Making Large Music Collections Accessible using Enhanced Metadata and Lightweight Visualizations

Florian Kleedorfer, Ulf Harr, Brigitte Krenn
Research Studios Austria
Studio Smart Agent Technolgies
Hasnerstrasse 123, 1060 Vienna, Austria
{florian.kleedorfer, ulf.harr, brigitte.krenn}@researchstudio.at

Abstract

We present stategies for improving the accessibility of large music archives intended for use in commercial environments like music download platforms. Our approach is based on metadata enhancement and on the augmentation of traditional browsing interfaces with concise data visualizations. New metadata attributes are created by recombination and summarization of existing ones as well as through web-based data mining.

1 Introduction

Today's commercial online music portals comprise large bodies of songs (e.g. 3+ millions at Napster and 3.5 millions at iTunes)[1]. In order to make them accessible to the user/customer, intelligent methods are required. One such means are music recommenders which provide the customer with suggestions about songs that she might be interested to listen to (and ideally to buy). Recommendations are based on metadata describing the artists and songs (cf. item-based filtering, [1]) and/or on information derived from the behaviour (which songs/artists are viewed, positively/negatively rated, bought, etc.) users frequenting the portal (cf. collaborative filtering, [2] [3] or hybrid combinations [4] [5]).

Another approach are visualization techniques that make the data available in the music collection visible in the user interface. Up to date, visualization of relations between tracks or artists is not yet an issue in current music portals. See for instance Pandora[2], Last FM[3], MyStrands[4], Music IP[5], with the latter featuring the most innovative representation of similar tracks in a spiral that intuitively suggests distance to the centered item. In commercial music portals typically only one dimension, if at all, is used for visualization.

In the research area of Music Information Retrieval, audio based feature extraction has mainly been used as a basis for organizing music collections. Pampalk et al. arrange songs on a self-organizing map (SOM) according to audio similarity [6]. In this visualization the whole music collection can be overlooked at a glance. Aligned SOMs have been proposed as an enhancement [7], allowing users to influence the weights of different audio features. Similarly, in the open source tool MusicMiner, Mörchen et al. allow for exploring a music collection using SOMs as well as traditional tree or list based controls [8]. Knees et al. create a 3D view based on the SOM [9] that allows free navigation using a game controller and plays back the music that corresponds to the current location in the music landscape. The music landscape is enriched with metadata obtained from the Internet. Based on audio similarity of songs, Pampalk and Goto arrange artists on a circle which can be navigated along; different regions of the circle are annotated with music related words extracted from the Internet [10]. Berenzweig et al. present a Web interface[6] for their collection, allowing to browse along metadata relations (artist, track, album) as well as based on relations computed from the audio signal [11].

Typical visualization techniques from the field of music information retrieval are hard to integrate in commercial applications. On the one hand, this is due to the multidimensionality of the shown data. The approaches that attempt to make such data accessible introduce new interaction paradigms that users still have to get accustomed to. On the other hand, creating and maintaining the visualizations are complex technical processes that consume much time in programming and computation and thus constrain

[1] `http://www.napster.com`, `http://www.itunes.com`
[2] `http://www.pandora.com`
[3] `http://www.last.fm`
[4] `http://www.mystrands.com`
[5] `http://www.musicip.com`
[6] `http://www.playola.org`

the applicability of these methods.

In this context, our goal was to design and implement what we term *lightweight visualization strategies* for browsing today's large music collections. The application we present adopts navigational structures that are well-established, such as views on artists and genres, as they are still most commonly employed in large music portals. The different views are enhanced with data visualizations of low dimensionality, designed for intuitive use and easy integration in a Web application. To create the visualizations, we make use of the standard metadata that are delivered by music content providers. Additional attributes are generated on the basis of this data. Others are generated making use of information extracted from the Internet.

The remainder of this paper is organized in two sections. In Section 2, we present our data sources and introduce methods for the generation of additional metadata based on the metadata available in the source data. The resulting data attributes form the basis for the visualizations proposed in Section 3. There we describe the mapping between data attributes and visual attributes and propose a selection of methods for visually combining different types of information such as the relatedness between artists, genre affiliation of artists, prototypicality of artists for particular genres, etc.

2. The Data

2.1. The Music Collection: Primary and Secondary Metadata

We work with a music archive that is a subset of the Napster[7] database. This subset comprises the approx. 60.000 most popular (= most frequently accessed) audio tracks from about 6.000 artists and comes for each track with a set of metadata: artist[8], album, duration, explicit lyrics information, label, date of creation, release date, and one or more genres.

While primary metadata in our terminology refer to the meta information delivered by the music providers, secondary metadata are computed from the primary metadata and from information extracted from the Internet. This way we gain a richer set of metadata which provides a valuable basis for visualization of the content of music archives.

2.2. Affiliation of Artists to Genres

Based on the genre classification of tracks and the association of artists to tracks, we compute values for the genre affiliation of artists. In the following, we briefly present the approaches pursued.

[7] `http://www.napster.com`

[8] Our data do not distinguish between individual artist and group.

Definitions used in the formalizations: Let A be the set of all artists, T the set of all tracks, G the set of all genres. Let T^a be the set of all tracks assigned to artist $a \in A$ and $\tau : T \mapsto \{0,1\}^{|G|}$ be a function computing the binary genre affiliation vector of length $|G|$ that defines which genres a track is assigned to. $\tau_g(t)$ refers to the component of the affilation vector computed by $\tau(t)$ that indicates affiliation to genre $g \in G$. The goal is to find a function $\gamma : A \mapsto [0,1]^{|G|}$ that outputs the affiliation vector for any artist a.

We considered the following approaches:

- *binary*:

$$\gamma_g^b(a) = sgn\left(\sum_{t \in T^a} \tau_g(t)\right) \tag{1}$$

  This approach results in the artist being assigned to a genre if at least one of her tracks is.

- *frequency based*

$$\gamma_g^f(a) = \frac{\sum_{t \in T^a} \tau_g(t)}{|T^a|} \tag{2}$$

  Here, we compute an affiliation vector where each component expresses the relative frequency of the respective genre in the artist's tracks. Hence, the affiliation value can be interpreted as a probability for an unknown track of a given artist to be associated with a given genre.

- *frequency based normalized*

$$\overline{\gamma}_g^f(a) = \frac{\sum_{t \in T^a} \tau_g(t)}{\sum_{h \in G} \sum_{t \in T^a} \tau_h(t)} \tag{3}$$

  This approach is similar to the frequency based one, except that the resulting vector is normalized so that the sum of its components is 1.

The binary method is a simple way of providing genre information for an artist by summarizing the genre information of the artist's tracks, but it lacks weighting of the genres. This is provided by the frequency based method and its normalized variant. The difference between the latter two approaches is the point of view from which the tracks of an artist and their genres are looked at. In the frequency based method genres are weighted independently of each other whereas in the normalized model, genre affiliation expresses the importance of the genre relative to all other genres in the music collection.

2.3. Genre Purity

The normalized variant of the frequency based method is well suited for the computation of genre purity, with a high value indicating that the tracks in the genre are associated with few or no other genres, and it can be interpreted as a measure of the genre's discriminating power. Genre purity is derived by averaging the normalized affilation values for the genre that are greater than zero:

$$\pi(g) = \frac{\sum_{a \in A} \overline{\gamma}_g^f(a)}{|\{a, a \in A : \overline{\gamma}_g^f(a) > 0\}|} \tag{4}$$

2.4. Genre Similarity

Based on the genre affiliation vector $\gamma(a)$[9] of an artist a, a similarity measure for artists can be defined as the cosine of their genre affiliation vectors, i.e.

$$\sigma(a, b) = \frac{\gamma(a).\gamma(b)}{\|\gamma(a)\|\|\gamma(b)\|} \tag{5}$$

for two artists a and b, where $x.y$ denotes the dot product of two vectors x and y and $\|x\|$ denotes the euclidian norm of vector x. Small σ values are interpreted as high similarities.

2.5. Metadata Obtained from the Internet

We use Web-based co-occurrence analysis to compute additional metadata for the tracks in our collection. The approach we use has been developed by Schedl et al.[12], in particular we use variant B, *BL/FL ratio with popularity penalization. Artist relatedness* is calculated based on co-occurrences of artists in Web documents. This measure of relatedness is asymmetric, so that for any artist a in our collection, two ordered lists of artists are obtained. The one list contains artists that are related to a, and the other one artists to whom a is related. In the context of our visualization techniques we call the first type *outgoing references* and the second *incoming references*.

In connection with the genre classification of an artist, a value indicating the artist's *prototypicality* with respect to each of her genres is derived from the co-occurrence matrix. Furthermore, an *overall prototypicality* of artists relative to the whole music collection is computed.

3 Visualization

The approaches for metadata computation presented in the previous section provide the data attributes for the visualizations described below.

3.1. Visual Attribute Mapping

A general problem for visualization is the mapping between visual attributes and data attributes. If the data attribute is uniformly distributed, a linear mapping to the visual attribute is feasible. In this case, the whole range of the visual attribute is used to produce the graphical representation. However, if the distribution of the data attribute is not uniform, the use of a linear mapping is inappropriate, as it leads to a small portion of the range of the visual attribute being used for a disproportionally large number of items. As a remedy, we employ a technique developed by Herman et al. [13], where mapping between data attributes and visual attributes is conceived as a two-step process: First, the empirical distribution function $F_X : \mathbb{R} \mapsto [0, 1]$ is computed for the data values:

$$F_{x_1,\ldots,x_n}(t) = \frac{1}{n} |\{i, 1 \leq i \leq n : x_i \leq t\}| \tag{6}$$

The value thus obtained is then mapped to the visual attribute by an attribute-specific function, the *attribute mapping*. The *attribute mapping* is a function $\alpha : [0, 1] \mapsto \mathbb{R}$ that depends on the type of visual attribute. It is not generally true that a linear mapping is always appropriate for this purpose. For example, linear grading of a certain visual attribute may not be perceived as linear by a human, in which case α should be nonlinear, compensating for the said effect. However, for our data sets we did not encounter any such problems. By choosing suitable linear functions for our various visual attributes, we obtained convincing results, as will be shown in the remainder of this paper.

3.2. Visualization: General Considerations

The visual representations presented were chosen to provide better orientation while browsing music archives on the web. In such a setting, the user is usually confronted with a mid-range amount of data (10-200 items), typically as a mixture of known and unknown terms (artist names, genre names, etc.) in the context of a rather short attention span and spontaneous action. Here, the main goal is to support the identification of the context of terms already known to the user, since a meaningful interaction with the data provided is only possible if one's own intellectual classification matches the classification provided by the system.

Where appropriate, the visual design of the different views supports micro-macro reading[14]. This means that the same view can be read with a broad focus, providing information about the whole collection of objects displayed, and it can be read in a detailed fashion, yielding more specific information on single objects. In creating the visualizations, principles of gestalt psychology[15] were embraced in order to make them as coherent as possible.

3.3. Tag Cloud

Tag clouds, or weighted lists in visual design, recently became popular as a depiction of content tags in Web 2.0

[9] Any variant of $\gamma(a)$ as described in Section 2.2 is applicable here.

Figure 1. Tag Cloud

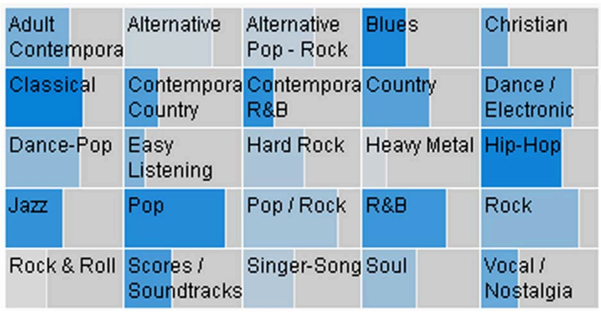

Figure 2. Bar Chart Matrix

applications. More frequently used tags are emphasized by font size and color. Tag clouds have been popularised by Websites such as Del.icio.us[10] and Flickr[11], and are a common representation for relative weights. However, from a typographic point of view, this method is partly misleading since terms such as "Soul" and "Adult Contemporary" on the same level of emphasis will never have the same level of visual importance due to the different typographical weight (length) of the terms.

In the context of our application, tag clouds are used to depict the size of the musical genres represented in the music collection (see Figure 1). Color saturation helps to emphasize the differences in font size, i.e. big, highly color saturated fonts as opposed to small ones with low saturation indicate a large amount of artists in the genre. Summing up, tag clouds are well suited for the visual representation of a single dimension. If a second dimension is to be rendered, bar chart matrices are a more intuitve means for presentation.

3.4. Bar Chart Matrix

Bar chart matrices are well suited to provide an overview when the number of individual items is high, see Figure 2

[10]http://del.icio.us
[11]http://www.flickr.com

Figure 3. Micro Bar Chart (Detail) displaying Jazz artists

presenting an overview of the 25 largest genres available in our music collection. The gray boxes equally sized for each genre compensate for the differing visual impact of the graphical weight of the genre names. The size of a genre in the music collection is displayed as the bar length. Genre purity is rendered as color saturation. An intense color still catches one's eye – rather independent of the size of the area. The direct juxtaposition of shades helps to distinguish even light differences which otherwise would not be noticeable.

3.5. Micro Bar Charts

When it comes to displaying a specific genre, genre prototypicality (red) and genre affiliation (blue) of the artists is rendered as a micro bar chart. See Figure 3 where artists (left column) are depicted according to their affiliation to a particular genre (in our case Jazz; middle column). In addition, an overview of the other genres an artist is affiliated to is given (right) in a comma separated list of <genre,bar chart> pairs. This kind of visualization accommodates the fact that in commercial music collections many artists are associated with more than 8 genres.

This view is an example of micro-macro reading: At a glance, one recognizes the continuously diminishing length of the red bars and the constant length of the blue bars for the genre Jazz. Affiliation to other genres can be examined in detail when focusing on a particular artist.

3.6. Similarity Spiral

Similarity spirals are visual equivalents of weighted lists. Figure 4 depicts a similarity spiral for Miles Davis. The names of related artists (i.e. outgoing references) are organized in a spiral emerging from the artist in focus displayed in the center, and with the most closely related artist as the nearest one in the spiral, which in our example is the Miles

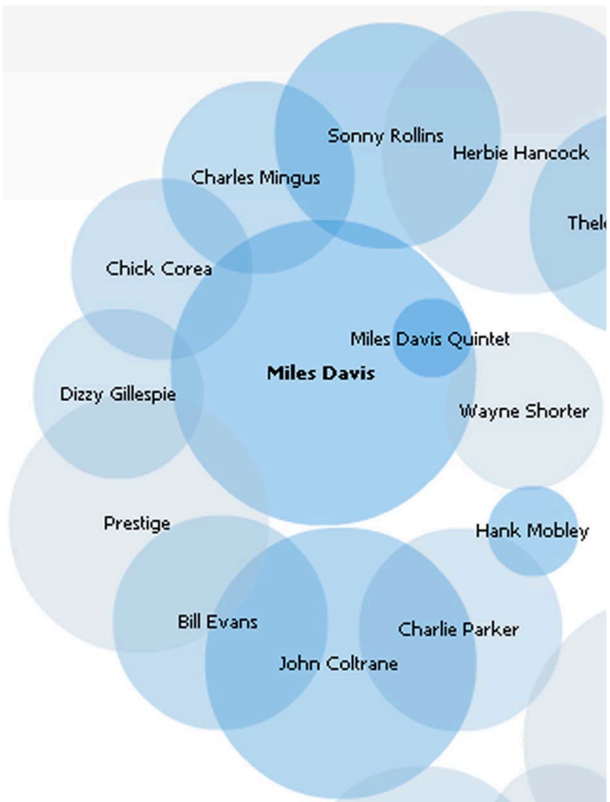

Figure 4. Similarity Spiral (Detail)

Davis Quintet[12].

The spiral is drawn on the screen in a short animation, starting with the centered sphere, by and by showing all of them. The inverse animation is run when the user clicks on one of the spheres (which leads to the display of a new spiral with the selected artist at the center). These movements emphasize the gestalt of the spiral and suggest reading it as a descending list.

The size of a sphere surrounding each artist name indicates the overall artist prototypicality, i.e. significance of the artist relative to the other artists in the collection. The genre similarity between the centered artist and the others in the spiral is rendered by means of color saturation. From our example we see that Miles Davis shows high genre similarity with Hank Mobley, John Coltrane and Sonny Rollins with respect to genre affiliation, whereas Prestige is quite dissimilar in this respect.

While bar charts are well suited to depict two value dimensions, similarity spirals are an appropriate means for depicting three dimensions. Moreover the spiral shape allows for a very compact representation of a large number of items on a single screen, and it is still comprehensible as a descending list. From a typographic point of view, the spheres help to compensate for the visual impact of different word sizes.

Among the presented visualization methods similarity spirals are by far the least *lightweight* ones, mainly because the list of similar artists is displayed in a not yet common form the user still has to become accustomed to. The realization of the spiral module requires flash[13], and thus raises technical issues when it comes to integrating it into a Web application.

3.7. Incoming/Outgoing References

Incoming/outgoing references for an artist facilitate a closer look at the significance of the artist and allow the user to explore the dynamics of musical influences. See Figure 5 for a depiction of the incoming and outgoing references of Miles Davis. Note that the outgoing references, depicted above the centered artist are those also shown in the similarity spiral. The references are represented as bar charts with the arrow head indicating an incoming reference and the tail indicating an outgoing reference. The strength of the relation with a certain artist is represented by the length and the alignment of the bars. Longer bars and bars closer to the artist in focus indicate a higher degree of relatedness.

[12] Note that our metadata do not distinguish between single artists, groups, composers etc.

[13] `http://www.adobe.com/products/flash/`

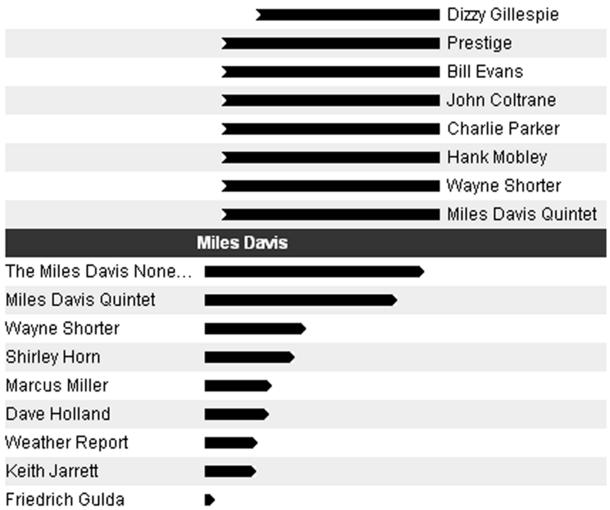

Figure 5. Incoming and Outgoing References

4 Conclusion

In this paper, we address two principal weaknesses of online music stores that limit their accessibility to users, namely (i) the absence of metadata that allow the user to grasp the structure and content of the archive , and (ii) the lack of visual navigation tools.

We enhance the existing metadata with *genre size*, *genre purity*, *genre affiliation* and *genre purity* of artists as well as *genre similarity* between artists, all of which are computed from existing metadata. Other new attributes, namely *artist-to-artist relatedness*, also interpreted as *artists' incoming and outgoing references*, *genre prototypicality* and *overall prototypicality* of artists are generated employing Web-based data mining techniques.

We show how the attributes can be used to augment classical Web-browsing applications with lightweight visualizations that employ commonly established principles of interaction and data visualization. More specifically, we demonstrate how one, two and three dimensions of metadata can be represented visually in a concise manner through *tag clouds* and *micro bar charts*, *bar chart matrices* or *similarity spirals*, respectively.

The resulting application combines different views on the music archive: an overview of the whole collection is given by the genre-centric bar chart matrix view; metadata attributes of single artists or genres are displayed using micro bar charts; displaying *incoming* and *outgoing* references, we throw a spotlight on the archive interpreted as a graph, where artists are nodes interconnected by edges weighted by their similarities. The similarity spiral is a condensed representation of the related artists enriched with metadata attributes not shown in the other views.

While the current application only comprises artists and genres,[14] future work will include songs as navigable elements in the user interface. Moreover, integration of metadata from other sources, especially additional similarity measures based on MIR techniques will be subject of future development.

5 Acknowledgements

This work was supported by the Austrian Federal Ministry of Economics and Labor and the Impulsprogramm Creativwirtschaft of Austria's Wirtschaftsservice (AWS).
We would like to thank David Mann for proficiently programming much of the application and Mandakini Pachauri for keeping us up to date with the latest developments in commercial music download platforms.
The code for performing Web-based co-occurrence analysis has been developed and provided by the Department of Computational Perception of the Johannes Kepler Universität Linz[15].

References

[1] R. v. Meteren and M. v. Someren. Using content-based filtering for recommendation. In *Proceedings of ECML 2000 Workshop: Machine Learning in New Information Age*, Barcelona, Spain, 2000.

[2] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and l J. T. Ried. *ACM Transactions on Information Systems*, volume 22, chapter Evaluating Collaborative Filtering Recommender Systems, pages 5–53. ACM Press, New York, USA, January 2004.

[3] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, pages 241–250. ACM, ACM Press, New York, NY, USA, 2000.

[4] E. Gstrein, F. Kleedorfer, R. Mayer, C. Schmotzer, G. Widmer, O. Holle, and S. Miksch. Adaptive personalization: A multi-dimensional approach to boosting a large scale mobile music portal. In *Fifth Open Workshop on MUSICNETWORK: Integration of Music in Multimedia Applications*. Vienna, Austria, 2005.

[5] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, November 2002.

[14] The presented methods and strategies have been implementated in a Web application which is publicly accessible at `http://cloudsofmusic.researchstudio.at`

[15] `http://www.cp.jku.at/`

[6] E. Pampalk, A. Rauber, and D. Merkl. Content-based organization and visualization of music archives. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 570–579, New York, NY, USA, 2002. ACM Press.

[7] E. Pampalk. Aligned self-organizing maps. In *Proceedings of the Workshop on Self-Organizing Maps, Kitakyushu, Japan*, September 2003.

[8] F. Mörchen, A. Ultsch, M. Nöcker, and C. Stamm. Databionic visualization of music collections according to perceptual distance. In *ISMIR*, pages 396–403, 2005.

[9] 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 ACM Multimedia 2006 (MM'06)*, Santa Barbara, California, USA, October 23-26 2006.

[10] E. Pampalk and M. Goto. Musicrainbow: A new user interface to discover artists using audio-based similarity and web-based labeling. In *ISMIR*, pages 367–370, 2006.

[11] A. Berenzweig, D. P. W. Ellis, and S. Lawrence. Anchor space for classification and similarity measurement of music. In *ICME '03: Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 2 (ICME '03)*, pages 29–32, Washington, DC, USA, 2003. IEEE Computer Society.

[12] M. Schedl, P. Knees, and G. Widmer. Investigating web-based approaches to revealing prototypical music artists in genre taxonomies. In *Proceedings of the 1st IEEE International Conference on Digital Information Management (ICDIM'06)*, Bangalore, India, December 2006.

[13] I. Herman, M.S. Marshall, and G. Melanon. Density functions for visual attributes and effective partitioning in graph visualization. In SF Roth and DA Keim, editors, *Proceedings of the IEEE Information Visualization Symposium 2000*, pages 49–56. IEEE CS Press, 2000.

[14] Edward R. Tufte. *Envisioning Information*. Graphics Press, 1990.

[15] David Katz. *Gestaltpsychologie*. Benno Schwabe & Co. Verlag, Basle, 1948.