

# USING LINKED OPEN DATA FOR NOVEL ARTIST RECOMMENDATIONS

Stephan Baumann and Rafael Schirru

German Research Center for Artificial Intelligence

Alt-Moabit 91c

10559 Berlin, Germany

firstname.lastname@dfki.de

## ABSTRACT

*As part of the Linking Open Data initiative many (community) platforms have made their data freely available in recent years. In these systems valuable information from the music domain can be found. We use this data for content-based music recommendations thus not requiring expensive metadata annotations from domain experts. In this demo paper we present a system that uses data from the Freebase platform to describe music artists. The artists are represented using metadata profiles that characterize different aspects of their works such as genres and instrumentations but also collaborations with other artists. The system uses a Lucene index for content-based artist recommendations aiming at Web-scalability.*

## 1. INTRODUCTION

Current music recommender systems often rely on collaborative filtering, i.e., they recommend music items that peers of the active user enjoyed in the past. Other systems apply content-based approaches using for instance audio features or expert annotations to describe music items. Recent work has shown that collaborative filtering has problems providing access to the long tail of seldom listened music [2]. Audio features on the other hand cannot describe content on a semantic level. Expert annotations allow for high quality content-based recommendations but the annotation process is expensive. As a consequence we propose HORST, a content-based recommender system for music artists using freely available semantic data. The system uses Linked Open Data<sup>1</sup> from the Freebase<sup>2</sup> platform to describe music artists thus not requiring expensive metadata annotations from domain experts. Each artist is represented using a metadata profile that characterizes different aspects of his/her work such as genres and instrumentations but also collaborations with other artists. The system uses a

<sup>1</sup><http://linkeddata.org/>

<sup>2</sup><http://www.freebase.com/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2012 International Society for Music Information Retrieval.

Lucene<sup>3</sup> index aiming at Web-scalability. In our previous work we could show that a content-based recommendation approach based on Linked Open Data can generate many high quality novel artist recommendations [1].

## 2. RELATED WORK

Passant introduces distance measures on Linked Data to determine the relatedness between resources [3]. These measures can be used to generate self-explanatory resource recommendations. For music recommendations Passant uses a distance measure that takes into account direct links and indirect links between resources (i.e., incoming links from two different resources or outgoing links to two different resources). Link types are weighted so that frequent link types obtain a lower weight. Passant determines the similarity between artists by applying a link-based distance measure on the Linked Open Data graph. In contrast our approach selects metadata that we consider relevant to describe the music styles of artists and uses Lucene to perform similarity calculations in the vector space.

In [1] we introduced an approach to identify similar artists based on Semantic Web metadata. The evaluation experiments have shown that the method can generate high quality, novel artist recommendations. In this contribution we propose a software architecture that is designed to meet industry requirements thus making such an approach applicable in real world scenarios.

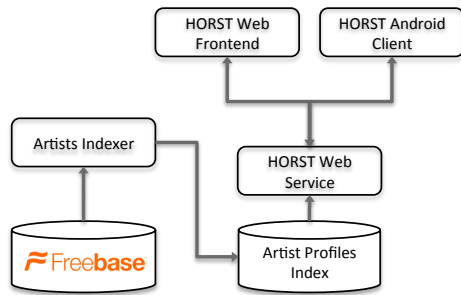
## 3. SYSTEM ARCHITECTURE

Our system follows a client/server architecture (Figure 1). Its components will be described subsequently.

### 3.1 Data Source

Freebase is a collaborative knowledge base mainly consisting of metadata provided by the community of users. It further contains data that has been harvested from many different sources. The data is provided in structured formats making it easily accessible for people and machines. Applications can access the data via a dedicated API, an RDF endpoint, as well as data base dumps. The available metadata to describe artists comprises among others: their

<sup>3</sup><http://lucene.apache.org/>



**Figure 1.** Components of the HORST system.

origin, instruments played together with the associated instrument families, genres, collaborations with other artists, and record release years.

A very important aspect of the Freebase data set is the fine granularity of the genre annotations. In the data set we used there were 2029 distinct genres organized in a hierarchical structure. It is reasonable to assume that artists sharing at least some of the very specific genre annotations are in some respects similar.

### 3.2 Artists Indexer

The indexing component aggregates for every artist in the system a metadata profile consisting of the following data: i) A fine-grained list of genres the artist is associated with. ii) The instruments played by the artist together with the associated instrument families. iii) Other associated artists as found in the Freebase data set. For efficient retrieval the artist profiles are stored in a Lucene index. Each metadata type (genres, instruments, associated artists) is stored in its own index field.

### 3.3 HORST Web Service

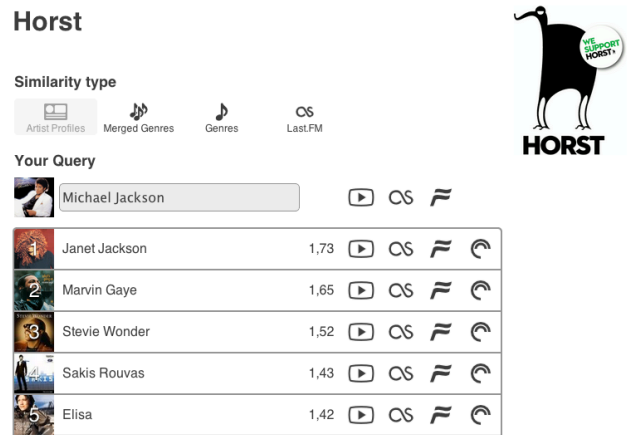
The HORST Web Service provides a RESTful<sup>4</sup> interface and uses JSON<sup>5</sup> as data exchange format. Whenever clients request artist recommendations they provide the name of the query artist, a starting position for paging and the desired number of similar artists. The Web Service then performs a lookup for the metadata profile of the query artist in the Lucene index. This profile is used as query for finding similar artists. A query is composed using the content of the seed artist profile. The metadata fields are preserved, i.e., genres are matched against genres, instruments against instruments, etc. We keep the ranking of the Lucene index and return the results in the specified range.

### 3.4 Client Applications

The client/server architecture of HORST allows an integration of our recommendations into various systems and applications. Currently, there are two client applications that allow access to the HORST artist recommendations. First, we implemented a Web interface that can be accessed via

<sup>4</sup>[http://www.ics.uci.edu/~fielding/pubs/dissertation/fielding\\_dissertation.pdf](http://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf)

<sup>5</sup><http://www.json.org/>



**Figure 2.** HORST Web Interface

standard browsers. A screenshot of the Web interface is depicted in Figure 2. Second, an Android application has been developed for mobile usage.

## 4. SUMMARY AND OUTLOOK

In this paper we described a recommender system for music artists exploiting freely available metadata from the Semantic Web. For content-based recommendations the system uses a Lucene index aiming at Web-scalability.

In our future work we intend to provide music recommendations taking the link structure of the giant graph of Linked Data into account. In this approach two artists are considered similar when sharing common experiences such as common roots, friendships, or performances at the same concert/festival. Link-based explanations constitute the main focus in this research by telling the user a story about how two artists are related. We plan to integrate these approaches based on Linked Data with social data for instance from Facebook. User data from social networks can be used for enhanced user profile learning. Further, we plan to add collaborative filtering in order to ensure the quality of the recommendations provided.

## 5. REFERENCES

- [1] Stephan Baumann, Rafael Schirru, and Bernhard Streit. Towards a storytelling approach for novel artist recommendations. In Marcin Detyniecki, Peter Knees, Andreas Nürnberger, Markus Schedl, and Sebastian Stober, editors, *Adaptive Multimedia Retrieval*, volume 6817 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2010.
- [2] Óscar Celma. *Music Recommendation and Discovery in the Long Tail*. PhD thesis, University Pompeu Fabra, Barcelona, Spain, 2008.
- [3] Alexandre Passant. Measuring semantic distance on linking data and using it for resources recommendations. In *Proceedings of the AAAI Spring Symposium "Linked Data Meets Artificial Intelligence"*, pages 93–98, 2010.