MUSIC SIMILARITY ANALYSIS IN A P2P ENVIRONMENT

STEPHAN BAUMANN
German Research Center for Artificial Intelligence
Erwin Schrödinger Str.
67663 Kaiserslautern, Germany
E-mail: stephan.baumann@dfki.de

In this paper we present our work towards an integrated P2P platform for convenient music retrieval. The main goal is to provide a distributed computation and storage of basic music features from MP3 encoded files and/or sources on the web. In this way we try to resolve the problem of missing meta data in order to provide intelligent services for convenient music querying. We believe that such value-based features are essential for a consumer-oriented success of future music distribution services.

1. Introduction

The digital distribution of music is one of the most attracting and challenging topics for musicians and computer scientists these days. In despite of the ongoing legal debates about consumer behaviour and illegal file sharing services we find a lot of potential for convenient man-machine-interfaces to music on the technical side. The focus of this paper is the presentation of a new P2P research framework handling essential features for future commercial scenarios, e.g.:

- Online digital music distribution (e.g. Musicnet, Pressplay, etc.)
- Mobile services (UMTS services, mobiles with MP3 players, PDAs)

For these applications our approach meet some goals with respect to a maximum of convenient usability and a minimum amount of manual indexing of underlying large-scale musical data. We subsumed these objectives under the term *super-convenience* in our recent work [1]. The main issues of this concept handled in this paper are: (1) Fuzzy interpretation of misspelled search queries by phonetic matching [2], (2) automatic acquisition of music features using standards (e.g. MP3 [3], MPEG-7 [4], RDF-S), so-called cultural metadata [5] gathered by web crawlers and automatic audio analysis [6], (3) recommendations for similar songs by distance computation of spectral feature profiles (either from MP3 [7]. or cepstral coefficients [6,8]).

The paper is outlined as follows: we describe the benefits of a P2P architecture in Section 2, details about the individual processing components

delivering enhanced functionality are presented in Section 3, the conclusion is given in Section 4.

2. Peer-to-Peer Architecture

We will make some remarks why a mainstream music service should follow a P2P paradigmin order to be succesful. It is a very natural paradigm to process the content where it is, in this case at the private computers of the consumers. Distributed computing and storage power, bandwidth, fault-tolerance and reliability are the characteristics of such a network. In addition to this a P2P protocol offers the possibility to implement peer clustering by content to realize such issues as collaborative retrieval on top. This step offers the potential to have automatic and dynamic community building (e.g. special genre interest groups) over the time the network evolves. Furthermore the P2P paradigm is the obvious way to implement the idea of the semantic web. The collaborative efforts of volunteers or expert groups in indexing audio, resp. musical content with semantic tags offers the full potential of transparent processing or inferring about musical queries between computers and/or human beings.

Beside this technical and sociological point of view there comes an economic effect also into play. Value-based approaches to e-commerce offer a quantitative analysis how to cope with the problem of illegal file sharing. It seems to be possible in a reasonable way to convert special consumer groups of illegal P2P file sharing platforms to legal offers. Required incredentials are reduction of search time, superior quality of service and innovative interactive features. This can be reached if one enhances existing P2P networks with intelligent processing components such as automated similarity recommendations based on semi-automated meta-tagging and user-friendly query processing.

We decided to set up such a convenient music file sharing network by re-using the components of a former system which has been implemented in a modular way but deployed as a typical service on a central server satisfying requests of thin clients, namely web browsers. After evaluating different P2P protocols we decided to realize our new system by using the JXTA open source framework of Sun.

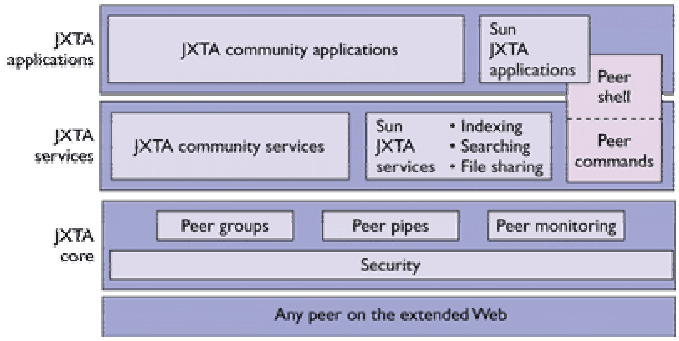

Figure 1. P2P Software Architecture (JXTA Framework © 2001 Sun Microsystems).

JXTA was choosen as the platform for developing our integrated approach because its main features support our concept and it offers several major advantages over other P2P protocols, namely handling firewall and NAT problems. A filesharing application with some basic features based on the JXTA protocol has been developed opensource in the MyJXTA2 project. We used straightforward the following functionalities of the client:

(1) Peer Groups: can be used to model different interest groups according to specific styles, genres or artist fan communities.

(2) Group Chat, 1:1 Chat: supporting the verbal communication about preferred music has been one of the cornerstones in our design of an intelligent P2P music platform since we aim at extracting cultural metadata from chats in the future.

(3) File Sharing: there is nothing special about this feature, but we realized that we would have a great benefit by adding our special additional tags (e.g. the audio profiles, extended metatags such as “similar artists”, etc.) into the existing MP3 file format which is supported at hand by its mime type.

(4) Metatag representation: by using ID3v2 as part of the MP3 file format we were able to supply the interesting musical features using a well-established format, having a critical mass of users, resp. standard applications such as MP3 players, playlist and cataloguing tools to read and store this format. In this way simple filesharing enters a totally new dimension of quality since each MP3 file can be interpreted as a self-contained music knowledge container.

(5) File Search: at this point we added our component for a fuzzy match of phonetic misspelled queries to correct entities.

(6) Presentation of metadata attributes and values: after a computation of similar songs by using the music similarity metrics we added the according

artists as a new meta tag. Furthermore standard ID3 tags such as artist and genre are presented.

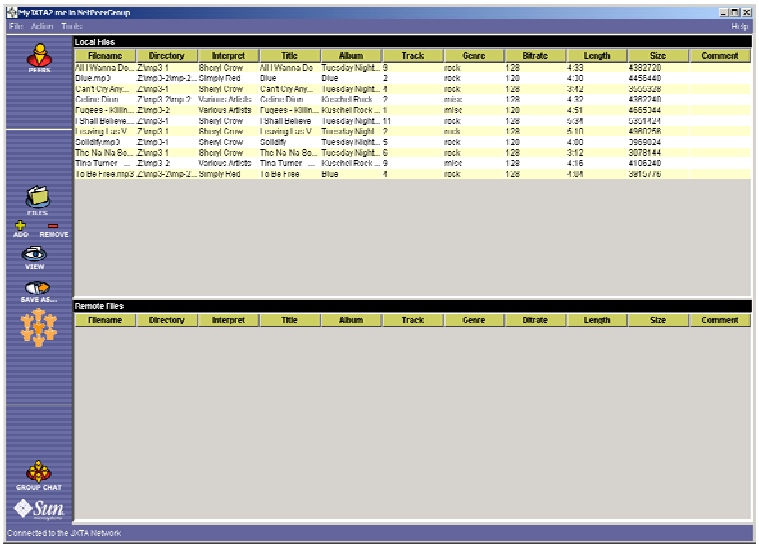

Figure 2. MyJXTA2 GUI with embedded features

Fig.2 shows the enhanced MyJXTA2 client where the most interesting metatags for music selection are presented and the recommendations for similar artists/ songs which are based on the computation of the audio similarity, resp. the usage of community metadata.

3. Details about the New Components

3.1. *Correction of phonetic misspellings*

The search field is not confused by entries containing typing errors. Additionally, the system is able to connect artist names which sound similar to each other, i.e. it is still able to produce results when there is phonetic similarity (such as e.g. „fil collins“ vs. „phil collins“).

Many general purpose sequence distance methods have been investigated. The phonetic fuzzy match used by our system is based on a PhD thesis on this subject 0. In contrast to other methods of non-exact search (such as Levensthein or Soundex) our method is optimized for application in the musical field. The

approach is used by a spinoff company (www.sonicson.com) in central server applications (e.g. www.musicline.de). In contrast to this the distributed manner of the P2P network allows for parallel processing of a misspelled query on several peers.

3.2. *Standard & Cultural Metadata*

The first try to receive meta information based on MP3 content is to take a look at the ID3 tags. Unfortunately, in real life data quality is insufficient for automatic processing. This is because meta information found in ID3 tags mainly come from databases like Gracenote or the FreeDB project. Those databases are generated by a large community of volunteers and their input – though very useful in many applications – is not quality assured. While the inconsistencies in artist, title and volume tags can be removed automatically, the genre information remains useless for automatic processing since the tagging is far too heterogeneous – if this ID3 tag is filled at all. In our work we aim at using high-quality data which can be found at sites such as www.allmusic.com. Furthermore there is hope that in near future data collections in standard formats will be merged and offer access to large-scale musical collections including high-level musical features. Meanwhile we make some experiments to cluster similar songs and artists by approaches using cultural metadata extracted from web crawls [5]. Previous work of others shows some potential to map these clusters to accepted genres or styles.

3.3. *Automatic Audio Analysis*

The automatic audio analysis recognizes properties about loudness, tempo and timbral features. These features can be used for the determination of music similarity. For the extraction of basic features such as loudness and psychoacoustic features, we used the approaches of Pfeiffer which have been compiled in a toolset under GPL license, available at CSIRO [3]. In parallel we use mel frequency cepstral coefficients and a clustering approach to generate audio profiles. The extracted features are stored as a feature vector in a special user-defined ID3 tag. In this way every peer makes a initial indexing of its files during setup or by entering entities missing this tag. In order to perform a similarity computation based on the audio profiles we use both the Earth Moving Distance, as suggested in [6] as well as a Gaussian Mixture Model. Even working with a standard Nearest Neighbor (NN) classifier delivered interesting results for cross-genre recommendations of music “sounding” similar. Subjective evaluation of these results have been performed by

professional musicians and non-musicians in a very limited scope since this process is very time-consuming. In the average 3 out of 5 top hits which we computed have been rated as similar in the subjective evaluations.

4. Conclusion

We proposed a P2P environment for the convenient sharing and discovering of music. The intelligent music processing components of this approach have been used so far mainly by the Music Information Retrieval community in central server approaches. In contrast a pure P2P framework for file sharing has been a common paradigm in the music domain in the past. The integration of the aforementioned components into the P2P framework delivers in our view the most powerful and natural approach to tackle technical and economic hurdles in the future.

References

1. S. Baumann, A. Klüter, Super Convenience for Non-Musicians: Querying MP3 and the Semantic Web, *Proceedings of the ISMIR 2002*, Paris, France, Oct 13-17, (2002).
2. A. Weigel , S. Baumann, A modified levensthein-distance for handwriting recognition*, Proceedings of the 7th International Conference on Image Analysis and Processing*, Bari, September, (1993).
3. S. Pfeiffer, T. Vincent, Formalisation of MPEG-1 compressed domain audio features, *Report No 01/96 of CSIRO Mathematical and Information Sciences,* Australia, Dec, (2001).
4. E. Allamanche, J. Herre, O. Hellmuth, B. Froeba, T. Kastner, M. Cremer, Content-based Identification of Audio Material Using MPEG-7 Low Level Description, *Proceedings of the ISMIR 2001*, Indiana University Bloomington, Indiana, USA, October 15-17, (2001).
5. B. Whitman., P. Smaragdis, Combining Musical and Cultural Feautures for Intelligent Style Detection, *Proceedings of the ISMIR 2002*, Paris, France, October 13-17, 2002.
6. B. Logan*,* Mel Frequency Cepstral Coefficients for Music Modeling *, Proceedings of the Int. Symposium on Music Information Retrieval (ISMIR) 2000*, Plymouth, USA, October 2000.
7. S. Baumann, A. Klüter, Using natural language and audio analysis for a human-oriented MIR system*, Proceedings of the Wedelmusic 2002*, Darmstadt, Germany, December 9-11, 2002.
8. G. Tzanetakis, Automatic Musical Genre Classification Of Audio Signals*, IEEE Transactions on Speech and Audio Processing,* July 2002.