A Personalized Audio Server using MPEG-7 and MPEG-21 standards

Emmanouil Skondras,
Dpt of Informatics & Computer Technology
Technological Educational Institute of Western Macedonia, 52100 Kastoria, Greece
e-mail: manolisskondras@yahoo.gr

Angelos Michalas
Dpt of Informatics & Computer Technology
Technological Educational Institute of Western Macedonia, 52100 Kastoria, Greece
e-mail: amichalas@kastoria.teikoz.gr

Malamati Louta
Dpt of Informatics and Telecommunications Engineering, University of Western Macedonia
50100, Kozani, Greece
e-mail: louta@uowm.gr

Ioannis Anagnostopoulos
Dpt of Computer Science & Biomedical Informatics,
University of Central Greece
35100, Lamia, Greece
e-mail: janag@ucg.gr

***Abstract*— This paper presents a semantic model which delivers personalized audio information. The personalization process is automated and decentralized. The metadata which support personalization are separated in two categories: the metadata describing user preferences stored at each user and the resource adaptation metadata stored at the server. The multimedia models MPEG-21 and MPEG-7 are used to describe metadata information. The Web Ontology Language (OWL) language is used to produce and manipulate the relative ontological descriptions.**

I. Introduction

Nowadays, the volume of multimedia data is increasing rapidly in many information channels. Network infrastructures enable information repositories to be accessed from users all over the world. The delivery of multimedia services is a common task. However more users tend to require information retrieval services which include high quality features such as semantic description and personalization of information. In this paper a prototype application that delivers personalized audio information to users is described.

The application uses MPEG-7 and MPEG-21 for the description of audio content as well as the users' preferences, which are the basic metadata for the semantic description of multimedia content.

The metadata information is managed using Web Ontology Language (OWL) ontologies. On one hand, the information that describes the user preferences is created and stored at the client. On the other hand, the server-side information contains audio resources and resource adaptation metadata, minimizing thus, the central storage requirements and complexity. This significantly reduces the response time of the audio server, handling multiple concurrent requests from users during normal operation.

The remainder of the paper is organized as follows. In section 2, the related research literature is revisited. Section 3 presents an overview of the standards followed in this study. Section 4 describes the software architecture that supports the prototype application, as well as the software elements and modules required. Finally, Section 5 concludes our work and presents possible future extensions and plans.

II. Related Work

The rapid increase in multimedia content has challenged the academic and industrial communities into the development of information retrieval and management tools enhanced with personalization and adaptation capabilities. An increasing number of these applications use defined standards to support personalization.

In the work described in [1], the authors propose a personalization process that customizes rich multimedia documents to the needs of an individual reader. Multimedia documents, such as textbooks, reference materials and leisure materials, inherently use techniques that make them accessible for people with disabilities, who are incapable of using printed materials. The authors address issues of establishing user personalization profiles, as well as adapting and customizing content, interaction and navigation. Customization of interaction and navigation leads to different user interfaces, as well as different structural content presentation. Customization of content includes insertion of a summary, synchronization of sign language video with highlighting of text, self-voicing capability, alternative support for screen readers, or reorganization of layout to accommodate large fonts.

The work described in [2] examines a metadata based approach, supporting the personalization process for knowledge workers who interact with distributed information objects. An architecture supporting the personalization process is described, along with a prototype personalization environment. Its metadata are decentralized, in terms that the information is stored locally on client-side. The authors discuss the advantages, as well as the challenges of the suggested approach.

The authors of the approach presented in [3] introduce a wide view of personalization and user profiles, making the preferences available to a range of services and devices. Behind every instance of personalization is a profile that stores the user preferences, context of use and other information capable to deliver user experiences that describe individual users' needs and preferences. It is based upon the fact that users' needs depend on the context and current situation, (e.g. “At home”, “In a Meeting”, “In the Car”).

In the approach proposed in [4] the user of a multimedia database returns relevance ranking to his retrieval intention for top n data of a retrieval result. Using this feedback information, the framework produces an adjustment data inherent to the user and utilizes it for personalization.

In [5], the design and the implementation of a MPEG-7 based Multimedia Retrieval System for Film Heritage is presented. The multimedia content has been indexed using an Annotation Tool based on MPEG-7 standard. An MPEG-7 Compliant Ontology in OWL DL has been developed to fulfill the requirements of the system. This ontology has been instantiated so that the retrieval process can be handled. This work has been assessed during the validation of the CINeSPACE project, which aims to design and implement a mobile rich media collaborative information exchange platform, accessible through a wide variety of networks (cities WiMax, WANs etc.) for the promotion of Film Heritage.

In the work described in [6], the issues associated with designing a video personalization and summarization system in heterogeneous usage environments are addressed, providing in parallel, a tutorial that introduces MPEG-7 and MPEG-21 within these contexts. The authors introduce a framework for a three-tier summarization system (server, middleware and client). The server maintains the content sources, the MPEG-7 metadata descriptions, the MPEG-21 rights expressions and content adaptability declarations. The client exploits the MPEG-7 user preferences and the MPEG-21 usage environments, in order to retrieve and display the personalized content. The middleware contains the personalization and adaptation engines, which select, adapt, and deliver the summarized rich media content to the user. The system includes MPEG-7 annotation tools, semantic summarization engines, real-time video transcoding and composition tools, application interfaces for PDA devices as well as browser portals.

In [7] a model for integrating semantic user preference descriptions within the MPEG-7/21 standard is presented. The approach preserves the hierarchical structure of the MPEG-21/7 user preference descriptions. The implementation of the model is presented, which allows descriptions of domain ontologies, semantic content descriptions and user preference descriptions in an OWL/RDF environment and also supports automatic conversion of the proposed extensions to MPEG-21/7 descriptions.

Finally, the work described in [8] presents an agent based multimedia broadcasting framework using MPEG-21/7 and Foundation for Intelligent Physical Agents (FIPA) standards [9]. A FIPA implementation is used as platform for exchanging user preferences and program information, based on the classical client-server architecture. The user preferences are modeled in respect to the MPEG-21/7 User Preference description scheme.

III. USED STANDARDS

This section makes an overview of the standards used for the development of the application prototype. These standards include MPEG-7 [10], MPEG-21 [11] and OWL [12].

MPEG-7 is a multimedia content description standard. The description is associated with the content itself, to allow fast and efficient searching for material that is of interest to the user. MPEG-7 is formally called Multimedia Content Description Interface. It does not deal with the actual encoding of moving pictures and audio, like MPEG-1, MPEG-2 and MPEG-4. It uses XML to store metadata such as timecoding of particular events, or synchronizing lyrics to a song.

The MPEG-21 standard aims at defining an open framework for multimedia applications. MPEG-21 uses the architectural concept of the Digital Item. A Digital Item is a combination of resources (such as videos, audio tracks, images), metadata (such as descriptors, identifiers), and structures describing the relationships between resources. Digital Items are declared using the Digital Item Declaration Language (DIDL). MPEG-21 Digital Item Adaptation (DIA) architecture and the MPEG-7 Multimedia Description Schemes (MDS) for content and service personalization provide a Usage Environment which models user preferences. The Usage Environment Description is part of the MPEG-21 DIA architecture and consists of the following description elements:

- The User Characteristics, which specify user features, including:
  - The User Info, where user information is stored.
  - The User Preferences, describing the user browsing, filtering and search preferences.
  - The Usage History, where the history of user interaction with digital items is presented.
  - The Presentation Preferences, which describe user preferences concerning the means of presentation of multimedia information.
  - The Accessibility Characteristics, responsible for content adaptation concerning users with auditory or visual impairments.
- The Terminal Capabilities, which describe the technical characteristics of user devices.
- The Natural Environment Characteristics, providing information about the location and time of a user in a particular environment, as well as audio-visual characteristics which may include noise levels and illumination properties of the natural environment.
- The Network Characteristics, which specify the network characteristics parameters including bandwidth utilization, packet delay and packet loss.

The Web Ontology Language (OWL) is adopted so as to create the relative ontologies and provide a common semantic understanding between the components involved in the personalization process. OWL is a family of knowledge representation languages for authoring ontologies endorsed by the World Wide Web Consortium.

They are characterised by formal semantics and RDF/XML-based serializations for the Semantic Web.

IV. Software Architecture

In this section we present the architecture of our model. The architecture is decentralized in respect to the information required to achieve personalization. User related preferences are created and stored locally at each client. Resource adaptation metadata along with the resources are the only to be composed and stored centrally at the server. As an effect, distribution of both computational load and personalization data is achieved improving framework scalability.

The server contains the music tracks and the respective audio metadata using MPEG-7 in an MPEG-21 structure. The music tracks are divided in sixteen different music categories (pop, classical, dance, electronic etc.). Audio metadata include user defined metadata (artist, producer, production year and category), technical oriented metadata (bitrate, sample rate, track duration, upload date and last download date, audio channels, audio format, file size) as well as usage history metadata (track's popularity in respect to all tracks, track's popularity in its category and recommended similar tracks). Table 1 presents a sample of the audio metadata structure.

```
<mpeg21:DIDL xmlns:mpeg21="urn:mpeg:mpeg21:2002:02-mpeg21-NS"
                 xmlns:mpeg7="http://www.mpeg.org/MPEG7/2000">
 <mpeg21:Container>
  <mpeg21:Item>
   <mpeg21:Descriptor>
    <mpeg21:Statement mpeg7:mimeType="text/plain">Metadata about  audio
     track.</mpeg21:Statement>
     </mpeg21:Descriptor>
      <mpeg21:Component>
       <mpeg21:Resource mpeg7:mimeType="application/xml">
        <mpeg7:Mpeg7>

         <mpeg7:CreationPreferences>
          <mpeg7:Title mpeg7:preferenceValue="12"
            xml:lang="en">track1.mp3</mpeg7:Title>
         </mpeg7:CreationPreferences>

         <mpeg7:CreationInformation>
          <mpeg7:Creation>
           <mpeg7:Creator>
            <mpeg7:Role
            mpeg7:href="urn:mpeg:mpeg7:cs:RoleCS:2001:AUTHOR" />
            <mpeg7:Agent xsi:type="PersonType">
             <mpeg7:Name>
              <mpeg7:GivenName>John</mpeg7:GivenName>
              <mpeg7:FamilyName>Johny</mpeg7:FamilyName>
             </mpeg7:Name>
            </mpeg7:Agent>
           </mpeg7:Creator>

           <mpeg7:Creator>
            <mpeg7:Role
            mpeg7:href="urn:mpeg:mpeg7:cs:RoleCS:2001:Publisher"/>
            <mpeg7:Agent xsi:type="PersonType">
             <mpeg7:Name>
              <mpeg7:GivenName>George</mpeg7:GivenName>
              <mpeg7:FamilyName>Smith</mpeg7:FamilyName>
             </mpeg7:Name>
            </mpeg7:Agent>
           </mpeg7:Creator>

           <mpeg7:Abstract>
            <mpeg7:FreeTextAnnotation>VeryGood
            </mpeg7:FreeTextAnnotation>
            <mpeg7:StructuredAnnotation>
             <mpeg7:What><mpeg7:Name>Music Track</mpeg7:Name>
             </mpeg7:What>
            </mpeg7:StructuredAnnotation>
           </mpeg7:Abstract>

           <mpeg7:CreationCoordinates>
            <mpeg7:CreationDate>
             <mpeg7:TimePoint>2010-05-11</mpeg7:TimePoint>
             <mpeg7:Duration>P7D</mpeg7:Duration>
            </mpeg7:CreationDate>
           </mpeg7:CreationCoordinates>
          </mpeg7:Creation>
         </mpeg7:CreationInformation>

         <mpeg7:ClassificationPreferences>
          <mpeg7:Genre mpeg7:preferenceValue="47"
            mpeg7:href="urn:mpeg:ContentCS:1">
           <mpeg7:Name xml:lang="en">Acappella</mpeg7:Name>
          </mpeg7:Genre>
         </mpeg7:ClassificationPreferences>

         <mpeg7:MediaLocator>
          <mpeg7:MediaUri>tracks/track1.mp3</mpeg7:MediaUri>
         </mpeg7:MediaLocator>
         <mpeg7:MediaTime>
          <mpeg7:MediaTimePoint>T00:00:00F100</mpeg7:MediaTimePoint>
          <mpeg7:MediaDuration>T00:13:07F100</mpeg7:MediaDuration>
         </mpeg7:MediaTime>

         <mpeg7:MediaFormat>
         <mpeg7:Content mpeg7:href="urn:mpeg:mpeg7:cs:ContentCS:2001:2">
          <mpeg7:Name xml:lang="en">audio</mpeg7:Name>
         </mpeg7:Content>
         <mpeg7:Medium
          mpeg7:href="urn:mpeg:mpeg7:cs:MediumCS:2001:2.1.1 ">
          <mpeg7:Name xml:lang="en">HD</mpeg7:Name>
         </mpeg7:Medium>
         <mpeg7:FileFormat
          mpeg7:href="urn:mpeg:mpeg7:cs:FileFormatCS:2001:3">
          <mpeg7:Name xml:lang="en">MP3</mpeg7:Name>
         </mpeg7:FileFormat>
         <mpeg7:FileSize>787082</mpeg7:FileSize>
         <mpeg7:BitRate mpeg7:minimum="N/A" mpeg7:average="8000"
          mpeg7:maximum="N/A"></mpeg7:BitRate>
         <mpeg7:AudioCoding>
          <mpeg7:Format
           mpeg7:href="urn:mpeg:mpeg7:cs:AudioCodingFormatCS:2001:1">
             <mpeg7:Name xml:lang="en">MP3</mpeg7:Name>
          </mpeg7:Format>
          <mpeg7:AudioChannels mpeg7:track="2"></mpeg7:AudioChannels>
          <mpeg7:Sample mpeg7:rate="22050" mpeg7:bitPer="0">
          </mpeg7:Sample>
         </mpeg7:AudioCoding>
        </mpeg7:MediaFormat>
       </mpeg7:Mpeg7>
      </mpeg21:Resource>
     </mpeg21:Component>
    </mpeg21:Item>
   </mpeg21:Container>
</mpeg21:DIDL>
```

Table 1. Sample of the audio metadata structure

Each client organizes its metadata using MPEG-21/7 user preferences element of MPEG-21 usage environment. The client's metadata rely on user's preferences (favorite music categories and top 10 music tracks in each category). Table 2 presents a sample of the user preferences metadata structure.

```
<mpeg21:DIDL xmlns:mpeg21="urn:mpeg:mpeg21:2002:02-mpeg21-NS">
<mpeg21:Container>
  <mpeg21:Item>
    <mpeg21:Descriptor>
      <mpeg21:Statement mimeType="text/plain">This item is a metadata block
                                        about John's preferences.</mpeg21:Statement>
     </mpeg21:Descriptor>
     <mpeg21:Component>
       <mpeg21:Resource mimeType="application/xml">
         <Mpeg7 xmlns= "http://www.w3.org/2000/XMLSchema-instance"
           type="complete">
          <UserPreferences>
            <UserIdentifier protected="true">
              <UserName>John</UserName>
            </UserIdentifier>
            <UsagePreferences allowAutomaticUpdate="true">
              <FilteringAndSearchPreferences protected="true">
                <ClassificationPreference>
                  <Genre href="urn:mpeg:GenreCS" preferenceValue="32">
                    <Name> Acappella </Name>
                  </Genre>
                  <Genre href="urn:mpeg:GenreCS" preferenceValue="75">
                    <Name> Classical </Name>
                  </Genre>
                  <Genre href="urn:mpeg:GenreCS" preferenceValue="40">
                    <Name> HipHop </Name>
                  </Genre>
                  <Genre href="urn:mpeg:GenreCS" preferenceValue="18">
                    <Name> Jazz </Name>
                  </Genre>
                  <Genre href="urn:mpeg:GenreCS" preferenceValue="24">
                    <Name> Pop </Name>
                  </Genre>
...
                  <Genre href="urn:mpeg:GenreCS" preferenceValue="63">
                    <Name> Rock </Name>
```

```
                    </Genre>
                  </ClassificationPreference>
                </FilteringAndSearchPreferences>
              </UsagePreferences>
            </UserPreferences>
          </Mpeg7>
        </mpeg21:Resource>
      </mpeg21:Component>
    </mpeg21:Item>
  </mpeg21:Container>
</mpeg21:DIDL>
```

Table 2. Sample of the user preferences metadata structure

Suitable OWL ontologies for metadata manipulation have been created. The server uses an OWL ontology for managing the audio metadata, while the client uses the personal metadata based on the relevant OWL ontology as it is presented in the class diagram of Figure 1.

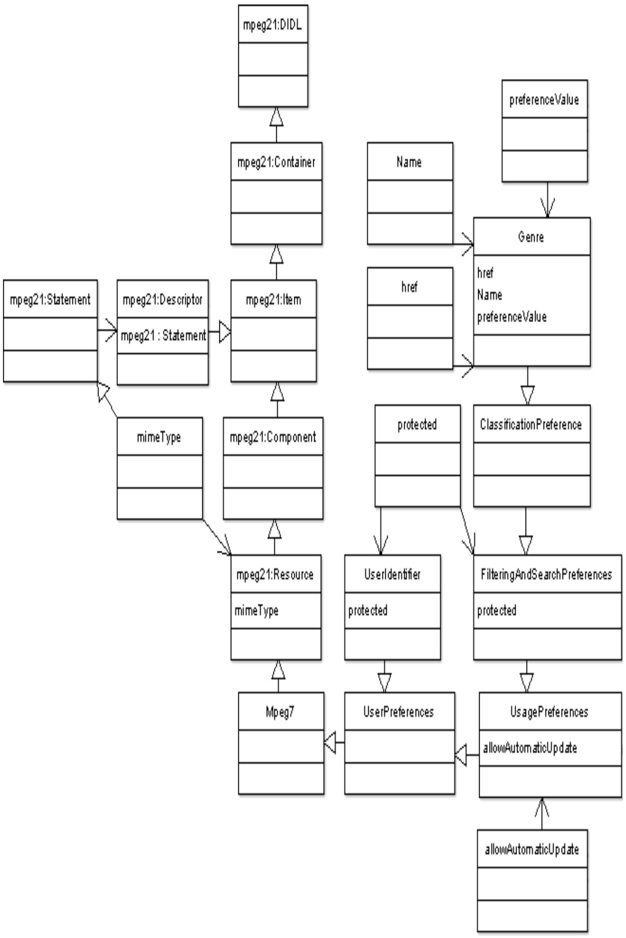

Figure 1. OWL ontology about user preferences metadata

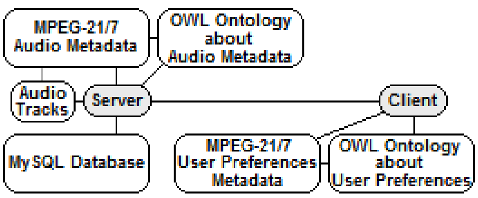

Figure 2. The basic modules of our architecture

A client requests to listen to a music track, upload a new music track or retrieve a catalog of music tracks that match specific criteria (e.g. belong in a specific music category, comply with user preferences). The server can propose music tracks to the clients, based on their choices and preferences.

The server is also capable to interact with the client and propose music files that share similar metadata with the requested music tracks (artist, music company, year, type, format, file-size, time duration, channels, sample rate, bitrate and audio quality).

Figure 2 presents the basic modules of our architecture. User defined metadata of a specific resource are created from the client when a new music track is uploaded to the server. The server uses the Java Media Framework to analyze the uploaded audio track and extracts technical oriented metadata. After these actions, the server formats and inserts all the audio metadata into the relevant metadata file according to the relative standards and to the OWL ontology. The file upload operation is described in Figure 3.

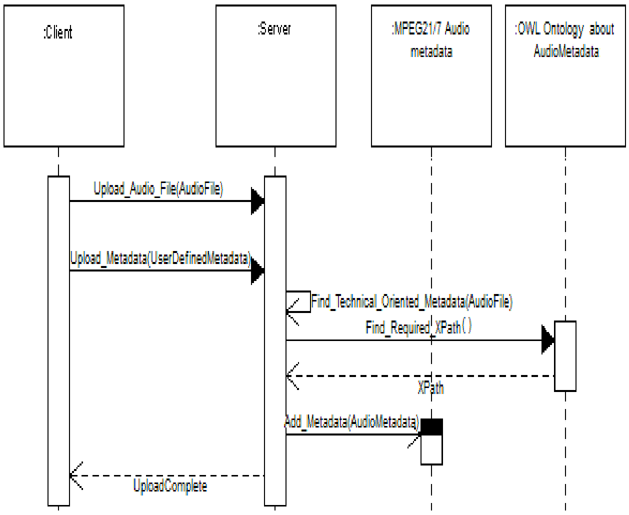

Figure 3. Updating the audio metadata

The client interacts with the server and sends the user's preferences along with the respective credentials. The server checks user credentials, establishes a session with the client and promotes audio tracks to the client according to its preferences. The client receives the response regarding the request sent, and updates its preferences. In case the user preferences are significantly modified throughout a session, the client sends its updated preferences to the server. The system operation is graphically illustrated in the sequence diagram of Figure 4. MySQL is used for handling the user's credentials at the server.

The audio server and client modules are developed using Java and Java Media Framework. The classes created are instantiated with the corresponding XML file description. Java objects are the instances of all the selected MPEG-21 and MPEG-7 concepts mapped to Java classes.

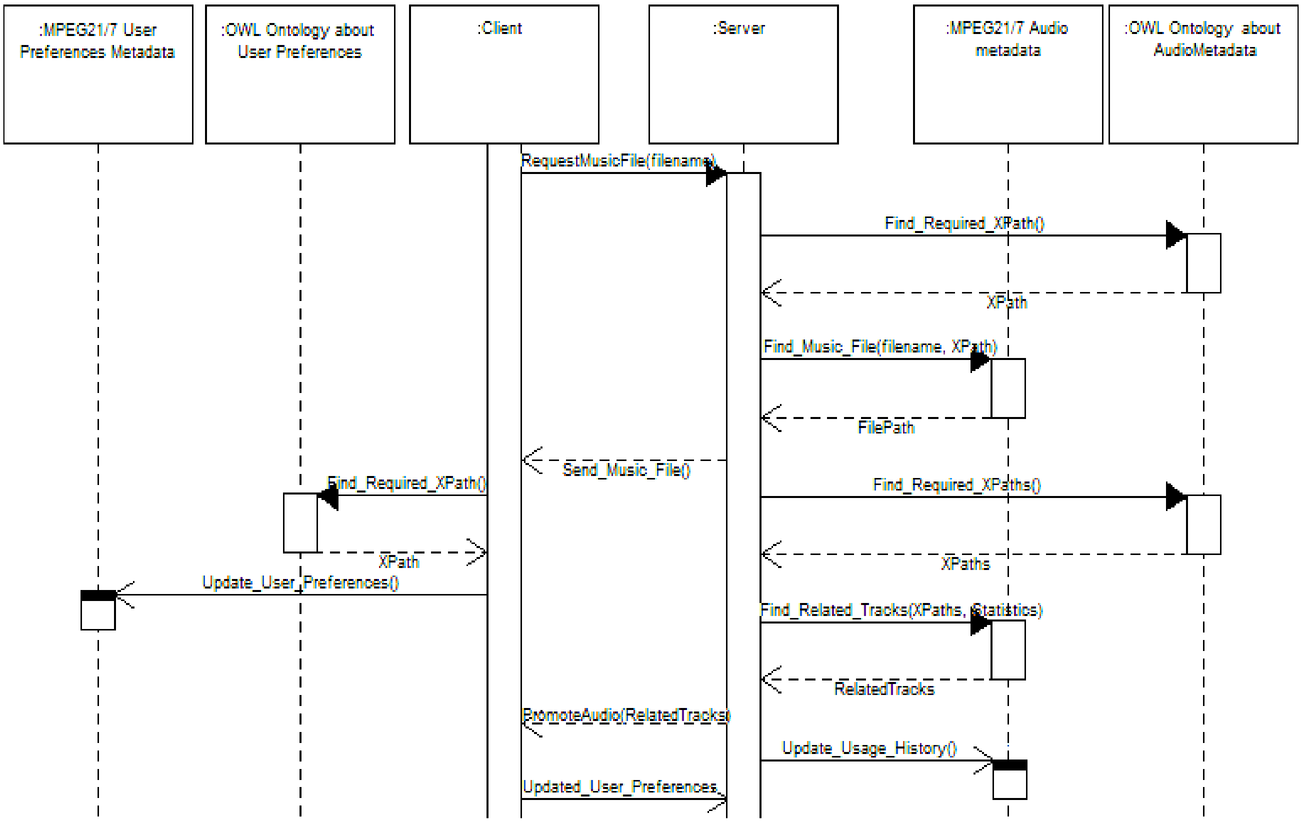

Figure 4. Server proposes audio tracks

V. CASE STUDY

This section presents an example of our framework's functionality. The client contacts the server and sends his preferences through the relative OWL ontology. Consequently the client requests a Jazz audio track, named 'Lullaby.mp3'. The server uses the relative OWL ontology to manipulate the audio metadata and sends the requested track to the client.

Figure 5 presents the form which plays the requested audio track and promotes similar trucks to the client. It contains the appropriate components for playing audio and two lists. The list at the top, presents the promoted audio tracks according to user's preferences. The list at the bottom presents the promoted tracks according to the usage history stored at the server. The user can request any track from the two lists.

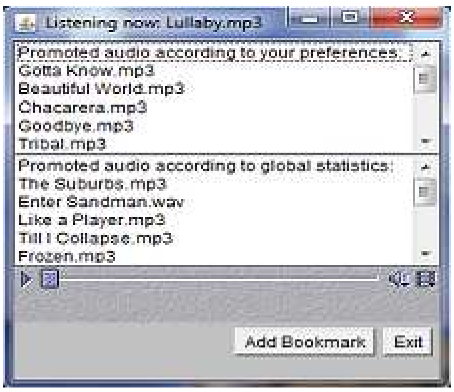

Figure 5. The client's form for playing audio

The selection of a music track results to an update of user preferences metadata at the client and the usage history resource adaptation metadata at the server. Figure 6 presents the relative user preference metadata block before and after the client's request. The 'preferenceValue' concerning music track's genre is increased.

```
before:
<Genre href="urn:mpeg:GenreCS" preferenceValue="18">
  <Name>Jazz</Name>
</Genre>...
after:
<Genre href="urn:mpeg:GenreCS" preferenceValue="19">
  <Name>Jazz</Name>
</Genre>...
```

Figure 6. The relative user preference block before and after the client's request

```
before:
<mpeg7:CreationPreferences>
  <mpeg7:Title mpeg7:preferenceValue="12" xml:lang="en">Lullaby.mp3</mpeg7:Title>
</mpeg7:CreationPreferences>...
<mpeg7:ClassificationPreferences>
 - <mpeg7:Genre mpeg7:preferenceValue="47" href="urn:mpeg:ContentCS:1">
     <mpeg7:Name xml:lang="en">Jazz</mpeg7:Name>
   </mpeg7:Genre>
</mpeg7:ClassificationPreferences>...
after:
<mpeg7:CreationPreferences>
  <mpeg7:Title mpeg7:preferenceValue="13" xml:lang="en">Lullaby.mp3</mpeg7:Title>
</mpeg7:CreationPreferences>...
<mpeg7:ClassificationPreferences>
 - <mpeg7:Genre mpeg7:preferenceValue="48" href="urn:mpeg:ContentCS:1">
     <mpeg7:Name xml:lang="en">Jazz</mpeg7:Name>
   </mpeg7:Genre>
</mpeg7:ClassificationPreferences>...
```

Figure 7. The relative audio metadata blocks before and after the client's request

Figure 7 presents the CreationPreferences and the ClassificationPreferences metadata blocks stored at the server before and after client's request. The 'preferenceValue' of the former block shows the number of times the relative audio track has been requested -from all users. Respectively, the 'preferenceValue' of the later shows the times the relative genre of tracks has been requested from all users. Both values are increased after client's request.

VI. CONCLUSIONS

Our approach relies on MPEG-21 and MPEG-7 standards to achieve personalization. MPEG-21 DIDL and DIA are used handling Digital Items declaration and user preferences, respectively. Moreover, the appropriate OWL ontologies are used for managing the metadata. The architecture is decentralized, since each client organizes its own metadata locally. The server hosts the resource adaptation metadata along with the resources, proposing audio tracks to the clients based on their choices and preferences.

Applications conforming to MPEG-21 and MPEG-7 may use the metadata produced by our framework. Additionally the model presented in this paper can be applied to any type of multimedia resources.

Future work includes the implementation of the proposed framework over web services. Furthermore, it would be useful to develop a mechanism capable of combining user preferences with resource adaptation metadata, according to content related preferences formally defined at the user. Content related preferences could include relevance importance specifications improving framework's functionality and increasing personalization intelligence.

REFERENCES

[1] H.L. Petrie, G. Weber, W. Fisher, "Personalization, interaction, and navigation in rich multimedia documents for print-disabled users", IBM Systems Journal, Vol 44, No 3, pp. 629-635, IBM, 2005 [2] David L. Hicks, Klaus Tochtermann, "Personalizing Information Spaces: A Metadata Based Approach", International Conference on Dublin Core and Metadata Applications, pp.213-220, Tokyo, Japan, October 24-26, 2001

[3] Tatiana Kovacikova, Francoise Petersen, Mike Pluke, Valentin Alonso Alvarez, Giovanni Bartolomeo, Antonella Frisiello, Erik Zetterstrom, Scott Cadzow, "Personalization and User Profile Standardization", Seventh IASTED International Conference on Communication Systems and Networks, Track 629-043, Palma de Mallorca, Spain, September 1-3, 2008

[4] Tsuyoshi Takayama, Hirotaka Sasaki, Shigeyuki Kuroda, "Personalization by Relevance Ranking Feedback in Impression-based Retrieval for Multimedia Database", Journal of Systemics, Cybernetics and Informatics, Vol 3, No 2, pp.85-89, IIISCI, 2005

[5] Yolanda Cobos, Cristina Sarasua, Maria Teresa Linaza, Ivan Jimerez, Ander Garcia, "Retrieving Film Heritage content using an MPEG-7 Compliant Ontology", Third International Workshop on Semantic Media Adaptation and Personalization, pp. 63-68, Prague, Czech Republic, December 15-16, 2008

[6] Belle Tseng, Ching Yung Lin, John Smith, "Using MPEG7 and MPEG-21 for Personilizing Video", IEEE Multimedia Journal, Vol. 11, Num. 1, pp. 42-53, IEEE Computer Society, January 2004

[7] Chrisa Tsinaraki, Stavros Christodoulakis, "Semantic User Preference Description in MPEG-7/21", Fourth Hellenic Data Management Symposium, pid 6, Athens, Greece, 25-26 August, 2005

[8] Munchurl Kim, Jeongyeon Lim, Kyeongok Kang, Jinwoong Kim, "Agent-Based Intelligent Multimedia Broadcasting within MPEG-21 Multimedia Framework", ETRI Journal, Vol 26, Num 2, Daejeon, South Korea, April 2004

[9] The Foundation for Intelligent Physical Agents, http://www.fipa.org/

[10] MPEG-7, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

[11] MPEG-21, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm

[12] Web Ontology Language, World Wide Web Consortium, http://www.w3.org/TR/owl-features/