Semi-Automatic Annotation of Music Collections

Mohamed Sordo

Master thesis submitted in partial fulfillment of the requirements for the degree:

Màster en Tecnologies de la Informació, la Comunicació i els Mitjans Audiovisuals

Supervisor: Xavier Serra

Department of Information and Communication Technologies
Universitat Pompeu Fabra
Spain
September 2007

To my uncle Taoufik, a wonderful man that left us
suddenly and unexpectedly. Rest in peace.
To my little sister Randa, my parents
and my whole family.

Acknowledgments

First of all, I would like to thank **Xavier Serra**, my supervisor, for giving me the opportunity to join the Music Technology Group and to work with this very interesting topic. I would like to thank also, and very specially, **Òscar Celma**, for giving me the chance to work with him, and for his priceless support whenever I need it. Truthfully, this Thesis would have been nothing without his help.

I am very grateful to all the people that I have been involved with during the development of this Thesis: **Cyril Laurier**, **Perfecto Herrera**, **Pedro Cano**. A special mention to my room colleagues **Oscar Mayor**, **José Pedro**, **Koppi**, **Pablo**, **Jens**, **Dmitry**, **Paul**, who contributed to a pleasant working environment.

I am grateful as well to the administration staff (**Cristina**, **Joana** and **Salvador**) and the *sysadmins* (**Jordi**, **Ramon** and **Guillem**) for solving many situations I would have never done alone; and the rest of (but not less important) **MTG** members.

Finally, I would like to thank all my master colleagues, specially **César de la Rosa** and **Adrià Tauste**, who have been very close during the master academic year.

Abstract

The amount of multimedia content in the World Wide Web is increasing very much, and music is one of the most outstanding. Every time, there are more and more songs, artists, and even new genres. Hence, it is really hard to manage this huge quantity, in terms of searching, filtering, navigating through the content, etc.

One of the solutions for this problem is keeping annotations of the music files, in order to facilitate the retrieval process. However, it is known that annotating songs manually has a huge cost and annotating them automatically is quite inaccurate.

The approach of this master thesis is to propose a semi-automatic strategy that allows to annotate huge music collections, based on audio similarity and a community of users that annotate music titles. This strategy allows to increase the efficiency regarding the manual annotation, and the accuracy regarding the automatic annotation.

The Thesis presents two experiments followed for the evaluation of the annotation process: the first experiment consists on testing how the content–based similarity can propagate labels. Using a collection of of ∼5500 songs, we show that with a collection annotated at 40% with styles, we can reach a 78% (40%+38%) annotated collection, with a recall greater than or equal to 0.4, only using content–based similarity. In the case of moods, with a 30% annotated collection we can automatically propagate up to 65% (30%+35%).

Regarding the second experiment, we use a collection of ∼258000 songs. With a 48% manually annotated collection we propagate the annotations up to 76% (48%+28%) and then evaluate a small set of the propagated annotations by means of user relevance feedback.

Contents

Chapter 1

Introduction

Music is unquestionably a fundamental part of the society. It generates large communities of artists and also large communities of listeners. The expansion of the World Wide Web has helped to promote and to make know many artists and bands, but since the amount of music was increasing a lot, it was obligatory to organize this huge amount music and make it easy for searching and retrieval purposes. From the fields of investigation in this scope, Music Information Retrieval, Music Recommendation Systems and similarities between songs and artists are three of the most outstanding ones.

1.1 Motivation

Nowadays, there is a vast amount of digital multimedia material available on the World Wide Web, in digital storage medias, etc. So there is a need to organize and make this music easy to search, navigate, filter and retrieve in an efficient way. Searching in digital libraries has been studied for several years, mostly using text–based methods. These methods can be complemented with new strategies of retrieval, like those focused on content–based descriptors — extracted directly from the music files. However, these descriptors do not refer to any object in the real word, so that means that music is not a strictly type of knowledge. Another way of describing music, usually called meta–data, could help — in combination with the music content descriptors — to create some musical knowledge management, classification and representation.

Manual annotations of multimedia data is an arduous task, and very time consuming. Automatic annotation methods, normally fine-tuned to reduced domains such as musical instruments or limited to sound effects taxonomies, are not mature enough to label with great detail any possible sound. Yet, in the music domain the annotation becomes more complex due to the time domain frame.

As a paradigmatic example, the Music Genome Project is a big effort to "capture the essence of music at the fundamental level" by using over 400 attributes to describe songs. To achieve this, more than 40 musicologists have been annotating thousands of files since 2000. Based on this knowledge, a well–known system named Pandora[1] creates playlists by exploiting these human–based annotations. It is clear that helping these musicologists can reduce both time and cost of the annotation task.

Thus, the main goal of this thesis is to ease the process of annotating music collections, by using content-based similarity — as a way to propagate labels among songs — and relevance feedback from the users — in order to moderate these propagated labels.

1.2 Objectives

The main goal of this Thesis is to facilitate the music annotation problem. To achieve this goal, the following objectives are outlined:

- Make the annotation of songs faster and easier, since it allows to propagate and propose tags, and refine the system by means of relevance feedback from the users
- Improve audio similarity distances, by a multi–modal approach, that is, not only using "pure" content–based, but a hybrid one.
- Improve the quality of music retrieval tasks, by means of combining content information, context information, and subjective information (tags) in order to get a better approximation of the similarities between songs.

1.3 Thesis Outline

This Thesis is structured as follows: chapter 2 introduces the basics and a brief background of multimedia annotation, and music in particular, and reviews some related work. Chapter 3 explains the motivation and the need of this work, and overviews all the technical and non-technical details of the system developed in order to demonstrate the approach.

[1] http://www.pandora.com

After that, chapter 4 shows some results obtained from two different experiments using the system developed in chapter 3. Finally, chapter 5 draws some conclusions and discusses open issues and future lines for the PhD Thesis.

Chapter 2

State of the Art

The structure of this chapter is as follows: section 2.1 describes very briefly some basis about the music information retrieval (MIR) field; section 2.2 covers the topic of cultural music annotations. In section 2.3 current research on music field and other multimedia fields (such as image, video and sound effects) is reviewed. Finally, in section 2.4, systems that are the inspiration and the point of departure of this Thesis are described.

2.1 Music Information Retrieval

Nowadays, there is a vast amount of digital multimedia material available on the World Wide Web, in digital storage medias, etc. So there is a need to organize and make this music easy to search, navigate, filter and retrieve in an efficient way. Searching in digital libraries has been studied for several years, mostly using text-based methods. These methods can be complemented with new strategies of retrieval, like those focused on content-based descriptors — extracted directly from the music files. However, these descriptors does not refer to any object in the real word, so that means that music is not a strictly type of knowledge. Another way of describing music, usually called meta–data, could help — in combination with the music content descriptors — to create some musical knowledge management, classification and representation.

The purpose of making all music easily accessible implies a condition of describing music in such a way that machine learning can understand it, as [Pachet 05] states. Specifically, these two steps must be followed:

- Build descriptions of music easy to maintain, and

- Exploit these descriptions to build efficient music access systems that help the users finding music in large collections.

2.1.1 The Music Information Plane

There are several ways for describing music content, but we can basically classify the descriptors in three groups [Celma 06a], [Pachet 05]:

- Editorial meta–data: this kind of meta–data is obtained by the editor. Editorial meta–data includes songs and albums, but also information about artists. It can be either objective (song name, artist name, etc.) or subjective, like artists' biographies[1], genre information, etc. Depending on the nature of the human source, editorial meta–data could be also described as:
  - prescriptive, where the information is decided by well-defined experts.
  - non-prescriptive, where the information is classified based on collaborative scheme (a community of users).
- Cultural meta–data: the meta–data is obtained by the environment or culture. The information is not explicitly entered in an information system, rather is calculated using user profiles — also known as the so-called collaborative filtering. However, it does not depend only on these profiles — since it is very poor — but on other sources like search engines, encyclopaedias, music radio programs, etc. The techniques — borrowed from natural language processing — are most of them based on co-occurrence analysis: associate items that are closer in some sense, for example, similar in genres, etc.
- Acoustic meta–data: obtained by the analysis of the audio file (no other kind of information is used), i.e, the content descriptors of the sounds. The intention is to have purely objective information about the music files. The descriptors can be either Tempo (in bps[2]) or other more complex descriptors like rhythm, timbre, instrument recognition [Herrera 05b], etc.

We can see these three groups as three different points of viewing the annotation of music (meta–data). So if we take into account more than one point of view at a time, the result could be a better description of music.

[1] Personal description of a human is almost always subjective information.

[2] beats per second

Another field of study and research is the similarity between music files, in terms of their content; and the retrieval of the descriptors that best define the music files.

2.1.2 Music information retrieval systems

Music information retrieval (MIR) systems, like other kind of information retrieval, try to satisfy users' queries retrieving music files that are related to the given query. For classic information retrieval systems [Baeza-Yates 00], every single document (in our case an audio file) is described by a set of features. The system will then retrieve the documents that are relevant to the user's query, that is to say, according to the set of features. This kind of retrieval is also known as keyword-based technique. These features then need an accurate annotation and extraction process, to make information retrieval easier. In the field of music, these features are related to the before mentioned editorial, cultural and acoustical meta–data.

It is clear that annotating music files give them more value, and eases the retrieval, search, navigation and filter processes. The following section presents different ways of doing this task.

2.2 Annotations based on a community of users

2.2.1 Annotation via Tags

We can annotate music files by means of tags, but what does this *word* means? Tags are keywords, category names, or meta data that describe web content. Tags[3] can be whatever words that better describe web content for users. But their job is not to organize all the information over the world wide web into tidy categories, rather it is to add value to the huge amount of data available nowadays [Beckett 06]. Tagging is then a process to describe web content using these tags. This process is actually a combination of 4 entities, as shown in figure 2.1.

- Person: who perform the operation, also called tagger.
- Tag: set of tags being used.
- Date-Time: when the tagging process was performed.

[3] In the rest of this Thesis, both "tag" and "label" words will refer to the same concept.

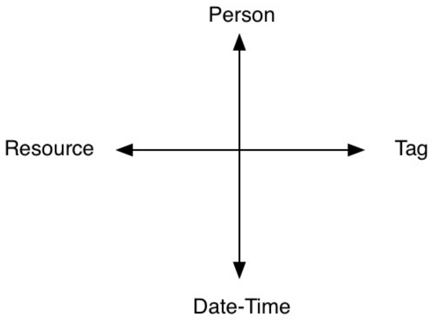


Figure 2.1: The tagging process

- Resource: the resource URI being described.

In the music field, tags are keywords that describe the audio files, and resource is the audio file itself.

On the other hand, annotations from a community of users are part of the next generation Web, also known as Web 2,0. Initially, the tagging process was used for simple objects (web pages), but now it is also used to annotate multimedia content.

Annotations consist of combination of natural language words. They can be free, that is to say, there is no rule that restricts their use; or controlled by a taxonomy or a classification scheme. We can observe two kind of free annotations, named Folksonomy and Personomy.

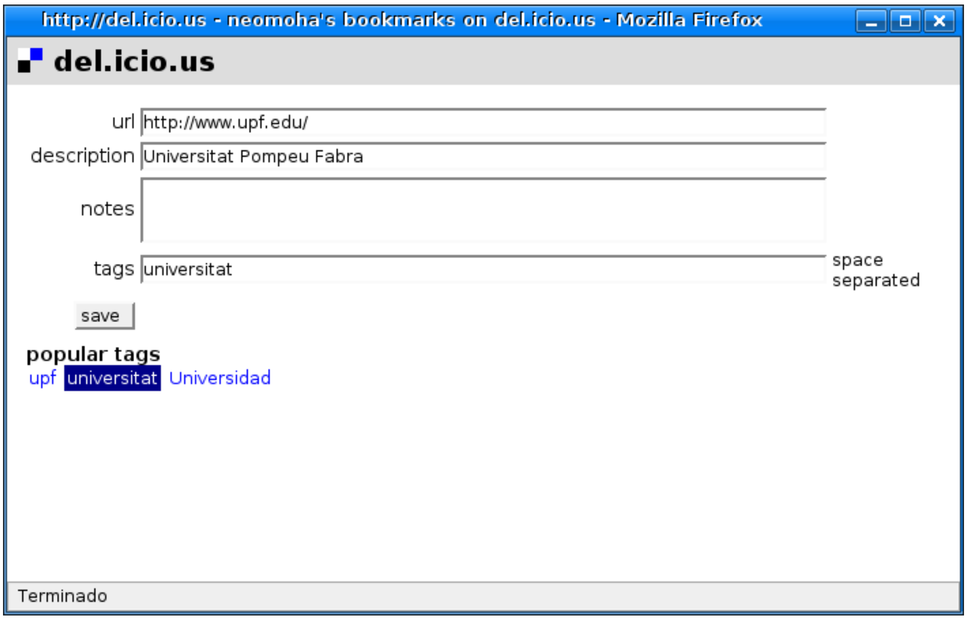
Folksonomy

The word “Folksonomy” is a combination of “folk” and “taxonomy”. It was first proposed by Wal in a mailing list[4] in 2004. It provides user–created meta–data rather than professional or author created meta–data [Mathes 04]. Examples of folksonomies are the web sites: del.icio.us, Flickr and Technorati.

del.icio.us. del.icio.us[5] is a social bookmarks website. Its objective is to annotate bookmarked URIs. For del.icio.us, tags are “one-word descriptors that you can assign to any bookmark”, and that “there is no such thing as a right or wrong tag. A tag is whatever you want it to be”.

[4] http://atomiq.org/archives/2004/08/folksonomy_social_classification.html

[5] http://del.icio.us/tag



http://del.icio.us - neomoha's bookmarks on del.icio.us - Mozilla Firefox

del.icio.us

url http://www.upf.edu/

description Universitat Pompeu Fabra

notes

tags universitat

space separated

popular tags

upf universitat Universidad

Terminado

Figure 2.2: Tagging a website with del.icio.us

The only required field for annotating is the "description" field. It encourages the use of tags in the bookmarking process, including a presentation of suggested tags on the form once a URI is given. Figure 2.2 shows an example of annotating a website with del.icio.us.

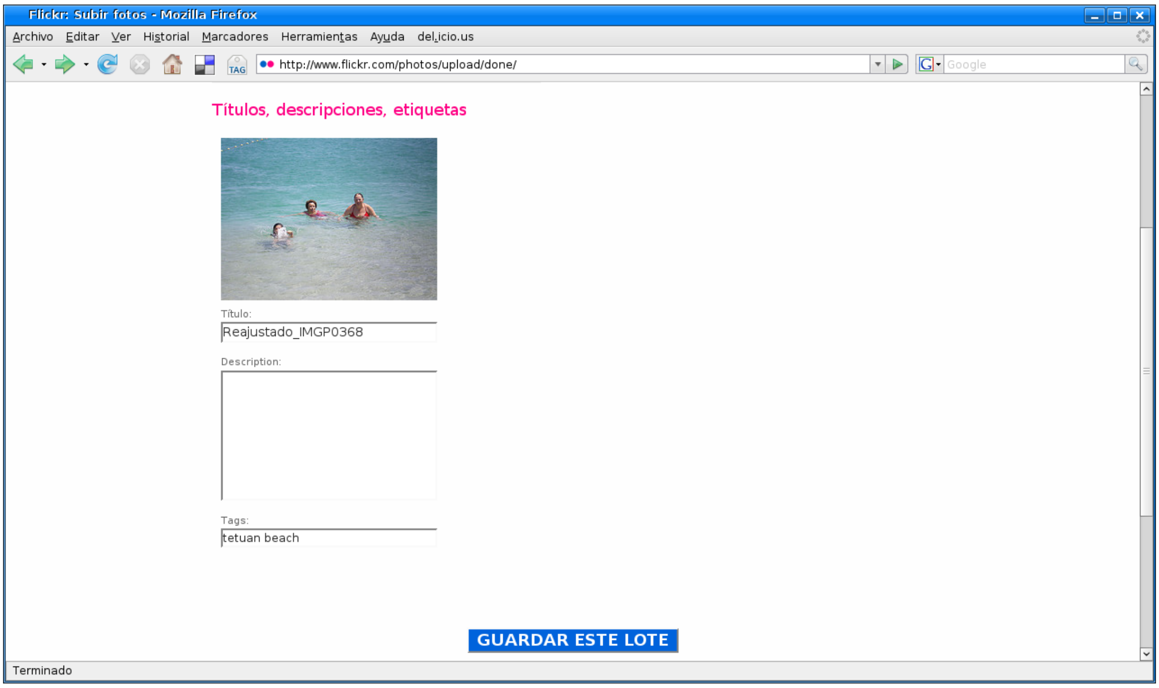
Flickr. Flickr[6] is a website that focuses on still images. Its primary goal is to annotate a picture (mainly the title and a description), and optionally for parts of an image, with additional descriptions. For Flickr tags are "like keywords or labels that you can add to a photo to make it easier to find later". Thus, its main use is for retrieval purposes.

Flickr encourages tagging — like del.icio.us — but does not offer suggestions. A user can tag whatever image present in the website — normally self ones or other users' uploaded images — and there are rich interfaces to browse pictures via tags, such as tag clouds, group tags and clusters. Figure 2.3 shows an example of tagging an image with Flickr.

Technorati. Technorati is a real–time search engine for user–generated media (including weblogs) by tag. For Technorati[7], tags and the tagging process are the reasons to exist. These tags are used inside HTML content as syndicated feeds. Technorati includes an

[6] http://www.flickr.com/tags/

[7] http://www.technorati.com/tag/


Flickr: Subir fotos - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda del.icio.us
http://www.flickr.com/photos/upload/done/
Google
Títulos, descripciones, etiquetas
GUARDAR ESTE LOTE
Terminado

Título: Reajustado_IMGP0368

Description:

Tags: tetuan beach

Figure 2.3: Tagging a picture with Flickr

aggregator that looks for the rel="tag" attribute on links and uses that for marking URIs that are relevant to a tag — known as the tagging context. The Technorati tags are mainly for retrieval and aggregation tasks, but there is no community to join.

Personomy

"Personomy" considers individuals' annotations. It has been used traditionally for organization and self-retrieval benefits, although there are other motivations for personal annotations, as we will see in subsection 2.2.2. An example of a Personomy is the main page of a del.icio.us user, with the tags that he/she has used so far.

2.2.2 Motivations for annotation

Another interesting point can be inferred from the following questions: "Why people started using tags?", and "Which are the motivations that lead us to annotate digital objects, specially web objects, or multimedia objects?" We can automatically suggest a logic one: for organizing and retrieving these objects afterwards; but maybe there are more reasons for annotating.

Ames and Naaman [Ames 07] have recently developed an experiment and arrived to offering a taxonomy of motivations for annotation. Their work is focused on annotating images, but it can be useful for other kind of multimedia objects. The taxonomy consists of 2 dimensions, "sociality" and "function", as it can be observed in table 2.1. Sociality refers to the purpose of the tags, either for personal use or for other users. The function dimension distinguishes the fact of using tags for organizing content or — as a new contribution from the authors — for communicating some additional context to the multimedia objects.

	Function	
	Organization	**Communication**
Self	∗ Retrieval, Directory ∗ Search	∗ Context for self ∗ Memory
Social	∗ Contribution, attention ∗ Ad hoc pooling	∗ Content descriptors ∗ Social signaling

Table 2.1: A taxonomy of tagging motivations

2.2.3 Links With The Thesis

The purpose of this Thesis is to ease the process of annotating large music collections by using content–based similarity distance as a way to propagate labels among songs, and thus reduce the effort, in both time and cost, of manual annotation. It is clear then that understanding the meaning and the motivation of annotation is crucial for this Thesis.

On the other hand, as it will be described in chapter 3, in order to validate the propagated labels, either a Ground Truth of manually annotated songs, or user feedback is required. For the latter case, users will interact with a search engine that retrieves songs based on these keywords. From these retrieved songs, users will be able to give feedback to the labels associated with each song.

2.3 Related Research

In this section, state-of-the-art of multimedia annotation, description and retrieval is reviewed. Subsection 2.3.1 reviews the fields of image, video and sound effects, whilst subsection 2.3.2 reviews the current state-of-the-art of the music information retrieval field.

2.3.1 Image, video and sound effects annotation

In other research areas, like image retrieval, there have been several proposals for retrieving image files, using keyword-based techniques [Baeza-Yates 00] and/or content-based techniques.

Wenyin et al. [Wenyin 01] proposed a semi–automatic strategy for semantically annotating images, that is better than manual annotation in terms of efficiency and better than automatic annotation in terms of accuracy. This strategy aims to create and refine annotations by "encouraging the user", to give *relevance feedback* [Lu 00] of the retrieved results. That is, let the user confirm if an annotation is correct or wrong for a given image. The conclusion they made is that images' annotation percentage would increase without too much user effort. This increase would be larger if an initial amount of the images' collection, for example 10%, is manually annotated.

In video techniques, Song et al. [Song 05] proposed a semi–automatic video annotation strategy for video semantic classification, using relevance feedback to refine the classification, and active learning process to speed up the automatic learning process of classifying videos, by labeling the most informative samples.

A similar approach to this master Thesis is Alipr[8], an image search engine that retrieves images relevant to a text–based query, or similar to an image — uploaded in real time. Each image has two links to get the most similar images to it. One of this links is "visual similar", which returns the most similar images based on the content of the images. The other link is "related", which returns the most related images based on the annotations (tags) of the images.

In the context of sound effects annotation — like other different domains — there is a need to manage the high amount of meta–data that corresponds to the audio files, in order to make it easier to search and retrieve. One of the main problems in annotation is the ambiguity of the natural language used to create these annotations. Thus, the use of taxonomies or classification schemes has become very important for this purpose, in order to avoid this ambiguity problem.

Cano et al. used an existing lexical network, WordNet[9], that contains over 100.000 concepts of the real word, as a set of synonyms, or *synsets*, with links among them; and extended it with new semantic descriptors and perceptual descriptors synsets of sounds in order to manage the taxonomy of sound effects (SFX) [Cano 04c]. In other related

[8] http://www.alipr.com/

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

work [Cano 05a] they used an extended WordNet version, and a k-NN[10] algorithm to classify the sound effects and to automatically annotate them. Both studies used a repository that centralizes audio content, corresponding meta data, taxonomies and algorithms [Cano 04b].

2.3.2 Music Field

In traditional music information retrieval (MIR) systems, i.e. search systems offered by commercial music portals (like for instance Audio Yahoo[11] or Altavista Audio[12]), user interaction is very limited. They make use of simple meta–data like artist name, album title, song title or year, and optionally subjective meta–data like the style of the music pieces. Therefore, users must already know in advance the music they want to find — at least one of the meta–data labels must be known. This option does not help to retrieve songs from unknown, new or not so popular artists, thus reducing their probability of becoming popular. We will not enter into a discussion, but just consider that the Internet is a network for everybody and therefore music search engines should be for everybody.

Hopefully, research in annotation and social communities in the MIR field has grown in interest in the last years. For the best of our knowledge there are three main groups doing research on music description, focusing on annotation and retrieval.

Whitman et al. in MIT labs and Ellis et al. at Columbia University are working on the extraction of acoustical and cultural information from music [Ellis 06] in order to understand the meaning of words and descriptions for music [Whitman 02a] [Whitman 02b] [Whitman 02c] [Whitman 04] [Berenzweig 04] [Whitman 05].

In [Whitman 02a] the authors propose some methods for unsupervised learning of unstructured music profiles retrieved from the web, with the purpose of understanding the "semantic profile" of an artist through a "feature space that maximizes generality and descriptiveness". These methods would help to infer artists' descriptions, represented as vector spaces, and similarity between artists by means of a peer–to–peer similarity.

[Berenzweig 04] examines and evaluates different approaches of both acoustic and subjective information of music pieces, for evaluating the performance of these approaches when computing similarity between artists.

Whitman and Rifkin present in [Whitman 02b] a query–by–description (QBD) system

[10] http://en.wikipedia.org/wiki/Nearest_neighbor_search

[11] http://audio.search.yahoo.com/

[12] http://www.altavista.com/audio/

that makes use of language processing, information retrieval and machine learning techniques for returning results to queries such as “rock with guitar riffs”. Their system treats the relation between web–based descriptions and music content as a 'severe multiclass' problem, using a novel machine learning technique: regularized least–squares classification (RLSC). In a posterior work [Whitman 03], Whitman et al. extend this technique by using a “linguistic expert”, Wordnet[13], a lexical database, for finding parameter spaces that would help to describe better and more precisely the artists' descriptions.

Knees, Schedl et al. within Johannes Kepler University Linz are working on the combination of audio content–based similarity with web–based data extracted from a web search engine in order to build a music search engine based on query by description [Knees 07b] — queries such as “rock with guitar riffs”— and automatic playlist generation [Knees 06]. In this sense, they are using both content and context information to improve retrieval quality. Their strategy proceeds as follows:

- Context information: from the ID3 meta–data tags associated with each song, retrieve relevant web pages using Google, and process these web pages in order to represent the music pieces by term vectors.

- Content information: for each audio track, they compute 19 Mel Frequency Cepstral Coefficients (MFCC) on short–time audio segments; then these MFCC's are represented as a Gaussian Mixture Model (GMM). The similarity between music pieces is obtained computing similarity between GMM's, using a symmetrized Kullback-Leibler divergence metric. The content information is used for reducing the dimensionality of the term vectors obtained from web–based data.

The main drawback of this approach is that the system is dependent of the usage and limitations of the Google API. In order to solve this limitation, in [Knees 07c] they modify the system by using a local web page index for query–vector construction.

Their latest work [Knees 07c], [Knees 07a], is the inclusion of Rocchio's relevance feedback method [Rocchio 71] to improve the quality of (personal) retrieval.

Barrington, Turnbull et al. at University of California, San Diego, are working on semantic annotations for audio information retrieval [Barrington 07], [Turnbull 07a].

In [Barrington 07], Barrington et al. propose a query-by-semantic-example audio information retrieval system, based on semantic concept models learned from a data set that

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

contains audio examples and their associated text captions. The semantic information is then used to compute similarity and to retrieve semantically meaningful audio pieces from the collection. In a similar work [Turnbull 07b] they take profit of the semantic concept models to propose a query-by-semantic-description audio information retrieval system, that learns the relationship between acoustic data and words by using an audio data set (which they call CAL500) with associated semantic annotations. For the task of learning, they adapt the supervised multiclass labeling (SML) model, with which they provide semantic multinomial distributions of words over a vocabulary.

2.4 Related Work

Within the music information retrieval domain, some of the challenges of the last years have been the development of search engines specialized in the music domain, and the automatic creation of playlists. For the formers, although still not comparable to the efficiency of web search engines, they are building the blocks for the next generation of music in the web.

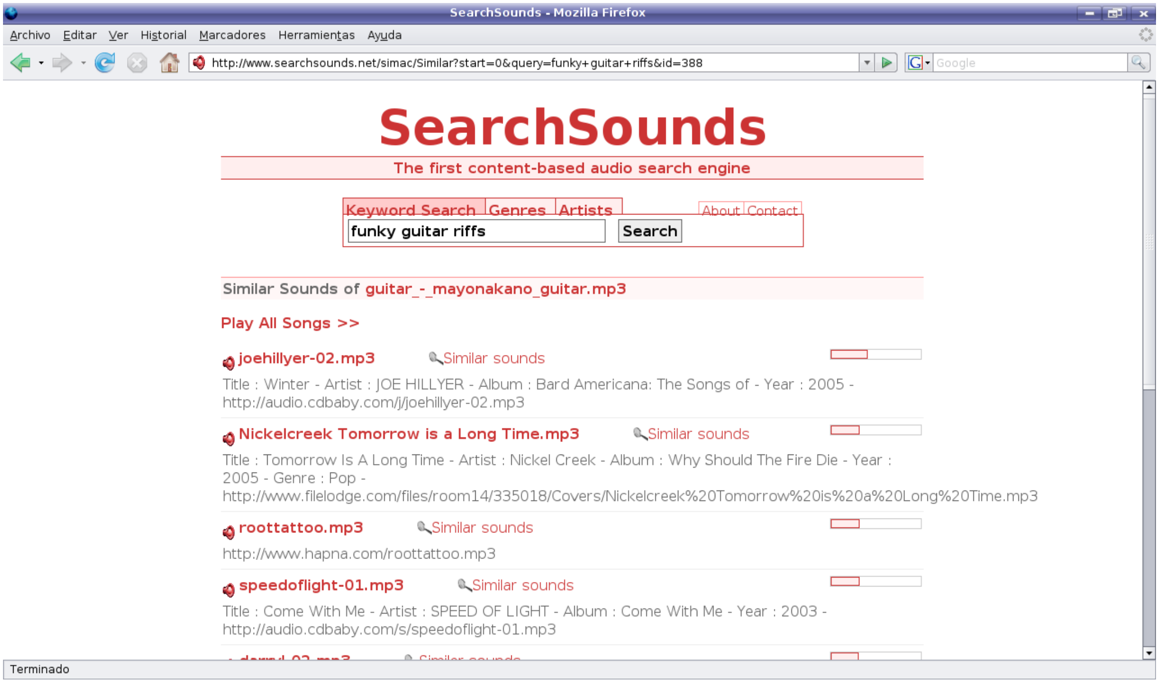
2.4.1 Music Search Engines

State-of-the-art of MIR field has led to the possibility of formulating a query in different ways, such as :

- Query by example
- Query by description
- Query by humming

Although the first and third types have been very important within MIR research, query by description stands out for being the simplest case. The simplicity refers to the ease of processing of queries by a search engine, it has nothing to do with ease of use by users. Interestingly enough, users that have little or no knowledge about music would find it difficult for formulating a query such as "funky guitar riffs", as it has been observed by Knees in [Knees 07b], [Knees 07c].

On the other hand, music search engines may be dependent of context–only information, content–only information or both content and context, in order to retrieve the most relevant songs for a given query.


SearchSounds - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda
http://www.searchsounds.net/simac/Similar?start=0&query=funky+guitar+riffs&id=388
Google
SearchSounds
The first content-based audio search engine
Keyword Search
Genres
Artists
About
Contact
funky guitar riffs
Search
Terminado

Similar Sounds of guitar_-_mayonakano_guitar.mp3

Play All Songs >>

joehillyer-02.mp3 Similar sounds
Title : Winter - Artist : JOE HILLYER - Album : Bard Americana: The Songs of - Year : 2005 - http://audio.cdbaby.com/j/joehillyer-02.mp3

Nickelcreek Tomorrow is a Long Time.mp3 Similar sounds
Title : Tomorrow Is A Long Time - Artist : Nickel Creek - Album : Why Should The Fire Die - Year : 2005 - Genre : Pop - http://www.filelodge.com/files/room14/335018/Covers/Nickelcreek%20Tomorrow%20is%20a%20Long%20Time.mp3

roottattoo.mp3 Similar sounds
http://www.hapna.com/roottattoo.mp3

speedoflight-01.mp3 Similar sounds
Title : Come With Me - Artist : SPEED OF LIGHT - Album : Come With Me - Year : 2003 - http://audio.cdbaby.com/s/speedoflight-01.mp3

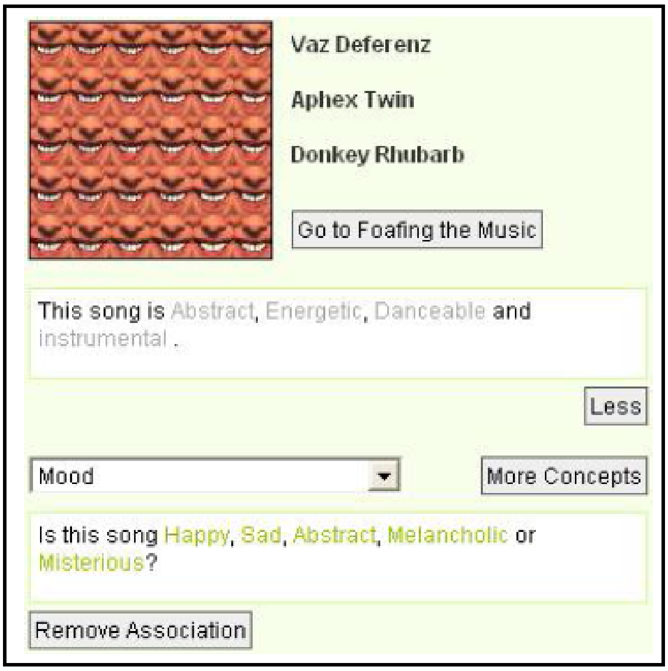
Figure 2.4: SearchSounds: User Interface.

SearchSounds

SearchSounds, presented by Celma et al. [Celma 06b], is a music search engine. It includes an audio crawler that mines audio web logs (also known as MP3 blogs) based on the syndicated web content of this web log, using RSS[14] to mine the artist or music information available in this RSS and also links to download the music files, if they are available and no copyright restriction is present.

The mined data is available to be retrieved by the search engine. A user can then make a text query, related to the artist's name, the song's name, or even more sophisticated queries like "funky guitar riffs" or "traditional folklore tunes". The result of the query is a set of songs relevant to the query. Based on this query, the system makes available a *Similar* button for each audio file, that retrieves the most similar audio files to this song, using a set of mid–level descriptors extracted directly from the audio file, like harmony, rhythm, timbre and instrumentation, intensity, structure and complexity. Please refer to [Herrera 05a] for a deeper explanation of these descriptors. This new kind of approximation makes users discover and explore new music that would have been very difficult to

[14] Really Simple Sindication (RSS 2.0), Rich Site Summary (RSS 0.91 and 1.0) or RDF Site Summary(RSS 1.0)



Vaz Deferenz

Aphex Twin

Donkey Rhubarb

This song is Abstract, Energetic, Danceable and instrumental .

Is this song Happy, Sad, Abstract, Melancholic or Misterious?

Figure 2.5: Good Vibrations: User Interface.

do by using only text-based queries.

2.4.2 Music Annotation Systems

In the context of music annotations, we can find Good Vibrations, developed by Sandvold et al. [Sandvold 06]. Good Vibrations is a tool (actually is a media player plug-in) for music annotation, exploration and discovery. Initially, user starts to annotate songs manually and the tool starts an active learning process. After some hours of this active annotation, the tool starts automatically proposing the possible proper tags to the user, who can obviously correct them. The goal is to finally generate and propose play lists according to the user's concepts. This is a clear example of personomy, since it is an individual's annotation.

Another approximation is Last.fm[15]. Last.fm is a community–based system for sharing music information. It has a downloadable media player plug-in that recollects information from users' listened songs and send it to Last.fm website to add it in users' profiles. Users can add manually tags to a song before the plug-in sends the information about it. The

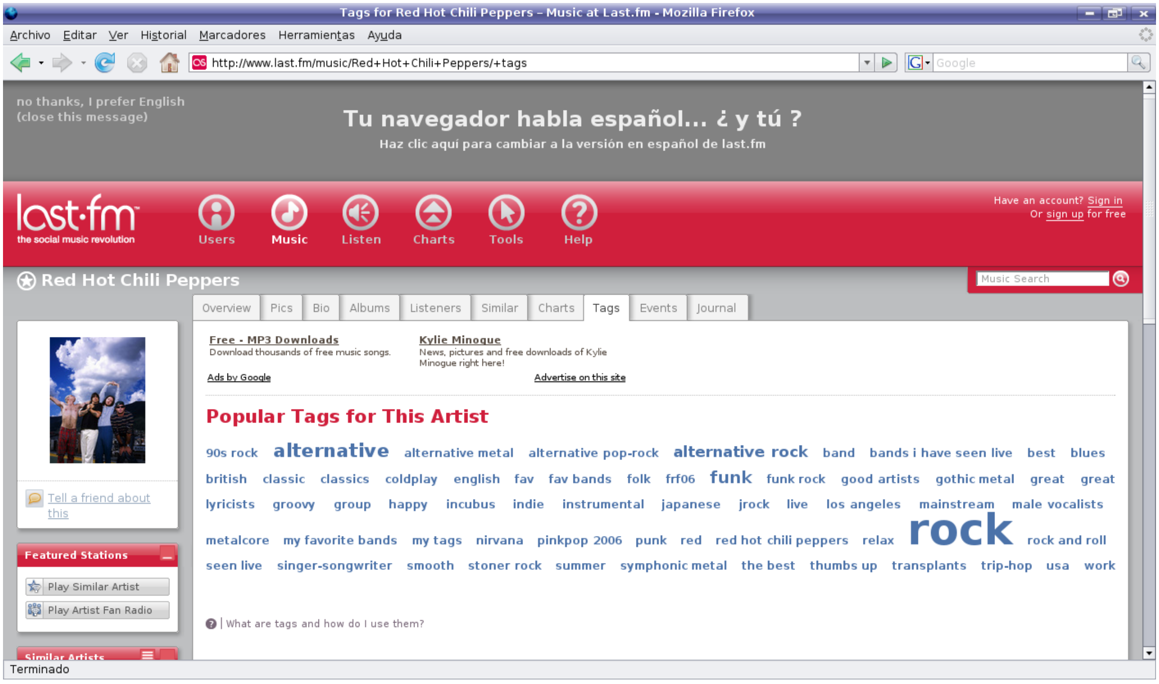
[15] http://www.last.fm/

tags are recollected from users' profiles and organized by artists. Here is an excerpt of an artist's collection of tags:

```
<toptags artist=“Metallica”>
    <tag>
        <name>metal</name>
        <count>100</count>
        <url>http://www.last.fm/tag/metal</url>
    </tag>
    <tag>
        <name>heavy metal</name>
        <count>22</count>
        <url>http://www.last.fm/tag/heavy%20metal</url>
    </tag>
    <tag>
        <name>seen live</name>
        <count>5</count>
        <url>http://www.last.fm/tag/seen%20live</url>
    </tag>
</toptags>
```

Another functionality is that given a tag, Last.fm returns a list of artist that are the most annotated with this tag. Last.fm is a clear example of folksonomy, since it uses annotations from a community of users.

However, it has some limitations: it is not song based, the annotations are organized by artists, not by songs. This has been a classical limitation of meta–data based music search engines, since there is very little information about a single song in web search engines compared to artist information. Another limitation is that it follows a collaborative filtering approach, no similarities in music content are taken into account, so it is not possible to semi–automatically propagate useful tags.



no thanks, I prefer English
(close this message)

Tu navegador habla español... ¿ y tú ?

Haz clic aqui para cambiar a la versión en español de last.fm

Have an account? Sign in
Or sign up for free

Red Hot Chili Peppers

Free - MP3 Downloads
Download thousands of free music songs.

Kylie Minogue
News, pictures and free downloads of Kylie Minogue right here!

Ads by Google

Advertise on this site

Popular Tags for This Artist

90s rock alternative alternative metal alternative pop-rock alternative rock band bands i have seen live best blues british classic classics coldplay english fav fav bands folk frf06 funk funk rock good artists gothic metal great great lyricists groovy group happy incubus indie instrumental japanese jrock live los angeles mainstream male vocalists metalcore my favorite bands my tags nirvana pinkpop 2006 punk red red hot chili peppers relax rock rock and roll seen live singer-songwriter smooth stoner rock summer symphonic metal the best thumbs up transplants trip-hop usa work

What are tags and how do I use them?

Figure 2.6: Artist tag cloud in last.fm

Chapter 3

Own approach

3.1 Motivation

In the previous chapter music information retrieval, music similarity based on content mid-level descriptors, and music annotation were briefly explained. Some related research, in the music field and other areas, were introduced; and some related work, that are the inspiration and the point of departure for this master thesis were explained.

The problem to be studied and solved is the following: the number of available media is increasing exponentially, and that makes it hard for the users to search, navigate, and retrieve. Annotation is a way for describing and giving more value to multimedia content. But annotating the huge amount of available music files is very hard and costly. The need is then to propose strategies to make this process lighter. Other ways of retrieval, based on the similarity between multimedia content have been created and turned to be very useful. Different approaches and strategies that combines these described features have been made for this purpose, most of them in other research areas. In the context of the music domain, there are also some strategies devoted to the topic. So why then this master Thesis exists?

3.1.1 Related systems

Let's go through the characteristics of the related work presented in section 2.4. The SearchSounds system retrieves music files similar to a given one taking into account only the content mid–level descriptors, no annotation is present. The work by Cano et. al

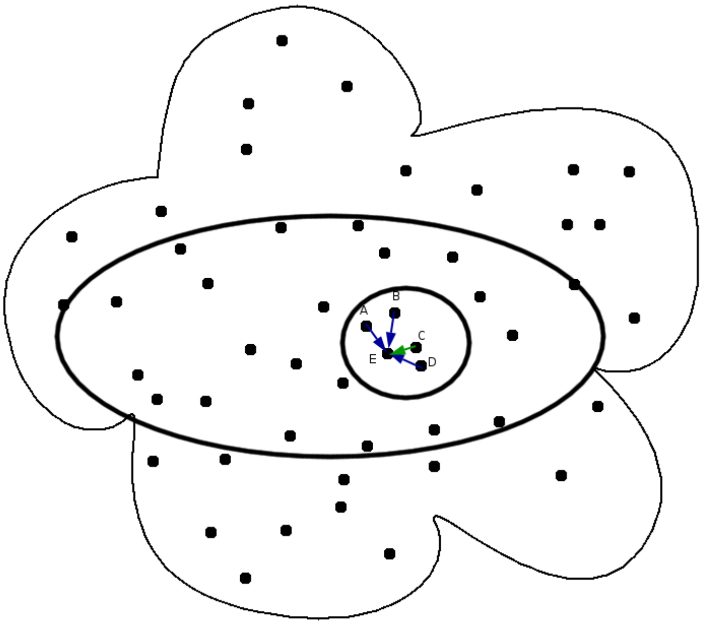


Figure 3.1: Possible structure of a music network

[Cano 05a] is addressed for sound effects and short audio excerpts, not for whole music files. Last.fm's annotations are used for playlist generation, or to refine the collaborative filtering approach, and no similarities in content are used at the moment. Good vibrations is more or less the idea of this Thesis, but it is individual, that is to say, is addressed to individual use, not for a community.

3.1.2 Description

The purpose of this master Thesis is then as follows: study and propose a semi–automatic strategy for annotating music collections, that would be better than manual annotation in terms of efficiency (speed), and better than automatic annotation in terms of accuracy.

Having annotated a large amount of music files makes it easier to retrieve and to manage them. There are several points that have to be taken into account. In figure 3.1 we can observe a possible "network" of the music collection. Each node represents a music file. The distance between the nodes is based on the content similarity of the music files: closer nodes are similar than the distant nodes. In the small circle we can see 5 nodes, represented by capital letters A,B,C,D and E. Suppose that A,B,C and D are already annotated and are the closer nodes to node E, which is one of the not-annotated nodes.

A propagation function could be applied to propagate tags present in A,B,C, and D to E. After annotating the node E, it would be useful (necessary) that a user can correct these propagated annotations if he/she thinks that they are wrong.

The propagation will be applied for:

- A song already in the collection (expanding annotations).
- A new added song in the whole collection (avoiding the cold-start problem for finding new songs).

Obviously, there are some annotations that could not be propagated, like lyrics' language of a music file, the song name, etc. As an example, suppose that two music files are close in distance: A song from the Beatles and a song from a Spanish group, "La oreja de van gogh"; they are both pop songs, but that does not imply that the lyrics of the song from La oreja de van gogh will be in English.

On the other hand, One of the main characteristics for this new approach is that it will be community–based, that is, a community of users will have contact — by means of an interface — with the music files. This characteristic raises a big problem and a big challenge at the same time. Annotated music files will have feedback from users, therefore some users can consider that some annotations of an audio file are not correct, and they will be able to change these annotations. It would be great that all the users describe the audio file in the same way, and no contradictions exist, but this is actually unfeasible and unreal. Figure 3.2 illustrates the problem. A new annotated music file through propagation from other music files, represented by letter E in that case, is supposed to have not been well annotated. Two users are trying to correct it, one of them says that the song is sad, the other one that the song is happy, so what will be the annotation of this music file? A validation process is a must to avoid this ambiguity. These issues are described in section 3.2.4.

3.2 System Overview

The development part of the Thesis has been focused on:

- The creation of a simple label propagation process (see subsection 3.2.2).

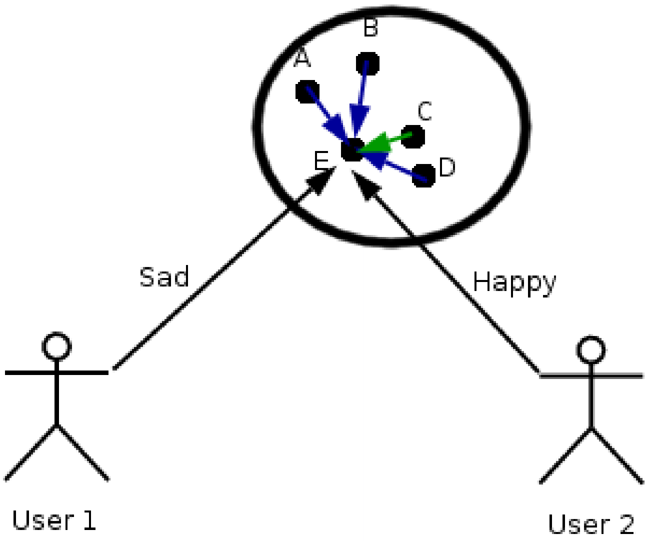


Figure 3.2: Annotation problems in a community

- The addition of annotation functionality to an already existing system: the SearchSounds[1] (see subsection 2.4.1 for more details about SearchSounds).

The following subsections describe all the steps followed for the development of this approach, starting from the collection of manual annotations (subsection 3.2.1), then moving to the tag propagation process, the User Interface (subsection 3.2.3) that allows users to annotate songs giving relevance feedback (subsection 3.2.4) about the propagated labels. Finally, subsection 3.2.5 describes some implementation details.

3.2.1 Gathering Information

The music collection used in this Thesis consisted of:

- A considerable number of music pieces' information that were already in the SearchSounds collection (around 258,000 songs).
- About 5,500 new songs borrowed from the Magnatune[2] creative–commons collection.

[1] http://www.SearchSounds.net/
[2] http://www.magnatune.com/

The Ground Truth

An important aspect while gathering information was to collect a set of manual annotations for the collection, in order to get a Ground Truth that would be useful during the evaluation of this approach. We obtained the manual annotations from three different sources:

- Cdbaby[3], a " little online record store that sells albums by independent musicians". This was the source that gave more annotations. When crawling this website some labels at artist level were also retrieved. These labels were then propagated to every artist's song. Maybe it is not an efficient way of creating a ground truth, however, as we will see in subsection 3.2.4, user feedback would help to refine this data.
- Magnatune: the list of 5,400 songs came with annotations about the style of the song. Many songs got more than one style label.
- SearchSounds collection: labels were inferred from the id3 genre tags from songs in the collection. Irrelevant ID3 genre tags found in the collection, such as *other* or *unknown*, were not taken into account.

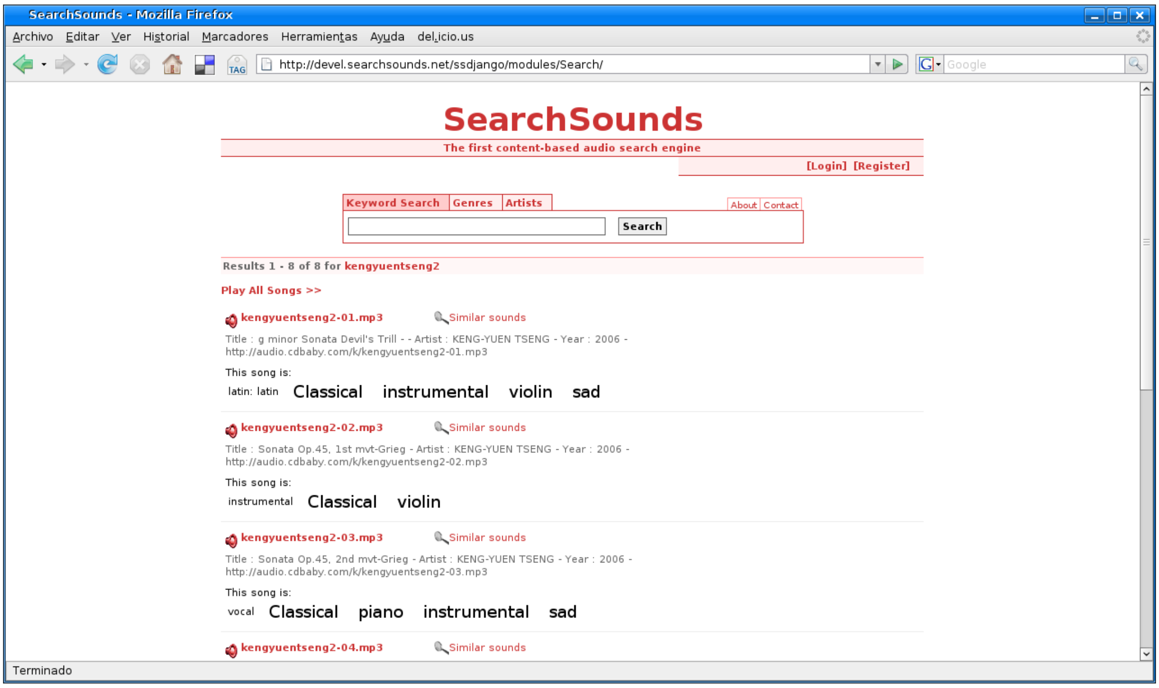
All these sources covered around 48% of the database, so to sum up, the Ground Truth consisted of ∼124,000 songs annotated with one label per song — those from cdbaby, and ID3 genre tags — plus ∼5,400 songs annotated with one or more labels — those from the magnatune collection.

3.2.2 The Tag Propagation process

The tag propagation process is a very simple algorithm that proposes labels for yet not annotated songs by using content–based similarity as a way to propagate labels.

The algorithm proceeds as follows: from a content–based similarity system, retrieve the *i-th* most similar songs to a given seed song. From this list of similar songs, compute the overlap of labels: if a label is used in more than 20% of the similar songs, then it is considered as a proposed label for the seed song. In other words, every label with a frequency greater or equal than 20% is propagated to the seed song, and the percentage frequency decides how important is the propagated tag.

[3] http://cdbaby.com/


SearchSounds - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda del.icio.us
TAG
http://devel.searchsounds.net/ssdjango/modules/Search/
Google
SearchSounds
The first content-based audio search engine
[Login] [Register]
Keyword Search
Genres
Artists
About
Contact
Search
Terminado

Results 1 - 8 of 8 for kengyuentseng2

Play All Songs >>

kengyuentseng2-01.mp3 Similar sounds

Title : g minor Sonata Devil's Trill - - Artist : KENG-YUEN TSENG - Year : 2006 -
http://audio.cdbaby.com/k/kengyuentseng2-01.mp3

This song is:

latin: latin Classical instrumental violin sad

kengyuentseng2-02.mp3 Similar sounds

Title : Sonata Op.45, 1st mvt-Grieg - Artist : KENG-YUEN TSENG - Year : 2006 -
http://audio.cdbaby.com/k/kengyuentseng2-02.mp3

This song is:

instrumental Classical violin

kengyuentseng2-03.mp3 Similar sounds

Title : Sonata Op.45, 2nd mvt-Grieg - Artist : KENG-YUEN TSENG - Year : 2006 -
http://audio.cdbaby.com/k/kengyuentseng2-03.mp3

This song is:

vocal Classical piano instrumental sad

kengyuentseng2-04.mp3 Similar sounds

Figure 3.3: Retrieved songs for a query in the system when a user is not logged in.

The content–based similarity can be seen as a black box. That is to say, given a seed song, the module returns a list of the *i-th* most similar songs. This study employs a content–based module that considers not only timbrical features (e.g. MFCC), but some musical mid–level and high–level descriptors, related to rhythm, tonality, etc [Cano 05b], [Gouyon 05], [Gómez 06].

3.2.3 User Interface

Once collected a set of manually annotated songs as a Ground Truth and propagated labels among not yet annotated songs based on the content–based similarity distance, the next step is to visualize the results in a user interface. For this purpose, we used an already existing music search engine — SearchSounds — and extended its functionality in order to support labels' visualization and user management.

The new features added to SearchSounds can be summarized as follows:

- Addition of a tag cloud for each song representing the song's labels. The labels' size is relative to each song. The biggest and boldest the label, the most representative the label is for the song.

- User management. A user can register a username and login in order to interact with the labels, so personalized recommendation can be made for each user.
- Confirmation/rejection button links below each label, that enable users confirm or reject the label, respectively.
- "Annotate this song" functionality, that expands the annotation form to add new labels for the song. This form consists of a select input with 6 different options representing 6 labels' categories (genre, instrumentation, intensity, mood, gender and lyrics).
- "Tag expansion" feature, which consists on expanding the number of tags for a given song with WordNet[4] (mainly using *hyponyms* and *hypernyms*). When a user formulates a query, the system will use this additional information in order to improve the retrieved results.

We can observe the first feature in figure 3.3. When a user is not logged in, he/she can only see the tag clouds, but not interact with them. Figure 3.4 shows the tag clouds with confirmation/rejection button links below each label, except one. That is because the user has already confirmed or rejected the tag and he/she cannot annotate the song with the same label twice.

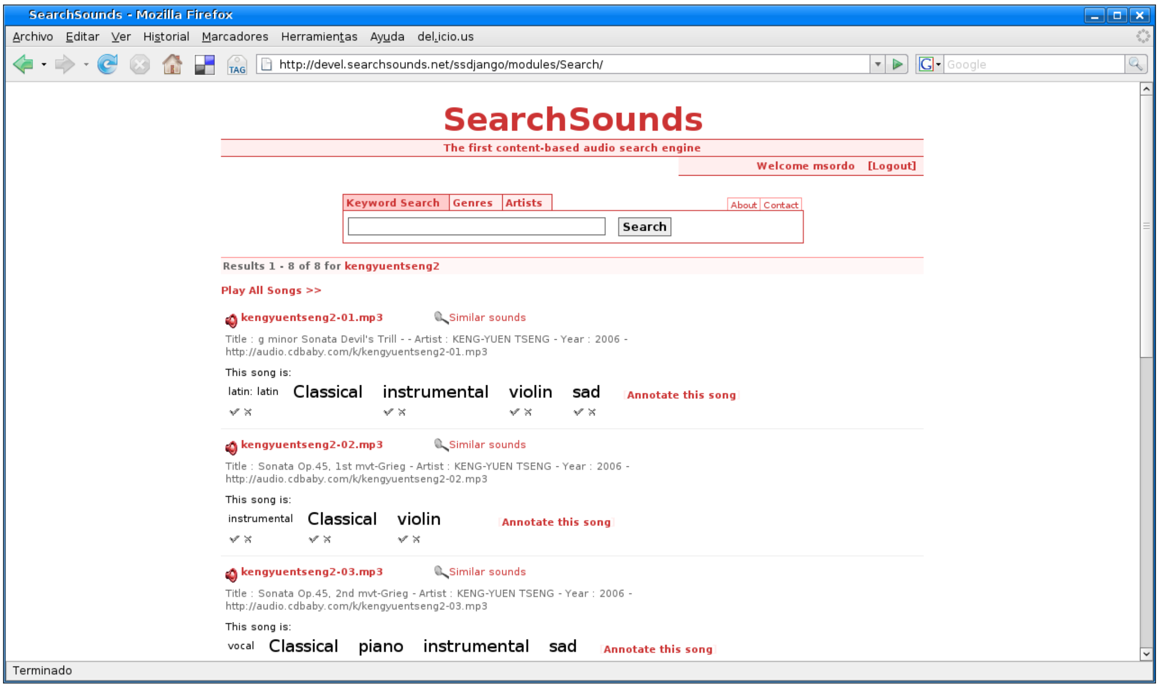
If the user thinks that he/she can add a new label for a song, he/she has to click on the "Annotate this song" link that will expand the annotation form, as it is shown in figure 3.5. We used 6 different label categories because we found it useful to classify the tags in categories. However, the annotation is still free, there is no restriction on using whatever natural words. Please refer to chapter 5 for a discussion about this topic.

3.2.4 Relevance Feedback

Relevance feedback[Rocchio 71] is a feature added originally to information retrieval systems that takes results retrieved from a given query, and uses some kind of information about whether or not the returned results are relevant to the query, in order to reformulate the query or refine the results. There are three types of feedback:

- Explicit feedback, where the user explicitly mark results as relevant or not, either ranking them or putting them in place.

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


SearchSounds - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda del.icio.us
http://devel.searchsounds.net/ssdjango/modules/Search/
Google
Keyword Search
Genres
Artists
About
Contact
Search
Terminado

SearchSounds

The first content-based audio search engine

Welcome msordo [Logout]

Results 1 - 8 of 8 for kengyuentseng2

Play All Songs >>

kengyuentseng2-01.mp3 Similar sounds

Title : g minor Sonata Devil's Trill - - Artist : KENG-YUEN TSENG - Year : 2006 - http://audio.cdbaby.com/k/kengyuentseng2-01.mp3

This song is:

latin: latin Classical instrumental violin sad Annotate this song

kengyuentseng2-02.mp3 Similar sounds

Title : Sonata Op.45, 1st mvt-Grieg - Artist : KENG-YUEN TSENG - Year : 2006 - http://audio.cdbaby.com/k/kengyuentseng2-02.mp3

This song is:

instrumental Classical violin Annotate this song

kengyuentseng2-03.mp3 Similar sounds

Title : Sonata Op.45, 2nd mvt-Grieg - Artist : KENG-YUEN TSENG - Year : 2006 - http://audio.cdbaby.com/k/kengyuentseng2-03.mp3

This song is:

vocal Classical piano instrumental sad Annotate this song

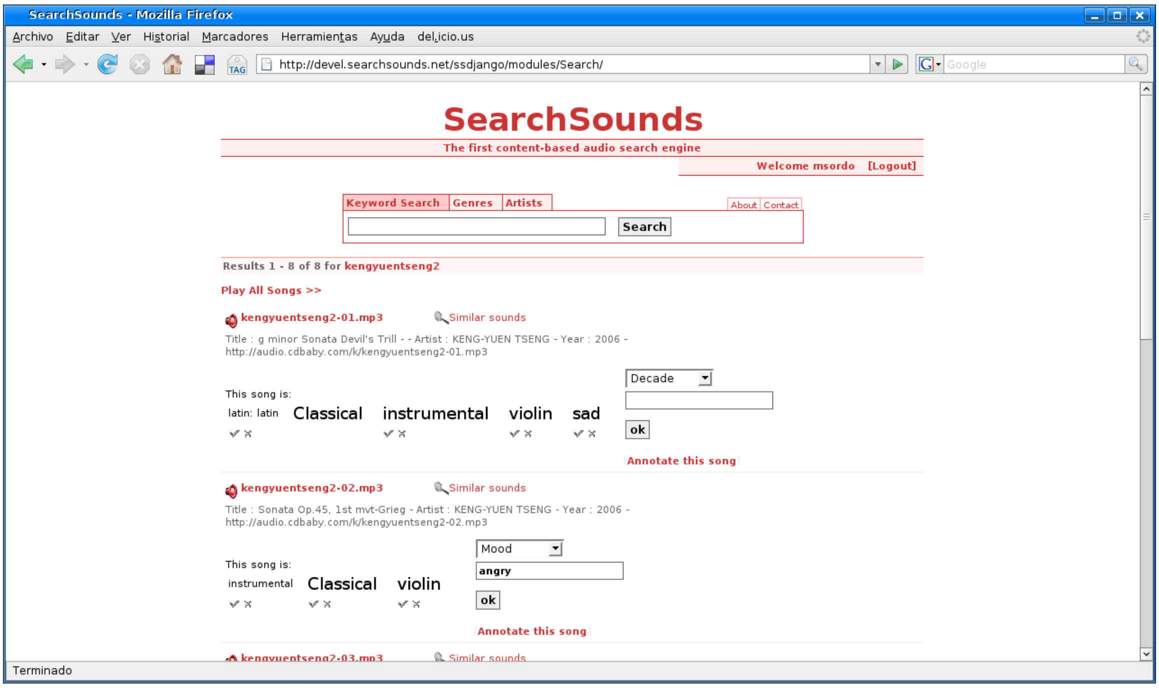
Figure 3.4: Interacting with the tag clouds in the system when a user is logged in.

- Implicit feedback, inferred from user behavior. The behavior could be which results are actually selected by the user, or how long does this user view/read/reproduce the results.

- Blind feedback, also known as "pseudo" feedback, where the top m documents are considered the most relevant.

The process of relevance feedback is iterative. Generally speaking, the user gets a list of the most similar songs to a given query. After examining the retrieved list — either listening to the songs if possible, or just consider the songs' meta–data — he/she can mark which are the most relevant in his/her opinion. This is a clear example of explicit feedback.

In the case of our system, the user is presented with a list of songs, each song with a tag cloud representing the labels for this song and a confirmation/rejection button for each label, so the user confirms or rejects that label. Each annotation is given a score the first time the label is used in a song. Confirming or rejecting this label will increase or decrease the score, respectively. Once a label in a song reaches the score of 0, it is considered not valid for the song anymore, and it is deleted. Thus, the task of the user is to decide which


SearchSounds - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda del.icio.us
TAG
http://devel.searchsounds.net/ssdjango/modules/Search/
Google
SearchSounds
The first content-based audio search engine
Welcome msordo [Logout]
Keyword Search Genres Artists
About Contact
Search
Play All Songs >>
Similar sounds
Annotate this song
ok
Terminado

Results 1 - 8 of 8 for kengyuentseng2

kengyuentseng2-01.mp3

Title : g minor Sonata Devil's Trill - - Artist : KENG-YUEN TSENG - Year : 2006 - http://audio.cdbaby.com/k/kengyuentseng2-01.mp3

This song is: latin: latin ✓ ✗ Classical instrumental ✓ ✗ violin ✓ ✗ sad ✓ ✗

Decade

kengyuentseng2-02.mp3

Title : Sonata Op.45, 1st mvt-Grieg - Artist : KENG-YUEN TSENG - Year : 2006 - http://audio.cdbaby.com/k/kengyuentseng2-02.mp3

This song is: instrumental ✓ ✗ Classical ✓ ✗ violin ✓ ✗

Mood

angry

kengyuentseng2-03.mp3

Figure 3.5: Annotating a song in the system when a user is logged in.

labels are relevant or not for a song. After a certain number of evaluations by different users over a same set of songs, the system will proceed to “re–propagate” labels within the affected songs. After that, users will be able to “re–evaluate” the songs’ labels, so it is a clear example of an iterative process.

3.2.5 Implementation details

The implementation language of the development part of this Thesis has been Python[5], a fast prototyping object–oriented compatible scripting language. The reason for using this language is because the implementation of the original SearchSounds is in Python. All the scripts used for collecting the manual annotations (subsection 3.2.1) and for propagating labels (subsection 3.2.2) were written in Python.

Regarding the data structure model, figure 3.6 shows the entity-relationship model (ERM). As can be observed, it is an M–N–P relationship. A special mention to the “Annotation” associative entity[6], that restricts the use of a tag for a song by a user to

[5] http://www.python.org/

[6] Annotation should be understood here as a tuple of (audio, tag, user).

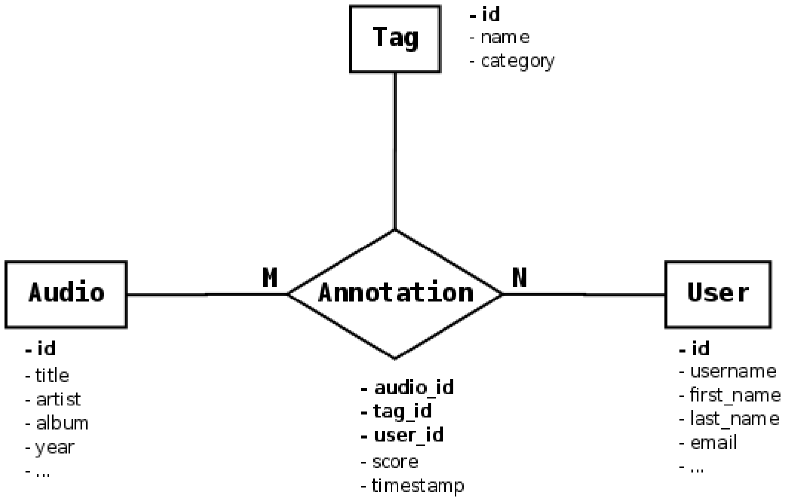


Figure 3.6: Entity-Relationship Model of the system.

one. That is, a user cannot annotate a song with the same tag twice. Another important aspect of the "Annotation" entity is the score attribute. Every new annotation is given a score of 20[7], except the propagated annotations[8] that have a given score of 10, since these annotations are in principle the less confident. The manual annotations that form the ground truth are given a score of 20 as well.

In the system's user interface, when a user formulates a query, a list of the most relevant music pieces is returned, each one with a tag cloud (except music pieces that do not have labels yet), and the user can confirm or reject these tags. When the user confirms a tag, the annotation score is incremented by one. It is decreased by one when the user rejects the tag. Moreover, the user can annotate a song with a new tag. In this case the annotation is new and thus it is given a score of 20.

[7] The scores may change in the future depending on the number of users.

[8] The propagated annotations are made by a "special" (automatic) user.

Chapter 4

Evaluation and Results

4.1 Approach I: Automatic Annotation

As a first approach, we have made some experiments on automatic annotation of music collections, by using a content–based similarity distance as a way to propagate labels among songs [Sordo 07]. The goal was to prove empirically how content–based similarity can help to propose labels to yet unlabeled songs, and thus reducing the hard effort of manually annotating songs. In our case we are using a content–based module that takes into account not only timbrical features (e.g. MFCC), but some mid–level descriptors such as: rhythm, tonality, etc.(see [Gómez 06], [Gouyon 05] for more details).

Two different experiments were done. The first one propagated labels that were related to the style of the piece, whereas the second experiment dealt with mood labels (happy, sad, angry, and mysterious).

4.1.1 Propagation of Music Style Labels

Procedure

We used two music collections from Magnatune as our Ground Truth, one with labels about styles of the songs, and the other with mood labels, both annotated by the Magnatune musicologist. The problem of Magnatune collections is that there is only one human that annotates the labels, when normally a Ground Truth of this nature should be pair-reviewed. Yet, we validated a large amount of the annotated songs by listening to them.

Style experiments

The Ground Truth for the style experiment consists of: 29 different labels (like Rock, Instrumental, Classical, Baroque, etc.), and 5,481 different annotated songs.

The process of evaluation was the following:

```
For each x-percent annotated collection
  For each i-number of similar songs
     For each song
        retrieve i-similar songs
        propose tags for this song based on the similar songs
        compute measures
     compute average measures
```

Where x goes from 10% to 50%, and i goes from 10 to 30. The proposed tags had a threshold of 0.2. That is to say, a tag to be proposed should appear in at least 20% of the i-th similar songs. The measures used will be explained in the next section.

For each experiment configuration we used an annotated set and a not yet annotated set. For example, with a 10% of annotated songs configuration, we had 548 annotated songs and 4,933 not–annotated songs, using the well–known leave–one–out cross–validation method.

A special case was for the whole collection (100%) annotated. In this case, we did the experiment with the whole annotated collection to test the reliability of our approach — that is based on propagating labels according to audio similarity. According to the results (see Table 4.1) with a recall of 84%, we conclude that the approach could be useful for the case that the collection is not fully annotated, which is a more real case. Finally, we studied as well whether the content–based similarity is affected by the "album and artist effect". That is to say that given a seed song, some of the most similar songs may belong to the same album, or to the same artist. To test this, we did not take into account these similar songs, as we will see in subsection 4.1.1.

Evaluation Metrics

The metrics used to evaluate the styles experiments were initially Precision/Recall and F-Measure. The Precision/Recall evaluation is defined as:

$$Precision = \frac{TP}{TP+FP} \quad Recall = \frac{TP}{TP+FN} \tag{4.1}$$

where TP represents True Positives (i.e. tags appearing on both the Ground Truth and the proposed tags), FP represents False Positives (i.e. tags appearing on the proposed tags but not on the Ground Truth), and FN represents False Negatives (i.e. tags appearing on the Ground Truth but not on the proposed tags). That is to say, Precision metric tells, from all the proposed tags, how many of them were correct (correctness means tags appearing on both the Ground Truth and the proposed tags); and Recall metric tells, from the Ground Truth, how many of them were correctly proposed. We did that for each not-annotated song (for the 10% annotated collection case, for example, we used a test collection that contains the other 90%, and we considered this 90% as not-annotated songs, following the leave–one–out cross–validation method), and then computed the average for each case.

The F-Measure is a weighted harmonic mean of Precision and Recall metrics. The general formula is:

$$F_\alpha = \frac{(1+\alpha)\cdot(Precision \cdot Recall)}{\alpha \cdot Precision + Recall} \quad \alpha \geq 0 \tag{4.2}$$

We used $\alpha = 2$ because we wanted to give more focus on the Recall value. In our case, Recall seems to be more informative since our purpose is to know how well the tags can be propagated.

Here is an example of evaluation for a single song:

Ground Truth for song id *usana31605*:

```
[Classical, Piano, Baroque, Instrumental]
```

Proposed tags and their frequency:

```
Instrumental: 0.55   Baroque:  0.25
Classical:    0.40   Invented: 0.20
```

The frequency of each tag is computed based on the tags collected from the i–th most similar songs of the seed song with id usana31605.

Precision/Recall measure:

```
Precision: 0.75 (Invented tag)
Recall: 0.75 (Piano is missing)
```

The Recall, as said before, is more informative in this case, but it does not take into account the frequencies of the tags. Thus, we used the Spearman ρ metric as well. The Spearman's rank correlation coefficient, better known as Spearman ρ (rho), is defined as:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{4.3}$$

Where d_i represents the distance between each rank of pair of values — in our case labels in the Ground Truth and labels in the proposed tags — and n the number of all possible pair of values. To compute the distances we assume that the manually annotated labels have frequency of 1.

Results

For the style experiment, we ran different configurations and we computed the average metrics. A special case is when using the 100% annotated songs (see the results in Table 4.1). This experiment is used to test whether the content–based similarity is good for propagating labels. There are four different configurations when retrieving the most similar songs to a given one: do not apply any constraint, or filter by artist/album. The constraints, then, are: filtering the similarity results by same Artist, same Album, or by same Artist and Album. The latter case makes only sense when the songs appear in compilations, various artists albums, etc. When filtering by artist or by album we make sure that the most similar songs to a given one are not from the same artist or the same album. That of course decreases the Precision/Recall measure. We can see from the results, that to achieve more precision and recall when applying a constraint, we need to increase the number of similar songs, which makes sense because we are not taking into account similar songs that are closer to a given one.

Now, table 4.2 shows the results of propagating a partially annotated collection. The Spearman ρ coefficient, as well as Precision/Recall and F_2-measure, grows when increasing

Similars	Constraint	Precision	Recall	F_2	Spearman ρ
10	None	0.56	0.84	0.72	0.51
	Artist	0.41	0.58	0.51	0.23
	Album	0.50	0.71	0.62	0.34
	Artist & Album	0.43	0.59	0.53	0.19
20	None	0.56	0.82	0.71	0.49
	Artist	0.48	0.61	0.56	0.26
	Album	0.53	0.72	0.64	0.35
	Artist & Album	0.48	0.61	0.56	0.24
30	None	0.60	0.77	0.70	0.45
	Artist	0.50	0.58	0.55	0.28
	Album	0.56	0.67	0.63	0.37
	Artist & Album	0.50	0.59	0.55	0.27

Table 4.1: Experiments with the 100% annotated collection. The Precision/Recall measure, the F_2-measure and the Spearman ρ measure are proportional to the number of similar songs. When constraints are present, these measures decrease.

the percentage of songs annotated in the collection. Interestingly enough, the values decrease when increasing the number of neighbors (from 10 to 30) for a given song.

Annotation	Similars	Precision	Recall	F_2	Spearman ρ
20%	10	0.32	0.29	0.30	0.24
	20	0.22	0.17	0.19	0.16
	30	0.08	0.05	0.06	0.06
40%	10	0.57	0.59	0.58	0.43
	20	0.56	0.52	0.53	0.41
	30	0.49	0.39	0.42	0.34
50%	10	0.61	0.67	0.64	0.47
	20	0.61	0.61	0.61	0.45
	30	0.57	0.51	0.53	0.41

Table 4.2: Experiments with the 20%, 40% and 50% annotated collection. The Precision, Recall and F_2-measure and the Spearman ρ values grow with a higher percentage of annotated songs, and a smaller number of similar songs.

Finally, we propose another experiment that is to automatically annotate songs in a music collection by means of the propagation process. The results are presented in Table 4.3. It is clear that the percentage of songs automatically annotated by content–based similarity increases when the number of already annotated songs grows. Nevertheless, we can see an interesting exception here, that is the 40% annotated collection performs better (up to 38.68% new propagated labels, with a low Recall (0.4) than the 50% one. This could be due to the random process of splitting the Ground Truth and the test set from the collection. Furthermore, we can see how the percentage of songs automatically annotated is inversely proportional to the number of similar songs used by the content–based similarity module — in contrast with the results from the 100% annotated collection, see Table 4.1, when applying any constraint.

		Propagation with Recall		
Annotation	Similars	> 0.8	> 0.6	> 0.4
20%	10	17.515%	21.365%	24.977%
	20	8.666%	12.352%	15.453%
	30	2.554%	3.758%	5.145%
40%	10	28.01%	33.46%	38.68%
	20	22.50%	28.92%	34.32%
	30	15.22%	20.82%	26.22%
50%	10	26.77%	31.62%	35.92%
	20	22.66%	28.74%	33.37%
	30	17.48%	23.15%	28.44%

Table 4.3: Extending annotations of a music collection by means of content–based similarity. We observe that the propagation grows with a smaller number of similars and a higher percentage of annotated songs, except for the case of 40% and 50%.

4.1.2 Propagation of Mood Labels

Procedure

For the moods experiment, the first issue is the choice of the taxonomy. As advised by Juslin et al. in [Juslin 01], in order to make our experiment and to build a Ground Truth that achieves the best agreement between people, we should consider few categories. We used a reduced version of the Magnatune online library. This collection offers a set of playlists based on mood[1]. We clustered the 150 mood playlists to fit in our few categories paradigm. The adjectives proposed by Juslin: happiness, sadness, anger and fear in

[1] http://www.magnatune.com/moods/

Mood	Songs
Happy	67
Sad	61
Angry	34
Mysterious	29

Table 4.4: Mood distribution of the Ground Truth

[Juslin 01] have been applied by Feng et al. in [Feng 03] and proved to give satisfying results. As the collection is mostly focused on popular and classical music, the "fear" adjective has been extended to a larger category called "mysterious". Using WordNet[2] we have joined the possible playlists together in the following four categories : happy, sad, angry and mysterious. Then, a listener was asked to validate each song label. We obtained a Ground Truth database of 191 songs with the distribution in mood shown in Table 4.4. For each song, there is only one mood label. It is not an equal distribution but there is enough data in each category to experiment with the content–based similarity.

Evaluation Metrics

To evaluate the mood results, we used two measures. First we wanted to check if the system was able to guess the correct mood label (there is only one possible label per song). We evaluated the Precision just considering the first result using Precision at 1, also called P@1.

$$P@1 = \begin{cases} 1, & best\, proposed\, label = real\, label \\ 0, & otherwise \end{cases} \tag{4.4}$$

We averaged this value over all the examples. This metric helps us to understand if the system can predict the correct mood label. However it does not take into account the relative frequencies. Then another measure would be needed to evaluate this aspect. We weighted the frequencies of the proposed label and normalized to compute a weighted Precision at 1, that we will call wP@1. It is equal to the frequency value of the correct label over the sum of all the proposed label frequencies:

$$wP@1 = \frac{freq.\, correct\, label}{\sum freq.\, proposed\, labels} \tag{4.5}$$

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

GT/Predicted	Angry	Happy	Mysterious	Sad
Angry	27	7	1	1
Happy	4	55	1	2
Mysterious	8	6	7	5
Sad	4	16	2	35

Table 4.5: Confusion matrix for the mood experiment with a 100% annotated collection.

Mood	P@1	wP@1
Angry	0.72	0.65
Happy	0.89	0.62
Mysterious	0.27	0.22
Sad	0.61	0.59
TOTAL	0.62	0.52

Table 4.6: P@1 and wP@1 values averaged for each mood

Results

To have an overview of the system performance for each mood, we built a confusion matrix in Table 4.5. It has been computed using 100% of the collection annotated. Each row gives the predicted mood distribution (considering only the best label) for each mood in the Ground Truth. Looking at the confusion matrix we observe that a content–based similarity approach can propagate relatively well the “happy”, “angry”, and “sad” labels. However the “mysterious” label does not give good results. We can explain this by the fact that it might be the most ambiguous concept of these categories. Table 4.6 presents the average P@1 and wP@1 values per mood.

It confirms what we have in the confusion matrix, the “happy” category gives the best result. However looking at the values of wP@1, we note that if “happy” is the most guessed mood, the system gives more reliability to its results about the label “angry”.

In our last experiment we wanted to evaluate how well the mood labels can be propagated if we annotate just partially the collection. We computed the P@1 for 70%, 50% and 30% of the database and we obtained the results written in Table 7. It shows that for 30% of the collection annotated, the system can propagate correctly the tags up to 65% of the collection.

As the content–based approach may not consider important aspects that can infer the mood, all these performances should be improved by using dedicated descriptors or meta–data, like information about the title, the style or the lyrics.

Initial annotation	70%	50%	30%
P@1	0.60	0.44	0.5
Correctly annotated after prop.	88%	72%	65%

Table 4.7: Evaluation of the mood label propagation with the initially percentage of annotated songs.

From the results presented in the previous tables, we can infer on one hand that taking 10 similar songs performs better for those cases where no filtering is present. On the other hand, taking 20 similar songs seems that performs better for the cases where there are restrictions, such as filtering results by same artist or by same album.

4.2 Approach II: Semi-Automatic Annotation

For the second experiment, we used the SearchSounds music collection, collected some manual annotations to form the Ground Truth and then propagated labels among not yet annotated songs. The collection consists of 257,738 songs and the manual annotations covered $\sim$48% of the collection.

Procedure

The procedure for the label propagation in this approach is very similar to the one described in section 4.1.1. The only difference is that, rather than taking all the not yet annotated songs that have proposed labels[3] without caring about the density of the similarity space where these songs appears, take into account this density and propagate labels for the songs appearing in the more dense spaces first. Density means, from a list of similar songs[4], how many of them are already annotated. The first approach did not care about that.

In this second approach, songs in a more dense space are propagated first. It turned to be an iterative process. More concretely, seven iterations were followed. The first two iterations only propagated labels among songs in a $\geq$60% dense space. The following three in a $\geq$50% dense space, and the last two in a $\geq$40% dense space. For each new iteration, the songs with propagated labels in the previous iteration are included to the

[3] Remember the restriction that a label should appear at least in 20 % of the similar songs to be considered a proposed label

[4] The number of similar songs was restricted to 10, because it gave better results in the first approach

set of already annotated songs.

In total, the SearchSounds collection consists of 257,738 songs. A set of collected manual annotations covered 123,681 songs, representing the ∼48% of the collection (subsection 3.2.1). The annotation propagation in each iteration is presented in table 4.8. The total number of songs with propagated labels after the seven iterations was 73,507, representing ∼28% of the collection. Thus, before starting the relevance feedback, the collection was already 48% + 28% = 76% annotated.

Iteration	Density	Number of songs with propagated labels
1	≥60%	21396
2	≥60%	11904
3	≥50%	12628
4	≥50%	8528
5	≥50%	5255
6	≥40%	7627
7	≥40%	5719

Table 4.8: Iterations of the annotation propagation process.

In another experiment we filtered the similar songs to a given one by same artist, for the same reasons that we expose in section 4.1. As we can see in table 4.9, the number of songs with propagated labels is reduced considerably to a total of 31,993 songs, representing ∼13%, that is, less than the half of the propagated annotations in the experiment without filters.

With respect to relevance feedback, we made a pilot study, preparing a closed collection of 94 songs that had propagated annotations. Then, we asked 10 people from our research group to confirm or reject the propagated annotations.

Evaluation Metrics

For the evaluation of this approach we computed some metrics for the evaluation of the propagated labels, without the relevance feedback from users yet. We computed the distribution of the labels in both the manual annotations (Ground Truth) and the propagated annotations. We also computed the distribution of the labels' categories (recall from section 3.2.3 that labels are classified in 6 categories).

Another important metric was the NN-precision of the propagated labels. This metric is defined as:

Iteration	Density	Number of songs with propagated labels
1	≥60%	10112
2	≥60%	4978
3	≥50%	4972
4	≥50%	3643
5	≥50%	2681
6	≥40%	2974
7	≥40%	2633

Table 4.9: Iterations of the annotation propagation process. In this case, the similar songs are filtered by same artist.

$$label_i = \frac{\sum_j overlap_{i,j}}{N} \tag{4.6}$$

Where N is the number of songs that are annotated with the label i, and $overlap_{i,j}$ represents the overlap of label i in the list of the most similar songs to song j. In other words, the frequency of label i in the list of similar songs to a given song j. With this metric we wanted to know the precision of the propagation of a label using our similarity distance function.

For the relevance feedback experiment, we computed the score of each annotation (recall from subsection 3.2.4 that each song annotation is given a score) after the feedback and observed which labels were more confirmed or more rejected as specific measures, and the score mean as a general measure.

Results

Regarding the distribution of labels in the manual annotations, there were 310 different labels used. The top-five labels were 'rock: modern rock' (6,6% of the collection), 'solo male artist' (6,17%), 'featuring guitar' (5,1%), 'instrumental' (4,8%) and 'acoustic' (4,4%). The less used labels were 'Fusion', 'Pranks', 'Primus', 'Showtunes' and 'Space Filler', representing the 0,001% of the collection each one. The distribution of labels' categories is shown in figure 4.1.

With respect to the distribution of labels in the propagated annotations, 224 different labels were used. The top-five labels were: 'rock: modern rock' (14,5%), 'solo male artist' (9,8%), 'Rock' (9%), 'featuring guitar' (6,6%) and 'acoustic' (5,8%). The less used labels

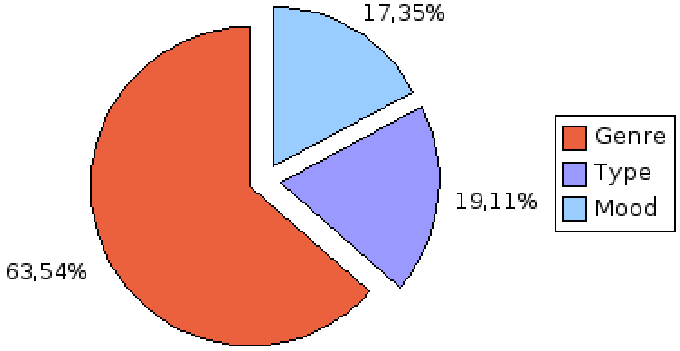


Figure 4.1: Distribution of labels' categories in the manual annotations.

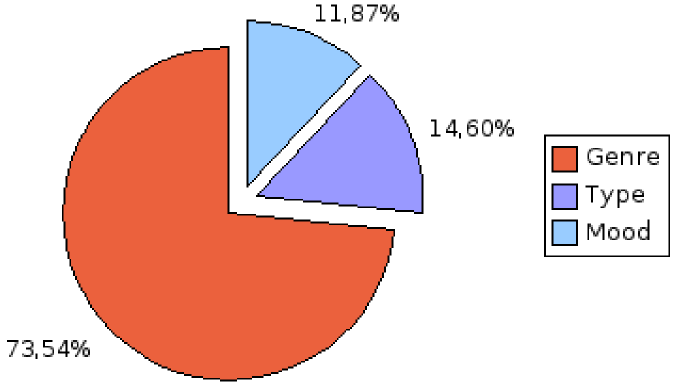


Figure 4.2: Distribution of labels' categories in the propagated annotations.

were 'rock: funk rock', 'rock: jam-band', 'rock: progressive rock', 'world: african' and 'world: reggae', each one representing the 0,001% of the collection. The distribution of labels' categories is shown in figure 4.2.

We can infer from these results that the most used labels in the manual annotations have more possibilities to be propagated — four of the top-five labels in the manual annotations. The less used labels in the manual annotations were not propagated at all, which seems very logical, since there is a restriction for the labels to be propagated.

We calculated the NN-precision for every label in the propagated annotations. In figure 4.3 we can see the result. The x-label represents the number of songs that had a given label propagated, and the y-label represents the precision (which goes from 0 to 1). The labels that had the best NN-precision were 'electronic' and 'speech', of ∼68% and ∼65%, respectively.

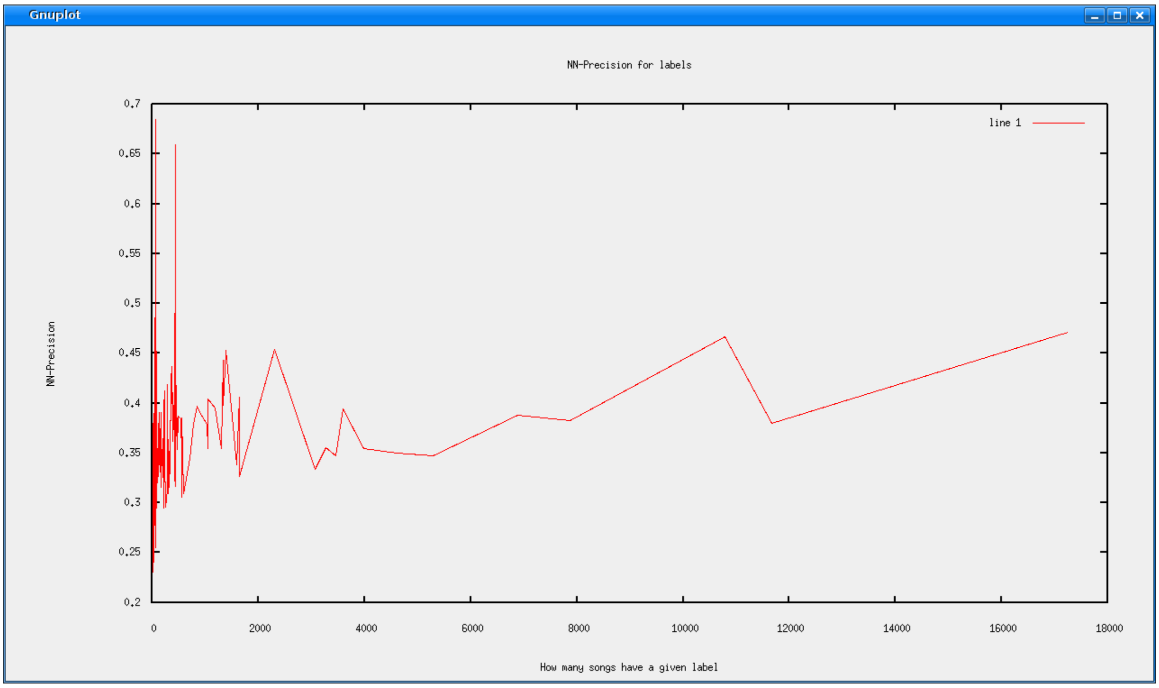


Figure 4.3: NN-precision for labels.

With respect to relevance feedback, figure 4.4 shows an histogram of the average final score for each tag after the relevance feedback process. Initially all propagated annotations had a score of 10. After feedback, the most confirmed tags were 'electronic' and 'world: world beat', with a score of 13, and the most rejected tags were 'Ambient', 'New Age' and 'Piano', with a score of 4.5, 4.67 and 5, respectively. From a total list of 209 annotations (with feedback), 103 of these annotations were confirmed, 21 were not given a feedback, and 85 were rejected. The mean average score after the feedback process was 10.01, that is, the same score given initially for a propagated annotation.

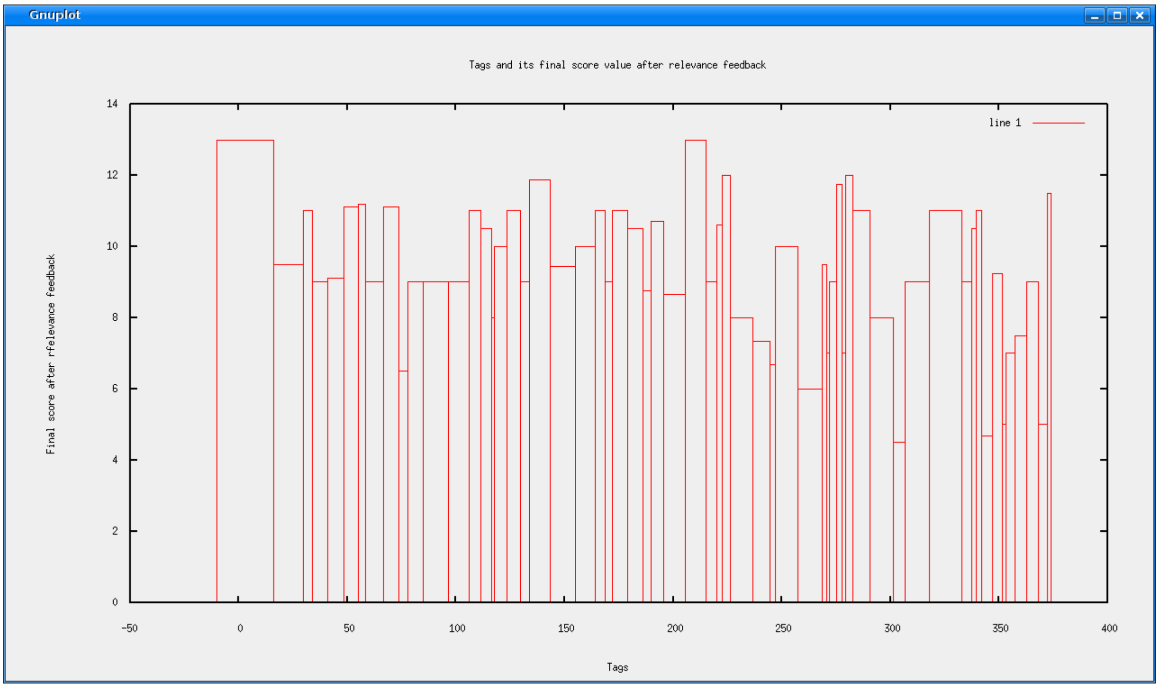

Gnuplot
Tags and its final score value after relevance feedback
Final score after rfelevance feedback
Tags
line 1
0
2
4
6
8
10
12
14
-50
0
50
100
150
200
250
300
350
400

Figure 4.4: Tags and their final score after relevance feedback.

Chapter 5

Conclusions and Future Work

The aim of this thesis is to propagate annotations in music collections by means of an audio–based similarity distance, and receive feedback from users about how good the propagated annotations are.

In chapter 4 we described two experiments followed for the evaluation of the labels' propagation. With respect to the first experiment, the objective was to test how the content–based similarity can propagate labels. Using a collection of of ∼5500 songs borrowed from Magnatune[1] collection, we showed that with a collection annotated at 40% with styles, we can reach a 78% (40%+38%) annotated collection, with a recall greater than or equal to 0.4, only using content–based similarity. In the case of moods, with a 30% annotated collection we can automatically propagate up to 65% (30%+35%).

Regarding the second experiment, we used a collection of ∼258000 songs from SearchSounds and Magnatune. With a 48% manually annotated collection we propagated the annotations up to 76% (48%+28%) and then evaluated a small set of the propagated annotations by means of user relevance feedback.

5.1 Summary of Contributions

Manual annotations of multimedia data is an arduous task, and very time consuming. Automatic annotation methods, normally fine-tuned to reduced domains such as musical instruments or limited to sound effects taxonomies, are not mature enough to label with great detail any possible sound. Yet, in the music domain the annotation becomes more

[1] http://www.magnatune.com/

complex due to the time domain frame.

The purpose of this Thesis has been to propose a semi-automatic strategy that allows to annotate huge music collections, based on audio similarity and a community of users that annotate music titles. This strategy would win in efficiency regarding the manual annotation and accuracy regarding the automatic annotation.

The contributions of this Thesis can then be summarized as follows:

- Automatic propagation of labels among yet not annotated songs, by using a content–based similarity distance as a way to propagate labels.
- Relevance feedback from users, in order to confirm or reject these labels.

This "semi–automatic" approach eases significantly the annotation process of large music collections. For every new song that enters the collection, a list of similar songs is computed and new labels are propagated to it. The only required (manual) effort for users is to confirm or reject tags — using the confirmation or rejection button links in the user interface, respectively. Annotating a song with a new label is optional — although very useful for posterior analysis of the results.

5.2 Discussion

In the introduction part of chapter 5 the two experiments of the Thesis were reviewed. For the first experiment with the Magnatune collection we found some limitations in the mood experiment (see subsection 4.1.2), the "mysterious" label, which did not give good results. The concept has to be clearly encoded in the music for the content–based propagation to work.

With respect to user relevance feedback process, we saw in subsection 3.2.4 that when a user is presented with a list of the most similar music pieces to a given query, each music piece with a tag cloud, he/she can optionally annotate a music piece with a new label. The problem is that although there is a restriction on using labels' categories, the label itself can be whatever natural language word. The reason for not restricting the use of labels is due to folksonomy–like nature of this Thesis' approach. Nevertheless, a special case that should be taken into account is when users from different countries annotate songs with words from their own language. In future work (subsection 5.3) we plan to integrate an auto–completion feature that would suggest labels to the users while they are typing, thus reducing the variation in vocabulary or languages used for annotating.

5.3 Future Work

There are still open issues that could not be studied or developed in this Thesis, and are clear work for the immediate future. Some of these open issues are:

- Generate music playlists automatically based on the labels. That is, the system could generate a playlist of songs with 'electronic new age happy' annotations, for example.
- Recommend music to users based on labels. In this case, user profiles based on the labels they use should be created.
- Rather than just creating tag clouds for each song, extend this idea to every artist.
- Take profit of labels in order to update the system for searching, using a Query-by-description model based on labels.
- Create a user profile based on his/her queries, tags used, and annotated songs, and more generally
- Study the user behavior and learn the relationship between acoustic data, semantic annotations (labels) and human conception of music semantic descriptions.

5.4 Closing Statement

We presented a semi–automatic strategy for annotating large music collections that proposes labels to yet not annotated songs by using content–based similarity as a way to propagated labels, and refine the annotations by means of relevance feedback.

We expounded that automatic annotations are not completely accurate; due to that, this Thesis proposes a relevance feedback option, with which annotations can be confirmed or rejected, with the aim of refining these annotations.

It can be inferred from the pilot study of user feedback that although the results are not perfect, they are very promising. This Thesis is intended to be a first state-of-the-art of semi–automatic annotation of music, and the point of departure for the author's PhD Thesis.

List of Figures

Bibliography

[Ames 07] M. Ames & M. Naaman. *Why we tag: motivations for annotation in mobile and online media.* Proceedings of the SIGCHI conference on Human factors in computing systems, pages 971–980, 2007.

[Baeza-Yates 00] R. Baeza-Yates & B. Ribeiro-Neto. Modern information retrieval. ACM Press [ua], 2000.

[Barrington 07] L. Barrington, A. Chan, D. Turnbull & G. Lanckriet. *Audio Information Retrieval Using Semantic Simlarity.* International Conference on Acoustic, Speech and Signal Processing (ICASSP), 2007.

[Beckett 06] D. Beckett. *Semantics Through the Tag.* In Proceedings of the XTech Conference, Amsterdam, The Netherlands, 2006.

[Berenzweig 04] A. Berenzweig, B. Logan, D.P.W. Ellis & B.P.W. Whitman. *A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures.* Computer Music Journal, vol. 28, no. 2, pages 63–76, 2004.

[Cano 04a] P. Cano & M. Koppenberger. *Automatic sound annotation.* In Proceedings of 14th IEEE workshop on Machine Learning for Signal Processing, SÃ£o Luís, Brazil, 2004.

[Cano 04b] P. Cano, M. Koppenberger, S. Ferradans, A. Martinez, F. Gouyon, V. Sandvold, V. Tarasov & N. Wack. *MTG-DB: A Repository for Music Audio Processing.* In Proceedings of 4th International Conference on Web Delivering of Music, Barcelona, Spain, 2004.

[Cano 04c] P. Cano, M. Koppenberger, P. Herrera, O. Celma & V. Tarasov. *Sound Effect Taxonomy Management in Production Environments.* In Proceedings of 25th International AES Conference, London, UK, 2004.

[Cano 05a] P. Cano, M. Koppenberger, S. Le Groux, J. Ricard, N. Wack & P. Herrera. *Nearest-Neighbor Automatic Sound Classification with a WordNet Taxonomy.* Journal of Intelligent Information Systems, vol. 24, no. 2, pages 99–111, 2005.

[Cano 05b] P. Cano, M. Koppenberger, N. Wack, J. G. Mahedero, J. Masip, O. Celma, D. Garcia, E. Gómez, F. Gouyon, E. Guaus, P. Herrera, J. Massaguer, B. Ong, M. Ramírez, S. Streich & X. Serra. *An Industrial-Strength Content-based Music Recommendation System.* In Proceedings of 28th Annual International ACM SIGIR Conference, Salvador, Brazil, 2005.

[Carneiro 05] G. Carneiro & N. Vasconcelos. *Formulating Semantic Image Annotation as a Supervised Learning Problem.* Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, 2005.

[Celma 06a] O. Celma. *Music Recommendation: a multi-faceted approach.* PhD thesis, 2006.

[Celma 06b] O. Celma, P. Cano & P. Herrera. *Search Sounds: An audio crawler focused on weblogs.* In Proceedings of 7th Intl. Conference on Music Information Retrieval, Victoria, Canada, 2006.

[Ellis 06] Daniel P.W. Ellis. *Extracting information from music audio.* Commun. ACM, vol. 49, no. 8, pages 32–37, 2006.

[Feng 03] Y. Feng, Y. Zhuang & Y. Pan. *Music information retrieval by detecting mood via computational media aesthetics.* Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on, pages 235–241, 2003.

[Gómez 06] E. Gómez. *Tonal Description of Music Audio Signals.* PhD thesis, 2006.

[Gouyon 05] F. Gouyon. *A computational approach to rhythm description — Audio features for the computation of rhythm periodicity functions and their use in tempo induction and music content processing.* PhD thesis, 2005.

[Handschuh 02] S. Handschuh, S. Staab & F. Ciravegna. *S-CREAM-Semi-automatic CREAtion of Metadata.* Proceedings of EKAW 2002, pages 358–372, 2002.

[Herrera 05a] P. Herrera, J. Bello, G. Widmer, M. Sandler, O. Celma, F. Vignoli, E. Pampalk, P. Cano, S. Pauws & X. Serra. *SIMAC: Semantic interaction with music audio contents.* In Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, Savoy Place, London, UK, 2005.

[Herrera 05b] P. Herrera, O. Celma, J. Massaguer, P. Cano, E. Gómez, F. Gouyon, M. Koppenberger, D. Garcia, J. G. Mahedero & N. Wack. *Mucosa: a music content semantic annotator.* In Proceedings of 6th International Conference on Music Information Retrieval, London, UK, 2005.

[Jeon 03] J. Jeon, V. Lavrenko & R. Manmatha. *Automatic image annotation and retrieval using cross-media relevance models.* Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pages 119–126, 2003.

[Juslin 01] P.N. Juslin & J.A. Sloboda. Music and emotion: theory and research. Oxford University Press, 2001.

[Knees 06] Peter Knees, Tim Pohle, Markus Schedl & Gerhard Widmer. *Combining Audio-based Similarity with Web-based Data to Accelerate Automatic Music Playlist Generation.* In Proceedings of the 8th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'06), Santa Barbara, California, USA, October 26-27 2006.

[Knees 07a] Peter Knees. *Search & Select - Intuitively Retrieving Music from Large Collections.* In Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07), Vienna, Austria, September 2007.

[Knees 07b] Peter Knees, Tim Pohle, Markus Schedl & Gerhard Widmer. *A Music Search Engine Built upon Audio-based and Web-based Similarity Measures.* In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07), Amsterdam, the Netherlands, July 23-27 2007.

[Knees 07c] Peter Knees & Gerhard Widmer. *Searching for Music Using Natural Language Queries and Relevance Feedback.* In Proceedings of the 5th International Workshop on Adaptive Multimedia Retrieval (AMR'07), Paris, France, July 2007.

[Lu 00] Y. Lu, C. Hu, X. Zhu, H.J. Zhang & Q. Yang. *A unified framework for semantics and feature based relevance feedback in image retrieval systems.* Proceedings of the eighth ACM international conference on Multimedia, pages 31–37, 2000.

[Mathes 04] A. Mathes. *Folksonomies-Cooperative Classification and Communication Through Shared Metadata.* Computer Mediated Communication, LIS590CMC (Doctoral Seminar), Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, December, 2004.

[Pachet 05] F. Pachet. *Knowledge Management and Musical Metadata.* Encyclopedia of Knowledge Management, 2005.

[Rocchio 71] J.J. Rocchio. *Relevance feedback in information retrieval.* The SMART Retrieval System: Experiments in Automatic Document Processing, pages 313–323, 1971.

[Sandvold 06] V. Sandvold, T. Aussenac, O. Celma & P. Herrera. *Good Vibrations: Music Discovery through Personal Musical Concepts.* In Proceedings of 7th Intl. Conference on Music Information Retrieval, Victoria, Canada, 2006.

[Song 05] Yan Song, Xian-Sheng Hua, Li-Rong Dai & Meng Wang. *Semi-automatic video annotation based on active learning with multiple complementary predictors.* In MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, pages 97–104, New York, NY, USA, 2005. ACM Press.

[Sordo 07] M. Sordo, C. Laurier & O. Celma. *Annotating Music Collections : How content-based similarity helps to propagate labels.* Proceedings of 8th International Conference on Music Information Retrieval, 2007.

[Staab 05] S. Staab, P. Domingos, P. Mika, J. Golbeck, L. Ding, T. Finin, A. Joshi, A. Nowak & R.R. Vallacher. *Social Networks Applied.* IEEE Intelligent Systems, vol. 20, no. 1, pages 80–93, 2005.

[Turnbull 07a] D. Turnbull, L. Barrington, D. Torres & G. Lanckriet. *Exploring the Semantic Annotation and Retrieval of Sound.* CAL Technical Report CAL-2007-01, 2007.

[Turnbull 07b] D. Turnbull, L. Barrington, D. Torres & G. Lanckriet. *Towards musical query-by-semantic-description using the CAL500 data set.* Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pages 439–446, 2007.

[Wenyin 01] L. Wenyin, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski & B. Field. *Semi-automatic image annotation.* Proc. of Interact 2001: Conference on Human-Computer Interaction, pages 326–333, 2001.

[Whitman 02a] B. Whitman & S. Lawrence. *Inferring descriptions and similarity for music from community metadata.* Proceedings of the 2002 International Computer Music Conference, pages 591–598, 2002.

[Whitman 02b] B. Whitman & R. Rifkin. *Musical query-by-description as a multiclass learning problem.* Multimedia Signal Processing, 2002 IEEE Workshop on, pages 153–156, 2002.

[Whitman 02c] B. Whitman & P. Smaragdis. *Combining musical and cultural features for intelligent style detection.* Proc. Int. Symposium on Music Inform. Retriev.(ISMIR), pages 47–52, 2002.

[Whitman 03] B. Whitman, D. Roy & B. Vercoe. *Learning Word Meanings and Descriptive Parameter Spaces from Music.* Proceedings of the HLT-NAACL 2003 workshop on Learning word meaning from non-linguistic data-Volume 6, pages 92–99, 2003.

[Whitman 04] B. Whitman & D. Ellis. *Automatic record reviews.* Proceedings of the 2004 International Symposium on Music Information Retrieval, 2004.

[Whitman 05] B.A. Whitman. *Learning the meaning of music.* PhD thesis, 2005.