

Opera and Classical Music

Project report

Knowledge Engineering
year 2020/21

Colagrande Pierpasquale, Cozzi Riccardo

Goal

This project aims at integrating different existing datasets encoded in heterogeneous formats about opera and classical music, into a semantic web knowledge graph (RDF). This knowledge graph will combine different sources which come in highly heterogeneous form.

Background

The subjects of this domain are the classic compositions, the composers, and the musical performances. The point of interest of this work is to align information coming from different sources to reconstruct as best as possible the story of musical composition, musical performance and the subjects involved in it.

Data are provided by different organizations each one with different goals and methodologies, so lacking a central organized structure. Plus, historical data are often retrieved from not exhaustive sources, leading to incomplete and/or inconsistent data.

Resources

The AudioLabs website provides information about song recordings and authors. We used two dataset: Cross-Composers [4] and Cross-Eras [5]. Cross-Composers dataset presented on AudioLabs website served as basis for studying the composer identification task for western classical music recordings. It is compiled from commercial audio recordings, totalling 1100 tracks where a track refers to the movement level of a piece. It provides annotations about composers, songs, and albums. Cross-Era dataset, from the same website, contains about 1000 tracks for piano and orchestra (without singing voice / solo instruments). In this case more specific information about the tracks are provided, as key and mode. Cross-Era also provides information about the author such as nationality and life period.

Operadatabase [6] is a comprehensive database of operas and arias, which aim is to to catalog arias and operas and providing a repository for librettos, scores,

translations, and synopses of songs and operas. For each song it is provided the author and the collection (album or opera) in which it is contained. It is composed of 4 datasetes: operas, zarzuelas, art-songs, opera-arias and zarzuela arias. The last one describes a Spanish composition similar to operas. Similarly, aria and zarzuela-arias have a similar structure. Art-songs dataset is a collection of singed songs which can have more than one voice, for each of which a link to the sheet is provided, if present.

We also used some available ontologies for alignment. The first one is Arco [8], which describes useful concepts such as musical events and artistic or musical properties and heritage. Then we exploited MuziekSchatten and Corago to align concepts about composers and performers.

Finally, the more general concepts, for instance about persons, songs, and operas and instruments, have been aligned to DBpedia [7].

User scenarios

We used the eXtreme Design methodology to delineate two user scenarios according to which we focused on different aspects of the ontology. We imagined two different types of users which can be interested in exploiting our ontology for different purposes. For each user we modelled a set of 3 competency questions, which will be listed and tested in the Results section.

Scenario 1 - Maria

Maria is a conservatory student writing a thesis about Opera in the 18th century. Her thesis concerns the analysis of trends in opera compositions in that period. She would like to understand if composers of the 19th century shared the same style of composition, analysing the types of instruments and voices used within their works, in order to understand what the trend was.

1. Which are the songs for piano composed in the 19th century?
2. What are Mozart's compositions that employ the usage of baritone singers?
3. Which are the operas with soprano or tenore singers?

Scenario 2 – Alejandro

Alejandro is a violinist, and he is going to perform with its orchestra in a theatrical event. In view of the preparation for the concert, Alejandro is preparing himself on Spanish pieces of 19th century that contain violin part and, in general, on pieces which contain violin parts. He also wants to improve its knowledge about the violinists of that period.

1. What are the Spanish works of the second half of the 1800s?
2. What are the Spanish operas with soprano voice?
3. What are the musicians who have played a piece written by an author that died in the 19th century containing a violin part?

Tools

While Cross-Era and Cross-Composers were directly available in a csv format downloadable from the website, to deal with operadatabase datasets we needed to add an additional step. In facts, the latter only provides a simple search engine based on keywords through which one can access the four dataset independently, producing results only available in a form of html table. To obtain useful csv files we used the python library BeautifulSoup4 which is supposed to scrape web pages.

The pre-processing has been done by using python to unify the names of attributes within the different tables. We managed to create specific python scripts for each dataset.

In these phases, all the datasets have been manipulated with Pandas library.

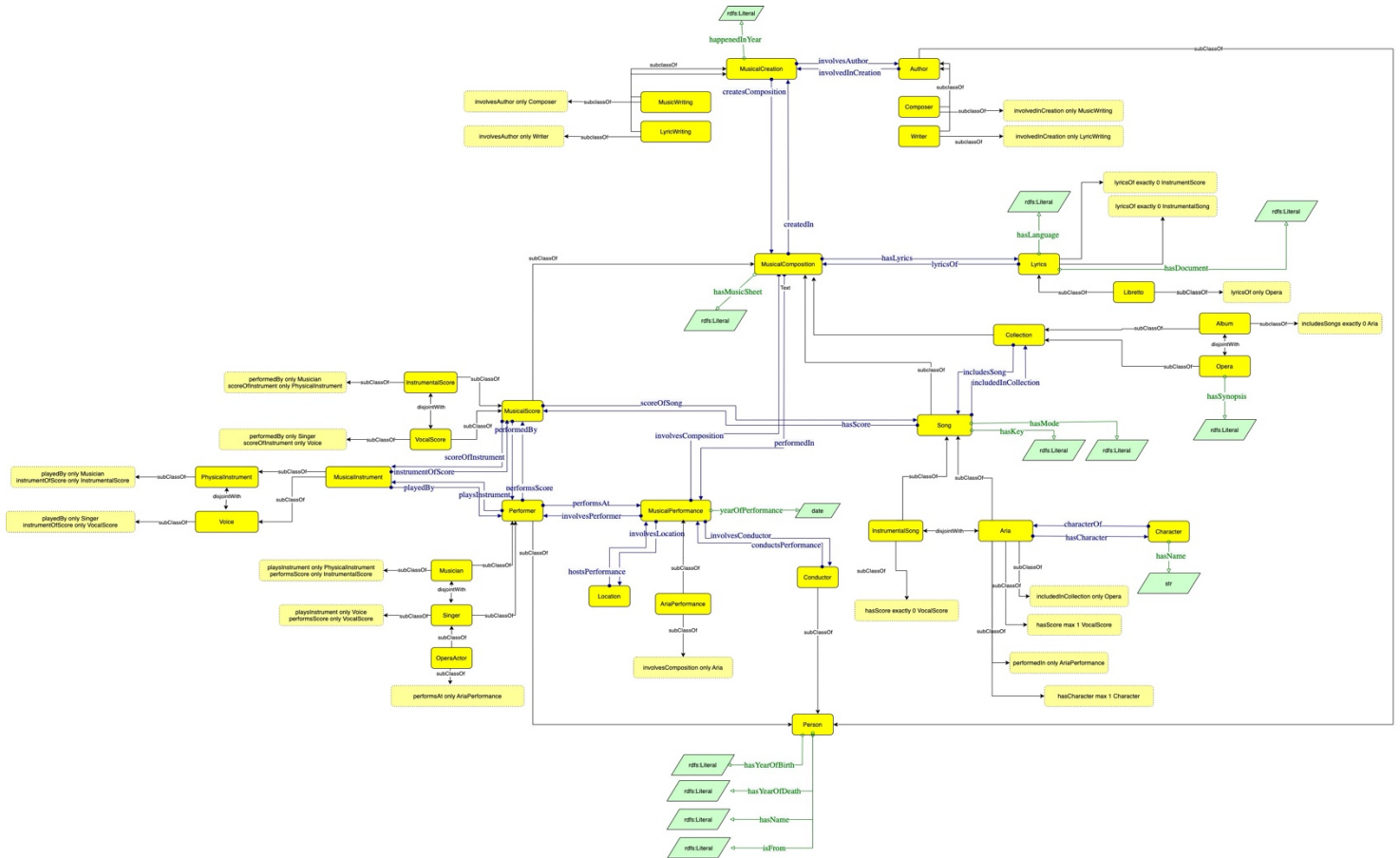
We used Protégé to define ontology entities and predicates and to export the ontology in an OWL file.

To map data into our knowledge base we used RDFLib for python, instantiating a main graph object to represent the ontology schema and another one for the assertions. Then the two named graphs have been bound and serialized that the knowledge base exported into an OWL file.

Finally, we used LIMES to produce triples for linking individuals over DBpedia.

We also used git and GitHub for collaborative development and Docker. Moreover, we also embedded in the docker container LODE for visualising ontologies as HTML, WebVOWL for visualising ontologies with the Visual

Notation for OWL Ontologies (VOWL) and LodView for browsing ontology entities as well as controlled vocabularies entities.



Ontology

The graphical representation of our ontology schema is available in our GitHub repository as high-resolution jpg file [2].

Musical Composition

A musical production of a process of music creation. It can be specialized in a Song, a Musical Score or a Collection of songs. The definition of a musical score is slightly more complex and less intuitive with respect to those for song and collection. By definition [3], a score is a part of a music sheet which may contain the single part for a solo work or the many parts that make up an orchestral or ensemble composition. In simpler words, the score for an

instrument describes, in music symbols, the “voice” of that instrument in a certain song, naming the “part that an instrument is supposed to execute”.

Musical Creation

With this entity we modelled the act [see odp section] of creating a musical composition. It involves an author, which can be demanded to compose the music or the lyrics. Depending on which parts are involved (composer-music, writer-lyrics) it can be specialized in MusicWriting and LyricWriting.

Musical Performance

It is an event in which a musical composition, such as a song or a collection of songs is performed. It also involves a location and a date for the premiere, one or more performers and possibly a conductor.

Musical Instrument

A musical instrument is a mechanical or natural tool used to make musical sounds. In this context, we can so include the human voice as subclass of musical instrument.

Person

Describes general human being, for which information such as name, nation of provenience and date of death and birth are known. In its subclasses it describes the roles of the different people involved in the various musical situations, such as music creation and musical performance. More specifically, within the context of a music creation the involved person is called author, which can be further specialized in composer and writer. In a performance situation, the involved people can be performers and conductors. The first one can be specialized in musician and singers. In the special case in which the performed composition is an aria, the performer takes the name of opera actor, which is considered as a special case of singer.

Lyrics

It is the text associated with a musical composition, namely the words pronounced in a singed song. In the special case of an opera, it is often available the libretto, namely the transcription of all the lyrics of its songs and arias.

Character

It is a fictional person who is interpreted by an opera actor during an aria.

Location

Indicates where a musical performance took place

Ontology design patters:

Situation design pattern

Since in classical music context there can be several adaptations for an opera, we needed to differentiate them in distinct situations. For example, many performances of the same opera or song can be done, involving different performers and directors. We designed both MusicCreation and MusicalPerformance as situation entities.

Collection (containment) design pattern

In this context there can be found musical scores as parts of songs, and songs as part of an opera. The need we had was to link all the parts of a song, and all the song of a collection to the same performance or to the same author. This pattern has been applied between Song and Collection and between Musical Score and Song.

Results

We successfully managed to upload the different datasets and interrogate our knowledge base to solve the required competency questions.

Scenario 1 (Maria)

Which are the songs for piano composed in the 19th century?

```
select distinct ?title where {
  ?creation ocm:involvesAuthor ?composer ;
             ocm:createsComposition ?composition .
  ?composer ocm:hasYearOfBirth ?year .
  ?song rdf:type ocm:Song ;
         ocm:hasScore ?score .
  ?instr ocm:instrumentOfScore ?score ;
         rdfs:label "Piano" .
  ?composition ocm:hasTitle ?title .
  filter(xsd:integer(?year) > 1800)
  filter(xsd:integer(?year) < 1900)
}
```

What are Mozart's compositions that employ the usage of baritone singers?

```
select ?title ?voicetype where {
  ?creation ocm:involvesAuthor ?composer;
             ocm:createsComposition ?song .
  ?composer ocm:hasName ?name .
  ?song ocm:hasScore ?score ;
         ocm:hasTitle ?title .
  ?instr ocm:instrumentOfScore ?score ;
         rdf:type ocm:Voice ;
         rdfs:label ?voicetype .
  FILTER regex(?voicetype, "Baritone")
  FILTER regex(?name, "Mozart") .
}
```

Which are the operas with soprano or tenore singers?

```
select ?title ?voicetype where {
  ?opera rdf:type ocm:Opera ;
         ocm:includesSong ?song ;
         ocm:hasTitle ?title.
  ?song ocm:hasScore ?score.
  ?instr ocm:instrumentOfScore ?score .
         rdf:type ocm:Voice ;
         rdfs:label ?voicetype .
  ?song ocm:hasLyrics ?lyrics .
  FILTER (?voicetype IN ("Soprano" , "Tenor")) .
}
```


Scenario 2 (Alejandro)

What are the Spanish operas of the second half of the 1800s?

```
select ?title ?language ?year where {
    ?creation ocm:involvesAuthor ?composer;
        ocm:createsComposition ?opera .
    ?composer ocm:hasYearOfBirth ?year.
    ?opera rdf:type ocm:Opera ;
        ocm:hasLyrics ?lyric ;
        ocm:hasTitle ?title .
    ?lyric ocm:hasLanguage ?language .
    FILTER (?language = "Spanish").
    FILTER (xsd:integer(?year) >= 1850) .
    FILTER (xsd:integer(?year) <= 1900) .
}
```

What are the Spanish operas with soprano voice?

```
select distinct ?title ?language ?voicetype where {
    ?creation ocm:involvesAuthor ?composer ;
        ocm:createsComposition ?opera .
    ?opera rdf:type ocm:Opera ;
        ocm:hasLyrics ?lyric ;
        ocm:includesSong ?song ;
        ocm:hasTitle ?title .
    ?lyric ocm:hasLanguage ?language .
    ?song ocm:hasScore ?score .
    ?instr ocm:instrumentOfScore ?score ;
        rdfs:label ?voicetype .
    ?song ocm:hasLyrics ?lyrics .
    FILTER (?voicetype = "Soprano") .
    FILTER (?language = "Spanish") .
}
```

What are the musicians who have played a piece written by an author that died in the 19th century containing a violin part?

```
select ?pname ?instrname ?cname ?title ?death where {
    ?performer ocm:playsInstrument ?instr .
        ocm:performsAt ?performance ;
        ocm:hasName ?pname .
    ?composition ocm:performedIn ?performance ;
        ocm:hasTitle ?title .
    ?creation ocm:createsComposition ?composition ;
        ocm:involvesAuthor ?author .
    ?author ocm:hasYearOfDeath ?death;
        ocm:hasName ?cname .
    ?instr rdfs:label ?instrname .
    filter(?instrname = "Violin") .
    filter(xsd:integer(?death) > 1800)
    filter(xsd:integer(?death) < 1900)
}
```

Conclusions and future works

We immediately realised that this context is way more complex than the one we described. We focused only on some aspects, but there are many others which can be involved to amplify the capabilities for interrogating the knowledge base. For instance, one can be interested in retrieving all the songs which talk about love in the period between 1800 and 1900. For this purpose, it can be interesting to actuate text mining processes over the lyrics. Another feature extraction it can be conduct could be about the music arrangements of the tracks (which we decided to ignore) in order to provide a more technical and detailed analysis of the nature of the songs, describing features such as speed, modes and so on.

References

1. github.com/pierclgr/Opera-Music-Ontology
2. https://github.com/pierclgr/Opera-Music-Ontology/tree/main/knowledge_graph
3. britannica.com/art/score-music
4. audiolabs-erlangen.de/resources/MIR/cross-comp
5. audiolabs-erlangen.de/resources/MIR/cross-era
6. theoperadatabase.com
7. dbpedia.org
8. corago.unibo.it
9. data.muziekschatten.nl
10. wit.istc.cnr.it/arco