

MusicWeb: music discovery with open linked semantic metadata

Alo Allik, Mariano Mora-McGinity, György Fazekas, and Mark Sandler

Queen Mary University of London,

{a.allik, m.mora-mcginity, g.fazekas, mark.sandler}@qmul.ac.uk

Abstract. This demo presents MusicWeb, a novel platform for linking music artists within a web-based application for discovering associations between them. MusicWeb provides a browsing experience using connections that are either extra-musical or tangential to music, such as the artists’ political affiliation or social influence, or intra-musical, such as the artists’ main instrument or most favoured musical key. The platform integrates open linked semantic metadata from various Semantic Web, music recommendation and social media data sources. The connections are further supplemented by thematic analysis of journal articles, blog posts and content-based similarity measures focussing on high level musical categories.

Keywords: Semantic Web, Linked Open Data, music metadata, semantic audio analysis, music information retrieval

1 Introduction

MusicWeb is a music discovery platform which offers users the possibility of exploring editorial, cultural and musical links between artists. It gathers, extracts and manages metadata from many different sources. The connections between artists are based on YAGO categories such as style, geographical location, instrumentation, record label, but also more obscure links, for instance, artists who have received the same award, have shared the same fate, or belonged to the same organisation or religion. The connections are further enhanced by thematic analysis of journal articles and blog posts, content-based music information retrieval similarity metrics and proximity measures in a 2-dimensional mood space. Information about artists is collated and processed from several different web knowledge content resources and presented for the user to navigate in a faceted manner [5]. MusicWeb as a discovery platform is subtly different in function from recommender systems like *dbrec* [7], which suggests music obtained from DBpedia by computing a measure of semantic distance as the number of indirect and distinct links between resources in a graph. This demo intends to show the methods of linking artists employed in the system and how these could help overcome issues such as infrequent access of lesser known artists in large music catalogues (the “long tail” problem) or the difficulty of recommending artists without user ratings in systems that employ collaborative filtering (“cold start” problem) [3].

musicweb beta
OPEN LINKED SEMANTIC METADATA FOR MUSIC

Search for artists:

Ella Fitzgerald

BIO
Ella Jane Fitzgerald (April 25, 1917 – June 15, 1996) was an American jazz singer often referred to as the First Lady of Song, Queen of Jazz and Lady Ella. She was noted for her purity of tone, impeccable diction, phrasing and intonation, and a "horn-like" improvisational ability, particularly in her scat singing. After tumultuous teenage years, Fitzgerald found stability in musical success with ... [read more](#)

AUDIO
Two Little Men in a Flying Saucer
How High The Moon
April In Paris
wait till you see him (de-phazz remix)
Cow Cow Boogie

VIDEO

The Best Of Ella Fitzgerald

The Best Of Ella Fitzgerald

Ella Fitzgerald - Pure Ella (1954) full album

Ella Fitzgerald & Louis Armstrong - Ella & Louis (Full Album)

Fig. 1. Example of a MusicWeb artist page.

2 System overview

The core functionality of the platform relies on available SPARQL endpoints as well as various commercial and community-run APIs. More recently, novel services complement the platform to provide alternative ways to forge connections using natural language processing and machine learning methods. The front portal includes suggested links to selected artists and a search functionality from where users can navigate to individual artists pages. Each artist page contains a biography, a playlist of online audio and a selection of Youtube videos, as shown in Fig. 1. The MusicWeb API uses a number of LOD resources and Semantic Web ontologies to process and aggregate information about artists:

MusicBrainz¹ provides reliable and unambiguous identifiers for entities in music publishing metadata, including artists.

DBPedia² is used for building profiles and querying socio-cultural links between artists.

SameAs.org³ manages URI co-references which is useful for mapping MusicBrainz identifiers to DBpedia entities.

YouTube API is used to query associated video content for the artist panel.

Last.fm⁴ provides recommendations based on crowd-sourced user listening habits.

YAGO [4] semantic knowledge base enables collation of Wikipedia categories

¹ <http://musicbrainz.org>

² <http://dbpedia.org>

³ <http://sameas.org>

⁴ <http://last.fm>

for linking artists.

the Music Ontology [8] provides main concepts and properties for describing musical entities, including artists, on the Semantic Web.

AcousticBrainz⁵ service, which gathers crowd-sourced acoustic information about music, facilitates content-based similarity calculation.

3 Artist similarity

There are many ways in which artists can be considered related. MusicWeb uses Semantic Web technologies and linked data to facilitate faceted searching and displaying of information [6]. This is done by modeling artist similarities in four different domains: socio-cultural, research and journalistic literature, crowd-sourced tag statistics and content-based information retrieval.

Socio-cultural connections between artists in MusicWeb are primarily derived from YAGO categories that are incorporated into entities in DBpedia. Many categories, in particular those that can be considered extra-musical or tangential to music, stem from the particular methodology used to derive YAGO categories from Wikipedia [4].

Literature-based linking is achieved by data-mining research articles and on-line publications using natural language processing. MusicWeb uses Mendeley⁶ and Elsevier⁷ databases for accessing research articles that are curated and categorised by keywords, authors and disciplines. Online newspapers, music magazines and blogs, on the other hand, constitute non-curated data. Relevant information in this case must be extracted from the body of the text by Web-crawling based on keywords or tags. The Alchemy API⁸ is then used for named entity recognition and keyword extraction.

Crowd-sourced tags enable modelling similarity based on projected mood. This method involves using the Semantic Web version of ILM10K music mood dataset that consists of over 4000 unique artists [1]. The dataset is based on crowd-sourced mood tag statistics from Last.fm users, which have been transformed to 2-dimensional coordinates reflecting energy and pleasantness. The similarity between artists is measured by first obtaining the average location of each artist based on their track coordinates. The average locations then enable computing distances between artists and using these as the similarity metric.

Content-based linking involves methodology of Music Information Retrieval (MIR) [2] which facilitate applications that rely on perceptual, statistical, semantic or musical features derived from audio using digital signal processing and machine learning methods. These features may include statistical aggregates computed from time-frequency representations extracted over short time windows. Higher-level musical features include keys, chords, tempo, rhythm, as well as semantic features like genre or mood, with specific algorithms to extract this information from audio. To exploit different types of similarity, we

⁵ <https://acousticbrainz.org/>

⁶ <http://dev.mendeley.com/>

⁷ <http://dev.elsevier.com/>

⁸ <http://www.alchemyapi.com>

model each artist using three main categories of audio descriptors: rhythmic, harmonic and timbral. The features are obtained from the AcousticBrainz Web service which provides descriptors in each category of interest. For each artist in our database, we retrieve features for a large collection of their tracks in the above categories, including beats-per-minute and onset rate (rhythmic), chord histograms (harmonic) and MFCC (timbral) features.

4 Conclusion

MusicWeb is an emerging application to explore the possibilities of linked data-based music discovery. It facilitates users to engage in interesting discovery paths through the space of music artists. The aim is to gather in one application various different approaches to music discovery and how they can benefit from linked music metadata. The next steps are directed toward evaluating its potential acceptance by end users, in particular, exploring which linking methods listeners find most appealing or interesting, and which they would use more often.

MusicWeb is accessible online: <http://musicweb.eecs.qmul.ac.uk/>

References

1. Mathieu Barthet, György Fazekas, Alo Allik, and Mark B. Sandler. Moodplay: an interactive mood-based musical experience. In George Kalliris and Charalampos Dimoulas, editors, *Proceedings of the Audio Mostly 2015 on Interaction With Sound, AM '15, Thessaloniki, Greece, October 7-9, 2015*, pages 3:1–3:8. ACM, 2015.
2. Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-based music information retrieval: Current directions and future challenges. volume 96. IEEE Proceedings, April 2008.
3. Ö. Celma. *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer Verlag, 2010.
4. M.S. Fabian, K. Gjergji, and W. Gerhard. YAGO: A core of semantic knowledge unifying wordnet and wikipedia. In *16th International World Wide Web Conference, WWW*, pages 697–706, 2007.
5. Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(9), 2006.
6. E. Oren, R. Delbru, and S. Decker. Extending faceted navigation for RDF data. In *ISWC*, pages 559–572, 2006.
7. Alexandre Passant. dbrec - music recommendations using DBpedia. In Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, Jeff Z. Pan, Ian Horrocks, and Birte Glimm, editors, *The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part II*, volume 6497 of *Lecture Notes in Computer Science*, pages 209–224. Springer, 2010.
8. Yves Raimond, Samer A Abdallah, Mark B Sandler, and Frederick Giasson. The music ontology. In *ISMIR*, pages 417–422. Citeseer, 2007.