FOAFING THE MUSIC: A MUSIC RECOMMENDATION SYSTEM BASED ON RSS FEEDS AND USER PREFERENCES

Òscar Celma
Music Technology Group
Universitat Pompeu Fabra
Pg. Circumval·lació 8, 08003
Barcelona, SPAIN
ocelma@iua.upf.edu

Miquel Ramírez
Music Technology Group
Universitat Pompeu Fabra
Pg. Circumval·lació 8, 08003
Barcelona, SPAIN
mramirez@iua.upf.edu

Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
Pg. Circumval·lació 8, 08003
Barcelona, SPAIN
pherrera@iua.upf.edu

ABSTRACT

In this paper we give an overview of the *Foafing the Music* system. The system uses the *Friend of a Friend* (FOAF) and *Rich Site Summary* (RSS) vocabularies for recommending music to a user, depending on her musical tastes. Music information (new album releases, related artists' news and available audio) is gathered from thousands of RSS feeds —an XML format for syndicating Web content. On the other hand, FOAF documents are used to define user preferences.

The presented system provides music discovery by means of: user profiling —defined in the user's FOAF description—, context-based information —extracted from music related RSS feeds— and content-based descriptions —extracted from the audio itself.

1 INTRODUCTION

The World Wide Web has become the host and distribution channel of a broad variety of digital multimedia assets. Although the Internet infrastructure allows simple, straightforward acquisition, the value of these resources lacks of powerful content management, retrieval and visualization tools. Music content is no exception: although there is a sizeable amount of text-based information about music (album reviews, artist biographies, etc.) this information is hardly associated to the objects they refer to, that is music music files (MIDI and/or audio). Music is an important vehicle for communicating other people something relevant about our personality, history, etc.

In the context of the Semantic Web, there is a clear interest to create a Web of machine-readable homepages describing people, the links among them, and the things they create and do. The FOAF (*Friend Of A Friend*) project[1] provides conventions and a language "to tell" a machine the sort of things that a user says about herself in her homepage. FOAF is based on the RDF/XML[2] vocabulary. We can foresee that with the user's FOAF profile, a system would get a better representation of the user's musical needs. On the other hand, the RSS vocabulary[3] allows to syndicate Web content on Internet. Syndicated content includes data such as news feeds, events listings, news stories, headlines, project updates, etc.

Foafing-the-music initiative is covered by the SIMAC IST project[4]. The main goal of SIMAC project is doing research on semantic descriptors of music contents, in order to use them, by means of a set of prototypes, for providing song collection exploration, retrieval and recommendation services. These services are meant for "home" users, music content producers and distributors, and academic users. One special feature is that these descriptions are composed by *semantic descriptors*. Music will be tagged using a language close to the user's own way of describing its contents —moving the focus from low-level to higher-level (i.e. semantic) descriptions.

2 BACKGROUND

Recommender Systems are software applications whose purpose is to deliver information to people that "needs" it. In this way, one cannot tell the difference between a recommender system and a search engine —both software types share the same purpose: to select objects (or items) from a repository whose features were found to satisfy the querying user's needs.

However, there are two subtle but meaningful differences between "search engines" and "recommender systems". The first one lies in the design intention, or better said: the wording of the problem to address when designing the system. Is that "information need" related to solving a contingent situation, or is that need something periodic or static? The second difference is also another design intention difference, which lies in the use of two different words to describe the system: does it *retrieve* information from a relatively static repository of information? Or does it *filter* objects embedded in an incoming stream of information?

The term "recommender system" emerged as a logical evolution of the research on information retrieval (IR)

[1] http://www.foaf-project.org

[2] http://www.w3.org/RDF

[3] http://web.resource.org/rss/1.0/

[4] http://www.semanticaudio.org

systems. This evolutive main feature was the emphasis put on the "query" concept definition and representation. Recommender systems were initially thought as *information filtering* systems, whose technological framework baseline stemmed from *information retrieval* systems (Belkin and Croft, 1997). This, then, effectively implies that a recommender system is an inherently dual purpose application: the user profiling of static information needs might be used to better understanding and attending immediate, unforeseen needs.

There are two main approaches to recommend items to users: collaborative filtering and content-based filtering. Next section explains the differences between both approximations.

2.1 Collaborative filtering versus Content-based filtering

Collaborative filtering consists of making use of feedback from users to improve the quality of material presented to users. Obtaining feedback can be explicit or implicit. Explicit feedback comes in the form of user ratings or annotations, whereas implicit feedback can be extracted from user's habits. One of the main caveats of this approach is the fact that the only way to recommend brand new items is that some user has to previously rate or review that item. There are some examples that succeed based on this approach. For instance, Amazon is a good illustration system (Linden et al., 2003).

On the other hand, content-based filtering tries to extract useful information —from the items of the user's collection— that could be useful to represent user's needs. This approach solves the limitation of collaborative filtering as it can recommend new items (even before the system does not know anything from that item), by comparing the actual set of user's items and calculating a distance with some sort of similariy measure. In the music field, extracting musical semantics from the raw audio and computing similarities between music pieces is a challenging problem. Traditional music similarity measures use low-level —mainly timbre-based— features. We belive that adding cultural metadata terms to such a similarity measure can help to get better results.

Pachet (2005) proposes a classification of musical metadata, and how this classification affects music content management, as well as the problems to face when elaborating a ground truth reference for music similarity (both in collaborative and content-based filtering).

2.2 Friend of a friend initiative

The Friend of a friend (FOAF) initiative provides a framework for representing information about people, their interests, relationships between them and social connections. The FOAF vocabulary contains terms for describing personal information —name, nick, mailbox, interest, images, etc.—, membership in groups —member, group, organization, etc. FOAF is based on the RDF/XML vocabulary. Listing 1 shows a possible input file[5]. A FOAF description, then, describes —in a machine readable format— a person. Currently, there are more then 6 million of FOAF documents describing people on the web (Golbeck, 2005).

```
<rdf:RDF xml:lang="en">
  <foaf:Person>
    <foaf:nick>a_user</foaf:nick>
    <foaf:dateOfBirth>04-17</
        foaf:dateOfBirth>
    <foaf:mbox_sha1sum>ce24ca...a1f0</
        foaf:mbox_sha1sum>
    <foaf:page>
      <foaf:Document
       rdf:about="http://www.livejournal.
           com/userinfo.bml?user=a_user">
        <dc:title>LJ Profile</dc:title>
      </foaf:Document>
    </foaf:page>
    <foaf:weblog rdf:resource="http://www.
       livejournal.com/users/a_user/" />
    <foaf:interest dc:title="gretsch"
      rdf:resource="http://www.livejournal
          .com/interests.bml?int=gretsch"
          />
    <foaf:interest dc:title="dogs d'amour"
      rdf:resource="http://www.livejournal
          .com/interests.bml?int=dogs+d%27
          amour" />
    <foaf:interest dc:title="social
        distortion"
      rdf:resource="http://www.livejournal
          .com/interests.bml?int=social+
          distortion" />
    <foaf:interest dc:title="beer"
      rdf:resource="http://www.livejournal
          .com/interests.bml?int=beer" />
    <foaf:interest dc:title="the misfits"
      rdf:resource="http://www.livejournal
          .com/interests.bml?int=the+
          misfits" />
    <foaf:interest dc:title="the pogues"
      rdf:resource="http://www.livejournal
          .com/interests.bml?int=the+
          pogues" />
  </foaf:Person>
</rdf:RDF>
```

Listing 1: Example of a user's FOAF profile

To our knowledge, nowadays it does not exist any system that recommends items to a user, based on her FOAF profile (Celma et al., 2004). The *FilmTrust* system[6] is a part of a research study aimed to understanding how social preferences might help web sites to present information in a more useful way. The system collects user reviews and ratings about movies, and holds them into the user's FOAF profile. Although it has not yet implemented a recommendation system, it includes a rating algorithm for films based on a trust-based algorithm (Golbeck and Parsia, 2005).

2.3 Music recommendation systems

The main goal of a music recommendation system is to propose, to the end-user, interesting and unknown music artists (and their available tracks —if possible—), based on her musical taste. But musical taste and music preferences are affected by several factors, even demographic and personality traits. Then, the combination of music

[5] A real example extracted from *http://www.livejournal.com*, only changing the user's name

[6] http://trust.mindswap.org/FilmTrust

preferences and personal aspects —such as: age, gender, origin, occupation, musical education, etc.— could improve music recommendations (Uitdenbogerd and van Schyndel, 2002).

Moreover, a music recommendation system should be able to get new music dynamically, as it should recommend new items to the user once in a while. In this sense, there is a lot of free available (in terms of licensing) music on Internet, performed by "unknown" artists that can suit perfectly for new recommendations. Nowadays, music websites are noticing the user about new releases or artist's related news, mostly in the form of RSS feeds. iTunes Music Store[7] provides an RSS (version 2.0) feed generator[8], updated once a week, that publishes new releases of artists' albums. A music recommendation system should take advantage of these publishing services, as well as integrating them into the system, in order to filter and recommend new music to the user.

Most of the current music recommenders are based on collaborative filtering approach, or on a hybrid version including clustering and users' communities Examples of such systems are: Audioscrobbler[9], iRate[10], Goombah Emergent Music[11], MusicStrands[12], and inDiscover[13].

The basic idea of a music recommender system based on collaborative filtering is:

1. To keep track of which artists a user listens to (through WinAmp, XMMS, etc. plugins),
2. To find other users with similar tastes, and
3. To recommend similar artists to the user, according to these similar listeners' taste.

But, digital music collections can be huge (thousands of files) and very heterogeneous. Thus, this approach to recommend music can generate some "silly" (or obvious) answers.

Contrastingly, the main goal of the *Foafing the Music* system is to recommend, to discover and to explore music content; based on user profiling —via FOAF descriptions—, context-based information —extracted from music related RSS feeds—, and content-based descriptions —extracted from the audio itself.

3 SYSTEM OVERVIEW

Music recommendations, in the *Foafing the Music* system, are generated through the following steps:

1. Get interests from user's FOAF profile,
2. Detect artists and bands,
3. Select related artists, from artists encountered in the user's FOAF profile, and
4. Rate results by relevance.

The system reads an input FOAF profile —that is, an RDF/XML file—, and extracts user's interests. Then, it queries a music repository in order to detect whether the interest is a music artist (or a band), and selects similar artists to the ones found. To get artists' similarities, a focused web crawled has been implemented to look for relationships between artists (such as: related to, influenced by, followers of, etc.). Moreover, a music similarity distance is used to recommend tracks that are similar to tracks composed or played by artists found in the FOAF profile.

Based on the previous FOAF example (see listing 1), the system detects the following artists from the user's profile: *Dogs d'Amour*, *Social Distortion*, *The Misfits* and *The Pogues*. Starting from these artists, the system searches for similar artists and for artists influenced by them. Then, it scores them in terms of counting artist occurrences. If there are any tracks in the music repository from artists declared in the FOAF profile, it computes the similarity and gets the most significant similar tracks from other artists. Figure 1 shows the output recommended artists.

Once the related artists have been selected, *Foafing the Music* filters music related information coming form RSS feeds to:

- Get new music releases,
- Download (or stream) audio from MP3-blogs and Podcast sessions, and
- Create, automatically, playlists based on audio similarity.

Another component of the system is the (music related) newsfeeds filtering. Next section explains the main characteristics of this component.

3.1 Music related news filtering

The music related news filtering component queries a newsfeeds system that filters news related artists found in user's FOAF profile. To do so, this component permits to communicate with the PubSub server[14], via the Jabber protocol, and creates an RSS feed with a given query —the user musical preferences found in the FOAF file. PubSub is a matching service that instantly notifies a user whenever new content matching user's subscription is created. PubSub reads over 13 million weblogs and more than 50,000 internet newsgroups. Jabber[15] is an open secure protocol, an ad-free alternative to consumer instant messaging services like ICQ, MSN, and Yahoo. Jabber makes use of XML protocols that enable any two entities on the Internet to exchange messages, presence, and other structured information in nearly real time.

Once the subscription has been created, it is possible to visualize all the music related news for a given user. Each news item has a bar score that shows how much it is

[7] http://www.apple.com/itunes
[8] http://phobos.apple.com/WebObjects/MZSearch.woa/wo/0.1
[9] http://www.audioscrobbler.com
[10] http://irate.sourceforge.net
[11] http://goombah.emergentmusic.com/
[12] http://www.musicstrands.com
[13] http://www.indiscover.net/
[14] http://www.pubsub.com
[15] http://www.jabber.org

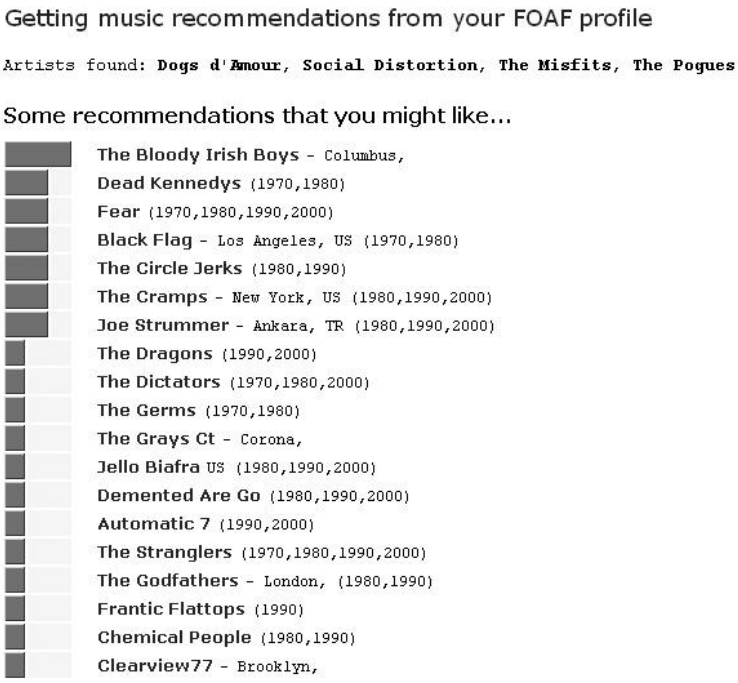

Figure 1: Recommended artists from artists detected in a user's FOAF profile.

related with user's musical interests. Scoring is done using the TF/IDF ranking algorithm (Baeza-Yates and Ribeiro-Neto, 1999). TF/IDF ranks documents by counting the number of ocurrences of user's term query into each document.

Foafing the Music system is available at: http://www.semanticaudio.org/foafin-the-music

4 CONCLUSIONS

We have proposed a system that filters music related information from RSS based on a given user's profile. A system based on FOAF profiles allows to "understand" a user in two complementary ways; psychological factors —personality, demographic preferences, socio-economics, situation— and explicit musical preferences. This system, then, is able to filter and to contextualize users' queries.

In the music field context, we expect that using news filtering about new music releases, artists' interviews, album reviews, etc. can improve a recommendation system in a dynamic way. Finally, this approach opens a wide range of possible usages and applications, such as notifying a user the forthcoming gigs by an artist —playing close to user's location— whose music is similar to user's musical taste.

Finally, the evaluation of the system is planned to be done by real users, by comparing the music recommendation results of our system (in the artist level) with a music recommender based on collaborative system, such as Audioscrobbler.

ACKNOWLEDGEMENTS

This work is partially funded by the SIMAC IST-FP6-507142 European project.

REFERENCES

R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, first edition, 1999.

N. J. Belkin and W. B. Croft. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12):29–39, December 1997.

O. Celma, M. Ramírez, and P. Herrera. Semantic interaction with music content using foaf. In *Proceedings of 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web*, Galway, Ireland, 2004.

J. Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, 2005.

J. Golbeck and B. Parsia. Trust network-based filtering of aggregated claims. In *International Journal of Metadata, Semantics, and Ontologies (to appear)*, 2005.

G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 4(1), 2003.

F. Pachet. *Knowledge Management and Musical Metadata*. Idea Group, 2005.

A. Uitdenbogerd and R. van Schyndel. A review of factors affecting music recommender success. In *ISMIR 3rd International Conference on Music Information Retrieval, October 13-17.*, 2002.