Toward MUI: Ontology assisted global identification of Audio Resources

Giovanni Tummarello[1], Christian Morbidoni[1], Dimitris Kourtesis, F. Piazza[1], P. Puliti[1]

[1] SeMedia – Semantic Web and Multimedi Group, Università Politecnica delle Marche Ancona (ITALY)
{g.tummarello,c.morbidoni}@deit.univpm.it
http://semedia.deit.univpm.it

Abstract. For machine and individuals to successfully exchange annotations about audio resources, a technique is needed to provide these with stable identifiers. In this paper we propose ontological definitions of some identification problems and discuss the preliminary conceptual structure of the Music URI Infrastructure (MUI) project. The idea behind MUI is that of designing a distributed infrastructure to provide stable or seldom changing identifiers to clients querying about audio recordings they might possess either in directly audible format (e.g MP3) or in low level metadata only (MPEG-7). In designing MUI, special care has been taken in efficiently addressing the most common use cases, e.g. naming of audio resources downloaded from the Internet. In this case, MUI attempts to make optimal use of the existing metadata, acknowledging that this might be incomplete, partially or totally incorrect.

1. Introduction

Identifiers for conceptual instances are the key concept of the Semantic Web initiative. Once identifiers have been agreed upon, it is straightforward for software agents and human alike to exchange knowledge (e.g in forms of RDF annotations) about resources in a given conceptual domain. For machine and individuals to successfully exchange annotations about audio resources, a technique is therefore needed to provide these with stable identifiers. In this paper we propose ontological definitions which pinpoint useful identification tasks and discuss the Music URI Infrastructure (MUI) project. MUI aims to be a distributed infrastructure designed to provide stable or seldom changing identifiers to clients querying about audio recordings they might possess either in directly audible format (e.g MP3) or in low level metadata only (MPEG-7 [1]). In designing MUI, special care has been taken in efficiently addressing the most common use cases, e.g. naming of audio resources downloaded from the Internet. In this case, MUI attempts to make optimal use of the existing metadata, acknowledging that this might be incomplete, partially or totally incorrect.

2. Resource identification

Based on the current W3C Semantic Web initiative standards (RDF [2], RDFS , OWL [3]), identifiers for resources can be given as URI (Uniform Resource Identifiers) or as Inverse Functional Properties (IFPs) [3]. In a few domains there are direct ways of assigning identifiers. An obvious example is the web itself, where if the conceptual object is the HTTP accessible data object, the URL themselves are URIs and act as natural identifier for Semantic Web annotations. Other notable examples are published titles, which can be directly mapped to the URN:ISBN: URI space etc.

There are however domains where mapping to a single identifier is all but simple. For example, and not surprisingly, an agreed procedure for mapping human beings to URI has so far failed to emerge. In this specific case, usually resources are identified based on IFPs they might have, that is, properties such as a Social Security Number or a mailbox, that can be said to uniquely identify that person [4].

When no central authority exists and when multiple "identification providers" entities coexist for the same conceptual domain, there exists the obvious problem of identical resources potentially being assigned non matching identifiers by the different identification providers. In this situation, this could be avoided if identifiers are chosen according to deterministic algorithms that evaluate function of the sole, common accessible, properties of the entity. One such example is identifying file resources using an agreed digital hash function as happens in P2P file sharing softwares.

In case of media objects, even limiting ourself to the field of audio, providing such identifiers is in general a very difficult problem.

3. Scenario and ontological definitions of interest in Music Identification

We start considering that a user has a digital audio resource and would like to somehow obtain an identifier for it. Such identifier will map that resource to a conceptual entity of a given class. In order for the identification to be useful, the class must naturally have properties that support the specific use case that the user or agent is interested into. Ideally, identifiers should also be given so that resources that map to identical conceptual entities in the specific use case obtain the same identifier. This, in turn, enables distributed annotations and semantic web scenarios in general.

This calls for conceptual analysis to define classes of identifier that share the same conceptual level. We propose the following meta-classification:

Table 1. Classes uses in MUI

Acoustic Entity Class	ClassIdentifier(A) = ClassIdentifier(B) IFF	Use cases
Digital audio file (DAF)	Bitwise file identity	Synchronizing remote musical filebases
Digital Encoded Audio Content (DEAC)	Encoded content bitwise identical (e.g. Mp3 audio frames)	Annotating encoding quality/performance by specific hardware

Reproducible rECCording (REC)	Once theoretically recovered from the encoding A and B would sound identical to a golden ear listener.	Author/Album/track annotations. Performance annotations.
Identifiable Musical Artwork (IMA)	Most listeners knowing all the nameable and distinguishable audio compositions ever produced would agree that both A and B are acoustical representations of the same, single, one.	Melody/tune/composition annotations

The classes are here listed in inverse order of generality but they do not form a hierarchy in the ontology/OO sense, e.g we cannot directly substitute a digital audio file in a reasoning about an Identifiable Musical Artwork, as in fact it *is not* a digital file. These classes have however the important property that instances belonging to a “narrow” class can act as identifiers for the more “general” classes, i.e. can act as Inverse Functional Properties.

For example one might annotate any DAF (e.g. file://c/mymusic/jingle.mp3) with annotations that would instead more properly pertain to the matching REC (e.g. about the performance - “Live in Camerano, Italy, December 2001”) , or even in the scope of the IMA instead (e.g. “a Christmas classic”). Given that by class definition if such mapping exist, the DAF will act as inverse functional property of both the REC and the IMA, it will be straightforward for a reasoning system to analyze the user annotations and move those pertaining to higher level classes to the correct instances so that they can be maximally useful to other users on the Semantic Web. In this case this would mean move the annotations about “Jingle bells” as a Christmas classic from the URI of the MP3 file to the URI indicating the theme (IMA).

Finally, we note how these classes are described directly by the definition explaining what does it mean for two audio resources to be assigned the same identifier (second column in the table). Such procedures are based solely on data or acoustical analysis of the audio representations and a priori knowledge about existing instances (e.g. knowing the existing compositions and the existing recordings).

3.1. Digital Audio File level identification

At client level identification can trivially happen at level of DAF, assuming that extremely rare but not impossible identifier collisions are acceptable, using standard digital hashing technique such as SHA1 or MD5. If collisions cannot be accepted then the only possibility is for MUI to centrally store a copy of each distinct DAF so to be sure that each new incoming one is not already in the DB. As such setting is highly unlikely, MUI clients by default rely only on digital hashes, in particular on SHA1. Obviously, to provide such digital hash identification there is no need for a centralized structure, as each client can evaluate it locally. The MUI infrastructure however will provide added level such as RDF metadata about the specific DAF and additional information such as known file names, first added on etc.

3.2. Digital Encoded Audio Content identification

This is somehow similar to the DAF identification but it is only concerned with the audible content of an audio file. This covers the common use case of recognizing files being digitally different due to minor metadata changes (e.g. Id3 changes in MP3 files) but having identical playable audio. This concept is useful for a number of purposes especially in the distribution/retrieval of digital content. For example, as more files with different file metadata are identified to be the same audible stream, an improved file sharing client could then use them all as download sources for the data stream.

Techniques for DEAC identification implemented in MUI are based on the MP3 file format analysis plus a digital content hash, a simplistic but effective heuristic being an SHA-1 hash of the last 1kb of the file.

3.3. Reproducible Recording Identification

At this level, we're interested in providing the same identifiers to all the DAF and DEAC that, if the loss of quality due to the encoding could be compensated, would sound identical to a golden ear listener. The existence of 2 files that can be matched according to this definition, implies that at least at some time a Reproducible Recording "master" existed and this is what we seek to identify. Clearly, RECs are all CD released tracks and all independently created digital audio files. This also implies that live recordings of audio events form their own REC and live capturing of RECs are other, independent, REC.

The reason for this is that, in theory, a live recording is always composed by multiple sources, e.g. the concert being performed and the ambient noise where someone in the crowd might be whistling some completely uncorrelated tune. In such case, the live recording would capture two different acoustic events that have similar dignity, there is nothing that could be "compensated" to remove the "whistling spectator" from the recording. Such a live recording will stand as its own REC.

As it is a common use case to locate the original track a live microphone has captured, this task is performed by mapping the REC (cellphone recording) to a IMA (see next session, e.g "Michael Jackson's "Bad") and providing the identifier of a predefine REC, if one exists, for that.

To be noticed that this definition implies that if the same track was released in 2 formats that have all above human hearing reproduction qualities, a single REC will be assigned as, for all the practical human purposes, they would be identical.

REC identification in MUI plays the most important part and happens making use of different techniques:

- Trough the identification of a DAF or DEAC - e.g. a file is recognized to be identical to one in the server that has previously been said to be acoustically equivalent to the REC in question, for example keeping a list of SHA-1 hashes of different known instances
- Through metadata identification , e.g. by identifying the id3 description of the piece to be identical to one known to be meaningful and representative to the REC.
- Through audio content analysis: by using acoustic fingerprinting algorithms.

Such procedure is described in section 4.

3.4. Identifiable Musical Artwork (IMA)

IMA identifiers are conceptually given to all individually identifiable musical compositions that can be given a name. Such task is clearly complex as it would mean to identify from the audio the musical structures composing it (e.g. The score) and then matching these with the known ones taking into consideration the ability that humans have to compensate for stylistic modifications. In practice, and in MUI, assigning IMA happens trough a loose identification of REC and providing the user with the IMA that is associated to them (if available). This means that a live recording of a song is tentatively matched to the REC of the original song and then the title is provided (if this is present in the MUI database) as REC and IMA identification are the same thing in case of most published music (but not, for example, in case of classical music CDs).

1. The MUI infrastructure overview

MUI operates in a "semi distributed" infrastructure to allow clients to receive stable or seldom changing URI or IFP for musical resources they might possess locally. By "semi distributed" we mean a mixed model where the end client interacts with multiple servers which are in turn cooperating (e.g. To resolve naming conflicts) but in a loose and decentralized way. Identifiers are assigned referring to the above mentioned classes tasks, in particular identifiers of the DAF, DEAC, REC, IMA.

At end client/server level, rather than relying on a single audio identification technique (sometime referred to as "acoustic fingerprinting techniques" [5]), MUI attempts to solve the problem by using a combination of techniques that address directly common use case scenarios.

In particular MUI makes maximal use of any form of metadata that is potentially attached to existing digital media. Metadata such as filenames, ID3 tags, digital hashes of file segments (MD5) are used by the MUI infrastructure as very efficient heuristics for speeding up and improving accuracy of the audio content based fingerprinting by several orders of magnitude. In case of wrong metadata (e.g. a completely misleading file metadata) the audio fingerprinting can almost always point out the anomaly and therefore fall back to more accurate, but time consuming, acoustic matching techniques. To be noticed how MUI takes in high consideration the amount of bandwidth transferred. While other identification services work by transmitting the full audio stream (albeit in highly compressed format such as in GSM identification services such as those offered by Shazam[1]) MUI works by transferring metadata only, be this "higher level" as mentioned or MPEG-7 low level metadata. This happens over an incremental protocol so that more detailed descriptions are transferred only if necessary for a difficult identification. Also, any MPEG-7 low level metadata processing happens based on the MPEG-7 Audio Enc [6] and MPEG-7 Audio DB [7], thus enjoy the ability to operate across heterogeneous MPEG-7 sources, e.g. to compare MPEG-7 streams also if they're not identical in the typology and resolution of the low level descriptors [8].

1 http://www.shazamentertainment.com

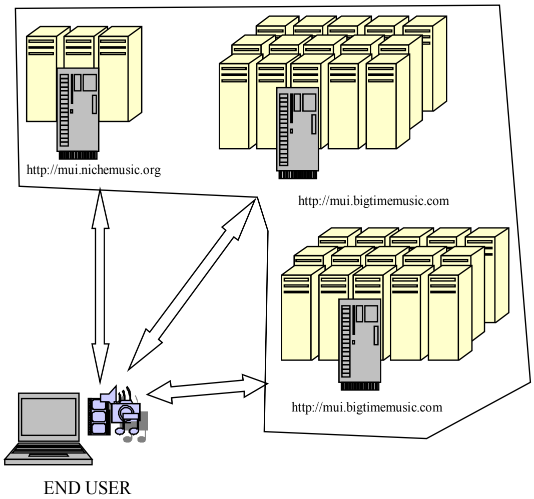

Figure 1The user may interrogate multiple MusicURI clusters set up independently.

The main acoustic fingerprinting algorithm is in fact based on MPEG-7 low level descriptions, in this sense similar to [9] and [10], but greatly improved over what made available in the MPEG-7 experimental model.

In case a MUI server is highly confident that the user is proposing an audio resource it does not know about, an import procedure will be initiated where the high definition MPEG-7 stream will be requested (or the audio resource itself) and a new identifier will be assigned. MUI servers then cooperate in P2P so that conflicting assignments are resolved based on the principles of "causing less problems" at identification levels, e.g. The identifiers that have been used the most among the servers will be the one to agree to.

While the client server database update interaction and the P2P server exchanges are opens to a number of security issues, these are not different from those from any open world Semantic Web application, so we will here consider the system to be performing in a trusted and good faith environment and leave security and social model issues for further studies.

MUI is being developed in Java (see [11], why this is in fact a good idea) on top of MetaMedia , an architecture for high performance distributed Java computing. Thanks to MetaMedia [12], MUI servers rely on cluster computing scalability in terms of both processor power and disk space. In the following figure, a client queries several MUI clusters at the same time.

4. Conclusions and related works

In this work we presented the preliminary theoretical and architectural foundations of MusicURI, tentatively a highly useful infrastructure for cooperative annotations about musical resources on the Semantic Web. The idea is that "connecting to the net" one can receive identifiers about the digital music resources one has on his/her machine so that annotations can then be shared and merged. As an identifier means little without a well specified ontological definition of the class and properties it has, we searched for previous efforts in this direction. Ontological initiatives in the domain of digital audio and music are currently undergoing development but can however be said to be in an early stage. In SUMO [13] and MILO [14], the term SoundRecording appears (albeit with the explanation that "has the form of a plastic or glass disk), which is then used and interrelated in the Digital Audio Ontology effort [15] with concepts such as "PsicologicalProcess/Hearing" or "ArtWork". As these definitions however does not directly cover the use case we focus on, we are working on a precise ontological definition of the issues with preliminary results which have been sketched in this paper. Works are in progress under multiple points of view, an early implementation of the MusicURI basic client server infrastructure is available as Open Source on the project home page [2].

References

[1] "ISO/IEC JTC1/SC29/WG11 N4031. MPEG-7" 2001
[2] "Resource Description Framework", http://www.w3.org/RDF/
[3] "OWL Web Ontology Language", http://www.w3.org/TR/owl-features/
[4] D Brickley, L Miller , "FOAF Namespace Document" 2004 http://xmlns.com/foaf/0.1/
[5] P Cano, E Batlle, E Gomez, LCT Gomes, M Bonnet , "Audio Fingerprinting: Concepts and Application" 2002 FSKD
[6] Holger Crysandt, Giovanni Tummarello, , "MPEG7AUDIOENC"
[7] G.Tummarello, C.Morbidoni, F.Piazza , "MPEG-7 Audio Db"
[8] G. Tummarello, C. Morbidoni, P. Puliti, A. F. Dragoni, F. Piazza , "From Multimedia to the Semantic Web using MPEG-7 and Computational Intelligence" 2004 WedelMusic 2004 , Barcellona
[9] Fröba, Bernhard; Allamanche, Eric; Herre, Jürgen, Kastner, Thorsten; Hellmuth, Oliver; Cremer, Markus: , "Content-based Identification of Audio Material Using MPEG-7 Low Level Description" 2001 ISMIR Conference, Bloomington, USA
[10] Holger Crysandt , "Music identification with MPEG-7" 2003 AES 115th Convention, New York
[11] Ronald F. Boisvert, Jose Moreira, Michael Philippsen, and Roldan Pozo , "Java and Numerical Computing " 2001
[12] P. Astrelli, M. Di Bono, M. Martinelli, O.Salvetti, M. Catasta, C. Morbidoni, F.Piazza, G. Tummarello, O. Signore , "Toward a scalable Multimedia Metadata infrastructure using distributed computing and Semantic Web technologies" 2005 EWIMT, London UK
[13] I. Niles, A. Pease , "Towards a standard upper ontology" 2001 FOIS
[14] I. Niles, A. Terry , "A General-purpose, Mid-level Ontology." 2004 IKE
[15] "Digital Audio Ontology initiative" http://dao.cim3.net/

2 http://semedia.deit.univpm.it/musicuri/