Information Processing and Management 45 (2009) 392–396

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Music search engines: Specifications and challenges

Alexandros Nanopoulos [a,*], Dimitrios Rafailidis [b], Maria M. Ruxanda [c,1], Yannis Manolopoulos [b]

[a] Hildesheim University, Institute of Computer Science, Germany
[b] Department of Informatics, Aristotle University of Thessaloniki, Greece
[c] Department of Computer Science, Aalborg University, Denmark

ARTICLE INFO

Article history:
Received 8 April 2008
Received in revised form 2 February 2009
Accepted 3 February 2009
Available online 19 March 2009

Keywords:
Multimedia information retrieval
Audio input/output
Sound and music computing
Web search

ABSTRACT

Nowadays we have a proliferation of music data available over the Web. One of the imperative challenges is how to search these vast, global-scale musical resources to find preferred music. Recent research has envisaged the notion of music search engines (MSEs) that allow for searching preferred music over the Web. In this paper, we examine the growing research topic of MSEs, and provide potential specifications to follow and challenges to face.

1. Introduction

Successful technologies, like the Web, file compression, or handheld music devices, changed the way that music is distributed and consumed, resulting to a unique characteristic: never before we had access to such global-scale, vast amount of musical resources. Novel opportunities are created for the music industry, by increasing the demand for music. Several commercial systems have been developed with varying objectives, e.g., online purchasing of tracks/albums or music recommender systems. As well, the research field of Music Information Retrieval (MIR) responded promptly by active research within the topics of music search and retrieval, such as the International Conferences on Music Information Retrieval (ISMIR) (http://ismir.net) supported by the emergence of the web service MIREX (the Music Information Retrieval Evaluation eXchange: http://music-ir.org/mirexwiki) that standardizes datasets used for the evaluation of the proposed MIR algorithms.

Several issues have emerged, both technological and economical. Among them there is the imperative need to address one big challenge: how to search the vast amount of available musical resources to find preferred music? Numerous existing commercial systems rely on simple searching by metadata (song titles or artist names). This has significant limitations (Berenzweig, Logan, Ellis, & Whitman, 2004), such as the ambiguity or the inability to search for songs with specific musical characteristics. By viewing the Web as a global-scale musical resource, we confront new issues that, so far, have not been thoroughly addressed. Motivated by this fact and inspired by Web search engines, recent researches (Celma, Cano, & Herrera, 2006; Knees, Pohle, Scheldl, & Widmer, 2007) have envisaged the development of *music search engines* (MSE). MSEs search preferred music over the Web. Other multimedia types, such as images, have already been included in web search engines.

* Corresponding author.
E-mail addresses: nanopoulos@ismll.de (A. Nanopoulos), draf@csd.auth.gr (D. Rafailidis), mmr@cs.aau.dk (M.M. Ruxanda), manolopo@csd.auth.gr (Y. Manolopoulos).

[1] This research was supported in part by the Danish Research Council for Technology and Production Sciences project no. 26-04-0092 Intelligent Sound (www.intelligentsound.org).

0306-4573/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2009.02.002

Due to the vehement competition between search-engine vendors and the proliferation of online musical content, developments in the field of MSEs will appear in the near future (Sahami, Mittal, Baluja, & Rowley, 2004).

The objective of this article is to tackle the growing research topic of MSEs, and to provide potential specifications to follow and challenges to face. For this reason, we have examined 60 existing commercial and research prototype systems, henceforth called *music exploration systems*. In this paper, we focus on 20 of the most prominent music exploration systems to identify possible solutions (Section 2). Next (Section 3), we present specifications for MSEs, based on findings from the present survey. Moreover, we describe the challenges that have not been considered so far by music exploration systems, which have to be addressed by MSEs. Finally, we furnish our conclusions (Section 4).

2. Characteristics of existing music exploration systems

2.1. Functionalities

The most common functionalities found in existing systems are the following:

- *Search by metadata for music-related information*: (artists biography, music reviews, new releases, concerts dates, etc.). Known examples are *Allmusic* and *Yahoo!Music*.
- *Search for Lyrics*: *Lyrics.com* and *SearchLyrics.com*.
- *Media Management and Track Identification*: identify metadata for tracks (*Gracenote*). Others (*MusicIP*) organize identified music. With track identification systems (*Shazam*), users record tracks and the system outputs metadata.
- *Search by Track/Artist for music browsing (discovery and listening of music):* popular systems (*Last.fm, iTunes, Yahoo!Music*) search with various criteria. More narrow criteria may also be used (*Mp3Realm, SkreemR*). There exist some systems (*Musipedia* and *Sloud)* where users query by humming.
- *Recommend Similar Music*: based on seed elements (artist, track) users are recommended similar tracks or artists. Popular recommender systems are *Pandora, Last.fm*, *Yahoo!Music*, *OneLlama*, and *Musicovery*.
- *Generate Playlists*: automatic generation of playlists, that satisfy user constraints (Aucouturier & Pachet, 2002).

2.2. Input/output

The types of input that are provided by existing music exploration systems can be broadly grouped in two categories: textual input and audio input.

Textual input refers to pure textual information, e.g., metadata (*SearchSounds* and *SkreemR*), lyrics (e.g., *Lyrics.com*), or high-level textual features, like "smart words" (Knees et al., 2007) ('driving music' or 'artist with aesthetic voice') or "tags" (Levy & Sandler, 2007), which have started to be used in *Amazon* and *Last.fm* (Fig. 1b). Other textual features represent high-level concepts, like mood or genre – see *Musicovery* (Fig. 1a) and *MusicRainbow* (Pampalk & Goto, 2006) (Fig. 1c).

Audio input can be a track or low-level features extracted from a track, to be used for Querying-by-Example. Some systems (*MusicIP*) allow users to input an existing track either from a catalog or to upload one (*Orpheus*), and others to generate audio by recording (*Shazam*), or humming (*Sloud* or *Musipedia* – Fig. 2e). An alternative way is to select various low-level features extracted directly from the audio (Fig. 2d for *Sony's Music Browser*).

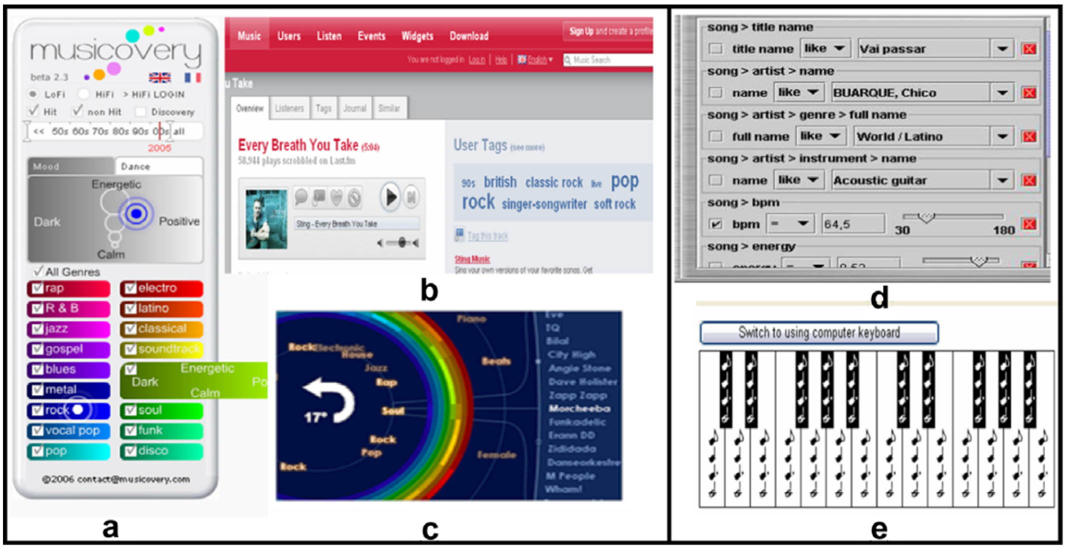

Fig. 1. Examples of input.

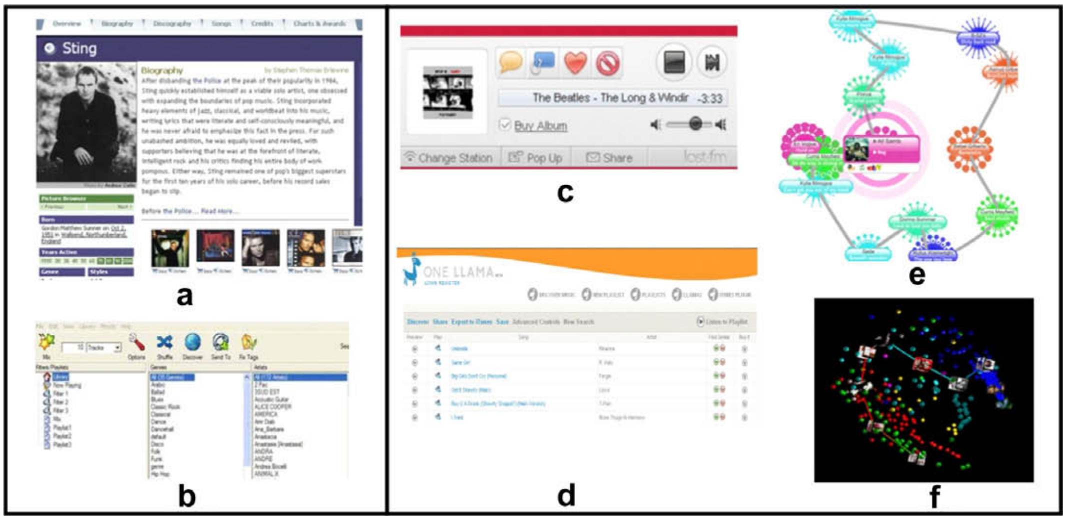

Fig. 2. Examples of output.

Input types can be characterized along the following three dimensions. (i) Easiness of interaction: how easily a user provides input; (ii) control on recall: ability to input various criteria and increase coverage by finding several relevant results; (iii) control on precision: ability to input exact criteria and find accurate results. Table 1 describes the three corresponding dimensions. Overall, high-level textual features can be considered as the best input type, because they do not have a low mark in any dimension.

Output should be displayed in an intuitive, artistic, and appealing manner. We can group output in two basic categories: *non-audio* output and *audio* output. Non – audio provides information about music (*Allmusic* Fig. 2a), not the music itself. Media management systems (*Gracenote* and *MusicIP*) display metadata (Fig. 2b). Audio-output is presented with links to archives of songs (*Mp3Realm* and *SkreemR*) An interactive radio box (Fig. 2c) is used by *Yahoo!Music*, *Pandora* and *Last.fm*. Other systems display playlists (*OneLlama* – Fig. 2d and *MusicIP*), or coloured networks (*Musicovery* – Fig. 2e and *Search Inside The Music* by Sun Microsystems – Fig. 2f).

2.3. Similarity measures

Similarity measures quantify how much alike songs (or artists etc.) are to each other. Three general types of similarity measures exist:

- *Audio similarity measures*: are based on audio features, like Mel-Frequency Cepstral Coefficients (MFCCs) and computed with statistical methods, like the Gaussian Mixture Models (GMMs). They cannot capture similarity in terms of human perception (Berenzweig et al., 2004), but they do not require any human input.
- *Social similarity measures*: are based on features, like tags, collaborative data, playlists, and human opinions, collected from social media sites. They can discover relationships that may be difficult to detect from audio (Berenzweig et al., 2004). However, they require a reasonable amount of reliable social features.
- *Hybrid similarity measures*: combine the advantages of audio and social similarity measures, by fusing social with audio features (Berenzweig et al., 2004;Knees, Schedl, Pohle, & Widmer, 2006; Knees et al., 2007; Pampalk & Goto, 2006).

Table 1
Comparison of input types (QBE stands for query-by-example).

	Query specification	Easiness of interaction	Control on recall	Control on precision
Textual input	Metadata	High	Low	Low
	Lyrics	Medium	Low–medium	Low–medium
	High-level textual features	High	High	Medium–high
Audio input	QBE based on audio	Low (humming, recording) – High (existing audio)	Low–medium	Low–medium
	QBE based on audio features	Low	Medium–high	High

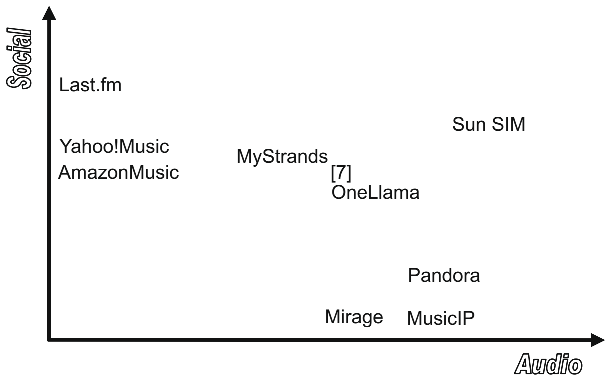

Fig. 3. The two dimensions of music similarity.

In Fig. 3, known music exploration systems are plotted, where placement is qualitative along two clear axes that represent the audio and the social similarity.

3. Specifications and challenges

3.1. Specifications

This section describes the most important specifications for MSEs, by evaluating possible solutions that have been followed. We avoid making exclusive decisions. For each specification we prioritize the presented solutions according to their appropriateness for MSEs.

3.1.1. Functionalities

An MSE can have many functionalities. The most important is to search for tracks and artists, because it helps users to find relevant music. Searching for music-related information is important too and is already provided by engines like GoogleMusic.

3.1.2. Input/output types

Metadata is indispensable for input, but may not be adequate. Based on conclusions from Table 1, the most important input type for an MSE is the high-level textual features (they encapsulate metadata). An additional graphical mechanism (see Fig. 1c) can help the input process. Audio files as an input type do not allow users to specify any detailed searching criteria. An alternative use is first to extract audio features from an input track, and then provide these features for an "advanced search". Output in the form of Fig. 2e and f is more appealing than simple lists of hyperlinks. Graphical output summarizes various aspects: musical characteristics (mood, genre), metadata and relationships between results.

3.1.3. Similarity measures

Hybrid similarity measures are the most promising ones and should be preferred.

3.1.4. Architecture and copyright issues

MSE can directly follow the paradigm of web search engines. Content Delivery Networks (CDNs) reduce delivery times (especially for video clips). MSEs could operate in decentralized architectures, such as P2P. Regarding copyright issues, an MSE can act as an intermediate layer. This way, audio output can be provided through external links, in a streaming form, or/and as music snippets.

3.2. Challenges

When considering the Web as a big musical resource, we face novel challenges. In this section we describe some of them, which are imperative to solve and have an implementation of real-world MSEs.

The first challenge is how to crawl for music data. Following web search engines, crawling in MSEs has to identify musical data, either from context or through other sources (catalogues). A prioritization scheme can address the volume of available music data and fast rate of their change. Focused crawlers (Chakrabarti, van den Berg, & Dom, 1999) can be used, but novel approaches are needed too (Celma et al., 2006). Due to copyright issues, we can either copy a track or extract features that will help just to index it.

The second challenge concerns the indexing of crawled data, to facilitate fast and accurate retrieval. We have to index huge volumes of acoustic features, which is hard. Established techniques, like GMMs of MFCCs are time demanding (Roy, Aucouturier, Pachet, & Beurivé, 2005) and not easily indexable.

The third challenge is about how to rank the results delivered by an MSE. Most existing systems rank search results only by music similarity. However, web search engines rank results according to authority induced from hyperlink analysis (e.g., PageRank). For multimedia data the issue of ranking is very important (Celino, Della Valle, Cerizza, & Turati, 2006; Yang & Chen, 2005). Nevertheless, in music browsing, serendipity and exploration are important distinctive factors. The balancing between authority ranking and factors like serendipity have just started to be considered (Ruxanda et al., 2008).

Finally, there is the issue of integration: MSEs can either be independent or can be integrated in existing web search engines. Web search engines have started to provide the ability for searching non-textual like images. The same can be expected for musical audio (Sahami et al., 2004). Nevertheless, independent MSEs are also expected to appear.

4. Conclusion

We examined the growing research topic of music search engines. According to latest research findings, we envisage MSEs as systems for searching preferred music over the Web. We believe that, like other multimedia types (i.e., mainly images), music searching will become a mainstream operation in search engines. The competition between search-engine vendors and the vast amount of available musical content that is online are promising for fast developments in this field.

In this article, we first provided a survey of the solutions adopted by existing music exploration systems. Next, according to this survey, we described possible solutions for the case of MSEs. Finally, we presented some of the main challenges that have to be confronted when implementing a real-world MSE.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ipm.2009.02.002.

References

Aucouturier, J.-J., & Pachet, F. (2002). Scaling up music playlist generation. In *Proceedings of the IEEE international conference on multimedia and expo (ICME)*.

Berenzweig, A., Logan, B., Ellis, D. P. W., & Whitman, B. (2004). A large – Scale evaluation of acoustic and subjective music similarity measures. *Computer Music Journal, 28*(2), 25–100.

Celino, I., Della Valle, E., Cerizza, D., & Turati, A. (2006). Squiggle: A semantic search engine for indexing and retrieval of multimedia content. In *Proceedings of the international workshop on semantic-enhanced multimedia presentation systems*.

Celma, O., Cano, P., & Herrera, P. (2006). Search sounds: An audio crawler focused on weblogs. In *Proceedings of the international conference on music information retrieval (ISMIR)*.

Chakrabarti, S., van den Berg, M., & Dom, B. (1999). Focused crawling: A new approach to topic-specific web resource discovery. *Computer Networks, 31*(11–16), 1623–1640.

Knees, P., Schedl, M., Pohle, T., & Widmer, G. (2006). An innovative three-dimensional user interface for exploring music collections enriched with meta-information from the web. In *Proceedings of the ACM international conference on multimedia*.

Knees, P., Pohle, T., Scheldl, M., & Widmer, G. (2007). A music search engine built upon audio-based and web-based similarity measures. In *Proceedings of the ACM conference on research and development in information retrieval (SIGIR)*.

Levy, M., & Sandler, M. (2007). A semantic space for music derived from social tags. In *Proceedings of the international conference on music information retrieval (ISMIR)*.

Pampalk, E., & Goto, M. (2006). MusicRainbow. A new user interface to discover artists using audio-based similarity and web-based labeling. In *International conference on music information retrieval (ISMIR)*.

Roy, P., Aucouturier, J.-J., Pachet, F., & Beurivé, A. (2005). Exploiting the tradeoff between precision and cpu-time to speed up nearest neighbor search. In *Proceedings of the international conference on music information retrieval (ISMIR)*.

Ruxanda, M. M., Nanopoulos, A., Jensen, C. S., & Manolopoulos, Y. (2008). Ranking music data by relevance and importance. In *Proceedings of the IEEE international conference on multimedia and expo (ICME'08)* (pp. 549–552).

Sahami, M., Mittal, V., Baluja, S., & Rowley, H. (2004). The happy searcher: Challenges in web information retrieval. In *Proceedings of the pacific rim international conference on artificial intelligence (PRICAI)*.

Yang, C. C., & Chen, K. Y. (2005). Retrieving multimedia web objects based on pagerank algorithm. In *Proceedings of the conference on world wide web (WWW)*.