Assigning and Visualizing Music Genres by Web-based Co-Occurrence Analysis

Markus Schedl[1], **Tim Pohle**[1], **Peter Knees**[1], **Gerhard Widmer**[1,2]
[1]Department of Computational Perception, Johannes Kepler University, Linz, Austria
[2]Austrian Research Institute for Artificial Intelligence, Vienna, Austria
markus.schedl@jku.at

Abstract

We explore a simple, web-based method for predicting the genre of a given artist based on co-occurrence analysis, i.e. analyzing co-occurrences of artist and genre names on music-related web pages. To this end, we use the page counts provided by Google to estimate the relatedness of an arbitrary artist to each of a set of genres. We investigate four different query schemes for obtaining the page counts and two different probabilistic approaches for predicting the genre of a given artist. Evaluation is performed on two test collections, a large one with a quite general genre taxonomy and a quite small one with rather specific genres.

Since our approach yields estimates for the relatedness of an artist to every genre of a given genre set, we can derive genre distributions which incorporate information about artists that cannot be assigned a single genre. This allows us to overcome the inflexible artist-genre assignment usually used in music information systems. We present a simple method to visualize such genre distributions with our *Traveller's Sound Player*. Finally, we briefly outline how to adapt the presented approach to extract other properties of music artists from the web.

Keywords: Web Mining, Co-Occurrence Analysis, Genre Classification, Evaluation, User Interface

1. Introduction and Motivation

The continuous growth of electronic music distribution increases the interest in automatic retrieval of meta-data for music. Today, meta-data like genre, instrumentation, or band members is usually provided by the music distributor who has to annotate the music. Unfortunately, this method has several drawbacks. First, for the distributor, it is a very labor intensive task. Second, even if annotation is performed by experts, it is usually influenced by subjective opinions and different local definitions, e.g. in Northern America the genre *Rock/Pop* is used in a broader sense than in Europe.

Intelligent methods for automatic music annotation that rely on global "knowledge" as encoded in the World Wide Web are therefore getting more and more important. To this end, we propose a very simple approach that automatically gathers descriptive information about an arbitrary artist from the web and, hence, incorporate opinions and knowledge of a huge number of people from all over the world.

In the following section, a brief overview of web mining approaches and co-occurrence analysis for tasks related to music information retrieval is given. In Section 3, we present our approach to inferring descriptive properties for music artists. We evaluate the approach on a genre assignment problem using two test collections and four query schemes. In Section 4, we show how to incorporate the extracted genre information in a music player, namely our *Traveller's Sound Player*, to facilitate browsing. Finally, Section 5 looks into the possibilities of inferring properties other than genre. We illustrate that with the property *tempo*.

2. Related Work

First experiments with co-occurrence analysis for tasks related to MIR can be found in [5], where playlists of radio stations and compilation CDs are used to find co-occurrences between titles and between artists. In [11, 2], first attempts to exploit the cultural knowledge offered by the web can be found. User collections taken from the music sharing service *OpenNap* [1] are analyzed, artist co-occurrences are extracted, and eventually, a similarity measure based on community meta-data is elaborated. This measure is evaluated by comparison with direct subjective similarity judgments obtained via a web-based survey. In contrast to this survey of non-professionals, in [1], expert opinions taken from the *All Music Guide* [2] and co-occurrences on playlists from *The Art of the Mix* [3] are used to create a similarity network of music artists.

Furthermore, co-occurrences of artist names on web pages have been successfully applied to the task of genre classification, e.g. in [8]. The approach presented in [8] uses the page counts returned by Google in reply to queries containing artist names. Based on these page counts, complete similarity matrices are determined, i.e. a similarity value is calculated for every pair of artists. However, this approach is computationally complex and hardly applicable for large music collections. An alternative that does not produce complete similarity matrices is proposed in [12]. Here,

[1] *http://opennap.sourceforge.net*
[2] *http://www.allmusic.com*
[3] *http://www.artofthemix.org*

the aim is to find similar artists to a given seed artist using Amazon's and Google's web services. A list of artists potentially related to the seed artist is used to calculate co-occurrences. Based on the number of web pages on which the seed artist and the potentially related artists co-occur, a "relatedness" is defined for every potentially related artist, and the artists are presented to the user in the order of their relatedness.

In contrast, the approach presented in [3] considers the content of artist-related web pages rather than only their page counts. The common text mining technique *TF* $\cdot$ *IDF* is applied to weight each of a set of words extracted from the particular web pages. The resulting term profiles are used for artist-to-genre classification.

Besides similarity measurement and genre classification, co-occurrence analysis has also been applied to the task of detecting prototypical artists for a given genre. In [9], we used a technique based on a similar idea as Google's *Page-Rank Citation Ranking* (cf. [6]) on page count estimates to derive the prototypicality of each of a set of artists for a given genre. In [10], this approach is extended to avoid distractions caused by artist names that equal common speech words.

The approaches to genre classification presented so far usually predict the genre of an unknown artist on the basis of similarities to already classified artists using, for example, *Support Vector Machines* or *k-Nearest Neighbor* classification. In contrast, our approach does not depend on an a priori assignment of artists to genres. In other words, we require no labeled training set. Indeed, lists of artist and genre names are sufficient since we directly investigate occurrences of artist and genre names on music-related web pages instead of deriving similarities between artists.

3. Genre Assignment by Co-Occurrence Analysis

Our approach to infer genre information about an arbitrary artist relies on the automatic analysis of results to specific queries raised to an arbitrary search engine. We use Google since it is the most popular search engine and provides a Web API [4]. Since we do not have access to artist collections that are annotated with meta-data other than genre, we must restrict evaluation to genre classification. As a result, we explain the approach for gathering genre information. However, we will show how to adapt the approach for extracting arbitrary properties in Section 5.

3.1. Methodology

The basic approach that we propose is very simple. Given two lists, one of artist names and one of genre names, we first query Google to estimate the total number of pages on which each single name of the two lists is mentioned. We denote the returned page counts as pc_a and pc_g, where a is the artist name and g is the genre name. We further investigate for every combination of artist and genre name, on how many web pages both can be found (denoted as $pc_{a,g}$). For the task of genre classification, we are indifferent of the order of the respective terms. [5]

To determine the genre of an artist, we investigate two different probabilistic approaches. Both use relative frequencies based on page counts. The first one estimates the conditional probability for the artist name to be found on a web page that mentions the genre name, more formally, $p(a|g) = \frac{pc_{a,g}}{pc_g}$. The second one estimates the probability for the genre name to be found on a page that contains the artist name, formally, $p(g|a) = \frac{pc_{a,g}}{pc_a}$. Both approaches yield, for every artist, a probability distribution for its relatedness to each genre and should therefore be able to deal with artists that cannot be assigned a single genre, for example, artists that produce music of very different styles. Having calculated $p(a|g)$ or $p(g|a)$ for the artist a to be classified and all potential genres g, we simply predict the most probable genre. Compared to the approach which we proposed in [8], the approach presented here usually has a much lower computational complexity since it only needs $a \cdot g$ queries and calculations (a being the number of artists, g the number of genres, which is usually much lower than a). The approach presented in [8] has complexity quadratic in a.

3.2. Experiments and Evaluation

We evaluated four different query schemes to obtain the page counts. They vary in regard to additional keywords added to the artist or genre name.

- *M*: "artist/genre name"+music
- *MG*: "artist/genre name"+music+genre
- *MS*: "artist/genre name"+music+style
- *MGS*: "artist/genre name"+music+genre+style

Since we aim at restricting the search results to web pages related to *music*, we use this keyword in all schemes. Additionally, we add the terms *genre* and/or *style* to describe the properties which we intend to capture.

For evaluation, two test collections were used. The first one comprises 1995 artists from 9 very general genres that were taken from the *All Music Guide*. [6] We abbreviate this collection as *C1995a* in the following. *C1995a* is used to test our approach on popular and mostly well-known artists. A list of the artists together with their assigned genres can be downloaded from *http://www.cp.jku.at/people/schedl/music/C1995a_artists_genres.txt*. Since we aimed at enriching our

[4] *http://www.google.at/apis*

[5] For predicting general properties, it may be better to take the order of the search terms into account, e.g. search for exact phrase "loud volume".

[6] The collection *C1995a* contains artists from the genres *Blues* (9.4%), *Country* (12.3%), *Electronica* (4.8%), *Folk* (4.1%), *Heavy Metal* (13.6%), *Jazz* (40.7%), *Rap* (2.1%), *Reggae* (3.0%), and *RnB* (10.1%).

Table 1. Accuracies in percent for the genre prediction task on the 1995-artist-collection for the different query schemes. The upper part of the table shows the accuracies obtained using $\frac{pc_{a,g}}{pc_g}$, the lower one those obtained with $\frac{pc_{a,g}}{pc_a}$. The last row shows the results obtained with the modified genre names (for $\frac{pc_{a,g}}{pc_a}$).

predictions	**1**	**2**	**3**	**4**	**5**
$pc_{a,g}/pc_g$					
M	42.01	65.87	76.09	82.21	87.37
MG	57.10	70.43	77.20	80.50	83.41
MS	36.89	65.36	73.13	79.55	84.86
MGS	23.96	39.35	50.63	61.86	72.48
$pc_{a,g}/pc_a$					
M	57.24	68.07	72.18	75.39	78.40
MG	62.31	68.07	72.78	77.04	79.80
MS	63.31	68.37	71.48	74.94	77.19
MGS	43.56	58.85	68.67	73.73	78.50
$pc_{a,g}/pc_a$ *with modified genre names*					
MS	71.33	81.75	86.27	93.13	95.14

Traveller's Sound Player with genre information extracted from the web, we needed a second collection that not only contains artist names, but real music tracks. To this end, we compiled an in-house collection containing 2545 tracks by 103 (partially quite unknown) artists that are clustered in 13 much more specific genres than in *C1995a*. Artist and genre names are available at *http://www.cp.jku.at/people/schedl/music/C103a_artists_genres.txt*. We denote this second collection *C103a*.[7]
We ran the evaluation experiments using each combination of query scheme, prediction approach, and test collection. Since genre is an ill-defined concept, it is often impossible to assign an artist to one particular genre. This issue together with the fact that our approach yields probabilities rather than boolean values for the relatedness of an artist to each genre permits us to predict more than one genre for an artist. However, our test collections only show a $1 : n$ assignment between genre and artist. Thus, we try to account for the probabilistic output of our genre classifier in the evaluation by investigating not only the most probable genre of an artist but up to 5 genres (those with maximum probability). Hence, if the correct genre with respect to our ground truth is within the 5 most probable genres predicted by our approach, we rate the classification result as correct. Of course, we also show the results when allowing only 1, 2, 3, and 4 genre(s) to be predicted.

3.3. Results and Discussion

In Table 1, the evaluation results for the collection *C1995a* are shown. It can be seen that the prediction approach that

[7] The collection *C103a* contains tracks from the genres *A Cappella* (4.4%), *Acid Jazz* (2.7%), *Blues* (2.5%), *Bossa Nova* (2.8%), *Celtic* (5.2%), *Electronica* (21.1%), *Folk Rock* (9.4%), *Italian* (5.6%), *Jazz* (5.3%), *Metal* (16.2%), *Punk Rock* (10.2%), *Rap* (13.0%), and *Reggae* (1.9%).

Table 2. Accuracies in percent for the genre prediction task on the 103-artist-collection for the different query schemes. The upper part of the table shows the accuracies obtained using $\frac{pc_{a,g}}{pc_g}$, the lower one those obtained with $\frac{pc_{a,g}}{pc_a}$.

predictions	**1**	**2**	**3**	**4**	**5**
$pc_{a,g}/pc_g$					
M	29.13	45.63	61.17	71.85	79.61
MG	44.66	57.28	64.08	71.85	78.64
MS	30.10	47.57	61.17	69.90	72.82
MGS	30.10	44.66	58.25	66.02	73.79
$pc_{a,g}/pc_a$					
M	36.89	41.75	48.54	58.25	67.96
MG	33.98	42.72	48.54	52.43	58.25
MS	35.92	40.78	48.54	51.46	65.05
MGS	33.98	37.86	48.54	53.40	62.14

relates the combined page counts to the page counts of the web pages containing artist information ($\frac{pc_{a,g}}{pc_a}$) yields better results than $\frac{pc_{a,g}}{pc_g}$ for this collection, at least when looking at only the 1 or 2 top-ranked predictions (columns 1 and 2). An explanation for this may be that the artists of *C1995a* are grouped in very general genres for which a disproportionally large number of web pages (with respect to the genre classification task) exists. Therefore, the occurrence of a genre name on a web page that mentions the artist under consideration is more likely to indicate a correct artist-genre assignment than vice versa. Furthermore, we can state that the query schemes *MG* and *MS* perform better than the simple *M* and the complex *MGS* schemes.
Table 2 shows the classification results for the collection *C103a*. These are obviously worse since the genre taxonomy used for this collection clusters the artists according to much more specific and partially overlapping genres. Another interesting fact is that, overall, the prediction approach $\frac{pc_{a,g}}{pc_g}$ yields better results than $\frac{pc_{a,g}}{pc_a}$ for this collection. The reason for this is contrary to the explanation given above for the collection *C1995a*. The best results are obtained when using the query scheme *MG* with the prediction approach $\frac{pc_{a,g}}{pc_g}$.
Since we also wanted to investigate which genres are often confused, we draw confusion matrices that can be found for the best performing settings (query scheme and prediction approach) in Figure 1 for the collection *C1995a* and in Figure 2 for the collection *C103a*. A closer look at Figure 1 reveals that the genres *Blues*, *Country*, *Jazz*, *Rap*, and *Reggae* are usually classified correctly, whereas the performance of *Electronica*, *Heavy Metal*, and *RnB* is very bad. Since we suspected this to be the result of ambiguous genre names (e.g. instead of *Electronica*, *Electronic* may be used to denote the same genre), we performed evaluation again with slightly modified genre names. More precisely, instead of *Electronica*, we used *Electronic*, instead of *Heavy Metal*, we used *Metal*, and instead of *RnB*, we used *R&B*, which is

correct genres / predicted genres	Bl	Cou	Ele	Folk	HM	Jazz	Rap	Reg	RnB
Bl	89.9	5.3	0.5	0.5	0.5	2.1		1.1	
Cou	0.8	96.3	0.4	1.2		0.8		0.4	
Ele		23.2				71.6	3.2	1.1	1.1
Folk	16	42	2.5	34.6	1.2	2.5	1.2		
HM	19.9	73.8	0.4	0.4	2.6	2.6			0.4
Jazz	3.7	3.3	0.1	0.4	0.5	90.8	0.6	0.2	0.4
Rap		4.9				2.4	92.7		
Reg	5	13.3				1.7		80	
RnB	42.6	34.2	1.5			19.3	2	0.5	

Figure 1. Confusions for the genre prediction task performed on the 1995-artist-collection using the settings *MS* and $\frac{pc_{a,g}}{pc_a}$.

correct genres / predicted genres	AC	AJ	Bl	BN	Ce	El	FR	Ita	Jaz	Met	PR	Rap	Reg
AC	75				25								
AJ				25								25	50
Bl			100										
BN				100									
Ce					100								
El	7.1	7.1		3.6		25	3.6				7.1	7.1	39.3
FR					40			20		40			
Ita				20							20		60
Jaz				44.4	11.1				33.3		11.1		
Met					28.6					64.3	7.1		
PR					16.7						83.3		
Rap				41.7	8.3					16.7	8.3	25	
Reg													100

Figure 2. Confusions for the genre prediction task performed on the 103-artist-collection using the settings *MG* and $\frac{pc_{ag}}{pc_g}$.

correct genres / predicted genres	Bl	Cou	Ele	Folk	Met	Jazz	Rap	Reg	R&B
Bl	89.9	5.3	0.5	0.5		2.1		1.1	0.5
Cou	1.2	95.5	0.8	1.2	0.4	0.8			
Ele		14.7	40		1.1	41.1	2.1	1.1	
Folk	16	43.2	1.2	33.3		2.5	1.2	1.2	1.2
Met	4.1	45.8	0.7	0.4	46.9	1.8			0.4
Jazz	4.1	3.3		0.4	0.1	91.2	0.5	0.2	0.1
Rap		4.9				2.4	92.7		
Reg	5	13.3			1.7	1.7		78.3	
R&B	42.6	34.7			0.5	19.3	1.5	0.5	1

Figure 3. Confusions for the genre prediction task performed on the 1995-artist-collection using the modified genre names. The settings *MS* and $\frac{pc_{a,g}}{pc_a}$ were applied.

a more common abbreviation. The accuracies obtained with these modified genre names can be found in the last row of Table 1, the confusion matrix is depicted in Figure 3. It can be seen that the slight modifications considerably improve performance (by more than 8% overall), especially for the genres *Electronic* and *Metal*. *R&B* still seems to be too specific an expression.

However, this modification cannot improve the following distortion that becomes obvious when inspecting the second column of Figure 1 or 3. The genre *Country* is incorrectly predicted for a large number of artists. This can be explained by the fact that many web pages contain the term "country", but not to denote a genre name but to describe the country of origin of an artist. Moreover, *Electronica* is often misclassified as *Jazz*. This is not very surprising since the genre *Electronica* contains many artists that may also be classified as *Acid Jazz*. Finally, *RnB* is often misclassified as *Blues* because of the similar genre names.

4. Visualizing Genre Distributions

In the following, we show how to integrate the gathered genre meta-data into an existing music player. First, we present our *Traveller's Sound Player*. Then, we elaborate on how we extended it to visualize the genre distribution of arbitrary music collections. We demonstrate it on the collection *C103a*, which we already used for evaluating our genre prediction approach.

4.1. The Traveller's Sound Player

Our *Traveller's Sound Player (TSP)* was originally presented in [7]. The basic idea of the *TSP* is to arrange the tracks of a music collection around a wheel that serves as a track selector (cf. Figure 4) such that consecutive tracks are maximally similar. For this purpose, a large circular playlist is created by applying a Traveling Salesman algorithm on audio similarities. Provided that the heuristic used to solve the Traveling Salesman Problem finds a good tour, stylistically coherent areas emerge around the wheel. A more detailed elaboration on the used similarity measure and evaluations of different TSP algorithms can be found in [7].

Figure 4. Our *Traveller's Sound Player* extended with the visualization of genre distributions.

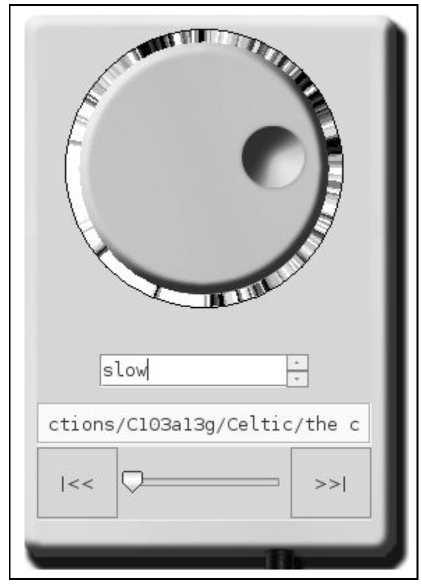

Figure 5. Our *Traveller's Sound Player* extended with the visualization of tempo distribution.

4.2. Visualization Technique

A drawback of the existing version of the *TSP* is that it does not guide the user in finding certain styles of music. Indeed, the user has to explore different regions of the playlist by randomly selecting different angular positions with the wheel.

To overcome this problem, we extended the *TSP* by visualizing distributions of meta-data, genre in our case, to facilitate browsing the collection. For this purpose, we use the genre distributions obtained by the approach which we presented in Section 3. We cluster the tracks of the collection in 360 bins, one for each degree. For every bin, we then calculate the mean of the probability values of the contained tracks. Performing this for every genre gives a smoothed distribution of each genre along the playlist. The values of the genre distributions are mapped to gray values and made available to the user via a ring which is visualized around the wheel. To switch between the visualizations of the particular genre distributions, the user is offered a choice box. In Figure 4, a screenshot of the extended *TSP* is depicted. In this example, the user has chosen to visualize the distribution of the genre *A Cappella* and can easily find music of that style.

5. Inferring General Properties

We also tried to apply our approach to inferring descriptive attributes for artists, e.g. period of activity/popularity, geographical origin, or the preferred tempo of their music. However, since most of the attribute values are mutually exclusive (e.g. tempo can be slow or fast), we found that calculating and visualizing probability distributions (like in the case of genres) did not yield good results in regard to the discriminability of the attribute values. We therefore adopt an alternative approach that assigns every artist the most probable value of the attribute under consideration. This produces only discrete values 0 and 1 for the attribute distribution of an artist. Following this approach for deriving the distribution of the tempo values *slow* and *fast* using the query scheme *"artist name"+music+tempo+*[*slow*/*fast*] and the prediction method $\frac{pc_{a,tempo=slow/fast}}{pc_a}$ on the collection *C103a* produces visualizations like the one depicted in Figure 5. Comparing this screenshot with Figure 4 reveals that areas predicted to contain music of the genre *A Cappella* also show high values for the property *slow tempo*. Likewise, *Bossa Nova*, *Blues*, and *Jazz* correspond to slow tempo, whereas the distribution of the attribute value *fast* correspond to the genres *Metal* and *Punk Rock*. Indeed, Pearson's linear correlation coefficient between the distribution of the genre *Metal* and that of *fast tempo* is 0.51. For *Punk Rock*, this correlation equals 0.36.

6. Conclusions and Future Work

We have presented a web-based artist-to-genre classification approach with computational complexity $a \cdot g$, where a is the number of artists to be classified, and g is the number of classes (genres). The approach investigates co-occurrences of artist and genre names on music-related web pages and uses a probabilistic model to predict the genre of an arbitrary artist.

We evaluated the approach on two test collections using four different query schemes for obtaining the page counts and two different probabilistic approaches for predicting the genre ($\frac{pc_{a,g}}{pc_a}$ and $\frac{pc_{a,g}}{pc_g}$). We found that $\frac{pc_{a,g}}{pc_a}$ seems to be better suited for genre taxonomies comprising general genres (like collection *C1995a*), whereas $\frac{pc_{a,g}}{pc_g}$ is better for taxonomies of specific genres (like *C103a*). As for the different query schemes, we can state that overall *MG* and *MS* perform better than the simple *M* and the complex *MGS* schemes.

Taking into account the simplicity of our approach, it performs quite well. However, we found that it depends strongly on proper genre names. Indeed, using different names for the same genre, e.g. *Electronica* vs. *Electronic*, may considerably change accuracy. On the whole, we can state that our approach is successfully applicable for genre classification as long as the used genre taxonomy is not too specific and genre names are reasonably unambiguous.

Moreover, we briefly described first steps to adapt the approach for predicting artist properties other than genre, and showed how to use the extracted meta-data, i.e. distributions of genres or other properties, to enrich our *Traveller's Sound Player*.

As for future work, we will investigate other visualization techniques for the obtained property distributions. For example, we plan to incorporate the meta-data into our SOM-based three-dimensional user interface for navigating in music collections (cf. [4]). Furthermore, methods should be investigated for dealing with synonymous genre names in order to overcome problems like the *Electronica* vs. *Electronic* case. Finally, we will intensify our efforts in automatically extracting arbitrary properties like those used, for example, in the music search engine *musiclens* [8]. Our ultimate aim is to automatically annotate music at the track level according to an arbitrary ontology.

7. Acknowledgments

This research is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) under project number L112-N04 and by the Vienna Science and Technology Fund (WWTF) under project number CI010 (Interfaces to Music). The Austrian Research Institute for Artificial Intelligence acknowledges financial support by the Austrian ministries BMBWK and BMVIT.

References

[1] P. Cano and M. Koppenberger. The Emergence of Complex Network Patterns in Music Artist Networks. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR'04)*, Barcelona, Spain, October 2004.

[2] D. P. W. Ellis, B. Withman, A. Berenzweig, and S. Lawrence. The Quest for Ground Truth in Musical Artist Similarity. In *Proceedings of the 3rd International Symposium on Music Information Retrieval (ISMIR'02)*, Paris, France, 2002.

[3] P. Knees, E. Pampalk, and G. Widmer. Artist Classification with Web-based Data. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR'04)*, pages 517–524, Barcelona, Spain, October 2004.

[4] P. Knees, M. Schedl, T. Pohle, and G. Widmer. An Innovative Three-Dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web. In *Proceedings of the 14th ACM Conference on Multimedia 2006*, Santa Barbara, CA, USA, October 2006. submitted.

[5] F. Pachet, G. Westerman, and D. Laigre. Musical Data Mining for Electronic Music Distribution. In *Proceedings of the 1st WedelMusic Conference*, 2001.

[6] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. In *Proceedings of the Annual Meeting of the American Society for Information Science (ASIS'98)*, pages 161–172, January 1998.

[7] T. Pohle, E. Pampalk, and G. Widmer. Generating Similarity-based Playlists Using Traveling Salesman Algorithms. In *Proceedings of the 8th International Conference on Digital Audio Effects (DAFx-05)*, pages 220–225, Madrid, Spain, September 20-22 2005.

[8] M. Schedl, P. Knees, and G. Widmer. A Web-Based Approach to Assessing Artist Similarity using Co-Occurrences. In *Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing (CBMI'05)*, Riga, Latvia, June 2005.

[9] M. Schedl, P. Knees, and G. Widmer. Discovering and Visualizing Prototypical Artists by Web-based Co-Occurrence Analysis. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR'05)*, London, UK, September 2005.

[10] M. Schedl, P. Knees, and G. Widmer. Improving Prototypical Artist Detection by Penalizing Exorbitant Popularity. In *Proceedings of the Third International Symposium on Computer Music Modeling and Retrieval (CMMR'05)*, Pisa, Italy, September 2005.

[11] B. Whitman and S. Lawrence. Inferring Descriptions and Similarity for Music from Community Metadata. In *Proceedings of the 2002 International Computer Music Conference*, pages 591–598, Goeteborg, Sweden, September 2002.

[12] M. Zadel and I. Fujinaga. Web Services for Music Information Retrieval. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR'04)*, Barcelona, Spain, October 2004.

[8] *http://www.musiclens.de*