Knolwedge Acquisition from Music Digital Libraries

Sergio Oramas, Mohamed Sordo
Motivation

- Musicological Knowledge is hidden between the lines
- Machines don’t know how to read
Why Knowledge Acquisition?

- Obtain knowledge automatically
- Make complex questions
- Visualize the information
- Improve navigation
- Share knowledge
Musical Libraries

**Items:** Books, scores, manuscripts, letters, journals, photos, etc.
Musical Libraries

**Digitized items:** pdf, jpg, wav, mp3

**Items:** Books, scores, manuscripts, letters, journals, photos, etc.

digital recording, scan
Musical Libraries

- **Items**: Books, scores, manuscripts, letters, journals, photos, etc.
- **Digitized items**: pdf, jpg, wav, mp3
- **Machine-readable items**: midi, xml, txt, html

- Digital recording, scan
- OCR, manual transcription
Musical Libraries

- **Items**: Books, scores, manuscripts, letters, journals, photos, etc.
- **Digitized items**: pdf, jpg, wav, mp3
- **Machine-readable items**: midi, xml, txt, html
- **Semantified items**: high level features, entities, relations, uris

- **Information Extraction, semantic annotation**
- **OCR, manual transcription**
- **Digital recording, scan**
Musical Libraries

**Items**: Books, scores, manuscripts, letters, journals, photos, etc.

**Digitized items**: pdf, jpg, wav, mp3

**Machine-readable items**: midi, xml, txt, html

**Semantified items**: high level features, entities, relations, uris

**Information Extraction, semantic annotation**

**OCR, manual transcription**

**digital recording, scan**

**Current DL**
Musical Libraries

- **Items**: Books, scores, manuscripts, letters, journals, photos, etc.
  - **Digitized items**: pdf, jpg, wav, mp3
  - **Machine-readable items**: midi, xml, txt, html
  - **Semantified items**: high level features, entities, relations, uris

- **Web search**
- **Current DL**
- **Information Extraction, semantic annotation**
- **OCR, manual transcription**
- **digital recording, scan**
Musical Libraries

- Digital recording, scan
  - OCR, manual transcription
  - Digital recording, scan

- Digitized items: pdf, jpg, wav, mp3
  - Current DL

- Machine-readable items: midi, xml, txt, html
  - Web search
  - Information Extraction, semantic annotation

- Semantified items: high level features, entities, relations, uris
  - Items: Books, scores, manuscripts, letters, journals, photos, etc.
Semantic Web

- The **Semantic Web** aims at converting the current web, dominated by unstructured and semi-structured documents into a web of linked data.

- Achievements useful for Digital Libraries
  - Common framework for data representation and interconnection (RDF, ontologies)
  - Semantic technologies to annotate texts (Entity Linking)
  - Language for complex queries (SPARQL)
Wikipedia and DBpedia

- Digital Encyclopedia
- Unstructured
- Keyword search

- Knowledge Base
- Structured
- Query search
Claude Debussy

From Wikipedia, the free encyclopedia

"Debussy" redirects here. For other uses, see Debussy (disambiguation).

Claude-Achille Debussy[1] (French: [klod aʃil debysi],[2] 22 August 1862 – 25 March 1918) was a French composer. Along with Maurice Ravel, he was one of the most prominent figures associated with impressionist music, though he himself disliked the term when applied to his compositions.[3] He was made Chevalier of the Legion of Honour in his native France in 1903.[4] Debussy was among the most influential composers of the late 19th and early 20th centuries, and his use of non-traditional scales and chromaticism influenced many composers who followed.[5]

Debussy's music is noted for its sensory content and frequent usage of atonality. The prominent French literary style of his period was known as Symbolism, and this movement directly inspired Debussy both as a composer and as an active cultural participant.[6]

Contents

1 Early life
   1.1 Musical development
2 Personal life
3 Death
4 Music
   4.1 Style
   4.2 List of works
   4.3 Early works
   4.4 Middle works
   4.5 Late works
   4.6 Mathematical structuring
   4.7 Influences

Claude Debussy in 1908
Claude-Achille Debussy (Saint-Germain-en-Laye, Francia, 22 de agosto de 1862 - París, 25 de marzo de 1918) fue un compositor francés y una figura central en la música europea de finales del siglo XIX y comienzos del siglo XX y, junto a Maurice Ravel, una de las figuras más prominentes de la música impresionista, aunque al propio compositor no le gustaba este término cuando se aplicaba a sus composiciones. Fue nombrado Caballero de la Legión de Honor en 1903.

Property | Value
---|---
dbpedia-owl:abstract | Achille-Claude Debussy (French: [ɛʃil kloːd dɛbusi]; 22 August 1862 – 25 March 1918) was a French composer. Along with Maurice Ravel, he was one of the most prominent figures associated with impressionist music, though he himself intensely disliked the term when applied to his compositions. In France, he was made Chevalier of the Legion of Honour in 1903. Debussy was among the most influential composers of the late 19th and early 20th centuries, and his use of non-traditional scales and chromaticism influenced many composers who followed. Debussy's music is noted for its sensory content and frequent eschewing of tonality. The French literary style of his period was known as Symbolism, and this movement directly inspired Debussy both as a composer and as an active cultural participant.

dpoezia-owl:alias | Debussy, Achille-Claude

dbpedia-owl:birthDate | 1862-08-22 (xsd:date)

dbpedia-owl:birthYear | 1862

dbpedia-owl:deathDate | 1918-03-25 (xsd:date)

dbpedia-owl:deathYear | 1918

dbpedia-owl:indivId | 118524166


dbpedia-owl:viafid | 6219836

dbpedia-owl:wikiPageExternalLink | http://debussepiano.com/
| http://www.musopen.com/
| http://www.uequebec.ca/musique/catal/debussy/debcat.htm

dbpedia-owl:wikiPageId | 6260 (xsd:integer)

dbpedia-owl:wikiPageRevisionId | 60504927 (xsd:integer)

ddprop:alternativeNames | Debussy, Achille-Claude

ddprop:alias | artist

ddprop:date | November 2012

ddprop:dateOfBirth | 1862-08-22 (xsd:date)

ddprop:dateOfDeath | 1918-03-25 (xsd:date)

ddprop:description | Performed by Robin Aictiore. Courtesy of Musopen
| Performed by Sarah Bassingthwaite
DBpedia

• Dbpedia example queries
  – Composers born in Vienna in XVIII Century
  – American jazz musicians that have written songs recorded by RCA Records

• Dbpedia graph applications
  – Entity Relevance
  – Entity Similarity
  – Entity Recommendation
Music Digital Libraries

• Current navigation
  – Indexes
  – Keyword-based search

• From searchable repositories to **knowledge environments**
  – Complex queries
  – Information visualization
  – Recommendation
  – Data Analytics
  – Q&A
Music Digital Libraries

Music Digital Library

- Audio
- Scores
- Texts
- Images
- Videos
Music Digital Libraries

Audio
Scores
Texts
Images
Videos
Dataset: The New Grove

- Encyclopedic dictionary
- One of the largest reference works in Western music
- Artist biographies crawled from the Grove Music Online
  - 16,707 biographies (1st paragraph)
  - From pre-medieval to contemporary
Dataset: The New Grove

- What are the most relevant music schools?
- What are the artists most similar to Schoenberg?
- Which are the most represented roles in the Grove?
- Is there a migration tendency in artists? To which cities?
- What is the best city to die for a musician?
Dataset: The New Grove

Webern, Anton (Friedrich Wilhelm von)

(b Vienna, 3 Dec 1883; d Mittersill, 15 Sept 1945). Austrian composer and conductor. Webern, who was probably Schoenberg's first private pupil, and Alban Berg, who came to him a few weeks later, were the most famous of Schoenberg's students and became, with him, the major exponents of 12-note technique in the second quarter of the 20th century. Webern applied the new technique more rigorously than either Schoenberg, who took many liberties, or Berg, who never used it exclusively; Webern's strictness, and his innovative organization of rhythm and dynamics, were seized upon eagerly by Boulez and Stockhausen and other integral serialists of the Darmstadt School in the 1950s and were a significant influence on music in the second half of the century.
Anton Webern
(b Vienna, 3 Dec 1883; d Mittersill, 15 Sept 1945). Austrian composer and conductor. Webern, who was probably Schoenberg's first private pupil, and Alban Berg, who came to him a few weeks later […] Boulez and Stockhausen and other integral serialists of the Darmstadt School […]
Anton Webern
(b Vienna, 3 Dec 1883; d Mittersill, 15 Sept 1945). Austrian composer and conductor. Webern, who was probably Schoenberg's first private pupil, and Alban Berg, who came to him a few weeks later […] Boulez and Stockhausen and other integral serialists of the Darmstadt School […]
Anton Webern  
(b Vienna, 3 Dec 1883; d Mittersill, 15 Sept 1945). Austrian composer and conductor. Webern, who was probably Schoenberg's first private pupil, and Alban Berg, who came to him a few weeks later [...] Boulez and Stockhausen and other integral serialists of the Darmstadt School [...]

Information Extraction

Anton Webern

GeoNames
Anton Webern
(b Vienna, 3 Dec 1883; d Mittersill, 15 Sept 1945). Austrian composer and conductor. Webern, who was probably Schoenberg's first private pupil, and Alban Berg, who came to him a few weeks later [...] Boulez and Stockhausen and other integral serialists of the Darmstadt School [...]

Information Extraction

GeoNames
Anton Webern
(b Vienna, 3 Dec 1883; d Mittersill, 15 Sept 1945). Austrian composer and conductor. Webern, who was probably Schoenberg's first private pupil, and Alban Berg, who came to him a few weeks later [...] Boulez and Stockhausen and other integral serialists of the Darmstadt School [...]
Information Extraction: Relation Extraction

- Webern, who was probably Schoenberg's first private pupil
Information Extraction: Relation Extraction

- Webern, who was probably Schoenberg's first private pupil
Methodology
Knowledge Graph

- **Webern**
  - hasRole: conductor
  - bornIn: Vienna
- **Berg**
  - hasRole: composer
  - mentioned: Schoenberg
- **Schoenberg**
  - hasRole: composer
  - bornIn: Vienna
  - diedIn: Los Ángeles
- **Jelly Roll Morton**
  - hasRole: pianist
  - bornIn: New Orleans
  - relatedGenre: jazz
Knowledge Graph: Data Analytics

- 16,707 biographies
- 434 roles
- Graph
  - 47,367 nodes
  - 274,333 edges

<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>
Birth Year

Death Year
Birth Year

Death Year
## Knowledge Graph: Data Analytics

<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>
Knowledge Graph: Entity Relevance

- City
- Venue
  - Covent Garden Theatre, King's Theatre, Drury Lane, Carnegie Hall, Théâtre de la Monnaie, Stadttheater, Theatre Royal, ...
- Educational Institution
  - Paris Conservatoire, Moscow Conservatory, Juilliard School, St Petersburg Conservatory, Bmus, Prague Conservatory, Leipzig Conservatory, Vienna Hochschule für Musik, ...
Knowledge Graph: Entity Relevance

- Biography subject
  - Haydn, Claude Debussy, Arnold Schoenberg, Robert Stevenson, Paul Hindemith, Giovanni Pierluigi da Palestrina, Gustav Mahler, Maurice Ravel, Jean-Philippe Rameau
  - Mozart?, Bach?, Wagner?
- Genre
  - chamber music, cappella, jazz, folk music, avant garde, baroque music, electronic music, musical theatre, plainchant
Knowledge Graph: Entity Similarity

- PageRank algorithm and Maximal Common Subgraph

- **Arnold Schoenberg**: Anton Webern, Paul Hindemith, Gustav Mahler, Alban Berg, Claude Debussy

- **Guido Adler**: Heinrich Jalowetz, Eusebius Mandyczewski, Robert Fuchs, Karl Weigl, Anton Wranitzky

- **Manuel de Falla**: Ricardo Viñes, Juan Vicente Lecuna, Enrique Granados, Miguel Llobet Soles, Luigi Russolo

- **Miles Davis**: Dizzy Gillespie, Herbie Hancock, Paul Chambers, Tony Williams, Cannonball Adderley
Relations Graph: Information Visualization
Relations Graph: Information Visualization

José Antonio De Donostia

Primary Connections
- received from Felipe Pedrell
- joined Capuchin College
- was influenced by Romanticism
- left Barcelona
- compiled books of song Basque Language
- wrote publishing monograph about Basque Music
- was active as Ethnomusicologist
Mixing Different Data Sources

- Creation and curation of a library from Web content
- FlaBase: Flamenco Knowledge Base
Mixing Different Data Sources

• Data Acquisition
  – APIs
  – Web crawling
  – SPARQL endpoints

• Entity Resolution
  – String similarity between labels
  – Graph similarity (context information)
FlaBase: Flamenco Knowledge Base
FlaBase: Flamenco Knowledge Base

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

• Knowledge Extracted
  – Place of birth
  – Date of birth
  – Entity mentions in text
FlaBase: Data Analytics

- Number of artists by year of birth
FlaBase: Data Analytics
FlaBase: 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</td>
</tr>
<tr>
<td>Manolo Caracol</td>
<td>Ramón Montoya</td>
<td>Rosario</td>
</tr>
<tr>
<td>La Niña de los Peines</td>
<td>Niño Ricardo</td>
<td>Antonio Gades</td>
</tr>
<tr>
<td>Antonio Chacón</td>
<td>Manolo Sanlúcar</td>
<td>Mario Maya</td>
</tr>
<tr>
<td>Camarón de la Isla</td>
<td>Sabicas</td>
<td>Carmen Amaya</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></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>

Precision Values
Conclusions

• Music Digital Libraries can benefit from semantic approaches

• Music Digital Libraries are still in an early stage of development compared to the Web (Linked Open Data, Google Knowledge Graph)

• Knowledge acquisition from Digital Libraries can help musicologists not only to search content, but also to discover new knowledge
Aknowledgments

• This work was partly funded by the COFLA2 research project (Proyectos de Excelencia de la Junta de Andalucía, FEDER P12-TIC-1362).
Bibliography


Knowledge Acquisition from Music Digital Libraries

Sergio Oramas, Mohamed Sordo

sergio.oramas@upf.edu
@sergiooramas