences resulting from the generality and flexibility of MPEG-7 is the necessity to define a subset of the standard to be used, which is formalized as a profile. Currently, the semantic constraints of a profile are only described in textual form. We thus propose an approach for formalizing these semantic constraints, so that automatic validation of conformance and mapping between different profiles is possible. We then describe the Detailed Audiovisual Profile, which we have defined for our internal metadata model.

In order to build a description infrastructure we have implemented a set of access tools. One of them is an API for working with MPEG-7 descriptions as an object hierarchy. The other is a client/server infrastructure to access and update whole or partial descriptions in a distributed system. After briefly describing two applications based on this metadata infrastructure, we conclude with the definition of the profile used and our experiences from using MPEG-7 as an internal metadata model.

2. REQUIREMENTS ON A DESCRIPTION INFRASTRUCTURE

This section describes the requirements imposed on the metadata infrastructure by a system as described above. All of them are technology independent in terms of storage technology, implementation language and tools. We have organized them into those concerning the metadata model and those concerning architectural aspects of the tools that will be part of the description infrastructure. These access tools will be based on the metadata model and enable applications to work with it.

2.1. Requirements on the AV description metadata model

Comprehensiveness

The most important goal is to design an internal metadata model that is capable of modeling a broad range of multimedia descriptions. This includes descriptions of different kinds of modalities, descriptions produced with different tools, such as results from automatic content-analysis, semantic interpretation and manual annotation. The latter are mainly in textual form, but it is nonetheless beneficial to structure these instead of having simple free text annotations.

Fine grained representation

The data model must allow to describe arbitrary fragments of media items. The scope of a description may vary from whole media items to small spatial, temporal or spatiotemporal fragments of the media item. The definition of these fragments must be flexible enough, to allow fragments that are based on audiovisual features (such as image regions representing objects or shots of a video), any higher-level features (e.g. scenes in a video) or manually defined by an annotator.

Structured representation

The metadata model must be able to hierarchically structure descriptions with different scopes and descriptions assigned to fragments of different granularity.

Modularity

The metadata model should avoid interdependencies within the description, such as between automatic analysis results from different modalities. For example, descriptions extracted only from visual or audio data and those based on audiovisual data shall be described separately. The metadata model shall separate descriptions which are on different levels of abstraction (e.g. low-level feature descriptions and semantic descriptions). This is important, as descriptions on higher abstraction levels are usually based on multiple modalities and thus cannot be assigned to one class of low-level features. Furthermore, higher level descriptions often use domain specific prior knowledge or are an interpretation of the underlying low-level descriptions, so that there can be multiple different, but equally valid high level descriptions based on the same low-level features.

Extensibility

It must be possible to easily extend the metadata model to support types of descriptions not foreseen at design time or which are domain or application specific. Extensions shall not hinder backward compatibility of existing descriptions.

Interoperability

It shall be easily possible to import metadata descriptions from other systems or to export to other systems. This means that the concepts used in the metadata model have to be sufficiently general, so that transformation between different data models is possible. This requirement can be met by basing the metadata model on an existing standard.

2.2. Requirements on the access tools

Distributed Architecture

From its very nature, a system as described above is distributed. There is typically a larger number of users accessing the system, and in most cases, different steps in the workflow (e.g. annotation, search) are done by different groups of users. On the other hand, it is desirable to have a central metadata repository, especially to facilitate search and ensure consistency of the data. Therefore access to metadata must be possible in a distributed system, with all consequences arising from possible concurrent access.

Fine grained access

The access components shall allow to access or modify not only whole descriptions but also parts thereof. One reason is that a comprehensive metadata description can become considerably large, while certain components just work on smaller fragments, so that it is more efficient to access just the required parts. The other reason is to allow parallel working of multiple automatic analysis tasks and manual documentation on the same metadata description.

Independency of Storage Technology

The data model shall be independent of the storage technology used (for example not influenced by the constraints of a relational database). The components for providing access should abstract the storage technology towards the components working on the metadata descriptions.

3. DEFINITION OF A METADATA MODEL

When defining the internal metadata model, we wanted to build as much as possible on existing standards for the description of audiovisual data. One reason is to facilitate interoperability with other systems and the other is that it simply does not make sense to reinvent the wheel. The following paragraphs describe the standards which may serve as candidates to be used as the basis of a metadata model and the rationale for choosing MPEG-7.

3.1. Standards that may serve as a basis for a AV description metadata model

The following standards have been identified as candidates for being used as the basis of our metadata model. Their strength and weaknesses have been reviewed. In the following, they are briefly described.

3.1.1. Dublin Core [8]

The Dublin Core Metadata Initiative (DCMI) has defined a set of elements for cross-domain information resource description. The set consists of a flat list of 15 elements describing common properties of resources, such as title, creator, identifier, etc. The content of the elements is primarily text without further inner structure. Dublin Core descriptions are represented using XML.

3.1.2. EBU P/Meta [4]

EBU P/Meta has been designed as a metadata vocabulary for programme exchange in the professional broadcast industry. It is not intended as an internal representation of a broadcaster's system but as an exchange format for programme-related information in a business-to-business use case. P/Meta consists of a number of attributes (some of them with a controlled list of values), which are organized into sets. P/Meta is technology independent, currently it can be represented in KLV (key, length, value) format or as XML.

3.1.3. BBC SMEF [1]

Standard Media Exchange Framework (SMEF) is a data model defined by the BBC to describe the metadata related to media items (media objects) and programmes and parts thereof (editorial objects), down to the shot level. In contrast to P/Meta, it was primarily designed for internal use and not as an exchange format. Thus only the data model, but no serialized representation is standardized.

3.1.4. MXF DMS-1 [20]

The Material Exchange Format (MXF) is a standard that defines an open file format for the exchange of audiovisual essence along with associated metadata. For describing this metadata, the Descriptive Metadata Scheme 1 (DMS-1) has been proposed. The metadata sets are defined in the SMPTE metadata dictionary. Metadata sets are organized in descriptive metadata (DM) frameworks. DMS-1 defines three DM frameworks, that correspond to different granularities of description: production (entire media item), clip (continuous AV essence part) and scene (narratively or dramatically coherent unit). When DMS-1 descriptions are embedded into MXF files they are represented in KLV format, but there exists also a mapping to a XML Schema.

3.1.5. MPEG-7 [9]

MPEG-7, formally named Multimedia Content Description Interface, is a standard for describing multimedia content, independent of the encoding of the content, and allows different levels of granularity of the description. MPEG-7 has been designed to support a broad range of applications. MPEG-7 descriptions can be represented either as XML (textual format, TeM) or in a binary format (binary format, BiM). A good overview can be found in [17].

3.2. Rationale for choosing MPEG-7 as basis of the AV description metadata model

We have chosen to use MPEG-7 as the basis of the internal metadata model of our system. In the following, we describe the rationale for this decision.

MPEG-7 has been designed as a metadata model, while some of the other candidates are mainly metadata dictionaries or have been designed as metadata exchange formats (as discussed for P/Meta in [2]). Other standards thus lack of comprehensiveness, as typically only certain subsets are needed for exchange. Many of the standards designed as exchange formats also lack of sufficient structuring capabilities, as they are rather modelled as flat lists of attributes than as hierarchical structures.

MPEG-7 has been designed for a broad range of applications. Thus most of the concepts are very general and widely applicable. Some of the other standards, for example those from the broadcast domain, are tailored towards this application area and thus lack of some generality.

MPEG-7 supports fine grained description of fragments of the content, and is the most flexible standard for describing different levels of abstraction. It allows defining arbitrary fragments of the content and does not limit structuring of these fragments.

The data model of MPEG-7 is very flexible, as far as structuring is concerned. This is especially true for the spatial, temporal and spatiotemporal structuring tools defined in part 5 [11] of the standard. This flexibility allows modularizing the description, which is a prerequisite for fine grained access to parts of the document. The structuring tools also allow hierarchical organization of description fragments with different scope.

The fact that MPEG-7 has been defined using XML Schema simplifies mapping MPEG-7 descriptions to object structures in order to build APIs. The use of XML Schema for the definition of the data model also facilitates extensibility. This is an important advantage, as no standard can fulfill all requirements of the internal metadata model of a system and some application specific extensions will be required. According to the MPEG-7 conformance guidelines [12], a data model based on MPEG-7 with some extensions still represents MPEG-7 compliant content descriptions.

The XML Schema based definition of the standard also supports a document oriented approach, which allows to model a relation between one multimedia document to one associated metadata document. The fact that MPEG-7 descriptions can also be serialized as XML documents also increases practical usability because of the number of available tools and the fact that it is human readable.

4. USING MPEG-7 AS AV DESCRIPTION METADATA MODEL

One of the strengths of MPEG-7 is its flexibility, which is provided by a high level of generality. It makes MPEG-7 usable for a broad range of application areas and does not impose too strict constraints on the metadata models of these applications. Some of the reasons described above for choosing MPEG-7 are directly related to this flexibility and openness. However, in the practical use of MPEG-7, two main problems arise from these features: complexity and hampered interoperability. The MPEG-7 requirements group has early recognized these issues [13].

The complexity arises from the use of generic concepts, allowing deep hierarchical structures, the high number of different descriptors and description schemes and their flexible inner structure, i.e. the variability concerning types of descriptors and their cardinalities. This complexity makes MPEG-7 difficult to learn and may thus sometimes cause hesitance in using the standard in products. It also makes it more difficult to implement tools for working with MPEG-7, and a lack of tools and implementation contributes to the hesitance mentioned before. Moreover the complexity of the constructs increases the documents in size, especially when serialized using XML.

The interoperability problem emerges from the openness in the definitions in the standard. There can be several standard conformant ways to structure and organize descriptions which are similar or even identical in terms of content. While conformance and interoperability can be checked on a level of used description schemes and descriptors and their structure, interoperability on a semantic level is not fully guaranteed by the standard. This means, that standard conformant MPEG-7 documents can only be understood correctly with the knowledge of how the standard has been used when creating the description. This means that an additional layer of definitions is necessary to enable full interoperability between systems using MPEG-7.

4.1. MPEG-7 Profiles

Recently, profiling has been proposed to partially solve these problems [14]. Based on the experience from other MPEG standards the means proposed are profiles and levels. Profiles are subsets of MPEG-7 tools which cover a certain functionality, while levels are further restrictions of profiles in order to reduce the complexity of the descriptions. However, the two means are related, and profiles will also significantly contribute to the reduction of complexity, as they will influence some of the complexity measures proposed in [14], e.g. the number of descriptors and description

schemes used. The definition of profiles will also facilitate interoperability between different applications working with MPEG-7 descriptions.

As MPEG-7 profiles have not been proposed when we started to define our metadata model, we did not formally define a profile, but a set of restrictions and guidelines for using MPEG-7. But for the definition of the constraints we have come to the three main steps that are also proposed in [14] for defining a profile:

1. Selection of tools supported in the profile, i.e. the descriptors and descriptions schemes used.
2. Constraints on the selected tools, e.g. reduction of cardinality of some elements.
3. Semantic constraints, i.e. defining the meaning of different elements of the description.

The profiles that will be defined in part 9 of the standard [15] will not be sufficient for a number of applications. If an application requires additional description tools, a new profile must be specified. It will thus be necessary to define further profiles for specific application areas. For interoperability it is crucial, that the definitions of these profiles are published, to check conformance to a certain profile and define mappings between the profiles.

If a profile is just defined according to the first two steps, conformance to the profile can only be ensured on the level of the structure of descriptors and description schemes. Without doubt this kind of conformance is important on a technical level and it is the only type of conformance for which validation tools are available today.

We believe that the third step is a crucial one for defining a profile. Describing the semantic constraints of MPEG-7 profiles are required for making descriptions interoperable between application and systems. Without sufficient semantic constraints mappings between different profiles cannot be defined automatically. If semantic constraints are only defined in textual form, as it is currently done for MPEG-7 profiles, conformance to a profile in terms of semantics cannot be checked automatically and mappings between different profiles can only be defined manually. In the following section, we propose an approach to formalize the semantic constraints of an MPEG-7 profile.

4.2. Formalizing semantic constraints

Using formalized vocabularies to achieve semantic interoperability in heterogeneous environments is a well known application of ontologies (see e.g. [6]). With the ongoing standardisation efforts of W3C and others the Semantic Web and its technologies, like the Description Logics-style language OWL [18] and rules languages as [21] are more and more used in diverse realms. Referring to [24], there is a strong interest that the Semantic Web is broadened to capture both the textual and the media-related digital world.

Some work in the area of semantic interoperability of multimedia metadata standards has already been done, like the widely known research of Jane Hunter [7], who proposed a core ontology for MPEG 7 and its application in various environments. Lately there have been some investigations in applying advanced techniques in this realm as in [23] and [22].

We are going to concentrate on the aspect of how the non-formal profile setups, described in textual form, can be expressed in a formal way using ontologies and rules. Having defined the framework, the resulting knowledge base can be used to perform semantic validation of a given MPEG7 document and further maintain semantic interoperability between MPEG 7 documents based on syntactically differing profiles.

We therefore propose the following steps to achieve a formal MPEG 7 profile handling:

- Define an MPEG 7 ontology for descriptors and description schemes with respect to a profile.
- Define rules for the contextual usage of descriptors in a profile to determine their semantics.
- Define mappings between MPEG 7 descriptors and description schemes and the ontological description using XSLT.
- Apply the resulting knowledge base to MPEG 7 documents using rules to classify and/or match semantic correspondent units.

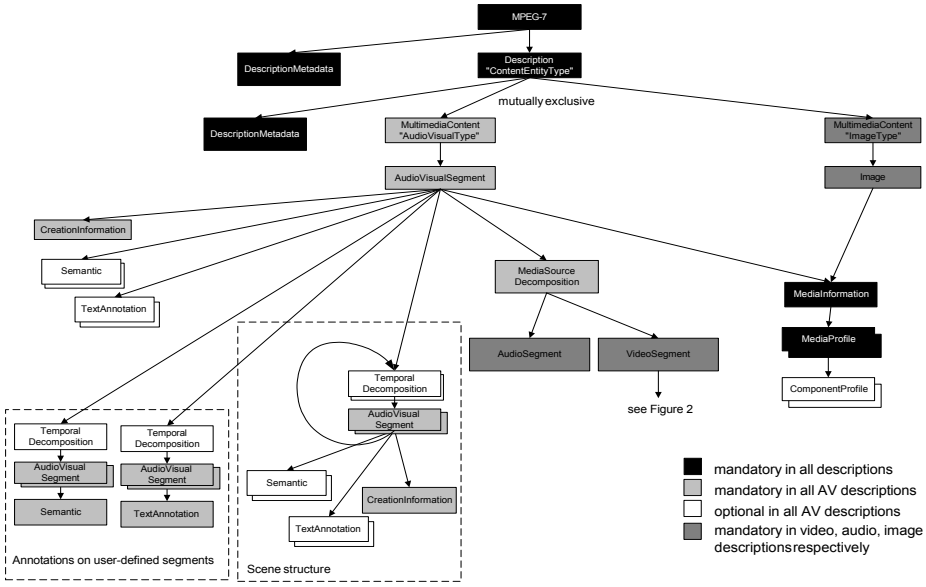

Figure 1: The top levels of a description using the Detailed Audiovisual Profile.

4.3. Description of Detailed Audiovisual Profile

Our profile defines the use of MPEG-7 in our internal metadata model. In general, it contains descriptions of video, audio and still images, which are described using a variety of automatic content analysis tools as well as manual annotation components.

In the definition of the profile we restrict the set of description schemes and descriptors that may be used at certain places in the description. We then define the semantics of the different parts of the description, whenever they are not already defined because of the use of descriptors and description schemes.

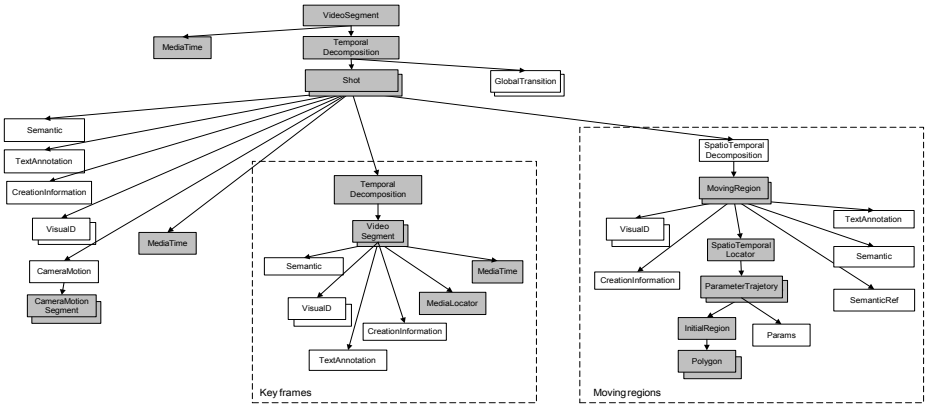

Figure 2: Description structure for the visual part.

Figure 3: Top level elements of the ontology describing the semantic constraints of the Detailed Audiovisual Profile.

4.3.1. Tool selection and constraints

A basic principle that we followed is that only one media item is described per MPEG-7 document. This means that our descriptions are allowed to contain one description of audiovisual content and the related description metadata. The profile allows the description of audiovisual content and still images (mutually exclusive, as otherwise the constraint of describing one media item per document would be violated). The top level description elements and those for audiovisual content are shown in Figure 1.

The profile allows the use of all spatiotemporal structuring tools. In the case of an audiovisual media item, we require a MediaSourceDecomposition at the top level, which splits into the parts describing the visual and audio related features only. This also includes structuring information, which is only based on these features, e.g. shots in the visual part, audio segments delimited by silence the audio part. All structuring information and other annotation, which relates to the audiovisual information (e.g. scene information derived from analysis of both visual and audio features), is attached directly to the top-level audiovisual segment. In the visual and audio part of the description, the use of the respective low-level descriptors (defined in part 3 and 4 of the MPEG-7 standard) is allowed.

Media information may only be used on the top level segment, a number of profiles and component profiles may be defined. CreationInformation, as well as semantic and text annotations are permitted on any segment.

As an example, how the detailed description of a part is organized, Figure 2 shows the description structure of the visual part. The main structuring criterion is the shot structure, as other features, such as camera motion, depend on it.

4.3.2. Semantic constraints

The main design criteria of the semantic constraints in our profile is to keep the description as modular as possible. This means that description fragments stemming from different sources or based on different modalities are kept in different parts of the description. The same approach is used for descriptions on different levels of abstraction, such as

descriptions based on features of the audiovisual content (e.g. shot boundaries) and descriptions on a higher level (e.g. scene structuring), which may have been generated using prior knowledge or other external sources.

Figure 3 depicts the formal description of the semantic constraints of some high-level elements of the Detailed Audiovisual Profile. The ontology uses the container-containee pattern in its core to model composition of entities and has been described using OWL. This domain ontology together with (external) rules, which can be encoded in e.g. SWRL allow to "understand" the profile's intention. The ontology and the rules model the following constraints for this part of the profile:

- The MPEG7 element has exactly one ContentEntityType description element.
- The content entity type element has either a AudioVisualType or a ImageType element (both are of type MultimediaContentType), each having one AudioVisualSegment or Image element respectively.
- The AudioVisualSegment at the top must have a media source decomposition, with either a AudioSegment or a VideoSegment or both. The root Audio- and VideoSegment must have the same start time and duration as the root AudioVisualSegment element.
- For scene structuring, the AudioVisualSegment may have a temporal decomposition with any recursive structure of AudioVisual/TemporalDecomposition below.
- The MediaInformation element must be present and attached to the top level AudioVisualSegment or to the Image element.
- The MediaInformation element has 1+ MediaProfiles.
- A MediaProfile may have component profiles only if it belongs to the MediaInformation of a AudioVisualSegment, that has both audio and video.

5. ACCESS TOOLS FOR THE DESCRIPTION INFRASTRUCTURE

In this section we describe the software tools and components we designed and implemented for establishing the MPEG-7 based description infrastructure.

5.1. MPEG-7 Library

For using the MPEG-7 based metadata model in the components of a system we have implemented an API for parts 3, 4 and 5 (visual, audio, MDS) of MPEG-7. The MPEG-7 library represents the types defined in the MPEG-7 schema as C++ classes. As we are using MPEG-7 as metadata model in all components of our system, we need an object-oriented and typed representation of the entities in the MPEG-7 standard and not just a generic representation of XML as it is for example provided by a DOM tree. The main advantages over using a concept like DOM are type safety, the possibility to add type specific implementation, such as specific access structures for collections of elements and increased efficiency. Because of the amount of types in the standard we decided to follow a generic approach. We have thus implemented a code generator that produces C++ classes from the MPEG-7 XSD files. This also makes the library future proof, as new code can be simply generated whenever new versions of the standard are released. The library is freely available [11].

MPEG-7 can be serialized as XML (TeM) or in a binary format (BiM). While the binary format is of course more efficient in terms of storage space, the XML representation has a number of advantages: there are a number of readily available tools for XML processing, it is human readable and it is easily extendible by way of schema extensions. Because of these advantages we decided to use the XML representation in our system. Our library therefore supports serialization to and parsing from XML.

Some additional features of the library include the support of late binding, so that application specific extensions can be added at runtime. Pattern types (such as time points) are explicitly modeled, so that inefficient string handling can be avoided. The library supports addressing nodes by XPath expressions.

5.2. Client/Server Infrastructure for Access to MPEG-7 Documents

As stated before, the analysis and retrieval system is distributed, while the metadata repository is centralized. Therefore tools for accessing metadata documents stored in the repository are required. We have designed a client/server infrastructure for access to metadata documents, with the focus of functionality on the server side. The approach we have chosen provides far more functionality than specified in the Systems part of MPEG-7 [10], which merely describes a delivery mechanism of single MPEG-7 descriptions, without the possibility to perform updates.

The document server provides read/write access to MPEG-7 documents for a number of clients. The server allows the exchange of whole documents or fragments thereof, which are addressed by XPath statements. The granularity of access is on node level. Access to parts of documents is crucial for the efficiency of the system, as MPEG-7 XML documents of larger media items tend to have considerable size. For the communication between client and server CORBA is used. Additionally, the interface has been implemented as a web service, which allows independency in terms of platform and implementation language for client and server side, using SOAP for exchanging the XML documents.

To allow concurrent operation in different parts of the document, the server supports locking of parts of the document to enable working in a multi-client environment. The document server also abstracts the infrastructure used to persistently store the document, i.e. it can serve as the interface to a file or database based storage of MPEG-7 documents. This is an advantage over direct access to a database, as partial access and especially update on a fine grained level is not supported by many XML enabled databases (cf. [19]).

6. APPLICATIONS

This section describes two applications that use the metadata description infrastructure presented in this paper.

6.1. Multimedia Mining Toolbox

The Multimedia Mining Toolbox provides users and application developers with tools for powerful combined text and content based search on multimedia data (video, audio, still images). The digital content is automatically analyzed and annotated. Manual annotation and usage of legacy metadata is included for text based search.

The media-analyze component is responsible for media import and automatic metadata extraction. During import fully automatic content analysis is performed (shot boundary detection, camera motion estimation, keyframe extraction, moving object segmentation, extraction of low-level visual features).

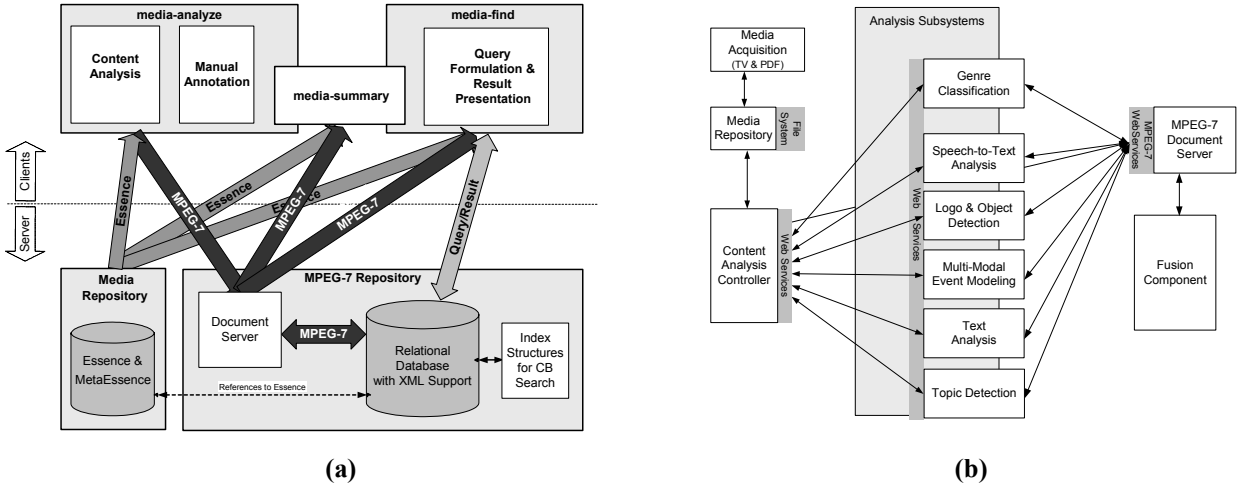

Figure 4: (a) System overview of Multimedia Mining Toolbox, (b) system architecture of DIECT-INFO media monitoring system.

The media-find component provides very fast access to the digital archive by supporting the formulation of combined text and content-based (similarity based) queries for visual content. The tool enables to efficiently search for all features automatically extracted by the media-analyze tool.

The innovative media summary viewer visualizes an entire video on one screen in terms of a temporal summary/overview and by providing efficient navigation functionality by shot structure and key frames. All content can be played back at various speeds and trimmed for further editing, the actual position is always synchronized with the temporal summary.

6.2. DIRECT-INFO Media Monitoring System

DIRECT-INFO aims to create a basic system for semi-automatic sponsorship tracking in the area of media monitoring. Its main goal is to offer an integrated system combining the output of basic media analysis modules to semantically meaningful trend analysis results, which shall give executive managers and policy makers a solid basis for their strategic decisions.

The system does 24/7 monitoring of audiovisual streams and detects segments which are relevant for analyzing. On these segments, a number of automatic content analysis tasks, such as genre classification, logo detection and event modeling are performed. These analysis tasks run in parallel to achieve the required throughput. The generated metadata is collected in a central repository, which is the basis for fusion and report generation.

7. CONCLUSION

We have implemented a description infrastructure for a distributed audiovisual content-analysis and retrieval system. The metadata model is based on MPEG-7, which has turned out to be an appropriate choice because of its generality and flexibility, especially of the concepts defined in part 5 of the standard. The hierarchical structure of the descriptions makes them modular and allows fine-grained access to parts of descriptions. The fact that the definition of MPEG-7 is based on XML Schema facilitates extensibility and allows a generic mapping to an object oriented model, which we have implemented in our MPEG-7 library. We have also established a client/server infrastructure for working with metadata documents, which allows access and update on partial documents. This is necessary for parallel annotation and to increase the efficiency when working with usually large MPEG-7 documents.

Our experience has shown, that it is necessary to restrict the generality (and thus the complexity) of MPEG-7 by defining a subset of the MPEG-7 tools to be used in a metadata model. The way to formalize this subset is the definition of a profile, and we have defined a profile for an internal metadata model. The profile is not intended to be used for the exchange of descriptions, but descriptions conforming to an exchange profile could be easily derived from the descriptions conforming to our profile.

It has turned out that for conformance on a higher level, the definition of semantic constraints in the profile is crucial. However, these constraints are currently only described in textual form. We proposed a methodology for formalizing the semantic constraints in a profile using ontologies and have demonstrated this approach for a part of our profile. This formalization allows automated validation of conformance to the semantic constraints of a profile and mapping between different profiles. It should be further investigated, how this approach can be applied to fully cover the semantic constraints in a profile and which tools can be used for automated validation and mapping.

ACKNOWLEDGEMENTS

The work described in this paper has been supported by several colleagues within JOANNEUM RESEARCH whom the authors would like to thank here. This work has been funded partially under the 5th Framework Programme of the European Union within the IST project "MECiTV" (IST-2001-37330, http://www.meci.tv) and partially under the 6th Framework Programme of the European Union within the IST project "DIRECT-INFO" (IST FP6-506898, http://www.direct-info.net).

REFERENCES

1. BBC, SMEF Data Model 1.5, 2000.
2. A. Carter, “Data-modelling terminology and P/Meta”, EBU Technical Review, Nr. 294, 2003.
3. M. Döller and H. Kosch, “An MPEG-7 Multimedia Data Cartridge”, In SPIE Conference on Multimedia Computing and Networking 2003 (MMCN03), Santa Clara, CA, January 2003.
4. EBU, The EBU Metadata Exchange Scheme, EBU Tech 3295, Mar. 2003.
5. L. Gagnon et al., “MPEG-7 Audio-Visual Indexing Test-Bed for Video Retrieval”, Proc. of Internet Imaging V Conference. San Jose, CA, USA, Jan. 2004.
6. J. Heflin and J. Hendler, “Semantic Interoperability on the Web”, Extreme Markup Languages 2000, http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf
7. J. Hunter, “Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology”, Proc. of First Semantic Web Working Symposium (SWWS), Stanford, USA (2001), pp. 261-281.
8. ISO 15836:2003: Information and documentation — The Dublin Core metadata element set, 2003.
9. ISO/IEC, Multimedia Content Description Interface, ISO/IEC 15938:2001.
10. ISO/IEC, Multimedia Content Description Interface, Part 1: Systems, ISO/IEC 15938-1:2001.
11. ISO/IEC, Multimedia Content Description Interface, Part 5: Multimedia Description Schemes, ISO/IEC 15938-5:2001.
12. ISO/IEC, Multimedia Content Description Interface, Part 7: Conformance, ISO/IEC 15938-7:2001.
13. ISO/IEC JTC 1/SC 29/ WG 11 N4039: MPEG-7 Interoperability, Conformance Testing and Profiling, Mar. 2001.
14. ISO/IEC JTC 1/SC 29/ WG 11 N6079: Definition of MPEG-7 Description Profiling, Oct. 2003.
15. ISO/IEC JTC 1/SC 29/ WG 11 N6263: Study of MPEG-7 Profiles Part 9 Committee Draft, Dec. 2003.
16. Joanneum Research MPEG-7 Library, http://iis.joanneum.at/mpeg-7
17. J. Martinez (ed.), “MPEG-7 Overview”, ISO/IEC JTC1/SC29/WG11 N4674, Jeju, March 2002.
18. OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004, http://www.w3.org/TR/owl-ref/
19. U. Westermann and W. Klas, “An Analysis of XML Database Solutions for the Management of MPEG-7 Media Descriptions”, ACM Computing Surveys, Vol. 35, No. 4, Dec. 2003, pp. 331-373.
20. SMPTE, Material Exchange Format (MXF) Descriptive Metadata Scheme - 1, SMPTE 380M.
21. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission 21 May 2004, http://www.w3.org/Submission/SWRL/
22. C. Tsinaraki, P. Polydoros, S. Christodoulakis, “Integration of OWL ontologies in MPEG-7 and TVAnytime compliant Semantic Indexing”, Proc. 16th International Conference on Advanced Information Systems Engineering (CAiSE), Riga, Latvia, June 2004.
23. G. Tummarello, C. Morbidoni, P. Puliti, A. F. Dragoni, F. Piazza, “From Multimedia to the Semantic Web using MPEG-7 and Computational Intelligence”, Fourth Intl. Conf. on Web Delivering of Music (WEDELMUSIC'04), pp. 52-59.
24. W3C Semantic Web Best Practices and Deployment (SWBPD) Working Group Charter, 2004-02-26, http://www.w3.org/2003/12/swa/swbpd-charter