Mario Kubek
4FriendsOnly.com Internet Technologies AG
mk@4fo.de

Jürgen Nützel
4FriendsOnly.com Internet Technologies AG
jn@4fo.de

NOVEL INTERACTIVE MUSIC SEARCH TECHNIQUES

Abstract: The discipline of finding and recommending new and similar multimedia content has gained a growing importance in multimedia information retrieval over the last years. The two classical approaches, namely Content-based filtering (CB) and Collaborative filtering (CF), can be used in conjunction with each other to get benefit from both content analysis and user feedback. This paper introduces a system, called GlobalMusic2one that implements a novel approach to hybrid music search by combining methods of Music Information Retrieval (MIR) and Web 2.0 techniques to reach a new quality in recommending and distributing global music collections. To be able to classify these collections and to determine relationships between them, textual descriptions are gathered by collaborative tagging and statistical Web document analysis, as relevant documents in the Web may contain a lot of freely accessible and descriptive information on music content. The resulting text corpus can be used to identify connected entities, e.g. which artist is connected or similar to another artist. Therefore, this approach sensibly complements CB and CF techniques and has already been successfully implemented in the interactive music search system NIMS, which will be introduced as well.

Keywords: interactive multimedia information retrieval, multimedia content recommendation, Web crawler, Web document analysis, search engine, text mining, GlobalMusic2one, NIMS

1. Motivation and Introduction

Since the last years recommender systems have become a commodity. Many e-commerce portals apply them to automate word-of-mouth advertising, e.g. Amazon.com. The most prominent kinds of classical recommender systems are using techniques of Content-based filtering (CB) and Collaborative filtering (CF) to suggest similar items. Pure CB systems typically analyze the content of items, store their relevant features in vectors and compare them to find items with similar vectors. Therefore, the result of this comparison is called item-to-item-correlation. Pure CF systems evaluate and compare inherent or explicit feedback, click paths and profile data of users. That is why this comparison

yields a user-to-user-correlation. CB systems do not rely on user profiles. Therefore, they can suggest items that are not yet rated by users (cold start problem). If there is no sufficient description of items available then the recommended results may not be of high quality. Even a qualitative evaluation is not always possible, except a qualitative description is part of the content to be analyzed. These two drawbacks can be addressed by CF systems, but they have specific problems as well. As an example, highly popular items are more likely to be recommended, disregarding a user's special taste. The above mentioned cold start problem is very immanent in this approach. Furthermore, the provider has to manage, maintain and secure sensitive user data. Therefore, hybrid recommender systems have been developed to combine both approaches in order to get benefit from their respective advantages and to reduce their drawbacks. In Music Information Retrieval (MIR) these approaches are applied in various ways. Music content analysis is used to extract features of music without user intervention with the help of digital signal processing. These extracted features are then mapped to defined classes by applying statistical models to identify similar content. A disadvantage of this solution is that computed content similarities often do not represent the perception of human listeners ("semantic gap"). Nowadays most recommender systems for music are based on the CF approach, e.g. Last.fm and iLike. Manual tagging is also a popular way to classify and evaluate music collections. This is usually conducted by experts (e.g. Pandora) or by users in a community (e.g. Musicbrainz). Collaborative tagging is a proper way to bridge the "semantic gap" in music recommendation. But as users are free in choosing suitable tags a complex statistical text analysis is needed to extract semantic relationships between the tagged objects. That implies that new content can not be easily recommended as there might not be an attribution to rely on. Another way to get semantic information on music in order to bridge the "semantic gap" is to analyze current textual music descriptions as lyrics or reviews. The World Wide Web contains a lot of this valuable information. Therefore, it is sensible to make use of this information as well in order to get meaningful descriptions or annotations for music content, even if it has not yet been reviewed or annotated. A further reason to get as much textual description on music as possible is that music cannot describe itself unless lyrics or other descriptions are attached to the content. Since users should be able to enter textual queries to retrieve music content a textual mapping from queries to the music and its annotations is necessary. This paper introduces the system GlobalMusic2one that implements a novel approach to hybrid music search by combining methods of content analysis, collaborative tagging and statistical Web document analysis in order to achieve a new quality in music recommendation. The next section focuses on the objectives and the technical architecture of GlobalMusic2one. Section three describes the framework for statistical text analysis that will be used in

GlobalMusic2one to extract relationships between music contents based on user annotations and Web document analysis. In section four the first showcase project for this approach, called NIMS, is introduced. Section five concludes the paper and provides a view on future research concerning hybrid music recommendation based on statistical text analysis of music descriptions.

2. GlobalMusic2one

Hybrid recommender systems for music content that combine community-based approaches and content-based music similarity analysis are a highly discussed topic in multimedia information retrieval. In the last years several publications on this have been made available [1][2][3]. Though, the real potential of such hybrid approaches has not yet been fully exploited. Global music content can be characterized by a large cultural diversity and a high dynamic. Therefore, the usage of above mentioned hybrid recommender systems to classify such content present a promising solution. In the project GlobalMusic2one [4] such a system is currently being created. It is funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) as part of the „KMU-Innovativ“ programme since November 2008. Participating members of the consortium are the Bach Technology GmbH, the Fraunhofer Institute for Digital Media Technology (IDMT), the 4FriendsOnly.com Internet Technologies AG and the Piranha Musik & IT AG.

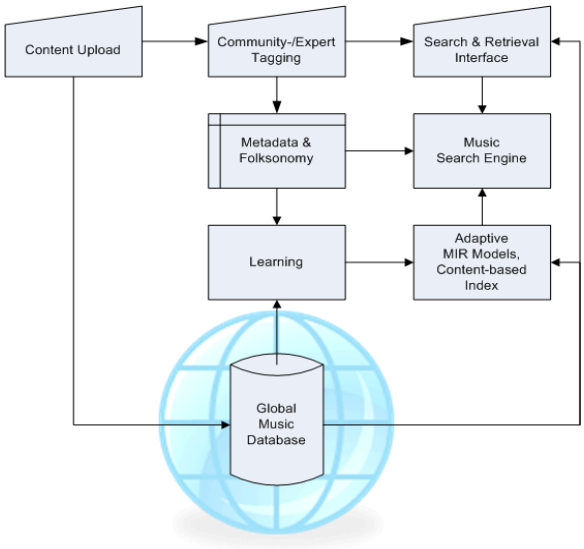

Figure 1. Architecture of GlobalMusic2one

The project's outcome will be an adaptive software program which users "train" by adding new musical categories and assigning them example songs. The software thereby gradually "learns" to recognize various musical qualities and relationships. GlobalMusic2one will also be able to automatically import new content into databases and to autonomically analyze recordings by identifying their respective musical qualities (for example genre, tempo, mood) and thereafter assign each recording to appropriate categories. Additionally, users will be able to freely annotate songs and categories by assigning describing catchwords. This tagging of content enriches the system with interactive elements and therefore closely relates to the Web 2.0 idea. The entirety of all assigned tags represents a continuously expanding musical knowledge base forming a folksonomy that shall be exploited to improve the quality of automatic searches on a semantic level. Furthermore, a statistical Web document analysis will be conducted to get additional information on music content out of freely accessible textual content on music as lyrics and reviews. This way it will be possible to gain relationships between music entities as artists and songs, even if users have not yet classified them. This solution also complements the used adaptive MIR framework by Fraunhofer IDMT and has already been successfully implemented in the interactive music search system NIMS which will be described in section four. For these two tasks a framework for statistical text analysis will be applied comprising a highly adaptable Web crawler and a component to calculate significant term frequencies and co-occurrences. A Web Service will be used to request these data from. This framework will be introduced in more detail in the next section. The overall system architecture of GlobalMusic2one is depicted in figure 1 presenting the most important components. Based on these approaches the system will be able to automatically adapt existing MIR models by evaluating user feedback, by recalculating and reweighing content relationships and by automatic reasoning. The realm of semantics represents one of the most important aspects of a future Web 3.0. This hybrid approach, namely the combination of music analysis with Web 2.0 technology and the "Semantic Web", is what distinguishes GlobalMusic2one from already existing solutions on the market and is a step towards new kinds of MIR systems in a Web 3.0.

3. Statistical Text Analysis of Music Descriptions

The World Wide Web contains a huge amount of textual descriptions on music as lyrics, CD reviews, discographies and fan sites. This knowledge from diverse sources presents a valuable and up-to-date resource to gather dynamically changing music trends from and to obtain relationships between musical entities. As music cannot describe itself textually per se, the importance of this knowledge is further underpinned when it comes to presenting appropriate search results in specialized music search engines that most often only accept textual queries. For a music search engine aiming at giving users a uniform access to this diverse knowledge, it is necessary to extract and sensibly combine this knowledge. Thus, the utilization of methods for statistical text analysis is a proper way to achieve this task. In the introduced project GlobalMusic2one we implement this approach to complement the music analysis. Furthermore, we use it to statistically analyze collaboratively acquired tags on music. The research project NIMS ("Neuartige Interaktive Multimedia-Suchverfahren") was the starting point to develop such a kind of music search engine. The project has been realized in collaboration with the NLP (Natural Language Processing) group [5] at the University of Leipzig and was funded by the German Federal Ministry of Economics and Technology (Bundesministerium für Wirtschaft und Technologie, BMWi). Its aim was to develop novel interactive multimedia search techniques based on text mining and statistical text analysis. While the interactive components of NIMS will be described in section four, we will now focus on the two main components used to acquire and analyze textual music descriptions and to calculate significant relationships between them.

3.1. The Web Crawler

To gather freely accessible music descriptions in the Web a specialized Web crawler was created. Its task is not only to identify and retrieve relevant documents with the topic music, but also to filter out unsolicited content such as advertisements that often decreased the quality of calculated relationships between music entities as artists during postprocessing, e.g. when they do not match the documents' topics. This filter has been gradually enhanced. To better identify music relevant websites a stopword list has been created that contains words that often occur in this context e.g. "love". Basically, it was necessary to manually find such relevant and interconnected websites as a seed to start the crawling process from in order to build up a meaningful music corpus as fast as possible. Thereby it was found that homepages of artists are not suitable for this purpose as they often do not contain any links to other similar artists. Also big music portals as MySpace and last.fm were found not to be a helpful resource,

as their pages often provided little descriptions and were full of misplaced user comments. Moreover, the crawling process has been made very difficult by them through technical restrictions. The resulting music corpus at the end of the project (October 2008) consisted of gathered information of the German Wikipedia and the portal Laut.de. Wikipedia provided a lot of useful information on the background of artists and often their whole discography was available as well. Laut.de was a proper choice because it offered very up-to-date information on artists in the form of CD reviews and links to similar artists. The Web crawler can work independently from the text analysis, so that the whole process can be easily controlled and adapted making it possible to update the data basis while the system is running.

3.2. Textual Analysis

The component for statistical text analysis, called Medusa [6][7], is used to determine significant occurrences of music entities and significant co-occurrences between them. For this purpose the significance measure log-likelihood is applied with the empirically validated thresholds 6.63 and 3.84. Terms in the above mentioned stopword list do not have an influence on these calculations. Additionally, a part of the freely available database for music metadata Musicbrainz [8] is used to further enhance the detection rate of music entities by matching them against a verified data basis before they reach the client. In the last step the resulting music corpus with its term frequencies and co-occurrence data is mapped to a database scheme which is used by the Web Service to forward these data to the client. The NIMS Web Service will be described in the next section.

4. NIMS

The second aim of NIMS was to visualize the gathered musical entities from the statistical text analysis on an interactive website in form of an individual map and to enrich them with matching advertisements of commercial products. First we take a look at the overall architecture of NIMS before the graph-based visualization and its navigation is discussed.

4.1. The Architecture of NIMS

The following figure depicts the main components of NIMS. The components for statistical text analysis and the graphical user interface are interconnected by a Web Service which calculates the visualization based on the data retrieved from the text technology. The graphical user interface consists of three components. The web site with search functionality embeds the interactive user

interface, integrates commercial advertisements and requests and receives visualization data depending on user interactions.

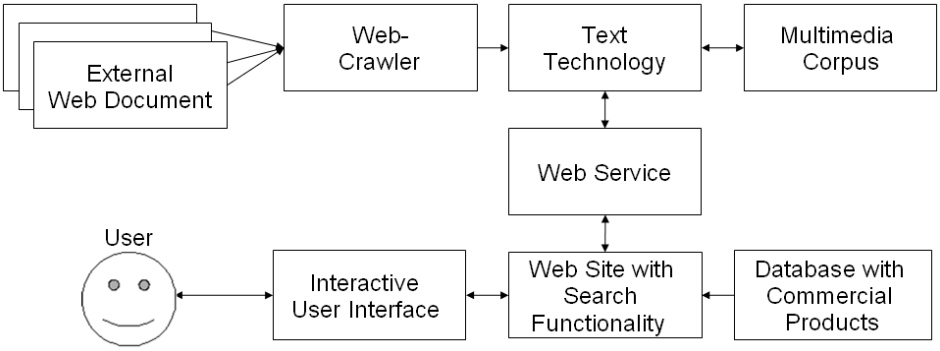

Figure 2. The Architecture of NIMS

We will now focus on the data provided by the Web Service, the interactive user interface and the graph-based visualization of musical entities.

4.2. Web Service

The main task of the Web Service is to calculate the graph-based visualization based on the previously calculated term frequencies and co-occurrences. An exemplary search for the German artist "Udo Lindenberg" could yield the following similarity data that would be used to calculate the visualization:

```
<nims_service>
<cooccurrences search="Udo Lindenberg" search-freq="304" search-freq-class="7.27">
<coocc term="Peter Maffay" label="ARTIST" dist="19.44" freq="335" freq-class="7.13"/>
<coocc term="Herbert Grönemeyer" label="ARTIST" dist="40.17" freq="299" freq-class="7.29"/>
<coocc term="Inga Rumpf" label="ARTIST" dist="52.29" freq="48" freq-class="9.93"/>
<coocc term="BAP" label="ARTIST" dist="54.16" freq="200" freq-class="7.87"/>
<coocc term="Ton Steine Scherben" label="ARTIST" dist="73.98" freq="158" freq-class="8.21"/>
<coocc term="Frumpy" label="ARTIST" dist="76.78" freq="47" freq-class="9.96"/>
<coocc term="Nena" label="ARTIST" dist="81.97" freq="310" freq-class="7.24"/>
<coocc term="Truck Stop" label="ARTIST" dist="101.1" freq="54" freq-class="9.76"/>
<coocc term="Peter Herbolzheimer" label="ARTIST" dist="106.42" freq="74" freq-class="9.31"/>
<coocc term="Nina" label="ARTIST" dist="123.8" freq="2631" freq-class="4.09"/>
<coocc term="Bruce Springsteen" label="ARTIST" dist="131.87" freq="314" freq-class="7.22"/>
</cooccurrences>
</nims_service>
```

Each dataset contains similar terms to the query, a specification on their semantic interpretation (artist, album, song), the semantic distance "dist" to the query, the absolute frequency "freq" of these terms in the corpus and the normalized frequency value "freq-class", which maps the absolute frequency to a range between 0 and 20. Terms with a frequency class near 0 have a high significance in the corpus, whereby terms with a frequency class near 20 are less important.

4.3. The Visualization

The interactive user interface of NIMS provides users a convenient way to navigate and zoom inside the graph of musical entities. For the prototypical implementation we decided to visualize artist similarities. The Google Maps API has been applied for this purpose because of its widely accepted interface supporting panning and zooming functionalities, its Ajax approach, so that only data that need to be updated have to be requested by the client (web browser), and because it is possible to create so-called "Custom"-Maps to visualize third party datasets. In our visualization close nodes suggest a high artist similarity. The thicker the edge between two nodes is the more important their represented artists are. This correlates with the previously mentioned frequency class and is an analogy to streets or highways between cities and to cartways between villages. Another important feature is that the map grows as the user searches inside it. Therefore, clusters of similar artists (and inherently music styles) will become visible. It would have been uneconomical to recalculate the whole map every time a new query is submitted. Thus, only a specific area around a requested node is recalculated using a modified version of the Fruchterman-Reingold algorithm [9] for force-directed placement in order to save CPU time and to gain a faster clustering of connected nodes.

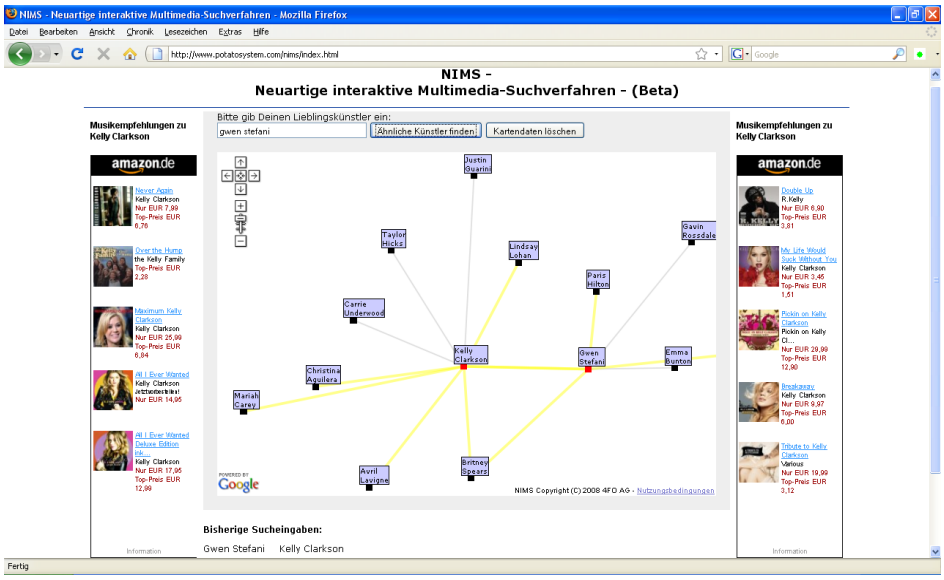

Figure 3. Graph-based Visualization in NIMS

The graph drawing itself is solely performed on the client and is based on XML-formatted data retrieved from the Web Service containing node identifiers, positions of nodes and edges and the color and thickness of edges.

The screenshot in figure 3 displays an exemplary graph after searching for the artist Gwen Stefani and then for Kelly Clarkson. Here the graph, previous queries underneath the graph (search history) and topically matching commercial advertisements are displayed. Nodes that have been used as a query are marked red.

Figure 4. Integration of Youtube-Videos in NIMS

To further enrich the multimedia experience users can view Youtube videos of selected artists on the map as depicted in figure 4. For this purpose users just have to move the mouse pointer over a node of interest.

5. Conclusion

In this paper we introduced the hybrid music recommender system GlobalMusic2one, which is currently under development, and described its overall architecture. Its novel approach lies in the combination of content-based music similarity analysis with user feedback evaluation and the acquisition of textual metadata of music in order to calculate content relationships by automatic reasoning. For this purpose, we proposed the utilization of a framework for statistical text analysis to gain these needed music metadata and their semantic relationships by analyzing freely accessible textual content offered in the World Wide Web. Then we presented NIMS as a first showcase project for this approach, described its system architecture, its interactive user interface and the visualization of the gained musical entities. Currently, GlobalMusic2one is still in an early stage and therefore many intended features have not yet been implemented. One example is the component to automatically recalculate and reweigh content relationships based on the above mentioned

input data. Another research field lies in the evaluation of methods for the statistical analysis of textual annotations of musical entities provided by users. It is necessary to determine if the described techniques for statistical Web document analysis can be applied on these mostly sparse data as well. This way sensible suggestions for possible annotations could not only be offered in the form of important and relevant terms provided by the Web document analysis but could also come from the community, even if users did not yet provide annotations in the first place. Also user generated playlists and personal ratings by users could be analyzed and compared using these techniques in order to not only find similar content but to find users with a similar taste as well. Another possible field of application for the text technology can be seen in the utilization of its feature to provide proper expansion terms or associated terms to users' queries [10] (e.g. synonyms) based on the aforementioned significant co-occurrences or based on frequently assigned annotations within the community. As an example, semantically connected terms could be automatically attached to query terms entered by users in order to narrow down search results and to compensate the vocabulary mismatch problem which is immanent in this system as users are free in choosing sensible annotations and a musical entity could be completely differently annotated within the community. This way expanded queries are likely to return more relevant results as they can be matched against a broader set of entities. The following table presents two exemplary queries with their potentially suggested expansion terms or associated queries.

Initial Query	Expanded or Associated Queries
guitars	acoustic guitars, electric guitars, heavy metal
female rock	Kelly Clarkson, Alanis Morissette, alternative rock

Table 1. Expanded or Associated Queries

Therefore, it is easy to see that these approaches to acquire and semantically analyze music metadata and user feedback present a proper basis to complement the calculation of content-based music similarities in a hybrid music recommender system. Another aim of GlobalMusic2one is to visualize relationships between musical entities similar to NIMS. The visualization will be improved in the way that multimodal similarities between musical entities like artists, albums, songs and even song segments will become visible. User feedback within the visualization will allow users to interactively rearrange entities according to their taste. The visualization as a component will also be exchangeable which makes it possible to provide a unique user experience adapted for different clients, e.g. mobile web browsers.

References

[1] K. YOSHII, M. GOTO, K. KOMATANI, T. OGATA, H. OKUNO, *Hybrid Collaborative and Content-based Music Recommendation Using Probabilistic Model with Latent User Preferences*, in Proc. ISMIR'06, 2006

[2] M. TIEMANN, S. PAUWS, F. VIGNOLI, *Ensemble Learning for Hybrid Music Recommendation*, in Proc. ISMIR'07, 2007

[3] D. TURNBULL, L. BARRINGTON, G. LANCKRIET, *Modeling Music and Words using a Multi-Class Naïve Bayes Approach*, in Proc. ISMIR'06, 2006

[4] WEBSITE OF GLOBALMUSIC2ONE, *http://www.globalmusic2one.net/*

[5] WEBSITE OF THE NATURAL LANGUAGE PROCESSING GROUP (NLP), University of Leipzig, *http://www.asv.informatik.uni-leipzig.de/opencms/asv/de/index.html*

[6] M. BÜCHLER, *Medusa: Performante Textstatistiken auf großen Textmengen: Kookkurrenzanalyse in Theorie und Anwendung*, Vdm Verlag Dr. Müller, 2008

[7] M. BÜCHLER, G. HEYER, *Kookkurrenzberechnungen mit UIMA und Medusa*, UIMA Workshop at the GLDV 2007 in Tübingen/Germany, 2007

[8] WEBSITE OF MUSICBRAINZ, *http://musicbrainz.org/*

[9] T. FRUCHTERMAN, E. REINGOLD, *Graph Drawing by Force-directed Placement*, SOFTWARE—PRACTICE AND EXPERIENCE, VOL. 21(1 1), p. 1129-1164, 1991

[10] M. KUBEK, H.F. WITSCHEL, *A System for Automatic Query Expansion in a Browser-based Environment*, In Proceedings of the 6th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods (VG'08), 2008