Linked Data Collection and Analysis Platform of Audio Features

Yuri Uehara, Takahiro Kawamura, Shusaku Egami,
Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga

Graduate School of Information Systems,
University of Electro-Communications, Tokyo, Japan
{uehara.yuri,kawamura,egami.shusaku}@ohsuga.is.uec.ac.jp
{seiuny,tahara,ohsuga}@uec.ac.jp

Abstract. Audio features extracted from music are commonly used in music information retrieval (MIR), but there is no open platform for data collection and analysis of audio features. Therefore, we build the platform for the data collection and analysis for MIR research. On the platform, we represent the music data with Linked Data. In this paper, we first investigate the frequency of the audio features used in previous studies on MIR for designing the Linked Data schema. Then, we build a platform, that automatically extracts the audio features and music metadata from YouTube URIs designated by users, and adds them to our Linked Data DB. Finally, the sample queries for music analysis and the current record of music registrations in the DB are presented.

Keywords: Linked Data, audio features, music information retrieval

1 Introduction

Recently, there are a large number of studies on music. Music Information Retrieval (MIR) deals with music on computers and has been studied in various ways [1]. In these studies, audio features extracted from music are frequently used, however, there is no open platform for collecting data including the audio features for music analysis. Therefore, we propose the platform for MIR research in this paper.

On the platform, we used Linked Data format, since it is suitable for complex searches for audio features and songs-related metadata. Note that this platform is designed for music-related researchers and developers, who intend to analyze music information and create their own applications, e.g., recommendation mechanism. Use of a listener is beyond the scope of this paper.

2 Schema Design of Music Information

In this section, designing Linked Data schema, including audio features and music metadata is described.

2.1 Selection of audio features

Audio features refer to the characteristics of the music, such as *Tempo* representing the speed of the track, the features used in MIR studies vary. For example, Osmalskyj et al. used *Tempo* and *Loudness* to identify cover songs [3]. Luo et al. used the audio features *Pitch*, *Zero crossing rate*, etc. to detect of common mistakes in novice violin playing [4].

Thus, we investigated the frequency of the audio features used in previous MIR studies. We collected 114 papers published in the International Society of Music Information retrieval (ISMIR)[1] in 2015, which is the top conference in the field of MIR. Then, we selected some of the audio features according to the policies: Features appeared just once in the publications can be ignored, etc.

2.2 Design of the schema

We defined original properties for selected audio features, excluding *Key* and *Mode*, since there were no existing properties for them, or the properties are not appropriate for our purpose. Then, we classified properties of audio features into some classes for making them easy to use. Table 1 shows the classes and properties corresponding to the audio features.

Table 1. Class and property of audio features

Class	Property	Audio Features	Explanation	Count
Tempo	tempo	Tempo	speed	28
Key	key	Key, Mode	tonality, difference of major and minor chord	5
Timbre	zerocross	Zero crossing rate	the rate at which the signal changes from positive to negative or back	3
	rolloff	Roll off	ratio of bass which accounts for 85 percent of the total	3
	brightness	Brightness	ratio of high-range (more than 1500Hz)	2
Dynamics	rmsenergy	RMS energy	the average of the volume (root mean square	2
	lowenergy	Low energy	ratio of sound low in volume	2

We designed the music schema with the video id (URI) of YouTube. In Fig. 1, the id: dvgZkm1xWPE indicates a song "Viva La Vida" by Coldplay. In the graph, the id node links to the classes of audio features and then links to each audio feature. Also, we added some degrees for categorizing numerical values in the features. The *tempo* has tmarks based on tempo values[2], which is a measure of the speed marks: Slow means 39 or less bpm, Largo means 40 − 49 bpm, etc.

In addition, we extended the schema of music metadata. In the graph of metadata, the video id node links to the class of metadata, and then links to the detailed value, as well as the graph for the audio features. Also, some nodes such as the artist name are linked to the external DBs like DBpedia.

There is the graphs of "Viva La Vida" by Coldplay, and thus the audio features graph and the metadata graph can be linked with the video id of YouTube.

[1] http://www.ismir.net/
[2] http://www.sii.co.jp/music/try/metronome/01.html

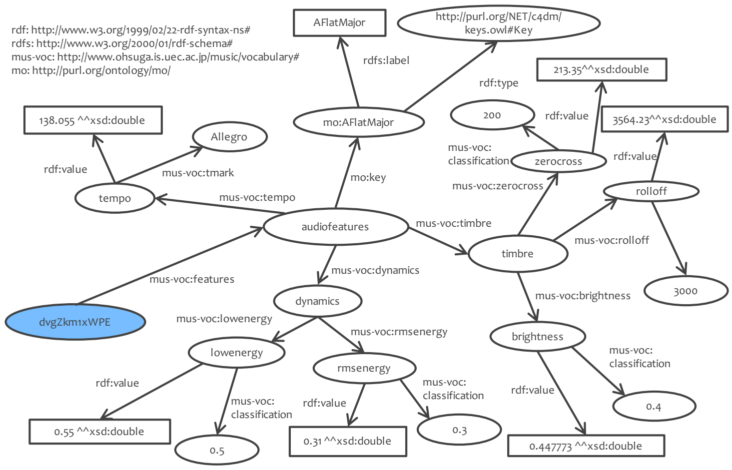

Fig. 1. Part of audio features in the Linked Data of music songs information

3 Music Information Extracting

The system architecture for our music information extraction is shown in Fig. 3, and its workflow is indicated by the number 1 to 11.

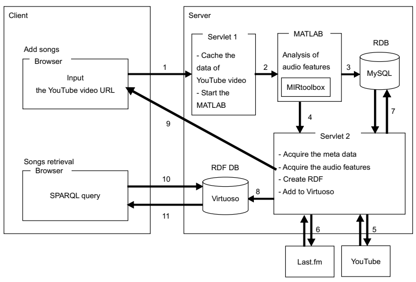

Fig. 2. Structure of the system

1. Download the video data from the YouTube video URI designated by a user in a web browser.
2. Call the MATLAB process that analyzes audio features in the video file.
3. Store the obtained audio features in an RDB, MySQL.
4. Call the RDF create program.
5. Obtain the music information for the video from the YouTube website.
6. Search the music metadata using Last.fm API.
7. Query the audio features of the video for MySQL.
8. Convert the metadata and audio features to RDF graphs, and store them in an RDF store, Virtuoso.
9. Notify the completion to the user.
10. Submit a simple SPARQL query for confirmation.
11. Returns the evidence of the inclusion of new sub-graphs corresponding to the video.

Our system obtains videos to analyze audio features from YouTube, and so public users can easily extend the music information on the platform. However, we discard the video files after extracting the audio features, and thus we believe this process does not cause any legal or moral problems.

The workflow is divided into several phases. The first phase is for analyzing the audio features of the YouTube video, and the second phase is for acquiring the metadata of the YouTube video. Then, the third phase converts the metadata and audio features to RDF graphs, and the RDF graphs are stored in Virtuoso database. Then, the final phase is for the confirmation of newly added graphs.

4 Example of Music Analysis

The current number of music registered in the platform is 1073 and the number of triples automatically extracted for representing the audio features and the metadata for that music is 20858. The platform is publicly available at http://www.ohsuga.is.uec.ac.jp/music/.

In this section, we show the results of some example queries on the platform, and how the music Linked Data can be used for MIR.In the SPARQL Query, we specified the audio feature *Brightness* of the song "Hello, Goodbye" by The Beatles, and search other songs, in which the value of the *Brightness* is similar to the specified song. As the result, we get 5 songs, in which the *Brightness* has the similar degree in Table 2.

```
[SPARQL Query]
PREFIX mus-voc:<http://www.ohsuga.is.uec.ac.jp/music/
vocabulary#>
PREFIX mo:<http://purl.org/ontology/mo/>

SELECT ?artist_x ?title_x ?brightness_x
WHERE { ?metadata rdfs:label ?title .
        ?resource mus-voc:meta ?metadata .
        ?resource mus-voc:features ?features .
        ?features mus-voc:timbre ?timbre .
        ?timbre mus-voc:brightness ?brightnessc .
        ?brightnessc rdf:value ?brightness.
        ?brightnessc_x rdf:value ?brightness_x.
        ?timbre_x mus-voc:brightness ?brightnessc_x .
        ?features_x mus-voc:timbre ?timbre_x .
        ?resource_x mus-voc:features ?features_x .
        ?resource_x mus-voc:meta ?metadata_x .
        ?metadata_x rdfs:label ?title_x .
        ?metadata_x mo:MusicArtist ?MusicArtist_x .
        ?MusicArtist_x rdfs:label ?artist_x .
        FILTER regex(?title, "Hello Goodby") . }
ORDER BY (
  IF( ?brightness < ?brightness_x,
      ?brightness_x - ?brightness,
      ?brightness - ?brightness_x )
) LIMIT 5
```

Table 2. Result of submitting the SPARQL query

artist_x	title_x	brightness_x
The Beatles	Can't Buy Me Love	0.553889
Whitney Houston	Never Give Up	0.559786
Coldplay	Princess Of China Ft. Rihanna	0.560039
Lady Gaga	Judas	0.550279
The Beatles	Penny Lane	0.550221

5 Conclusion and Future Work

In this paper, we proposed a platform for providing audio features and the music metadata to MIR research.

In future, we plan to provide more sophisticated examples and applications of music information analysis, which will encourage the expansion of the music Linked Data to music researchers and developers.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Numbers 16K12411, 16K00419, 16K12533.

References

1. Kitahara, T., Nagano, H.: Advancing Information Sciences through Research on Music:0. Foreword. IPSJ magazine "Joho Shori". 57 (6), 504–505. (2016)
2. Wang, M., Kawamura, T., Sei, Y., Nakagawa, H., Tahara, Y., Ohsuga, A.: Context-aware Music Recommendation with Serendipity Using Semantic Relations. Proceedings of 3rd Joint International Semantic Technology Conference. 17–32. (2013)
3. Osmalskyj, J., Foster, P., Dixon, S., Embrechts, J.J.:Combining Features for Cover Song Identification. Proceedings of the 16th International Society for Music Information Retrieval Conference. 462–468. (2015)
4. Luo, Y-J., Su, L., Yang, Y-H., Chi, T-S.:Real-time Music Tracking using Multiple Performances as a Reference. Proceedings of the 16th International Society for Music Information Retrieval Conference. 357–363. (2015)