

MPEG7ADB: Automatic RDF annotation of audio files from low level low level MPEG-7 metadata

Giovanni Tummarello, Christian Morbidoni, Francesco Piazza, Paolo Puliti

DEIT - Università Politecnica delle Marche, Ancona (ITALY)

Abstract. MPEG-7, a ISO standard since 2001, has been created recognizing the need for standardization within multimedia metadata. While efforts have been made to link the higher level semantic content to the languages of the semantic web, a big semantic gap remains between the machine extractable metadata (Low Level Descriptors) and meaningful, concise RDF annotations. In this paper we address this problem and present MPEG7ADB, a computational intelligence/signal processing based toolkit that can be used to quickly create components capable of producing automatic RDF annotations from MPEG-7 metadata coming from heterogeneous sources.

1 Introduction

While MPEG-7 and the tools of the Semantic Web (Notably RDF/S) were developed concurrently, the two efforts have been largely independent resulting in several integration challenges. At data model level, MPEG-7 is directly based on XML+Schema while the tools of Semantic Web use these just as an optional syntax format while conceptually relying on graph structures. At the semantic description level, it is thanks to a later effort [8][24] that RDF/DAML+OIL mappings have been made to allow interoperability. While such mappings are possible, their scope (semantic scene description) is currently beyond anything that can be machine automated. Previous works have also shown [4] that pure XML tools are very ineffective for handling MPEG-7 data. Although the syntax is well specified by the standard, generalized MPEG-7 usability is not simple. In fact, while it is relatively easy to create syntactically compliant MPEG-7 annotations, the freedom in terms of structures and parameters is such that generally, understanding MPEG-7 produced by others is difficult or worse. For the same reason, computational intelligence techniques, which are bound to play a key role in the applications envisioned for the standard, are not easy to apply directly. As MPEG-7 descriptions of identical objects could in fact be very different from each other when coming from different sources. Recognizing the intrinsic difficulty of full interoperability, work is currently under way [3] to standardize subsets of the base features as “profiles” for specific applications, generally trading off generality and expressivity in favor of the ease and lightness of the implementation. Necessarily, this also means to give up on interesting scenarios. In this paper we address the hard problem of “semantic mismatch”, that is, techniques to “distill” concise RDF annotations from raw, low level, MPEG-7 metadata. These techniques are implemented in a set of

tools (MPEG7ADB) by which it is possible to simply build powerful RDF audio automatic annotation components feeding on MPEG-7 low level descriptors (LLDs).

2 The MPEG7ADB

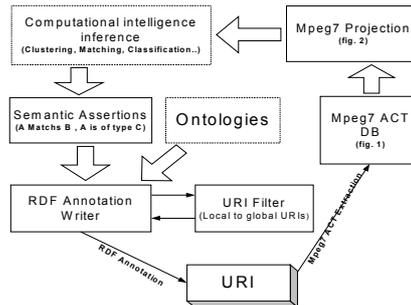


Figure 1. The overall structure of the proposed architecture.

The simplified representation of the proposed architecture (as currently implemented by the MPEG7ADB project [7]) is depicted in Figure 1. URIs are both used as references to the audio files and become the subjects of the annotations produced in standard RDF/OWL format.

When the database component is given the URI of a new audio clip to index, it will first try to locate an appropriate MPEG-7 resource describing it. At this logical point it is possible to envision several alternative models of metadata research including calls to Web Services, queries on distributed P2P systems or lookup in a local storage or cache. If this preliminary search fails to locate the MPEG-7 file, a similar mechanism will attempt to fetch the actual audio file if the URI turns out to be a resolvable URL and process it with the included, co-developed MPEG7ENC library[6].

Once a schema valid MPEG-7 has been retrieved, the basic raw sequences of data belonging to Low Level Descriptors are mapped into flat, array structures. These will not only serve as a convenient and compact container, but also provide abstraction from some of the basic free parameters allowed by MPEG-7. As an example, the MPEG7 ACT type provides the basic time interpolation/integration capabilities to handle the cases when LLDs have different sampling periods and different grouping operators applied.

To exploit the benefits of computational intelligence (e.g. neural networks) and perform clustering, matching, comparisons and classifications, each MPEG-7 resource will have to be projected to a single, fixed dimension vector in a consistent and mathematically justified way. The projection blocks performs this task, best understood as driven by a “feature space request”. A “feature space” deemed suitable for the desired computational intelligence task will be composed of pairs, one per dimension, of feature names and functions capable of projecting a series of scalars or vectors into a single scalar value. Among these, the framework provides a full set of classical statistical operators (mean, variance, higher data moments, median, percentiles etc..) that can be

cascaded with other “pre processing” such as, i.e. a time domain filter. Since MPEG-7 coming from different sources and processes could have different low level features available and not necessarily those that we have selected as the application “feature space”, the projection block will attempt to recursively predicting the missing features by means of those available (cross prediction). It is also interesting to notice that when a direct adaptation algorithm is not available, cross prediction based on neural networks proves to be, for a selected number of features, a viable alternative. For a more detailed tractation see .

Once a set of uniform projections have been obtained for descriptions within the database, classical computational intelligence methods, such as those provided in the framework and used in the example application (section 9), can be applied to fulfill the desired annotation task. Once higher level results have been inferred (e.g. piece with URI “file://c:/MyLegalMusic/foo.mp3” belongs to the genre “punk ballade”) they can be saved into “semantic containers” which will, hiding all the complexity, provide RDF annotations using terms given in an appropriate ontology pre-specified in OWL notation. Finally, prior to outputting the annotation stream, the system will make sure that local URIs (e.g. “file://foo.mp3”) are converted into globally meaningful formats like binary hash based URIs (e.g. hash “urn:md5: “ , “ed2k:// “ , etc.).

6 Producing annotations for the Semantic Web

Once obtained the mathematically homogeneous projection vectors representing the MPEG-7 files in the db, these can easily processed using a variety of well known techniques. While MPEG7ADB provides internal tools such as neural networks classifiers and clustering, many more can be interfaced at this point.

Among the tools provided by MPEG7ADB are those allowing the production of RDF annotations. Annotations produced by the MPEG7ADB will be of “rdf quality” that is, much more terse and qualitatively different than the original LLD metadata. Finally it is important to stress the importance of explicit context stating when delivering computational intelligence derived results on the Semantic Web. Virtually all the computational intelligence results are in fact subjects to change or revision according to the local state of the entity providing the annotation (e.g. the extraction settings). As new knowledge or settings could make previously obtained results invalid, this sort of inference is by nature nonmonotonic. Although the RDF framework is monotonic, it is known that results coming from nonmonotonic processes can be still mapped as long as context information are provided .

8 Implementation and conclusions

In this paper we discussed some of the challenges associated with making use of MPEG-7 low level audio descriptors to provide RDF annotations. Furthermore, we introduce MPEG7ADB, a library by which it is possible to create automatic RDF annotation components feeding not on actual (e.g. PCM or MP3) audio sources but on low

level MPEG-7 metadata descriptions. Sophisticated adaptation capabilities are provided to compensate for the many free parameters of the MPEG-7 standard itself. With these capabilities, “profile less” use can be made which fits the picture of the Semantic Web as also made of heterogeneous devices

MPEG7ADB has been implemented in Java (see [5] on why this is also computationally acceptable) and is available [7] for public use, review, suggestions and collaborative enhancement in the free software/open source model. Among the examples provide in the MPEG7ADB is a *Voice recording quality annotation component*. This purely demonstrative example, shows how a full RDF/MPEG-7/Neural Network audio annotation component can be built in approximately 40 lines of source code using MPEG7ADB. For lack of space the source code or an accurate description cannot be given directly here but is available at [7] and . Being, to the best of our knowledge, currently the only available tool with these capabilities, MPEG7ADB is hard to compare it directly but we believe it to be a good starting point for both implementation and research into audio MPEG-7 / Semantic Web annotation components.

References

- [1] ISO/IEC JTC1/SC29/WG11 N4031. MPEG-7 (2001)
- [2] ISO/IEC JTC1/SC29/WG11 N5527, MPEG-7 Profiles under Consideration, March 2003, Pattaya, Thailand.
- [3] Utz Westermann, Wolfgang Klas. “An analysis of XML database Solutions for the management of MPEG-7 media descriptions”ACM Computing Surveys (CSUR) Dec. 2003.
- [4] Ronald F. Boisvert, Jose Moreira, Michael Philippsen, and Roldan Pozo. “Java and Numerical Computing” IEEE Computing in Science and Engineering March/April 2001
- [5] Holger Crysandt, Giovanni Tummarello, MPEG7AUDIOENC – <http://sf.net/projects/mpeg7audioenc>
- [6] G.Tummarello, C.Morbidoni, F.Piazza – <http://sf.net/projects/MPEG7ADB>
- [7] Jane Hunter, “Enhancing the semantic interoperability through a core ontology”,. IEEE Transactions on circuits and systems for video technologies, special issue. Feb 2003.
- [8] Ralf Klamma, Marc Spaniol, Matthias Jarke “Digital Media Knowledge Management with MPEG-7”. WWW2003, Budapest.
- [9] G. Tummarello, C. Morbidoni, P. Puliti, A. F. Dragoni, F. Piazza “From Multimedia to the Semantic Web using MPEG-7 and Computational Intelligence”, Proceedings of Wedelmusic 2004, IEEE press, Barcellona
- [10] J. Lukasiak, D. Stirling, M.A. Jackson, N. Harders. “An Examination of practical information manipulation using the MPEG-7 low level Audio Descriptors” 1st Workshop on the Internet, Telecommunications and Signal Processing
- [11] Classification Schemes used in ISO/IEC 15938-4:Audio, ISO/IEC JTC 1/SC 29/WG 11N5727, Trondheim, Norway/Jul 2003
- [12] J. Hunter, “An RDF Schema/DAML+OIL Representation of MPEG-7 Semantics”, MPEG Document: ISO/IEC JTC1/SC29/WG11 W7807, December 2001, Pattaya
- [13] H. Crysandt, G. Tummarello, F. Piazza “An MPEG7 Library for Music”, 3rd MUSICNETWORK Open Workshop. Munich, 13-14 March 2004.
- [14] J. Hunter, C. Lagoze, “Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles”, WWW10, HongKong, May 2001.
- [15] J. van Ossenbruggen, F. Nack and L. Hardman.” That Obscure Object of Desire: Multimedia Metadata on the Web (Part I and II)”, IEEE Multimedia, to be published in 2004,