An Ontology-based 3D Audio Metadata Management System

Jaemin Hwang, Jeonghyeok Kim, Joohyung Song, Jaekwon Kim, Jongsik Lee and Sanggil Kang
Department of Computer Science, Inha University, Incheon, Korea
Intelligence Mobile Laboratory for Information Science, Incheon, Korea
{nulpis1, sspwiz, pc3103103, jaekwonkorea, jslee, sgkang}@ inha.edu

ABSTRACT

In this paper, we develop an ontology-based 3D audio metadata management system. We also define an ontology hierarchy structure for 3D audio metadata. Conventional audio metadata databases (DBs) have to be worked with manually to search and manage the metadata, because audio metadata is not indexed and not supported by most systems. To solve this problem, we propose a 3D audio metadata database management system (DBMS). In the experiment section, we show the feasibility of our system.

KEYWORDS

Ontology, multimedia, metadata, database

1. INTRODUCTION

In accordance with the growth in the 3D audio market, a lot of research has been done on 3D audio systems [1]. 3D audio has many types of metadata, such as duration, position, and listener's information [2]. The metadata has a lot of recording information, which is managed separately in an XML file. However, searching through and management of metadata in the XML file are analog processes. Maintenance and management of the metadata is inefficient. Searching for information within the XML file is also done by hand, according to a search history of relevant sources [3]. In order to solve this problem, metadata in an XML file is added to a database (DB) in ontology form. With this method, it is possible to efficiently manage the information in the source and the metadata through the Database Management System (DBMS).

This paper is organized as follows. Section 2 talks about the background and related work. In Section 3, we discuss how to structure the 3D audio metadata management system by using ontology, and Section 4 explains the experimental example configured on the basis of Section 3. Finally, we conclude and discuss future challenges in Section 5.

2. RELRATED WORKS

2.1 3D audio databases

Research on 3D audio has been actively carried out since the 1980s [1]. Among the studies, the essential sound source in order to create 3D audio is a head-related-transfer function (HRTF) sound source [4].

HRTF is data containing information, and measures the frequency of the sound reaching the ears of the listener from the beginning of the sound source. HRTF is the most important element of the 3D audio database. Metadata, such as the position of the listener, the location of the sound output, and the system information, is recorded in an XML file.

2.2 Ontology

Ontology is specifications on the target concepts explicitly achieved, as expressed in a computer model, to address the current constraints as to the usability and the concept types [5]. It is explicitly defined, as in the technical environment of the Internet. For an ontology, all group members must agree. It not up to an individual. And there are many ways to formalize understanding of a computer program. There are several types of markup language for building an ontology, and in this study, we used Extensible Markup Language (XML) to construct the 3D audio data ontology.

2.3 Other works

A 3D audio database was constructed by CIPIC [6]. It included, in total, 1250 entries of 25 azimuths and heights of 45 to 50 human subjects. Photos were taken by high resolution digital photograph. Pinna was measured with a Polhemus 3D stylus digitizer. Flare angle was defined via 3D data acquisition with the stylus digitizer. It recorded more angular and other information than in previous studies. For this reason, the recorded metadata was further increased.

IRCAM measured 10 heights from -10 to 90 degrees, and variable azimuths from 0 to 360 degrees, for a total of 187 points [7]. They added compensated data with a diffuse field. These files were written to a Subject ID number for the person measured, and then divided and recorded in the metadata files and in a WAV MAT file. The XML file comprises five categories: general, subject, room, system and data. The general category is composed of information such as title, date, copyright team goal, and notes. The subject category is composed of information such as gender, hair style, notes, and ID. The room category is composed of information such as room type and dimensions. The system category is composed of information like software, microphone, speaker and head tracker. Data is information like measured grid, quantification, sampling rate, and measured length.

While CIPIC recorded detailed anthropometric values to find the correlation between the overhead and the HRTF through detailed anthropometric values, IRCAM focused the HRTF DATA of a plurality of persons to be recorded. Both studies recorded the detailed metadata in XML form, but there are several disadvantages in that they have too much information, and it is difficult to use and control it, when viewing a number of attributes.If the metadata can be recorded via ontology and managed through a DBMS, an advantage would be the ability to manage it easily and efficiently. We propose a system that can efficiently manage the 3D audio metadata required by the researcher.

3. ARCHITECTURE OF DB

Figure 1 is a structure of the audio database system. The system consists of five parts, such as an audio DB, an XML descriptor, an XML parser & editor, the DBMS, and a UI. The audio database consists of 3D audio sources and XML metadata. The 3D audio source is usually an HRTF source. The XML descriptor explains the metadata of the. The audio files and saves it as an XML file

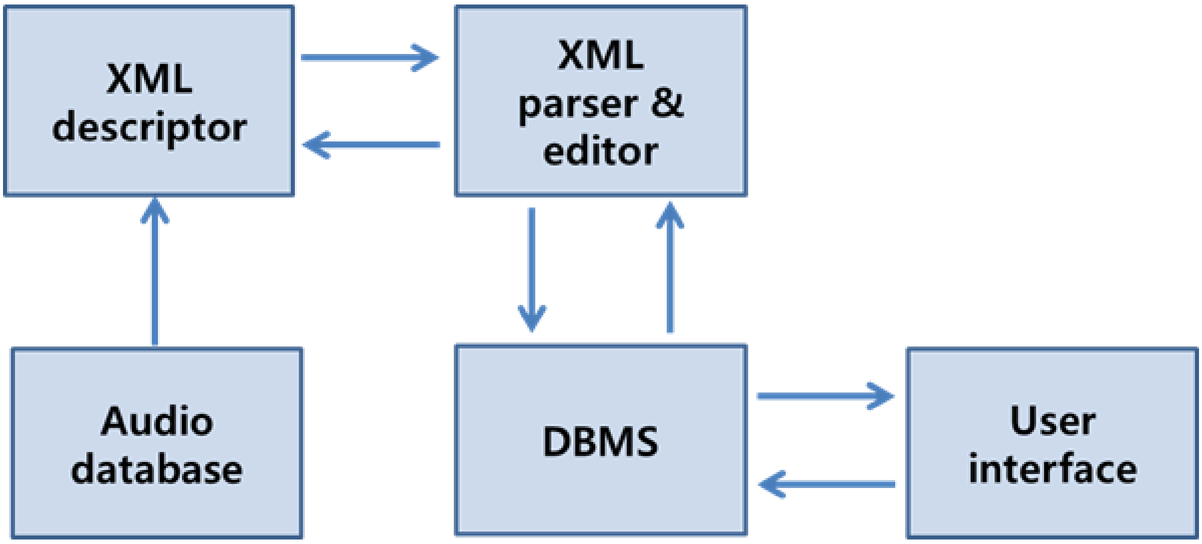

Figure 1: structure of DB

Metadata is described as a form of ontologies. The ontology hierarchy is described in Figure 2, and the examples of Web Ontology Language (OWL) [8] are shown in Figure 3. In the XML parser & editor, an XML parser reads the metadata described via the XML descriptor.

The XML editor revises the metadata according to the ontology hierarchy. Metadata modified to fit the ontology structure are stored in a DB by using ADO.NET [9]. The DBMS manages the metadata stored in the DB through the XML parser & editor. The data are stored in table form, which is shown in Figure 4. Also, the data management in the DB is described in Figure 4. The 3D audio source and metadata that are managed through the DBMS can be transmitted to the user through the UI manager. The metadata of the 3D audio source is described by the XML descriptor. It is to be stored in the DBMS through the XML Parser. The metadata is managed through the DBMS. The UI manager plays the 3D audio source and displays information from the 3D audio metadata in the DB.

Figure 2 is the ontology composition for the 3D audio database. It consists of a property to represent the class, and a class to represent each object. A straight arrow indicates each subclass and property. The general class creates general information about the 3D audio source in Figure 1. The date class is a subclass of the general class. The general class has necessary attributes, such as title, team, goal, copyright, and note. The listener class creates the listener information about the 3D audio source. The listener class has class information, position, and unit as subclasses. The information class has necessary attributes, such as sex, head length, hair, and age. The position class has necessary attributes, such as x, y, and z. The unit class has a unit size attribute. The source class creates the source information about the 3D audio source. The listener class includes class environment, position, and information as subclasses.

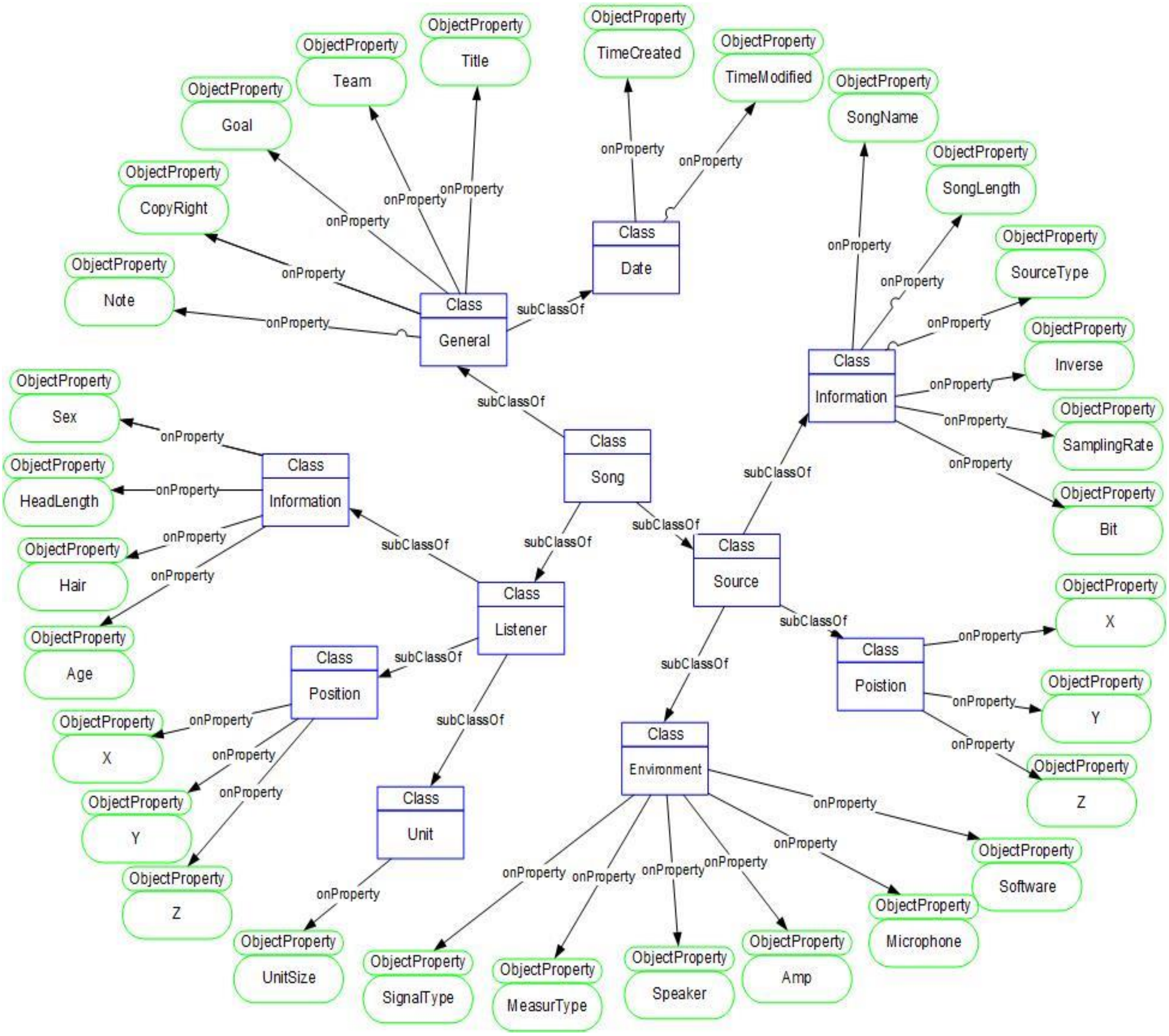

Figure 2: Ontology hierarchy structure

The environment class has necessary attributes, such as signal type, measure type, speaker, amp, microphone, and software. The position class has necessary attributes, such as x, y, and z. The information class has necessary attributes such as song name, song length, source type, inverse sampling rate, and bit.

Figure 4 is the ontology DB schema for managing the metadata. It reads the metadata from the source XML file and stores it in the DB. Each table records the defined metadata. It is possible to systematically support the management and use of the 3D audio metadata, because the metadata are managed via the DBMS.

```xml
<rdf:RDF
xmlns:rdf="http://www.w3.org/2004/02/10-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2004/02/rdf-schema#"
xmlns:owl="http://www.w3.org/2004/02/owl#"
xmlns:songs="http://www.inha.ac.kr/songs#">

<!-- OWL Class Definition - Song Information -->

<owl:Class rdf:about="http://www.inha.ac.kr/songs#songinfo">
    <rdfs:label>The song info</rdfs:label>
    <rdfs:comment>The class of all song informations.</rdfs:comment>
</owl:Class>

<!-- OWL Subclass Definition - general -->
<owl:Class rdf:about="http://www.inha.ac.kr/songs#general">

<!-- General is a subclassification of songinfor -->
    <rdfs:subClassOf rdf:resource="http://www.inha.ac.kr/songs#songinfo"/>
    <rdfs:label>General of songs</rdfs:label>
    <rdfs:comment>General, Information of general about songs.</rdfs:comment>
</owl:Class>

<!-- OWL Subclass Definition - listener -->
<owl:Class rdf:about="http://www.inha.ac.kr/songs#listener">

<!-- Listener is a subclassification of songtinfo -->
    <rdfs:subClassOf rdf:resource="http://www.inha.ac.kr/songs#songinfo"/>
    <rdfs:label>Listener of songs</rdfs:label>

    <rdfs:comment>Listener,  Information of listener about song.</rdfs:comment>
</owl:Class>

<!-- OWL Subclass Definition - source -->
<owl:Class rdf:about="http://www.inha.ac.kr/songs#source">

<!-- Source is a subclassification of songinfo -->
    <rdfs:subClassOf rdf:resource="http://www.inha.ac.kr/songs#songinfo"/>
    <rdfs:label>Source of songs</rdfs:label>

    <rdfs:comment>Source, Information of source about song.</rdfs:comment>
</owl:Class>
</rdf:Description>
</rdf:RDF>
```

Figure 3: OWL definition

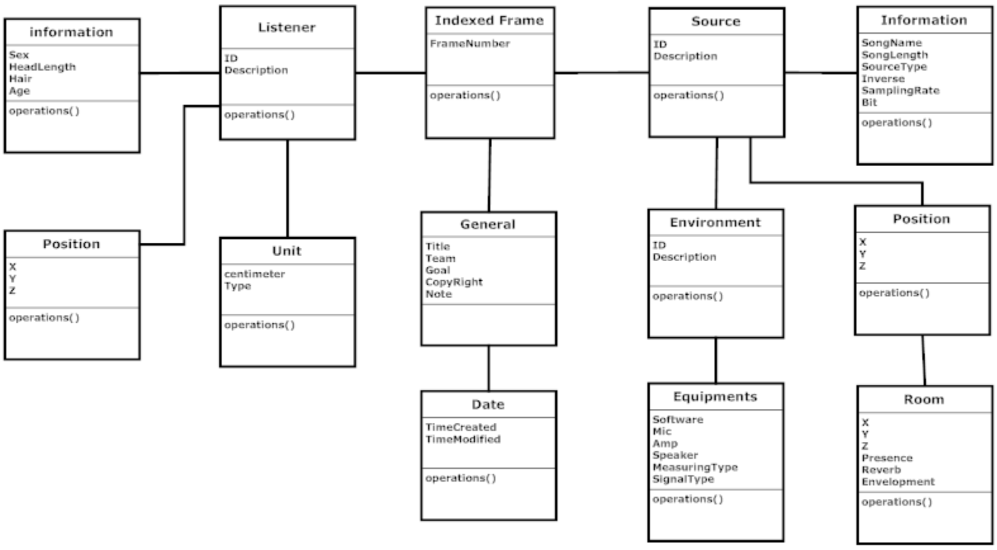

Figure 4: DB schema

4. EXPERIMENT

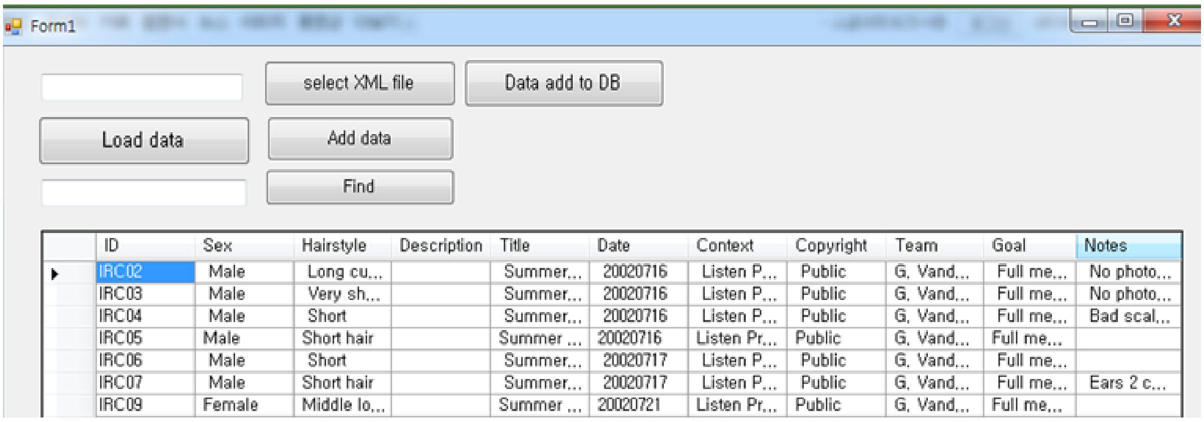

Figure 5: XML parser & editor

Figure 5 is a program for importing the metadata into the DBMS and editing the metadata. It consists of the control segment and the display segment. The control segment is located at the top. The display segment is located at the bottom. Control consists of five buttons: select XML file, load data, add data, data add to DB, and find. The select XML file button selects the XML metadata file to be added. The add data button adds XML metadata to the DBMS. The load data button loads metadata from the DBMS. The display can show information from the DBMS, and allows the user to edit the information.

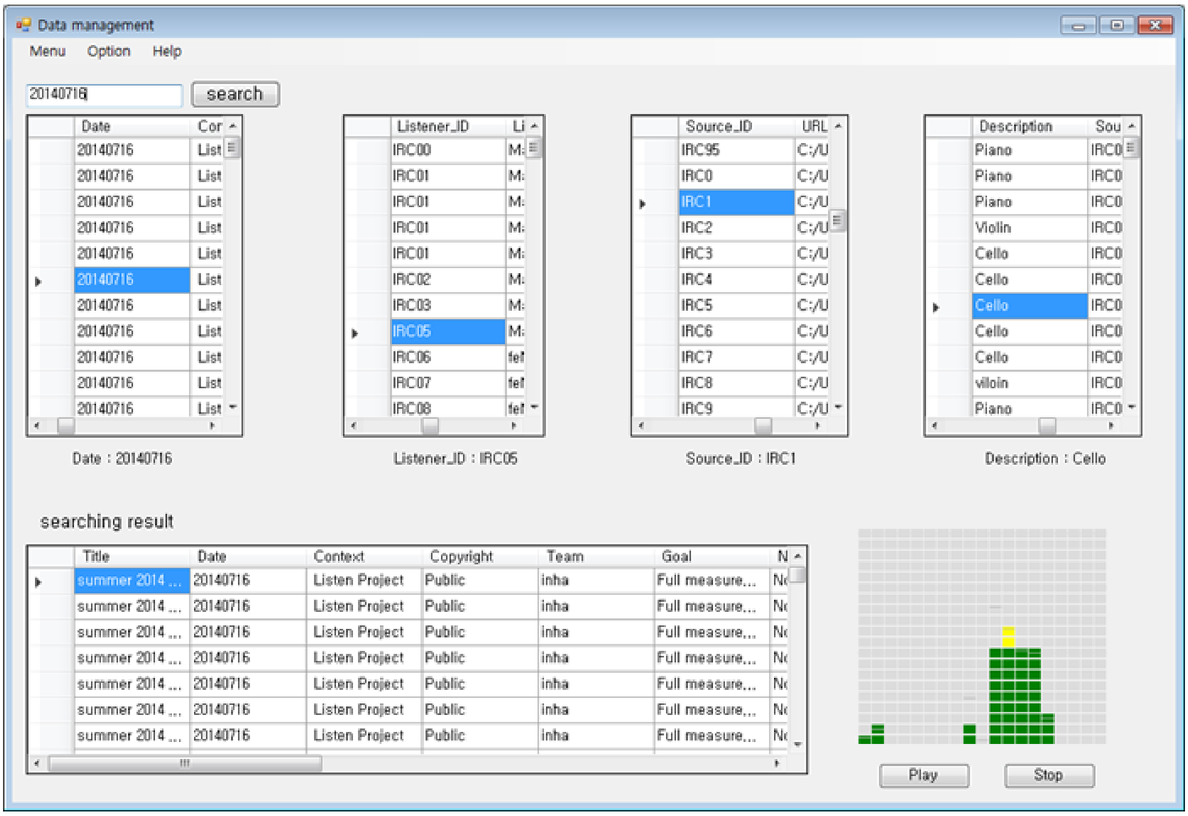

Figure 6: 3D audio metadata management program

The data add to DB button saves the modified data to the DBMS. If you want to find data from the DBMS, use the find button.

Figure 6 is the 3D audio metadata management program. It consists of a search function and the results display. The search is located at the top, and the results display is located at the bottom. The search consists of four display areas. First, find the information by pressing the search button at the top. Search results are output in the first display portion. If you select information from the first display, the ontology search result is displayed in the secondary display. The same happens with the third and fourth display sections. After selecting the relevant information in the search, the final search results are shown in the results display.

5. CONCLUSION

Conventional 3D audio metadata are recorded in an analog manner. It is controlled manually for managing and searching. Because of this, the metadata of the audio sources are used only for recording. Its search and management are very inconvenient. In this paper, we propose a 3D audio metadata management system that can easily and conveniently manage 3D audio metadata. We also define the types of 3D audio metadata in an ontology. We expect this technique can be used for other types of data. However, it is necessary to modify data that does not fit into the defined ontology. That is a limitation of the ontology technique, so it is also necessary to study more convenient and automatic ontology techniques

ACKNOWLEDGMENT

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2015-H8501-15-1016) supervised by the IITP (Institute for Information & communications Technology Promotion)

REFERENCES

[1] H Takada, M Date, Y Kurokawa, “Synergistic effect between image and sound in 3D audio visual communication system,” Perception ECVP, perceptionweb.com, 2014

[2] G Potard, I Burnett, “An XML-based 3D audio scene metadata scheme”, University of Wollongong, 2004

[3] N Gupta, A Barreto, M Joshi, “HRTF database at FIU DSP lab,” in Magnetism, Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, 2010, pp. 169–172.

[4] VN Ganesh, “Implementation of 3D Audio Effects using Head Related Transfer Function (HRTF) for Real Time Application using Blackfin Processor,” searchdl.org, 2014

[5] T Ashino, M Fujit, “Definition of a web ontology for design-oriented material selection,” Data Science Journal., 2006

[6] VR Algazi, RO Duda, DM Thompson, “The cipic hrtf database,” Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop pp. 99–102,2001

[7] V Larcher, O Warusfel, JM Jot, J Guyard “Study and comparison of efficient methods for 3-d audio spatialization based on linear decomposition of hrtf data” Audio Engineering Society, 2000.

[8] D Esposito,” Building web solutions with ASP. NET and ADO. NET, dl.aco.org, 2002”

[9] DL McGuinness, “OWL web ontology language overview” W3C