

# Distributional models vs. Linked Data: exploiting crowdsourcing to personalize music playlists

Cataldo Musto<sup>1</sup>, Fedelucio Narducci<sup>2</sup>, Giovanni Semeraro<sup>1</sup>,  
Pasquale Lops<sup>1</sup>, and Marco de Gemmis<sup>1</sup>

<sup>1</sup> Department of Computer Science  
University of Bari Aldo Moro, Italy  
`name.surname@uniba.it`

<sup>2</sup> Department of Information Science, Systems Theory, and Communication  
University of Milano-Bicocca, Italy  
`narducci@disco.unimib.it`

**Abstract.** This paper presents Play.me, a system that exploits social media to generate personalized music playlists. First, we extracted user preferences in music by mining Facebook profiles. Next, given this preliminary playlist based on explicit preferences, we enriched it by adding new artists related to those the user already likes. In this work two different enrichment techniques are compared: the first one relies on knowledge stored on DBpedia while the latter is based on the similarity calculations between semantic descriptions of the artists. A prototype version of the tool was made available online in order to carry out a preliminary user study to evaluate the best enrichment strategy. This paper summarizes the results presented in EC-Web 2012 [3].

## 1 Introduction and Related Work

According to a recent study<sup>3</sup>, 31,000 hours of music (and 28 million songs) are currently available on iTunes Store. As a consequence, the problem of information overload is currently felt for online music libraries and multimedia content, as well. However, the recent spread of social networks provides researchers with a rich source to draw to overcome the typical bottleneck represented by user preferences elicitation.

Given this insight, in this work we propose Play.me, a system that leverages social media for personalizing music playlists. The filtering model is based on the assumption that information about music preferences can be gathered from Facebook profiles. Next, explicit Facebook preferences may be enriched with new artists related to those the user already likes. In this paper we compare two different enrichment techniques: the first leverages the knowledge stored on DBpedia while the second is based on similarity calculations between semantics descriptions of artists. The final playlist is then ranked and finally presented

---

<sup>3</sup> <http://www.digitalmusicnews.com/permalink/2012/120425itunes>

to the user that can express her feedback. A prototype version of Play.me was made available online and a preliminary user study to detect the best enrichment technique was performed. Generally speaking, this work can be placed in the area of music recommendation (MR), a topic that has been widely covered in literature: an early attempt of handling MR problem is due to Shardanand [5], who proposed collaborative filtering to provide music recommendations. Similarly to our work, in [1] Lamere analyzed the use of tags as source for music recommendation, while the use of Linked Data is investigated in [4].

## 2 Play.me: personalized playlists generator

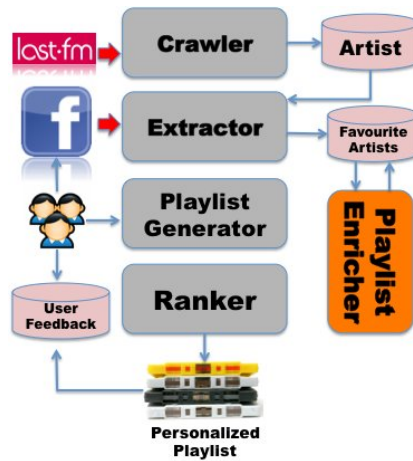
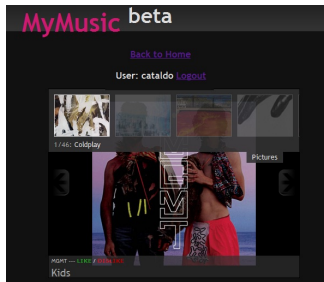


Fig. 1. Play.me architecture

The general architecture of Play.me is depicted in Figure 1. The generation is directly triggered by the user, who invokes the PLAYLIST GENERATOR module. The set of her favourite artists is built by mapping her preferences gathered from her own Facebook profile (specifically, by mining the links she posted as well as the pages she likes) with a set of artists extracted from Last.fm. Given this preliminary set, the PLAYLIST ENRICHER adds new artists by using different enrichment strategies. Finally, for each artist in that set, the most popular tracks are extracted and the final playlist is shown to the target user, who can express her feedback. A working implementation of Play.me has been made available online (Figure 2). For a complete description of the system it is possible to refer to [3], while in this paper we just focus on the enrichment algorithms.

**Enrichment based on Linked Data.** The first technique for enriching user preferences extracted from Facebook relies on the exploitation of DBpe-



**Fig. 2.** Play.me screenshot

dia<sup>4</sup>. Our approach is based on the assumption that each artist can be mapped to a DBpedia node. The incentive idea is that the similarity between two artists can be computed according to the number of properties they share (e.g. two Italian bands playing rock music are probably similar). Thus, we decided to use DBPEDIA-OWL:GENRE (describing the genre played by the artist) and DC-TERMS:SUBJECT, that provides information about the musical category. Operationally, we queried a SPARQL endpoint to extract the artists that share as many properties as possible with the target one. Finally, we ranked them according to their playcount in Last.fm. The first  $m$  artists returned by the endpoint are considered as related and added to the set of the favourite artists.

**Enrichment based on Distributional Models.** Each artist in Play.me is described through a set of tags (extracted from Last.fm), where each tag provides information about the genre played by the artist or describes features typical of her songs (e.g. *melanchonic*). According to the insight behind distributional models [2], each artist can be modeled as a point in a semantic vector space, and the position depends on the tags used to describe her and the co-occurrences between the tags themselves. The rationale behind this strategy is that the relatedness between two artists can be calculated by comparing their vector-space representation through the classical cosine similarity. So, we compute the cosine similarity between the target artist and all the other ones in the dataset, and the  $m$  with the highest scores are added to the list of favourite ones.

### 3 Experimental Evaluation

In the experimental evaluation we tried to identify the technique able to generate the most relevant playlists. We carried out an experiment by involving 30 users against a Last.fm crawl containing data on 228k artists. In order to identify the best enrichment technique, we asked users to use the application for three weeks. In the first two weeks the system was set with a different enrichment technique, while in the last a simple baseline based on the most popular artists was used. Given the playlist generated by the system, users were asked to express their

<sup>4</sup> <http://dbpedia.org>

feedback only on the tracks generated by the enrichment process. Results are reported in Table 1. The parameter  $m$  refers to the number of artists added by the enrichment algorithm for each one extracted from Facebook. It is worth to notice

**Table 1.** Results: each score represent the ratio of positive feedbacks.

Strategy	Artists		
	m=1	m=2	m=3
Linked Data	65.9%	64.6%	63.2%
<b>Distributional Models</b>	<b>76.3%</b>	<b>75.2%</b>	<b>69.7%</b>
Popularity	58%		

that both enrichment strategies outperform the baseline. This means that the social network data actually reflect user preferences. The enrichment technique that gained the best performance is that based on *distributional models*. However, even though this technique gained the best results, a deeper analysis can provide different outcomes. Indeed, with  $m=3$  the gap between the approaches drops down: this means that a pure content-based representation introduces more noise than DBpedia, whose effectiveness stays constant. The good results obtained by the baseline can be justified by the low diversity of the users involved in the evaluation. More details about the experimental settings are reported in [3].

## 4 Conclusions and Future Work

In this work we presented Play.me, a system for building music playlists based on social media. Specifically, we compared two techniques for enriching the playlists, the first based on DBpedia and the second based on similarity calculations in vector spaces. From the experimental session it emerged that the approach based on distributional models was able to produce the best playlists. Generally speaking, there is still space for future work since the enrichment might be tuned by analyzing different DBpedia properties or different tags. Furthermore, context-aware personalized playlists could be a promising research direction.

## References

1. P. Lamere. Social Tagging and Music Information Retrieval. *Journal of New Music Research*, 37(2):101–114, 2008.
2. A. Lenci. Distributional approaches in linguistic and cognitive research. *Italian Journal of Linguistics*, (1):1–31, 2010.
3. C. Musto, G. Semeraro, P. Lops, M. de Gemmis, and F. Narducci. Leveraging social media sources to generate personalized music playlists. In *EC-Web 2012*.
4. A. Passant and Y. Raimond. Combining Social Music and Semantic Web for Music-Related Recommender Systems. In *Social Data on the Web, ISWC Workshop*, 2008.
5. U. Shardanand. Social information filtering for music recommendation. Bachelor thesis, Massachusetts Institute of Technology, Massachusetts, 1994.