Universitat
Pompeu Fabra
Barcelona

Knowledge-based and Multimodal Deep Learning Approaches for Music Recommendation and Classification

Sergio Oramas Martín

TESI DOCTORAL UPF / 2015

Director de la tesi:

Dr. Xavier Serra Casals
Dept. of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain

Dissertation submitted to the Deptartment of Information and Communication Technologies of Universitat Pompeu Fabra in partial fulfillment of the requirements for the degree of

DOCTOR PER LA UNIVERSITAT POMPEU FABRA

Music Technology Group (http://mtg.upf.edu), Dept. of Information and Communication Technologies (http://www.upf.edu/dtic), Universitat Pompeu Fabra (http://www.upf.edu), Barcelona, Spain.

Acknowledgements

Great time.

This thesis has been carried out at the Music Technology Group of Universitat Pompeu Fabra (UPF) in Barcelona, Spain, from October 2013 to September 2017, at the Politecnico di Bari, Italy, from September 2014 to October 2014, at the Insight Centre for Data Analytics of University College of Dublin (UCD), Ireland, from May 2015 to July 2015, and at Pandora Media Inc., USA, from June 2016 to September2016. This work has been supported by Obra Social "La Caixa" under their graduate fellowship program, and by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502). The research stay at UCD has been also funded by the Keystone COST Action IC1302 under grant number SFI/12/RC/2289.

Abstract

Music content creation, publication and dissemination has changed dramatically in the last few decades. On the one hand, huge amounts of information about music are being daily published in online repositories such as web pages, forums, wikis, and social media. However, most of this content is still unusable by machines due to the fact that it is mostly created by humans and for humans. On the other hand, online music services currently offer ever-growing collections with dozens of millions of music tracks. This vast availability has posed two serious challenges. First, how can a musical item be properly annotated and classified within a large collection? Second, how can a user explore or discover preferred music from all the available content? In this thesis, we address these two questions focusing on the enrichment of item descriptions, and the exploitation of the heterogeneous data in large music collections. We first leverage semantic information present in online repositories. To this end, we focus on the problem of linking musical texts and metadata descriptions with online knowledge repositories via entity linking, and on the automatic construction of music knowledge bases via relation extraction. Then, we investigate how extracted knowledge may impact recommender systems, classification approaches, and musicological studies. We show how modeling semantic information contributes to outperform text-based approaches in music similarity and classification, and to achieve significant improvements with respect to state of the art collaborative algorithms in music recommendation, while promoting long tail recommendations. Then, we research on learning new data representations from multimodal content using deep learning architectures. Following this approach, we address the problem of cold-start music recommendations by combining audio and text. We show how the semantic enrichment of texts and the late fusion of feature embeddings learned separately improve the quality of recommendations. Moreover, we tackle the problem of single and multi-label music genre classification from different data modalities (i.e. audio, text and images). Experiments show major differences between modalities, which not only introduce new baselines for multi-label genre classification, but also suggest that combining them yields superior results. As an outcome of this thesis, we have collected and released eight different datasets and two software libraries. Our findings have significant implications for music recommendation and classification systems. Although our research is motivated by particularities of the music domain, we believe the proposed approaches can be easily generalized to other domains.

CHAPTER 1
Introduction

1.1. Motivation

Today, we are witnessing an unprecedented information explosion thanks to the dramatic technological advancement brought by the Information Age. This technological (r)evolution has set the foundations for the release and publication of huge amounts of data onto online repositories such as web pages, forums, wikis and social media. Art and culture have benefited dramatically from this context, which allows potentially anyone with an available Internet connection to access, produce, publish, comment or interact with any form of media.

In this context, music content creation, publication and dissemination has changed dramatically. Online music services, such as Pandora, Spotify or Apple Music, benefit from this situation and currently offer ever-growing catalogs with dozens of millions of music tracks, which are in turn just one click away from hundreds of millions of users. This vast availability of music has posed two serious challenges: (1) how can a musical item be properly annotated and classified within a large collection? Since manually managing these large libraries is not feasible due to size constraints, automatic methods for the annotation and classification of large-scale music collections have been an active area of research in recent years Schedl et al. (2014). (2) how can a user explore or discover preferred music from all the available content? Traditionally, users have relied on their friends, their favourite music radio host, a music expert in their local retail store, etc. to obtain recommendations on artists or albums they might like. Although this traditional approach is still valid and used by many people, its ability to cover the vast amount of available music nowadays is seriously hindered. Therefore, automatic approaches to music recommendation have become necessary Celma & Herrera (2008).

Large music collections combine information from multiple data modalities, such as audio, images, text or videos. In addition, music collections can be enriched with user generated content published online on a daily basis. However, most of this content is still unusable by machines due to the fact that it is mostly created by humans and for humans, and hence it only exists in human readable form. In this context, Natural Language Processing plays a key role, as one of its main lines of research is precisely to transform unstructured information in machine readable data Cowie & Lehnert (1996), allowing discovery of new facts and trends hidden in, for example, Music libraries, blogs, web pages, journals or social networks.

The way multimodal data from large music collections is represented and combined in computational models poses numerous challenges. Artificial intelligence methods, such as machine learning, heavily rely on the choice of data representation. Therefore, finding representations that maximize the different explanatory factors of variation behind the data is a fundamental task. Traditional approaches rely on handcrafted features to represent the variability of the data, whereas more recently, and thanks the raise of deep learning techniques, representation-learning approaches have demonstrated their superiority in multiple domains Bengio et al. (2013).

In this thesis, we focus on the problem of how to enrich and exploit multimodal data present in large music collections from two different stand points. (1) Leveraging semantic information present in online knowledge repositories and unstructured text sources. (2) Learning new data representations from heterogeneous data using deep learning architectures and further combining these representations in multimodal networks. Both ideas are in turn applied to the aforementioned problems of classification and recommendation of musical items.

1.2. Processing Language in Music Information Research

Music Information Retrieval/Research (MIR) is a multidisciplinary field of research concerned with the extraction, analysis, and usage of information about any kind of music entity (e.g. song, artist, album) on any representation level (e.g. audio signal, symbolic MIDI, metadata) Schedl (2008). As stated in Schedl et al. (2013), factors that influence human music perception can be categorized into music content, music context, user context and user properties. Music context relates to all musical aspects that are not encoded in the audio signal, such as song lyrics, artist's biography, album cover artwork or music video clips, whereas music content is defined as human perceptual aspects that can be extracted from the audio signal. Following this distinction, research methodologies within the MIR community that deals with data modalities different from audio are often called context-based approaches. Although we agree with this classification criteria, in concordance with the nomenclature used within the Recommender Systems community Ostuni et al. (2013), in this dissertation either audio signal, text (e.g. metadata, artist's biographies, song lyrics), images (e.g. album cover artwork, artist's photographies), and video (e.g. music video clips) are simply considered as different modalities of content information.

According to Humphrey et al. (2012), MIR approaches are typically based on a two-stage architecture of feature extraction and semantic interpretation, e.g., classification, regression, clustering, similarity ranking, etc. Traditionally, MIR has been mainly focused on the use of features extracted from audio, underestimating other data modalities. However, in recent years several studies have shown the benefits of using *context-based* and multimodal approaches Schedl et al. (2014).

Audio features are often classified into low, mid and high-level representations **?**. Low-level representations (e.g. spectral flux, cepstrum, MFCCs) are measured directly on the audio signal. Mid-level representations (e.g. chords, onsets) represent musical attributes extracted from the audio combining machine learning and musical knowledge. High-level representations (e.g. mood, form, genre) are related to human interpretations of the data, and are typically built on top of low and mid-level representations. The extraction and exploitation of features from these three representation levels have been widely studied.

Following this feature hierarchy, when dealing with textual data, we can also differentiate between low, mid and high-level representations (see Figure **??**). Low-level representations (e.g. word frequencies, word co-occurrences, n-grams) are measured directly on text. Mid-level representations (e.g. part-of-speech tags, named entities) combine linguistic knowledge and statistical analysis of text corpora. High-level representations (e.g. syntactic dependencies, semantic relations) involves a semantic understanding of text. In the context of MIR, most of the literature is focused on low-level representations, few in mid-level, and almost none in high-level. Little attention has been paid in the semantic of words, nor in the context they are being used. Thus, the epistemic potential of text has not been exploited yet.

In the first and second parts of this thesis, we focus on text-based approaches from two standpoints. On the one hand, we work on new methodologies for the extraction of high level semantic representations from unstructured texts. On the other hand, we put the emphasis on the development of approaches that exploits these semantic representations in MIR tasks, such as music recommendation and classification. In addition, we study how semantic information may impact musicological studies.

1.3. Representation Learning in Music Information Research

As stated before, MIR approaches are commonly based on a two-stage architecture of feature extraction and semantic interpretation. In this context, data representations are generally obtained following a traditional feature extraction process, which involves a combination of music domain-knowledge, psychoacoustics, and audio engineering Humphrey et al. (2012). Feature engineering compensate the inability of traditional machine learning algorithms to extract the discriminative information of the data. However, it involves a labor-intensive human effort, and also all the different explanatory factors of variation behind the data are not represented Bengio et al. (2013). Huge efforts have been put in the last two decades in the definition and extraction of audio features, which has given rise to comprehensive software libraries that assemble many of these feature extraction techniques Bogdanov et al. (2013a); Mcfee et al. (2015).

Representation learning (or feature learning) is a technique that allows a learning system to automatically discover the variation behind the data directly from raw data. As identified in Humphrey et al. (2012), MIR approaches can benefit from the use of these learning approaches using deep neural networks. This methodology has two main advantages. First, blurring the boundaries between the two-stage architecture, which implies fully-automated optimization of both stages at once. Second, it results in general-purpose architectures that can be applied to different MIR problems and data modalities. In the last years, several works have been published where end-to-end learning approaches using deep learning architectures have been applied to MIR tasks such as music recommendation van den Oord et al. (2013) and music classification Choi et al. (2016a), among others.

In the third part of this dissertation, we focus on representation learning approaches using deep neural networks. We apply this methodology to different data modalities (audio, text and images) and their combination, and in the context of music recommendation and classification tasks.

1.4. Annotation and Classification of Music Collections

The advent of large music collections has posed the challenge of how to access the information - in terms of retrieval, browsing, and recommendation -. One way to ease the access of large music collections is to keep annotations of all music resources Sordo et al. (2012). Annotations can be added either manually or automatically. Manual annotation of huge music collections is too costly due to high human effort required. Therefore, the implementation of automatic annotations processes has become mandatory.

We distinguish two ways of automatically enhance annotations: (i) gathering annotations from external sources, and (ii) learning annotations from the collection's data. To address (i), information can be obtained from online knowledge repositories (e.g. Wikipedia, MusicBrainz), or extracted from collections of unstructured documents. This imposes the challenge of how to properly map collection's items with external entities. To address (ii), machine learning techniques can be applied over the collection's data. When annotations are learned from audio this task is often called auto-tagging. However, annotations can be learned from different data modalities, such as album cover artworks, tags, editorial metadata, video clips, etc.

Among the different categories of annotations used in music collections, the most prototypical are: music genres, instruments, and moods. Music genre labels are useful categories to organize and classify songs, albums and artists into broader groups that share similar musical characteristics. Music genres have been widely used for music classification, from physical music stores to streaming services. Automatic music genre classification thus is a widely explored topic (Sturm, 2012). However, almost all related work is concentrated in the classification of music items into broad genres (e.g., Pop, Rock), assigning a single label per item. This is problematic since there may be hundreds of more specific music genres Pachet & Cazaly (2000), and these may not be necessarily mutually exclusive (i.e., a song could be Pop, and at the same time have elements from Deep House and a Reggae grove).

In this thesis, we focus on the problem of enriching annotations in music collections from the two above defined standpoints, i.e. gathering and learning. We study how semantic technologies may be useful to improve the annotations of musical items. In addition, we tackle the problem of single-label and multi-label music genre classification from different data modalities (i.e. audio, text and images) and their combination.

1.5. Music Recommendation

Information overload in modern Web applications challenges users in their decision-making tasks. Recommender systems have emerged in the last years as fundamental tools in assisting users to find, in a personalized manner, what is relevant for them in overflowing knowledge spaces.

Music Recommendation is a relatively young but continuously growing research topic, in both MIR and Recommender Systems communities Celma (2010a). Several research approaches and commercial systems have been proposed in the last decade. However, many of them are adaptations from other domains Celma (2010a). Music has its own specificities with respect to other domains. For instance, a user may consume a musical item several times, or very different items according to the user context (e.g. working, dinning, excercising). Therefore, music recommendation is a challenging and still unsolved problem.

Although music online services make available almost all existing music, only a small percentage of these catalogues is actually consumed by the vast majority of users. Music consumption follows what is called a long tail distribution Celma (2010a) (see Figure ??). Therefore, one of the main challenges in music recommendation is how to make this long tail of musical items profitable. Moreover, as music creation is continuously growing, new artists and releases appear every day. Hence, another important challenge in recommender systems is how to deal with these new items, which is often called the cold-start problem.

The web is full of documents, knowledge repositories and user generated content with relevant information about music and musicians. This information may have the potential to impact in the performance of music recommender systems. In addition, up to now, audio content has been barely exploited in comercial recommender systems. However, thanks to the advent of novel deep learning approaches van den Oord et al. (2013), audio content is becoming a key factor in order to provide accurate long tail recommendations.

Most research in Music Recommendation has been dedicated to developing algorithms that provide *good* and *useful* recommendations Celma (2010a), neglecting the importance of the novelty and diversity of recommendations Adomavicius & Kwon (2012); Bellogín et al. (2010). In addition, very few approaches are able to provide explanations of the recommendations to the users Passant & Raimond (2008); Passant (2010). According to Celma & Herrera (2008), giving explanations of the recommendations provides transparency to the recommendation process and increases the confidence of the user in the system.

In this thesis we dig into the Music Recommendation problem from three different perspectives. First, we investigate how information extracted from large collections of documents *talking* about music may be useful to provide explanations of recommendations to users. Second, we tackle the problem of recommending long tail items by leveraging semantic information from knowledge repositories and combining that with users feedback data. Finally, we address the problem of cold-start music recommendations by combining different data modalities using deep neural networks.

1.6. Objectives and outline of the thesis

In the previous sections we have explained the motivations and context of our thesis. According to that, the main goal of this dissertation is to contribute to advancing the state of the art in music recommendation and classification by leveraging knowledge repositories and unstructured text sources, and learning novel data representations from multimodal data. Although this thesis is focused on the music domain, the work we present can be easily adapted to other multimedia domains. Figure**??** shows a conceptual organisation of some of the chapters of this thesis according to the different approaches, tasks, and learning process.

This thesis is structured as follows: Chapter 2 presents some background knowledge and related work on Knowledge Extraction and MIR. Hereafter, the work in this dissertation is divided in three Parts: In Part **??** we explore different techniques and approaches to extract semantic information from unstructured text sources *talking* about music. Within this Part, Chapter **??** illustrates the problem of linking musical texts and knowledge repositories. In Chapter **??** we address the automatic generation of Music Knowledge Bases from unstructured text sources. This Chapter encloses with an experiment on explanations of music recommendations based on an extracted Knowledge Base. In Chapter **??**, three experiments study the potential impact of knowledge extraction techniques in musicological studies. Then, in Part **??**, the semantic representations described in Part **??** are exploited in Music Classification, Similarity and Recommendation problems. Chapter **??** presents the application of a semantic-based approach to music similarity and classification problems, whereas Chapter **??** address the problem of long tail recommendations by enriching annotations with semantic information. Then, in Part **??** an approach to learn data representations from different data modalities using deep neural networks is applied to the Music Recommendation and Classification problems. In Chapter **??** we address the problem of cold-start music recommendations using audio and text. Finally, in Chapter **??** we apply a similar approach to music genre classification, and two experiments are presented: a single-label classification problem over audio and images, and a multi-label classification problem over audio, text and images. At the end of each chapter, we include a focused discussion about the relevant results and conclusions. We conclude this thesis in Chapter **??** with a summary of our work, our main conclusions, and a discussion about open issues and future perspectives.

CHAPTER 2 Background

2.1. Introduction

The literature review presented in this chapter is divided into two stages. (i) we summarise existing work on several areas of Natural Language Understanding, with special focus on its application to the music domain. We define what a Knowledge Base (KB) is and the different existing types. Moreover, we deepen into the available KBs that contain music information. Then, we explain what Entity Linking is and briefly describe some state-of-the-art systems. Additionally, we outline different existing approaches for Relation Extraction. (ii) we dig into the available literature on Music Information Retrieval (MIR). More specifically, we first review the state-of-the-art of text-based approaches. Then, we focus on three specific tasks: music genre classification, artist similarity, and music recommendation.

2.2. Natural Language Understanding

Natural language understanding (NLU) is a subtopic of Natural Language Processing (NLP) that deals with machine reading comprehension. Knowledge Representation and Reasoning is a key enabler of Intelligent Systems Suchanek et al. (2007), and plays an important role in Natural language Understanding (NLU) Baral & De Giacomo (2015). In this disertation, we focus on an important aspect of NLU, which is *how to make sense* of the data that is generated and published online on a daily basis. This data is mostly produced in human-readable format, which makes it unsuitable for automatic processing. Considering that deep understanding of natural language by machines seems to be very far off Cambria & White (2014), there is great interest in formalizing unstructured data, and Knowledge Bases (KBs) are a paradigmatic example of large-scale content processed to make it machine readable.

Information Extraction (IE) is the task of automatically extracting structured information from unstructured or semi-structured text sources. It is a widely studied technique within the Natural Language Processing (NLP) research community Cowie & Lehnert (1996). A major step towards understanding language is the extraction of meaningful terms (entities) from text as well as relationships between those entities. This statement involves two different tasks. The former is to determine the identity and category of entity mentions present in text. This task is called Named Entity Recognition (NER). However, when this task involves a latter step of disambiguation of entities against a KB it is called Named Entity Disambiguation (NED) or Entity Linking (EL). The second task is to identify and annotate relevant semantic relations between entities in text. This task is called Rellation Extraction.

The work described in this thesis strongly focuses on the exploitation of linguistic and semantic properties of text collections. For this reason, we deem relevant to cover related work in the following areas: (1) KB construction and curation; (2) Music KBs; (3) Entity Linking, and (4) Relation Extraction.

2.2.1. Knowledge Base Construction

We may define a KB as a repository of knowledge organized in a predefined taxonomic or ontologic structure, potentially compatible with other KBs, thus contributing to the Linked Open Data initiative[1]. These KBs may be designed to represent unconstrained knowledge, or a single domain of interest. This representation is formalized either manually, automatically, or with a combination of both.

We understand language by making sense of the connections between words, concepts, phrases and thoughts (Havasi et al., 2007). KBs constitute a resource for encapsulating this knowledge. Previous efforts on KB construction may be characterized as: (1) Handcrafted KBs; (2) Integrative projects (automatic in design, but reliant on manually validated data); and (3) Fully automatic, also in the RE process.

Among the first group, the best known is probably WordNet (Miller, 1995), a lexical database which groups concepts in "synonym sets", and encodes predefined relations among them such as *hyponymy/hypernymy*, *meronymy*, *holonymy*, or *instantiation*. Manually constructed KBs, however, are mostly developed in specific domains, where the degree of ambiguity is lower and there is more availability of trained knowledge engineers.

Next, integrative projects are probably the most productive, as they are the most ambitious attempts in terms of content coverage and community involvement, not only users, but also contributors. Examples of these include Yago (Suchanek et al., 2007), an automatically created KB derived from integrating Wikipedia and WordNet; DBpedia (Lehmann et al., 2014), a collaboratively maintained project aimed at exploiting information present in Wiki-

[1] http://linkeddata.org/

pedia, both structured and in free text; Freebase (Bollacker et al., 2008), also a collaborative effort mainly based on extracting structured knowledge from Wikipedia; or BabelNet (Navigli & Ponzetto, 2012), a semantic network which started as a seamless integration of Wikipedia and WordNet, and today constitutes the largest multilingual repository of words and senses.

With regard to the third group we refer to approaches where knowledge is obtained automatically. Endeavours in this area include TextRunner (Banko et al., 2007a), widely regarded as the first *Open Information Extraction* (OIE) system; ReVerb (Fader et al., 2011), particularly designed to reduce noise while keeping a wide coverage, thanks in part to a set of syntactic and lexical constraints; NELL (Carlson et al., 2010), which incorporates semantic knowledge in the form of a handcrafted taxonomy of entities and relations; PATTY (Nakashole et al., 2012) and WiseNet (Moro & Navigli, 2012, 2013), in which a shared vision to integrate semantics is applied both at the entity and relation level; DefIE (Bovi et al., 2015b), a recent development in OIE tested on the whole set of BabelNet glosses; and KB-Unify (Bovi et al., 2015a), not an actual IE implementation, but rather a unification framework for IE systems.

2.2.2. Music Knowledge Bases

MusicBrainz and Discogs are two paramount examples of manually curated MKBs. They are not strictly KBs, but open music encyclopedias of music metadata, which are built collaboratively and are openly available. MusicBrainz, in addition, has been published as Linked Data by the LinkedBrainz project[2].

As for generic KBs based on Wikipedia, such as the ones described earlier, these include a remarkable amount of music data, such as artist, album and song biographies, definitions of musical concepts and genres, or articles about music institutions and venues. However, their coverage is biased towards the best known artists, and towards products from Western culture. Finally, let us refer to the notable case of Grove Music Online[3], a music encyclopedia containing over 60k articles written by music scholars. However, it has the drawback of not being freely open, as it runs by subscription. Other than the aforementioned curated repositories, to the best of our knowledge, there is not a single automatically learned open MKB. A first step in this direction is taken in this disertation.

Despite their scarcity, MKBs are becoming increasingly popular in MIR applications, such as artist similarity and music recommendation (Celma & Serra, 2008; Leal et al., 2012; Ostuni et al., 2013). MKBs have also been exploited as sources of explanations in music recommender systems. For instance, in

[2] http://linkedbrainz.org/

[3] http://www.oxfordmusiconline.com

(Passant, 2010), explanations of recommendations are created by exploiting DBpedia's structured information.

2.2.3. Entity Linking

The advent of large knowledge repositories and collaborative resources has contributed to the emergence of Entity Linking (EL), i.e. the task of discoveing mentions of entities in text and link them to a suitable knowledge repository (Moro et al., 2014b). It encompasses similar subtasks such as Named Entity Disambiguation (Bunescu & Pasca, 2006), which is precisely linking mentions of entities to a KB, or Wikification (Mihalcea & Csomai, 2007), specifically using Wikipedia as KB. There have been a great development of EL systems for unconstrained domains. Among these systems we focus on three of them in this thesis:

DBpedia Spotlight (Mendes et al., 2011) is a system for automatically annotating text documents with DBpedia URIs, finding and disambiguating natural language mentions of DBpedia resources. DBpedia Spotlight is shared as open source and deployed as a Web service freely available for public use[4].
TagMe (Ferragina & Scaiella, 2012) is an EL system that matches terms with Wikipedia link texts and disambiguates them using the Wikipedia in-link graph. Then, it performs a pruning process by looking at the entity context. TagMe is available as a web service [5].
Babelfy (Moro et al., 2014a) is an EL and Word Sense Disambiguation (WSD) system based on non-strict identification of candidate meanings (i.e. not necessarily exact string matching), together with a graph based algorithm that traverses the BabelNet graph and selects the most appropriate semantic interpretation for each candidate [6].

In the context of Open Data, the need for benchmarking datasets and evaluation frameworks for EL is clear. However, while general purpose datasets and benchmarks exist (Usbeck et al., 2015), dealing with highly specific domains (e.g. chemistry) or ever-evolving areas (e.g. videogames or music) poses a greater challenge due to linguistic idiosincrasies or under-representation in general purpose knowledge-bases. This is true in the music domain as well, where available data is scarce (Gruhl et al., 2009). Among the few works on EL for the music domain, let us refer to (Gruhl et al., 2009), who describe an approach for detecting musical entities in informal text. In addition, (Zhang et al., 2009) describe a system for musical EL in the Chinese language based on Hidden Markov Models.

[4] https://github.com/dbpedia-spotlight/dbpedia-spotlight/
[5] https://tagme.d4science.org/tagme/
[6] http://babelfy.org/

2.2.4. Relation Extraction

A large portion of the knowledge contained in the web is stored in unstructured natural language text. In order to acquire and formalize this heterogeneous knowledge, methods that automatically process this information are in demand. Extracting semantic relations between entities is an important step towards this formalization (Wang, 2008). Relation Extraction is an established task in Natural Language Processing (Bach & Badaskar, 2007). It has been defined as the process of identifying and annotating relevant semantic relations between entities in text (Jiang & Zhai, 2007).

Relation Extraction (RE) approaches are often classified according to the level of supervision involved. Supervised learning is a core-component of a vast number of RE systems, as they offer high precision and recall. However, the need of hand labeled training sets makes these methods not scalable to the thousands of relations found on the Web (Hoffmann et al., 2011). More promising approaches, called semi-supervised approaches, bootstrapping approaches, or distant supervision approaches do not need big hand labeled corpus, and often rely on existent knowledge bases to heuristically label a text corpus (e.g., (Carlson et al., 2010; Hoffmann et al., 2011))

Open Information Extraction methods do not require an annotated corpus nor a pre-specified vocabulary, as they aim to discover all possible relations in the text (Banko et al., 2007b). However, these unsupervised methods have to deal with uninformative and incoherent extractions. In (Fader et al., 2011) part-of-speech based regular expressions are introduced to reduce the number of these incoherent extractions. Less restrictive pattern templates based on dependency paths are learned in (Mausam et al., 2012) to increase the number of possible extracted relations.

Dependency Parsing is an NLP technique that provides a tree-like syntactic structure of a sentence based on the linguistic theory of Dependency Grammar (Tesnière, 1959). One of the outstanding features of Dependency Grammar is that it represents binary relations between words (Ballesteros & Nivre, 2013).Dependency relations have been successfully incorporated to RE systems. For example, (Bunescu & Mooney, 2005) describe and evaluate a RE system based on shortest paths among named entities. (Culotta & Sorensen, 2004) focus on the smallest dependency subtree in the sentence that captures the entities involved in a relation, and (Gamallo et al., 2012) propose a rule-based dependency-parsing Open IE system. Moreover, in (Nakashole et al., 2012; Moro & Navigli, 2012; Bovi et al., 2015b) syntactyc and semantic information is exploited to reduce inconsistent relations, by means of the combination of Dependency Parsing and Entity Linking techniques.

2.3. Music Information Retrieval

As stated in Section 1.2, Music Information Retrieval (MIR) is a multidisciplinary field of research that is concerned with the extraction, analysis, and usage of information about music. Although MIR approaches have traditionally been focused on audio content, there have been a growing interest in text-based and multimodal approaches along the years. However, most of these text-based approaches are focused on low and mid-level text representations, ignoring the full epistemic potential expressed in texts. In addition, most of audio-based approaches have traditionally relied on handcrafted features, underexploiting all the factors of variation behind the data. In this disertation we focus on knowledge-based and multimodal feature learning approaches about three MIR tasks: music genre classification, artist similarity, and music recommendation.

2.3.1. Natural Language Processing in MIR

Early work in NLP in the context of MIR is related to the extraction of music artist information from artist-related web pages, using search engines to gather those pages and then parsing their DOM trees Cohen & Fan (2000). Other studies Ellis et al. (2002); Whitman & Lawrence (2002) use weighted term profiles based on specific term sets for recommendation and classification tasks. Co-occurrence of artist names in web pages content and page count based on results provided by search engines have been used for artist similarity and recommendation tasks Schedl et al. (2005). Song lyrics and tweets **?** are other commonly used text sources in MIR. Two comprenhensive reviews on *context-based* approaches can be found in Knees & Schedl (2013); Schedl et al. (2014).

There have been also some initial atempts to work with mid and high-level text representations in the context of MIR. In **?** a methodology for extracting semantic information from music-related forums is proposed, inferring semantic relations from the co-occurrence of musical concepts in forum posts. In **?** a set of semantic facets is automatically obtained and anchored upon the structure of Wikipedia, and tags from the folkosonomy of Last.fm are then categorized with respect to the obtained facets. In **?** a methodology to automatically extract semantic information and relations about musical entities from arbitrary textual sources is proposed. In Tata & Di Eugenio (2010) a method to extract information about indidual songs from album reviews is proposed, combining syntactic, semantic and sentiment analysis. Finally, the C@amerata task Sutcliffe et al. (2016, 2015), part of the MeidaEval evaluation campaigns from 2013 to 2017, is focused on music Question & Answering (Q&A) systems. In this task the input is a natural language phrase, together with a music score in MusicXML, and the required output should be one or more matching passages in the score.

Semantic representations have been studiend in the context of MIR, but instead of extracted from text, they have been retrieved from knowledge repositories within the Semantic Web. ...

There have been also some interesting works trying to understand the semantics behind the audio signal using natural language text. In this sense, Whitman & Ellis (2004) combines text analysis with acoustic descriptors in order to automatically generate music reviews from the audio signal. **?** shows a method for the automatic creation of an ontology of musical instruments using formal concept analysis over audio features.

2.3.2. Music Classification

Music Genre Classification

Music genre labels are useful categories to organize and classify songs, albums and artists into broader groups that share similar musical characteristics. They have been widely used for music classification, from physical music stores to streaming services. Automatic music genre classification thus is a widely explored topic (Sturm, 2012).

Most published music genre classification approaches rely on audio sources (Sturm, 2012; Bogdanov et al., 2016). Traditional techniques typically use handcrafted audio features, such as Mel Frequency Cepstral Coefficients (MFCCs) (Logan & Others, 2000), as input of a machine learning classifier (e.g., SVM, k-NN) (Tzanetakis & Cook, 2002; Seyerlehner et al., 2010). More recent deep learning approaches take advantage of visual representations of the audio signal in form of spectrograms. These visual representations of audio are used as input to Convolutional Neural Networks (CNNs) (Dieleman et al., 2011; Dieleman & Schrauwen, 2014; Pons et al., 2016; Choi et al., 2016a,b), following approaches similar to those used for image classification.

Text-based approaches have also been explored for this task. For instance, one of the earliest attempts on genre classification of music reviews is described in (Hu et al., 2005), where experiments on multiclass genre classification and star rating prediction are described. Similarly, (Hu & Downie, 2006) extend these experiments with a novel approach for predicting usages of music via agglomerative clustering, and conclude that bigram features are more informative than unigram features. Moreover, part-of-speech (POS) tags along pattern mining techniques are applied in (Downie & Hu, 2006) to extract descriptive patterns for distinguishing negative from positive reviews. Additional textual evidence is leveraged in (Choi et al., 2014), who consider lyrics as well as texts referring to the meaning of the song, and used for training a kNN classifier for predicting song subjects (e.g. war, sex or drugs).

There are a limited number of papers dealing with image-based genre classification (Libeks & Turnbull, 2011). Regariding multimodal approaches found in the literature, most of them combine audio and song lyrics as text (Laurier et al., 2008; Neumayer & Rauber, 2007). Moreover, other modalities such as audio and video have been explored (Schindler & Rauber, 2015).

Almost all related work about Music Genre Classification is concentrated in multi-class classification of music items into broad genres (e.g., Pop, Rock), assigning a single label per item. This is problematic since there may be hundreds of more specific music genres (Pachet & Cazaly, 2000), and these may not be necessarily mutually exclusive (i.e., a song could be Pop, and at the same time have elements from Deep House and a Reggae grove). Multi-label classification is a widely studied problem (Tsoumakas & Katakis, 2006; Jain et al., 2016). Although there are not many approaches for multi-label classification of music genres (Sanden & Zhang, 2011; Wang et al., 2009), there is a long tradition in MIR for tag classification, which is a highly related multi-label problem (Choi et al., 2016a; Wang et al., 2009).

2.3.3. Artist Similarity

Although artist similarity may be seen as a subtask of music recommendation where no user information is ivolved, we decided to address its literature review separately, given that it has become a proper task in the context of MIR. Music artist similarity has been studied from the score level, the acoustic level, and the cultural level (Ellis et al., 2002). In this disertation, we focus on the latter approach, and more specifically in text-based approaches. Literature on document similarity, and more specifically on the application of text-based approaches for artist similarity is discussed next.

The task of identifying similar text instances, either at sentence or document level, has applications in many areas of Artificial Intelligence and Natural Language Processing (Liu & Wang, 2014). In general, document similarity can be computed according to the following approaches: surface-level representation like keywords or n-grams (Chim & Deng, 2008); corpus representation using counts (Rorvig, 1999), e.g. word-level correlation, jaccard or cosine models; Latent factor models, such as Latent Semantic Analysis (Deerwester et al., 1990); or methods exploiting external knowledge bases like ontologies or encyclopedias (Hu et al., 2009).

The use of text-based approaches for artist and music similarity was first applied in (Cohen & Fan, 2000), by computing co-occurrences of artist names in web page texts and building term vector representations. By contrast, in (Schedl et al., 2005) term weights are extracted from search engine's result counts. In (Whitman & Lawrence, 2002) n-grams, part-of-speech tagging and noun phrases are used to build a term profile for artists, weighted by employing tf-idf. Term profiles are then compared and the sum of common terms weights gives the similarity measure. More approaches using term weight vectors have been developed over different text sources, such as music reviews (Hu et al., 2005), blog posts (Celma et al., 2006), or microblogs (Schedl et al., 2013). In (Logan & Ellis, 2003) Latent Semantic Analysis is used to measure artist similarity from song lyrics. Domain specific ontologies have also been applied to the problem of music recommendation and similarity, such as in (Celma & Serra, 2008). In (Leal et al., 2012), paths on an ontological graph extracted from DBpedia are exploited for recommending music web pages.

2.3.4. Recommender Systems

Within the recommender systems arena, there are two main approaches for computing recommendations: collaborative filtering (CF) and contend-based ones. The most popular is collaborative filtering which provides recommendations to a user by considering the preferences of other users with similar tastes. Matrix factorization techniques are currently CF state-of-the-art (Koren et al., 2009). As CF methods rely only on users feedback information, they may suffer from the so-called cold-start problem (Saveski & Mantrach, 2014). That is, when new items are introduced in the system, they can not be initially recommended as there is no feedback information related to them. Content-based systems recommend items sharing similar features to those a user has preferred in past. Both approaches can be combined to build hybrid systems Burke (2002). When available, the usage of side information about items has proven to boost the performances of pure collaborative-filtering techniques **?**.

Semantic-based Approaches

Ontology-based and semantics-aware recommendation systems have been proposed in many works in the past. In (Middleton et al., 2009) an ontological recommender system is presented that makes use of semantic user profiles to compute collaborative recommendations with the effect of mitigating cold-start and improving overall recommendation accuracy. In (Mobasher et al., 2004) the authors present a *semantically enhanced collaborative filtering* approach, where structured semantic knowledge about items is used in conjunction with user-item ratings to create a combined similarity measure for item comparisons. In (Ziegler et al., 2004) taxonomic information is used to represents the user's interest in categories of products. Consequently, user similarity is determined by common interests in categories and not by common interests in items. In (Anand et al., 2007) the authors present an approach that infers user preferences from rating data using an item ontology. The system collaboratively generates recommendations using the ontology and infers preferences during similarity computation. Another hybrid ontological recommendation system is proposed in (Cantador et al., 2008) where user preferences and item features are described by semantic concepts to obtain users' clusters corresponding to implicit *Communities of Interest.* In all of these works, the experiments prove an accuracy improvement over traditional memory-based collaborative approaches especially in presence of sparse datasets.

In the last few years with the availability of Linked Open Data (LOD) datasets, a new class of recommender systems has emerged which can be named as LOD-based recommender systems. One of the first approaches that exploits Linked Open Data for building recommender systems is (Heitmann & Hayes, 2010). In (Fernández-Tob\'\ias et al., 2011) the authors present a knowledge-based framework leveraging DBpedia for computing cross-domain recommendations. In (Di Noia et al., 2012a,b) a model-based approach and a memory-based one to compute content-based recommendations are presented leveraging LOD datasets. Another LOD content-based method is presented in (Ostuni et al., 2014) which defines a neighborhood-based graph kernel for matching graph-based item representations. Two hybrid approaches have been presented lately. In (Ostuni et al., 2013) the authors show how to compute top-N recommendations from implicit feedback using linked data sources and in (Khrouf & Troncy, 2013) the authors propose an event recommendation system based on linked data and user diversity. Finally, another interesting direction about the usage of LOD for content-based RSs is explored in (Musto et al., 2014) where the authors present Contextual eVSM, a content-based context-aware recommendation framework that adopts a semantic representation based on distributional models and entity linking techniques. In particular entity linking is used to detect entities in free text and map them to LOD.

Music Recommendation

An extensive description of the music recommendation problem and a comprenhensive summarization of the initial attempts to tackle it is presented in (Celma, 2010b). An overview about techiques for music recommendation and similarity based on music contextual data is given in (Knees & Schedl, 2013). In (Kaminskas & Ricci, 2012) the authors provide a description of various tools and techniques that can be used for addressing the research challenges posed by context-aware music retrieval and recommendation. A survey about techniques for the generation of music playlists is given in (Bonnin & Jannach, 2014). In particular, the authors provide a review of the literature on automated playlist generation and a categorization of the existing approaches. A context-aware music recommender system which infers contextual information based on the most recent sequence of songs liked by the user is presented in (Hariri et al., 2012). More recently, a playlist generation algorithm with the goal of maximizing coherence and personalization of the playlist has been presented in (Jannach et al., 2015). Finally, in (Aghdam et al., 2015) a technique for adapting recommendations to contextual changes based on hierarchical hidden Markov models is presented.

Social tags have been extensively used as a source of content features to recommend music (Knees & Schedl, 2013). However, these tags are usually collectively annotated, which often introduce an artist popularity bias (Turnbull et al., 2008). Artist biographies and press releases, on the other hand, do not necessarily require a collaborative effort, as they may be produced by artists themselves. However, they have seldom been exploited for music recommendation. Part of the work on this disertation focuses on the exploitation of artist biographies.

Audio signals have also been widely used as a source of content features. Content-based approaches have shown useful when user feedback information is scarce, as in cold-start scenarios. Traditional audio-based approaches rely on handcrafted features obtained from audio signals (Bogdanov et al., 2013b). However, as in many other disciplines and MIR tasks, the application of deep learning approaches has suppossed a boost in the performance of music recommendation systems (van den Oord et al., 2013).

Multimodal approaches for content-based Music Recommendation typically combines audio and textual data, which most commonly consists of web documents, lyrics and social tags (Liem et al., 2011). In (Bogdanov & Herrera, 2011), for instance, it is evaluated how much metadata is necessary to use in order to improve the quality of audio-based recommendations. In (Eck et al., 2008), tags are first learned from audio separately and then combined with the audio in a recommendation system. However, to the best of our knowledge, there is not a multimodal system that make use of deep learning approaches for music recommendation nor music classification.

Sergio Oramas Martin, Barcelona, 11 March 2017.

Bibliography

Adomavicius, G. & Kwon, Y. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, *24*(5), 896–911.

Aghdam, M. H., Hariri, N., Mobasher, B., & Burke, R. D. (2015). Adapting Recommendations to Contextual Changes Using Hierarchical Hidden Markov Models. In *Proceedings of the 9th {ACM} Conference on Recommender Systems, RecSys 2015, Vienna, Austria, September 16-20, 2015*, pp. 241–244.

Anand, S. S., Kearney, P., & Shapcott, M. (2007). Generating semantically enriched user profiles for Web personalization. *ACM Trans. Internet Technol.*, *7*(4).

Bach, N. & Badaskar, S. (2007). A Review of Relation Extraction. *Literature review for Language and Statistics II*.

Ballesteros, M. & Nivre, J. (2013). Going to the Roots of Dependency Parsing. *Computational Linguistics*, *39*(1), 5–13.

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007a). Open Information Extraction for the Web. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 7, pp. 2670–2676.

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007b). Open Information Extraction from the Web. In *International Joint Conferences on Artificial Intelligence*, pp. 2670–2676.

Baral, C. & De Giacomo, G. (2015). Knowledge Representation and Reasoning: What's Hot. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 4316–4317.

Bellogín, A., Cantador, I., & Castells, P. (2010). A Study of Heterogeneity in Recommendations for a Social Music Service. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, HetRec '10, pp. 1–8. ACM.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

Bogdanov, D. & Herrera, P. (2011). How much metadata do we need in music recommendation? a subjective evaluation using preference sets. In *International Society for Music Information Retrieval Conference (ISMIR)*. Miami, Florida, USA.

Bogdanov, D., Porter, A., Herrera, P., & Serra, X. (2016). Cross-collection evaluation for music classification tasks. In *Proc. of the Int. Conf. on Music Information Retrieval (ISMIR)*.

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., Serra, X. et al. (2013a). Essentia: An audio analysis library for music information retrieval. In *ISMIR*, pp. 493–498.

Bogdanov, D., Wack, N., & Others (2013b). ESSENTIA: an Open-Source Library for Sound and Music Analysis. In *ACM International Conference on Multimedia (MM{\textquoteright}13)*, pp. 855–858.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pp. 1247–1250. New York, NY, USA: ACM.

Bonnin, G. & Jannach, D. (2014). Automated Generation of Music Playlists: Survey and Experiments. *ACM Comput. Surv.*, *47*(2), 26:1—-26:35.

Bovi, C. D., Espinosa-Anke, L., & Navigli, R. (2015a). Knowledge Base Unification via Sense Embeddings and Disambiguation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 726–736.

Bovi, C. D., Telesca, L., & Navigli, R. (2015b). Large-Scale Information Extraction from Textual Definitions through Deep Syntactic and Semantic Analysis. *Transactions of the Association for Computational Linguistics (TACL)*, *3*, 529–543.

Bunescu, R. & Pasca, M. (2006). Using Encyclopedic Knowledge for Named Entity Disambiguation. In *Proceesings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, pp. 9–16. Trento, Italy.

Bunescu, R. C. & Mooney, R. J. (2005). A Shortest Path Dependency Kernel for Relation Extraction. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pp. 724–731.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, *12*(4), 331–370.

Cambria, E. & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research. *Computational Intelligence Magazine, IEEE*, *9*(2), 48–57.

Cantador, I., Bellogín, A., & Castells, P. (2008). A multilayer ontology-based hybrid recommendation model. *AI Commun. Special Issue on Rec. Sys.*, *21*(2-3), 203–210.

Carlson, A., Betteridge, J., Wang, R. C., Hruschka Jr, E., & Mitchell, T. M. (2010). Coupled Semi-Supervised Learning for Information Extraction. In *Proceedings of the third ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 101–110.

Celma, Ò. (2010a). *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer.

Celma, Ò. (2010b). *The Long Tail in Recommender Systems*, pp. 87–107. Berlin, Heidelberg: Springer Berlin Heidelberg.

Celma, Ò., Cano, P., & Herrera, P. (2006). Search Sounds An audio crawler focused on weblogs. In *7th International Conference on Music Information Retrieval (ISMIR)*.

Celma, Ò. & Herrera, P. (2008). A new approach to evaluating novel recommendations. In *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 179–186. ACM.

Celma, Ò. & Serra, X. (2008). FOAFing the music: Bridging the semantic gap in music recommendation. *Web Semantics*, *6*, 250–256.

Chim, H. & Deng, X. (2008). Efficient phrase-based document similarity for clustering. *Knowledge and Data Engineering, IEEE Transactions on*, *20*(9), 1217–1229.

Choi, K., Fazekas, G., & Sandler, M. (2016a). Automatic tagging using deep convolutional neural networks. *International Society for Music Information Retrieval Conference*, pp. 805–811.

Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2016b). Convolutional Recurrent Neural Networks for Music Classification. *arXiv preprint arXiv:1609.04243*.

Choi, K., Lee, J. H., & Downie, J. S. (2014). What is this song about anyway?: Automatic classification of subject using user interpretations and lyrics. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, pp. 453–454.

Cohen, W. W. & Fan, W. (2000). Web-collaborative filtering: recommending music by crawling the Web. *Computer Networks*, *33*, 685–698.

Cowie, J. & Lehnert, W. (1996). Information extraction. *Communications of the ACM*, *39*(1), 80–91.

Culotta, A. & Sorensen, J. (2004). Dependency Tree Kernels for Relation Extraction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *JAsIs*, *41*(6), 391–407.

Di Noia, T., Mirizzi, R., Ostuni, V. C., & Romito, D. (2012a). Exploiting the Web of Data in Model-based Recommender Systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys '12, pp. 253–256. New York, NY, USA: ACM.

Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M. (2012b). Linked open data to support content-based recommender systems. In *Proceedings of the 8th International Conference on Semantic Systems*, I-SEMANTICS '12, pp. 1–8. New York, NY, USA: ACM.

Dieleman, S., Brakel, P., & Schrauwen, B. (2011). Audio-based music classification with a pretrained convolutional network. In *12th International Society for Music Information Retrieval Conference (ISMIR-2011)*, pp. 669–674. University of Miami.

Dieleman, S. & Schrauwen, B. (2014). End-to-end learning for music audio. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 6964–6968. IEEE.

Downie, J. S. & Hu, X. (2006). Review mining for music digital libraries: phase II. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, p. 196.

Eck, D., Lamere, P., Bertin-Mahieux, T., & Green, S. (2008). Automatic Generation of Social Tags for Music Recommendation. In *Advances in Neural Information Processing Systems 20*, pp. 385–392. Cambridge, MA: MIT Press.

Ellis, D. P. W., Ellis, D. P., Whitman, B., Berenzweig, A., & Lawrence, S. (2002). The quest for ground truth in musical artist similarity. In *Proc. International Symposium on Music Information Retrieval (ISMIR 2002)*, pp. 170–177.

Fader, A., Soderland, S., & Etzioni, O. (2011). Identifying relations for open information extraction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP '11*, pp. 1535–1545.

Fernández-Tob\'\ias, I., Cantador, I., Kaminskas, M., & Ricci, F. (2011). A generic semantic-based framework for cross-domain recommendation. In *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, HetRec '11, pp. 25–32. New York, NY, USA: ACM.

Ferragina, P. & Scaiella, U. (2012). Fast and accurate annotation of short texts with Wikipedia pages. *Software, IEEE*, *29*(1).

Gamallo, P., Garcia, M., & Fernández-Lanza, S. (2012). Dependency-based Open Information Extraction. In *Proceedings of the Joint Workshop on Unsupervised and Semi-Supervised Learning in NLP*, ROBUS-UNSUP '12, pp. 10–18.

Gruhl, D., Nagarajan, M., Pieper, J., Robson, C., & Sheth, A. (2009). Context and Domain Knowledge Enhanced Entity Spotting In Informal Text. In *The Semantic Web-ISWC*, pp. 260–276. Springer.

Hariri, N., Mobasher, B., & Burke, R. D. (2012). Context-aware music recommendation based on latenttopic sequential patterns. In *Sixth {ACM} Conference on Recommender Systems, RecSys '12, Dublin, Ireland, September 9-13, 2012*, pp. 131–138.

Havasi, C., Speer, R., & Alonso, J. (2007). ConceptNet 3: A Flexible, Multilingual Semantic Network for Common Sense Knowledge. In *Proceedings of Recent Advances in Natural Language Processing*, pp. 27–29. Citeseer.

Heitmann, B. & Hayes, C. (2010). Using Linked Data to Build Open, Collaborative Recommender Systems. In *AAAI Spring Symposium: Linked Data Meets Artificial Intelligence.*

Hoffmann, R., Zhang, C., Ling, X., Zettlemoyer, L., & Weld, D. S. (2011). Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations. *Network*, pp. 541–550.

Hu, X. & Downie, J. (2006). Stylistics in customer reviews of cultural objects. *SIGIR Forum*, pp. 49–51.

Hu, X., Downie, J., West, K., & Ehmann, A. (2005). Mining Music Reviews: Promising Preliminary Results. In *ISMIR*, pp. 536–539.

Hu, X., Zhang, X., Lu, C., Park, E. K., & Zhou, X. (2009). Exploiting Wikipedia as external knowledge for document clustering. In *Proceedings of the*

15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 389–396. ACM.

Humphrey, E. J., Bello, J. P., & LeCun, Y. (2012). Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In *ISMIR*, pp. 403–408.

Jain, H., Prabhu, Y., & Varma, M. (2016). Extreme Multi-label Loss Functions for Recommendation, Tagging, Ranking & Other Missing Label Applications. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 935–944. ACM.

Jannach, D., Lerche, L., & Kamehkhosh, I. (2015). Beyond "Hitting the Hits": Generating Coherent Music Playlist Continuations with the Right Tracks. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pp. 187–194. New York, NY, USA: ACM.

Jiang, J. & Zhai, C. (2007). A Systematic Exploration of the Feature Space for Relation Extraction. In *HLT-NAACL*, pp. 113–120.

Kaminskas, M. & Ricci, F. (2012). Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, *6*(2-3), 89–119.

Khrouf, H. & Troncy, R. (2013). Hybrid Event Recommendation Using Linked Data and User Diversity. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pp. 185–192. New York, NY, USA: ACM.

Knees, P. & Schedl, M. (2013). A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, *10*(1).

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, *42*(8), 42–49.

Laurier, C., Grivolla, J., & Herrera, P. (2008). Multimodal music mood classification using audio and lyrics. In *Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on*, pp. 688–693. IEEE.

Leal, J. P., Rodrigues, V., & Queirós, R. (2012). Computing Semantic Relatedness using DBPedia. *1st Symposium on Languages, Applications and Technologies, SLATE 2012.*

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., & Bizer, C. (2014). {DBpedia} - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web Journal.*

Libeks, J. & Turnbull, D. (2011). You can judge an artist by an album cover: Using images for music annotation. *IEEE MultiMedia*, *18*(4), 30–37.

Liem, C., Müller, M., Eck, D., Tzanetakis, G., & Hanjalic, A. (2011). The need for music information retrieval with user-centered and multimodal strategies. In *Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pp. 1–6. ACM.

Liu, H. & Wang, P. (2014). Assessing Text Semantic Similarity Using Ontology. *Journal of Software*, *9*(2), 490–497.

Logan, B. & Ellis, D. P. W. (2003). Toward Evaluation Techniques for Music Similarity. *SIGIR 2003: Workshop on the Evaluation of Music Information Retrieval Systems*, pp. 7–11.

Logan, B. & Others (2000). Mel Frequency Cepstral Coefficients for Music Modeling. In *ISMIR*.

Mausam, Schmitz, M., Bart, R., Soderland, S., & Etzioni, O. (2012). Open Language Learning for Information Extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

Mcfee, B., Raffel, C., Liang, D., Ellis, D. P. W., Mcvicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and Music Signal Analysis in Python. *PROC. OF THE 14th PYTHON IN SCIENCE CONF*, (Scipy), 1–7.

Mendes, P. N., Jakob, M., García-Silva, A., & Bizer, C. (2011). DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*, pp. 1–8. ACM.

Middleton, S. E., Roure, D. D., & Shadbolt, N. R. (2009). Ontology-Based Recommender Systems. *Handbook on Ontologies*, *32*(6), 779–796.

Mihalcea, R. & Csomai, A. (2007). Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pp. 233–242. ACM.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, *38*(11), 39–41.

Mobasher, B., Jin, X., & Zhou, Y. (2004). Semantically Enhanced Collaborative Filtering on the Web. In B. Berendt, A. Hotho, D. Mladenic, M. Someren, M. Spiliopoulou, & G. Stumme (Eds.) *Web Mining: From Web to Semantic Web*, *Lecture Notes in Computer Science*, vol. 3209, pp. 57–76. Springer Berlin Heidelberg.

Moro, A., Cecconi, F., & Navigli, R. (2014a). Multilingual Word Sense Disambiguation and Entity Linking for Everybody. In *Proceedings of the 13th Internation Conference on Semantic Web (P&D).*

Moro, A. & Navigli, R. (2012). WiSeNet: Building a Wikipedia-based Semantic Network with Ontologized Relations. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1672–1676.

Moro, A. & Navigli, R. (2013). Integrating Syntactic and Semantic Analysis into the Open Information Extraction Paradigm. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2148–2154. AAAI Press.

Moro, A., Raganato, A., & Navigli, R. (2014b). Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, *2*, 231–244.

Musto, C., Semeraro, G., Lops, P., & de Gemmis, M. (2014). Combining Distributional Semantics and Entity Linking for Context-Aware Content-Based Recommendation. In *User Modeling, Adaptation, and Personalization - 22nd International Conference, {UMAP} 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings*, pp. 381–392.

Nakashole, N., Weikum, G., & Suchanek, F. M. (2012). PATTY: A Taxonomy of Relational Patterns with Semantic Types. *EMNLP-CoNLL*, (July), 1135–1145.

Navigli, R. & Ponzetto, S. P. (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network. *Artificial Intelligence*, *193*, 217–250.

Neumayer, R. & Rauber, A. (2007). Integration of text and audio features for genre classification in music information retrieval. In *European Conference on Information Retrieval*, pp. 724–727. Springer.

Ostuni, V. C., Di Noia, T., Di Sciascio, E., & Mirizzi, R. (2013). Top-N Recommendations from Implicit Feedback Leveraging Linked Open Data. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pp. 85–92. New York, NY, USA: ACM.

Ostuni, V. C., Di Noia, T., Mirizzi, R., & Di Sciascio, E. (2014). A Linked Data Recommender System using a Neighborhood-based Graph Kernel. In *The 15th International Conference on Electronic Commerce and Web Technologies*, Lecture Notes in Business Information Processing. Springer-Verlag.

Pachet, F. & Cazaly, D. (2000). A taxonomy of musical genres. In *Content-Based Multimedia Information Access-Volume 2*, pp. 1238–1245. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE.

Passant, A. (2010). dbrec: music recommendations using DBpedia. In *Proc. of 9th Int. Sem. Web Conf.*, ISWC'10, pp. 209–224.

Passant, A. & Raimond, Y. (2008). Combining Social Music and Semantic Web for music-related recommender systems. In *Social Data on the Web Workshop*.

Pons, J., Lidy, T., & Serra, X. (2016). Experimenting with musically motivated convolutional neural networks. In *Content-Based Multimedia Indexing (CBMI), 2016 14th International Workshop on*, pp. 1–6. IEEE.

Rorvig, M. (1999). Images of similarity: A visual exploration of optimal similarity metrics and scaling properties of TREC topic-document sets. *Journal of the American Society for Information Science*, *50*(8), 639–651.

Sanden, C. & Zhang, J. Z. (2011). Enhancing Multi-label Music Genre Classification Through Ensemble Techniques. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pp. 705–714. New York, NY, USA: ACM.

Saveski, M. & Mantrach, a. (2014). Item cold-start recommendations: learning local collective embeddings. *RecSys '14 Proceedings of the 8th ACM Conference on Recommender systems*, pp. 89–96.

Schedl, M. (2008). *Automatically extracting, analyzing, and visualizing information on music artists from the World Wide Web*. na.

Schedl, M., Gómez, E., & Urbano, J. (2014). Music Information Retrieval: Recent Developments and Applications. *Foundations and Trends® in Information Retrieval*, *8*(2-3), 127–261.

Schedl, M., Hauger, D., & Urbano, J. (2013). Harvesting microblogs for contextual music similarity estimation: a co-occurrence-based framework. *Multimedia Systems*, *20*(6), 693–705.

Schedl, M., Knees, P., & Widmer, G. (2005). A Web-Based Approach to Assessing Artist Similarity using Co-Occurrences. In *Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing {(CBMI'05)}*.

Schindler, A. & Rauber, A. (2015). An audio-visual approach to music genre classification through affective color features. In *European Conference on Information Retrieval*, pp. 61–67. Springer.

Seyerlehner, K., Schedl, M., Pohle, T., & Knees, P. (2010). Using block-level features for genre classification, tag classification and music similarity estimation. *Submission to Audio Music Similarity and Retrieval Task of MIREX, 2010.*

Sordo, M. et al. (2012). Semantic annotation of music collections: A computational approach.

Sturm, B. L. (2012). A survey of evaluation in music genre recognition. In *International Workshop on Adaptive Multimedia Retrieval*, pp. 29–66. Springer.

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: A Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, pp. 697–706. ACM.

Sutcliffe, R., Crawford, T., Fox, C., Root, D. L., & Hovy, E. (2015). Relating natural language text to musical passages. *Proceedings of the 16th International Society for Music Information Retrieval Conference.*

Sutcliffe, R. F., Collins, T., Hovy, E. H., Lewis, R., Fox, C., & Root, D. L. (2016). The c@merata task at mediaeval 2016: Natural language queries derived from exam papers, articles and other sources against classical music scores in musicxml. In *Proceedings of the MediaEval 2016 Workshop.*

Tata, S. & Di Eugenio, B. (2010). Generating Fine-Grained Reviews of Songs from Album Reviews. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1376–1385. Association for Computational Linguistics.

Tesnière, L. (1959). *Elements de Syntaxe Structurale.* Editions Klincksieck.

Tsoumakas, G. & Katakis, I. (2006). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, *3*(3).

Turnbull, D., Barrington, L., & Lanckriet, G. (2008). Five approaches to collecting tags for music. *Proceedings of 9th the International Society for Music Information Retrieval Conference*, pp. 225–230.

Tzanetakis, G. & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302.

Usbeck, R., Röder, M., Ngomo, A.-C. N., Baron, C., Both, A., Brümmer, M., Ceccarelli, D., Cornolti, M., Cherix, D., Eickmann, B., Ferragina, P., Lemke, C., Moro, A., Navigli, R., Piccinno, F., Rizzo, G., Sack, H., Speck, R., Troncy, R., Waitelonis, J., & Wesemann, L. (2015). GERBIL – General Entity Annotator Benchmarking Framework. *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22, 2015*, pp. 1133–1143.

van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. *Electronics and Information Systems department (ELIS)*, p. 9.

Wang, F., Wang, X., Shao, B., Li, T., & Ogihara, M. (2009). Tag Integrated Multi-Label Music Style Classification with Hypergraph. In *ISMIR*, pp. 363–368.

Wang, M. (2008). A Re-examination of Dependency Path Kernels for Relation Extraction. In *IJCNLP*, pp. 841–846.

Whitman, B. & Ellis, D. P. W. (2004). Automatic record reviews. *Proceedings of 5th International Conference on Music Information Retrieval.*

Whitman, B. & Lawrence, S. (2002). Inferring descriptions and similarity for music from community metadata. In *Proceedings of the 2002 International Computer Music Conference*, pp. 591–598.

Zhang, X., Liu, Z., Qiu, H., & Fu, Y. (2009). A Hybrid Approach for Chinese Named Entity Recognition in Music Domain. *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, pp. 677–681.

Ziegler, C.-N., Lausen, G., & Schmidt-Thieme, L. (2004). Taxonomy-driven computation of product recommendations. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, CIKM '04, pp. 406–415. New York, NY, USA: ACM.

Appendix A: Datasets and Knowledge Bases

Introduction

Appendix B: publications by the author

In press

Journal papers

Oramas S., Espinosa-Anke L., Sordo M., Saggion H. & Serra X. (2016). Information Extraction for Knowledge Base Construction in the Music Domain. *Data & Knowledge Engineering, Volume 106*, Pages 70-83.

Oramas S., Ostuni V. C., Di Noia T., Serra, X., & Di Sciascio E. (2016). Music and Sound Recommendation with Knowledge Graphs. *ACM Transactions on Intelligent Systems and Technology, Volume 8*, Issue 2, Article 21.

Oramas S., Sordo M. (2016). Knowledge is Out There: A New Step in the Evolution of Music Digital Libraries. *Fontes Artis Musicae, Vol 63, no. 4*.

Conference papers

Oramas S., Espinosa-Anke L., Lawlor A., Serra X., & Saggion H. (2016). Exploring Music Reviews for Music Genre Classification and Evolutionary Studies. *In Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016).*

Oramas S., Espinosa-Anke L., Sordo M., Saggion H., & Serra X. (2016). ELMD: An Automatically Generated Entity Linking Gold Standard in the Music Domain. *In Proceedings of the 10th Conference on Language Resources and Evaluation (LREC 2016).*

Espinosa-Anke, L., Oramas S., Camacho-Collados J., & Saggion H. (2016). Finding and Expanding Hypernymic Relations in the Music Domain. *In Proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence (CCIA 2016).*

Oramas S., Sordo M., Espinosa-Anke L., & Serra X. (2015). A Semantic-based approach for Artist Similarity. *In Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015).*

Oramas S., Gómez F., Gómez E., & Mora J. (2015). FlaBase: Towards the creation of a Flamenco Music Knowledge Base. *In Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015).*

Ostuni V. C., Oramas S., Di Noia T., Serra, X., & Di Sciascio E. (2015). A Semantic Hybrid Approach for Sound Recommendation. *In Proceedings of the 24th International World Wide Web Conference (WWW 2015, Poster track).*

Oramas S., Sordo M., & Espinosa-Anke L. (2015). A Rule-based Approach to Extracting Relations from Music Tidbits. *In Proceedings of the 2nd Workshop on Knowledge Extraction from Text (KET 2015).*

Sordo, M., Oramas S., & Espinosa-Anke L. (2015). Extracting Relations from Unstructured Text Sources for Music Recommendation. *In Proceedings of the 20th International Conference on Applications of Natural Language to Information Systems (NLDB 2015).*

Oramas S., Sordo M., & Serra X. (2014). Automatic Creation of Knowledge Graphs from Digital Musical Document Libraries. *In Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM 2014).*

Oramas S. (2014). Harvesting and Structuring Social Data in Music Information Retrieval. *In Proceedings of the Extended Semantic Web Conference (ESWC 2014, PhD Symposium).*

Font, F., Oramas, S., Fazekas, G., & Serra, X. (2014). Extending Tagging Ontologies with Domain Specific Knowledge. In *Proceedings of the International Semantic Web Conference (ISWC 2014, Poster track).*

Tutorials and Challenges

Oramas S., Espinosa-Anke L., Zhang S., Saggion H., & Serra X. (2016). Natural Language Processing for Music Information Retrieval. *17th International Society for Music Information Retrieval Conference (ISMIR 2016).*

Conference presentations

Oramas, S. (2017). Discovering Similarities and Relevance Ranking of Renaissance Composers. *The 63rd Annual Meeting of the Renaissance Society of America (RSA)*, Chicago.

Oramas S. (2015). Information Extraction for the Music Domain. *The 2nd International Workshop on Human History Project: Natural Language Processing and Big Data*, CIRMMT, Montreal.

Oramas, S., & Sordo M. (2015). Knowledge Acquisition from Music Digital Libraries. *The International Association of Music Libraries and International Musicological Society Conference (IAML/IMS 2015)*, New York.