BUILDING AN INTERACTIVE NEXT-GENERATION ARTIST RECOMMENDER BASED ON AUTOMATICALLY DERIVED HIGH-LEVEL CONCEPTS

Tim Pohle[1], *Peter Knees*[1], *Markus Schedl*[1] *and Gerhard Widmer*[1,2]

[1)]Dept. of Computational Perception
Johannes Kepler University, Linz, Austria
[2)]Austrian Research Institute for Artificial Intelligence (OFAI)
Vienna, Austria

ABSTRACT

We present a new way of accessing large sets of musical artists based on high-level concepts. The concepts are derived and assigned to individual artists by an automatic procedure: Using a list of music-related words and phrases, the well-known TF×IDF approach is applied to analyse the 100 top web pages related to each artist, as delivered by a web search engine. This data then is decomposed into a number of "archetypical" bases or "concepts" by Non-Negative Matrix Factorisation (NMF). Each artist is then described by the amount by which it is related to each of these concepts. In our browser application presented here, such a representation allows for independently adjusting the weight of each of these concepts, to recommend those artists that best match the desired query profile.

1. INTRODUCTION

Have you ever heard somebody characterizing a musical artist or band that is unknown to the one he is talking to? Likely he will do it by saying something like "their music sounds like *(put some artist here that is known to both)*, but it is more *(put distinctive attribute here: aggressive / jazzy / funky...)*". In this paper, we present an approach to a computer program offering the user such an intuitive way to discover new music she may like (cf. Fig. 1). First, the user selects a "seed artist" from the dropdown list located at the center of the top area. Based on the chosen seed artist, the program then displays two types of information. On the right panel, a list is shown that contains the artists most similar[1] to the seed artist. The left panel shows attributes assigned to the seed artist and their respective weights. The user is free to modify the weights of the attributes, and as he changes the weights, the list of best-matching artists is dynamically adapted.

For such a software to be of use, of course it is crucial that the underlying artist description matches the human view. Thus, in the following, we describe the techniques we applied for obtaining them, and some experiments to assess their applicability. The remainder of this paper is organized as follows: First, we give an overview of related work in the field of Music Information Retrieval (MIR). Then, after describing the applied techniques and presenting some experimental results, we discuss our observations with the implemented prototype.

[1]as defined by an underlying similarity measure

2. RELATED WORK

Pioneering work for deriving information about musical artists from data available on the internet was done in [1]. In this work, techniques from text information retrieval were suggested to to be used for Music Information Retrieval (MIR). In [2, 3, 4], these techniques have been further investigated and additional applications have been proposed. The task of finding typical artists from a set of artists was approached in [5]. In [6], groups of similar artists are formed by clustering, enabling the user to browse through various artist categories. In [7], we have applied a self organizing map (SOM) to cluster songs based on their audio content. The music contained in each cluster is then characterized by combining the descriptions of the artists appearing in the cluster. Terms are selected based on how well they discriminate the music in the cluster to label and the other clusters. In [8], it is investigated in which way artists are connected and related in the large scale in artist recommendation systems. Recent work in web-based artist similarity computation is [9]. An example of MIR-related work applying Non-Negative Matrix Factorization is [10].

3. PROPOSED PROCEDURE

In this section, we first give an overview of the proposed procedure. Afterwards, some of the steps are discussed in more depth, including motivations for the choices we made while realizing the approach. Experimental results are given in the next section.

The outline of the proposed procedure is as follows:

1. *Obtain artist names.* This step is quite straightforward. The most common source of artist names are precompiled artist catalogues, or artist lists from the web. If an application such as the proposed artist browser should be used e.g. in a record store, the list of artists would contain the artists whose music can be bought at the store.

2. *Obtain artist descriptions.* For this step, we can think of three different possibilities for realizing it: Manually compiled, community-derived, and automatically extracted from the web. These three are discussed in Section 3.1. The output of this step is a long list of words or concepts associated with each artist. Each of the words or concepts has a weight associated.

3. *Analyse artist descriptions for common properties.* As the description (a long vector of weighted terms) that is output by the previous step is too long, it is necessary to compress it to few – ideally meaningful – concepts that make a high-level interaction feasible. This step is elaborated on in Section 3.2.

4. *Represent each artist as a mixture of the common properties.* This means to apply the transformation calculated in the previous step to obtain a compressed representation for each artist. For example, the output of step Step 2 may a length of about 2000, which is mapped to e.g. 16 concepts here.

The last step above yields a vector for each artist. As the user query also is represented by a vector of the same length, it is possible to calculate a similarity between the query vector and each artist by applying the cosine similarity measure. The user query can be seen as a user-generated artist description. Similarities are presented to the user in the form of an artist list.

In the following sections, we will focus on the description of Steps 2 and 3, because the other steps are trivial.

3.1. Obtaining Artist Descriptions

Given a list of artists (obtained in Step 1), it is necessary to obtain a description for each artist on the list (Step 2). This description has to be done uniformly for all artists, and should contain as many aspects as possible. Only a few binary labels per artist (as e.g. found on All Music Guide[2]) seem not to be sufficient for our purpose. Instead, we think of the artist descriptions as taking the form of a long (at least several hundred) list of terms with associated weights. We see three different approaches for obtaining such artist descriptions: Manually compiled, community-derived, and automatically extracted from the web.

3.1.1. Manually Compiling Artist Descriptions

To describe artists in a uniform way, one could think of a similar strategy as applied by the *Music Genome Project*[3] for music tracks. In this project, for each track a large number of annotations is created by specially trained persons. This approach has the advantage that the descriptions can be assumed to be meaningful. However, we see a number of disadvantages. The annotators need to know not only the music by the artists they are annotating, but also their socio-cultural background. For example, when thinking of Reggae, many people also think of Jamaica. Only listening to the music may not reveal such relationships. Particularly for rather unknown artists such issues may be a problem, and this may also be a source of inconsistent annotations. Furthermore, the way an artist is referred to changes over time. For example, during the *70*^{s}, nobody would have thought of the music of the *70*^{s} as “oldies”, which nowadays is one very common description for this kind of music. Also, as trends change, new artists appear and new genres emerge. Thus, a constant work would be necessary to keep the annotations up to date.

3.1.2. Community-Derived Artist Descriptions

There are web services such as Musicstrands[4] and Audioscrobbler[5] that collect user-assigned data about musical artists. In these systems, each artist is assigned a number of weighted tags by the users. Audioscrobbler offers a public web API that can be used to obtain this data. For our experiments, we used this web service, as described in Section 4.

3.1.3. Extracting Artist Descriptions from the Web

The third way we see to obtain the artist representations is to use data available on the internet that was not created with MIR applications in mind. Most notably, these are real-language music reviews that were written for human readers. As described in [1, 3, 4, 6], a search engine can be used to query text documents related to an artist name, and the top-ranked pages returned by the search engine can be analysed with text information retrieval techniques to obtain a vector of weighted terms describing each artist[6]. Such a *TF*$\times$*IDF* vector is constructed based on the frequency certain words appear on the analysed web pages.

For our experiments, we finally opted for a combination of the techniques described in Sections 3.1.2 and 3.1.3.

[2] allmusic.com

[3] http://pandora.com/mgp.shtml

[4] http://www.livestrands.com/

[5] http://www.lastfm.com

[6] Note that data obtained this way is called *community metadata* by [1]

3.2. Analyze Artist Descriptions for Common Properties

As the number of terms associated with each artist description vector is too large to be individually adjusted via the user interface, this large amount is reduced by a computational technique (Step 3).

For dividing collections of documents (represented by lists of words) into categories, a number of approaches have been used, for example Latent Semantic Indexing (LSI), k-Means Clustering and Bottom-Up Clustering Techniques. In [11], these techniques are discussed with regard to their ability to cluster documents based on their main topics, and Non-Negative Matrix Factorisation (NMF [12]) is found to be favourable.

Here, we are not interested in grouping similar artists, but in grouping similar terms to compress the long artist descriptions. In principle, all of the above techniques can be used for this task. We investigated Principal Component Analysis (PCA), clustering of term similarities based on term co-occurrences, and Non-Negative Matrix Factorization (NMF). We found NMF to work best, which seems in accordance with [11].

Given the result of an NMF, an artist represented as *TF*×*IDF* can be projected to a low-dimensional representation (Step 4).

4. EXPERIMENTAL SETUP

4.1. Obtaining Artist Names (Step 1)

For our experiments, we chose an artist repository of 1979 artists, taken from the web site All Music Guide[7]. The artists are labelled with genres as follows: *Jazz* (40.9%), *Heavy Metal* (13.2%), *Country* (12.4%), *RnB* (10.2%), *Blues* (9.4%), *Electronica* (4.8%), *Folk* (4.1%), *Reggae* (3.0%), and *Rap* (2.1%).

4.2. Calculating Artist Descriptions (Step 2)

The artist list contains a large number of rather unknown artists, so the attempt to obtain community-assigned artist tags from the Audioscrobbler[8] web service failed for most of them. After cleaning the obtained data (which included removing tags that appear only for one artist), only 331 of the 1979 artists had valid tags.

Although the tag data itself could not be used in our experiments, we assume that in general the tags that users assign to artists reflect the way people think and talk about music. So we compiled a list of all tags appearing for these 331 artists, and merged it with the Audioscrobbler list of "most frequent tags". The result is a list of 3026 tags. Thus, the terms allowed for describing the artists are now fixed. In the next step, for each artist we assign a weight to these terms, as described in the next section.

[7] allmusic.com

[8] www.audioscrobbler.net

4.2.1. *TF*×*IDF* Artist Analysis

To obtain *TF*×*IDF* for the artists, we apply a similar approach as in [3]. For each artist name, a search engine[9] is queried with the terms +"`artist name`" +*music* + *review*. The 100 top-ranked web pages are retrieved and stored locally. *TF*×*IDF* calculation is accomplished by regarding all web pages belonging to one artist as one single document (i.e., the texts of all web pages of the artist are concatenated). The resulting *TF*×*IDF* vector representing an artist assigns each term a specific weight, i.e., it is a characteristic artist profile.

k-NN	1-NN	5-NN	10-NN	20-NN
Accuracy	90.9%	90.1%	89.1%	87.8%

Tab. 1. Average k-NN leave one out classification accuracy when calculating artist similarities on *TF*×*IDF* vectors based on 2048 terms. Baseline: 40.9%. Classification accuracy is probabilistic, e.g., if 3 out of 5 closest neighbours have the same genre as the seed artist, this counts as $\frac{3}{5}$.

At this stage of the experiments, we conduct an intermediate test to estimate how well the artists are described by the calculated *TF*×*IDF* data. We assume that artists that are similar belong to the same genre. Thus, when comparing artists by calculating the cosine distance of their *TF*×*IDF* vectors, ideally the closest artists should belong to the same genre. This is quantified by a k-NN genre classification experiment. The average genre classification accuracy for $k = \{1, 5, 10, 20\}$ is given in Table 1. In our experimental setup, the obtained values of up to 90.9% give an indication that the artists are accurately described by the *TF*×*IDF* representation.

4.3. Extracting Concepts (Step 3)

We apply NMF to compress the 2048 terms that remain after cleaning[10]. When applying NMF, one can choose into how many factors r (i.e., bases) the data should be divided. We calculated NMF for $r = \{2, 4, 8, 16, 32, 40, 50\}$, each of these with 64 random initializations, and 100 iteration steps.

NMF yields a projection matrix that can be used to project the long *TF*×*IDF* representation of an artist down to few dimensions. In more detail, this means to multiply the artist's *TF*×*IDF* vector with the r (non-negative) basis vectors. Each

[9] we used Google.com

[10] I.e., removing terms that appear only for zero or one artist, and removing terms that appear for more than 99% of the artists. This is an automated procedure, no manual selection has taken place.

basis vector has the same length as the *TF*$\times$*IDF* vector, and can be thought of to represent a particular *topic* (cf. [11]), or “concept”.

In general, for deciding which model fits the data best, *information criteria* can be applied. In our experiments, we both used the Akaike Information Criterium (AIC) and the Bayesian Information Criterium (BIC). However, like other researchers before us, we found that these formulae do not well fit large models (as at hand), and the results are not of use. Thus, we proceed by manually selecting the model that appears to be best. It turns out that for $r = 8$, each of the found factors (or “concepts”) could clearly be assigned a genre label, except for one concept, which seemed to be both related to *Pop* and *Rap / Hip Hop*. Thus, $r = 8$ seems to be too small, and for further evaluation, we concentrate on $r = 16$.

4.4. Applying the projection (Step 4)

To confirm that instead of the 2048 dimensional *TF*$\times$*IDF* vector, the 16 dimensional concept vector found by NMF can be used to calculate artist similarity, we repeat the classification experiment presented in Section 4.2.1 using the 16 dimensional artist vectors. Again, for calculating the similarity of two vectors, we apply the cosine measure. Classification accuracy dropped only slightly, i.e. less than 2 percentage points (e.g., for 1-NN, the accuracy dropped from 90.9% to 89.2%).

After applying the projection, each artist is not represented by a long *TF*$\times$*IDF* vector any more. Instead, each artist is represented by a vector of lenght 16. Each of these 16 dimension is not a single word (as it is the case for the *TF*$\times$*IDF* vector), but rather is associated with a number of (weighted) words.

5. RESULTS

In Table 2, the most important terms belonging to each of the found categories are given. Each of the boxes contains the most dominant terms associated with the vectors used to project the high-dimensional *TF*$\times$*IDF*-vectors down to the 16 categories.

When looking at these categories, interesting observations can be made. Most of the categories are clearly related to a genre. We see this as an indication that genre categories are the dimensions that are best suited to describe musical artists.[11] Also, there is one category that contains mostly terms related to geographic entities. The fact that these are seen as a distinct category has no immediately obvious reason. One possible explanation is that “local flavour” or “geographic locality” is an additional attribute of artists. When using the browser, and setting all query concepts to zero except this, the top artist in the suggestion list is Lars Gullin. A look at how Lars Gullin is described on web pages reveals that even short descriptions of him stress the fact that he was an European saxophonist who never visited the United States, and that his impact would have been bigger if he had done so. Thus, one can indeed say that this is an artist with a strong “locality” attribute.

Although geographical regions may be an important aspect when characterizing an artist, in the way it is represented here it may not contribute to artist browsing. In such a case – i.e., if a found concept should not be considered for artist recommendation – it may be advisable to remove this concept by ignoring it during *TF*$\times$*IDF* computation. This is easily implemented by setting all values of the corresponding dimension to zero.

5.1. Assessing the Assignment of Artist to Categories

When several of the found categories are related to the same genre, they can be associated with different sub-genres. Most notably, this can be observed for the genre *Jazz*, which is the genre most artists are assigned to according to the All Music Guide data.[12] First, there is a category that could be best described by *Swing / 40s Jazz*. The most important terms of this category are *Jazz Vocals* and *40s*. Looking at the low-dimensional artist data created by NMF, we find that indeed artists like Glenn Miller, Jimmi and Tommy Dorsey, Sarah Vaughan, Louis Armstrong, and Ella Fizgerald have this category as their most dominant category.

Another category that is obviously strongly related to Jazz is the category we call *Bebop / Hard Bop / Free Jazz* category. The most important terms of this category are *Hard Bob*, *Blue Note* and *Free Jazz*. Well-known artists that have this as their most dominant category in the 16-dimensional representation are Art Blakey, McCoy Tyner, and John Coltrane.

Comparable results can be obtained for most of the other categories. Thus, from our usage of the system, we find that in the majority of the cases, the most dominant category of an artist is a category where one would intuitively expect the artist to be. Together with the results of the experiments presented in Section 4.4, we see this as a clear indication for the actual usability of the system.

5.2. Browsing Experience

When changing slider values, the user interface is quite responsive. In most cases, the order of suggested artist changes even for rather small slider changes. This is an important improvement over the first experimental version containing

[11] This may sound trivial. However, other categorizations such as instrumentation or mood also could have resulted.

[12] These genre labels were not used in this experiment.

all terms of the $TF \times IDF$ vector, where in most cases changing even several values did not have a visible effect. We see this as a final confirmation that the reduction to few (here: 16) concepts is a crucial step for the realisation of the system.

Finally, we want to give an example of how the suggested artists change when slider values are modified. When all slider values are set to zero, except the *Rap* category, the six top-ranked artists are Dr. Dre, LLCoolj, Boogie Down Productions, Slick Rick, Ice Cube and 2pac – as expected, predominantly Rap artists. When also moving the slider associated with the category best described as *Alternative* to the same extent as the *Rap* category slider, the top suggested artists gradually change to Rage Against the Machine, Eminem, Cypress Hill, Kid Rock, Beastie Boys and Outkast. These artists have Rap influences, but most of them also have strong Alternative aspects. Finally, when moving the *Rap* category slider towards zero, suggested artists change to such artists that have Alternative, but not Rap influences. The top-ranked artists are then the Melvins, Rammstein, Monstermagnet, Godsmack, the Deftones and Powerman 5000. During the described browsing procedure, the top-ranked artist changes nine times, i.e. the various slider mixtures produce nine different top-recommended artists, depending on their respective weight.

6. CONCLUSIONS AND FUTURE WORK

We have presented a system targeted to offer an intuitive way to discover new musical artists. The system automatically creates artist descriptions, and finds the categories that "best"[13] describe the concepts that are intrinsic factors for describing artists. At the end of the proposed procedure, each artist is represented as the amount he belongs to each of these categories. Based on this category representation, we build an interactive interface that allows the user to create his own "ideal" artist, for which the best-matching artists are displayed.

Our experiments show that the categories found by the algorithm are closely related to genres[14]. Although a genre-related categorization may be well-suited for a system as the one we proposed, in the future we will investigate if it is possible to find other categories more immediately related to human moods, such as *sad*, *happy*, *relaxed*. This may be accomplished by only allowing for terms that describe human moods, but such terms in a great variety.

[13] regarding the output of our algorithms, and as far as our experiments show, also in a semantic sense

[14] Note that the associations are not binary, i.e., each artist is assigned to several categories, which makes this representation favourable over a strict genre categorization.

7. ACKNOWLEDGEMENTS

This research is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) under project number L112-N04, and by the EU 6th FP project S2S2 ("Sound to Sense, Sense to Sound", IST-2004-03773). The Austrian Research Institute for Artificial Intelligence also acknowledges financial support by the Austrian Federal Ministries BMBWK and BMVIT. The authors would like to thank the reviewers for their valuable comments.

8. REFERENCES

[1] B. Whitman and S. Lawrence, "Inferring Descriptions and Similarity for Music from Community Metadata," in *Proc. Intern. Computer Music Conference*, 2002.

[2] S. Baumann and O. Hummel, "Using Cultural Metadata for Artist Recommendations," in *Proc. Conf. on Web Delivering of Music (Wedelmusic2003)*, 2003.

[3] P. Knees, E. Pampalk, and G. Widmer, "Artist classification with web-based data," in *Proc. Int. Symp. on Music Information Retrieval (ISMIR'04)*, 2004.

[4] S. Baumann, *Artificial Listening Systems: Modellierung und Approximation der subjektiven Perzeption von Musikähnlichkeit*, Ph.D. thesis, Technical University of Kaiserslautern, 2005.

[5] M. Schedl, P. Knees, and G. Widmer, "A web-based approach to assessing artist similarity using co-occurrences," in *Proc. Fourth Int. Workshop on Content-Based Multimedia Indexing (CBMI'05)*, Riga, Latvia, 2005.

[6] Elias Papalk, Arthur Flexer, and Gerhard Widmer, "Hierarchical organization and description of music collections at the artist level," in *Proceedings of the 9^{th} European Conference on Digital Libraries*, Vienna, Austria, 2005.

[7] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer, "Automatically describing music on a map," in *Proc. 1st Workshop on Learning the Semantics of Audio Signals (LSAS) 2006*, Athens, Greece, December 6 2006.

[8] P. Cano, O. Celma, M. Koppenberger, and J. Martin-Buldú, "Topology of music recommendation networks," in *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2006.

[9] G. Geleijnse and J. Korst, "Web-based artist categorization," in *Proc. Int. Symp. on Music Information Retrieval (ISMIR'06)*, 2006.

[10] S. Vembu and S. Baumann, “Separation of vocals from polyphonic audio recordings,” in *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR’05)*, 2005.

[11] Wei Xu, Xin Liu, and Yihong Gong, “Document clustering based on non-negative matrix factorization,” in *Proc. SIGIR03*, Toronto, Canada, July 28August 1 2003.

[12] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, pp. 788791, 1999.

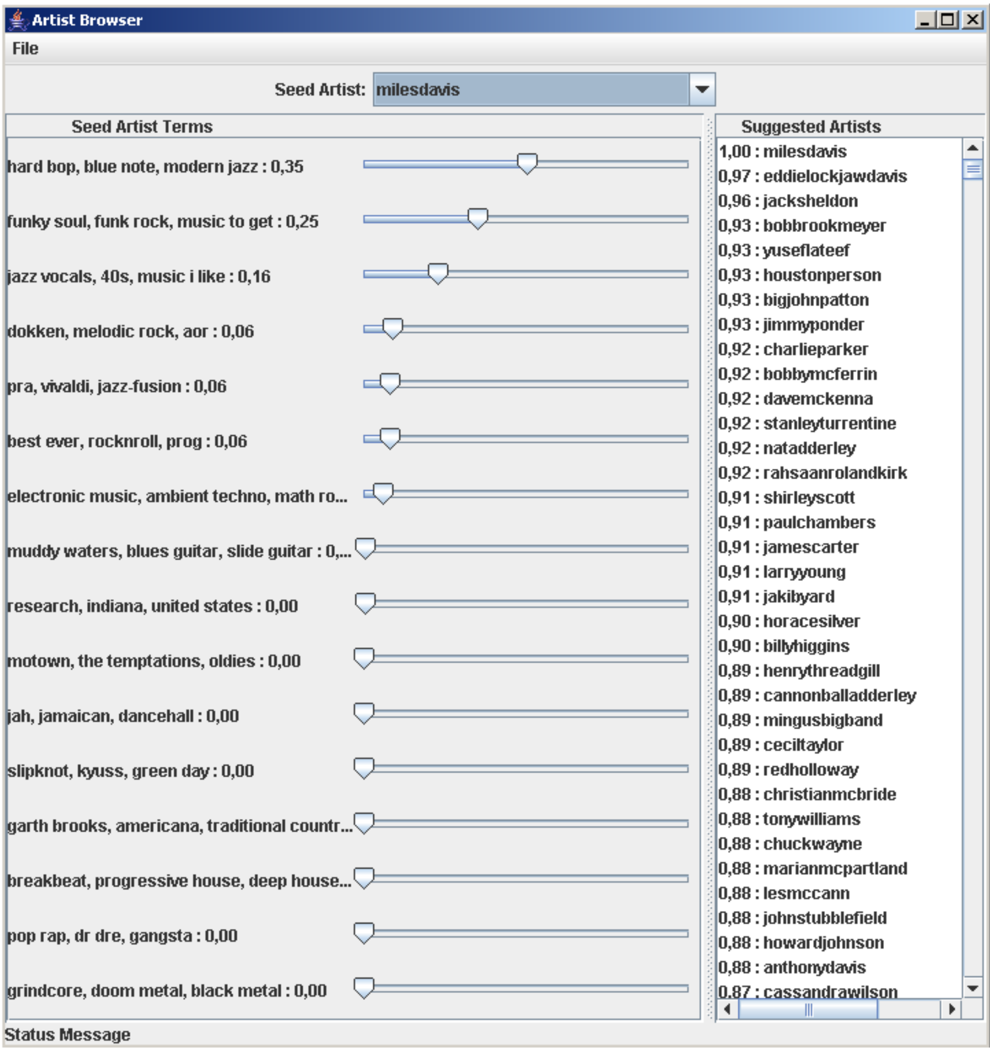

Fig. 1. Demonstration of the artist browser. Internally, each artist is represented as a 16-dimensional vector giving the amount this artist is related to each of the 16 categories given on the left panel. When the user selects a seed artist (via the dropdown list located at the top), this artist's representation is transferred to the sliders. Here, the representation of Miles Davis is shown. At the right, the artists most similar to the current slider positions are listed. When the user modifies the slider positions, the list of most similar artists gets updated.

jazz vocals	garth brooks	breakbeat	grindcore	funky soul	muddy waters
40s	americana	progressive house	doom metal	funk rock	blues guitar
music i like	traditional country	deep house	black metal	music to get	slide guitar
traditional pop	alan jackson	fatboy slim	century media	funkadelic	delta blues
jazz vocal	johnny cash	dnb	death metal	favorite artist	electric blues
jump blues	bluegrass	progressive trance	speed metal	funk	classic blues
30s	alt country	drum n bass	sepultura	psychedelic soul	mississippi

dokken	motown	hard bop	jah	best ever	electronic music
melodic rock	the temptations	blue note	jamaican	rocknroll	ambient techno
aor	oldies	modern jazz	dancehall	prog	math rock
deep purple	classic soul	free jazz	jamaica	progressive rock	ambient
dio	stax	tenor sax	rasta	dance pop	krautrock
hair metal	soul artists	post-bop	rocksteady	frank zappa	synthpop
whitesnake	northern soul	the jazz	ragga	70s	post-rock

research	slipknot	pop rap	pra
indiana	kyuss	dr dre	vivaldi
united states	green day	gangsta	jazz-fusion
massachusetts	pop punk	rappers	peter white
european	audioslave	gangsta rap	soothing
scandinavia	rock alternative	def jam	larry carlton
sweden	punk rock	old school rap	spyro gyra

Tab. 2. The 16 categories that were found by the NMF decomposition of the *TF*$\times$*IDF* vectors. For each category the 7 terms with the largest weight are given.