Objectives

Provide a general introduction to NLP.

Identify areas of NLP with potential application in MIR.

Address the extraction of semantic information from music text corpora.

Show methodologies for exploiting semantic information in MIR.

Illustrate latest tendencies in NLP
Why semantic information?
Introduction
NLP for MIR
Sergio Oramas

MIR tasks
Introduction

NLP for MIR

Sergio Oramas
Introduction

NLP for MIR

Sergio Oramas

LOREM IPSUM DOLOR SIT AMET CONSECTETUR ADIPISCING ELIT NUNC VITAE NEQUE DUIS VEL FEUGIAT NUNC PRAESENT SOLlicitudi N ALIQUET SAGITTIS.
Introduction

NLP for MIR

Sergio Oramas

MIR tasks

semantic / high level representations

surface features

LOREM IPSUM DOLOR SIT AMET CONSECTETUR ADIPISCING ELIT NUNC VITAE NEQUE DUIS VEL FEUGIAT NUNC PHARESENT SOLICITUDI N ALIQUET SAGITTIS.
Introduction

NLP for MIR

Sergio Oramas

MIR tasks

semantic / high level representations
Introduction

This tutorial focuses on the use of NLP for MIR. MIR tasks include working with surface features and semantic/high-level representations. This tutorial aims to provide insights into how these representations can be effectively utilized in MIR tasks.
Corpora in MIR Related Work

Lyrics

Biographies, blogs, forums, encyclopedias, digital libraries, social networks
Corpora in MIR Related Work

Lyrics

- Biographies, blogs, forums, encyclopedias, digital libraries, social networks

This tutorial
Outline

- Introduction to NLP (20 mins)
- Information Extraction (10 mins)
  - Construction of Music Knowledge Bases (15 mins)
  - Semantic Enrichment of Musical Texts (5 mins)
- Applications in MIR (25 mins)

--- break ---

- Applications in Musicology (10 mins)
- Lexical Semantics (15 mins)
- Deep Learning (10 mins)
- Conclusions and Future (5 mins)
Outline

- **Introduction to NLP**
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Introduction to NLP
Outline

· What is Natural Language Processing?
· NLP Core Tasks
· Applications
· Knowledge Repositories
· Resources
What is Natural Language Processing?

- NLP is a field of Computer Science and Artificial Intelligence concerned with the interaction between computers and human (natural) language.

- Alan Turing’s paper *Computing Machinery and Intelligence* is believed to be the first NLP paper. It stated that a computer could be considered intelligent if it could carry on a conversation with a human being without the human realizing he/she were talking to a machine.
What is Natural Language Processing?

· There are over 7k languages in the world. Cultural and sociological traces

· “In the future, the most useful data will be the kind that was too unstructured to be used in the past.” [“The future of big data is quasi-unstructured,” Chewy Chunks, 23 March 2013] (from Wired.com).

· NLP is a core component in daily life technologies: web search, speech recognition and synthesis, automatic summaries in the web, product (including music) recommendation, machine translation...
Why is it hard?

I’m a huge metal fan!
Why is it hard?

I’m a huge metal fan!
Why is it hard?

I’m a huge metal fan!
NLP is not a large uniform task

- Core NLP Tasks
  - * Part-of-speech Tagging
  - * Syntactic Parsing
  - * Semantic Parsing
  - * Named Entity Recognition
  - * Coreference Resolution
  - * Word Sense Disambiguation (WSD) & Entity Linking (EL)

Successful NLP: “Will a computer program ever be able to convert a piece of English text into a programmer friendly data structure that describes the meaning of the natural language text? Unfortunately, no consensus has emerged about the form or the existence of such a data structure” (Collobert et al., 2011).
Core elements in NLP - Part-of-Speech Tagging

I like jazz music, it’s like being alive for a second.
I like jazz music, it’s like being alive for a second.

NOUN VERB NOUN NOUN PUNCT NOUN VERB ADP VERB ADJ ADP DET ADJ PUNCT
Core elements in NLP

One morning I shot an elephant in my pajamas.
How he got into my pajamas I'll never know.

(Groucho Marx)

http://www.nltk.org/book/ch08.html
Core elements in NLP - Syntactic Parsing

- Identify relations holding between words or phrases in the sentence, and what is their function.

- By analyzing sentence structure, we understand the underlying meaning in a sentence.

http://www.nltk.org/book/ch08.html
Core elements in NLP - Constituency Parsing

· Identify relations holding between words or phrases in the sentence, and what is their function.

· By analyzing sentence structure, we understand the underlying meaning in a sentence.

http://www.nltk.org/book/ch08.html
Core elements in NLP - Dependency Parsing

- Identify relations holding between words or phrases in the sentence, and what is their function.

- By analyzing sentence structure, we understand the underlying meaning in a sentence.

http://www.nltk.org/book/ch08.html
Core elements in NLP - Semantic Parsing

- A level of parsing above morphology and syntax. Capture underlying semantics expressed in language. Most focus on verbs and their arguments.

- A PropBank ([http://propbank.github.io/](http://propbank.github.io/)) Example:

  -> Mary *left* the room

  * Arg0: **Entity leaving**, Arg1: **Place left**

  -> Mary *left* her daughter her pearls

  * Arg0: **Giver**, Arg1: **Thing given**, Arg2: **Beneficiary**.
**Core elements in NLP - Named Entity Recognition**

- **Manfred Mann's Earth Band** is a British **progressive rock** group formed in 1971 by **Manfred Mann**, a keyboard player born in **South Africa** best known as a founding member and namesake of 60s group **Manfred Mann**.

- **Band**, **Music Genre**, **Artist**, **Country**
Core elements in NLP - Coreference Resolution

“I voted for Nader because he was most aligned with my values,” she said.

Core elements in NLP - **WSD** and EL

- “The performance of that bass player was outstanding”
Core elements in NLP - **WSD** and **EL**

- “The performance of that bass player was outstanding”

[Image: https://tackyraccoons.com/2011/11/21/all-your-bass-are-belong-to-us/]
NLP is not a large uniform task

- NLP Tasks
  - Summarization
  - Author Profiling
  - Machine Translation
  - Sentiment Analysis
NLP Tasks - Summarization

· Extractive
  * Retains most important sentences.

· Abstractive
  * Reformulates most important info.
NLP Tasks - Author Profiling

- Revealing demographic traces behind the writer of a message (cybersecurity), aka digital text forensics.

* From PAN 2016

```xml
<author id="{author-id}" lang="en|es|nl" age_group="18-24|25-34|35-49|50-64|65-xx" gender="male|female"/>
```
NLP Tasks - Machine Translation

· Given text in L1, translate it into L2.

· One of the most widely known NLP tasks

· Originally it was approached as a rule-based task. Today, statistical approaches have taken over.

· Apertium is one of the best known RBMT systems (www.apertium.org).

· SMT is, by far, the most studied MT discipline. Challenges include *sentence alignment*, *word alignment*, *statistical anomalies*, *idioms*, *different word orders*, *OOV*. 
Sentiment Analysis

Computational study of **opinions**, **sentiments**, **subjectivity**, evaluations, attitudes, appraisal, affects, views, **emotions**, etc., expressed in text.

**Complex NLP task**


Sentiment Analysis

Computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in text.

Complex NLP task


go read the book!
Knowledge Repositories and Knowledge Bases

- A Knowledge Base (KB) is a rich form of Knowledge Repository (KR), term coined to differentiate from traditional *databases*.

- The term KB may be used to refer to terminological or lexical databases, ontologies, and any graph-like KR.

- KBs are essential for AI tasks such as reasoning, inference or semantic search. Also for Word Sense Disambiguation, Entity Linking, Machine Translation, Semantics…

- They may be constructed manually in specific domains (e.g. *Chemistry*), but the general preference is to learn them (semi) automatically.
Knowledge Bases

· Hand-crafted KBs
  → From generic to domain-specific. E.g. WordNet, CheBi, SnomedCT.

· Integrative Projects
  → Unify in one single resource manually curated KRs and KBs.
    ⇒ BabelNet (originally, WordNet + Wikipedia), DBPedia, Yago…

· Open Information Extraction for KB construction
  → NELL, PATTY, WiseNet, DefIE, KB-Unify…
Music Knowledge Bases

- **MusicBrainz and Discogs**
  - Open encyclopedias of music metadata
  - MB is regularly published as Linked Data by the LinkedBrainz project.

- **Grove Music Online**
  - Music *scholar* encyclopedia

- **Flamenco MKB**
Tools

Alchemy API
http://www.alchemyapi.com/products/alchemylanguage/entity-extraction

AYLIEN API http://aylien.com/text-api


Gensim python library https://radimrehurek.com/gensim/

Senti WordNet http://sentiwordnet.isti.cnr.it/
Software

Standalone

- OpenNLP: https://opennlp.apache.org/
- Stanford CoreNLP: http://stanfordnlp.github.io/CoreNLP/
- Freeling: http://nlp.lsi.upc.edu/freeling/node/1
- Gate: https://gate.ac.uk/
- Mate Parser: http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html

Python Libraries

- Spacy: https://spacy.io
- Pattern: http://www.clips.ua.ac.be/pattern
- NLTK: http://www.nltk.org/
- Gensim: https://radimrehurek.com/gensim/
- Rake: https://www.airpair.com/nlp.keyword-extraction-tutorial
Software

ML toolkits/libraries widely used in NLP

- CRF++: https://taku910.github.io/crfpp/
- Mallet: http://mallet.cs.umass.edu/
- Networkx: https://networkx.github.io
- Weka: http://www.cs.waikato.ac.nz/ml/weka/

Deep Learning:

- Keras https://keras.io/
- Tflearn http://tflearn.org/
- Tensorflow https://www.tensorflow.org/
- Theano http://deeplearning.net/software/theano/
- DyNet (formerly cnn) https://github.com/clab/dynet
References - NLP


References - KBs


**Discogs**: [www.discogs.com](http://www.discogs.com)

References


Outline

- Introduction to NLP
- **Information Extraction**
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Information Extraction
Information Extraction

Information extraction (IE) is the task of automatically extracting **structured** information from **unstructured** and/or semi-structured machine-readable documents.
“Hate It Here” was written by Wilco frontman, Jeff Tweedy.
Information Extraction

Entity Identification

“Hate It Here” was written by Wilco frontman, Jeff Tweedy.
"Hate It Here" was written by Wilco frontman, Jeff Tweedy.
Information Extraction

Wilco (disambiguation)

From Wikipedia, the free encyclopedia

Wilco is an American rock band.

Wilco may also refer to:

- Wilco (voice procedure), a radio procedure word, short for "Will Comply"; origin of the term
- Wilco (The Album), an album by the band Wilco, or the title song, "Wilco (The Song)"
- Wilco: Learning How to Die, a book about the band, by Greg Kot
- Wilco (farm supply cooperative), an American chain of agricultural cooperative stores
- Wilco (tree), *Anadenanthera colubrina*, a South American tree
- Wilkinson County, Georgia, sometimes abbreviated as "Wilco"
- Williamson County, Texas, sometimes abbreviated as "Wilco"
- WilcoHess, the chain of gas stations
“Hate It Here” was written by Wilco frontman, Jeff Tweedy.
“Hate It Here” was written by Wilco frontman, Jeff Tweedy.

Hate It Here was written by Jeff Tweedy

Jeff Tweedy is the frontman of Wilco.
Information Extraction

Relation Extraction

Unstructured

“Hate It Here” was written by Wilco frontman, Jeff Tweedy.

Structured

Hate It Here \(\xrightarrow{\text{was written by}}\) Jeff Tweedy

Jeff Tweedy \(\xrightarrow{\text{frontman}}\) Wilco
Entity Linking

Entity linking is the task to associate, for a given candidate textual fragment, the most suitable entry in a reference Knowledge Base.

- Also referred to as Entity Disambiguation
- Typically Wikipedia, DBpedia, YAGO, Freebase as reference KB
Entity Linking

Entity linking is the task to associate, for a given candidate textual fragment, the most suitable entry in a reference Knowledge Base.

- Also referred to as Entity Disambiguation
- Typically Wikipedia, DBpedia, YAGO, Freebase as reference KB

Entity linking is typically broken down into two main phases:

- Candidate selection
- Reference disambiguation
Entity Linking

The entity linking system can either return:

- Matching entry (e.g. DBpedia URI, Wikipedia URL)
- NIL (no matching in the Knowledge Base)

But most of the systems make the closed world assumption, i.e. there is always a target entity in the knowledge base.
Entity Linking

Entity linking needs to handle:

- **Name variations** (entities are referred to in many different ways)
  - e.g. Elvis, Elvis Presley, Elvis Aaron Presley, The King of Rock and Roll

- **Entity ambiguity** (the same string can refer to more than one entity)
  - e.g. Prince, Debut, Bach, Strauss

- **Missing entities** (there is no target entity in the knowledge base)
  - e.g. Supertrópica is not in Wikipedia
Entity Linking

Entity linking needs to handle:

- **Name variations** (entities are referred to in many different ways)
  - e.g. Elvis, Elvis Presley, Elvis Aaron Presley

- **Entity ambiguity** (the same string can refer to more than one entity)
  - e.g. Prince, Debut, Bach, Strauss

- **Missing entities** (there is no target entity in the knowledge base)
  - e.g. Supertrópica is not in Wikipedia
Entity Linking: Tools


**DBpedia Spotlight**: Installable web service. KB: DBpedia. [https://github.com/dbpedia-spotlight/dbpedia-spotlight](https://github.com/dbpedia-spotlight/dbpedia-spotlight)
Relation Extraction

Detection and classification of semantic relations within a set of artifacts (e.g. entities, noun phrases) from text.

Numerous variants:

- Supervision: {fully, un, semi, distant}-supervision
- Undefined vs. pre-determined set of relations
- Binary vs. n-ary relations
Relation Extraction

Typical features:

- morphologic, syntactic, semantic, statistical
- context words + part-of-speech tags, dependency paths, named entities
Relation Extraction

- **Input:**
  - Large corpus of unstructured text
  - Set of semantic relations
  - Labelled training data

- **Output:**
  - Knowledge Base of triples
  - \( \langle \text{entity, relation, entity} \rangle \)
Relation Extraction

- **Input:**
  - Large corpus of unstructured text
  - Set of semantic relations
  - High-precision seeds/examples

- **Output:**
  - Knowledge Base of triples
  - \( \langle \text{entity}, \text{relation}, \text{entity} \rangle \)
Relation Extraction

- **Input:**
  - Large corpus of unstructured text
  - Set of semantic relations
  - Labelled training data

- **Output:**
  - Knowledge Base of triples
  - \( \langle \text{entity, relation, entity} \rangle \)
  - Set of semantic relations
Relation Extraction

- Degree of supervision
- Semantic information
- Semi-supervised
- Supervised
- Underlying KB
- NER categories
Relation Extraction

Traditional IE
Relation Extraction

- **Open IE (ReVerb, TextRunner)**
- **Self Supervision (NELL)**
- **Distant Supervision (DeepDive)**
- **Weak Supervision**
- **Traditional IE**

Degree of supervision:
- **semi-supervised**
- **supervised**

Semantic information:
- **underlying KB**
- **NER categories**
Relation Extraction

- **Semantic OIE** (Patty, DefIE, KBSF)
- **Open IE (ReVerb, TextRunner)**
- **Self Supervision** (NELL)
- **Weak Supervision** (DeepDive)
- **Distant Supervision** (DeepDive)
- **Traditional IE**

Degree of supervision vs. semantic information.

Information Extraction

NLP for MIR

Sergio Oramas
Relation Extraction

Further information in [http://wwwusers.di.uniroma1.it/~dellibovi/talks/talk_OIE.pdf](http://wwwusers.di.uniroma1.it/~dellibovi/talks/talk_OIE.pdf)
Semantic Open IE

Entity Linking + Open Information Extraction

Advantages

- **Not restricted** to a set of predefined relations
- **Unsupervised**: no need of training samples
- Use of semantic information reduces imprecision of Open IE
- Useful for KB construction and KB expansion (no need of mapping)

Semantic Open IE

- **Entity linking** -> Semantic Information
- **Dependency parsing** -> Syntactic Information
- **Semantic-Syntactic integration**
- **Shortest path** between entities
- **Filtering** of relations
`"Hate It Here" was written by Wilco frontman, Jeff Tweedy.
`` Hate It Here '' was written by Wilco frontman, Jeff Tweedy.
`` Hate It Here '' was written by Wilco frontman, Jeff Tweedy.
"Hate It Here" was written by Wilco frontman, Jeff Tweedy.
`` Hate It Here '' was written by Wilco frontman, Jeff Tweedy.
Relation Extraction (References)

Traditional IE

Weak Supervision

Self Supervision

Distant Supervision
Relation Extraction (References)

Open IE

Semantic Open IE
Relation Extraction (Tools)


Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Construction of Music KBS
Outline

· Motivation

· The Challenge of EL in the Music domain
  → ELMD and ELVIS

· Towards MKB Learning from Scratch
Motivation - Why you should care

- Structuring information in the Information Age is the big thing.
- Making sense of what people say about music has the potential to contribute dramatically to musicology and MIR.
  - Obtain knowledge automatically
  - Ask complex questions
  - Information Visualization
  - Improve navigation and personalization
Motivation - Why you should care

- Structured information about music is incomplete
- (almost) Only popular artists and western music
- (almost) Only editorial and some biographical information
Motivation - Why you should care

- Huge amount of music information remains implicit in unstructured texts

* Artists biographies, articles, reviews, web pages, user posts.
Motivation - Why you should care

- Huge amount of music information remains implicit in unstructured texts

* Artists biographies, articles, reviews, web pages, user posts.
Motivation - Why you should care

- Huge amount of music information remains implicit in unstructured texts
  * Artists biographies, articles, reviews, web pages, user posts.
Challenges - Entity Linking

- **Entity Recognition.**

- Typical procedure: Gazetteers or knowledge repositories with musical information.

  - Efficient in idiosyncratic and unambiguous cases: *The Symphony No. 9 in D minor*.
  - **But what if there is variation?** For example, *The 9th is one of Beethoven’s best.*
  - One same mention may refer to different musical entities. E.g. *Carmen* the opera, and *Carmen* the opera’s main character.
  - Variability in musical entities. E.g. *The Rolling Stones* or *Their Satanic Majesties*.
  - Musical entities with common names.
    - E.g. *Madonna* (artist or representation of Mary)
Challenges - Entity Linking

- Album and especially artist names get shortened in casual language.
- Album and artist names being the same.
- Generic software for Entity Linking don’t do well. Lack of sensitivity to musical text. Also, most of them exploit context, but this can be counterproductive.
## Challenges - Entity Linking

<table>
<thead>
<tr>
<th>System</th>
<th>Song</th>
<th>Album</th>
<th>Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Babelify</strong></td>
<td>Carey</td>
<td>Debut</td>
<td>John_Lennon</td>
</tr>
<tr>
<td></td>
<td>Stephen</td>
<td>Song_For</td>
<td>Eminem</td>
</tr>
<tr>
<td></td>
<td>Rap_Song</td>
<td>Song_Of</td>
<td>Paul_McCartney</td>
</tr>
<tr>
<td><strong>Tagme</strong></td>
<td>The_Word</td>
<td>Up</td>
<td>John_Lennon</td>
</tr>
<tr>
<td></td>
<td>The_End</td>
<td>When_We_On</td>
<td>Do</td>
</tr>
<tr>
<td></td>
<td>If</td>
<td>Together</td>
<td>Neil_Young</td>
</tr>
<tr>
<td><strong>DBpedia Spotlight</strong></td>
<td>Sexy_Sadie</td>
<td>The_Wall</td>
<td>Madonna</td>
</tr>
<tr>
<td></td>
<td>Helter_Skelter</td>
<td>Let_It_Be</td>
<td>Eminem</td>
</tr>
<tr>
<td></td>
<td>Cleveland_Rocks</td>
<td>Born_This_Way</td>
<td>Rihanna</td>
</tr>
</tbody>
</table>
ELMD: Entity Linking in the Music Domain

ELMD: Entity Linking in the Music Domain

- We envisioned a text corpus annotated with a vast number of music entities (Album, Song, Artist and Record Label).

- While not all occurrences in text would be annotated, those who were should have very high Precision. Good for propagation, semi supervised learning, etc.

- We took advantage of artist biographies in lost.fm

- And annotated dozens of thousands of entities with very high precision thanks to ELVIS!
ELMD: Entity Linking in the Music Domain

- We envisioned a text corpus annotated with a vast number of music entities (Album, Song, Artist and Record Label).
- While not all occurrences in text would be annotated, those who were should have very high Precision. Good for propagation, semi supervised learning, etc.
- We took advantage of artist biographies in last.fm
- And annotated dozens of thousands of entities with very high Precision thanks to ELVIS!
ELVIS: Entity Linking Voting and Integration System

· Assume agreement among generic tools can be leveraged to detect entities with *high precision*.

https://github.com/sergiooramas/elvis
ELMD: Entity Linking in the Music Domain

**Dataset**

- 13k artist biographies
- Collaborative effort
- Biographies are connected via 92,930 inner hyperlinks

**ELMD: Entity Linking in the Music Domain**

- From hyperlinks to annotated named entities
- Entities are then linked to DBpedia using ELVIS with 97% of precision
## ELMD: Entity Linking in the Music Domain

<table>
<thead>
<tr>
<th></th>
<th>ELVIS Score</th>
<th>Precision</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>type-equivalent</strong></td>
<td>= 3</td>
<td>0.97</td>
<td>31,180</td>
</tr>
<tr>
<td></td>
<td>&gt;= 2</td>
<td>0.96</td>
<td>46,544</td>
</tr>
<tr>
<td></td>
<td>&gt;= 1</td>
<td>0.94</td>
<td>59,680</td>
</tr>
<tr>
<td><strong>all</strong></td>
<td>= 3</td>
<td>0.94</td>
<td>33,455</td>
</tr>
<tr>
<td></td>
<td>&gt;= 2</td>
<td>0.90</td>
<td>51,802</td>
</tr>
<tr>
<td></td>
<td>&gt;= 1</td>
<td>0.81</td>
<td>72,365</td>
</tr>
</tbody>
</table>
ELMD 2.0: Bigger and Better

- Novel entity disambiguation mapping to MusicBrainz.
- Existing annotations are heuristically propagated.
- Different output formats: JSON, XML GATE, NIF.
- 144,593 Annotations and 63,902 Entities.
- Full details and download available at: http://mtg.upf.edu/download/datasets/elmd
Towards MKB Learning from Scratch

Towards MKB Learning from Scratch

· Starting from songfacts.com as a source for raw musical text, and after performing entity linking...

· The task lies now on how to leverage this information as the cornerstone of a music knowledge graph, the backbone of an MKB.

· The approach: Combine linguistically motivated rules over syntactic dependencies along with statistical evidence.
Towards MKB Learning from Scratch

- Shortest path doesn’t always work
  → **Nile Rodgers** *told* NME that the first album he bought was 300 Impressions by **John Coltrane**.
  ⇒ **nile_rodgers** told that was impressions by **john_coltrane**

- Consider special cases of:
  * Reported speech (“say”, “tell”, “express”)
  * Enforce certain syntactic relations between entity and first relation word.
  * etc
## Towards MKB Learning from Scratch

- **Relation Clustering: Syntactic Dependencies + Type Filtering**

<table>
<thead>
<tr>
<th>Cluster Pattern</th>
<th>Typed cluster pattern</th>
<th>Relation triple</th>
</tr>
</thead>
<tbody>
<tr>
<td>was written by artist</td>
<td>song was written by artist</td>
<td><em>song was written by artist</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*song was written by composer *artist</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*song was written by *artist</td>
</tr>
<tr>
<td>album was written by</td>
<td></td>
<td>*album was written by frontman *artist</td>
</tr>
<tr>
<td>artist</td>
<td></td>
<td>*album was written by guitarist *artist</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*album was written by *artist</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*album was written by newcomer *artist</td>
</tr>
</tbody>
</table>
Towards MKB Learning from Scratch

· Relation Scoring

· The relevance of a cluster may be inferred by the number and proportion of triples it encodes, and whether these are evenly distributed.

· Degree of specificity. ⇒ <artist$_d$, performed_with, artist$_r$>

· Frequency, length and fluency. Reward those relations which preserve the original sentence’ word order.
Towards MKB Learning from Scratch
Towards MKB Learning from Scratch

Our most sophisticated KB extracts novel information in the form of triples for the same pair of entities in other KBs.

Our KB: **3633** vs. MB: 1535, DBpedia: 1240, DeflE: 456.
Towards MKB Learning from Scratch
Towards MKB Learning from Scratch

- Bruce Springsteen covered Jersey Girl
Towards MKB Learning from Scratch

- Bruce Springsteen *covered* Jersey Girl

- Bruce Springsteen *player* Clarence Clemons
Towards MKB Learning from Scratch

- Bruce Springsteen *covered* Jersey Girl
- Bruce Springsteen *player* Clarence Clemons
- Hair (Lady Gaga) *features* Clarence Clemons
Towards MKB Learning from Scratch

- Bruce Springsteen covered *Jersey Girl*
- Bruce Springsteen player *Clarence Clemons*
- Hair (Lady Gaga) features *Clarence Clemons*
Towards MKB Learning from Scratch

· Conclusion

· Lots of unstructured information about music in the form of natural language

· We have barely scratched the surface. No Social Networks, no Wikipedia, no lyrics, no subtitles...

· Potential for improving MIR and musicological resources by integrating automatically acquired knowledge via Natural Language Processing.
References


Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - **Semantic Enrichment of Musical Texts**
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Semantic Enrichment of Musical Texts
Semantic Enrichment of Musical Texts

**Approach**: Create a **Knowledge Graph** and then apply graph-based methodologies or linear embeddings.

Several **types of graphs**:

- Knowledge Base Graph
- Graph of Entities
- Semantically Enriched Graph
Knowledge Graphs
Knowledge Graphs
Knowledge Base Graph

**Wilco**
dbo:bandMember -> dbr:Jeff_Tweedy
dbo:genre -> dbr:Alternative_country
dbo:hometown -> dbr:Illinois

**Son Volt**
dbo:genre -> dbr:Alternative_country
dbo:hometown -> dbr:St._Louis,_Missouri
dbo:recordLabel -> dbr:Warner_Bros._Records
Knowledge Base Graph

**Wilco**
- dbo:bandMember -> dbr:Jeff_Tweedy
- dbo:genre -> dbr:Alternative_country
- dbo:hometown -> dbr:Illinois

**Son Volt**
- dbo:genre -> dbr:Alternative_country
- dbo:hometown -> dbr:St._Louis,_Missouri
- dbo:recordLabel -> dbr:Warner_Bros._Records
Knowledge Base Graph

Wilco
- dbo:bandMember -> dbr:Jeff_Tweedy
- dbo:genre -> dbr:Alternative_country
- dbo:hometown -> dbr:Illinois

Son Volt
- dbo:genre -> dbr:Alternative_country
- dbo:hometown -> dbr:St._Louis,_Missouri
- dbo:recordLabel -> dbr:Warner_Bros._Records

Jeff Tweedy
- bandMember

Wilco
- genre
- hometown

Illinois

Alternative country

Warner_Bros. Records
- recordLabel

Son Volt
- genre
- hometown

St. Louis, Missouri
Knowledge Graphs

KB Construction

KB

Query

Knowledge Base Graph

Semantic Enrichment | NLP for MIR | Sergio Oramas
Knowledge Graphs

Entity Linking

Graph of Entities
Graph of Entities

**Wilco**
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.

**Son Volt**
It is an American alternative country group, formed by Jay Farrar in 1994.
Graph of Entities

Wilco
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.

Son Volt
It is an American alternative country group, formed by Jay Farrar in 1994.

Entity Linking
Wilco
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.

Son Volt
It is an American alternative country group, formed by Jay Farrar in 1994.
**Wilco**
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.

**Son Volt**
It is an American alternative country group, formed by Jay Farrar in 1994.
Knowledge Graphs

- Entity Linking
- Knowledge Base Graph
- Query
Knowledge Graphs

- Entity Linking
- Knowledge Base Graph
- Graph of Entities
- Semantically Enriched Graph

Semantic Enrichment

NLP for MIR

Sergio Oramas
Semantically Enriched Graph

**Wilco**
This *alternative rock* band was formed in 1994 by the remaining members of *Uncle Tupelo* following singer *Jay Farrar’s* departure.

**Son Volt**
It is an *American alternative country* group, formed by *Jay Farrar* in 1994.

(Graph of Entities)
Wilco
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.

Son Volt
It is an American alternative country group, formed by Jay Farrar in 1994.

Uncle Tupelo
dbo:hometown -> dbr:Belleville,_Illinois
dbo:genre -> dbr:Alternative_country

Jay Farrar
dbo:formerBandMember -> dbr:Uncle_Tupelo
dbp:birthPlace -> dbr:Belleville,_Illinois

American

Son Volt

Alternative Country

DBpedia
Wilco
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.

Son Volt
It is an American alternative country group, formed by Jay Farrar in 1994.
Semantically Enriched Graph
Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- **Applications in MIR**
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Applications in MIR
Applications

Similarity

Classification

Recommendation

Applications in MIR

NLP for MIR

Sergio Oramas
Similarity

\[ \begin{array}{cccc}
  f_1 & f_2 & \ldots & f_m \\
  i_1 \\
  i_2 \\
  \vdots \\
  i_n \\
\end{array} \quad \leftrightarrow \quad \begin{array}{cccc}
  i_1 & i_2 & \ldots & i_n \\
  i_1 \\
  i_2 \\
  \vdots \\
  i_n \\
\end{array} \]
Classification
Recommendation
Applications in MIR

NLP for MIR

Sergio Oramas

Items

Items: artist, song, sound, album

item = document
Typical Document-based approach

Vector Space Model

BoW

tf-idf
Graph Embedding
h-hop Item Neighborhood Graph
h-hop Item Neighborhood Graph
h-hop Item Neighborhood Graph
Embedding parameters

Distance to the root node

Frequency of the node inside the subgraph

Tf-idf of the node

Number of in and out links

Paths: sequences of nodes from the root
Flat Embedding

- Select $h$ for the $h$-hop subgraphs
- Create a bag-of-nodes binary vector for each subgraph
Entity-based Embedding

- One feature per entity

- Weight according to:
  - Distance to root
  - Number of in-links
Entity-based Embedding

- One feature per entity
- Weight according to:
  - Distance to root
  - Number of in-links
Path-based Embedding

- **Path**: sequence of entities
- Each **feature** refers to several variants of paths rooted in the item node
Path-based Embedding

- **Path**: sequence of entities
- Each **feature** refers to several variants of paths rooted in the item node
Path-based Embedding

- **Path**: sequence of entities
- Each **feature** refers to several variants of paths rooted in the item node
Path-based Embedding

- **Path**: sequence of entities
- Each **feature** refers to several variants of paths rooted in the item node
Path-based Embedding

- **Path**: sequence of entities
- Each **feature** refers to several variants of paths rooted in the item node
Path-based Embedding

- **Path**: sequence of entities
- Each *feature* refers to several variants of paths rooted in the item node
Path-based Embedding

- **Path**: sequence of entities
- Each **feature** refers to several variants of paths rooted in the item node
Artist Similarity
Artist Similarity

Maximal Common Subgraph

\[ d(G_1, G_2) = 1 - \frac{|mcs(G_1, G_2)|}{max(|G_1|, |G_2|)} \]
Artist Similarity


- Artist biographies gathered from Last.fm
- Entity Linking tool used: Babelfy
- Build different knowledge graphs
- Two Experiments:
  - MIREX: 188 artists, MIREX Audio and Music Similarity evaluation dataset
  - Last.fm API: 2,336 artists, Last.fm API similarity

SAS dataset: [http://mtg.upf.edu/download/datasets/semantic-similarity](http://mtg.upf.edu/download/datasets/semantic-similarity)
**Artist Similarity**

*Gorillaz* are a **british virtual band** formed in 1998 by *Damon Albarn* of *Blur*, and *Jamie Hewlett*, co-creator of the comic book *Tank Girl*.
## Artist Similarity

<table>
<thead>
<tr>
<th>Approach</th>
<th>P@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-based approach (BoW)</td>
<td>0.090</td>
</tr>
<tr>
<td>Extracted KB Graph</td>
<td>0.055</td>
</tr>
<tr>
<td>Graph of Entities</td>
<td>0.136</td>
</tr>
<tr>
<td>Semantically Enriched Graph</td>
<td>0.160</td>
</tr>
</tbody>
</table>
Genre Classification
Genre Classification

MARD (Multimodal Album Reviews Dataset):

New dataset of album customer reviews from:

Amazon + MusicBrainz + AcousticBrainz

Genre Classification

Features:

- **Textual**: BoW uni-grams and bi-grams
- **Semantic**: Entities and Wikipedia categories (Entity Linking), flat embedding
- **Sentiment**: positiveness ratio, emotion ratio, average emotion strength
- **Acoustic**: low-level descriptors (loudness, dynamics, spectral shape, etc.)

SVM classifier 5-fold cross validation, 1300 albums, 13 genres

https://github.com/sergiooramas/music-genre-classification
Genre Classification

Accuracy %

- **BoW+SEM+SENT**: 67.62%
- **BoW+SEMb**: 68.08%
- **BoW+SEM**: 69.08%
- **BoW+SENT**: 63.39%
- **BoW**: 62.92%
- **NB**: 55.0%
- **AB**: 38.69%
Genre Classification

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+SEM+SENT</td>
<td>67.62</td>
</tr>
<tr>
<td>BoW+SEMb</td>
<td>68.08</td>
</tr>
<tr>
<td>BoW+SEM</td>
<td>69.08</td>
</tr>
<tr>
<td>BoW+SENT</td>
<td>63.39</td>
</tr>
<tr>
<td>BoW</td>
<td>62.92</td>
</tr>
<tr>
<td>NB</td>
<td>55.0</td>
</tr>
<tr>
<td>AB</td>
<td><strong>38.69</strong></td>
</tr>
</tbody>
</table>
Genre Classification

Accuracy %

- BoW+SEM+SENT: 67.62%
- BoW+SEMb: 68.08%
- BoW+SEM: 69.08%
- BoW+SENT: 63.39%
- BoW: 62.92%
- NB: 55.0%
- AB: 38.69%

Applications in MIR  NLP for MIR  Sergio Oramas
Genre Classification

![Accuracy Chart]

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+SEM+SENT</td>
<td>67.62</td>
</tr>
<tr>
<td>BoW+SEMb</td>
<td>68.08</td>
</tr>
<tr>
<td>BoW+SEM</td>
<td>69.08</td>
</tr>
<tr>
<td>BoW+SENT</td>
<td>63.39</td>
</tr>
<tr>
<td>BoW</td>
<td>62.92</td>
</tr>
<tr>
<td>NB</td>
<td>55.0</td>
</tr>
<tr>
<td>AB</td>
<td>38.69</td>
</tr>
</tbody>
</table>
# Genre Classification

## Audio / Text

<table>
<thead>
<tr>
<th></th>
<th>Alt. Rock</th>
<th>Classical</th>
<th>Country</th>
<th>Electronic</th>
<th>Folk</th>
<th>Jazz</th>
<th>Latin</th>
<th>Metal</th>
<th>New Age</th>
<th>Pop</th>
<th>R&amp;B</th>
<th>Hip-Hop</th>
<th>Rock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alt. Rock</td>
<td>28 / 42</td>
<td>1 / 3</td>
<td>3 / 1</td>
<td>10 / 10</td>
<td>7 / 1</td>
<td>1 / 2</td>
<td>2 / 0</td>
<td>18 / 12</td>
<td>10 / 2</td>
<td>4 / 10</td>
<td>3 / 6</td>
<td>3 / 2</td>
<td>10 / 9</td>
</tr>
<tr>
<td>Classical</td>
<td>0 / 0</td>
<td>87 / 95</td>
<td>1 / 0</td>
<td>0 / 0</td>
<td>1 / 1</td>
<td>1 / 1</td>
<td>2 / 2</td>
<td>1 / 0</td>
<td>5 / 1</td>
<td>1 / 0</td>
<td>0 / 0</td>
<td>0 / 0</td>
<td>1 / 0</td>
</tr>
<tr>
<td>Country</td>
<td>2 / 1</td>
<td>0 / 0</td>
<td>51 / 84</td>
<td>3 / 0</td>
<td>9 / 1</td>
<td>9 / 0</td>
<td>3 / 0</td>
<td>0 / 1</td>
<td>3 / 0</td>
<td>8 / 8</td>
<td>6 / 4</td>
<td>1 / 0</td>
<td>5 / 1</td>
</tr>
<tr>
<td>Electronic</td>
<td>7 / 3</td>
<td>3 / 1</td>
<td>1 / 2</td>
<td>40 / 61</td>
<td>4 / 1</td>
<td>1 / 2</td>
<td>2 / 2</td>
<td>6 / 0</td>
<td>7 / 5</td>
<td>6 / 5</td>
<td>6 / 7</td>
<td>13 / 5</td>
<td>4 / 7</td>
</tr>
<tr>
<td>Folk</td>
<td>4 / 6</td>
<td>11 / 0</td>
<td>13 / 10</td>
<td>7 / 0</td>
<td>27 / 55</td>
<td>6 / 1</td>
<td>7 / 3</td>
<td>4 / 2</td>
<td>6 / 9</td>
<td>5 / 9</td>
<td>6 / 4</td>
<td>1 / 0</td>
<td>3 / 1</td>
</tr>
<tr>
<td>Jazz</td>
<td>7 / 0</td>
<td>10 / 1</td>
<td>6 / 2</td>
<td>2 / 2</td>
<td>5 / 0</td>
<td>45 / 82</td>
<td>6 / 3</td>
<td>3 / 0</td>
<td>8 / 2</td>
<td>3 / 5</td>
<td>4 / 1</td>
<td>1 / 1</td>
<td>0 / 1</td>
</tr>
<tr>
<td>Latin</td>
<td>4 / 3</td>
<td>6 / 4</td>
<td>9 / 2</td>
<td>1 / 2</td>
<td>5 / 1</td>
<td>10 / 2</td>
<td>28 / 78</td>
<td>3 / 0</td>
<td>6 / 2</td>
<td>11 / 4</td>
<td>7 / 2</td>
<td>5 / 0</td>
<td>5 / 0</td>
</tr>
<tr>
<td>Metal</td>
<td>13 / 5</td>
<td>1 / 0</td>
<td>1 / 1</td>
<td>2 / 2</td>
<td>1 / 0</td>
<td>0 / 1</td>
<td>1 / 0</td>
<td>63 / 87</td>
<td>1 / 0</td>
<td>1 / 0</td>
<td>3 / 1</td>
<td>1 / 0</td>
<td>12 / 3</td>
</tr>
<tr>
<td>New Age</td>
<td>9 / 2</td>
<td>7 / 6</td>
<td>9 / 0</td>
<td>7 / 4</td>
<td>10 / 10</td>
<td>9 / 2</td>
<td>7 / 6</td>
<td>3 / 3</td>
<td>15 / 53</td>
<td>10 / 7</td>
<td>6 / 1</td>
<td>2 / 1</td>
<td>6 / 5</td>
</tr>
<tr>
<td>Pop</td>
<td>6 / 2</td>
<td>9 / 1</td>
<td>10 / 2</td>
<td>9 / 2</td>
<td>5 / 3</td>
<td>9 / 2</td>
<td>5 / 2</td>
<td>2 / 0</td>
<td>7 / 1</td>
<td>19 / 73</td>
<td>7 / 6</td>
<td>2 / 2</td>
<td>10 / 5</td>
</tr>
<tr>
<td>R&amp;B</td>
<td>8 / 2</td>
<td>0 / 1</td>
<td>16 / 3</td>
<td>8 / 4</td>
<td>2 / 0</td>
<td>5 / 3</td>
<td>5 / 0</td>
<td>1 / 0</td>
<td>3 / 0</td>
<td>7 / 10</td>
<td>24 / 71</td>
<td>17 / 5</td>
<td>4 / 1</td>
</tr>
<tr>
<td>Hip-Hop</td>
<td>8 / 2</td>
<td>0 / 0</td>
<td>2 / 1</td>
<td>8 / 2</td>
<td>0 / 1</td>
<td>0 / 1</td>
<td>1 / 0</td>
<td>4 / 3</td>
<td>2 / 0</td>
<td>4 / 1</td>
<td>7 / 2</td>
<td>61 / 86</td>
<td>3 / 1</td>
</tr>
<tr>
<td>Rock</td>
<td>17 / 15</td>
<td>1 / 2</td>
<td>6 / 8</td>
<td>4 / 7</td>
<td>10 / 5</td>
<td>2 / 4</td>
<td>7 / 1</td>
<td>12 / 13</td>
<td>4 / 1</td>
<td>9 / 7</td>
<td>7 / 4</td>
<td>6 / 2</td>
<td>15 / 31</td>
</tr>
</tbody>
</table>
Music Recommendation
Music Recommendation

Recommendation approaches:

- **Collaborative filtering** - only users matrix
- **Content-based** - only item features matrix
- **Hybrid** - both matrices
Music Recommendation

**Hybrid approach**: Aggregation of features

| Item vector | Knowledge Graph vector | Collaborative vector |

Train a regression model on every user


Source code: [https://github.com/sisinflab/lodreclib](https://github.com/sisinflab/lodreclib)
Music Recommendation

Two **experiments:**

- **Sounds Recommendation**
  - Freesound **tags** and **descriptions** + **Implicit feedback** (downloads)
  - 21,552 items and 20,000 users

- **Music Recommendation**
  - Last.fm **tags** and Songfacts **descriptions** + **Implicit feedback** (Last.fm listening habits)
  - 8,640 items and 5,199 users

Datasets: [http://mtg.upf.edu/download/datasets/knowledge-graph-rec](http://mtg.upf.edu/download/datasets/knowledge-graph-rec)
Music Recommendation

Knowledge Graph approach

- **Semantically Enriched Graph** over tags and text descriptions
- Using Babelfy for **Entity Linking**
- Using **Wikipedia** categories and **WordNet** synsets and hypernymy relations for **semantic expansion**

<table>
<thead>
<tr>
<th>dataset</th>
<th>items</th>
<th>avg. tags</th>
<th>avg. keywords</th>
<th>resources</th>
<th>synsets</th>
<th>categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freesound</td>
<td>21,552</td>
<td>6.44</td>
<td>11.36</td>
<td>16,407</td>
<td>20,034</td>
<td>54,419</td>
</tr>
<tr>
<td>Last.fm</td>
<td>8,640</td>
<td>42.09</td>
<td>77.33</td>
<td>46,109</td>
<td>27,708</td>
<td>96,942</td>
</tr>
</tbody>
</table>
## Music Recommendation

<table>
<thead>
<tr>
<th>KG features</th>
<th>Collab features</th>
<th>P@10</th>
<th>R@10</th>
<th>EBN@10</th>
<th>ADiv@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-based</td>
<td>si</td>
<td>0.118</td>
<td>0.067</td>
<td>2.426</td>
<td>0.361</td>
</tr>
<tr>
<td>Path-based</td>
<td>si</td>
<td>0.111</td>
<td>0.061</td>
<td>1.618</td>
<td>0.532</td>
</tr>
<tr>
<td>Path-based</td>
<td>no</td>
<td>0.049</td>
<td>0.028</td>
<td>0.369</td>
<td>0.670</td>
</tr>
<tr>
<td>-</td>
<td>si</td>
<td>0.110</td>
<td>0.062</td>
<td>2.890</td>
<td>0.181</td>
</tr>
<tr>
<td>VSM</td>
<td>si</td>
<td>0.116</td>
<td>0.066</td>
<td>2.621</td>
<td>0.305</td>
</tr>
<tr>
<td>Audio Sim</td>
<td>no</td>
<td>0.004</td>
<td>0.002</td>
<td>0.382</td>
<td>0.044</td>
</tr>
</tbody>
</table>

EBN: Entropy-based Novelty  
ADiv: Aggregated Diversity
## Music Recommendation

<table>
<thead>
<tr>
<th>KG features</th>
<th>Collab features</th>
<th>P@10</th>
<th>R@10</th>
<th>EBN@10</th>
<th>ADiv@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-based</td>
<td>si</td>
<td>0.118</td>
<td>0.067</td>
<td>2.426</td>
<td>0.361</td>
</tr>
<tr>
<td>Path-based</td>
<td>si</td>
<td>0.111</td>
<td>0.061</td>
<td>1.618</td>
<td>0.532</td>
</tr>
<tr>
<td>Path-based</td>
<td>no</td>
<td>0.049</td>
<td>0.028</td>
<td>0.369</td>
<td>0.670</td>
</tr>
<tr>
<td>-</td>
<td>si</td>
<td>0.110</td>
<td>0.062</td>
<td>2.890</td>
<td>0.181</td>
</tr>
<tr>
<td>VSM</td>
<td>si</td>
<td>0.116</td>
<td>0.066</td>
<td>2.621</td>
<td>0.305</td>
</tr>
<tr>
<td>Audio Sim</td>
<td>no</td>
<td>0.004</td>
<td>0.002</td>
<td>0.382</td>
<td>0.044</td>
</tr>
</tbody>
</table>

EBN: Entropy-based Novelty  
ADiv: Aggregated Diversity
## Music Recommendation

<table>
<thead>
<tr>
<th>KG features</th>
<th>Collab features</th>
<th>P@10</th>
<th>R@10</th>
<th>EBN@10</th>
<th>ADiv@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-based</td>
<td>si</td>
<td>0.118</td>
<td>0.067</td>
<td>2.426</td>
<td>0.361</td>
</tr>
<tr>
<td>Path-based</td>
<td>si</td>
<td>0.111</td>
<td>0.061</td>
<td>1.618</td>
<td>0.532</td>
</tr>
<tr>
<td>Path-based</td>
<td>no</td>
<td>0.049</td>
<td>0.028</td>
<td>0.369</td>
<td>0.670</td>
</tr>
<tr>
<td>-</td>
<td>si</td>
<td>0.110</td>
<td>0.062</td>
<td>2.890</td>
<td>0.181</td>
</tr>
<tr>
<td>VSM</td>
<td>si</td>
<td>0.116</td>
<td>0.066</td>
<td>2.621</td>
<td>0.305</td>
</tr>
<tr>
<td>Audio Sim</td>
<td>no</td>
<td>0.004</td>
<td>0.002</td>
<td>0.382</td>
<td>0.044</td>
</tr>
</tbody>
</table>

EBN: Entropy-based Novelty  
ADiv: Aggregated Diversity
Music Recommendation

<table>
<thead>
<tr>
<th>KG features</th>
<th>Collab features</th>
<th>P@10</th>
<th>R@10</th>
<th>EBN@10</th>
<th>ADiv@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-based</td>
<td>si</td>
<td>0.118</td>
<td>0.067</td>
<td>2.426</td>
<td>0.361</td>
</tr>
<tr>
<td>Path-based</td>
<td>si</td>
<td>0.111</td>
<td>0.061</td>
<td>1.618</td>
<td>0.532</td>
</tr>
<tr>
<td>Path-based</td>
<td>no</td>
<td>0.049</td>
<td>0.028</td>
<td>0.369</td>
<td>0.670</td>
</tr>
<tr>
<td>-</td>
<td>si</td>
<td>0.110</td>
<td>0.062</td>
<td>2.890</td>
<td>0.181</td>
</tr>
<tr>
<td>VSM</td>
<td>si</td>
<td>0.116</td>
<td>0.066</td>
<td>2.621</td>
<td>0.305</td>
</tr>
<tr>
<td>Audio Sim</td>
<td>no</td>
<td>0.004</td>
<td>0.002</td>
<td>0.382</td>
<td>0.044</td>
</tr>
</tbody>
</table>

EBN: Entropy-based Novelty
ADiv: Aggregated Diversity
## Music Recommendation

<table>
<thead>
<tr>
<th>KG features</th>
<th>Collab features</th>
<th>P@10</th>
<th>R@10</th>
<th>EBN@10</th>
<th>ADiv@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-based</td>
<td>yes</td>
<td>0.118</td>
<td>0.067</td>
<td>2.426</td>
<td>0.361</td>
</tr>
<tr>
<td>Path-based</td>
<td>yes</td>
<td>0.111</td>
<td>0.061</td>
<td>1.618</td>
<td>0.532</td>
</tr>
<tr>
<td>Path-based</td>
<td>no</td>
<td>0.049</td>
<td>0.028</td>
<td>0.369</td>
<td>0.670</td>
</tr>
<tr>
<td>-</td>
<td>yes</td>
<td>0.110</td>
<td>0.062</td>
<td>2.890</td>
<td>0.181</td>
</tr>
<tr>
<td>VSM</td>
<td>yes</td>
<td>0.116</td>
<td>0.066</td>
<td>2.621</td>
<td>0.305</td>
</tr>
<tr>
<td>Audio Sim</td>
<td>no</td>
<td>0.004</td>
<td>0.002</td>
<td>0.382</td>
<td>0.044</td>
</tr>
</tbody>
</table>

EBN: Entropy-based Novelty  
ADiv: Aggregated Diversity
Music Recommendation (Conclusions)

Semantically Enriched Graph improves novelty and diversity → better explore the long tail

Combination with collaborative features ensures high accuracy

**Path-based embedding**: better novelty and diversity, slightly lower accuracy

**Entity-based embedding**: better accuracy, slightly lower novelty and diversity
Interpreting Music Recommendations

Building natural language **explanations** of the relation between two entities

- **Using labels of a Knowledge Graph**


- **Using sentence texts where entities co-occur**

Applications in MIR

NLP for MIR

Sergio Oramas

SONG #18

You Know My Name (Look Up The Number) (The Beatles)

SONG #10

Sanctuary (Iron Maiden)

RECOMMENDED SONG

Fourth Time Around (Bob Dylan)

You Know My Name (Look Up The Number) <-- The Beatles <-- Fourth Time Around

The Beatles started recording You Know My Name (Look Up The Number) in 1967, adding all the instrumentation and a saxophone part played by Brian Jones from The Rolling Stones. Fourth Time Around was written in response to "Norwegian Wood"-LRB- This Bird Has Flown -RRB-" by The Beatles, since it is similar, both melodically and lyrically.

Set The World Afire (Megadeth)

Sanctuary <-- Iron Maiden <-- Jump In The Fire (Metallica) <-- Dave Mustaine --> Set The World Afire

Iron Maiden version of Sanctuary
Jump In The Fire (Metallica) was inspired by Iron Maiden
Dave Mustaine helped write Jump In The Fire (Metallica)
Dave Mustaine started writing Set The World Afire

Give a score to the provided recommendation:
1 2 3 4 5

Did you know the recommended song?
Yes No
Interpreting Music Recommendations

Challenges

- Select the **best path** (many possible paths between 2 entities)
- Generate a natural language **explanation**
  - Use relation labels
  - Use sentence texts
Interpreting Music Recommendations

Interpreting Music Recommendations

User Experiment:

● 35 subjects
● 3 different recommendations
  ○ no explanation (3.08)
  ○ original sentences (3.20)
  ○ predicate labels (3.04)
● Higher differences in average ratings on musically untrained subjects
Other Applications

● Question & Answering


● Entity Retrieval / Semantic Search

http://edgar.meij.pro/entity-linking-retrieval-semantic-search-wsdm-2014/
References


Oramas S., Gómez F., Gómez E., &
References


Supplementary Material

Download supplementary material:

http://mtg.upf.edu/nlp-tutorial

Create a BabelNet account:

http://babelnet.org/register
Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Applications in Musicology
Musicology

Musicology embraces the study of history, theory, and practice of music from many points of view.

Musicology is part of the humanities

Musicologists have to read a lot!
Musicology

Musicology embraces the study of history, theory, and practice of music from many points of view.

Musicology is part of the humanities

Musicologists have to read a lot!
Applications in Musicology

Entity Relevance  Analytics  Information Visualization
Entity Relevance

See a Graph of Entities as network of hyperlinks

Use Pagerank or HITS to compute entity relevance

Wilco
This alternative rock band was formed in 1994 by the remaining members of Uncle Tupelo following singer Jay Farrar's departure.
FlaBase


1,174 Artists (text biography)
76 Palos (flamenco genres)
2,913 Albums
14,078 Tracks
771 Andalusian locations

We built a Graph of Entities
Artist Relevance

**Flamenco expert evaluation**

<table>
<thead>
<tr>
<th>Cantaor</th>
<th>Guitarist</th>
<th>Bailaor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antonio Mairena</td>
<td>Paco de Lucía</td>
<td>Antonio Ruiz Soler Rosario</td>
</tr>
<tr>
<td>Manolo Caracol</td>
<td>Ramón Montoya</td>
<td>Antonio Gades</td>
</tr>
<tr>
<td>La Niña de los Peines</td>
<td>Niño Ricardo</td>
<td>Mario Maya</td>
</tr>
<tr>
<td>Antonio Chacón</td>
<td>Manolo Sanlúcar Sabicas</td>
<td>Carmen Amaya</td>
</tr>
<tr>
<td>Camarón de la Isla</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>PageRank</td>
<td>0.933</td>
<td>0.633</td>
</tr>
<tr>
<td>HITS Authority</td>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Applications in Musicology

NLP for MIR

Sergio Oramas
Analytics

- Extract attributes, events, entity mentions, relations, sentiment, etc.
- Compute analytics

Useful insights for musicologists
Analytics: Grove Dictionary


- **Grove Dictionary**: one of the largest reference works on Western music
- 16,707 biographies were gathered from Grove Music Online
- **Extracted information**: roles, birth and death, entity mentions, relations
Analytics: Grove Dictionary

<table>
<thead>
<tr>
<th>Role</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>composer</td>
<td>2618</td>
</tr>
<tr>
<td>teacher</td>
<td>1065</td>
</tr>
<tr>
<td>conductor</td>
<td>968</td>
</tr>
<tr>
<td>pianist</td>
<td>704</td>
</tr>
<tr>
<td>organist</td>
<td>676</td>
</tr>
<tr>
<td>singer</td>
<td>404</td>
</tr>
<tr>
<td>violinist</td>
<td>285</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>musicologist</td>
<td>144</td>
</tr>
<tr>
<td>critic</td>
<td>133</td>
</tr>
</tbody>
</table>
## Analytics: Grove Dictionary

<table>
<thead>
<tr>
<th>Country</th>
<th>Births</th>
<th>Deaths</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>2317</td>
<td>2094</td>
<td>-10%</td>
</tr>
<tr>
<td>Italy</td>
<td>1616</td>
<td>1279</td>
<td>-21%</td>
</tr>
<tr>
<td>Germany</td>
<td>1270</td>
<td>1292</td>
<td>2%</td>
</tr>
<tr>
<td>France</td>
<td>991</td>
<td>1058</td>
<td>7%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>882</td>
<td>877</td>
<td>-1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>City</th>
<th>Births</th>
<th>Deaths</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>322</td>
<td>507</td>
<td>57%</td>
</tr>
<tr>
<td>Paris</td>
<td>304</td>
<td>720</td>
<td>137%</td>
</tr>
<tr>
<td>New York</td>
<td>266</td>
<td>501</td>
<td>88%</td>
</tr>
<tr>
<td>Vienna</td>
<td>177</td>
<td>292</td>
<td>65%</td>
</tr>
<tr>
<td>Rome</td>
<td>159</td>
<td>256</td>
<td>61%</td>
</tr>
</tbody>
</table>
## Analytics: Grove Dictionary

<table>
<thead>
<tr>
<th>City</th>
<th>Births</th>
<th>Deaths</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>322</td>
<td>507</td>
<td>57%</td>
</tr>
<tr>
<td>Paris</td>
<td>304</td>
<td>720</td>
<td>137%</td>
</tr>
<tr>
<td>New York</td>
<td>266</td>
<td>501</td>
<td>88%</td>
</tr>
<tr>
<td>Vienna</td>
<td>177</td>
<td>292</td>
<td>65%</td>
</tr>
<tr>
<td>Rome</td>
<td>159</td>
<td>256</td>
<td>61%</td>
</tr>
</tbody>
</table>
Analytics: Diachronic study of affective language

MARD Multimodal Album Reviews Dataset

- **Amazon** (~66k albums / ~250k customer reviews)
  - Album customer reviews
  - Genre tags (16 genres and 287 subgenres)
  - Star Ratings
  - Metadata: title, artist, record label
- **MusicBrainz**: ids, song titles, year of publication
- **AcousticBrainz**: audio descriptors of songs

MARD: [http://mtg.upf.edu/download/datasets/mard](http://mtg.upf.edu/download/datasets/mard)
Aspect-based Sentiment Analysis

Beautiful Drug has great guitar riffs but the vocals are shrill

- **Entities:** Beautiful Drug
- **Aspects (also called features):** guitar riffs, vocals
- **Opinion words:** great, shrill


Aspect-based Sentiment Analysis


Rule-based approach using a sentiment lexicon

- Identification of aspects: bi-grams and single-noun
- Identification of opinion words: adjectives
- Context rules: distance, POS tags and negations between opinion words and aspects
- Sentiment Lexicon: SentiWordNet (http://sentiwordnet.isti.cnr.it/)
Diachronic Study of Affective Language

**Sentiment score**: Average sentiment score of all aspects in a review

Two perspectives:

- **Average of all reviews by** review publication year **(2000-2014)**
  - Evolution of affective language from a customer perspective

- **Average of all reviews by** album publication year **(1950-2014)**
  - Evolution of affective language from a musical perspective
Study by **review** publication year

Average sentiment

Average rating
Study by **review** publication year

**Average sentiment**

![Graph showing average sentiment over time from 2000 to 2014.](image)

**USA - Real GDP Trend**

- **Q1 2000 - Q3 2014**
  - Trend growth rate: 2.7% (Q2 09) 2.3% (annual rate)
  - Trend calculated on Q1 2000 - Q3 2008 and extended
  - Billions of chained (2009) dollars

![Graph showing real GDP trend from 2000 to 2014.](image)

Source: Datastream, Nabhns AM - philippweichler.en.nabnhs.com
In November 2008, at least for a time, hope prevailed over fear. The wall of racial prejudice fell as surely as the wall of oppression had fallen in Berlin twenty years earlier [...] Yet the emotional dimension of this election and the sense of pride it created in many Americans must not be underestimated.

In November 2008, at least for a time, hope prevailed over fear. The wall of racial prejudice fell as surely as the wall of oppression had fallen in Berlin twenty years earlier [...] Yet the emotional dimension of this election and the sense of pride it created in many Americans must not be underestimated.


Average sentiment
Study by **review** publication year

**Average sentiment by genre**

- **Country**
- **Jazz**
- **Latin Music**

**Average sentiment by aspect**

- **album**
- **music**
- **song**
- **sound**

**Applications in Musicology**  **NLP for MIR**  **Sergio Oramas**
Study by review publication year

Further studies necessary to validate any of these suggestions

Correlation ≠ Causation

Interesting insight for Musicologists
Study by album publication year

Average sentiment

Average rating

Pearson’s correlation $r = 0.75$, $p \ll 0.001$
Study by **album** publication year

Average sentiment by genre (trend curve)
Study by album publication year

Average sentiment by genre (trend curve)

- Bob Marley
  - Reggae
  - Pop

The Beatles

Applications in Musicology | NLP for MIR | Sergio Oramas
Study by *album* publication year

Approach useful to study evolution of music genres

Strong correlation between average sentiment and average rating

Again useful insights for musicologists
Information Visualization

Extract a Knowledge Base from the documents of a Digital Library.

Build a Knowledge Graph to navigate through the library.

Create a visual representation of the graph.

Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- **Lexical Semantics**
- Deep Learning
- Conclusions and Future
Lexical Semantics
Introduction

· “What is it about the representation of a lexical item that gives rise to sense extensions and to the phenomenon of logical polysemy?” - Pustejovsky, 1995. *Introduction: Lexical Semantics in Context, Journal of Semantics.*

· Lexical Semantics is about understanding the “units of meaning” of the language. Not only words, but also compound words, phrases, affixes, etc.

· In NLP: formal (logic), path-based and **distributional semantics**.

· Distributional semantics intersects with *Relational Semantics*, i.e. establishing relationships between pairs of lexical units.
Distributional Lexical Semantics

· “You shall know a word by the company it keeps”, Firth (1957).
Distributional Lexical Semantics

· “You shall know a word by the company it keeps”, Firth (1957).

wampimuk
Distributional Lexical Semantics

· “You shall know a word by the company it keeps”, Firth (1957).

He filled the wampimuk with the substance, passed it around we all drunk some.
Distributional Lexical Semantics

- “You shall know a word by the company it keeps”, Firth (1957).

He filled the **wampimuk** with the substance, passed it around we all drunk some.

We found a little, hairy **wampimuk** sleeping behind the tree.
Distributional Lexical Semantics

· “You shall know a word by the company it keeps”, Firth (1957).

He filled the wampimuk with the substance, passed it around we all drunk some.

We found a little, hairy wampimuk sleeping behind the tree.

(McDonald and Ramscar, 2001)

Distributional Hypothesis: words that appear in similar contexts exhibit similar semantics.
Distributional Lexical Semantics

- Project linguistic items in vector space.

- Predictive models vs count-based models (Baroni et al., 2014).

- **word2vec** (Mikolov et al., 2013), Glove (Pennington et al., 2014) …
Distributional Lexical Semantics

Male-Female

Verb tense

Country-Capital

Lexical Semantics

NLP for MIR

Luis Espinosa-Anke
Distributional Lexical Semantics

```python
>>> from gensim.models import Word2Vec
>>> model = Word2Vec.load(PATH)
```

https://radimrehurek.com/gensim/models/word2vec.html
Distributional Lexical Semantics
- Word similarity, relatedness
  or analogy tasks.

```python
>>> model.most_similar(positive=['woman', 'king'], negative=['man'])
[(u'queen', 0.71), ('monarch', 0.61), (u'princess', 0.59) ... ]
```
Distributional Lexical Semantics

- Can be used to discover facts about music. Representative instruments!

Hendrix is to guitar as Mozart is to x
Distributional Lexical Semantics

- Can be used to discover facts about music. **Representative instruments**

\[ \text{Hendrix is to guitar as Mozart is to } x \]

```python
>>> model.most_similar(positive=['Mozart', 'guitar'], negative=['Hendrix'])
[(u'piano', 0.52), (u'accordion', 0.47), (u'mandolin', 0.47), (u'banjo', 0.47),
 (u'trombone', 0.46), (u'flute', 0.44) ... ]
```
Distributional Lexical Semantics

- Can be used to discover facts about music. Associated Music Genres

*Enrique Iglesias is to Pop as Elvis Presley is to …*

```python
model.most_similar(positive=['Elvis', 'Pop'], negative=['Enrique_Iglesias'])
```
Distributional Lexical Semantics

- Can be used to discover facts about music. **Associated Music Genres**

*Enrique Iglesias is to Pop as Elvis Presley is to …*

model.most_similar(positive=['Elvis', 'Pop'], negative=['Enrique_Iglesias'])

[(u'Country', 0.57), (u'Rock', 0.57), (u'Reggae', 0.57), (u'Blues', 0.55), (u'Metal', 0.55), (u'Jazz', 0.54), (u'Punk', 0.54), (u'Hip_Hop', 0.54), (u'Rap', 0.53), (u'Bluegrass', 0.53)]
A word2vec model in the Music domain

- The model has a restricted vocabulary of 21635 words.
- Trained over 19850433 raw words and 861414 sentences.
- Trained on the following datasets (overall +72k documents):
  - Grove music encyclopedia, 16708 biographies.
  - Last.fm, 23015 biographies.
  - Songfacts trivia, biographies and tidbits, 32326 documents.
- Available at (we will upload further versions trained on larger corpora and additional preprocessing): [http://mtg.upf.edu/nlp-tutorial](http://mtg.upf.edu/nlp-tutorial)
A word2vec model trained on music corpora

```python
>>> model.most_similar(positive=['beatles','mick_jagger'],negative=['john_lennon'])
[(u'rolling_stones', 0.6256111860275269), ... ]

>>> model.most_similar(positive=['dance-pop','zz_top'],negative=['lady_gaga'])
[(u'jazz-rock', 0.6238052845001221) ... ]

>>> model.most_similar(positive=['syd_barrett','roger_waters'])
[(u'david_gilmour', 0.7655651569366455) ... ]

>>> model.most_similar(positive=['iggy_pop'])
[(u'patti_smith', 0.7802923917770386) ... ]
```
Other uses of embeddings for music lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.
Other uses of embeddings for lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.
  
  - The influence of *sisters of mercy* became evident in later *poetry*.
Other uses of embeddings for lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.
  - The influence of *sisters of mercy* became evident in later *poetry*. 
Other uses of embeddings for lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.
  - The influence of *sisters of mercy* became evident in later *poetry*.

Exploit sense-level embeddings using **BabelNet** (Navigli and Ponzetto, 2012) as a reference sense inventory (e.g. SensEmbed, by Iacobacci et al. 2015)

[https://iiacobac.wordpress.com/2015/09/02/senseembed/](https://iiacobac.wordpress.com/2015/09/02/senseembed/)
Other uses of embeddings for lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.
  - The influence of *sisters of mercy* is evident in many later *poetry* acts.

(Delli Bovi et al., 2015)

\[
\text{COS} \left( n, \varphi \right) = \max_{v_n \in V(n), v_{\varphi} \in V(\varphi)} \frac{v_n \cdot v_{\varphi}}{||v_n|| ||v_{\varphi}||}
\]
Other uses of embeddings for lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.

```python
>>> import sensembed_api as sensembed
>>> sister_senses = sensembed.getLemmaSenses('sisters_of_mercy')
>>> sisters_senses
[u'sisters_of_mercy_bn:00424887n', u'sisters_of_mercy_bn:03828439n']
>>> poetry_senses = sensembed.getLemmaSenses('poetry')
>>> sensembed.closest_senses(sisters_senses, poetry_senses)
(u'sisters_of_mercy_bn:03828439n', u'poetry_bn:00063195n',
0.08942216509947952)
```
Other uses of embeddings for lexical semantics

- Word Sense Disambiguation and Entity Linking in the music domain.

```python
>>> import sensembed_api as sensembed
>>> sister_senses = sensembed.getLemmaSenses('sisters_of_mercy')
>>> sisters_senses
[u'sisters_of_mercy_bn:00424887n', u'sisters_of_mercy_bn:03828439n']
>>> poetry = sensembed.getLemmaSenses('poetry')
>>> sensembed.closest_senses(sisters_senses, poetry)
(u'sisters_of_mercy_bn:03828439n', u'poetry_bn:00063195n', 0.08942216509947952)
```
Conclusion

- Lexical semantics is a *buzzword* in NLP.

- VSMs, lexical semantics and advances in neural approaches have opened up a vibrant area of research.

- EMNLP2015 (Conference with A rating according to Google Scholar):
  * Empirical Methods in Natural Language Processing
  * “The insider joke in Lisbon was that the E in EMNLP now stands for Embedding (instead of Empirical) (...)” (https://wit.ai/blog/2015/09/23/emnlp)
References


Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- **Deep Learning**
- Conclusions and Future
Deep Learning
Deep Learning in Natural Language Processing

Deep Learning improves almost all tasks in NLP!! (as in many other fields)

Deep Network Architectures: LSTM y CNN

**LSTM**: parsing, entity recognition, sentiment analysis

**CNN**: classification, sentiment analysis

**More than words**: end-to-end, character level processing, word embeddings
Word2vec

Predict a context word $c \in (w_{i-L}, \ldots, w_{i-1}, w_{i+1}, \ldots, w_{i+L})$ given a word $w_i$

$$P(D = 1|w, c) = \sigma(v_w \cdot v_c)$$

$$\sigma(a) = \frac{1}{1+e^{-a}}$$
Skip-Gram Negative Sampling (SGNS)

Maximize $P(D = 1|w, c)$ for observed $(w, c)$

Maximize $P(D = 0|w, c)$ for randomly sampled “negative” examples $(w, c)$

$$\arg \max_{\theta} \sum_{(w,c) \in D} \log \sigma(v_c \cdot v_w) + \sum_{(w,c) \in D'} \log \sigma(-v_c \cdot v_w)$$
Word2vec as Matrix Factorization

Word and context embeddings matrices $\mathbf{W}$ and $\mathbf{C}$ are learnt.

$\mathbf{W}$ is typically used in NLP, while $\mathbf{C}$ is ignored.

$\mathbf{C} \cdot \mathbf{W}^T = \mathbf{M}$ what is $\mathbf{M}$?

According to Levy et al. 2014

$$M_{i,j}^{\text{SGNS}} = W_i \cdot C_j = \tilde{w}_i \cdot \tilde{c}_j = PMI(w_i, c_j) - \log k$$

$$PMI(x, y) = \log \frac{P(x, y)}{P(x)P(y)}$$
More about Word2vec

https://www.tensorflow.org/versions/r0.10/tutorials/word2vec/index.html


http://hduongtrong.github.io/2015/11/20/word2vec/


Beyond words

\( C \) and \( W \) can be different from words

Ej.:

- \( W \) songs or artists, \( C \) playlists
- \( W \) tags, \( C \) items

We can learn vector embeddings of musical items
Word2vec in Playlists

Trained with Gensim in Art of the Mix playlists
(http://labrosa.ee.columbia.edu/projects/musicsim/aotm.htm)

model.most_similar('miles davis')
[('john clotrane', 0.88384414), ('dizzie gillespie', 0.78484219), ('charlie walker', 0.74520659)]

model.most_similar('marilyn manson')
[('godsmack', 0.93274206), ('white zombie', 0.91064525), ('drowning pool', 0.90275443)]

model.most_similar('nirvana')
[('soundgarden', 0.84231329), ('pearl jame', 0.8271907), ('oysterhead', 0.81855756)]
Deep Learning for Music Recommendation
Deep Learning for Music Recommendation

Matrix Factorization

<table>
<thead>
<tr>
<th>u_1</th>
<th>u_2</th>
<th>...</th>
<th>u_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>i_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i_n</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

= Users

Items
Cold start problem

No user’s information for new items → Collaborative filtering doesn’t work

Need of content-based or hybrid approaches:

- Aggregation of feature vectors
- Learn item factors from content features

Deep Learning for Music Recommendation

Deep Learning for Music Recommendation

BRITNEY SPEARS

Deep Learning

NLP for MIR

Sergio Oramas
Dataset

**Million Song Dataset +** Artist biographies and tags from **Last.fm**

Artists: ~27k  
Users: 1 million  
Sparsity: 0.9990

<table>
<thead>
<tr>
<th>Input</th>
<th>Embed.</th>
<th>Learning</th>
<th>MAP@500</th>
<th>ROC-AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>VSM</td>
<td>Random Forest</td>
<td>0.015</td>
<td>0.664</td>
</tr>
<tr>
<td>Text</td>
<td>VSM</td>
<td>Feed Forward</td>
<td>0.030</td>
<td>0.748</td>
</tr>
<tr>
<td><strong>Text + Semantic Graph</strong></td>
<td>VSM</td>
<td>Feed Forward</td>
<td><strong>0.035</strong></td>
<td><strong>0.748</strong></td>
</tr>
<tr>
<td>Text</td>
<td>avg-w2v</td>
<td>Feed Forward</td>
<td>0.010</td>
<td>0.686</td>
</tr>
<tr>
<td>Text</td>
<td>w2v</td>
<td>LSTM</td>
<td>0.010</td>
<td>0.697</td>
</tr>
<tr>
<td><strong>Text + Semantic Graph</strong></td>
<td>n2v</td>
<td>Feed Forward</td>
<td>0.028</td>
<td>0.763</td>
</tr>
<tr>
<td>Random</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>0.495</td>
</tr>
<tr>
<td>Tags</td>
<td>VSM</td>
<td>Feed Forward</td>
<td>0.057</td>
<td>0.786</td>
</tr>
<tr>
<td>Upperbound</td>
<td>-</td>
<td>-</td>
<td><strong>0.519</strong></td>
<td>0.955</td>
</tr>
<tr>
<td>Input</td>
<td>Embed.</td>
<td>Learning</td>
<td>MAP@500</td>
<td>ROC-AUC</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>-----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Text</td>
<td>VSM</td>
<td>Random Forest</td>
<td>0.015</td>
<td>0.664</td>
</tr>
<tr>
<td>Text</td>
<td>VSM</td>
<td>Feed Forward</td>
<td>0.030</td>
<td>0.748</td>
</tr>
<tr>
<td><strong>Text + Semantic Graph</strong></td>
<td>VSM</td>
<td>Feed Forward</td>
<td><strong>0.035</strong></td>
<td><strong>0.748</strong></td>
</tr>
<tr>
<td>Text</td>
<td>avg-w2v</td>
<td>Feed Forward</td>
<td>0.010</td>
<td>0.686</td>
</tr>
<tr>
<td>Text</td>
<td>w2v</td>
<td>LSTM</td>
<td>0.010</td>
<td>0.697</td>
</tr>
<tr>
<td>Text + Semantic Graph</td>
<td>n2v</td>
<td>Feed Forward</td>
<td>0.028</td>
<td>0.763</td>
</tr>
<tr>
<td>Random</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>0.495</td>
</tr>
<tr>
<td>Tags</td>
<td>VSM</td>
<td>Feed Forward</td>
<td>0.057</td>
<td>0.786</td>
</tr>
<tr>
<td>Upperbound</td>
<td>-</td>
<td>-</td>
<td>0.519</td>
<td>0.955</td>
</tr>
<tr>
<td>Input</td>
<td>Embed.</td>
<td>Learning</td>
<td>MAP@500</td>
<td>ROC-AUC</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------</td>
<td>-----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Text</td>
<td>VSM</td>
<td>Random Forest</td>
<td>0.015</td>
<td>0.664</td>
</tr>
<tr>
<td>Text</td>
<td>VSM</td>
<td>Feed Forward</td>
<td>0.030</td>
<td>0.748</td>
</tr>
<tr>
<td><strong>Text + Semantic Graph</strong></td>
<td>VSM</td>
<td>Feed Forward</td>
<td><strong>0.035</strong></td>
<td><strong>0.748</strong></td>
</tr>
<tr>
<td>Text</td>
<td>avg-w2v</td>
<td>Feed Forward</td>
<td>0.010</td>
<td>0.686</td>
</tr>
<tr>
<td>Text</td>
<td>w2v</td>
<td>LSTM</td>
<td>0.010</td>
<td>0.697</td>
</tr>
<tr>
<td><strong>Text + Semantic Graph</strong></td>
<td>n2v</td>
<td>Feed Forward</td>
<td><strong>0.028</strong></td>
<td><strong>0.763</strong></td>
</tr>
<tr>
<td>Random</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>0.495</td>
</tr>
<tr>
<td>Tags</td>
<td>VSM</td>
<td>Feed Forward</td>
<td>0.057</td>
<td>0.786</td>
</tr>
<tr>
<td>Upperbound</td>
<td>-</td>
<td>-</td>
<td>0.519</td>
<td>0.955</td>
</tr>
</tbody>
</table>
Multimodal Approach

Audio and text can be combined in a deep neural network
Outline

- Introduction to NLP
- Information Extraction
  - Construction of Music Knowledge Bases
  - Semantic Enrichment of Musical Texts
- Applications in MIR
- Applications in Musicology
- Lexical Semantics
- Deep Learning
- Conclusions and Future
Conclusions and Future
Conclusions

- The extraction of high level semantic representations from text have been shown useful in different MIR and Musicological problems.
- There is already a need of new methodologies that better exploit these semantic representations.
- Word Embeddings and Deep Learning opens a new world of barely exploited possibilities.
- This tutorial is an initial attempt to boost the interaction between the NLP and MIR communities.
# Datasets Overview

<table>
<thead>
<tr>
<th>Name</th>
<th>Documents</th>
<th>Task</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS</td>
<td>artist biographies</td>
<td>similarity</td>
<td><a href="http://mtg.upf.edu/download/datasets/semantic-similarity">http://mtg.upf.edu/download/datasets/semantic-similarity</a></td>
</tr>
<tr>
<td>MARD</td>
<td>album reviews</td>
<td>classification</td>
<td><a href="http://mtg.upf.edu/download/datasets/mard">http://mtg.upf.edu/download/datasets/mard</a></td>
</tr>
<tr>
<td>KGRec-sound</td>
<td>sound descriptions</td>
<td>recommendation</td>
<td><a href="http://mtg.upf.edu/download/datasets/knowledge-graph-rec">http://mtg.upf.edu/download/datasets/knowledge-graph-rec</a></td>
</tr>
<tr>
<td>KGRec-music</td>
<td>song stories</td>
<td>recommendation</td>
<td><a href="http://mtg.upf.edu/download/datasets/knowledge-graph-rec">http://mtg.upf.edu/download/datasets/knowledge-graph-rec</a></td>
</tr>
<tr>
<td>ELMD</td>
<td>artist biographies</td>
<td>entity recognition</td>
<td><a href="http://mtg.upf.edu/download/datasets/elmd">http://mtg.upf.edu/download/datasets/elmd</a></td>
</tr>
</tbody>
</table>

---

**Conclusions and Future**

**NLP for MIR**
## KBs Overview

<table>
<thead>
<tr>
<th>Name</th>
<th>Source documents</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>KBSF</td>
<td>songs stories</td>
<td><a href="http://mtg.upf.edu/download/datasets/kbsf">http://mtg.upf.edu/download/datasets/kbsf</a></td>
</tr>
<tr>
<td>FlaBase</td>
<td>flamenco music webs</td>
<td><a href="http://mtg.upf.edu/download/datasets/flabase">http://mtg.upf.edu/download/datasets/flabase</a></td>
</tr>
</tbody>
</table>
Open Knowledge Extraction Challenge @ European Semantic Web Conference’17

- We are currently annotating and validating a gold standard dataset in the context of Task 3 in the OKE challenge @ ESWC 2017:
  - Focused Musical NE Recognition and Linking
- A good opportunity to develop and evaluate an EL system in the music domain.
- Reference inventory is MusicBrainz (instead than the classic DBpedia URIs).
Open Knowledge Extraction Challenge @ European Semantic Web Conference’17

Call for Participation - 2 Tasks

Musical NE Recognition

Identification of musical entities: Artist, Album, Song

Musical NE Linking

Linking of identified entities to MusicBrainz

Open Knowledge Extraction Challenge @ European Semantic Web Conference’17

Call for Participation - 2 Tasks

Musical NE Recognition
Identification of musical entities: Artist, Album, Song

Musical NE Linking
Linking of identified entities to MusicBrainz


Cash prize!
When Simon & Garfunkel split in 1970, Simon quickly began his solo career with the release of the self-titled album “Paul Simon”. This was followed by “There Goes Rhymin’ Simon” and “Still Crazy After All These Years”, both of which featured chart-topping hits such as “Loves Me Like A Rock” and “Kodachrome”.

<table>
<thead>
<tr>
<th>identified named entity</th>
<th>classified type</th>
<th>generated URI</th>
<th>indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simon &amp; Garfunkel</td>
<td>MusicArtist</td>
<td>artist:5d02f264-e225-41ff-83f7-d9b1f0b1874a</td>
<td>5,22</td>
</tr>
<tr>
<td>Simon</td>
<td>MusicArtist</td>
<td>artist:fc0a5289-4b77-4246-9c8d-857c8b617f5d</td>
<td>38,43</td>
</tr>
<tr>
<td>Paul Simon</td>
<td>SignalGroup</td>
<td>release-group:a1cc3fbd-609b-323c-95e2-435dfceb51e9</td>
<td>117,127</td>
</tr>
<tr>
<td>There Goes Rhymin’ Simon</td>
<td>SignalGroup</td>
<td>release-group:fb1e90a8-4461-382b-9081-183abb3c8997</td>
<td>152,176</td>
</tr>
<tr>
<td>Still Crazy After All These Years</td>
<td>SignalGroup</td>
<td>release-group:cd0c17f4-ff8d-3b1d-ac36-397ebeb069e9</td>
<td>183,216</td>
</tr>
<tr>
<td>Loves Me Like A Rock</td>
<td>MusicalWork</td>
<td>work:bc76594b-b113-4a57-b929-b9911531108e</td>
<td>270,290</td>
</tr>
<tr>
<td>Kodachrome</td>
<td>MusicalWork</td>
<td>work:c9ad17e6-440e-40b6-b4f9-58b74b006c20</td>
<td>297,307</td>
</tr>
</tbody>
</table>
Future

Chatbots

Deep Learning + Semantics

Multimodality

Deep Generative Models

  Text generation from audio

  Audio generation from text
Thanks!

Questions? Ideas? Suggestions?

@sergiooramasa @luisanke