Visually Profiling Radio Stations

Thomas Lidy **Andreas Rauber**
Vienna University of Technology
Department of Software Technology and Interactive Systems
Favoritenstrasse 9-11/188, A-1040 Vienna, Austria
lidy,rauber@ifs.tuwien.ac.at

Abstract

The overwhelming number of radio stations, both online and over the air, makes the choice of an appropriate program difficult. By profiling the program content of radio stations using Self-Organizing Maps we provide a reflection of a station's program type and give potential listeners a visual clue for selecting radio stations. Profiles of current broadcasts indicate which program type a station is currently playing. By creating radio station maps it is possible to directly pick a specific program type instead of having to search for a suitable radio station.

Keywords: radio stations, online streams, broadcast, audio feature extraction, genre discrimination, profiles, Self-Organizing Map

1. Introduction

Music information retrieval is a fast growing research area. To date many issues for intelligent retrieval of music have been addressed. Music can be retrieved by similarity, in terms of melody, rhythm, pitch, brightness, etc. Artists similar to a given one can be found either by acoustic analysis or data mining on the web, or also combined approaches. Songs can be retrieved by humming, by audio examples, etc. Many more techniques have been presented in recent years, and many of them deal with efficient retrieval of pieces of music from large repositories.

Yet, there is a plethora of audio streams emerging: new on-line radio stations are created, and the number of stations using traditional broadcast over the air is growing as well. Also, the trend to podcasts increases the problem of "channel overflow". Choosing a station with a program that matches one's personal taste thus is not an easy task, a situation that one might know if one travels to another country. Only from a station's name (e.g. "Soundportal") it is very hard to guess what music it is playing. While listening a while to one or another station and switching through several stations might be the traditional solution, this will hardly be feasible given the multitude of new stations. The European Broadcasting Union has addressed this problem within the Radio Data System (RDS; RDBS in the USA) [1], which provides additional information to a traditional broadcast. Besides broadcasting the station name and other textual information, automatically switching to traffic broadcast or alternative frequencies, it also broadcasts meta-data about the program type, allowing one to pick radio stations playing classical music or news. However, this information is hardly used by the consumers, and also many stations do not even broadcast this meta-information.

In this paper we present a method for profiling radio stations, both traditional ones and online radios, which assists in solving this problem by letting the computer index the radio stations and create profiles, or "fingerprints", of them. With these profiles it is possible to get a quick impression of a station's focus. We use the technique of Self-Organizing Maps to organize the program coverage of radio stations on a two-dimensional map. From the radio station map we derive visual profiles of specific radio stations by considering their distribution within specific areas of the musical horizon the music map spans. This approach allows to profile the complete program of a radio station, getting an overview of the station's program, or to view an activity profile of e.g. the last 10 minutes of the broadcast, in order to see where the station is currently active. Moreover, the "now playing" function can show immediately the kind of music a station is playing by marking the location on the radio station map. This enables completely novel access to the plethora of radio stations, allowing to select always the program or kind of music one likes instead of having to have a favorite radio station.

In Section 2 we discuss related work. Section 3 presents the method of profiling radio stations. Section 4 discusses different application scenarios for our technique. Section 5 gives a summary and outlines future work.

2. Related Work

The two techniques underneath our profiling approach are audio descriptors for genre classification and Self-Organizing Maps.

Feature extraction for genre discrimination is a research area that has seen tremendous attention within the last years. While a state-of-the-art comparison is done in the annual

MIREX evaluation [2], a broad overview upon existing audio descriptors for music classification is given in [3, 4]. In this work we employ Rhythm Histograms and Statistical Spectrum Descriptors [5], both because they performed very well in the MIREX2005 evaluation [6] and because they are sufficiently efficient for our approach. A feature set for genre discrimination seems to be well-suited to distinguish different radio stations, nevertheless, we will investigate the use of other feature sets for this novel kind of application.

A Self-Organizing Map (SOM) is an unsupervised neural network providing a topology-preserving mapping from a high-dimensional input space onto a two-dimensional output space [7]. The input data, in our case the audio features, are arranged on a two-dimensional grid whose units are iteratively activated, and thus "trained" by the input features. This results in an organization where similar data is mapped close to each other, building clusters.

The earliest works that use Self-Organizing Maps to organize sounds, based on pitch, duration and loudness, date back to [8, 9]. In [10] MFCCs are used for retrieval of sound events from a SOM.

Automatic organization of music collections on SOMs has been first demonstrated in [11], and later in [12, 13, 14]. Another work on exploring music collections [15] uses Aligned-SOMs, which allow for interactively changing the focus of organization among different aspects, like e.g. timbre or rhythm. In [16, 17] SOMs are applied to organize music at the artist level using artist information mined from the web.

In [18] so-called Emergent SOMs are employed for visualization of music collections which are particularly suitable for creating large maps. The usage of very large maps, however, is not appropriate in our context, as this would possibly result in a too detailed discrimination of audio features, discriminating parts of songs rather than radio stations.

Another interface to audio collections has been presented with MARSYAS3D [19], a framework that contains, among other visualizations, a 3D space with audio files mapped according to timbre. Among audio from different sources also clips recorded from FM radio have been visualized in MARSYAS3D. In [20] applications of the FastMap algorithm for visualizing audio similarity and improving browsing of music archives are discussed. Torrens et. al [21] present new interfaces for exploring personal music libraries in form of disc- and tree-map-based visualizations.

In [22] a semantic web application is described, that indexes (mines) websites and retrieves collections of data useful for a variety of applications. One of them is "Now Playing!", which shows the current song playing on a number of radio stations, based on the information found on the radio stations' websites. With our radio station profiling technique we can provide a "Now Playing" function based on the content of the audio stream.

3. Generating Radio Station Profiles

The program content of a selected set of radio stations is organized on a two-dimensional Self-Organizing Map, according to similarity or dissimilarity of audio that is played. To achieve this, we extract features from the audio signal received from the radio station, in order to be able to recognize the content of a radio station's program. Research on Music Information Retrieval has generated a wealth of descriptors (feature sets) for computing audio similarity. For profiling radio stations we apply Rhythm Histograms and Statistical Spectrum Descriptors [5]. The feature extraction algorithm tries to replicate the human perception by incorporating psycho-acoustics, and thus should be close to what a human listener perceives while listening to radio stations. It then computes descriptors containing information about the audio spectrum and rhythmics contained in it. One feature vector is calculated for every 6 seconds received from a radio station's broadcast.

A Self-Organizing Map (SOM) is an unsupervised neural network that maps a high-dimensional feature space onto a two-dimensional output space, in our case a rectangular map. The neural network is trained by the feature vectors extracted from the audio. We collect an appropriate number of feature vectors (between 1000 and 4000) from each radio station and put them together as the training database for the Self-Organizing Map. No information about which feature vector belongs to which radio station is provided to the SOM algorithm. We obtain a map organized by acoustic similarity, where each feature vector has been mapped to the map unit best representing it.

A radio station's profile is derived from the organization of its program content on the two-dimensional map. Nearly every radio station serves specific user groups and thus has a focus on a specific type of program, being classical music, pop music, rock music, report-based, etc. As a consequence, the radio station will also have a focus on a specific region of the music map. The more focused a radio station is, the more concentrated it appears on the map. We compute the frequency each radio station hits each unit on the map. The graphical representation of this "hit histogram" reveals the areas in which a radio station is active. The visual profile is enhanced by taking the logarithm of the hit histogram, in order to reduce elevated concentrations, and by subsequent smoothing. Such a representation of a radio station's program recorded for several hours contains a very good reflection of the station's focused music area. Many radio stations build clearly visible clusters, which immediately give an idea of the radio station's concentration on a certain music style (or show whether the station is mainly speech-based). By comparing the visual profiles of several stations, the user gets an immediate overview on the coverage of a radio station's program, and if he or she will like it or not.

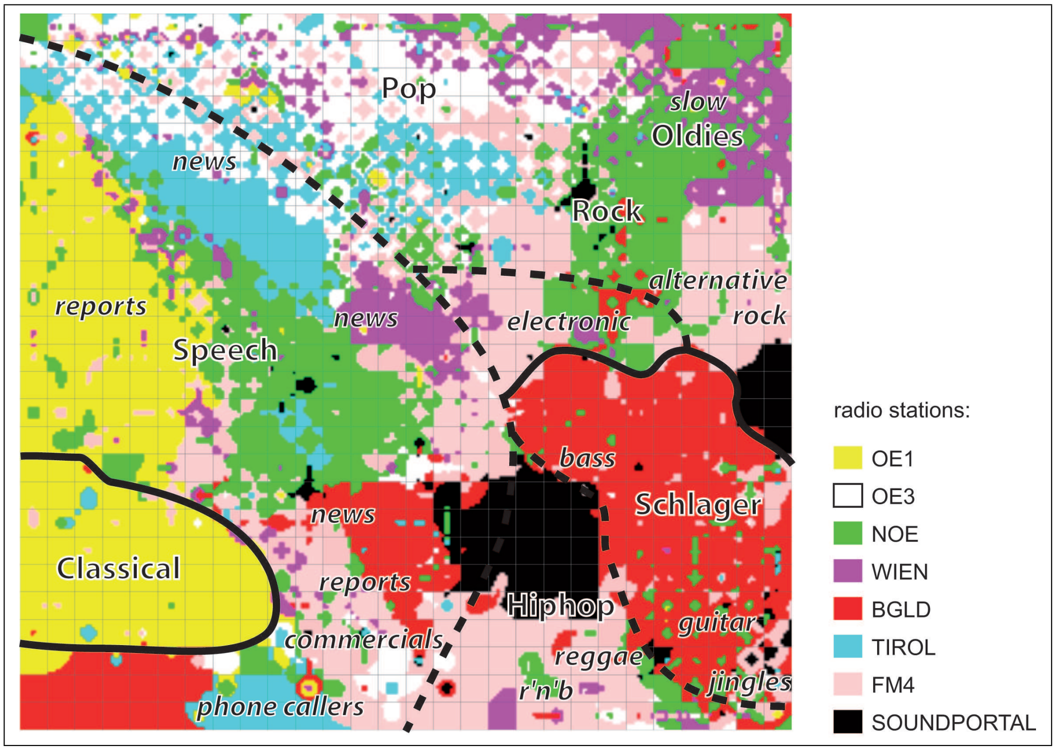

Figure 1. Overview of the radio station map containing 18,600 segments: the colored tessellation shows the distribution of the 8 radio stations, the labels are explaining the acoustic organization (independently from radio station assignments); dashed lines indicate smooth transitions between the genres.

If we reassign a station's name as class label to each feature vector, we can obtain a colored visualization, which shows the distribution of the radio stations on the map (c.f. Figure 1). Through acoustically exploring the map, one can find out rather quickly the locations of aggregations of a specific type of music. Thus, it is possible to explore the coverage, or "horizon", of a radio station's program.

4. Application Scenarios

We recorded the program of 8 Austrian radio stations and extracted audio features of every 6-second segment of the recorded audio. We recorded a total of 31 hours program and consequently extracted about 18,600 feature vectors. The radio station map (or RadioSOM) we created contains 28x26 units, which means that each unit on average contains 25 segments (which is about 2.5 minutes of program). We did not want to create a larger map in this case, as this would result in a much more detailed discrimination of audio features, which means that a single song (usually roughly 4 minutes long) would be discriminated and distributed to a much larger extent.

4.1. The RadioSOM

The RadioSOM we are presenting in Figure 1 shows the musical horizon of 8 Austrian radio stations. There are radio stations that are clearly represented by two or three clusters, while others are more interweaved. The former are those with a strong focus of their program, e.g. on news reports, classical music or rock music. The latter kind of radio stations are mostly those broadcasting "pop" music, a "genre", which is in fact not a genre by itself, but represents the broad spectrum of popular music from different decades. On the one hand we still can make out differences in the areas covered by "pop" music. On the other hand, most people whose favorite music is "pop" music are used to the broad variety that it represents.

In order to provide a more detailed description of the radio station map given in Figure 1 we have manually labeled the image with areas of clearly distinctive sound characteristics. The image shows a Voronoi-like tessellation of the RadioSOM color-coding each radio station, in order to give an overview of the distribution of the 8 radio stations. Evidently, the goal is not to get a uniform class (color) separation in this application, as many of the stations overlap in their program (clearly visible from Figure 1). Note that in contrast to this figure the visual profiles given in Figure 2 also include the frequency of mappings per station to a unit.

A rather large part of the map consists of recordings containing speech, which is not really surprising, as all of the radio stations broadcast news, and some of them additionally bring reports, sports, telephone games, and so on. Among the radio stations with a clearly distinctive area is OE1, whose program consists of about 50 % classical music and 50 % spoken contributions, reports and news. This

(a) Radio OE1: Classical music and speech. (b) Radio OE3: pop songs. (c) Radio Soundportal: hip hop and rock.

Figure 2. Profiles of three of the eight stations on the RadioSOM.

Figure 3. Profile of a yet unknown radio station, Radio Stephansdom, mapped on the RadioSOM.

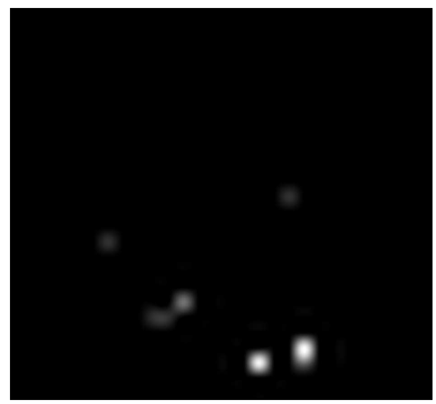

Figure 4. “Now Playing” profile of 10 minutes of radio FM4. The profile shows activity in the hiphop/reggae/r’n’b area.

is clearly reflected on the map with one out of two clusters containing all recordings with speech and the other one containing classical music. Note that the classical cluster is actually located within the huge speech cluster, nevertheless it is clearly separated by strong borders (no smooth transitions).

The large area labeled “speech” is covered effectively by all radio stations. The reason why the stations are rather well separated and not completely mixed up is that with the features extracted from the audio signal the computer is able to distinguish the voice of different speakers. Thus, the various sub-areas of the speech cluster contain different pitches of voices, and besides distinguishing male and female speakers, the computer also finds differences in the type of spoken program: news, reports, weather, telephone calls, and others. Commercials are located alongside the borderline between speech and music as the majority of them contain both speech and music. A cluster with a particularly large number of commercials has been labeled on the map in Figure 1.

The pop cluster is covered by 5 of the radio stations, predominantly by OE3, but also by radio TIROL, partly by FM4, by radio NOE and radio WIEN, with smooth transitions both towards the rock and “slow oldies” cluster. Two of our recorded stations - radio NOE and WIEN - have a strong focus on oldies. “Schlager” music is an Austrian / German specialty: a kind of simply structured pop music with a catchy harmonic melody and modest humorous or sentimental lyrics [1]. Many Austrian regional radio stations like to broadcast this kind of music, and radio BGLD seems to focus fully on it. Also radio NOE plays this kind of music, but tends more to oldies, similar to radio TIROL, whose music we find in the middle of pop, rock and oldies. Radio FM4 defines itself to be the radio station for “alternative music” and combines elements from alternative rock, electronics and also hip hop and reggae music. Radio Soundportal has two clearly visible areas: hip hop (plus a little electronic music) and rock (more specifically “new metal”).

Music maps have been previously presented mainly with the intention to organize music archives (see Section 2). The RadioSOM is facing some new challenges, one of them being the high number of situations where speech plays a role. We have reports with moderately spoken text, while

[1] see `http://[de/en].wikipedia.org/wiki/Schlager`

news are usually spoken much faster. There is text accompanied by background music as in commercials, program announcements, etc. There is even a cluster with jingles on its own. There are phone games and interviews with phone callers, which means that the frequency bandwidth of the signal is narrowed under these circumstances. The RadioSOM also distinguishes segments with transitions in the flow of the program, i.e. a piece of music ends and the station's speaker starts to talk, or vice versa, and has aggregated them in the lower right corner. This is especially convenient when using the "Sticking To Favorite Music" feature described in sub-section 4.5, because the separation of those segments enables avoiding discontinuities.

4.2. Obtaining Radio Station Profiles

As described in Section 3 we compute hit histograms as the basis for a radio station's profile. Based on the RadioSOM described in Section 4.1 we now describe the visual profiles of three of the stations, depicted in Figure 2. The profile of radio OE1 shows that this station is almost equally active in two areas: The upper cluster belongs to the speech area of the map while the lower cluster represents classical music.

Radio OE3 is *the* pop station among our radio stations and its profile is thus almost entirely active in the upper (rather left) area. A few outlying peaks show that OE3 makes a few exceptions like playing for instance a hip hop song. The small cluster on the lower center corresponds to a sports report.

The profile of radio Soundportal reflects very nice its two program focuses: The cluster on the right border belongs to the rock area, the other cluster is the center of the hip hop area.

In this example we have shown three very distinctive profiles. In the next sub-section we will see that similar radio stations will result in similar profiles.

4.3. Profiling Unknown Radio Stations

Let us suppose we find out about a new radio station, but do not yet know what kind of program it is broadcasting. With our technique we can record the radio station for a while and map its characteristics on the already existing radio station map. In this application scenario we do not train a new RadioSOM, rather we want to see which kind of programs of the spectrum given by the map we already know is covered by the new station. Put in short, we want to find out the focus(es) of the new station.

We recorded 3 hours from Radio Stephansdom, segmented the recordings and extracted audio features, as we did with the other radio stations before. Now, for each segment, we find the optimal mapping on the RadioSOM, i.e. the unit with the closest distance to the segment's feature vector. Doing this with the 1800 segments we extracted from Radio Stephansdom, we get the profile visualized in Figure 3. Immediately we recognize a strong overlap between this profile and the one of Radio OE1, depicted in Figure 2(a). Indeed, Radio Stephansdom also consists mainly of classical music and spoken contributions, both reflected in the visual profile.

Applying this technique one can get a good impression of previously unknown radio stations.

4.4. Now Playing

Profiling only the last e.g. 10 minutes of a radio station shows us the area of the RadioSOM where a certain station has been recently active, and thus indicates which program type the station is currently broadcasting. If we reduce the time recorded for creating the profile to 3 minutes or even one single 6 second segment we get kind of a "Now Playing" function, showing exactly what a station is broadcasting at the moment. This would constitute an audio-based equivalent to the "Now Playing" feature offered by semantic web applications monitoring and integrating the current program website of several radio stations [22]. Figure 4 shows a "fingerprint" of what FM4 broadcast within the 10 minutes on 2006-02-13, between 20:08 and 20:18 (CET). This activity profile reveals that the station was mainly active in the hiphop/reggae/r'n'b area, with a few speech sections in between. If this feature is performed in real-time the "Now Playing" function creates an animation-like live-view of the radio station's current type of program.

4.5. Sticking To Favorite Music

We can use the RadioSOM to narrow the music or program that we want to listen to by selecting a specific area of the map and telling the computer to play only stations that are currently in the focus of the selected area. This allows a completely novel access to choosing radio stations, always listening to the favorite type of program or music instead of having to search for a radio station. Furthermore, with this technique we can not only limit the scope of music to listen to, it also may enable functions like skipping commercials and other parts of the program that one dislikes, by switching to another program with similar music. The "Sticking To Favorite Music" feature is not yet implemented as an automatic real-time feature in our application, but is one of the main features to be integrated next.

5. Conclusions

We presented a method for creating profiles of radio stations as well as creating radio station maps, which give an overview of the horizon of a group of radio stations. The visual profiles allow quick discovery of the program diversity of new/unknown radio stations. Moreover, this method allows for identification of the program that is *currently* broadcast on a radio station and provides a novel way of selecting radio stations according to one's preferences. The approach is useful for discovering new radio stations for one's personal taste, both online and traditional ones.

We will investigate the selection of appropriate features to be extracted from radio stations improving the discrimination of program types of different radio stations and addressing a better distinction of the large number of stations broadcasting "pop" music.

We would like to integrate on-the-fly feature extraction and visualization into one application which then supports the "now playing" function in real-time and allows ad-hoc selection of a favorite program area having the computer to select automatically the station that best covers one's personal taste.

6. Acknowledgments

Part of this work was supported by the European Union in the 6. Framework Program, IST, through the MUSCLE NoE on Multimedia Understanding through Semantics, Computation and Learning, contract 507752.

References

[1] D. Kopitz and B. Marks, *RDS: Radio Data System*. Artech House, November 30 1998.

[2] "Annual Music Information Retrieval Evaluation eXchange (MIREX)," Website, 2005, http://www.music-ir.org/mirexwiki/index.php/Main_Page.

[3] J. Downie, *Annual Review of Information Science and Technology*. Medford, NJ: Information Today, 2003, vol. 37, ch. Music information retrieval, pp. 295–340.

[4] R. Typke, F. Wiering, and R. C. Veltkamp, "A survey of music information retrieval systems," in *Proceedings of 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11-15 2005.

[5] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11-15 2005, pp. 34–41.

[6] "Music Information Retrieval Evaluation eXchange - audio genre classification," Website, 2005, http://www.music-ir.org/evaluation/mirex-results/audio-genre/index.html.

[7] T. Kohonen, *Self-Organizing Maps*, 3rd ed., ser. Springer Series in Information Sciences. Berlin: Springer, 2001, vol. 30.

[8] P. Cosi, G. D. Poli, and G. Lauzzana, "Auditory modelling and self-organizing neural networks for timbre classification," *Journal of New Music Research*, vol. 23, no. 1, pp. 71–98, March 1994.

[9] B. Feiten and S. Günzel, "Automatic indexing of a sound database using self-organizing neural nets," *Computer Music Journal*, vol. 18, no. 3, pp. 53–65, 1994.

[10] C. Spevak and R. Polfreman, "Sound spotting - a frame-based approach," in *Proceedings of the 2nd International Symposium on Music Information Retrieval (ISMIR 2001)*. Indiana University, Bloomington, IN, USA: Indiana University, October 15-17 2001, pp. 35–36. [Online]. Available: http://ismir2001.indiana.edu/posters/spevak.pdf

[11] A. Rauber and M. Frühwirth, "Automatically analyzing and organizing music archives," in *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, Darmstadt, Germany, September 4-8 2001.

[12] A. Rauber, E. Pampalk, and D. Merkl, "Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by musical styles," in *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, Paris, France, October 13-17 2002, pp. 71–80.

[13] E. Pampalk, A. Rauber, and D. Merkl, "Content-based organization and visualization of music archives," in *Proceedings of ACM Multimedia 2002*. Juan-les-Pins, France: ACM, December 1-6 2002, pp. 570–579.

[14] R. Neumayer, M. Dittenbach, and A. Rauber, "Playsom and pocketsomplayer, alternative interfaces to large music collections," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11-15 2005, pp. 618–623.

[15] E. Pampalk, S. Dixon, and G. Widmer, "Exploring music collections by browsing different views," *Computer Music Journal*, vol. 28, no. 2, pp. 49–62, 2004.

[16] P. Knees, E. Pampalk, and G. Widmer, "Artist classifiction with web-based data," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, October 10-14 2004.

[17] E. Pampalk, A. Flexer, and G. Widmer, "Hierarchical organization and description of music collections at the artist level," in *Proceedings of the 9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2005)*, Vienna, Austria, September 18-23 2005.

[18] F. Mörchen, A. Ultsch, M. Nöcker, and C. Stamm, "Databionic visualization of music collections according to perceptual distance," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 11-15 2005.

[19] G. Tzanetakis and P. Cook, "MARSYAS3D: A prototype audio browser-editor using a large-sclae immersive visual and audio display," in *Proceedings of the International Conference on Auditory Display (ICAD 2001)*, Espoo, Finland, July 29 - August 1 2001.

[20] P. Cano, M. Kaltenbrunner, F. Gouyon, and E. Battle, "On the use of fastmap for audio retrieval and browsing," in *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, Paris, France, 2002.

[21] M. Torrens, P. Hertzog, and J. L. Arcos, "Visualizing and exploring personal music libraries," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, October 2004.

[22] G. Gottlob, C. Koch, R. Baumgartner, M. Herzog, and S. Flesca, "The lixto data extraction project - back and forth between theory and practice." in *Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS'04)*. New York, NY, USA: ACM Press, 2004, pp. 1–12.