Harvesting and Structuring Social Data in Music Information Retrieval

Sergio Oramas

Music Technology Group
Universitat Pompeu Fabra, Barcelona, Spain
sergio.oramas@upf.edu

Abstract. An exponentially growing amount of music and sound resources are being shared by communities of users on the Internet. Social media content can be found with different levels of structuring, and the contributing users might be experts or non-experts of the domain. Harvesting and structuring this information semantically would be very useful in context-aware Music Information Retrieval (MIR). Until now, scant research in this field has taken advantage of the use of formal knowledge representations in the process of structuring information. We propose a methodology that combines Social Media Mining, Knowledge Extraction and Natural Language Processing techniques, to extract meaningful context information from social data. By using the extracted information we aim to improve retrieval, discovery and annotation of music and sound resources. We define three different scenarios to test and develop our methodology.

Keywords: social media mining, knowledge extraction, natural language processing, information retrieval, music

1 Introduction

Online communities of users sharing multimedia content have become a cornerstone of the World Wide Web. Millions of users are handing out videos, photos, audios, and documents. Thus, large collections of multimedia resources have been gathered in web sites. This imposes challenges on how to deal with these data in an effective manner [1].

Music Information Retrieval (MIR) is a multidisciplinary field of research that is concerned with the extraction, analysis, and usage of information about music and audio. Traditionally, MIR has been more focused on the use of audio content, underestimating context information. However, in recent years several studies have showed the benefits of using a multimodal approach [2].

As stated by Schedl [3], factors that influence human music perception can be categorized into music content, music context, user context and user properties. According to this classification, music context, and user context seem to be a key aspect of MIR, and online communities are a very suitable place to look for this kind of information.

Sound and music related sites can be classified according to the presence or absence of music content, and the presence or absence of a community of users involved in the creation and edition of context information (Table 1).

User Community	Music Content	Example Sites
No	No	artist web pages, magazines
No	Yes	Internet Archive, iTunes, Spotify
Yes	No	Facebook, Twitter, MusicBrainz, Last.fm
Yes	Yes	Freesound, SoundCloud

Table 1. Categorization of sound and music related sites

Music context and user context information can be found in the Web in a structured or an unstructured form. On one hand, relational databases, web APIs or SPARQL endpoints are typical sources of structured content. On the other hand, unstructured content can be found in web documents, forum posts, user comments, microblogs, etc. Web content mining techniques can be applied to deal with these unstructured sources of information, harvesting relevant data from web content.

Ontologies have shown its utility to structure information in the Web, but in the creation process it is not easy to find clear agreement between different information sources. Thus, there is always the need to involve domain experts and to account for the fact that there are no single and long-lived formalizations [4]. Collaborative tagging has led to another data structure, the folksonomy. The analysis of folksonomies has demonstrated its utility [5]. However, this data structure suffers from a lack of semantic meaning.

Ontologies can also be exploited and enriched using natural language processing. Academic and industrial applications of this technique are usually called semantic technologies [6]. The combination of knowledge extraction and text mining can be addressed in two directions. First, learning ontology classes or instances in a semi-automatic way by using text mining techniques. Second, using ontologies as a guide that details what type of information to harvest, improving the process and the results of text mining [7].

In this proposal we take these ideas in order to develop a methodology to improve the process of annotation and retrieval of large audio collections. Our overall goal is to extract knowledge from structured and unstructured social data, using text mining techniques together with formal representations of the domain. For this purpose, ontologies will be created and enriched in a semi-automatic way from the analysis of context information generated by communities of experts, and they will be used to guide in the process of information extraction from user-generated content.

We will focus our research in the extraction of knowledge from music-related context information sources in the Web. To this end, we will take information

from structured (Wikipedia, WordNet, MusicBrainz), semi-structured (Freesound, SoundCloud, Last.fm, Internet Archive) and unstructured sources (Facebook, Twitter).

As a first step, our intention is to take some elements from folksonomies and ontologies to improve annotation, searching and browsing in Freesound.org. Freesound is an online audio clip-sharing website with more than three million registered users and more than 200.000 user-contributed samples [8].

In addition, we plan to harvest less structured social media content from Facebook, Twitter, music-related websites, and music forums. With the obtained information we aim to improve the annotation of large collections of audio, and use this new metadata in music recommendation and artist similarity tasks.

New Music Information Systems can be created using the harvested and structured social data. Those systems would provide data in a machine-readable format, making knowledge available on the Web in a structured way. Hence, new music context information would be added to the Semantic Web.

The rest of the proposal is organized as follows. First we comment on research found in related work. Then we propose our research questions. After that we outline our plan of research and describe the methodology. Preliminary results are then reported and an evaluation plan is proposed. We conclude the proposal with an outline of future benefits derived from our research.

2 State of the Art

As this proposal is strongly related to web content mining and knowledge extraction, we will review related work on those topics in the context of MIR. Although web content mining research has been an emergent topic in the MIR community over the last years, there is scant research related to knowledge discovery. Before addressing this related research, we want to briefly summarize some relevant concepts and perspectives related to ontologies and folksonomies.

2.1 Folksonomies and Ontologies

Folksonomies are the result of a collaborative annotation process [9]. They are composed of tags, resources, users, and their three-fold relations. They are generated in websites, where users attach tags to annotate resources, usually without any restriction or predefined hierarchy. Problems associated with them are related with the linguistic and semantic limitations of tags. Synonyms, misspelt words, or semantic relations between terms are not reflected in the folksonomy [5].

An ontology represents knowledge as a set of concepts within a given domain, and the relationships between those concepts. It provides a framework to deal with structured information, making implicit knowledge explicit, describing relevant parts of a domain and making data understandable and processable by machines. To define an ontology it is required consensual agreement from community members. Therefore, creation and maintenance of ontologies are more expensive than folksonomies, which are easier to create, edit, use and reuse [1].

Ontologies are commonly created by a small set of experts, and users are not usually involved in the creation process. On the contrary, folksonomies are created by final users directly. However, they do not contain a precise representation of the relations between concepts of the domain. Hence, on one hand we have experts wisdom in ontologies, and on the other hand we have the wisdom of the crowds in folksonomies [10].

Although both ways of knowledge representations has its pros and cons, they are not absolutely opposite; they can be used in combination to create better ways of organizing information on the Internet. There are several approaches on how they can cooperate, combining the flexibility of use and cooperation of folksonomies and the structured model of knowledge of ontologies. One approach is to create an ontology that supports a folksonomy like in [9], [11] and [12]. Another approach is to use a folksonomy to create an ontology [5]. There is a third approach where tagging, taxonomy, and ontology are mixed. Here folksonomies are used to find concepts, and ontologies are used as a schema, in a way that the ontology is modified by the community, given to a socially driven ontology [10].

2.2 Web Content Mining

Early work in text mining in the context of MIR is mainly related to extraction of music artist information from artist-related web pages, using search engines to gather those pages and then parsing their DOM trees [13]. Other studies [14] [15] use weighted term profiles based on specific term sets for recommendation and classification tasks. Co-occurrence of artist names in web pages content and page count based on results provided by search engines have been used for artist similarity and recommendation tasks [16]. Another interesting application of text mining techniques is the analysis of music artist-related microblogging posts for artist similarity estimation and artist labeling [17].

Sordo et al. [18] propose a methodology for extracting semantic information from music-related forums, inferring semantic relations from the co-occurrence of musical concepts in forum posts, and using network analysis. Other application of web content mining in MIR is automatic generation of Music Information Systems [19]. Here, information about music artists and bands is automatically gathered from various sources in the Web, processed, and published.

2.3 Knowledge Extraction

The boundary between natural language processing techniques and knowledge extraction is somehow fuzzy. We address here some research related to the use of structured knowledge representations in the context of MIR. In [20] a set of semantic facets is automatically obtained and anchored upon the structure of Wikipedia, and tags from the folkosonomy of Last.fm are then categorized with respect to the obtained facets. In [21] a methodology to automatically extract semantic information and relations about musical entities from arbitrary textual sources is proposed. Although more related with music content than music context, [22] shows a method for the automatic creation of an ontology

of musical instruments using formal concept analysis to build the hierarchical structure of the ontology.

3 Problem Statement and Contributions

After a concise study of the state of the art, our research questions are: Can we extract meaningful musical knowledge from social data in online communities? How can we use expert-based knowledge information and user generated content to better structure context information in audio repositories? How can we improve retrieval and discovery using the harvested and structured context information?

The creation of new methodologies to harvest and structure meaningful information from social data is a hot topic in the Big Data era. The Music Information Retrieval field has experienced an increase of related research in the last few years. However, scant studies have taken advantage of ontology-based knowledge extraction techniques. The Web is full of communities of domain experts creating meaningful knowledge in a crowd-sourced way. Therefore, it would be very valuable to extract and structure this community knowledge. By using it, we could improve structuring, browsing and annotation in music and sound repositories, and also ameliorate accuracy in some typical issues of Music Information Retrieval. This will require a combination of methodologies coming from distinct areas: Social Media Mining, Information Retrieval, Knowledge Extraction, Natural Language Processing and Semantic Multimedia Web.

4 Research Methodology and Approach

The ultimate end of this PhD work is contribute to the improvement of Music Information Systems, exploiting structured context information with semantic meaning. To obtain this domain knowledge, a two step process is defined. First, structured and unstructured social data related to the music domain is gathered from online communities using web content mining techniques. This information can be classified as expert or non-expert content. Expert generated content is considered especially suitable for knowledge extraction. However, both of them are valuable data sources for information extraction using natural language processing. User generated information can be extracted from any of the music related online-community types described above.

Second, gathered information is then structured and semantically annotated. For this purpose, it is necessary a combination of natural language processing and knowledge extraction, using ontologies as a formal knowledge representation. Ontologies play a key role in this step, working as a background for the natural language processing, and at the same time, being enriched with new extracted knowledge.

Finally, the structured and semantically annotated information can be used in a Music Information System to improve Music Information Retrieval tasks,

such as music recommendation, artist similarity or the annotation of sound and music resources.

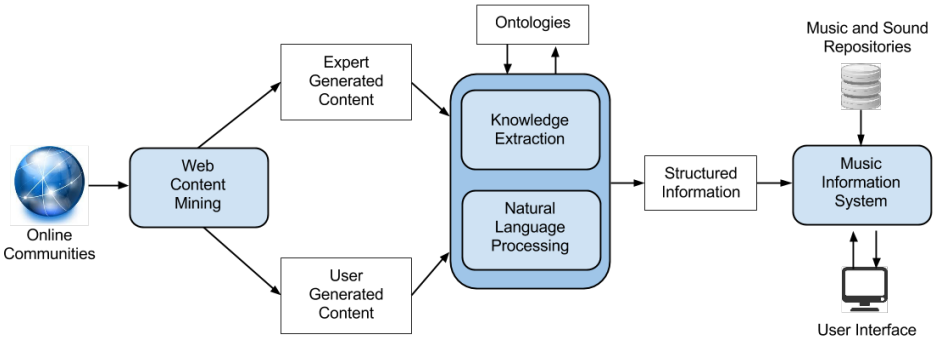

Fig. 1. Overview of the proposed methodology

We plan to apply this methodology in three different scenarios. Thus, we want to prove the importance of the use of experts knowledge in Music Information Retrieval. We plan to use structured, semi-structured and unstructured information as an input of our system, and try to use the obtained structured information to help in different tasks.

First, we plan to improve the annotation quality and the searching process of an audio-clip sharing site, Freesound.org. We will gather all the information related to resources, users and tags conforming the folksonomy. For the knowledge extraction step, we will design an ontology which is able to represent the information from the folksonomy together with domain specific semantic relations between tags and resources.

Second, we plan to improve artist similarity and genre classification tasks by using information extracted from user generated content in Facebook posts and comments. Natural language processing techniques such as sentiment analysis or topic modeling are going to be applied to this data. Expert generated content related to genre will be also gathered from Wikipedia. With the semantic information obtained we also plan to generate a new Music Information System, publishing gathered content automatically in HTML for navigation and in a machine readable way using RDF.

Third, we will use harvested user and expert generated content from different web sources to get structured information for the improvement of the annotation of the Internet Archive music collection. At this moment the context information of the collection is scant. We plan to use structured information from MusicBrainz and DBpedia, together with semi-structured and unstructured information gathered from SoundCloud, Last.fm, Twitter and Facebook.

5 Preliminary Results

The first step has been the identification of the problem for the first scenario. Members of our research group have already done some research in this area. Freesound.org has been developed at the Music Technology Group, and there are various publications analyzing its resources, folksonomy and community [23], [8], [24] and [25].

According to [25], there are 971,561 tag applications provided by 6,802 users and including 143,188 sounds, resulting in an average of 6.79 tags per resource. The folksonomy of Freesound is continuously growing, but it is quite noisy (misspelt words, synonyms, homonyms, ...). Hence, a tag recommendation system has been implemented to increase tag reuse.

As a starting point in the ontology design, we plan to reuse some concepts from the MUTO (Modular Unified Tag Ontology) ontology [11], which is, according to the literature, the most recent ontology of a folksonomy, and it suits our needs.

This ontology is only focused on the annotation process, storing all relevant information about the tripartite relations of the folksonomy. However, it does not add any semantic information about tags and resources. Therefore, our intention is to reuse some concepts of this ontology and add some domain specific semantic relations between tags and resources. For this purpose, we added a set of subclasses derived from the resource and tag classes, and some semantic properties relating them (Fig. 2).

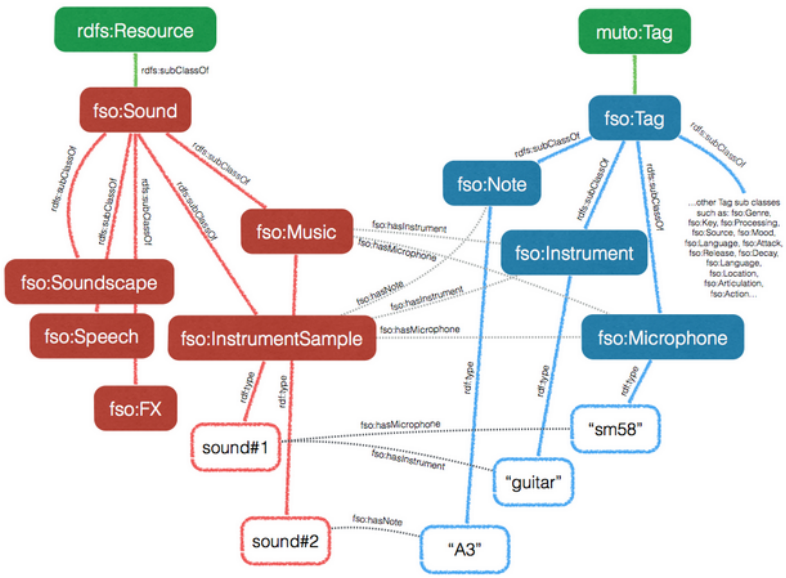

Fig. 2. Schema of the Freesound Ontology

Using this ontology, we aim to perform an automatic classification of tags. In addition, we plan to modify the web interface of Freesound, letting the user to choose the category of a tag. Thus, membership of a specific subclass of tags will be determined by the user with the use of semantically enhanced tags in the annotation process. These tags will have two members, an attribute and a value with syntax *attribute:value*, where each attribute corresponds to a specific subclass of tag.

Starting from previous studies, an initial analysis has been done to define the subclasses of resources and tags [25]. According to this we have determined five different categories of sounds and thirteen categories of tags as an starting point.

Using these categories a first version of the Freesound Ontology has been created. The ontology is already defined in OWL, and an RDF triplestore have been created to store all the information of the folksonomy following our ontology design.

6 Evaluation Plan

Our methodology implies different types of evaluation for each processing step. On one hand we need to evaluate the knowledge extraction and natural language processing step, and on the other hand, we have to measure the improvement of the Music Information System.

To evaluate a knowledge extraction system, we need to measure the quality of the inferred knowledge. The creation of gold standards based on existing ontologies, and available expert resources such as WordNet are a crucial step in our evaluation process.

To evaluate the quality of the information extracted, we may use as ground truth structured information already present in expert communities such as MusicBrainz or well annotated music repositories.

Finally, to evaluate the improvement of an MIR task, we should measure its performance with and without the use of the extracted information. Precision and recall are typical measures used for MIR evaluation. Moreover, each specific MIR task may require its specific evaluation process. User feedback is also a key value in other to evaluate Music Information Systems. User-centric evaluation experiments involving real users will be carried on to measure the performance of our systems.

7 Conclusion

Combining concepts from Information Retrieval, Social Media Mining and Knowledge Extraction in the analysis and improvement of Music Information Systems is an open field not very much explored, and with an enormous potential. We have proposed a methodology that takes advantage of this combination in order to transform social data into structured and meaningful information. With this

information we plan to improve annotation in sound and music repositories, and some related MIR tasks.

Adding structured and semantic information to sound and music collections would be useful not only for users, but also for researchers. For instance, in the case of Freesound and the Internet Archive, well annotated subsets of audio files would be excellent datasets to develop and test MIR algorithms.

We expect that methodologies and prototypes created for this purpose will be applicable to other multimedia online communities, and even more, to any type of online community. Moreover, semantic technologies applied to extract structured information will be reusable in other frameworks and research fields. The Big Data era has arrived, and expert knowledge should play a key role in information retrieval tasks.

References

1. M. Sordo, "Semantic Annotation of Music Collections : A Computational Approach," *PhD Thesis*, 2011.
2. M. Schedl and D. Schnitzer, "Location-Aware Music Artist Recommendation," *Lecture Notes in Computer Science*, vol. 8326, pp. 205–213, 2014.
3. M. Schedl, "On the Use of the Web and Social Media in Multimodal Music Information Retrieval," *Postdoctoral Thesis (Habilitation)*, 2013.
4. X. Serra, "Exploiting Domain Knowledge in Music Information Research," *Stockholm Music Acoustics Conference 2013 and Sound and Music Computing Conference 2013*, 2013.
5. P. Mika, "Ontologies are us: A unified model of social networks and semantics," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, pp. 5–15, Mar. 2007.
6. A. Gangemi, "A Comparison of Knowledge Extraction Tools for the Semantic Web," in *The Semantic Web: Semantics and Big Data*, pp. 351–366, 2013.
7. H. Alani, D. Millard, M. Weal, W. Hall, P. Lewis, and N. Shadbolt, "Automatic ontology-based knowledge extraction from Web documents," *IEEE Intelligent Systems*, vol. 18, pp. 14–21, 2003.
8. F. Font, G. Roma, P. Herrera, and X. Serra, "Characterization of the Freesound online community," *2012 3rd International Workshop on Cognitive Information Processing (CIP)*, pp. 1–6, 2012.
9. F. Echarte, J. J. Astrain, A. Córdoba, and J. Villadangos, "Ontology of Folksonomy : A New Modeling Method," *SAAKM 2007*, p. 8, 2007.
10. A. Sharif, "Combining ontology and folksonomy : An Integrated Approach to Knowledge Representation," *Emerging trends in technology libraries between Web 20 semantic web and search technology*, pp. 1–13, 2007.
11. S. Lohmann, P. Díaz, and I. Aedo, "MUTO: the modular unified tagging ontology," *Proceedings of the 7th International Conference on Semantic Systems - I-Semantics '11*, pp. 95–104, 2011.
12. S. Angeletou, M. Sabou, L. Specia, and E. Motta, "Bridging the Gap Between Folksonomies and the Semantic Web : An Experience Report," *European Semantic Web Conference*, pp. 30–43, 2007.
13. W. W. Cohen and W. Fan, "Web-collaborative filtering: recommending music by crawling the Web," *Computer Networks*, vol. 33, pp. 685–698, 2000.

14. D. P. W. Ellis, D. P. Ellis, B. Whitman, B. Whitman, A. Berenzweig, A. Berenzweig, S. Lawrence, and S. Lawrence, "The quest for ground truth in musical artist similarity," in *Proc. International Symposium on Music Information Retrieval (ISMIR 2002)*, pp. 170–177, 2002.
15. B. Whitman and S. Lawrence, "Inferring descriptions and similarity for music from community metadata," in *Proceedings of the 2002 International Computer Music Conference*, pp. 591–598, 2002.
16. M. Schedl, P. Knees, and G. Widmer, "A Web-Based Approach to Assessing Artist Similarity using Co-Occurrences," in *Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing {(CBMI'05)}*, 2005.
17. M. Schedl, D. Hauger, and J. Urbano, "Harvesting microblogs for contextual music similarity estimation: a co-occurrence-based framework," *Multimedia Systems*, vol. -, pp. –, 2013.
18. M. Sordo, J. Serrà, and X. Serra, "A Method for Extracting Semantic Information from on-line Art Music Discussion Forums," in *2nd CompMusic Workshop*, pp. 55–60, 2012.
19. M. Schedl, G. Widmer, P. Knees, and T. Pohle, "A music information system automatically generated via Web content mining techniques," *Information Processing and Management: an International Journal*, vol. 47, 2011.
20. M. Sordo, F. Gouyon, L. Sarmento, O. Celma, and X. Serra, "Inferring Semantic Facets of a Music Folksonomy with Wikipedia," *Journal of New Music Research*, vol. 42, pp. 346–363, 2013.
21. P. Knees and M. Schedl, "Towards Semantic Music Information Extraction from the Web Using Rule Patterns and Supervised Learning," *Colocated with ACM RecSys 2011 Chicago, IL, USA October 23, 2011*, p. 18, 2011.
22. S. Kolozali, M. Barthet, G. Fazekas, and M. B. Sandler, "Automatic Ontology Generation for Musical Instruments Based on Audio Analysis," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, pp. 2207–2220, 2013.
23. F. Font, X. Serra, M. T. Gorup, and U. P. Fabra, "Folksonomy-based tag recommendation for online audio clip sharing," *Proceedings of 13th International Society for Music Information Retrieval Conference, Oporto*, 2012.
24. F. Font and X. Serra, "Analysis of the folksonomy of freesound," in *Proceedings of the 2nd CompMusic Workshop*, pp. 48–54, 2012.
25. F. Font and J. Serr, "Audio clip classification using social tags and the effect of tag expansion," *AES 53rd International Conference on Semantic Audio*, pp. 1–9, 2014.