MUSICWEB: SIMILARITY MODELLING STRATEGIES FOR ARTIST DISCOVERY

Alo Allik, Mariano Mora-Mcginity, György Fazekas, Mark Sandler
Queen Mary University of London
a.allik, m.mora-mcginity, g.fazekas, mark.sandler@qmul.ac.uk

ABSTRACT

MusicWeb is a Web application for music artist discovery that 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 connections are further enhanced by thematic analysis of journal articles and blog posts as well as content-based similarity measures focussing on high level musical categories.

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 linked open data sources as well as editorial and content-derived information. The connections between artists are based on several 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. These connections are further enhanced by thematic analysis of journal articles, blog posts and content-based similarity measures focussing on high level musical categories.

2. SYSTEM OVERVIEW

The core functionality of the platform relies on available linked open data sources 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. Further it provides lists of categories linking each artist to other similar artists by various commonalities. The MusicWeb API uses a number of LOD resources and Semantic Web ontologies to process and aggregate information about artists: MusicBrainz [1], DBPedia [2], SameAs [3], YouTube, Last.fm [4], YAGO [2], AcousticBrainz [5], and the Music Ontology [5].

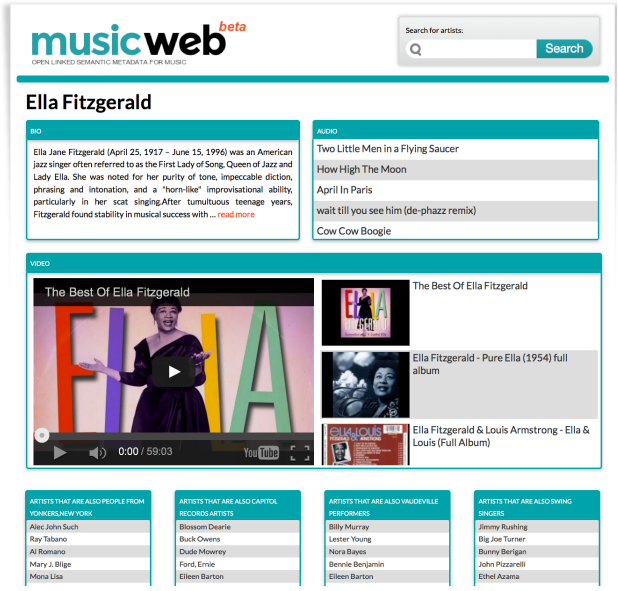

Figure 1. Example of a MusicWeb artist page.

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 [4]. This is done by modeling artist similarities using four different strategies: 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

[1] http://musicbrainz.org
[2] http://dbpedia.org
[3] http://sameas.org
[4] http://last.fm
[5] https://acousticbrainz.org

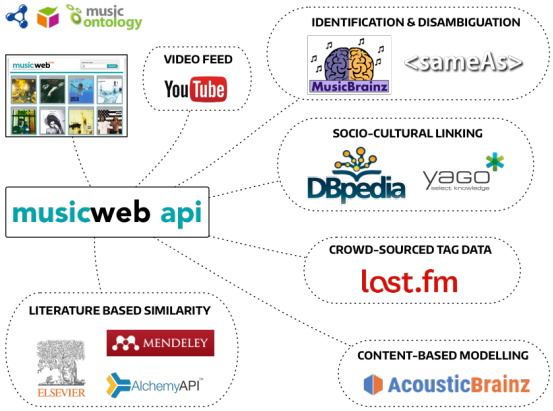

Figure 2. Architecture of the MusicWeb application.

tangential to music, stem from the particular methodology used to derive YAGO categories from Wikipedia [2].

Literature-based linking is achieved by data-mining research articles and online publications using natural language processing. MusicWeb uses Mendeley [6] and Elsevier [7] 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 API [8] 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. 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 models similarity using three main categories of audio descriptors: rhythmic, harmonic and timbral. The artists in our dataset are selected based on the number of tracks that have features available in the AcousticBrainz Web service. For each such artist, we retrieve features for their tracks, including beats-per-minute and onset rate (rhythmic), chord histograms (harmonic) and MFCC (timbral). Then we fit a Gaussian Mixture Model (GMM) with full covariances on each set of aggregated features in each category across the tracks and compute symmetrised Kullback-Leibler divergences for the selected category. The divergences for each artist are then ranked and identifiers of the top N closest artists are stored.

[6] `http://dev.mendeley.com`
[7] `http://dev.elsevier.com`
[8] `http://www.alchemyapi.com`

4. CONCLUSION

MusicWeb is an emerging application that 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 keys. It does this by pulling data from several different Web resources and presenting them for the user to navigate in a faceted manner [3]. It facilitates users to engage in interesting discovery paths through the space of music artists and strives to 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) [1]. MusicWeb is accessible online: `http://musicweb.eecs.qmul.ac.uk/`

5. REFERENCES

[1] Ò. Celma. *Music Recommendation and Discovery:The Long Tail, Long Fail, and Long Play in the Digital Music Space.* Springer Verlag, 2010.

[2] MS 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.

[3] Gary Marchionini. Exploratory search: from finding to understanding. *COMMUNICATIONS OF THE ACM*, 49(9), 2006.

[4] E. Oren, R. Delbru, and S. Decker. Extending faceted navigation for RDF data. In *ISWC*, pages 559–572, 2006.

[5] Yves Raimond, Samer A Abdallah, Mark B Sandler, and Frederick Giasson. The music ontology. In *ISMIR*, pages 417–422. Citeseer, 2007.