

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

Electronic music distribution, the internet success of MP3 and the actual activities concerning the semantic web of music require for convenient music information retrieval, resp. question-answering systems. 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 future electronic devices such as cd-changers for cars, or PDAs. It is a full-fledged architecture combining state-of-the-art approaches from different research disciplines. We customized in a cross-discipline approach techniques from natural language understanding phonetic matching, automatic analysis of audio for meta tag construction, content-based classification and music ontologies as a backbone for the representation of musical knowledge. Beside the basic framework we present a novel idea to incorporate the processing of lyrics based on standard information retrieval methods, i.e the vector space model.

This work has been performed at the German Research Center for AI and the authors spin-off company – sonicson -, specialized in music web 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 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 our research framework, which has been partially applied to real-world scenarios.

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 a multi-component approach. In this way our system approach can be applied for the main application areas in this field:

- Digital music distribution [1] (e.g. Musicnet, Pressplay)
- Online retailers [2] (e.g. Amazon, JPC)
- Music information services [3] (e.g. All-music-guide)
- Retrieval and playlist generation for hardware music devices
- Mobile services (future UMTS services, PDAs)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2002 IRCAM – Centre Pompidou

For all of 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 subsume these objectives under the term *super-convenience*:

- Human-oriented interface paradigm: utilizing the expressive power and simplicity of natural language requests
- Fuzzy interpretation of misspellings: phonetic matching
- Beginner and expert handling: expressiveness, transparent inferring from symbolic to sub symbolic concepts
- Uniform feature handling: administrative metatags, lyrics, symbolic and sub symbolic audio features
- Automatic acquisition of features: using standards (e.g. MP3, MPEG-7), web crawling, automatic audio and text analysis
- Retrieval, recommendations: similarity metrics and classification in a VSM model (Vector Space Model)
- Usage of extendible common-sense knowledge about music: music ontology alignment

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. ARCHITECTURE

The functional elements of the MIR system BEAGLE are realized

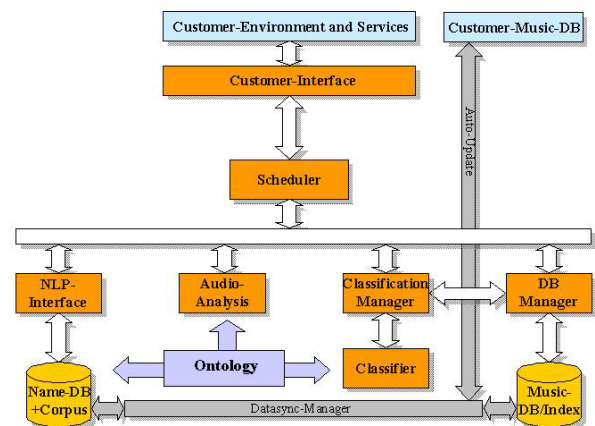


Figure 1. System Architecture.

in the system design as independent components. Figure 1 shows the interrelationship of *classifier*, *audio analysis* and *natural language interpretation*.

The components *DB manager* and *customer interface* implement flexible interfaces that simplify the integration work in existing customer environments (eCommerce solutions, databases, etc.). The service module *scheduler* performs the executive control of the whole system as well as the quality of service management for simplified scalability of the MIR system. The music data of the customers are accessible in the *music-DB*. *Name-DB + corpus* determine the country language and the natural language processing. They are interwoven with the language-independent musical concepts specified in the ontology.

A possible query “like lucky star by Madonna, but faster” is processed as follows: Initiated by the *Customer Environment and Service* and handed over by the *Customer Interface*, any query is passed to the *Scheduler*. An instance of the *NLP-Interface* identifies the semantic parts of the input sentence: (a) DB-ID: “madonna/lucky_star”, (b) result should be similar to “lucky star” and (c) but faster. Since a similarity fingerprint of that was already extracted in a preprocessing step, the *Audio Analysis* component is not employed in this example. Based on its’ internal know-how and information from the *DB Manager*, the *Classification Manager* extracts a set of matching titles. To finish the query, the results are passed back to the user via *Scheduler* and *Customer Interface*.

At this point our extension to lyrics comes into play: future similarity measurements may be based also on lyrics. A user is allowed to enter queries such as “I am interested in songs about sadness in a slow tempo” or “I need something really hot about summer”. Beside the pure MP3 data the text contained in the lyrics files are also taken into account for both: classic retrieval purposes and furthermore “textual similarities”.

In the following subchapters the individual components are highlighted in short, chapter 3 describes the novel lyrics component.

2.1 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 [4]. Furthermore the collaborative effects of a broad user base can be used to enrich or even substitute the current techniques for making recommendations or computing the similarity between musical tracks. *KANDEM* is such an approach as described by the research group of B.Vercoe at MIT media lab [5]. Further interesting cross-fertilization can be seen if one looks at the peer-to-peer activities in the context of the semantic web. Distributed processing power may be an option for shared signal processing or collaborative agent frameworks on top of the semantic web [6].

But the driving force behind our decision to put an ontology at the heart of the system has been the following: Answering real life questions 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 this sense we apply to a very basic definition of ontologies:

“An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary”.

In our application scenario we noted as terms the concepts of required know-how in the music domain. The relations consist of several types, namely 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*, *organ* is-a *keyboard*) while part-of-relations denote required parts (e.g. *track* part-of *compilation*, *member* part-of *band*). In this way we were able to set up an initial ontology for further refinement and alignment with other existing ontologies. The aspect of sharing knowledge about conceptualizations with others is indeed the most relevant aspect today. While in the beginning only our own agents were able to talk with each other through the ontology, for the future different agents will share access to the semantic web of music. These activities are still in their infancy but the power of collaborative work behind these forces will surely lead to large-scale musical data [7]. The encountered problems in this process are described thoroughly in a recent publication of Pachet [8]. He focuses on the aspect of genre taxonomies, which is the most difficult but also most important part when defining a music ontology. A further difficult question is how to handle the acquisition of instances, which represent beside the generic concepts the actual state of the “musical world”. This will be an ongoing and dynamic process since “music never stops”: new artists, new releases, the latest trends in genres, etc. pop up at different corners of the world and have to be assembled.

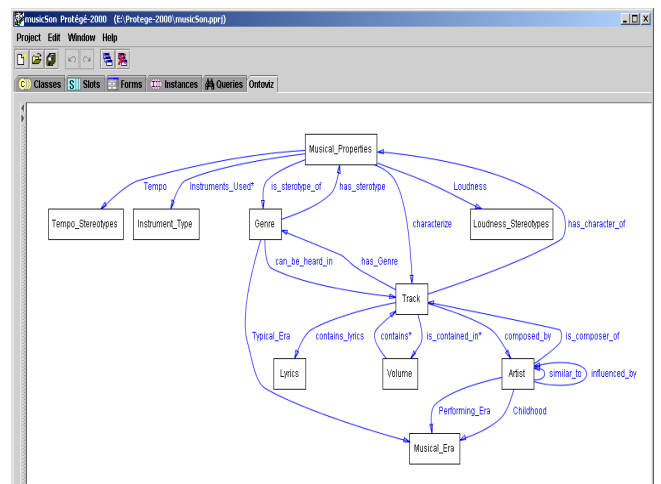


Figure 2. Excerpt of ontology (top-level).

In Figure 2 the top-level of our ontology is presented. It 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/main instruments). As a novelty we introduced recently a semantic link *contains_lyrics*, which is grounded by the ASCII-text in our document database.

Technically we use the Protégé 2000 tool for convenient design of ontologies. This tool and the underlying approach have been developed at Stanford University, mainly driven by needs in medical departments. Protégé 2000 is a cross-platform tool under widespread usage in different areas. It is mainly its flexible plug-in-architecture and the powerful import and export of standard formats such as XML and RDF (S) making it the tool of choice when it comes to ontology editors. Beside the definition of classes, automatically generated form editors allow the definition of the instances.

Especially the options to import other know-how domains via importing RDF (S) files and the open plug-in architecture have been the major reasons to choose PROTÉGÉ 2000 [9].

2.2 Grounding by 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 simple database 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. The genre tags have been set manually.

For the experiments at hand about 500 lyrics have been added as plain ASCII text by linkage to the database.

2.3 Intelligent agents for information acquisition and processing

A bunch of different components is used to bridge the gaps between user's utterances and intended needs to the concepts in the ontology and finally to access the musical data in the database. In this paper we give only a short overview since our focus is mainly the novel approach handling lyrics.

2.3.1 Usage of ID3 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 CDDB or the FreeDB project. A large community of volunteers generates those databases 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; the genre information remains useless for automatic processing since the tagging is far too heterogeneous – if this ID3 tag is filled at all.

2.3.2 Automatic audio analysis

The automatic audio analysis recognizes properties such as loud/quiet, fast/slow etc, as well as more sophisticated features for the determination of 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 [10]. For the purpose of tempo analysis we developed our own approach, which consists of techniques described in [11]. The extracted features are stored as a feature vector in the database. More sophisticated approaches have been developed by the group of Brandenburg [12] and Scheirer at MIT media lab [13].

2.3.3 Natural language interpretation

Approaches to processing natural language lie between the two extremes *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* [14] 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 (requests whose complexity reach the limits of the machine processing of natural language).

2.3.4 Correction of phonetic misspellings

Our query interface 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. The phonetic fuzzy match used by our system is based on former work at the German research center for artificial intelligence on this subject [15]. In contrast to other methods of non-exact our method is optimized for application in the music field.

2.3.5 Content-based retrieval and classification

Large music databases should provide efficient and easy access. So-called content-based music retrieval and classification is a well-known research topic being explored by different authors in the past. Still the most challenging task is the selection of the most appropriate low-level features in combination with a well-suited similarity measure, resp. classification approach. An excellent overview of related work and promising results could be found in the dissertation of E. Scheirer [13].

Our architecture may be used as a testbed for different classifiers being coordinated or voted by a classification manager, which synchronizes the different results. Actually we are working with a standard Nearest Neighbor (NN) classifier to deliver cross-genre recommendations for music “sounding” similar. The basic audio features are MFCCs, bpm and loudness. A temporal clustering is performed in combination with the NN classifier. In parallel we are working on the evaluation of support-vector-machines acting on our feature set as trainable genre-classifiers [16].

3. 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 [17].

3.1 Lyrics and our music ontology

We used the Protégé 2000 tool for convenient design of ontologies as mentioned in chapter 2.1. 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.

Technically our document management system is based on a standard ODBC-Database, here Informix/Solaris. The database tables reflect meta information such as song title, author and text body. The unstructured full-text document content is stored as plain ASCII in the file system and accessible via document filename identifiers.

3.2 Document retrieval and classification tools

State-of-the-art document retrieval and classification approaches are still missing an in-depth ontological support, which is described in our approach for a non-music scenario [18]. 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*. In this way a lot of the users questions about the content of a song can be handled.

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.

3.3 Some examples

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.

- Reference Song 193: Phil Collins - One More Night.txt

Most-relevant terms: *forever wait night cos ooh*

1. Similar Lyrics: Phil Collins - YOU CAN'T HURRY LOVE.txt
2. Similar Lyrics: Phil Collins - Inside Out.txt
3. Similar Lyrics: Phil Collins - This must be Love.txt

- Reference Song 297: Cat Stevens - Father And Son.txt

Most-relevant terms: *fault decision marry son settle*

1. Similar Lyrics: Phil Collins - We're Sons Of Our Fathers.txt
2. Similar Lyrics: Sheryl Crow - No One Said It Would Be Easy.txt
3. Similar Lyrics: George Michael - Father Figure.txt

- Reference Song 112: Lucy pearl - Dance tonight.txt

Most-relevant terms: *toast spend tonight dance money*

1. Similar Song : Lucy Pearl - you (feat. snoop dogg and Q-tipp).txt
2. Similar Song: Phil Collins - Please Come Out Tonight.txt
3. Similar Song: Madonna - Into the groove.

- Reference Song 56: Das Kind Vor Dem Euch Alle Warnten.txt - die fantastischen vier

Most-relevant terms: *wollten euch sehn entsetzt selben*

1. Similar Song: Die fantastischen Vier - Auf Der Flucht.txt
2. Similar Song: Freundeskreis - Mit Dir.txt Artist:
3. Similar Song: Die fantastischen Vier - Populär

- Reference Song 145: madonna - Paradise.txt

Features: *remains pas encore fois moi*

Zero Hits

4. DISCUSSION

The novel approach for similarity based on lyrics showed at a first glance some opportunities but also severe problem areas. The approach offers possibilities to enhance the “super-convenience” of our system.

- Non-musicians may query 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 whole bunch of their 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.
- Some 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. The corpus, resp. document collection has a great influence on the effects of lyrics similarity. Different corpora and term-weighting metrics have to be evaluated for this purpose.
- 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 information retrieval techniques to create automatically meaningful terms and topics. The emotional perception of such a topic (*war vs. peace, love vs. hate*) may be coupled with the emotional perception of the audio surface structure (*major vs. minor*, etc). In such a way the concept of *moods* (as characterizations in the tow-dimensional space of *valence* and *arousal*) [19] could be provided automatically for end-user queries instead of using labor-intensive manual tagging of songs with metadata.

5. FUTURE WORK

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 artificial intelligence. Natural Language Processing, Information Retrieval and Ontology (Semantic Web) issues are the most prominent ones which have been incorporated in this work to get close to our goal.

The presented work is still on its way. The novel lyrics tool has only been tested in isolation and has to be integrated into our system in the future. Our initial tests with students interested in music showed how people liked to play around with such a tool. This seems to be an indicator for its importance to create “stickiness” in future digital music services.

Our future work will remain two-folded. State-of-the-art research results will be evaluated partially in real-life scenarios. This can be done on-demand in cooperation with the spinoff company sonicson. Currently the phonetic fuzzy match is online for a large scale music information system [20].

The system architecture is flexible enough to incorporate new agents and enhancements of the underlying knowledge representation. Especially the latter is open for different formats, which seems very important to stay on the track with future developments in the scope of MPEG-7, MPEG-21, resp. the semantic web.

6. ACKNOWLEDGMENTS

Our thanks to H.P. Guebert for fetching lots of the lyrics and providing substantial help to conduct the experiments at sonicson GmbH.

7. REFERENCES

- [1] www.musicnet.com
- [2] www.amazon.com
- [3] www.allmusic.com
- [4] Hunter J., *Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology*, in International Semantic Web Working Symposium (SWWS), Stanford, July 30 - August 1, 2001
- [5] Whitman B., *KANDEM: Community Metadata for Informed Music Retrieval*, Project Description MIT MediaLab, May 2002, <http://web.media.mit.edu/~bwhitman/kandem/>
- [6] Neidl W., Wolf B., Qu Ch., Decker S., Sintek M., Naeve A., Nilsson M., Palmer M., Risch T., *Edutella: A P2P Networking Infrastructure Based on RFD*, in Proceeding of the 11th International World Wide Web Conference, Honolulu, USA, May 2002.
- [7] Swartz A., *MusicBrainz: A Semantic Web Service*, in IEEE Intelligent Systems, Intelligent Web Services, January/February 2002.
- [8] Pachet F., *A Taxonomy of Musical Genre*, Proceedings of ISMIR 2001, Paris, France, 2001
- [9] <http://protege.stanford.edu>
- [10] Pfeiffer S., Vincent T., *Formalisation of MPEG-1 compressed domain audio features*, Technical Report Number 01/196, CSIRO Mathematical and Information Sciences, Australia, December 2001.
- [11] Wang Y., Vilermo M., *A compressed domain beat detector using MP3 audio bitstreams*, Proc. ACM Multimedia 2001, Sep.30-Oct 5, Ottawa, Ontario, Canada, pp. 194-202, 2001
- [12] Allamanche E., Herre J., Hellmuth O., Froeba B., Kastner T., Cremer M., *Content-based Identification of Audio Material Using MPEG-7 Low Level Description*, Indiana University Bloomington, Indiana, USA, October 15-17, 2001
- [13] Scheirer E., *Music Listening Systems*, Ph.D. Thesis, MIT Media Lab, June 2000
- [14] Wahlster W. (Ed.), *VerbMobil: Foundations of Speech-to-Speech Translation*, Springer, Berlin, 2000
- [15] Weigel A., Baumann S., *A modified levensthein-distance for handwriting recognition*, in Proceedings of the 7th International Conference on Image Analysis and Processing, Bari, September 1993.
- [16] Joachims T., *Text Categorization with Support Vector Machines: Learning with many relevant Features*, Proceedings of the 10th European Conference on Machine Learning, pp. 137-148, Chemnitz, Germany, April 1998
- [17] Junker M., *Heuristisches Lernen von Regeln für die Textkategorisierung*, Ph. D. Thesis University of Kaiserslautern, Germany, March 2001
- [18] Baumann S., Dengel A., Junker M., Kieninger T., *Combining ontologies and document retrieval techniques: a case study for an e-learning scenario*, in Third International Workshop on Theory and Applications of Knowledge Management, TAKMA 2002 In Conjunction with DEXA 2002, Aix-en-Provence, France, September 2 - 6, 2002
- [19] Huron D., *Perception and Musical Applications in Music Information Retrieval*, in Proceedings of the ISMIR 2000, Plymouth, Massachusetts, October 23-25, 2000
- [20] http://www.sonicson.com/german/03_products/products.html