Freesound Technical Demo

Frederic Font, Gerard Roma and Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
frederic.font@upf.edu, gerard.roma@upf.edu, xavier.serra@upf.edu

ABSTRACT

Freesound[1] is an online collaborative sound database where people with diverse interests share recorded sound samples under Creative Commons licenses. It was started in 2005 and it is being maintained to support diverse research projects and as a service to the overall research and artistic community.

In this demo we want to introduce Freesound to the multimedia community and show its potential as a research resource. We begin by describing some general aspects of Freesound, its architecture and functionalities, and then explain potential usages that this framework has for research applications.

Categories and Subject Descriptors

H.3.5 [**On-line Information Services**]: Web-based services; H.5.5 [**Sound and Music Computing**]: Systems

Keywords

Freesound; online databases; sound; audio clips

1. INTRODUCTION

Freesound's initial goal was to give support to sound researchers, who often have trouble finding large royalty-free sound databases to test their algorithms, and to sound artists, who use pre-recorded sounds in their pieces. After eight years since its inception, Freesound has become one of the most popular sites for sharing sound snippets. Freesound serves around 40,000 unique visits per day and has 3.5 million registered users accessing more than 160,000 uploaded sounds[2]. More importantly, there is a highly engaged community of users continuously contributing to the site, not only uploading sounds but also commenting, rating and discussing in the forums about relevant topics for the community. All sounds in Freesound are manually moderated by a group of users (users of Freesound) that check that descriptions are correct and sounds are not illegal.

All the content in Freesound is released under Creative Commons licenses. When uploading sounds, Freesound users can choose between CC0 (public domain), Attribution and Attribution-NonCommercial[3]. The reason to offer these licenses is to ensure that all the content uploaded in Freesound can be reused by other users, developers and researchers, but at the same time we provide users the option to require the attribution of their work or to restrict the use of their sounds to non-commercial activities. Furthermore, the source code of the web application is being released[4] as open source under the GNU AGPL license[5].

Freesound was built with high load and scalability in mind. Figure 1 shows the block diagram of the architecture. Retrieval of sounds can be performed using text queries, content-based similarity search (Query by Example) or by browsing tags or geotags. The front-end is a Django[6] application which includes basic social interaction features (forum, sound comments, sound ratings, private messaging...). Text indexing is supported by an Apache Solr[7] server including text descriptions and tags, which allows for sophisticated text queries using the Solr query syntax. A distributed architecture is used for processing incoming sounds, producing compressed previews and waveform/spectrogram images, as well as for feature extraction. Frame-level and clip-level descriptors are available for each sound. Feature extraction and similarity search services are supported with Essentia and Gaia[8], both open-source technologies developed also at the Music Technology Group.

In 2011, an API was introduced to facilitate access to the Freesound content to researchers and developers. The API runs as a Django application based on the RESTful principles. The API allows accessing a basic set of resources (sounds, users, sound packs) to http clients. All of the mentioned methods for searching sounds are supported, and also the API provides more complex search functionalities such as filtering with audio descriptors. Responses from the API can be obtained in a variety of formats (JSON, XML...). Content descriptors can be obtained through the API both at the frame level (descriptors calculated at every frame of the audio files) and clip level (statistics across frames). A

[1] `http://www.freesound.org`
[2] Freesound data current as of April 2013.

MM'13, October 21–25, 2013, Barcelona, Spain.
ACM 978-1-4503-2404-5/13/10.
http://dx.doi.org/10.1145/2502081.2502245.

[3] `http://www.creativecommons.org/licenses`
[4] `http://www.github.com/MTG/freesound`
[5] `http://www.gnu.org/licenses/agpl.html`
[6] `http://www.djangoproject.com`
[7] `http://lucene.apache.org/solr`
[8] `http://essentia.upf.edu`

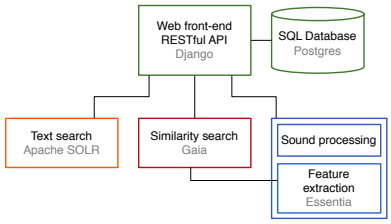

Figure 1: Freesound architecture.

number of clients have already been implemented for different programming languages. At the time of this writing, Python, Javascript, Actionscript, PhP, C# and SuperCollider clients are available.

2. FREESOUND AS A RESOURCE FOR RESEARCH

Freesound data has been used for research purposes in many different areas. Inside the Music Technology Group there have been publications dealing with topics such as soundscape generation and content-based audio-clip classification [7, 4], database discovery [9], community characterization and network analysis [8] or folksonomy analysis and tag recommendation [2, 3]. Moreover, outside the Music Technology Group Freesound data has also been used for research on creative interfaces, composition [1, 10], annotation of environmental sounds [6] and statistical methods for dimensionality reduction (using freesound data without considering domain-specific information, [5]).

To get an idea of the potential of Freesound as a research resource, here we briefly describe the basic types of content and information that can be gathered either using the API or by requesting us an anonymized version of the SQL database. The most obvious type of content in Freesound are sounds (audio files) and their descriptions extracted using Essentia. As Freesound was started with research purposes in mind, from the very beginning the quality of sounds and their descriptions has been prioritized over the quantity. For that reason, when sounds are uploaded, they all have to be described by adding some tags and a textual description (currently the average is 6.5 tags for sound with almost 56,000 unique tag concepts). Users can also add a geotag to indicate the location where the sound was recorded (there are more than 12,000 geotagged sounds). All this information is available with corresponding timestamps since 2005. Furthermore, 75% of uploaded sounds are grouped into packs of sounds that users can create. Hence, a big amount of sounds is grouped into almost 10,000 packs which are also available.

On the other side, users also are a relevant type of content in Freesound which provide a different perspective than that of the sounds. Although there are more 3.5 million registered users, only the 10% has contributed by uploading, commenting or rating a sounds. Indeed, only ∼8,500 of the registered users have uploaded at least one sound. This means that the community that is uploading to Freesound is very small compared to the number of people that are downloading (Freesound has already served more than 43 million downloads since its inception). All the data about downloads, ratings and comments is available with corresponding timestamps since 2005.

In the Freesound forums many types of conversations may be identified where, for example, users request sound samples, discuss about recording gear and techniques or talk about work done using Freesound samples. Currently, the forums have more than 8,000 threads and almost 37,000 posts. The forums reflect very relevant aspects of the Freesound community and are therefore an interesting source of information to be explored. Again, all this data is available with timestamps since 2005.

3. CONCLUSION

Freesound has become a consolidated audio clip sharing site with a lot of content and a strong community behind it. With the inclusion of the API, access to Freesound content has become much easier for the development of third-party applications and for research purposes. We strongly believe in the value of Freesound for the scientific community of information retrieval, sound and music computing and multimedia in general.

4. REFERENCES

[1] N. N. Correia. AV CLASH: Online Tool for Mixing and Visualizing Audio Retrieved from Freesound.org Database. In *Proc. of the Sound and Music Computing Cong. (SMC)*, 2010.

[2] F. Font, J. Serrà, and X. Serra. Folksonomy-based tag recommendation for online audio clip sharing. In *Proc. of the Int. Conf. on Music Information Retrieval (ISMIR)*, pages 73–78, 2012.

[3] F. Font and X. Serra. Analysis of the folksonomy of freesound. In *Proc. of the CompMusic Workshop*, pages 48–54, 2012.

[4] J. Janer, S. Kersten, M. Schirosa and G. Roma. An online platform for interactive soundscapes with user-contributed content. In *Proc. of the 41st AES Int. Conf. on Audio for Games*, 2011.

[5] M. Karasuyama and S. Masashi. Canonical dependency analysis based on squared-loss mutual information. *Neural Networks*, 34:46–55 2012.

[6] B. Mechtley, P. Cook and A. Spanias. Shortest Path Techniques for Annotation and Retrieval of Environmental Sounds. In *Proc. of the Int. Conf. on Music Information Retrieval (ISMIR)*, pages 541–546, 2012.

[7] G. Roma, J. Janer, S. Kersten, M. Schirosa, P. Herrera, and X. Serra. Ecological Acoustics Perspective for Content-Based Retrieval of Environmental Sounds. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010:7, 2010.

[8] G. Roma, M. Zanin, P. Herrera, S. L. Toral, F. Font and X. Serra. Small world networks and creativity in audio clip sharing. *International Journal of Social Network Mining (IJSNM)*, 1(1):112-127, 2012.

[9] G. Roma, P. Herrera and X. Serra. Freesound Radio: supporting music creation by exploration of a sound database. In *Computational Creativity Support Workshop (CHI09)*, 2009.

[10] M. Thorogood, P. Pasquier and A. Eigenfeldt. Audio Metaphor: Audio Information Retrieval for Soundscape Composition. In *Proc. of the Sound and Music Computing Cong. (SMC)*, 2012.