

Extending Sound Sample Descriptions through the extraction of Community Knowledge

Frederic Font and Xavier Serra

Music Technology Group, Universitat Pompeu Fabra
Roc Boronat, 138, 08018 Barcelona, Spain
{frederic.font, xavier.serra}@upf.edu

Abstract. Sound and music online services driven by communities of users are filled with large amounts of user-created content that has to be properly described. In these services, typical sound and music modeling is performed using either content-based or context-based strategies, but no special emphasis is given to the extraction of knowledge from the community. We outline a research plan in the context of Freesound.org and propose ideas about how audio clip sharing sites could adapt and take advantage of particular user communities to improve the descriptions of their content.

Keywords: sound and music computing, online communities, folksonomies, emergent semantics, freesound.

1 Introduction and Related Work

The proliferation of domain specific online services with social networking capabilities has created new ways of sharing information items such as pictures, videos and audio recordings. These services are typically filled with content contributed by communities of users with very different motivations and interests, generating large amounts of information that has to be properly described. Examples of this kind of services are Flickr (photo sharing), YouTube (video sharing), SoundCloud (music sharing) and Freesound (audio clip sharing)¹.

Traditional approaches of information modeling in sound and music online services are either focused on content-based or context-based techniques. Freesound is a collaborative database of contributed audio samples that uses content-based analysis - via manual tagging or sound waveform analysis - to perform queries and similarity search among audio files. On the other side, music services like Last.fm or Spotify² use context-based collaborative filtering techniques for their recommendation systems. In this case, explicit feedback provided by users in the form of comments, ratings or annotations is being analyzed, but also implicit information that can be extracted from their habits - such as the most common listened artists - can be considered, allowing the creation of user

¹ www.flickr.com, www.youtube.com, www.soundcloud.com, www.freesound.org

² www.last.fm, www.spotify.com

models for a personalized recommendation. Current research trends are proposing combinations of both content-based and context-based analysis to perform those tasks.

There has been extensive research on the analysis, modeling and knowledge extraction from communities of users. Some studies have performed group modeling through the aggregation of individual user profiles with the purpose of improving TV content recommendations [6, 9]. Other papers have proposed community models to study users behavior (*e.g.* to analyze document sharing patterns among an online community of researchers [5]). In the context of the semantic web, the trend of *emergent semantics* proposes the analysis of online communities for the extraction of ontologies that structure the concepts and terms of a particular domain [1]. Peter Mika [7] demonstrates ontology emergence through the analysis of tag folksonomies in a social bookmarking tool. Community knowledge extraction has been also investigated in [3], where the folksonomy of tags in Flickr is categorized using the concepts of a collaboratively built ontology with the purpose of improving tag-based recommendations. Similar ideas have been studied for music annotation and retrieval purposes [2, 10], but none of them gives a special emphasis to community-based approaches for the description of sound and music.

There is a need to move the emphasis of sound and music information modeling *from audio* - namely sound files - *to user* - people that use sound and music services - *to community* - understanding sounds and users as members of common particular contexts -, and be able to integrate and take advantage of all them. This does not only mean to perform better analysis of community knowledge, but also to use this knowledge for the adaptation of the technology to the particular needs of user communities in order to improve the “quality” of the information we can gather from them. In our thesis, we will address this problem aiming to propose community-based methods for the description of sound samples in Freesound.

2 Freesound and its Community

Freesound.org is a collaborative sound database where people from different disciplines share recorded sound samples under an open Creative Commons license³. It was started in 2005 and it is being further developed within our research group. Nowadays, there are 1.8 million registered users that have uploaded more than 100.000 sound samples (although only 1.5% of the users have uploaded at least one sample), and there is a rate of 30.000 unique visits per day. Registered users can provide information through comments and ratings on sounds (average of 0.5 comments and 5.5 ratings per sound), forum posts (already more than 25.000) or sound tags (with more than 26.000 different concepts used, average of 6.24 tags per sound). The search process is either based on raw tags attached to sound samples (without any post-processing for “cleaning” or classifying these tags) or

³ www.creativecommons.org/licenses/sampling+/1.0

in acoustic audio similarity (using automatic feature extraction). Freesound does not support the explicit grouping of users into communities, nevertheless it has been observed that different implicit communities exist related to the interests of users [8]. As a consequence, Freesound turns out to be a highly heterogeneous database in types of sounds, their descriptions and user interests. This makes Freesound interesting as a framework for researching community-based approaches of improving sound descriptions.

3 Research Challenges and Work Plan

In our current approach for addressing the problem of improving sound sample descriptions using community knowledge, we can identify two research challenges: *i*) how can we better analyze and gather knowledge from the community? and *ii*) how can we improve sound descriptions using this knowledge? A work plan for addressing these challenges is now described.

As a first step, we will focus in one of the implicit communities of interest found in Freesound [8]. Applying the Actor-Concept-Instance model described in [7], we expect to come upon a lightweight ontology emerging from the community and resembling the most important used tags - and their relations - for the description of sounds. We have already tried to apply this model to the whole database of Freesound and we have seen that, although many interesting semantic relations between tags arise (*e.g.* the connection of the tag *drums* with *snare*, *hi-hat*, *percussion...*), we obtain a huge cloud of small clusters that does not bring much information as a whole. We expect to get more interesting results by just considering sounds and users belonging to a particular community of Freesound.

Our next step will be to “post-process” tag descriptions based on the previously obtained ontology and other ontologies such as WordNet⁴ (a lexical database that relates English terms in sets of cognitive concepts). Similarly to what is done in [2,3], the idea is to map existing tags to concepts of the ontology and get more insight in their meaning. Then, we can extend the description with related and complementary tags (*e.g.* extending the tag *guitar* with the tag *musical-instrument*). We will carry on user-based evaluation to asses the expected improvement of sound descriptions.

We will then investigate how can we influence users behavior in describing sounds with the re-design of some Freesound functionalities, thus shaping the information we gather from the community. As an example, the system could recommend tags based on the relevant concepts identified for the particular community (similarly to [4]). In that way, when tagging sound samples of recorded musical instruments, the interface could recommend the use of tags describing important perceptual qualities such as *key* or *pitch*. Again, we will carry on user testing to evaluate the improvement of the descriptions.

Later stages of our research will include the extension of the work done to include other communities of Freesound and the evaluation of the overall impact

⁴ <http://wordnet.princeton.edu>

of the new functionalities. For that purpose, we plan to build a community model (similarly to [5]) to quantify the activity with respect to a number of previously decided key aspects (*e.g.* sound sample sharing, users collaboration in sound description, reuse of existing tags...).

4 Conclusions

With our thesis we expect to develop methodologies to better analyze and extract knowledge from communities of users. Although we are particularly focused in the case of sound sample descriptions, we believe that our work will be of relevance to other types of media. Furthermore, by implementing those methods in Freesound, we expect to drive the site into a more dynamic, rich and creative platform for users, and a more valuable source of information for researchers.

Acknowledgments. This work is partially supported under BES-2010-037309 FPI grant from the Spanish Ministry of Science and Innovation for the TIN2009-14247-C02-01 DRIMS project.

References

1. Aberer, K., et al: Emergent Semantics Principles and Issues. Proc. of the International Conference on Database Systems for Advanced Applications, 25-38 (2004)
2. Cano, P., et al: Knowledge and Content-based Audio Retrieval Using WordNet. Proc. of the International Conference on E-business and Telecommunication Networks, 301-308 (2004)
3. Cantador, L., Konstas, I., Jose, J.M.: Categorising Social Tags to Improve Folksonomy-based Recommendations. Web Semantics: Science, Services and Agents on the World Wide Web, 9(1):1-15 (2011)
4. Jschke, R., et al: Tag Recommendations in Folksonomies. Proc. of the European conference on Principles and Practice of Knowledge Discovery in Databases, 506-514 (2007)
5. Kleanthous, S., Dimitrova, V.: Analyzing Community Knowledge Sharing Behavior. Proc. of the International Conference on User Modeling, Adaptation, and Personalization, 231-242 (2010)
6. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. User Modeling and User-Adapted Interaction, 14(1):37-85 (2004)
7. Mika, P.: Ontologies are Us: A Unified Model of Social Networks and Semantics. Web Semantics: Science, Services and Agents on the World Wide Web, 5(1):5-15 (2007)
8. Roma, G., Herrera, P.: Community Structure in Audio Clip Sharing. Proc. of the International Conference on Intelligent Networking and Collaborative Systems, 200-205 (2010)
9. Senot, C., et al: Analysis of Strategies for Building Group Profiles. Proc. of the International Conference on User Modeling, Adaptation, and Personalization, 40-51 (2010)
10. Wang, J., Chen, X., Hu, Y., Feng, T.: Predicting High-level Music Semantics using Social Tags via Ontology-based Reasoning. Proc. of the International Conference on Music Information Retrieval, 405-410 (2010)