WASABI: a Two Million Song Database Project with Audio and Cultural Metadata plus WebAudio enhanced Client Applications

Gabriel Meseguer-Brocal, Geoffroy Peeters, Guillaume Pellerin, Michel Buffa, Elena Cabrio, Catherine Faron Zucker, Alain Giboin, Isabelle Mirbel, Romain Hennequin, Manuel Moussallam, et al.

WASABI: a Two Million Song Database Project with Audio and Cultural Metadata plus WebAudio enhanced Client Applications

Gabriel Meseguer-Brocal, Geoffroy Peeters, Guillaume Pellerin
IRCAM
France
1st.lastname@ircam.fr

Michel Buffa, Elena Cabrio, Catherine Faron Zucker, Alain Giboin, Isabelle Mirbel
Université Côte d'Azur, CNRS, INRIA
Nice, France
1st.lastname@i3s.unice.fr

Romain Hennequin, Manuel Moussallam, Francesco Piccoli
DEEZER
France
rhennequin@deezer.com

Thomas Fillon
PARISSON
France
thomas@parisson.com

ABSTRACT

This paper presents the WASABI project, started in 2017, which aims at (1) the construction of a 2 million song knowledge base that combines metadata collected from music databases on the Web, metadata resulting from the analysis of song lyrics, and metadata resulting from the audio analysis, and (2) the development of semantic applications with high added value to exploit this semantic database. A preliminary version of the WASABI database is already online[1] and will be enriched all along the project. The main originality of this project is the collaboration between the algorithms that will extract semantic metadata from the web and from song lyrics with the algorithms that will work on the audio. The following WebAudio enhanced applications will be associated with each song in the database: an online mixing table, guitar amp simulations with a virtual pedalboard, audio analysis visualization tools, annotation tools, a similarity search tool that works by uploading audio extracts or playing some melody using a MIDI device are planned as companions for the WASABI database.

1. INTRODUCTION

Streaming professionals such as Deezer, Spotify, Pandora or Apple Music enrich music listening with biography and offer suggestions for listening to other songs/albums from the same or similar artists. A journalist or a radio presenter uses the Web to prepare his programs. A professor in a sound engineering school will use analytical tools to explain the production techniques to his students. These three scenarios have one thing in common: they use knowledge bases ranging from the most empirical — the result of a keyword search in Google — to the most formalized and programmatically usable through a REST API — such as Spotify's use of LastFM, MusicBrainz, DBPedia and audio extractors from The Echo Nest. Then, the need for more precise musical knowledge bases and tools to explore and exploit this knowledge becomes evident.

The WASABI project (Web Audio Semantic Aggregated in the Browser for Indexation), started in early 2017, is a 42-month project founded by the French National Agency for Research (ANR) which aims to answer this need. The partners in this project are the I3S laboratory from Université Côte d'Azur, IRCAM, DEEZER, and the Parisson company. Other collaborators are Radio France (journalists, archivists), music composers, musicologists, music schools, sound engineering schools. The primary goal of this multidisciplinary project is to build a 2 million song knowledge base that contains metadata collected from music databases on the Web (artists, discography, producers, year of production, etc.), from the analysis of song lyrics (What are they talking about? Are locations or people mentioned? Which emotions do they convey? What is the structure of the song lyrics?), and from the audio analysis (beat, loudness, chords, structure, cover detection, source separation / unmixing, etc.). A preliminary version of the WASABI database is already online and will be enriched all along the project (https://wasabi.i3s.unice.fr). Figures 1 and 2 show the current version of the WASABI database front-end. The database comes with a complete REST API[2] and will include a SPARQL endpoint.

The WASABI project has the originality to use the algorithms from the Music Information Retrieval research domain that work on the audio content together with the Web of Data / Semantic Web and plain text analysis algorithms

[1] https://wasabi.i3s.unice.fr

[2] https://wasabi.i3s.unice.fr/apidoc/

Web Audio Conference WAC-2017, August 21–23, 2017, London, UK.

to produce a more consistent knowledge base. WebAudio client applications will then take benefit of this huge amount of integrated heterogeneous knowledge. Semantic Web databases can be used to extract cultural data, linking music to elements such as producer, recording studio, composer, year of production, band members, etc. Free text data such as song lyrics or the text of pages linked to a song can be used to extract non-explicit data such as topics, locations, people, events, dates, or even the emotions conveyed. The analysis of the audio signal enables to extract different kinds of audio descriptors for music such as beat, loudness, chords, emotions, genre, and the structure of a song. These extractions are prone to errors and uncertainties and we aim at improving them by combining the extracted information with the knowledge extracted from the Semantic Web and from the song lyric analysis: for example the temporal structure, the presence and characterization of the voice, the emotions, and to spot covers / plagiarism, even to facilitate the unmixing. In the case of music that can be accessed in unmixed form (separate tracks), a more accurate audio analysis can be performed and richer data can be extracted (notes, instruments, type of reverberation, etc.).

The WASABI project will specify the different use cases thanks to the presence of users of our future search results: Deezer, Radio France, music journalists, composers and musicologists. For this purpose, WASABI will offer a suite of open source software packages and open-data services for:

- visualization of audio metadata from Music Information Retrieval, listening to unmixed tracks in the browser using the WebAudio API (real-time mixing, audio effects, audio analysis),
- automatic song lyric processing, named entity recognition and binding, annotation and collaborative editing,
- access to Web services with an API offering a musical similarity study environment combining audio and semantic analyzes.

These software bricks will be used to develop formalized demonstrators with our partners and collaborators using the WebAudio API standard and allowing the development of music applications accessible to the public from a Web browser.

The rest of the paper is organized as follows: Section 2 presents the context of the WASABI project and related work. Section 3 presents the research questions and objectives of the project. Section 4 presents the first results of the project. Section5 concludes.

2. CONTEXT AND STATE-OF-THE-ART

In the field of extracting music information from the audio signal, there is little previous work on the sharing of information from the Semantic Web, speech analysis and audio altogether. The Million Song Dataset project ran a classification mainly based on audio data [2], but did not take advantage of the structured data available for example on DBpedia, to remove some uncertainties. Information such as group composition or orchestration can be very relevant to informing unmixing algorithms, but is only available in certain data sources (BBC, MusicBrainz, ...), and for many little-known artists this information is not available. It is

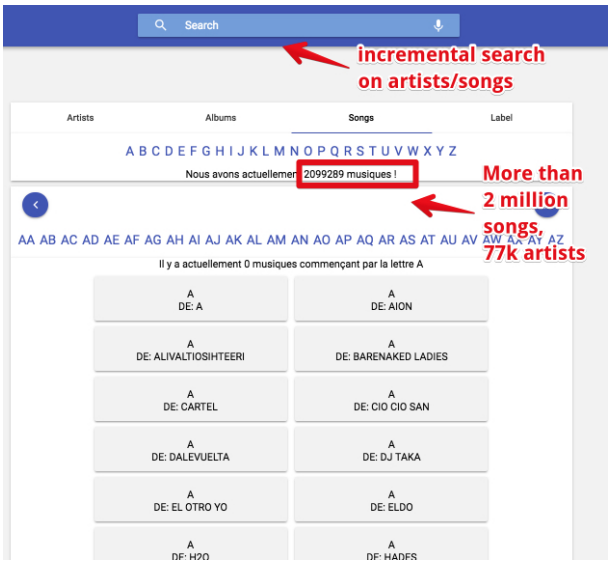

Figure 1: WASABI database search GUI.

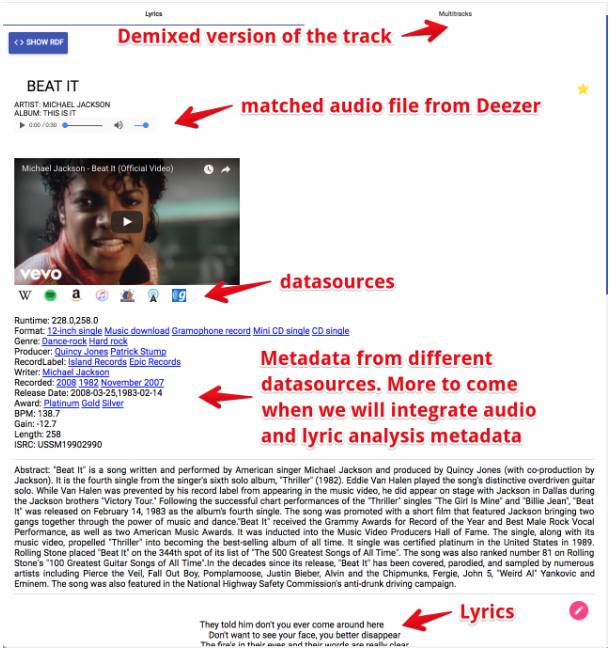

Figure 2: An example of an entry in the 2M song WASABI database.

here that the collaboration between audio and semantics finds its meaning, one taking advantage of the other.

Research centers such as the C4DM of Queen Mary University of London have developed collaborations with the BBC in the use of RDF and proposed musical ontologies in several fields (including audio effects and organology - see semanticaudio.ac.uk-). On the other hand, companies such as Spotify (having acquired The Echo Nest), Grace Note, or Pandora (having recently engaged several audio researchers), Apple Music (which has just created its audio research team in London) have developed some expertise in this area but that is not being returned to the public domain, unfortunately. MusicWeb [5] links music artists within a Web-based application for discovering connections between them and provides a browsing experience using connections that are either extra-musical or tangential to music. It integrates open linked semantic metadata from various music recommendation and social media data sources including DBpedia.org, sameas.org, MusicBrainz, the Music Ontology, Last.FM and YouTube as well as content-derived information. The project shares some ideas with WASABI, but does not seem to address the same scale of data, and does not perform analyses on the audio and lyrics content.

The WASABI project will allow scale-up on a wider scale than the Million Song Dataset (through the sharing of the Deezer audio base and the I3S Semantic Web / database built from the Web of data and lyrics analysis) with public domain development of open source tools. It will also draw on the results and developments of two previous projects funded by the ANR: WAVES[3] and DIADEMS[4], the last ones using the Telemeta open source music platform.

3. RESEARCH QUESTIONS AND OBJECTIVES

3.1 Identification and evaluation of semantic and textual / cultural / audio datasources

3.1.1 Semantic datasources that can be used to enrich the initial database

This scientific objective aims to (1) feed MIR algorithms with data coming from the Web of data and enabling to bootstrap initial parameters (for example, if we know the producer, the artist profile, the band members and the instruments they used on a particular song, we can search for a guitar signal directly), and (2) enrich data collected from the Web of data with information extracted from MIR algorithms (for example, if little or no metadata are available on the Web about some artists/bands/songs, no Wikipedia page, nearly empty MusicBrainz entry, etc. but the audio analysis detected drums/bass/guitar). The heart of the WASABI project is its knowledge base, that will be enriched all along the project, both with new metadata and with new models/ontologies.

We identified several datasources that will provide data with different structuring levels, from plain text to semantic databases, that will be used to build the WASABI knowledge base. Among them we can distinguish:

- Online datasets: the Deezer database is an aggregation of data coming from music companies/labels (40M songs); it is up to date, but suffers from a lack of coherence: the level of details changes a lot from one music label to another, semantic variations are important, classification is inaccurate (terms such as “pop”, “rock”, “electro” are too generic), and produces lots of metadata collisions (e.g. homonyms between artists). The Million Song Dataset is rich in audio metadata, but does not focus on cultural metadata. The Linked Data Catalog references also other datasets that we will evaluate (date of the last update, quality of metadata).

- Online databases with a REST API: DBPedia (different versions from each country), musicbrainz.org (free musical encyclopedia, that collects musical metadata), last.fm and libre.fm, seevl.fm, musicmight.com (hard rock/metal database), discogs.com (one of the biggest online database), soundfacts.org, etc.

These databases provide song/artist metadata with a lot of variability in their content: some musical genres are more represented than others, different DBpedia chapters may have complementary or conflicting contents (it happens that DBPedia.org proposes more content than DBPedia.it even about an Italian band). An important challenge in this project is the identification of relevant metadata, merging them and resolving conflicts (people names with different abbreviations, etc.), and curating them. We will compare some metadata obtained from these databases with the results from pure audio analyses by IRCAM’s algorithms.

3.1.2 Audio datasources

We mainly rely on the stereo audio files provided by Deezer but also on a set of unmixed songs from sound engineer schools, from recording studios, and from the Radio France archives (thousands of artists recorded live in multitrack format). We also use audio data from 7digital and YouTube to complete the dataset.

Matching these audio titles (2 million stereo tracks and about two thousand unmixed tracks) with the current WASABI database entries is an ongoing task that should be completed by June 2017. We already matched 87 percent of the songs. Our preliminary results are further described in Section 4.

3.1.3 Vocabularies

We will need to use/reuse one or more vocabularies for the project database, and define the WASABI vocabulary as a set of complementary vocabularies. As a first step, we are conducting interviews with final users, to define their needs, while a state of the art about vocabularies and models in the musical industry is being conducted. We adopt an incremental approach driven by the user needs: we will integrate new metadata and update the vocabularies we use as new user needs appear (directly from users, or specific to the tools we are developing). In particular, we already identified the following needs that should be modelled in the WASABI vocabulary:

- the segmentation of a music track in its temporal structure (verse, chorus, etc),

- the segmentation of a music track into singing parts and the characterization of its vocal quality,

[3]http://wave.ircam.fr/

[4]https://www.irit.fr/recherches/SAMOVA/DIADEMS/

- the estimation of a music track emotion,
- the identification of cover (or plagiarism) versions,
- the use of prior information (such as the organological composition of a group obtained from the Semantic Web) to help source separation algorithms,
- the use of multi-channel audio (either clean or estimated) to improve content-estimation.

This project will also give us the opportunity to work on the definition of an ontology for describing efficiently a song over time.

Yves Raimond's papers about the BBC dataset are precursors on the way one can build a cultural dataset about music [7]. Actually, a large research project, the SAFE project[5] is also addressing this topic [1, 12]. The Music Ontology Specification provides the main concepts and properties for describing music (i.e. artists, albums, songs, lyrics,tracks), schema.org also contains a vocabulary about music, built by the main search engine providers. Music is a vocabulary for classical music and stage performances. The SAFE project also created an ontology for classifying sounds (hot, fuzzy, cold, etc.). IRCAM has developed several vocabularies for musical feature extraction, in particular for the genre classification and labeling, mood, instrumentation/orchestration and rhythmic analysis (metric, tempo and time and bar segmentation). The Million Song Dataset has developed vocabularies for genre classification too [8].

3.2 Song lyric analysis and consolidation with data from audio analysis

As introduced before, the goal of the WASABI project is to jointly use information extraction algorithms and the Semantic Web formalisms to produce more consistent musical knowledge bases. Then, Web Audio technologies are applied to explore them in depth. More specifically, textual data such as song lyrics or free text related to the songs will be used as sources to extract implicit data (such as the topics of the song, the places, people, events, dates involved, or even the conveyed emotions) using Natural Language Processing algorithms. Jointly exploiting such knowledge, together with information contained in the audio signal can improve the automatic extraction of musical information, including for instance the tempo, the presence and characterization of the voice, musical emotions, identify plagiarism, or even facilitate the music unmixing.

With respect to the song lyric analysis, we have identified the following tasks that we will address by extending existing approaches in Natural Language Processing (NLP) and adapting them to the music domain.

Structure detection: identification of the structure of the song (e.g. intro, verse, refrain), following [4].

Event detection: analysis of the text of the song to extract the context (both explicit or implicit references), as well as the extraction of entities, geographic locations and time references directly or indirectly expressed in the text. For instance, being able to link the "We came down to Montreux" (from the song "Smoke on the Water") to the Jazz festival of Montreux. One of the challenges will be to deal with the abundant use of metaphor in text lyrics, that we will address by combining resources describing them (whenever possible), and NLP approaches to metaphor detection [9].

Topic modelling: implementation of probabilistic models to identify topics or abstract themes in the lyrics by establishing relationships between a set of documents and the terms they contain [4, 10].

Sentiment analysis: classification of the emotions in the songs, using both music and song lyrics in a complementary manner. We will test and adapt machine learning algorithms to capture information of emotions expressed by the text of a song, exploiting both textual features, and data extracted from the audio [6].

Plagiarism detection: measure the similarity degree of the texts of two songs, and identification of recurrent structures, paraphrases, multi-sources plagiarism [11, 3]. An interesting aspect to explore is the combination of similarity measures calculated over both the text (both lexical and syntactic similarity) and the audio features in particular to detect the plagiarism among songs in different languages. For instance, if there is a high similarity in the audio features, we benefit from machine translation methods to calculate also the textual similarity among texts in different languages. Moreover, multilingual resources such as DBpedia could be exploited to recognize the presence of the same entities in the lyrics.

3.3 Improving music information retrieval using both audio, lyrics and semantic Web

One of the main scientific questions that WASABI will try to answer is "how to benefit from various knowledge sources on a given music track to improve its description". In WASABI, we will study the use of three sources of knowledge: the audio content of a music track (this is often denoted by content-information), the lyrics for non-instrumental tracks and semantic information taken from the Web (such as from DBpedia, MusicBrainz, Discogs or others). To demonstrate that using the three sources of information is beneficial, we will study six different use cases of their joint usage:

1. Music structure estimation (i.e. the estimation of the position in time of the verse, chorus or bridge) based on audio repetition and the structure of the lyrics
2. Singing voice segmentation over time and characterization based on audio content (using prior on singer characteristic) and the structure of the lyrics
3. Music emotion estimation using audio, lyrics analysis and semantic Web
4. Cover-version identification using prior on lyrics
5. Informed audio source separation (using band composition prior from DBpedia or others)
6. Instrument recognition using separated sources

Additionally, the unmixed dataset (separated audio tracks) provides a rich environment for MIR problems. From these tracks, it is possible to have access individually to the different instruments or even to specify combinations of them. This scenario is ideal for improving tasks such as melody and harmonic extraction, singing voice extraction, etc. subproblems of our six use cases.

[5]http://www.semanticaudio.co.uk/

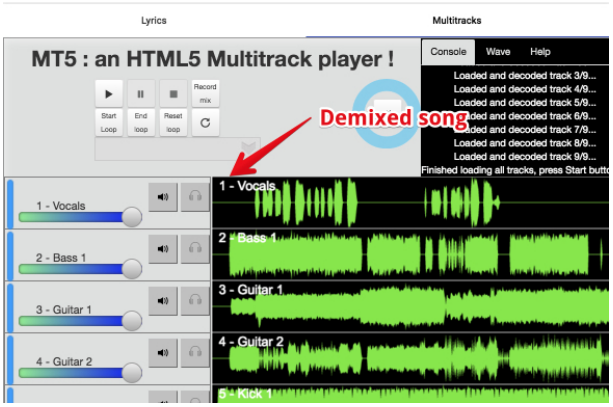

Figure 3: WASABI embeds WebAudio applications such as a multi-track audio player.

Figure 4: A guitar tube amplifier simulation made with WebAudio. The WASABI database embeds tools for re-creating/studying the sound of the instruments played in classic rock songs.

3.4 Innovative user experience

We have many contacts with composers, musicologists, journalists from Radio France and Deezer, who want to use the technologies being developed, both on the back office (for archivists) and front office (for final users). Parisson will propose the definition and use of multi-track interfaces to the communities of researchers and teachers for which it is already developing collaborative Web tools for automatic, semi-automatic and manual audio indexing: the Telemeta platform (figure 5). Telemeta is based on the audio analysis Web framework TimeSide[6]). Wimmics/I3S, IRCAM and Parisson have already published or presented demonstrations in the previous Web Audio Conferences and have produced innovative products with this technology, that will be included in the WASABI project (such as the multi-track audio player and the guitar tube amp simulation shown in figures 3 and 4).

These two partners will also participate in the realization of interactive demonstrators in the form of rich Web applications, capable of real-time sound manipulations:

1. Web applications to browse musical knowledge bases: an audio analysis engine, a search engine, faceted navigation browser with the ability to manually edit / correct text data (title of a song or misspelled lyrics, for example, correct or add an annotation), but also an API to be interrogated by external programs.

[6]https://github.com/Parisson/TimeSide

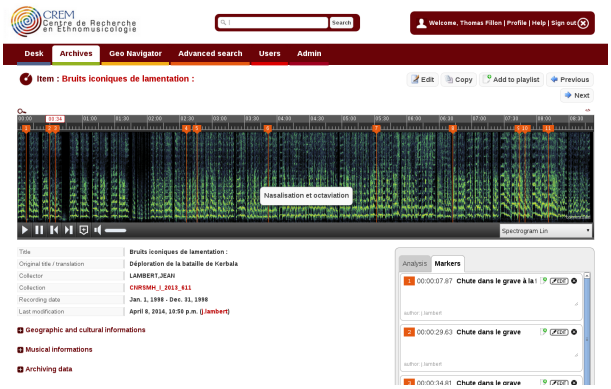

Figure 5: Telemeta platform deployed for the CREM-CNRS sound archives.[8]

2. Demonstrators linked to scenarios:
   - A tool to help with music composition,
   - A tool for data journalists,
   - A tool for the musicological analysis of a work,
   - Interactive examples integrable in MOOCs.
3. One or more datasets containing the set of RDF data describing the final corpus.
4. Open source bricks allowing (1) to develop applications exploiting this dataset: audio analyzers, parsers, WebAudio widgets, etc., and (2) facilitating interconnection with the on-line database.

IRCAM and I3S are involved in the WebAudio API standard and in the development of applications using this technology (organization of the first Web Conference Conference WAC2015, publications on the subject, ANR WAVE project, members of the W3C WebAudio working group that makes the standard, etc.). These demonstrators will make the most of this technology, which allows to develop musical applications of quality close to what is known in the world of native applications, but especially taking advantage of the possibilities of the Web: collaboration and hypermedia.

4. FIRST RESULTS: THE WASABI DATABASE AS IN APRIL 2017

Currently, the WASABI database comprises 77K artists, 200k albums, more than two million songs. We collected for each artist his complete discography, band members with their instruments, time line, etc. For each song we collected its lyrics[9], the synchronized lyrics when available[10], the word count, DBpedia abstracts and categories the song belongs to, genre, label, writer, release date, awards, producers, artist and/or band members, the stereo audio track from Deezer, when available, the unmixed audio tracks of the song, its ISRC, bpm, duration.

We matched song ids from the WASABI database with ids from MusicBrainz, iTunes, Discogs, Spotify, amazon, AllMusic, GoHear, YouTube. The matching is done using multiple

[9]from http://lyrics.wikia.com/
[10]from http://usdb.animux.de/

criteria to avoid false matches. We managed to match 87 percent of the WASABI songs with songs from the Deezer database (this was critical, as we needed the corresponding audio tracks). We used multiple criteria: metadata collected from DBpedia and MusicBrainz, but also, when we had the ids for this song from other datasources (iTunes, Spotify etc.), we used this information as Deezer also had collected such ids. We have 1732169 songs with lyrics, 73444 that have at least an abstract on DBpedia, 7227 that have been identified as "classic songs" (they have been number one, or got a Grammy award, or have lots of cover versions, etc.), and that will get some particular attention for one of the WASABI use case (sound engineer and music schools). We matched 2k songs with their multi-track version. This will be used as references for the source separation algorithms that IRCAM will run on the whole dataset.

WASABI will rely on multiple database engines: currently, we run on a MongoDB server altogether with an indexation by Elasticsearch. This database comes with a rather complete REST API[11].

In the next months metadata from the lyrics and audio analyses will be integrated, and the current nosql database will be associated with a full RDF semantic database.

We are also currently conducting interviews of WASABI end users, and developing ontologies that will match the different needs/use cases.

5. CONCLUSION

The WASABI project, started in January 2017, proposes a 2 million song metadata database that contains metadata from DBpedia, MusicBrainz and Deezer, lyrics, and matched audio files (stereo and some unmixed versions). Along the course of the project, we will add other datasources, run MIR algorithms on the audio files and NLP algorithms on the lyrics to enrich this database with new metadata, build a SPARQL endpoint that will work altogether with the current nosql database. Access to the database is public[12] through the WASABI Web GUI or available programmatically through the WASABI REST API. The database will be exploited by WebAudio client applications such as a multi-track player, the Telemeta application for automatic, semi-automatic and manual audio indexing, and comes with tools for studying the way songs have been recorded and produced: e.g. guitar tube amp simulations and multiple real time audio effects (pedalboard), mixing table, frequency analysers, oscilloscopes, audiograms and more to come (an integrated DAW is on the way).

6. ACKNOWLEDGMENTS

The WASABI project is supported by the French National Research Agency (contract ANR-16-CE23-0017-01).

7. REFERENCES

[1] A. Allik, G. Fazekas, M. Barthet, and M. Swire. mymoodplay: an interactive mood-based music discovery app. 2016.

[2] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere. The million song dataset. In *ISMIR*, volume 2, page 10, 2011.

[3] C. Leung and Y. Chan. A natural language processing approach to automatic plagiarism detection. In *Proceedings of the 8th Conference on Information Technology Education, SIGITE 2007, Destin, Florida, USA, October 18-20, 2007*, pages 213–218, 2007.

[4] J. P. G. Mahedero, A. Martinez, P. Cano, M. Koppenberger, and F. Gouyon. Natural language processing of lyrics. In *Proceedings of the 13th ACM International Conference on Multimedia, Singapore, November 6-11, 2005*, pages 475–478, 2005.

[5] G. Ì. F. Mariano Mora-Mcginity, Alo Allik and M. Sandler. Musicweb: an open linked semantic platform for music metadata. 2016.

[6] R. Mihalcea and C. Strapparava. Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea*, pages 590–599, 2012.

[7] Y. Raimond, S. A. Abdallah, M. B. Sandler, and F. Giasson. The music ontology. In *ISMIR*, pages 417–422. Citeseer, 2007.

[8] H. Schreiber. Improving genre annotations for the million song dataset. In *ISMIR*, pages 241–247, 2015.

[9] M. Schulder and E. Hovy. Metaphor detection through term relevance, 2014.

[10] L. Sterckx, T. Demeester, J. Deleu, L. Mertens, and C. Develder. Assessing quality of unsupervised topics in song lyrics. In *Advances in Information Retrieval - 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings*, pages 547–552, 2014.

[11] T. Tashiro, T. Ueda, T. Hori, Y. Hirate, and H. Yamana. EPCI: extracting potentially copyright infringement texts from the web. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 1151–1152, 2007.

[12] F. Thalmann, A. Perez Carillo, et al. The semantic music player: A smart mobile player based on ontological structures and analytical feature metadata. 2016.

[11] https://wasabi.i3s.unice.fr/apidoc/

[12] We mean here full access to metadata. There is no public access to copyrighted data such as lyrics and full length audio files. Audio extracts of 30s length are nevertheless available for nearly all songs.