MUCOSA: A MUSIC CONTENT SEMANTIC ANNOTATOR

Perfecto Herrera, Òscar Celma, Jordi Massaguer, Pedro Cano, Emilia Gómez,
Fabien Gouyon, Markus Koppenberger, David García,
José-Pedro García, Nicolas Wack
Universitat Pompeu Fabra
Pg. Circumval·lació 8
Barcelona, 08003 Spain
pherrera, ocelma, jmassaguer, pcano, egomez,
fgouyon, koppi, dgarcia, jpgarcia, nwack@iua.upf.es

ABSTRACT

MUCOSA (Music Content Semantic Annotator) is an environment for the annotation and generation of music metadata at different levels of abstraction. It is composed of three tiers: an annotation client that deals with micro-annotations (i.e. within-file annotations), a collection tagger, which deals with macro-annotations (i.e. across-files annotations), and a collaborative annotation subsystem, which manages large-scale annotation tasks that can be shared among different research centres. The annotation client is an enhanced version of WaveSurfer, a speech annotation tool. The collection tagger includes tools for automatic generation of unary descriptors, invention of new descriptors, and propagation of descriptors across sub-collections or playlists. Finally, the collaborative annotation subsystem, based on Plone, makes possible to share the annotation chores and results between several research institutions. A collection of annotated songs is available, as a "starter pack" to all the individuals or institutions that are eager to join this initiative.

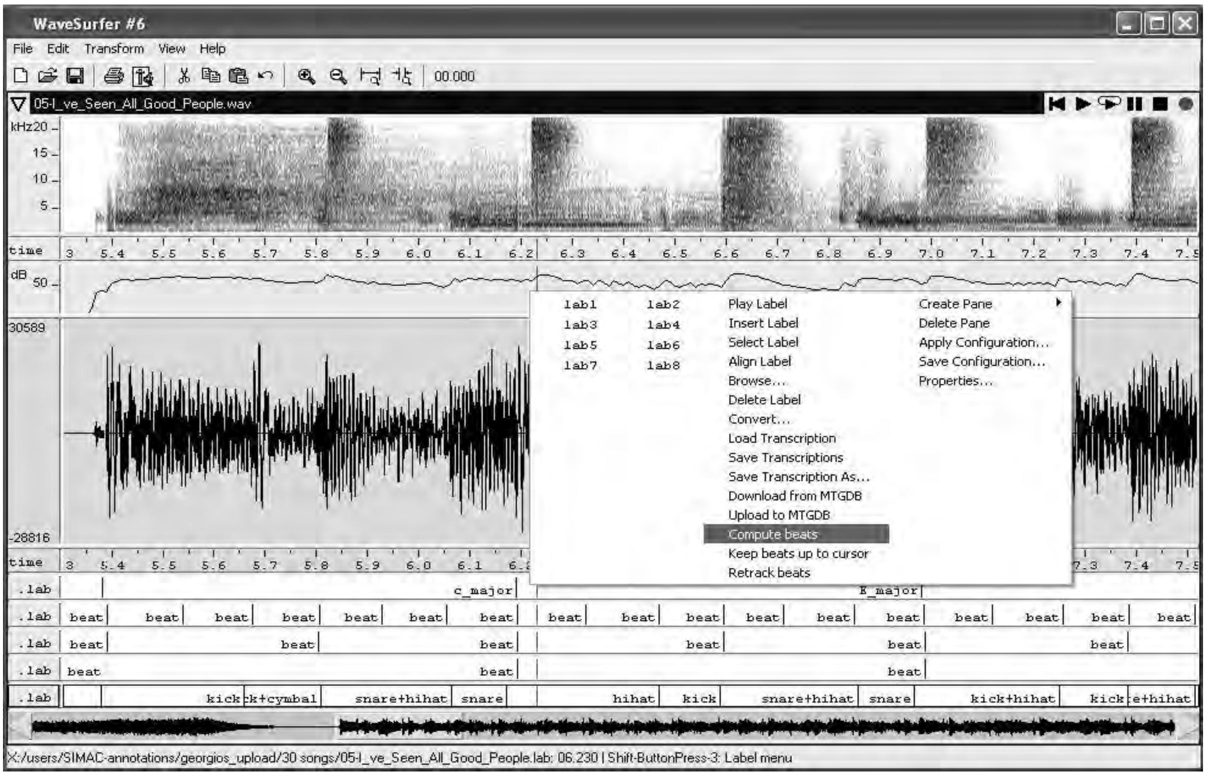
Keywords: Semantic descriptors, music tagging, audio annotations, audio music content processing, music databases.

1 INTRODUCTION AND MOTIVATION

The growing amount of digital music is driving the need for effective methods for indexing, searching, and retrieving of music based on its content. While recent advances in content analysis, feature extraction, and classification are improving the capabilities for effectively searching and filtering digital music content, the process of reliably and efficiently indexing multimedia data is still a challenging issue. Manual indexing of music collections has been attempted in different moments and contexts, but most of the attempts have been tagging artists and songs in a global way (i.e. assigning tags to a whole song, artist or recording). Micro-annotations, on the other hand, are required in order to compute predictive models of certain musical features. Micro-annotations may provide solid ground-truths for training artificial systems to automatically compute features like beats, chords, instruments and structural units. The models induced by these artificial systems can be exploited, in a second phase, for accelerating the annotation process itself, which, on its turn, should help to improving the quality of the inductive models, and so on. Descriptors generated thanks to micro-annotations are also used as building blocks for computing models for automatic labelling of higher-level descriptors that can be then exploited in macro-annotations.

Back in 1992, the visionary Marvin Minsky stated: "the most critical thing, in both music research and general AI research, is to learn how to build a common (…) database" [1]. It seems that, more than a decade later, we are recognizing its value, and some useful reflections and recommendations have been discussed in [2]. Apart from lacking of a clear methodology and theory for annotating, and from having to deal with an ill-posed problem, building an annotated database of music, specially in the case of micro-annotations, is very expensive and time-consuming. Motivated by that, we have devised an environment that alleviates the individual cost of annotating songs and music collections by maximizing the synergies between different research groups.

In this paper we present MUCOSA (MUsic COntent Semantic Annotator), an annotation environment to allow authors to semi-automatically annotate music content with semantic descriptions. MUCOSA is a three-tiered environment consisting on an annotation client, a collection tagger, and a collaborative annotation manager. The tools included in this environment explore or will explore a number of interesting capabilities as automatic segmentation, summarization, automatic label propagation, and template annotation propagation to similar segments and files. MUCOSA also includes an administrative web interface for the management of descriptors, users, groups of annotators, and annotation tasks. One of its most interesting features for the Music Information Retrieval community is that it allows an incremental shared-cost-and-benefit approach to getting a universally available corpus of annotated audio music files (i.e. *the more, the merrier*). We are taking advantage, in this respect, of Creative Commons' licensing



ing

Figure 1. A song annotated using Wavesurfer with different types of descriptors; values for some of them have been automatically computed by means of plug-ins.

schemes[1], that make possible to get and give "free-for research-purposes" access to moderate-size music collections in audio format.

In the forthcoming sections, we first present some related annotation systems, then we move on to the presentation of the three components of the proposed environment, that is, the MUCOSA annotation client, the MUCOSA collection tagger, and the Collaborative Annotation subsystem.

2 BACKGROUND

There have been several annotators available for the video world: VideoAnnEx [3], developed by IBM, is one of the most famous, and it has operated as an inspirational tool for the MUCOSA environment. Between 2001 and 2003, a more-than-a-hundred community of annotators, from twenty research centres, amassed a total amount of 100 annotated hours of video using VideoAnnEx [4]; the annotated labels were served as a foundation for several TREC-video retrieval systems [5]. IBM also released a a *Multimodal Annotation Tool*, which was derived from an earlier version of *VideoAnnEx* including special features such as audio signal graphs and manual audio segmentation functions [6]. Other video annotators that deserve a mention are *MovieToo*l[2], developed by Ricoh for interactively creating (directly using XML) video content descriptions conforming to MPEG-7 syntax, or Vannotea [7], a prototype for the real-time collaborative indexing, browsing, description, annotation and discussion of digital films and videos.

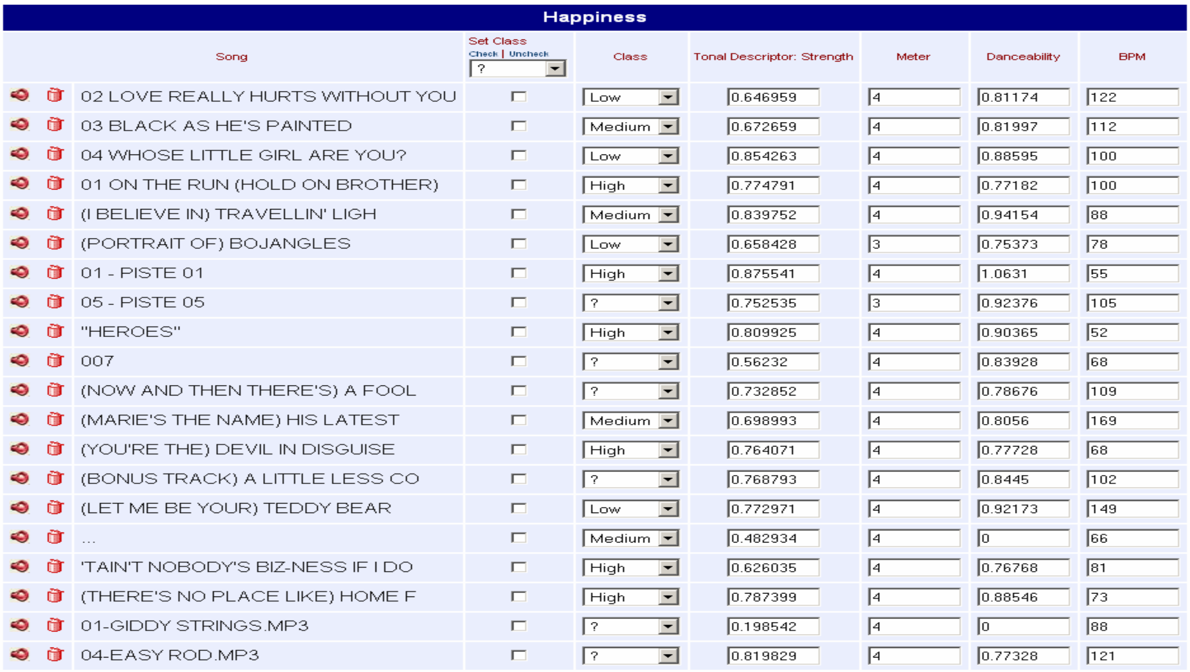
In the musical audio side, annotators have been a rare species. Let us mention Timeliner [8], the Acousmograph[3], MiXA [9], Marsyas [10] or the CLAM Annotator [11]. The first one is integrated in the Indiana University digital music library (Variations2), and is intended for pedagogical functions related to the structural description of music files. The Acousmograph, Marsyas and the CLAM Annotator are focused on micro-annotations of an audio file, but they do not incorporate automatic or semi-automatic annotation capabilities, or the functionalities required to share large annotation tasks among different teams of annotators. MiXA, on the other hand, is intended to help the annotation of scores by means of musicXML [12] descriptions.

In a different category, the MTG-DB [13], a database of audio material that offers functionalities for adding audio content, content browsing, adding metadata and dealing with taxonomies and algorithms, provides most of the infrastructure upon which we have built MUCOSA, which can be considered as a complementary subsystem, specifically focused on music annotation under collaborative requirements.

[1] http://creativecommons.org/about/licenses/

[2] http://www.ricoh.co.jp/src/multimedia/MovieTool/

[3] http://www.ina.fr/grm/outils_dev/acousmographe/



Happiness

Song	Set Class Check | Uncheck ?	Class	Tonal Descriptor: Strength	Meter	Danceability	BPM
02 LOVE REALLY HURTS WITHOUT YOU	□	Low	0.646959	4	0.81174	122
03 BLACK AS HE'S PAINTED	□	Medium	0.672659	4	0.81997	112
04 WHOSE LITTLE GIRL ARE YOU?	□	Low	0.854263	4	0.88595	100
01 ON THE RUN (HOLD ON BROTHER)	□	High	0.774791	4	0.77182	100
(I BELIEVE IN) TRAVELLIN' LIGH	□	Medium	0.839752	4	0.94154	88
(PORTRAIT OF) BOJANGLES	□	Low	0.658428	3	0.75373	78
01 - PISTE 01	□	High	0.875541	4	1.0631	55
05 - PISTE 05	□	?	0.752535	3	0.92376	105
"HEROES"	□	High	0.809925	4	0.90365	52
007	□	?	0.56232	4	0.83928	68
(NOW AND THEN THERE'S) A FOOL	□	?	0.732852	4	0.78676	109
(MARIE'S THE NAME) HIS LATEST	□	Medium	0.698993	4	0.8056	169
(YOU'RE THE) DEVIL IN DISGUISE	□	High	0.764071	4	0.77728	68
(BONUS TRACK) A LITTLE LESS CO	□	?	0.768793	4	0.8445	102
(LET ME BE YOUR) TEDDY BEAR	□	Low	0.772971	4	0.92173	149
...	□	Medium	0.482934	4	0	66
'TAIN'T NOBODY'S BIZ-NESS IF I DO	□	High	0.626035	4	0.76768	81
(THERE'S NO PLACE LIKE) HOME F	□	High	0.787399	4	0.88546	73
01-GIDDY STRINGS.MP3	□	?	0.198542	4	0	88
04-EASY ROD.MP3	□	?	0.819829	4	0.77328	121

Figure 2. A screenshot of the descriptor creator that is included in the collection tagger

3 THE MUCOSA ANNOTATION CLIENT

The MUCOSA client is in charge of:

- computing descriptors for a given song,
- depicting descriptors as time-varying lines or as labelled segments,
- computing a fingerprint for a given song

The core of MUCOSA is another annotation tool, WaveSurfer, developed at the Stockholm's Centre for Speech Technology (KTH) [14]. WaveSurfer was originally designed for tasks such as viewing, editing, and labeling of audio data, and it is built around a small core to which most functionality is added in the form of plug-ins. The tool was designed to work on most common platforms and with the aims that it should be easy to configure and extend. WaveSurfer is provided as open source, under the GPL license, with the explicit goal that the speech community jointly will improve and expand its scope and capabilities. The WaveSurfer tool is built using the Tcl/Tk [15] scripting language[4], with scripts and dynamic link libraries wrapped into a single executable. The tool consists of a simple core, combined with a novel plug-in architecture for all task-specific functionality. Wavesurfer also incorporates analysis and visualization of pitch, spectrogram and formants.

The MUCOSA client exploits the WaveSurfer functionalities plus added features coming from specific plug-ins, in order to categorize the semantic content of each music file or their extracted segments and upload the description to the central database.

There are four major operations in a MUCOSA client working session:

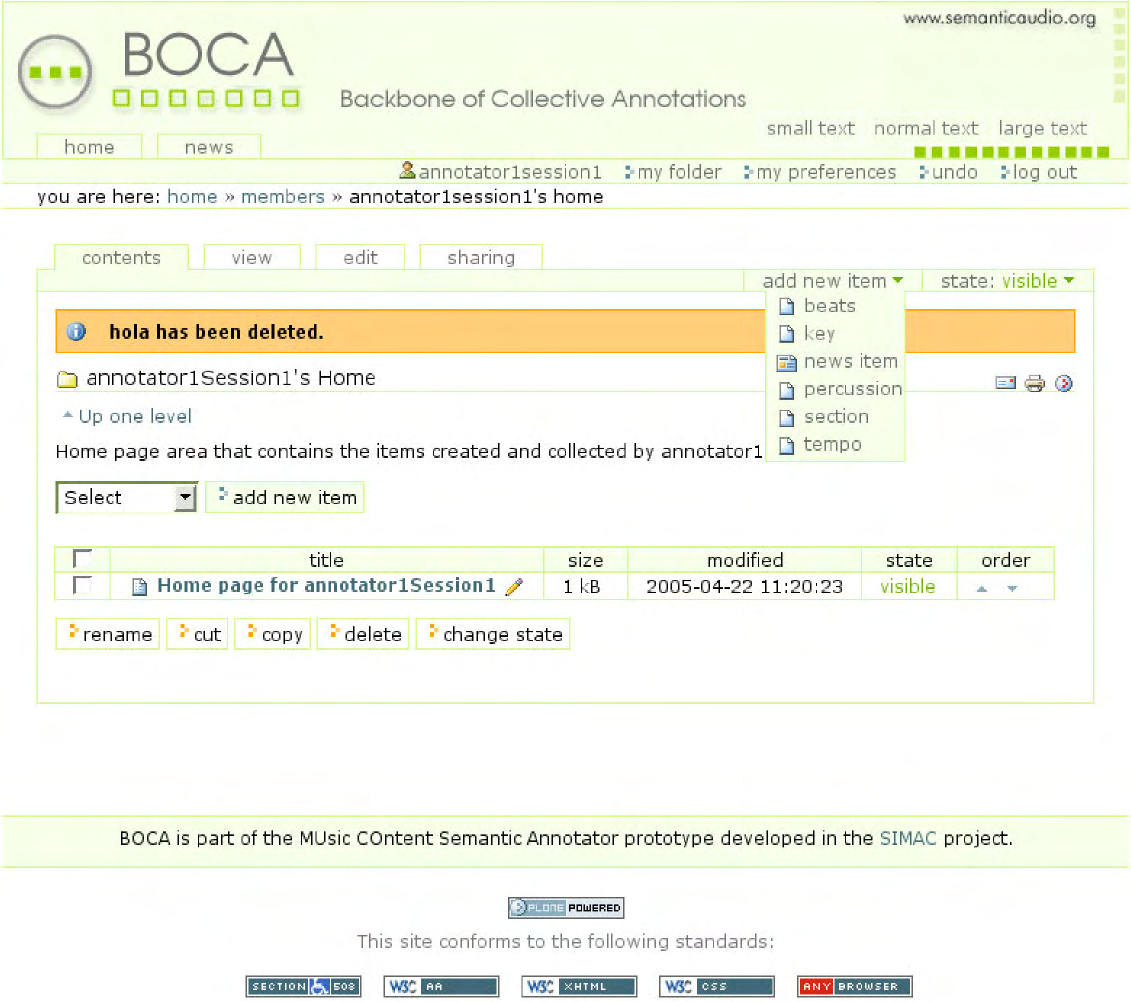
1. Music segmentation can be performed to cut up the file into smaller units.
2. A pre-defined semantic lexicon is used in order to regulate the music content descriptions.
3. A human annotator labels the music segments with its semantic labels. Automatic annotation-learning components can be used to speed up the annotation task. These components are integrated as WaveSurfer plug-ins.
4. The resulting descriptions of the annotation process are uploaded from the MUCOSA client to a central server but they can also be locally outputted in a structured format.

Descriptors that are currently extracted or in the way to be extracted include:

- Low-level descriptors such as spectral centroid, skewness, or Mel Cepstrum coefficients; an MPEG-7 subset of audio low-level descriptors can be specifically computed thanks to the integration of the MPEG7AudioEnc library[5] [16]
- Rhythm descriptors such as tempo, beat or metric.
- Tonality descriptors.
- Instrumentation descriptors such as the occurrence of percussive events.
- Miscellaneous descriptors such as danceability, subjective energy, or dynamics complexity.

[4] http://www.tcl.tk/

[5] http://www.ient.rwth-achen.de/team/crysandt/software/



you are here: home » members » annotator1session1's home

hola has been deleted.

annotator1Session1's Home

Up one level

Home page area that contains the items created and collected by annotator1

☐	title	size	modified	state	order
☐	Home page for annotator1Session1	1 kB	2005-04-22 11:20:23	visible	▲ ▼

BOCA is part of the MUsic COntent Semantic Annotator prototype developed in the SIMAC project.

This site conforms to the following standards:

Figure 3. Screenshot of BOCA annotator's homepage

- Section segments, which sometimes correspond to structural units like intro, verse or chorus.
- Genre assignment probabilities.

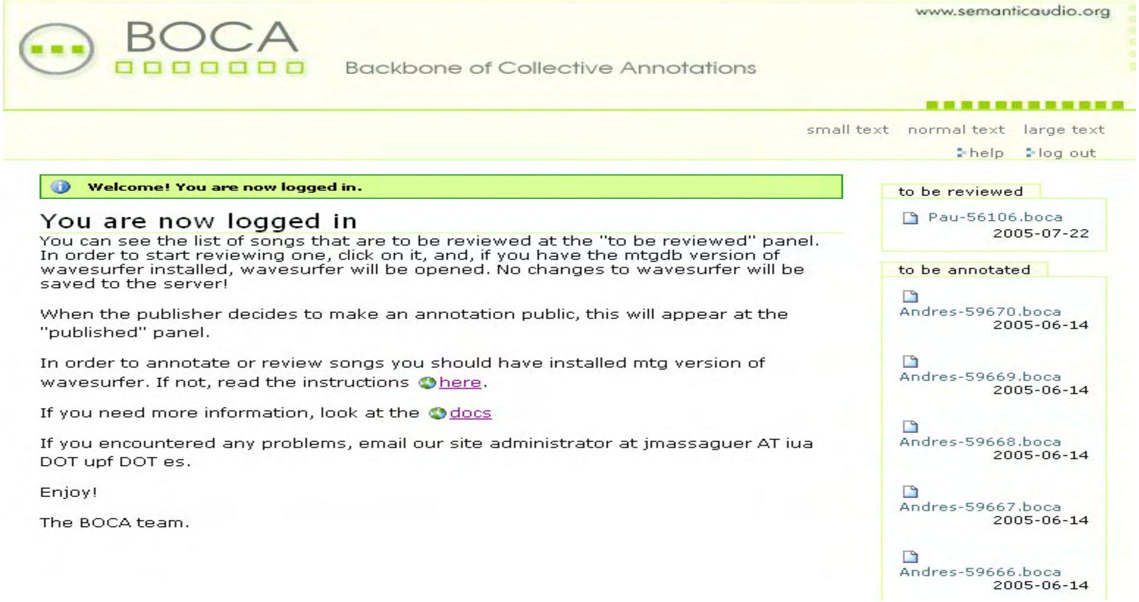
Time-varying real-valued descriptors (i.e. spectral centroid, MFCCs, etc.) can be visualized as colour-coded functions (several of them can be stacked on the same window). Labelled segments (according to beat or chord marks, presence of a given instrument, or structural sections) are stacked below the waveform representation, according to the requirements of the user generating the annotations (see figure 1). Segments can be "extracted" and converted into independent soundfiles for additional processing. Label prediction utilizes different statistical and machine learning techniques [17] to suggest labels for further segments or for other songs that have not been annotated yet. Predictions can also be propagated to other songs according to user-specified filters available in the MUCOSA collection tagger (see section 4).

There are two working modes, independent annotation and collaborative annotation. In the independent mode, where the audio is located in the user's computer, the annotations are stored locally in WaveSurfer format (text). In case that the song has not been previously registered in the database, a fingerprint is computed and centrally stored for further recognition and for fast downloading of its description to anybody requesting for it.

In the collaborative mode the audio file is downloaded from the BOCA server to the annotators' machine (see section 5) and, once it has been annotated using the WaveSurfer client, the annotations are uploaded to the central server, which checks if the song has been previously annotated, updates the annotation management checklists, stores the data in the database, and communicates all the participants in the project that a song has been annotated.

4 THE MUCOSA COLLECTION TAGGER

A second facet of MUCOSA is collection tagging, which deals with macro-descriptions or the assignment of "unary" descriptors (i.e. those that describe a song with a



Welcome! You are now logged in.

You are now logged in

You can see the list of songs that are to be reviewed at the "to be reviewed" panel. In order to start reviewing one, click on it, and, if you have the mtgdb version of wavesurfer installed, wavesurfer will be opened. No changes to wavesurfer will be saved to the server!

When the publisher decides to make an annotation public, this will appear at the "published" panel.

In order to annotate or review songs you should have installed mtg version of wavesurfer. If not, read the instructions here.

If you need more information, look at the docs

If you encountered any problems, email our site administrator at jmassaguer AT iua DOT upf DOT es.

Enjoy!

The BOCA team.

to be reviewed

Pau-56106.boca 2005-07-22

to be annotated

Andres-59670.boca 2005-06-14

Andres-59669.boca 2005-06-14

Andres-59668.boca 2005-06-14

Andres-59667.boca 2005-06-14

Andres-59666.boca 2005-06-14

Figure 4. BOCA control page for the annotation reviewer

single value like, for example, *tempo*). There are three strategies whereby a collection can be tagged with categorical (i.e. discrete) labels:

- Batch-assigning a given categorical value to a subset of the available songs (i.e. "wild" songs). This is done first by means of creating an M3U playlist (with Winamp or XMMS) that is named using a given concept, then by importing the playlist. This operation automatically tags all the songs in the playlist with the concept given.
- Creating a predictive model for a given concept[6]. The user provides examples for each one of the declared values for the concept (by means of submission of playlists, as explained above). With all these information, the system makes several calls to a remote Weka[7] server and computes a predictive model based on the signal-based descriptors that have been extracted for the songs (see figure 2).
- Retrieving words that have been used to describe that song or that artist and using them as descriptors. For a given artist, a web crawling system has gathered a series of words that are frequently associated with it. Wordnet[8] is used to expand this set with synonyms and related words. The most significant are offered as acceptable tags for a given song. A "propagate" button makes possible the propagation of the label to other similar songs, or to other similar artists' songs. Regrettably this functionality is not properly integrated into the system yet.

5 BOCA: THE MUCOSA COLLABORATIVE ANNOTATION SUBSYSTEM

MUCOSA allows collaborative annotation among multiple users through the Internet (see Figure 3). Music that is being collaboratively annotated is stored in a central server and is downloaded accordingly to the requests of the annotator. As these music titles have been issued under the Creative Commons licensing scheme, they can be distributed, jointly with their annotations, in a way that is "free of legal concerns". The currently available collection comes from Magnatune[9] but other collections are currently prospected.

A central server, called BOCA (Backbone Of Collaborative Annotation), takes care of:

- assigning user IDs and passwords to access the music and annotation files
- storing the collaboration checklists
- storing the annotation sessions
- storing the data files
- making possible the centralized management of all that.

BOCA has been developed using Plone[10], an open source content management system (CMS) that is built on top of Zope, a Python-based open source application server. Plone can be easily extended to meet specific needs like those generated by MUCOSA. It is also easy to create new content types as those used for BOCA files, and manage the client annotator while displaying the audio and its descriptors. Contrastingly to other

[6] This functionality has been implemented in collaboration with OFAI

[7] http://www.cs.waikato.ac.nz/ml/weka

[8] http://wordnet.princeton.edu/

[9] http://www.magnatune.org

[10] http://plone.org

similar CMS, Plone offers faster and more flexible workflow management building capabilities.

For collaborative annotation, there are three categories of user access to the BOCA Server: (1) annotation task leader, (2) annotator, and (3) annotation reviewer. The annotation task leader, by means of a web interface, sets up the project on the BOCA server, selects the songs to be annotated, registers the annotators, and distributes the annotation tasks and files among them. The annotators are the persons who perform the annotation task on the MUCOSA client and belong to different research institutions that join forces to get the job done in a fraction of the total amount of time. The annotation reviewers are the quality checking agents of the system: they review all the annotated songs and mark or comment those that should be revised. Every annotated song is reviewed by one annotation reviewer and by the task leader, in order to ensure a quality standard. The reviewing process is also intended as a mechanism for achieving convergence in case of conflicting annotations.

A project could consist of, for example, annotating the beats in 100 songs from Magnatune, selected by uniform sampling of genres. Here the annotation task leader would select those songs and would distribute them across the available annotators. The annotators, after logging into the collaborative annotation system (a specific web page will request for the user ID and password), would select the proper annotation task from the assignments they get from the BOCA server assignments page. The annotators can choose a file to annotate, can leave it temporally unfinished, and can see the existing annotations for that music file.

The usual workflow goes as follows:

1. The Annotation task leader adds annotators and reviewers to the system, and selects the songs to be annotated.
2. An annotator logs in, finds the songs to be annotated in a "to be annotated" list, and selects one of them to be downloaded and annotated.
3. When one song is annotated, it appears at the "to be reviewed" list.
4. The reviewer checks the completeness and correctness of the annotation, and accepts it if everything is alright.
5. The song appears at the "to be published" list.
6. When the task leader decides, it publishes the annotation to the annotators enrolled in that task.

In case of observing different annotation speeds, the task leader may re-assign songs or issue warning messages. When the 100 songs have been annotated, all the annotators are permitted to download the complete project (i.e. all the annotations) or some portions of it, depending on the existing agreements.

There is an annotation availability page where one can see the group task status list by the different groups and institutions supporting the project. Entries include the group or institution name, administrator's name, their allocated assignments, and the annotation status.

The BOCA server is currently under internal testing and for the moment we have gathered tempo, beats, percussion hits, structural sections, and key annotations for 100 songs from the Magnatune collection. This annotation effort required more than 160 work hours of specialized annotators (i.e. trained music students and one musicologist for coordinating them and reviewing their annotations). We hope other Institutions join this initiative and we all share the time and the outputs of a joint annotation effort. In order to make the collaboration even more attractive to them, the system provides a starting pack of 10 annotated songs, to be downloaded after providing a minimal amount of input to the existing collection.

6 CONCLUSIONS

We have presented the Music Content Semantic Annotator (MUCOSA), a three-tired environment that has been devised for annotating music in automatic, semi-automatic and totally manual modes. With MUCOSA, song files can be micro- and macro- annotated, and the descriptions may range from low-level to semantic labels. Some of the presented functionalities still require additional testing and improvement; other interesting ones are under consideration as, for instance, the possibility to play the audio file and a linked midi file, or showing the lyrics in the timeline.

One of the most interesting features included in the environment is the collaborative annotation whereby the chores of annotating a collection of songs can be shared between groups of researchers which, in the end, get the whole corpus of annotated music by a fraction of the effort required to do the task alone. MUCOSA can be accessed through the following link: http://www.semanticaudio.org/mucosa.

7 ACKNOWLEDGMENTS

The research and development reported here was partially funded by the EU-FP6-IST-507142 project SIMAC (Semantic Interaction with Music Audio Contents) project. The authors would like to thank Edgar Barroso, and the Audioclas and CLAM teams for their support to the project.

REFERENCES

[1] Minsky, M. and Laske, O. “A conversation with Marvin Minsky”. AI Magazine, 31-45, 1992.

[2] Lessaffre, M., Leman, M., De Baets, B. and Martens, J.-P. “Methodological considerations concerning manual annotation of musical audio in function of algorithm development”. Proceedings of the 4th International Conference on Music Information Retrieval, Barcelona, 2004, 64-71.

[3] Lin, C-Y., Tseng, B. L. and Smith, J. R. "VideoAnnEx: IBM MPEG-7 annotation tool for multimedia indexing and concept learning". Proceedings of the IEEE Intl. Conf. on Multimedia and Expo (ICME), Baltimore, MD, July 2003.

[4] Lin, C-Y., Tseng, B. L. and Smith, J. R. "Video collaborative annotation forum: Establishing ground-truth labels on large multimedia datasets. Proceedings of the NIST TREC Video 2003.

[5] Amir, A., Berg,M., Chang, S.F., Iyengar, G., Lin, C.Y., Natsev, A. P., Neti, C., Nock, H., Naphade, M., Hsu, W.,Smith, J. R., Tseng, B., Wu, Y. and Zhang, D. "IBM research TRECVID-2003 Video Retrieval System", Proceedings of the TRECVID 2003 Workshop.

[6] Adams, W.H., Lin, C.-Y., Iyengar, G., Tseng, B. L. and Smith, J. R.. "IBM multimodal annotation tool", IBM Alphaworks, August 2002.

[7] Schroeter, R., Hunter, J. and Kosovc, D. "Vannotea - A collaborative video indexing. Annotation and discussion system for broadband networks", K-CAP 2003 workshop on knowledge markup and semantic annotation, Florida, October 2003.

[8] Notess, M. and Swann, M. B. "Timeliner: Building a Learning Tool into a Digital Music Library", Proceedings of ED-MEDIA, Lugano, Switzerland, 2004.

[9] Kaji, K. and Nagao, K. "MiXA: A Musical Annotation System". Proceedings of the 3rd International Semantic Web Conference, Hiroshima, Japan, 2004.

[10] Tzanetakis, G. and Cook, P. R. "Experiments in computer-assisted annotation of audio", Proceedings of the ICAD, Atlanta, GE, 2000.

[11] Amatriain, X., Massaguer, J., García, D. and Mosquera, I. "The CLAM Annotator: A cross-platform audio descriptors editing tool". Proceedings of the 6th International Conference on Music Information Retrieval, London, UK, 2005.

[12] Good, M. "MusicXML in practice: Issues in translation and analysis". Proceedings of the First International Conference MAX 2002: Musical Application Using XML, Milan, Italy, 2002.

[13] Cano, P., Koppenberger, M., Ferradans, S., Martínez, A., Gouyon, V., Sandvold, V., Tarasov, V. and Wack, N. "MTG-DB: A repository for music audio processing". Proceedings of the 4th Intl. Conf. on Web Delivering of Music, Barcelona, 2004.

[14] Sjölande, K. and Beskow, J. "Wavesurfer - An open source speech tool", Proceedings of the Intl. Conf. on Spoken Language Processing, 2000, IV: 464-467.

[15] Ousterhout, J.K., Tcl and the Tk Toolkit. Addison Wesley, 1994.

[16] Crysand, H., Tummarello, G. and Piazza, F. "MPEG-7 encoding and processing: MPEG7AUDIOENC+MPEG7AUDIODB. 3rd Musicnetwork Open Workshop, Munich, 2004.

[17] Gouyon, F., Wack, N. and Dixon, S. "An open source tool for semi-automatic rhythmic annotation". Proceedings of the 7th Intl. Conference on Digital Audio Effects, Naples, 2004.