Super-Convenience for Non-musicians: Querying MP3 and the Semantic Web

Stephan Baumann
German Research Center for AI (DFKI)
Erwin Schrödinger Str.
67608 Kaiserslautern
+49-631-205-3447
Stephan.Baumann@dfki.de

Andreas Klüter
sonicson GmbH
Luxemburger Str. 3
67657 Kaiserslautern
+49-631-303-2800
Andreas.Klueter@sonicson.de

ABSTRACT

Digital music distribution, the success of MP3 and the actual activities concerning the semantic web of music require for convenient music information retrieval. In this paper we will give an overview about the concepts behind our "super-convenience" approach for MIR. By using natural language as input for human-oriented queries to large-scale music collections we were able to address the needs of non-musicians. The entire system is applicable for future semantic web services, existing music web-sites and mobile devices. Beside the framework we present a novel idea to incorporate the processing of lyrics based on standard information retrieval methods, i.e the vector space model.

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 we find a lot of potential for convenient man-machine-interfaces to music on the technical side. Our long-term goal is the provision of a system architecture giving as much flexibility as needed to build powerful applications as customized instances of such an approach.

Our goal is to reach a maximum of convenient usability and a minimum amount of manual indexing of underlying large-scale data, we subsume this as *super-convenience*: (1) Human-oriented interface paradigm, (2) uniform feature handling and automatic metadata generation, (3) retrieval and recommendations.

Our overall approach is targeted to hybrid processing ranging from pure surface structure recognition to symbolic inferences among the concepts of the ontologies. As a unique novelty we present the seamless incorporation of lyrics in this approach in order to get – in the upper end - insight experiences about the perception of moods. We focus on naïve listeners or non-musicians in order to provide applications for the masses.

2. ONTOLOGICAL BACKBONE

The semantic web is on its way to enter the masses. Real killer applications may be convenient music information retrieval systems for naïve listeners. These contributions can be seen in the tradition of established standards such as MPEG-7. Indeed, authors report about successful transformations of MPEG-7 to the RDF (S) standard used for the semantic web [1]. Furthermore the collaborative effects of a broad user base can be used to make recommendations or computing the similarity between musical tracks. KANDEM is such an approach as described by the group at MIT media lab [2]. Answering real life questions of non-musicians requires real life knowledge in the music domain to be used within the MIR. For this purpose we modelled an ontology about the domain of music. In our application scenario we noted as terms the concepts of required know-how in the music domain. The relations consist of several types; is-a and part-of relations are used quite often. Is-a-relations are used to indicate specializations of concepts (e.g. *acid jazz is-a jazz*) while part-of-relations denote required parts (e.g. *track part-of compilation*). The aspect of sharing knowledge about conceptualizations with others is the most relevant aspect when building ontologies. In such a way different agents can share access to the semantic web of music. These activities are still in their infancies and the problems of the status quo are described thoroughly in a recent publication of Pachet [3]. At present our ontology is able to handle multiple inheritances for the concepts of tracks, albums and artists who outperforms standard subsumption hierarchies as found on many MP3-sites. Further concepts are the musical properties, which are linked to the automatic audio processing (i.e. *loudness, tempo, timbre*). As a novelty we introduced a semantic link *contains_lyrics*, which is grounded by the ASCII-text in our document database.

3. MUSIC DATABASE

The MIR system accesses the musical data from an underlying database. In our first prototype we ripped a private CD collection to MP3 format at 128kbps. The scope of this dataset is about 1000 tracks covering 60 artists and approx. 50 different genres. The administrative information about artist, title, and album has been gathered by usage of the CDDB. Unfortunately data quality was insufficient for automatic processing. While the inconsistencies in artist, title and volume tags could be removed; the genre information remained useless for automatic processing. Therefore the genre tags have been set manually. For the experiments at hand about 500 lyrics have been added as plain ASCII text.

4. AUDIO ANALYSIS AND NLP

The automatic audio analysis recognizes properties such as loud/quiet, fast/slow and MP3 subband features for the determination of similarity. For the extraction we used the approaches of Pfeiffer [4]. Natural Language Processing (NLP) approaches lie between the two extremes of *key word processing* (= disregard for word relations and context) and *complete understanding*. Both are not applicable for pragmatic processing of natural language music queries. The approach of *example-based processing with partial abstraction* is especially suited for music search requests (limited domain, high speed requirements) and offers an optimum trade-off between processing speed and good-natured reaction to off-scope requests. Our query interface in front of the NLP component is not confused by 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 in the past. The phonetic fuzzy match used by our system is based on former work at the German Research Center for Artificial Intelligence on this subject. Currently the phonetic fuzzy match is online for a large-scale music information system acting on 50.000 different artists accessible via a webservice at www.sonicson.com. An evaluation of recall and precision will be possible by logfile analysis in the future.

5. EXPLORING LYRICS

We started some experiments concerning the similarity of lyrics and the implications for the perception of music similarity. Here we use state-of-the-art document retrieval and classification approaches, which have been recently commercialized and successfully adopted to real-world problems. We used both, an API to a commercial tool as well as the text classification workbench and its submodules developed at our institute. We used the Protégé 2000 tool for convenient design of ontologies. The top-level concept lyrics is broken down into a taxonomy of typical topics covered by mainstream music. In the future such handcrafted topic ontology may be supported by semi-automatic ontology learning through document clustering approaches. For the current experiments we focused first on the “subsymbolic” level of lyrics. State-of-the-art document retrieval and classification approaches are still missing an in-depth ontological support. Nevertheless the basic techniques have a long-standing tradition in information retrieval and could be applied to the domain of lyrics. These tools allow for different functionalities. A query in the boolean retrieval model consists of a boolean combination of tests on the occurrence of specific words. For instance, the query *(hate or love) and girls* tests whether a document contains one of the words *hate* or *love* as well as the word *girls*. To go beyond the boolean retrieval, additional functionality, which we integrated, is based on the vector space model (VSM). In this model, lyrics as well as queries are represented as vectors. The dimension of the vectors indicate specific terms, the value of a vectors component indicates the number of times the respective term occurs in the lyrics/query to be represented. Defining a similarity measure between vectors does standard document retrieval based on queries in the VSM. The most frequently used measure here is the cosine-measure, which computes the angle between two vectors. Having a vector representing the query, the documents corresponding to the most similar document vectors are returned as answer documents. In this way we realize the computation of similarity among lyrics. Since queries and lyrics in the VSM are represented as vectors, also the similarity between vectors representing just lyrics can be computed. Roughly spoken, those lyrics, which share many important words, will have a high similarity. Computing the most relevant terms can perform a kind of summarization. As a further functionality the similarity between terms is computable allowing for automated term expansion and mapping to the taxonomy of topics in the music ontology. The lyrics collection contains 500 documents. While the querying for terms or topics is easy to perform, the more challenging approach is to examine term similarities or even document similarities. For the latter we show some typical results as stereotypes for the most common result cases of the approach in the following. For simplification we reduced the presentation on the 5 most-relevant terms of a given reference song and the top 3 similar songs by applying standard metrics of the vector space model.

- Song 193: Phil Collins - One More Night

Most-relevant terms: *forever wait night cos,* Similar: *P. Collins – You Cant Hurry Love, P.Collins - Inside Out, P.Collins - This must be Love*

- Reference Song 297: Cat Stevens - Father And Son

Most-relevant terms: *fault decision marry son settle,* Similar: *P.Collins - We're Sons Of Our Fathers, Sheryl Crow - No One Said It Would Be Easy, George Michael - Father Figure*

- Reference Song 112: Lucy pearl - Dance tonight

Most-relevant terms: *toast spend tonight dance money*, Similar : *Lucy Pearl - you (feat. snoop dogg and Q-tipp), Phil Collins - Please Come Out Tonight, Madonna - Into the groove.*

- Reference Song 56: Fanta4 - Das Kind Vor Dem Euch ...

Most-relevant terms: *wollten euch sehn entsetzt selben*, Similar: *Fanta4 - Auf Der Flucht, Freundeskreis - Mit Dir, Fanta4– Populär*

- Reference Song 145: Madonna - Paradise

Features: *remains pas encore fois moi,* Similar *: Zero Hits*

6. DISCUSSION AND FUTURE WORK

Non-musicians may query the musical database by remembering parts of the lyrics. In a recent evaluation with 100 naïve listeners we found this class of queries being essentially often used in a non-restricted user interface. The integrated approach can handle these queries. Some artists seem to cope with an overall theme on a complete album or even for a set of albums. Similarity metrics for term frequencies deliver appropriate results for these phenomena (see example 193). Some topics can be found across genre-boundaries (see example 297), which is indeed the intention for topic-based queries neglecting musical genres. Other topics are more often represented in specific genres (see example 112). *Dancy* music often talks about *dancing, parties, good vibes.* Specific vocabularies are typical for some very specific styles, e.g. German hip-hop (see example 56). This is a first impression, which has to be evaluated thoroughly in the future. Large corpora with multi-lingual entities are obviously necessary to cope with lyrics in different languages. Our initial corpus has been too small to cope with languages being different from English or German (see example 145). We still see a lot of potential in this kind of work if combined with the theory of *affective computing*. We could use lyrics and IR techniques to create automatically meaningful terms and topics. The emotional perception of such a topic (*war* vs. *peace*) may be coupled with the emotional perception of the audio surface structure (*minor* vs. *major*). In such a way the concept of *moods* [5] could be provided automatically for end-user queries. We presented the concept of *super-convenience* in this work for the first time. Our framework could be established by using cross-fertilization from different research disciplines, mainly in the area of AI. NLP, IR and Ontologies are the most prominent ones which have been incorporated in this work to get close to our initial goal.

7. REFERENCES

[1] Hunter J., Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology, in Int. SWWS, Stanford, July 30 - August 1, 2001

[2] Whitman B., KANDEM: Community Metadata for Informed Music Retrieval, MIT MediaLab, May 2002, Website <URL: http://web.media.mit.edu/~bwhitman/kandem/>

[3] Pachet F., A Taxonomy of Musical Genre, Proc. of ISMIR 2001, Paris, France, 2001

[4] Pfeiffer S., Vincent T., Formalisation of MPEG-1 compressed domain audio features, Technical Report Number 01/196, CSIRO, Australia, 2001

[5] Huron D., Perception and Musical Applications in Music Information Retrieval, in Proc. of the ISMIR 2000, Plymouth, Massachusetts, 2000