UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

FIB

Semantic Data Management

Lab 3 – Knowledge Graphs

Victor Gonzalez Monge
victor.gonzalez.monge@estudiantat.upc.edu

Emmanuel Werr
emmanuel.werr@estudiantat.upc.edu

Facultat d'Informàtica de Barcelona
Master of Data Science

May 18, 2023

B Ontology Creation

B.1 TBOX Definition

In this section, we use Protege to define a TBOX for the research publication domain following the idea of Lab 1, while modeling the concepts of paper, authorship, publication, and review. The resulting file for the generated TBOX is included in the zipped submission for this lab and is titled **"12I-B1-GonzalezWerr.ttl"**.

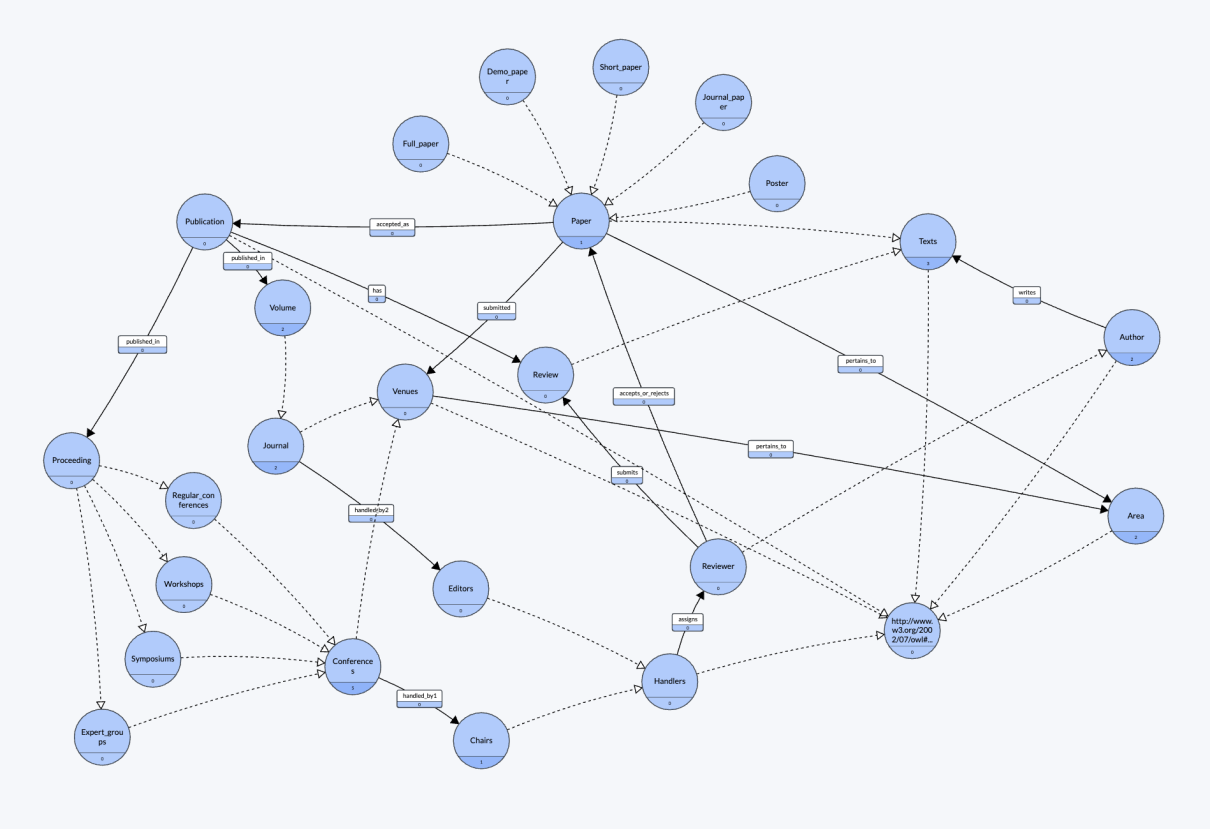

Figure 1: Visualization of TBOX using Gra.fo

B.2 ABOX Definition

For this section, we have decided to create the ABOX programatically using the RDFLib Library in Python. The python script is included in the zipped submission for this lab and is titled **"12I-B2-GonzalezWerr.py"**. In order to generate the ABOX from non-semantic data (reusing our data from Lab1) it was necessary to add new properties throughout the different entities in order to make the final dataset compatible with this project.

B.3 Create the Final Ontology

In this section, we have decided to link the TBOX and ABOX programatically using RDFLib. The python script is included in the zipped submission for this lab and is titled **"12I-B3-GonzalezWerr.py"**. The resulting ABOX is also included in the zipped submission and is titled **"12I-B3-GonzalezWerr.owl"**.

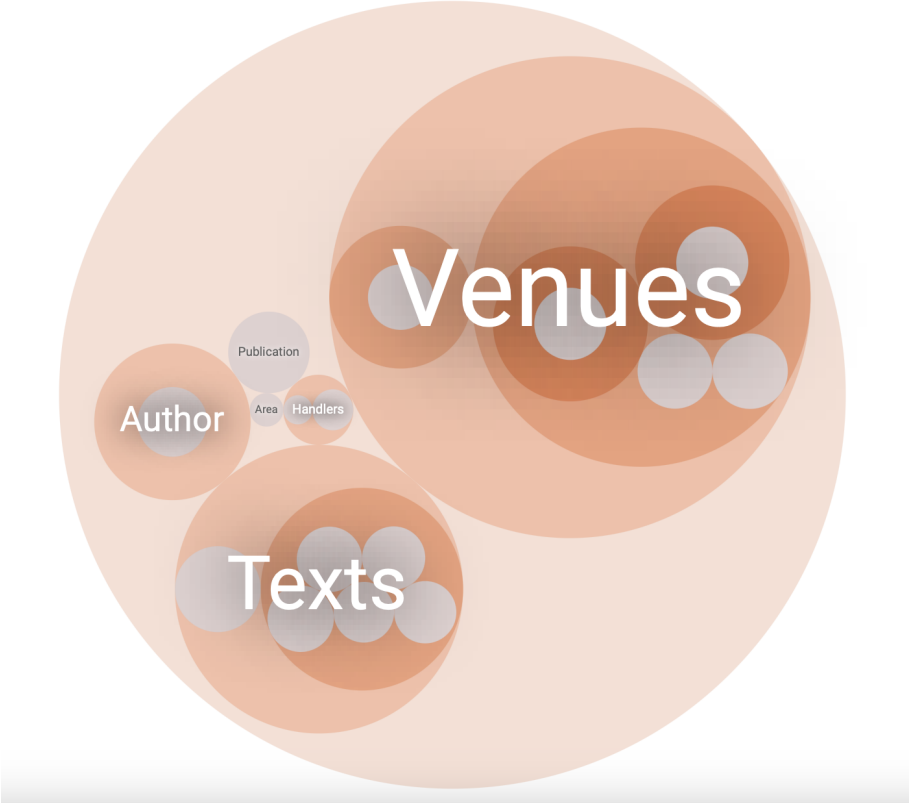

Figure 2: Visualization of Class Hierarchy of Knowledge Graph

The figure above shows a screenshot of the visualization of Class Hierarchies provided by GraphDB in the "Explore" panel. Simple statistics have also been calculated following the provided examples by the Official Lab3 Document. Due to their size the result tables, as well as the SPARQL queries used to obtain them, are **provided in the Appendix at the end of this document**.

We are considering RDFS Regime Entailment Inference Rules. The rdf:type links we saved to explicitly generate thanks to reasoning are the following:

- :Volumes is a subclass of :Venues
- all the subclasses of :Conferences are also subclass of :Venues

B.4 Querying the Ontology

B.4.1. Find all Authors.

```
PREFIX : <http://graph.org/abox/>

SELECT ?author
WHERE {
  ?author a :Author .
}
```

Figure 3: Partial Output from SPARQL Query B.4.1

B.4.2. Find all properties whose domain is Author.

```
PREFIX : <http://graph.org/abox/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?property
WHERE {
  ?property rdfs:domain :Author .
}
```

Figure 4: Full Output from SPARQL Query B.4.2

B.4.3. Find all properties whose domain is either Conference or Journal.

```
PREFIX : <http://graph.org/abox/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?property
WHERE {
  { ?property rdfs:domain :Conferences }
  UNION { ?property rdfs:domain :Journal }
}
```

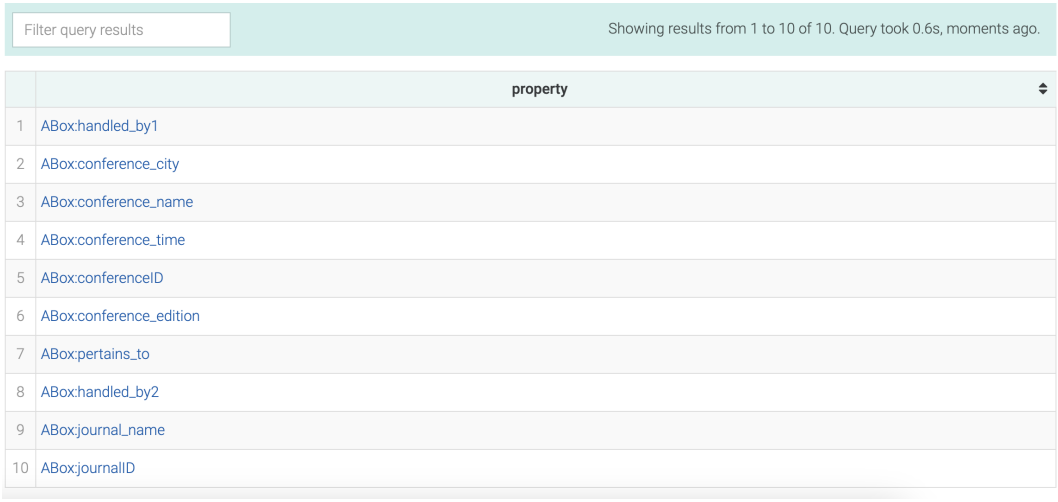

Figure 5: Full Output from SPARQL Query B.4.3

B.4.4. Find all the papers written by (:Author_14) that where published in database conferences.

```
PREFIX : <http://graph.org/abox/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?papers
WHERE {
  ?conference rdf:type :Proceeding .
    :Author_14 :writes ?papers .
  ?papers :accepted_as ?pub .
  ?pub :published_in ?conference .
  ?conference :pertains_to ?area .
  ?area :area ?areaLabel .
  FILTER (str(?areaLabel) = "databases"
    && langMatches(lang(?areaLabel), "en"))
}
```

Figure 6: Full Output from SPARQL Query B.4.4

Appendix

This appendix section contains some simple statistics about our resulting knowledge graph as well as the SPARQL queries used to generate them.

A.3.1 Total Number of Distinct Classes and Properties

There are 37 total distinct classes (24 Main Classes), and 31 total distinct properties (22 Main Properties) in the knowledge graph. The queries used to obtain those results are showcased below.

```
PREFIX : <http://graph.org/abox/>

SELECT (COUNT(DISTINCT ?class) as ?count) WHERE {
  ?s a ?class .
}
```

```
PREFIX : <http://graph.org/abox/>

SELECT (COUNT(DISTINCT ?p) as ?count) WHERE {
  ?s ?p ?o .
}
```

A.3.2 Number of Instances for each of the Main Classes

```
PREFIX : <http://graph.org/abox/>

SELECT ?class (COUNT(?s) as ?count) WHERE {
  ?s a ?class .
  FILTER(STRSTARTS(STR(?class), "http://graph.org/abox/"))
} GROUP BY ?class ORDER BY DESC(?count)
```

	class	count
1	ABox:Texts	"980"^^xsd:integer
2	ABox:Venues	"523"^^xsd:integer
3	ABox:Paper	"490"^^xsd:integer
4	ABox:Review	"490"^^xsd:integer
5	ABox:Conferences	"362"^^xsd:integer
6	ABox:Publication	"354"^^xsd:integer
7	ABox:Author	"302"^^xsd:integer
8	ABox:Symposiums	"231"^^xsd:integer
9	ABox:Workshops	"229"^^xsd:integer
10	ABox:Regular_conferences	"223"^^xsd:integer
11	ABox:Expert_groups	"222"^^xsd:integer
12	ABox:Proceeding	"181"^^xsd:integer
13	ABox:Journal	"161"^^xsd:integer
14	ABox:Reviewer	"151"^^xsd:integer
15	ABox:Journal_paper	"119"^^xsd:integer
16	ABox:Full_paper	"111"^^xsd:integer
17	ABox:Volume	"110"^^xsd:integer
18	ABox:Demo_paper	"94"^^xsd:integer
19	ABox:Short_paper	"88"^^xsd:integer
20	ABox:Poster	"78"^^xsd:integer
21	ABox:Handlers	"25"^^xsd:integer
22	ABox:Chairs	"18"^^xsd:integer
23	ABox:Area	"10"^^xsd:integer
24	ABox:Editors	"7"^^xsd:integer

Figure 7: Full Output from SPARQL Query of Instances in the Main Classes

A.3.3 Number of Triples Using the Main Properties

```
PREFIX : <http://graph.org/abox/>

SELECT ?property (COUNT(*) as ?count) WHERE {
  ?s ?property ?o .
  FILTER(STRSTARTS(STR(?property), "http://graph.org/abox/"))
}
GROUP BY ?property ORDER BY DESC(?count)
```

	property	count
1	ABox:submits	"2196"^^xsd:integer
2	ABox:accepts_or_rejects	"2196"^^xsd:integer
3	ABox:writes	"1969"^^xsd:integer
4	ABox:assigns	"1616"^^xsd:integer
5	ABox:text_date	"980"^^xsd:integer
6	ABox:pertains_to	"875"^^xsd:integer
7	ABox:text_name	"844"^^xsd:integer
8	ABox:text_valid	"490"^^xsd:integer
9	ABox:submitted	"481"^^xsd:integer
10	ABox:accepted_as	"354"^^xsd:integer
11	ABox:has	"354"^^xsd:integer
12	ABox:published_in	"354"^^xsd:integer
13	ABox:Name	"302"^^xsd:integer
14	ABox:conference_city	"181"^^xsd:integer
15	ABox:conference_name	"181"^^xsd:integer
16	ABox:conference_time	"181"^^xsd:integer
17	ABox:handled_by1	"159"^^xsd:integer
18	ABox:volumeID	"110"^^xsd:integer
19	ABox:Date	"110"^^xsd:integer
20	ABox:journal_name	"51"^^xsd:integer
21	ABox:handled_by2	"47"^^xsd:integer
22	ABox:area	"10"^^xsd:integer

Figure 8: Full Output from SPARQL Query of Triples Using the Main Properties