

USING ARTIST SIMILARITY TO PROPAGATE SEMANTIC INFORMATION

Joon Hee Kim, Brian Tomasik, Douglas Turnbull

Department of Computer Science,

Swarthmore College

{joonhee.kim@alum, btomasik@alum, turnbull@cs}.swarthmore.edu

ABSTRACT

Tags are useful text-based labels that encode semantic information about music (instrumentation, genres, emotions, geographic origins). While there are a number of ways to collect and generate tags, there is generally a *data sparsity* problem in which very few songs and artists have been accurately annotated with a sufficiently large set of relevant tags. We explore the idea of *tag propagation* to help alleviate the data sparsity problem. Tag propagation, originally proposed by Sordo et al., involves annotating a novel artist with tags that have been frequently associated with other similar artists. In this paper, we explore four approaches for computing artists similarity based on different sources of music information (user preference data, social tags, web documents, and audio content). We compare these approaches in terms of their ability to accurately propagate three different types of tags (genres, acoustic descriptors, social tags). We find that the approach based on collaborative filtering performs best. This is somewhat surprising considering that it is the only approach that is not explicitly based on notions of semantic similarity. We also find that tag propagation based on content-based music analysis results in relatively poor performance.

1. INTRODUCTION

Tags, such as “hair metal”, “afro-cuban influences”, and “grl power”, are semantic labels that are useful for semantic music information retrieval (IR). That is, once we annotate (i.e., index) each artist (or song) in our music database with a sufficiently large set of tags, we can then retrieve (i.e., rank-order) the artists based on relevance to a text-based query.

The main problem with tag-based music IR is *data sparsity* (sometimes referred to as the *cold start* problem [1]). That is, in an ideal world, we would know the relevance (or lack thereof) between every artist and every tag. However, given that there are millions of songs and potentially thousands of useful tags, this is an enormous anno-

tation problem. For example, Lamere [2] points out that Last.fm, a popular music-oriented social network, has a database containing over 150 millions songs each of which have been tagged with an average of 0.26 tags. This problem is made worse by *popularity bias* in which popular songs and artists tend to be annotated with a heavily disproportionate number of tags. This is illustrated by the fact that Lamere found only 7.5% of artists in his corpus of 280,000 artists had been annotated with one or more tags.

One potential solution to the *data sparsity* problem is to *propagate* tags between artists based on artist similarity. To annotate tags for an artist a , we find the most similar artists to a (referred to as neighbors) and transfer the most frequently occurring tags among the neighbors to artist a . Note that while we focus on artist annotation in this paper, our approach is general in that it could also be used to propagate tags between songs as well as other non-music related items such as movies and books.

Tag propagation has two potential uses. First, it allows us to index an unannotated artist if we can calculate the similarity between the artist and other annotated artists. Second, tag propagation allows us to augment and/or improve an existing annotation for an artist.

This idea was originally proposed by Sordo et al. who explore tag propagation of social tags based on acoustic similarity [3]. This content-based approach is compelling because we can automatically calculate artist similarity without relying on human input. However, as we will show in Section 5, the content-based tag propagation performs poorly relative to other music information sources.

In this paper, we extend their initial exploration by comparing alternative approaches to compute similarity: collaborative filtering of user preference data, similarity based on social tags, text-mining of web documents, and content-based analysis of music signals. In addition, we experiment with tag propagation on three different types of tags: acoustic descriptors, genres, and social tags.

While our focus is on the use of tag propagation for text-based music IR, we can also view our system as a way to evaluate artist similarity metric. That is, the approach that results in the best transfer of semantic information between artists may be considered a good approach for accessing artist similarity. Since artist similarity is often used for music recommendation, evaluating tag propagation performance is an automatic alternative to using labor-intensive human surveys when determining the quality of a music

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.

© 2009 International Society for Music Information Retrieval.

recommendation system.

1.1 Related Work

The importance of annotating music with tags is underscored by large investments that have been made by various companies in recent years. Companies like Pandora and AMG Allmusic employ dozens of professional music editors to manually annotate music with a small and structured vocabulary of tags. While this approach tends to produce accurate and complete characterizations of some songs, this labor-intensive approach does not scale with the rapidly increasing amount of available music online. For example, 50 Pandora experts annotate about 15,000 songs per month and would take over 83 years to annotate the 15 million songs that are currently in the AMG Allmusic database¹

Last.fm and MyStrands use an alternative “crowdsourcing” approach in which millions of registered users are encouraged to label songs with any open-end free-text tags. As of September 2008, Last.fm had collected over 25 million song-tag annotations and 20 million artist-tag annotations using a vocabulary of 1.2 million unique tags (although only about 11% had been used more than 10 times) [4]. Each month, about 300 thousand unique users contribute more than 2.5 million new song-tag or artist-tag annotations. However, as mention above, a relatively small percentage of artists and songs have ever been tagged and even fewer have been thoroughly annotated.

Academic research has also focused on the music annotation problem in recent years. Turnbull et al. suggest that there are five general distinct approaches to annotating music with tags: conducting a survey (e.g., Pandora), harvesting social tags (e.g., Last.fm), playing annotation games [5, 6], text-mining web documents [7, 8], and analyzing audio content with signal processing and machine learning [9–11]. In some sense, tag propagation represents a sixth approach because it is based on the notions of artist similarity. That is, propagation can incorporate other forms of music information, such as user preference data, to generate tags for music. However, it cannot be used in isolation from these other approaches because it makes direct use of an initial set of annotated artists.

In the next section, we present the general tag propagation algorithm. We then introduce four different music information sources that are individually useful for calculating artist similarity. Section 4 describes the two evaluation metrics that we use to test our system with a database of 3,500 artists, four similarity metrics, and three types of tags. We discuss the results in Section 5, and conclude in Section 6.

2. TAG PROPAGATION

Compared with other automatic tagging algorithm, tag propagation is relatively straightforward. Suppose that we

¹ Pandora statistics are based on personal notes for a public talk by Pandora founder Tim Westergren. AMG statistics were found at <http://www.allmusic.com>.

want to annotate a novel artist a . We find the most similar artists of a , combine existing annotations of them, and select the tags that appear frequently.

More formally, tag propagation requires two matrices: a *similarity* matrix \mathbf{S} and a *tag* matrix \mathbf{T} . \mathbf{S} is an artist-by-artist similarity matrix where $[\mathbf{S}]_{i,j}$ indicates similarity score between artist i and j . \mathbf{T} is an artist-by-tag matrix where $[\mathbf{T}]_{a,t}$ represents the strength of association between artist a and tag t . In this paper, we consider the entries in \mathbf{T} to be a binary number of 0 or 1, where 0 represents unknown or weak association, and 1 indicates a strong association. We call the a -th row of \mathbf{T} the *tag annotation vector*, and denote as \mathbf{t}_a .

Once we have a similarity matrix \mathbf{S} (as described in Section 3), we can use the standard k -Nearest Neighbor (kNN) algorithm to propagate tags. For the artist a in question, we find the k most similar artists (i.e., the neighbors), which we denote as \mathcal{N}_a . The neighbors are the columns corresponding to the k largest values in the a -th row of \mathbf{S} . We average the annotation vectors from \mathbf{T} of \mathcal{N}_a to estimate the annotation vector $\hat{\mathbf{t}}_a$ of a .

$$\hat{\mathbf{t}}_a = \frac{\sum_{i \in \mathcal{N}_a} \mathbf{t}_i}{k} \quad (1)$$

Based on an exponential grid search with $k \in \{2^i | 0 \leq i \leq 6\}$, we find that k between 8 and 64 results in comparable performance for each of our approaches. As such, we set $k = 32$ for each of our experiments in Section 5.

3. ARTIST SIMILARITY

In this section, we describe ways in which we can calculate artist similarity matrices from four different sources of music information.² In that our goal is to evaluate tag propagation, we primarily make use of existing music IR approaches [12–15].

3.1 Collaborative Filtering (CF)

Collaborative filtering (CF) is a popular commercial technique for calculating artist similarity [16] that is based on user preference data. The idea is that two artists are considered similar if there is a large number of users that listen to both artists. In this paper, we consider two forms of user preference data: *explicit* feedback and *implicit* feedback. Feedback is explicit if a user has indicated directly that he or she “likes” an artist. This information is often recorded by a user through a button on a music player interface. Implicit feedback is found by tracking user listening habits. For example, Last.fm monitors which songs each of their users listens to over a long period of time. Implicit feedback assumes that two artists are similar if many users listen to songs by both artists.

We aggregate user preference data from 400,000 Last.fm users, and build an artist similarity matrix, *CF-Explicit*, by counting the number of users who have explicitly indicated that they like both artists. We construct

² The data that we describe in this paper was collected from the Internet in April of 2008.

Table 1. Most similar pairs of artists based on CF (explicit) and their top social tags.

Tex Ritter Red Foley	country country	classic country classic country	country roots boogie	oldies rock	old timey american
Unwound Young Widows	noise rock noise rock	post-hardcore post-hardcore	indie rock math rock	math rock experimental	post-rock heavy
DLG Puerto Rican Power	salsa salsa	latin latin	dlg mambo	bachata latino	spanish cuba
Starkillers Kid Dub	dance electro	house electro house	trance electronic	electro house dub	electronica electro-house
Lynda Randle George Jones	gospel country	female vocalists classic country	christian americana	southern gospel singer-songwriter	female vocalist traditional country
An Albatross See You Next Tuesday	experimental grindcore	grindcore deathcore	noisecore mathcore	hardcore experimental	noise noisecore

a second similarity matrix, *CF-Implicit*, by counting the number of users who listen to both artists at least 1% of the time.

One issue that arises when using the raw co-occurrence counts is that the popular artists tend to occur frequently as a “most similarity” artist [16]. A standard solution is to *normalize by the popularity* of each artists:

$$[\mathbf{S}]_{i,j} = \frac{co(i,j)}{\sqrt{\sum_{k \in \mathcal{A}} co(i,k)} \sqrt{\sum_{k \in \mathcal{A}} co(k,j)}} \quad (2)$$

where \mathcal{A} is the set of 3,500 artists, $co(i,j)$ is the number of users that have given feedback for both artist i and artist j (explicit or implicit depending on the matrix type). Note that this equation is equivalent to the cosine distance between two column vectors of a User-by-Item rating matrix if we assume that users give binary rating [16].

It could be the case that similarity based on CF is not strongly related to semantic similarity, and thus might not be useful for tag propagation. However, if we look at a couple of examples (see Table 1), we find that similar artists share a number of common tags. This is confirmed in Section 5.1, when we quantitatively compare the performance of tag propagation using *CF-Explicit* and *CF-Implicit*. We also report on the effect of popularity normalization for these two approaches.

3.2 Social Tags (ST)

As described in Section 1.1, social tags (ST) are socially generated semantic information about music. Lamere and Celma [13] show that computing artist similarity using social tags produces better performance for music recommendation than other approaches such as collaborative filtering, content-based analysis, or human expert recommendations.

Following their approach, we collect a set of social tags (represented as a tag annotation vector t_a) for each artist a from Last.fm. However, when collecting this data set, we found a total of about 30,000 unique tags for our 3,500 artists from Last.fm. Since Last.fm allows anyone to apply

any tag, this vocabulary of tags contains many rare tags that seemed to be (inconsistently) applied to a small number of artists [1]. In an attempt to clean up the data, we choose to prune tags that are associated with less than .5% of the artists. This resulted in vocabulary of 949 unique tags.

The ST artist similarity matrix \mathbf{S} is built by calculating cosine similarity between each annotation vector:

$$[\mathbf{S}]_{i,j} = \frac{\mathbf{t}_i \cdot \mathbf{t}_j}{|\mathbf{t}_i| |\mathbf{t}_j|} \quad (3)$$

where each annotation vector \mathbf{t} is a vector over 949 dimension.

3.3 Web Documents (WD)

Web documents represent a third source of music information that can be used to calculate music similarity. For each artist a , we collect 50 documents from the Google Search Engine³ with query ‘‘artist name’’ music. We combine the top 50 results into a single document and then represent that document as a bag-of-words. This bag-of-words is converted into the term-frequency-inverse-document-frequency (TF-IDF) *document* vector \mathbf{d}_a over a large vocabulary of words [17]. TF-IDF is a standard text-IR representation that places more emphasis on words that appear frequently in the given document and are less common in the entire set of documents.

We build the WD artist similarity matrix \mathbf{S} by calculating cosine similarity score on each pair of TF-IDF document:

$$[\mathbf{S}]_{i,j} = \frac{\mathbf{d}_i \cdot \mathbf{d}_j}{|\mathbf{d}_i| |\mathbf{d}_j|} \quad (4)$$

where i, j are artists.

3.4 Content-Based Analysis (CB)

Lastly, we explore two content-based (CB) approaches for calculating artist similarity that have performed well in various MIREX tasks [12, 15, 18] in recent years. For both approaches, we begin by extracting a bag of Mel-Frequency

³ www.google.com

Cepstral Coefficients (MFCCs) feature vectors from one randomly selected song by each artist.

Our first approach, which was proposed by Mandel and Ellis [12] (referred to as *CB-Acoustic*), models the bag-of-MFCCs with a single Gaussian distribution over the MFCC feature space. To calculate the similarity between two artists, we calculate the symmetric KL divergence between the two Gaussian distributions for the songs by the two artists. For this approach, we use the first 20 MFCCs and estimate the Gaussian distribution using a full covariance matrix. This approach is chosen because it is fast, easy to compute, and a popular baseline within the music-IR community.

The second approach, proposed by Barrington et al. [15] (referred to as *CB-Semantic*), involves estimating the KL-divergence between the two *Semantic Multinomial* distributions corresponding to the selected songs for each pair of artists. A semantic multinomial is a (normalized) vector of probabilities over a vocabulary of tags. To calculate the semantic multinomial, we first learn one Gaussian Mixture Model (GMM) over the MFCC feature space for each tag in our vocabulary. The GMMs are estimated using training data (e.g., songs that are known to be associated with each tag) in a supervised learning framework. We then take a novel song and calculate its likelihood under each of the GMMs to produce a vector of unnormalized probabilities. When normalized, this vector can be interpreted as a multinomial distribution over a semantic space of tags. We choose a vocabulary of 512 genres and acoustic tags and use 39-dimensional MFCC+Delta feature vectors. MFCC+Delta vectors include the first 13 MFCCs plus each of their 1st and 2nd instantaneous derivatives. This approach is chosen because it is based on a top performing approach in the 2007 MIREX audio similarity task and is based on a top performing approach in the 2008 MIREX audio tag classification task.

4. EXPERIMENTAL SETUP

4.1 Data

Our data set consists of 3,500 artists with music that spans 19 top-level genres (e.g., Rock, Classical, Electronic) and 123 subgenre (e.g., Grunge, Romantic Period Opera, Trance). Each artist is associated with 1 or more genre and 1 or more subgenres. The set of 142 genres and subgenres make up our initial *Genre* vocabulary.

For each artist, we collect a set of acoustic tags for songs by the artist from Pandora’s Music Genome Project. This *Acoustic* tag vocabulary consists of 891 unique tags like “dominant bass riff”, “gravelly male vocalist”, and “acoustic sonority”. In general, these acoustic tags are thought to be *objective* in that two trained experts can annotate a song using the same tags with high probability [19]. Lastly, we collect social tags for each artist using the Last.fm public API as discussed in Section 3.2. After pruning, the *Social* tag vocabulary, it consists of 949 unique tags.

In all three cases, we construct a binary ground truth tag matrix \mathbf{T} where $[\mathbf{T}]_{s,a} = 1$ if the tag is present for the

artists (or in one of the songs by the artists), and 0 otherwise.

4.2 Evaluation Metrics

We use leave-one-out cross-validation to test our system. For each artist a , we hold out the ground truth tag annotation vector \mathbf{t}_a and calculate the estimated vector $\hat{\mathbf{t}}_a$ by k NN algorithm. In the *artist annotation* test, we test how well we can propagate relevant tags to a novel artist by comparing the estimated vector with the ground truth.

In the *tag-based retrieval* test, we generate a ranked list of the artists for each tag based on their association strength to a tag. Then we evaluate how high the relevant artists are placed on the ranked list. Each test is described in detail below.

One of our artist similarity metric is based on the similarity of socially generated tags as discussed in Section 3.2. We use tags generated by Last.fm users as our data source because it provides the largest data set of social tags. Unfortunately, we evaluate our system on the same data as well. Therefore, we use 10-fold cross-validation to evaluate the propagation of *social* tags based on the similarity of *social* tags. That is, for each of 10 folds, we use 90% of the tags to estimate a song similarity matrix. This similarity matrix is used to propagate the other 10% of the tags. We can combine the 10 estimated annotation tag vectors from each of the 10 folds into one complete annotation vector.

4.2.1 Artist Annotation

For each artist a , we evaluate the relevance of the estimated annotation vector $\hat{\mathbf{t}}_a$ by comparing it to the ground truth \mathbf{t}_a . As described earlier, the ground truth data is in binary format. We transform the estimated annotation vector into the same binary vector by setting each value that is above a threshold to 1, and zero otherwise.

By doing so, we move from the estimation problem to the standard retrieval problem [17]. That is, we predict a set of relevant tags to describe the artist. We can then calculate precision, recall and f-measure for the given threshold. By varying threshold, we compute a precision-recall curve as shown in Figure ??.

4.2.2 Tag-Based Retrieval

In this experiment, we evaluate the performance of tag-based retrieval of relevant artists. For each tag, we generate a ranked-list of 3,500 artists. The rank is based on the association score of the tag in each artist’s estimated annotation vector. Using the ground truth annotations, we calculate R-precision, 10-Precision, MAP (mean average precision) and AUC (area under the ROC curve) for each tag [17]. We then average the performance of the tags in each of our three tag vocabularies: Pandora Genre, Pandora Acoustic, and Last.fm Social.

Table 2. Exploring variants of collaborative filtering (CF): We report the average f-measure / area under the ROC curve (AUC) for explicit or implicit user preference information when we have either normalized or not normalized for popularity. Each evaluation metric is the average value over the three tag vocabularies.

	Unnormalized	Normalized
Explicit	.438 / .867	.495 / .885
Implicit	.410 / .824	.502 / .891

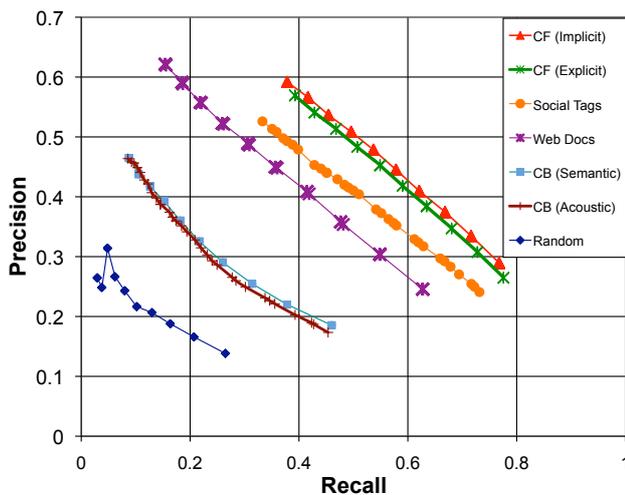


Figure 1. Semantic annotation and retrieval model diagram.

5. RESULTS

5.1 CF Comparison

The collaborative filtering approach has four variants with two sets of varying conditions. First, we compare using explicit and the implicit user preference data. Second, the similarity matrix S was generated with and without the popularity-normalization. We evaluate the performance of each variant by comparing f-measure from the *artist annotation* test and area under the ROC curve (AUC) from the *tag-based retrieval* test.

The result of each test is illustrated in Table 2. In our experiments, we observe no significant difference between the explicit and the implicit user preference data. However, in both cases, the normalization improves the performance. It is interesting that the normalization boosts the performance of the implicit data more significantly than the explicit data. This could be due to the fact that implicit data may be more prone to the popularity bias since Last.fm radio playlists tend to recommend music from *popular* artists [16].

5.2 Artist Annotation

The precision-recall curves for artist annotation are plotted in Figure ???. For each test, we varied the threshold from

0.1 to 0.4 with the interval of 0.01 and calculated precision, recall, and f-measure. The baseline *Random* performance is calculated by estimating each annotation vector with $k = 32$ distinct random neighbors. Except for the random baseline, the f-measure was maximized at around a threshold of 0.3.

In general, the two variants of the collaborative filtering (CF) approach perform best, with the *implicit* feedback approach performing slightly better. This is surprising because the collaborative filtering approach does not explicitly encode semantic information whereas social tag, web documents, and CB-Semantic are based on the similarity of semantic information. This suggests that collaborative filtering is useful for determining semantic similarity as well as music recommendation.

5.3 Tag-based Retrieval

We evaluate tag-based music retrieval based on tag propagation using seven approaches to computing music similarity. We report the performance for three vocabularies of tags (*Genre*, *Acoustic*, and *Social*) in Table 3.

As was the case with artist annotation, both CF-Implicit and CF-Explicit show strong performance for all four metrics and all three vocabularies. However, ST has the best performance for R-Precision, 10-Precision, and MAP when propagating social tags.

Since area under the ROC curve (AUC) is an evaluation metric that is not biased by the prior probability of relevant artists for a given tag, we can safely compare average AUC values across the different tag vocabularies. Based on this metric, we see that all of the approaches (except for the CB-Acoustic) have higher AUC values in the order of *Genre*, *Acoustic*, and *Social* tag sets. This suggests that it may be easiest to propagate genres and hardest to propagate social tags to novel artists.

Both CB approaches show relatively poor performance (though much better than random), which is disappointing since all of the other methods require additional human input to calculate music similarity for a novel artist. That is, if either CB approached showed better performance, we could remedy the data sparsity problem for novel artists with a fully automatic tag propagation approach.

6. CONCLUSION

In this paper, we have explored tag propagation as a technique for annotating artists with tags. We explored alternative ways to calculate artist similarity by taking advantage of the existing sources of music information such as user preference data (CF), social tags (ST), web documents (WD), and audio content (CB). Each similarity metric was tested on three distinct tag sets: *genre*, *acoustic*, and *social*. Both *artist annotation*, and *tag-based retrieval tests* show that CF generally performs the best, followed by ST, WD, and CB. This result is somewhat surprising because collaborative filtering (CF) is solely based on the aggregate trends of listening habits and user preferences, rather than explicitly representing music semantics. It confirms the idea that CF similarity (e.g., user behavior) can be

Table 3. Tag-based music retrieval performance. Each evaluation metric is averaged over all tags for each of the three vocabularies. R-precision for a tag is the precision (the ratio of correctly-labelled artists to the total number of retrieved artists) when R documents are retrieved, where R is the number of relevant artists in the ground-truth. Similarly, 10-precision for a tag is the precision when 10 artists are retrieved (e.g., the “search engine metric”). Mean average precision (MAP) is found by moving down the ranked list of artists and averaging the precisions at every point where we correctly identify a relevant artist based on the ground truth. The last metric is the area under the receiver operating characteristic (ROC) curve (denoted AUC). The ROC curve compares the rate of correct detections to false alarms at each point in the ranking. A perfect ranking (i.e., all the relevant songs at the top) results in an AUC equal to 1.0. We expect the AUC to be 0.5 if we randomly rank songs. More details on these standard IR metrics can be found in Chapter 8 of [17].

Approach	Genre (142 tags)				Acoustic (891 tags)				Social (949 tags)			
	r-prec	10-prec	MAP	AUC	r-prec	10-prec	MAP	AUC	r-prec	10-prec	MAP	AUC
Random	0.012	0.015	0.017	0.499	0.025	0.023	0.029	0.495	0.030	0.029	0.033	0.498
CF (implicit)	0.362	0.381	0.342	0.914	0.281	0.306	0.254	0.882	0.409	0.543	0.394	0.876
CF (explicit)	0.362	0.388	0.329	0.909	0.282	0.304	0.246	0.878	0.410	0.562	0.396	0.869
ST	0.344	0.349	0.311	0.889	0.267	0.274	0.237	0.874	0.428	0.584	0.413	0.874
WD	0.321	0.393	0.282	0.861	0.244	0.300	0.200	0.814	0.318	0.478	0.286	0.797
CB (acoustic)	0.101	0.127	0.076	0.701	0.118	0.132	0.088	0.692	0.117	0.159	0.092	0.661
CB (semantic)	0.087	0.103	0.069	0.687	0.115	0.123	0.091	0.714	0.107	0.126	0.084	0.662

used to capture the semantic similarity (e.g., tags) among artists. We also found that two content-based approaches (CB) performed poorly in our experiments. This is unfortunate because content-based similarity can be calculated for novel artists without human intervention, and thus would have solved the data sparsity problem.

7. REFERENCES

- [1] D. Turnbull, L. Barrington, and G. Lanckriet. Five approaches to collecting tags for music. *ISMIR*, 2008.
- [2] P. Lamere. Social tagging and music information retrieval. *JNMR*, 2008.
- [3] M. Sordo, C. Lauier, and O. Celma. Annotating music collections: How content-based similarity helps to propagate labels. In *ISMIR*, 2007.
- [4] P. Lamere and E. Pampalk. Social tags and music information retrieval. *ISMIR Tutorial*, 2008.
- [5] D. Turnbull, R. Liu, L. Barrington, D. Torres, and G. Lanckriet. Using games to collect semantic information about music. In *ISMIR '07*, 2007.
- [6] E. Law and L. von Ahn. Input-agreement: A new mechanism for data collection using human computation games. *ACM CHI*, 2009.
- [7] B. Whitman and D. Ellis. Automatic record reviews. *ISMIR*, 2004.
- [8] P. Knees, T. Pohle, M. Schedl, and G. Widmer. A music search engine built upon audio-based and web-based similarity measures. In *ACM SIGIR*, 2007.
- [9] M. Mandel and D. Ellis. Multiple-instance learning for music information retrieval. In *ISMIR*, 2008.
- [10] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE TASLP*, 16(2):467–476, February 2008.
- [11] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In *Neural Information Processing Systems Conference (NIPS)*, 2007.
- [12] M.I. Mandel and D.P.W. Ellis. Song-level features and support vector machines for music classification. *ISMIR*, 2005.
- [13] P. Lamere and O. Celma. Music recommendation tutorial notes. *ISMIR Tutorial*, September 2007.
- [14] A. Berenzweig, B. Logan, D. Ellis, and B. Whitman. A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, pages 63–76, 2004.
- [15] L. Barrington, A. Chan, D. Turnbull, and G. Lanckriet. Audio information retrieval using semantic similarity. *ICASSP*, 2007.
- [16] O. Celma. *Music Recommendation and Discovery in the Long Tail*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2008.
- [17] C.D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [18] S. J. Downie. The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, 2008.
- [19] T. Westergren. Personal notes from Pandora get-together in San Diego, March 2007.