

A C++ Library for Handling MPEG-7 Descriptions

Werner Bailer, Hermann Fürntratt, Peter Schallauer, Georg Thallinger, Werner Haas
JOANNEUM RESEARCH, DIGITAL – Institute for Information and Communication Technologies
Steyrergasse 17, 8010 Graz, Austria

mpeg7lib@joanneum.at

SUBMITTED to ACM MULTIMEDIA 2011 OPEN SOURCE SOFTWARE COMPETITION

ABSTRACT

We present a C++ library implementing part 2, 3, 4 and 5 of the MPEG-7 multimedia content description standard, including the updated version finalized in 2004. The library supports handling of MPEG-7 descriptions as trees of typed objects, supporting (de)serialization from/to XML. It has convenient and powerful features such as creation of subtrees by XPath statements and is extensible at runtime. The library is available for Windows, Linux/Unix and Mac OS X. It has been provided under a free use license for several years, downloaded more than 5,200 times and used in a large number of projects. It has been published under GNU LGPL in 2009. This paper discusses the key functionalities of the library as well as some exemplary applications.

Categories and Subject Descriptors

E.2 [Data Storage Representations], H.3.7 [Information Storage and Retrieval]: Digital Libraries – standards, H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems.

General Terms

Management, Standardization.

Keywords

MPEG-7, metadata, description scheme, API, library, XPath.

1. INTRODUCTION

MPEG-7 [5] is a comprehensive standard for describing multimedia content in a diverse range of applications. The complexity of the standard and the lack of tools have hindered the wide adoption of the standard. Creating, modifying and navigating the structures of an MPEG-7 description is a key functionality needed in every application creating, modifying or consuming MPEG-7 documents. This has motivated us to develop an easy to use C++ class library for handling MPEG-7 descriptions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

Copyright 2011 ACM 978-1-4503-0616-4/11/11...\$10.00.

The rest of this paper is organized as follows. The remainder of this section gives a short overview of the MPEG-7 standard and then introduces the MPEG-7 library. Section 2 describes the concept and key features of the library, and in Section 3 we present three applications of the library. Section 4 concludes the paper.

1.1 About MPEG-7

The ISO/IEC standard *Multimedia Content Description Interface* (MPEG-7) [5] has been defined as a format for the description of multimedia content in a wide range of applications. MPEG-7 defines a set of description tools, called description schemes (DS) and descriptors (D). Descriptors represent single properties of the content description, while description schemes are containers for descriptors and other description schemes. The definition of description schemes and descriptors uses the *Description Definition Language (DDL)*, which is an extension of XML Schema. MPEG-7 descriptions can be either represented as XML (textual format, TeM) or in a binary format (binary format, BiM).

A core part of MPEG-7 are the *Multimedia Description Schemes (MDS, part 5)*, which provide support for the description of media information, creation and production information, content structure, usage of content, semantics, navigation and access, content organization and user interaction. Especially the structuring tools are very flexible and enable the description of content on different levels of granularity. In addition, the *Audio* (part 4) and *Visual* (part 3) parts define low- and mid-level descriptors for these modalities.

1.2 The MPEG-7 Library

We present a C++ library, which covers parts 2, 3, 4 and 5 of the MPEG-7 standard (ISO/IEC 15938:2001 – version 1 and ISO/IEC 15938:2004 – version 2). With this library application developers are able to create multimedia content descriptions, manipulate them, serialize to XML and de-serialize – with validation – from XML. Target operating systems are Windows, Linux/Unix and Mac OS X systems, supporting the 32 and 64 bit versions of the systems respectively.

The library is intended for application developers who want to handle MPEG-7 metadata, and who do not want to struggle with complex and not type-safe XML DOM [4] programming. MPEG-7 metadata, which is stored in a hierarchical XML representation, is made available to a developer in an object oriented hierarchical class tree. The MPEG-7 library avoids the time consuming task of implementing hundreds of necessary MPEG-7 classes before the implementation of MPEG-7 functionality can start. One major design goal is to simplify

extending the implementation of classes to allow enrichment of interface functionality for certain descriptors.

Due to the lack of available implementations of the then new MPEG-7 standard, JOANNEUM RESEARCH has started the development of the library in 2002 as an API for handling MPEG-7 descriptions in own tools and projects. Realizing the need for such a library in the community of researchers and application developers, JOANNEUM RESEARCH has released the library in binary form under a free use license, first for Windows and in 2007 also for Linux and Mac OS X. In 2009 we have decided to release the library under the GNU LGPL. LGPL has been chosen over other open source licenses due to the compatibility with free software licenses on the one and commercial use of the library on the other side. The library has been downloaded by over 5,200 users as of April 2011.

The MPEG-7 library, including documentation and samples can be downloaded from [7] as source release and as binaries. Instructions how to build and use the library are included in the documentation. The web site of the MPEG-7 library also contains further MPEG-7 related resources and pointers to other MPEG-7 related web resources.

2. CONCEPT AND KEY FEATURES

The MPEG-7 library represents the MPEG-7 XML document as a tree structure. Each element or attribute, which has an XML type defined in the MPEG-7 schema, is represented by a node in this tree. Basically, each XML type is mapped to a class. This applies both for global types (those defined in the schema as a named simple or complex type) and local types (elements which do not have a named type but are defined where they are used). Sequences and choices with a maximum occurrence attribute larger than one and lists are mapped to collection classes. The inheritance hierarchy of the XML schema, i.e., extensions and restrictions of base types, are modeled as a class inheritance hierarchy. The implementation is separated by defining interfaces that model the XML types and classes implementing these interfaces.

There is one generic node type, which defines the common interface of all node classes. All classes that represent XML types are derived from the generic node type. The factory pattern is heavily used in the MPEG-7 library. For each class a factory class is generated, which handles creating and destroying instances of this class. This mechanism is especially useful due to the heavy use of type extensions in MPEG-7.

The factories return smart pointers to the nodes, which implement reference counting. Thus nodes or subtrees that are no longer referenced will be automatically destroyed.

Serialization and parsing is implemented based on Xerces C++ [2]. This XML library is licensed under the Apache License [1] and is available for a broad range of platforms. Parsing, serialization and validation are transparent to the user of the library.

2.1 Typed Access to Elements

In contrast to handling MPEG-7 documents with generic tools such as a DOM [4] API, the MPEG-7 library provides typed access to the elements in the document. For structured types, this provides a first level of conformance check as only certain types of elements and attributes may be added. For unstructured types,

this enables access to numeric or enumeration types without string processing and ensures conformance of the value to the data type.

MPEG-7 uses a number of regular expression patterns to represent structured data as strings. For commonly used types, such as media time points and durations, the MPEG-7 library provides functionality to access fields using the appropriate (numeric) data type. This saves a lot of tedious string processing in the application.

2.2 XPath Support

The MPEG-7 library supports absolute and relative XPath statements to access nodes along the child axis. In the XPath statements, numeric indices and `xsi:type` attributes are supported.

There is also a powerful feature that combines the functionality of getting a node from an XPath statement with the option to create that fragment of the tree, if it is not yet there. This feature has been added based on our experience, that it is often necessary for users to add an element at a certain position in a tree, and most of the code is needed to check whether the elements on the path to this element are already there and to create them if needed. Doing this by XPath is a very efficient way that allows creating several hierarchy levels with correct node types within one statement. The sample code that comes with the library shows the saving in terms of lines of code between creating a subtree by XPath or node by node.

2.3 Extensibility

The MPEG-7 library can be easily extended by replacing the classes of the core library with other classes providing additional (application specific) functionality. This can be done at runtime (late binding). MPEG-7 part 7 specifies that MPEG-7 documents containing proprietary types which are extensions of MPEG-7 types do conform to the standard. The MPEG-7 library supports this concept by wrapping all unknown types it encounters during parsing into a special type of node. It can be handled like any other node in the tree and will contain the unknown subtree, which can be serialized again without modification.

Another type of extensions is to use other XML schemata together with MPEG-7. The library is prepared to handle this by managing namespace prefixes for the different schemata of a document. This is for example useful when defining schemata that extend MPEG-7 by descriptors for new features. This mechanism has been used to support novel visual descriptors in an application for visual quality analysis (the respective schema is published at [7]).

3. APPLICATIONS

The range of potential applications of the library is naturally as wide as that of MPEG-7 itself, i.e., including all areas dealing with production, analysis, modification and consumption of multimedia content. From the feedback we get from users it seems that analysis of audiovisual content, applications in audiovisual media production and surveillance applications are dominant. The majority of users come from academia, however, there is a significant share of users from industry.

The library has been used in a number of large research projects in content analysis and annotation tools, authoring tools, retrieval

systems, summarization tools and much more. Some examples of these projects are listed at [7].

We discuss three application of the MPEG-7 library in more detail in the following.

3.1 Generic Metadata Engine

The Generic Metadata Engine [10] is a part of the emerging MPEG Extensible Middleware (MXM, MPEG-M) standard [8], which aims at providing APIs for parts of different MPEG standards, including MPEG-7. The Generic Metadata Engine is a read/write API for basic properties of media items, compatible with the set of core properties defined by the W3C Media Annotations WG [13]. The C++ reference implementation of the Generic Metadata Engine uses the MPEG-7 library. Source and binary distributions as well as build instructions for the Generic Metadata Engine are available at [10].

3.2 iPhoto MPEG-7 Exporter

iPhoto [3] is Apple's photo management software. It offers a plugin interface, among others for exporting photos to photo sharing sites, social networks, etc.

The MPEG-7 Exporter Plugin [9] uses this mechanism to export the photo metadata, which is kept in a proprietary file within iPhoto, as standardized MPEG-7 document. All the images selected when initiating the export are described in a single MPEG-7 document. Both metadata extracted from the EXIF information as well as metadata entered by the user are included in the exported document.

The plugin builds on the Mac OS X version of the MPEG-7 library and shows how the C++ code can be accessed from an application written in Objective C. Binaries as well as sources and an Xcode project can be downloaded from [9].

3.3 Semantic Video Annotation Suite

The Semantic Video Annotation Suite (SVAS) [11] is a suite of tools for analysis and annotation of video content by professional users, consisting of an automatic analysis tool and an interface for validation of analysis results and manual annotation.

The analysis tool automatically extracts metadata from video files for navigation and structuring (e.g., shot boundaries, key frames) and metadata which enables fast semi-automatic annotation, e.g., camera motion and face occurrences. Figure 1 shows the user interface for batch analysis.

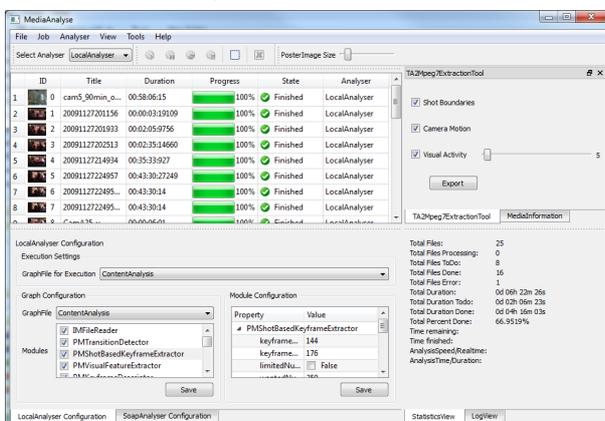


Figure 1: Interface of the analysis tool of SVAS.

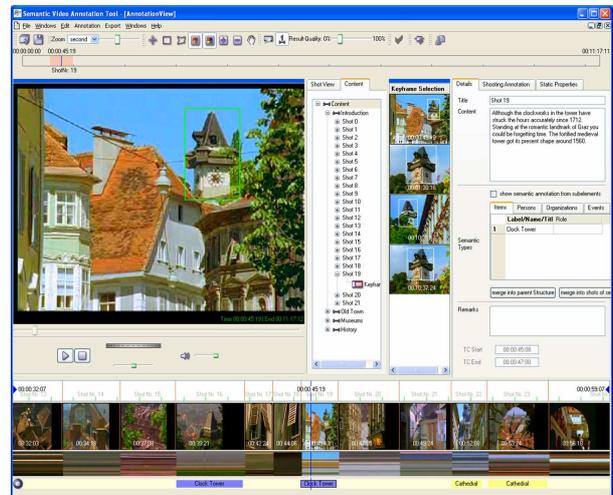


Figure 2: Interface of the interactive video annotation tool.

The following content analysis modules are included in the tool: shot boundary detection, detecting hard cuts and dissolves, key-frame extraction, camera motion estimation, face detection and extraction of global and local visual features for the key frames.

The interactive annotation tool allows viewing and editing the results of the automatic metadata extraction. Furthermore it supports to manually structure content, adding textual and semantic annotation, and the annotation of objects (image regions). As soon as content analysis results are available they can be displayed and modified.

Otherwise annotation can start from scratch, i.e. by creating a new MPEG-7 document for the content to be filled with manual annotation, including content structuring. A screenshot of the video annotation tool is shown in Figure 2.

All content structuring information, technical metadata, automatic analysis results and manual annotations are represented using MPEG-7. The MPEG-7 library is a core component of both applications and serves as internal representation of the underlying data model.

The Semantic Video Annotation Suite can be downloaded from [12].

4. CONCLUSION

We have developed a C++ class library for efficient handling of MPEG-7 documents, covering parts 2, 3, 4 and 5 of the standard. The standard has again gained more interest recently, by inclusion of parts of MPEG-7 into the MPEG MXM APIs and by specification of the Audiovisual Description Profile [6], an MPEG-7 profile targeting information extraction in audiovisual media production.

By providing typed access to the elements in the MPEG-7 document the MPEG-7 library simplifies handling documents and also facilitates conformance to the standard. The library also offers features that are not possible in generic DOM APIs, such as the efficient creation of subtrees by XPath statements. A number of quite diverse applications have been implemented using the library.

ACKNOWLEDGMENTS

The authors would like to thank their former colleagues at the Institute of Information Systems at JOANNEUM RESEARCH, who contributed to early versions of the library: Georg Mittendorfer, Rudolf Schlatte and Herwig Rehatschek. The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreements no. FP6-027122, "SALERO" (<http://www.salero.eu/>), FP7-231161, "PrestoPRIME" (<http://www.prestoprime.eu/>), and FP7-215475, "2020 3D Media" (<http://www.20203dmedia.eu/>).

5. REFERENCES

- [1] Apache License 2.0, <http://www.apache.org/licenses/LICENSE-2.0> (accessed 2011-05-08)
- [2] Apache Xerces C++. <http://xerces.apache.org/> (accessed 2011-05-08)
- [3] Apple iPhoto, <http://www.apple.com/ilife/iphoto/> (accessed 2011-05-08)
- [4] Document Object Model (DOM), <http://www.w3.org/DOM/> (accessed 2011-05-08)
- [5] Information Technology — Multimedia Content Description Interface, ISO/IEC standard 15938, 2001.
- [6] Information technology — Multimedia content description interface — Part 9: Profiles and levels, AMENDMENT 1: Extensions to profiles and levels, ISO/IEC 15938-9:2005/PDAM 1, 2010.
- [7] JOANNEUM RESEARCH MPEG-7 Resources, <http://mpeg-7.joanneum.at> (accessed 2011-05-08)
- [8] MPEG Extensible Middleware (MXM), <http://mxm.wg11.sc29.org/> (accessed 2011-05-08)
- [9] MPEG-7 Export Plugin for Apple iPhoto, <http://www.wbailer.com/work/mpeg7export> (accessed 2011-05-08)
- [10] MXM Generic Metadata Engine, <http://mxm.wg11.sc29.org/2010/08/generic-metadata-engine/> (accessed 2011-05-08)
- [11] H. Rehatschek, W. Bailer, H. Neuschmied, S. Ober, and H. Bischof. A tool supporting annotation and analysis of video. In *Reconfigurations. Interdisciplinary Perspectives on Religion in a Post-Secular Society*, Vienna, pp. 253–268, 2007.
- [12] Semantic Video Annotation, <http://www.joanneum.at/en/digital/products-solutions/semantic-video-annotation.html> (last accessed 2011-05-08)
- [13] W3C Media Annotations WG (MAWG), <http://www.w3.org/2008/WebVideo/Annotations/> (last accessed 2011-05-08)