

# Semantic Data Management

## *Lab 3: Knowledge Graphs*



**Submitted to:**

Prof. Oscar Romero

SDM - [Semantic Data Management](#)

**Submitted by:**

Dwi Prasetyo Adi NUGROHO

Kunal ARORA

## **CONTENTS**

1. Exploring Dbpedia
2. Analytical queries on top of QBAirbase
3. Ontology creation
  - 3.1 TBOX definition
  - 3.2 ABOX definition
  - 3.3 Linking ABOX to TBOX
  - 3.4 Queries on top of the Ontology
4. References

## A. Exploring Dbpedia

---

- 1. Find a class representing an Actor in the dataset (using Filters)

```
PREFIX owl:<http://www.w3.org/2002/07/owl#>
select (?x as ?class), (?y as ?label)
Where
{
    ?x rdf:type owl:Class .
    ?x rdfs:label ?y
    FILTER(?y = "actor"@es)
}
```

- 2. Find the superclass for the class Actor

```
PREFIX owl:<http://www.w3.org/2002/07/owl#>
select (?x as ?SubClass), (?z as ?SuperClass)
Where
{
    ?x rdf:type owl:Class .
    ?x rdfs:label ?y
    FILTER(?y = "actor"@es)
    ?x rdfs:subClassOf ?z
}
```

- 3. Find all the actors in the dataset

```
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX dbo:<http://dbpedia.org/ontology/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT (?x as ?ActorResource), (?y as ?ActorLabel)
WHERE
{
    ?x rdf:type dbo:Actor .
    ?x rdfs:label ?y .
    FILTER (langMatches(lang(?y), "en"))
}
```

- **4. Get different classes that are defined as range of the properties that have the class Actor defined as their domain.**

```
SELECT distinct(?b) WHERE
{
  ?a rdfs:domain dbo:Actor .
  ?a rdfs:range ?b
}
```

- **5. Find the super property of the goldenRaspberryAward property**

```
SELECT (?x as ?subProperty), (?z as ?superProperty)
WHERE {
  ?x a rdf:Property .
  ?x rdfs:label ?y .
  FILTER(regex(?y, "Golden Raspberry Award", "i")) .
  ?x rdfs:subPropertyOf ?z
}
```

- **6. Return all the properties that have the class Actor as either their range or domain**

```
PREFIX dbo:<http://dbpedia.org/ontology/>
select *
where
{
  {
    ?x a rdf:Property .
    ?x rdfs:domain dbo:Actor
  }
  UNION
  {
    ?x rdfs:range dbo:Actor
  }
}
```

- 7. Return all persons that are not actors.

```
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX dbo:<http://dbpedia.org/ontology/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT ?x
WHERE
{
  ?x rdf:type dbo:Person .
  MINUS { ?x rdf:type dbo:Actor }
}
```

## B. Analytical queries on top of QBAirbase

---

Endpoint: <http://lod.cs.aau.dk:8891/sparql>

1. List the country, station type, latitude, and longitude details of each station.

```
PREFIX schema: <http://qweb.cs.aau.dk/airbase/schema/>
PREFIX property: <http://qweb.cs.aau.dk/airbase/property/>

SELECT STR(?stationName) as ?stationName, STR(?type) as
?stationType, ?lat, ?long, STR(?countryName) as ?countryName
WHERE {
  ?station property:station ?stationName .
  ?station property:longitudeDegree ?long .
  ?station property:latitudeDegree ?lat .
  ?station property:type ?type .
  ?station schema:inCountry ?country .
  ?country property:country ?countryName

}
```

2. List the 10 highest averages of C6H6 emission and the country and the year on which they were recorded.

```
PREFIX schema: <http://qweb.cs.aau.dk/airbase/schema/>
PREFIX property: <http://qweb.cs.aau.dk/airbase/property/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?c6h6 as ?C6H6emision, ?yearNum as ?year, STR(?countryName) as
?countryName WHERE {
  ?obs schema:C6H6 ?c6h6 .
  ?obs schema:station ?station .
}
```

```

?obs schema:year ?year .
?station schema:inCountry ?country .
?country property:country ?countryName .
?year property:yearNum ?yearNum .
?obs schema:sensor ?sensor .
?sensor property:statisticShortName "Mean"^^xsd:string .
}
ORDER BY DESC(?c6h6)
LIMIT 10

```

**3. For each city and property type, give the yearly average emission for NO2, SO2, PB, and PM10.**

```

PREFIX schema: <http://qweb.cs.aau.dk/airbase/schema/>
PREFIX property: <http://qweb.cs.aau.dk/airbase/property/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT STR(?cityName) as ?cityName,
       STR(?type) as ?type,
       ?yearNum as ?year,
       avg(?no2) as ?avgno2,
       avg(?so2) as ?avgso2,
       avg(?pb) as ?avgpb,
       avg(?pm10) as ?avgpm10
WHERE {
  ?obsNo2 schema:NO2 ?no2 .
  ?obsNo2 schema:station ?station .
  ?obsNo2 schema:year ?year .

  ?obsSo2 schema:S02 ?so2 .
  ?obsSo2 schema:station ?station .
  ?obsSo2 schema:year ?year .

  ?obsPb schema:Pb ?pb .
  ?obsPb schema:station ?station .
  ?obsPb schema:year ?year .

  ?obsPm10 schema:PM10 ?pm10 .
  ?obsPm10 schema:station ?station .
  ?obsPm10 schema:year ?year .

  ?station schema:inCity ?city .
  ?station property:type ?type .
  ?city property:city ?cityName .
  ?year property:yearNum ?yearNum .

  ?obsNo2 schema:sensor ?sensorNo2 .
  ?obsSo2 schema:sensor ?sensorSo2 .

```

```

?obsPb schema:sensor ?sensorPb .
?obsPm10 schema:sensor ?sensorPm10 .

?sensorNo2 property:statisticShortName "Mean"^^xsd:string .
?sensorSo2 property:statisticShortName "Mean"^^xsd:string .
?sensorPb property:statisticShortName "Mean"^^xsd:string .
?sensorPm10 property:statisticShortName "Mean"^^xsd:string .

}
GROUP BY ?type ?cityName ?yearNum
ORDER BY ?cityName ?type ?year

```

**4. Define 3 additional SPARQL queries (and their corresponding interpretation) that you think could be interesting for the domain of analyzing air quality/pollution.**

**i. Average station's age per country**

This query will help us to know which country that have old station, and may need to be renewed.

```

PREFIX schema: <http://qweb.cs.aau.dk/airbase/schema/>
PREFIX property: <http://qweb.cs.aau.dk/airbase/property/>

SELECT str(?countryName), avg(?age) as ?avgAge WHERE {
    ?station schema:inCountry ?country .
    ?country property:country ?countryName .
    ?station property:establishedDate ?dateEstablished .
    ?station property:shutDownDate ?dateShutDown .
    bind( year(?dateShutDown) - year(?dateEstablished) -
if(month(?dateShutDown)<month(?dateEstablished) ||
(month(?dateShutDown)=month(?dateEstablished) &&
day(?dateShutDown)<day(?dateEstablished)),1,0) as ?age )
}
GROUP BY ?countryName ORDER BY ?countryName

```

**ii. Cities and year where annual average NO2 is higher than the limit set by EU Air Quality Directive**

(<https://www.eea.europa.eu/themes/air/air-quality-standards>) after it is set.

This query will help us to identify which country still violate the recommendation after it is enforced. Hence the regulator can give more attention and support to these countries in order to meet the standard.

```

PREFIX schema: <http://qweb.cs.aau.dk/airbase/schema/>
PREFIX property: <http://qweb.cs.aau.dk/airbase/property/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

```

```

SELECT STR(?cityName), ?yearNum, AVG(?no2) AS ?avgco WHERE {
  ?obs schema:NO2 ?no2 .
  ?obs schema:station ?station .
  ?obs schema:sensor ?sensor .
  ?obs schema:year ?year .

  ?year property:yearNum ?yearNum .
  ?station schema:inCity ?city .
  ?city property:city ?cityName .

  ?sensor property:statisticShortName "Mean"^^xsd:string

  FILTER(?no2 > 40)
  FILTER(?yearNum > 2008)
}
GROUP BY ?cityName ?yearNum ORDER BY ?cityName ?yearNum

```

### iii. Estimated number of people affected by CO pollution in United Kingdom.

By knowing estimated number of affected people, we can predict further effect of the pollution. It also be a measure of a success or failure of action when trying to reduce CO pollution.

```

PREFIX schema: <http://qweb.cs.aau.dk/airbase/schema/>
PREFIX property: <http://qweb.cs.aau.dk/airbase/property/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX yago: <http://yago-knowledge.org/resource/>

SELECT STR(?cityName) AS ?cityName, AVG(?co) AS ?avgco,
avg(?pop) AS ?population WHERE {
  ?obs schema:CO ?co .
  ?obs schema:station ?station .
  ?obs schema:sensor ?sensor .
  ?obs schema:year ?year .

  ?year property:yearNum ?yearNum .
  ?station schema:inCity ?city .

  ?city schema:locatedIn
  <http://qweb.cs.aau.dk/airbase/data/country/United_Kingdom/> .
  ?city owl:sameAs ?cityDb .
  ?city property:city ?cityName

SERVICE <https://linkeddata1.calcul.u-psud.fr/sparql> {

```



```

    ?cityDb yago:hasNumberOfPeople ?pop
  }

  ?sensor property:statisticShortName "Mean"^^xsd:string
  FILTER (?yearNum = 2012)
}
GROUP BY ?cityName ORDER BY ?cityName
LIMIT 100

```

## C. Ontology creation

### C.1 TBOX definition

The TBOX was created using the Protégé tool (<https://protege.stanford.edu/>). The reason to use Protégé as our TBOX creation tool is because, it is very simple, easy and straightforward with clear explanations and examples. Protégé has also a wide community support. There's a forum dedicated to protégé users. In addition, there are tutorial examples to guide you on your first ontology.

Further details of different classes that were extended from owl:Thing class to make research publication domain ontology as complete as possible is as below:

- Person:** The person concept is reused from the already existing foaf ontology ([http://xmlns.com/foaf/spec/#term\\_Person](http://xmlns.com/foaf/spec/#term_Person)). The **Author** class was extended from the Person class as an Author shares the same kind of properties as Person ontology. **Author** class specify the first name and last name.  
[Author-subClassOf->Person](#)
- Organization:** The organization concept is also reused from the already existing foaf ontology ([http://xmlns.com/foaf/spec/#term\\_Organization](http://xmlns.com/foaf/spec/#term_Organization)). There are two classes **Company** and **University** which are extended from organization class. **Company** class specify the name of the company and it's industry. **University** class specify the name of the university and it's homepage.  
[Company-subClassOf->Organization](#)  
[University-subClassOf->Organization](#)
- Document:** The document concept is also reused from the already existing foaf ontology ([http://xmlns.com/foaf/spec/#term\\_Document](http://xmlns.com/foaf/spec/#term_Document)). The **Paper** class is extended directly from the Document class as a research paper also signifies a document. The **DemoPaper**, **SurveyPaper**, **ShortPaper** and **FullPaper** are further extended from the Paper class to specify the specific properties of these different type of research papers.  
[Paper-subClassOf->Document](#)  
[DemoPaper-subClassOf->Paper](#)  
[SurveyPaper-subClassOf->Paper](#)

ShortPaper-subClassOf->Paper

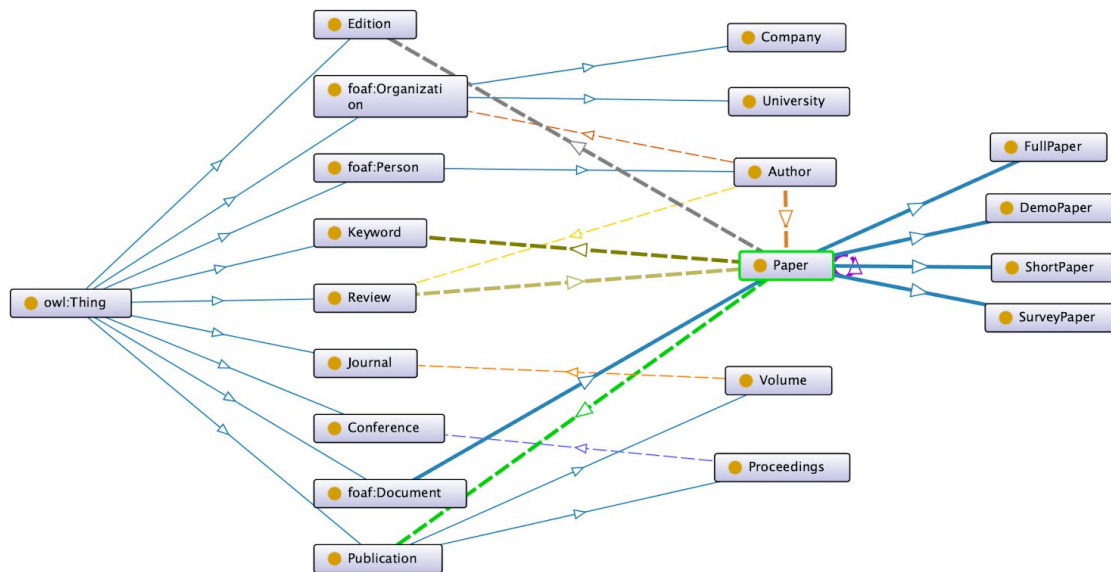
FullPaper-subClassOf->Paper

- **Keyword:** The keyword class is a newly created concept for the research publication ontology domain. This class consists of various keywords and links them with a paper.
- **Review:** The review class is also a newly created concept for the research publication ontology domain. The review class consists of comment and decision that an author provides regarding the paper he/she reviews.
- **Conference:** The conference class is also a newly created concept for the research publication ontology domain. The conference signifies a formal meeting of people with a shared interest in the real world sense. In this ontology, it consists of name of the conference and the publisher details.
- **Edition:** The edition class is also a newly created concept for the research publication ontology domain. A conference usually have several editions so, it is linked to conference class to specify which edition belongs to which conference. The edition also have it's unique edition number for a specific conference, the year when it was held and the venue details.
- **Journal:** The journal class is also a newly created concept for the research publication ontology domain. The journal signifies a newspaper or magazine that deals with a particular subject or professional activity in the real world sense. In this ontology, it consists of name of the journal and the publisher details.
- **Publication:** The publication class is also a newly created concept for the research publication ontology domain. A publication usually refers to the published material out of a conference or journal in the real world sense. There are two subclasses **Volume** and **Proceedings** from this publication class. **Volume** class is also a newly created class and specify the volume number and the journal it belongs to.  
**Proceedings** class is also a newly created class and is linked to conference. It usually consists of all the papers that were presented in an edition of a conference.

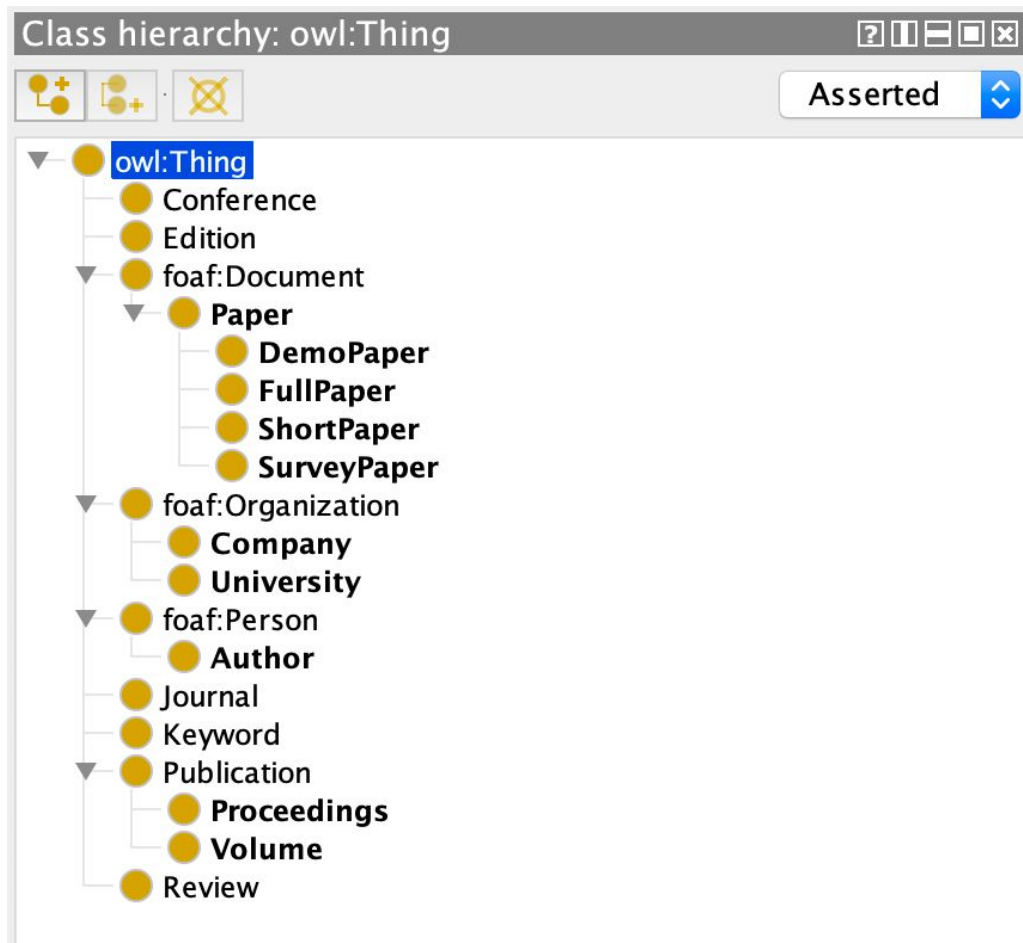
Volume-subClassOf->Publication

Proceedings-subClassOf->Publication

The visual representation of TBOX created using the Protégé tool and creating the above classes is as follows:



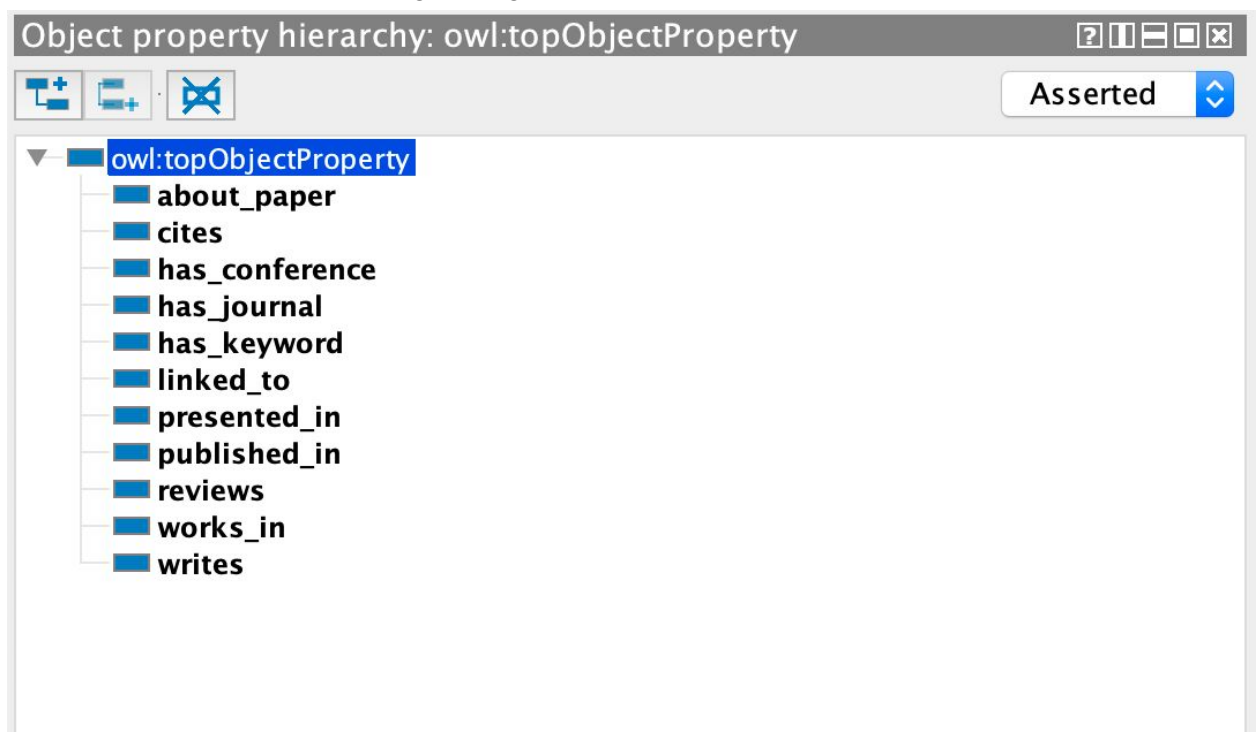
Information about the class hierarchy created using the Protégé tool is as follows:



In the above classes, we had some disjoint classes as well based on our assumptions:

- Proceedings and Volume are disjoint because we assumed a publication can either be in a Volume of a Journal or Proceedings of a Conference. We assumed that the same research paper cannot be published in a Journal or presented at the conference. It is either of those two.
- Company and University are disjoint as well. An author is either teaching or doing research in the university or working and doing research in a company (could be a research lab).

These classes are linked to each other through various object properties. These Object properties are also defined using Protégé tool and are as shown below:



The object properties or relationship between different classes are extended from the generic owl:topObjectProperty. The details of different object properties is as follows:

- **about\_paper:** This property defines the relationship between Review and Paper class. The review provided by an Author is in the Review class and that review relates to a certain paper. The domain of this property is Review and range is Paper.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/reviews">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Author"/>
```

```
    <rdfs:range rdf:resource="http://www.semanticweb.org/Review"/>
</owl:ObjectProperty>
```

- **cites:** This property defines the relationship of one Paper to another Paper. A Paper can have one or more citations. Both the domain and range is Paper class.

```
<owl:ObjectProperty rdf:about="http://www.semanticweb.org/property/cites">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Paper"/>
</owl:ObjectProperty>
```

- **has\_conference:** This property defines the relationship between Edition and Conference class. An edition is always a part of certain conference. The domain and range of this property are Edition and Conference respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/has_conference">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Edition"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Conference"/>
</owl:ObjectProperty>
```

- **has\_journal:** This property related the Volume and Journal class. A Volume also belongs to a certain Journal. The domain and range of this property are Volume and Journal respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/has_journal">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Volume"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Journal"/>
</owl:ObjectProperty>
```

- **has\_keyword:** This property is used to link a Paper with its various keywords that can help identify the type of paper or its community. The domain and range of this property are Paper and Keyword class respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/has_keyword">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Keyword"/>
</owl:ObjectProperty>
```

- **linked\_to:** This property links the Proceedings to its associated Conference. The domain and range of this property are Proceedings and Conference class respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/linked_to">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Proceedings"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Conference"/>
</owl:ObjectProperty>
```

- **presented\_in:** This property links the Paper to an Edition class. It represents those Paper that are presented in an Edition of some conference. The domain and range are Paper and Edition class respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/presented_in">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Edition"/>
</owl:ObjectProperty>
```

- **published\_in:** This property links the Paper to a Publication class. It represents those Paper that are published either in the Volume of some Journal or in the Proceedings of some Conference. The domain and range are Paper and Publication class respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/published_in">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Publication"/>
</owl:ObjectProperty>
```

- **reviews:** This property links the Review to an Author. It relates a review provided by an Author. The domain and range are Author and Review class respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/reviews">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Author"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Review"/>
</owl:ObjectProperty>
```

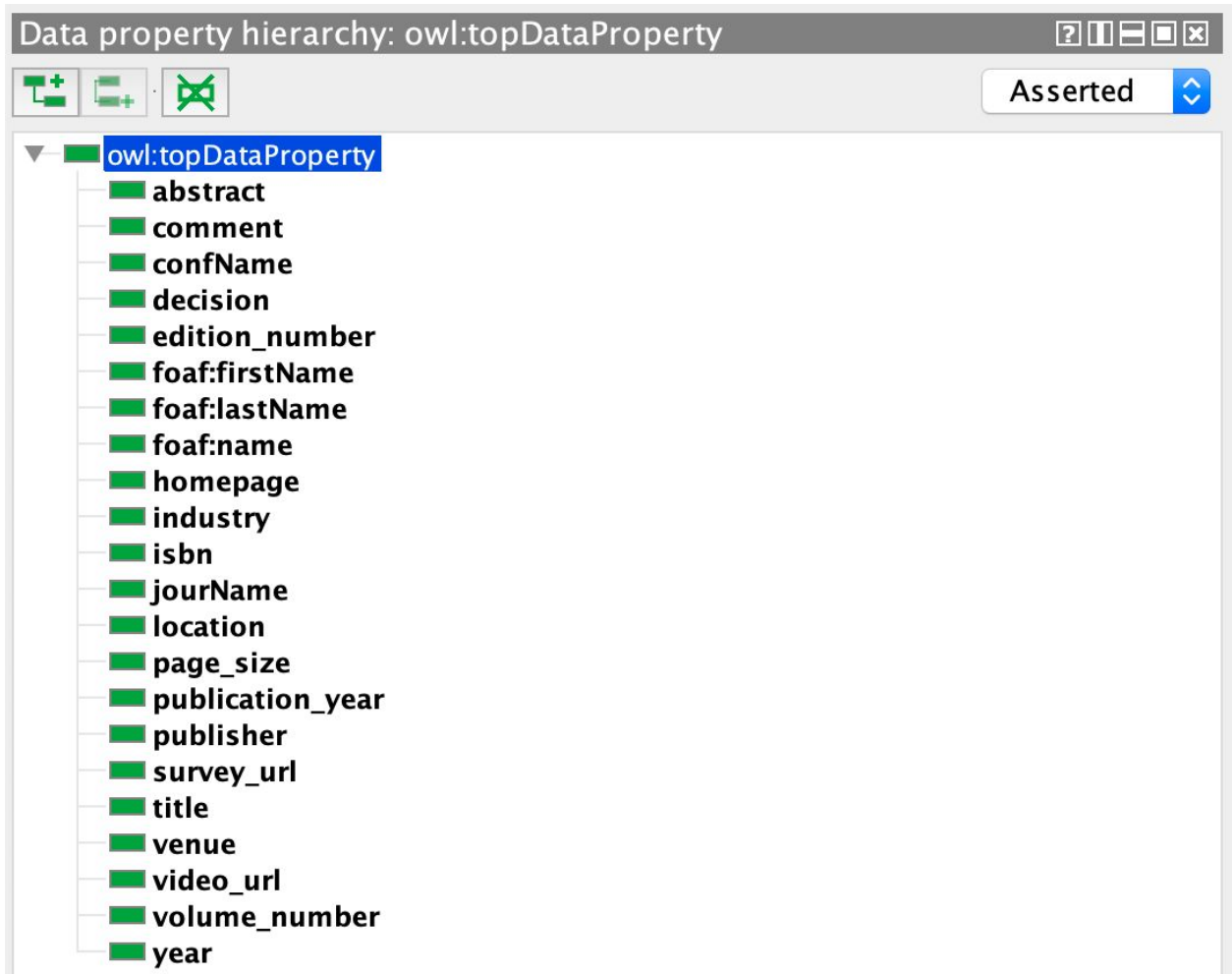
- **works\_in:** This property relates an Author to an Organization class. It links either the Company or University that Author is working at currently. The domain and range are Author and Organization class respectively.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/property/works_in">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Author"/>
    <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
</owl:ObjectProperty>
```

- **writes:** This property is a link between Author and Paper class. It signifies all the Papers that an Author has written. The domain and range are Author and Paper respectively.

```
<owl:ObjectProperty rdf:about="http://www.semanticweb.org/property/writes">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Author"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/Paper"/>
</owl:ObjectProperty>
```

The classes also consists of various data properties that defines the content or label information. These datatype properties are also defined using Protégé tool and are as shown below:



The data properties are extended from the generic owl:topDataProperty. The details of various data properties are as follows:

- **abstract:** This property specifies the abstract of information from the Proceedings or Paper class. It has two domain values Paper and Proceedings. The range is xsd:string.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/abstract">
```

```
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
```

```
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Proceedings"/>
```

```
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```



- **comment:** This property consists of the comments given by the Author for a certain Paper. The domain and range are Review and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/comment">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Review"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **confName:** This property defines the name of the conference. The domain and range are Conference and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/confName">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Conference"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **decision:** The property specifies the decision of an Author on the Paper he/she has reviewed. Currently, in the research publication ontology graph we have respective values for decision, *“Approved”*, *“Rejected”*, *“InProgress”* and *“InHalt”*. The domain and range are Review and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/decision">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Review"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **edition\_number:** This property defines the number of the Edition of a particular Conference. The domain and range are Edition and xsd:integer respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/edition_number">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Edition"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty>
```

- **homepage:** The property defines the homepage url of University. The domain and range are University and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/homepage">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/University"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```



- **industry:** The property defines the specific industry of Company. The domain and range are Company and xsd:string respectively.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/industry">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Company"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
```

- **isbn:** This property defines the unique isbn number of Paper. The domain and range are Paper and xsd:string respectively.

```
<owl:DatatypeProperty rdf:about="http://www.semanticweb.org/property/isbn">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **jourName:** This property defines the name of the Journal. The domain and range are Journal and xsd:string respectively.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/jourName">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Journal"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
```

- **location:** This property defines the location where an Organization either a Company or University is situated. The domain and range are Organization and xsd:string respectively.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/location">
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
```

- **page\_size:** This property contains the size of the paper. This property is specific to ShortPaper and FullPaper class. The domain is FullPaper and ShortPaper. The range is xsd:integer.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/page_size">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/FullPaper"/>
    <rdfs:domain rdf:resource="http://www.semanticweb.org/ShortPaper"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
  </owl:DatatypeProperty>
```

- **publication\_year:** This property defines the year in which the research paper was published. The domain and range are Paper and xsd:integer respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/publication_year">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty>
```

- **publisher:** This property defines the publisher of a Conference or Journal. The domain is Conference and Journal. The range is xsd:string.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/publisher">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Conference"/>
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Journal"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **survey\_url:** This property defines the url of the survey that is conducted for the research paper. This property is specific to SurveyPaper. The domain and range are SurveyPaper and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/survey_url">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/SurveyPaper"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **title:** This property defines the title of a research Paper. The domain and range are Paper and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/title">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Paper"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **venue:** This property defines the city or country where the Edition of some Conference took place. The domain and range are Edition and xsd:string respectively.

```
<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/property/venue">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Edition"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **video\_url:** This property defines the url or the video link which is kind of demo for the research paper. The property is specific to DemoPaper class. The domain and range are DemoPaper and xsd:string respectively.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/video_url">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/DemoPaper"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
```

- **volume\_number:** This property defines the number of the Volume of a particular Journal. The domain and range are Volume and xsd:integer respectively.

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/property/volume_number">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/Volume"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
  </owl:DatatypeProperty>
```

- **year:** This property defines the year in which the Edition of a Conference happened. The domain and range are Edition and xsd:integer respectively.

```
<owl:DatatypeProperty rdf:about="http://www.semanticweb.org/property/year">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Edition"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty>
```

- **foaf:firstName:** This property is re-used from the foaf ontology ([http://xmlns.com/foaf/spec/#term\\_firstName](http://xmlns.com/foaf/spec/#term_firstName)). This property defines the first name of a Person. The domain and range are Person and xsd:string respectively.

```
<owl:DatatypeProperty rdf:about="http://xmlns.com/foaf/0.1/firstName">
  <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **foaf:lastName:** This property is re-used from the foaf ontology ([http://xmlns.com/foaf/spec/#term\\_lastName](http://xmlns.com/foaf/spec/#term_lastName)). This property defines the last name of a Person. The domain and range are Person and xsd:string respectively.

```
<owl:DatatypeProperty rdf:about="http://xmlns.com/foaf/0.1/lastName">
  <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

- **foaf:name:** This property is re-used from the foaf ontology ([http://xmlns.com/foaf/spec/#term\\_name](http://xmlns.com/foaf/spec/#term_name)). This property defines the name of Keyword and Organization. The domain is Keyword and Organization. The range is xsd:string.

```
<owl:DatatypeProperty rdf:about="http://xmlns.com/foaf/0.1/name">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/Keyword"/>
  <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

## C.2 ABOX definition

We create the ABOX using DBLP publication data same as Assignment of Lab 1. We only take the data that are relevant to our TBOX. We also use additional synthetic data such as using randomization, dummy values, and external data when it is not available in the DBLP dataset. The process is done using **Apache Jena API**.

For each TBOX class that we want to instantiate, we create the ABOX using the following steps:

1. Read the CSV.
2. Create a Jena RDF Model.
3. Get the respective column from the CSV that we want to use as property.
4. Create URI using one of the column in the CSV (e.g name for author, title for conference, DBLP paper key for paper). Note that we need to avoid the use of space, and replace it with underscore ("\_").
5. Put the property in the model using add property.
  - a. Use predefined property such as FOAF when possible,
  - b. Otherwise, create a new property using addProperty()
  - c. When property is an objectProperty, we need to pass a Jena Resource data type as the second parameter of addProperty. Therefore we need to instantiate it using model.createResource()
6. Save the created model by redirecting the write function of the Jena model to an output file.

Below are an example of creating ABOX for volume:

```
// title, volume, year
Model model = ModelFactory.createDefaultModel();

// read the csv line by line
BufferedReader csvReader = new BufferedReader(new FileReader(Config.JOURNAL_PATH));
String row;
while ((row = csvReader.readLine()) != null) {
    String[] row_data = row.split(",");

    String volumeUri = Config.RESOURCE_URL+row_data[0].replace(" ", "_") + "_Volume_"
+ row_data[1];
    String journalUri = Config.RESOURCE_URL+row_data[0].replace(" ", "_");
    Resource currentVolume = model.createResource(volumeUri)
        .addProperty(model.createProperty(Config.PROPERTY_URL+"volume_number"),
row_data[1])

    .addProperty(model.createProperty(Config.PROPERTY_URL+"has_journal"),model.createRes
ource(journalUri));
}
csvReader.close();

model.write(new PrintStream(
    new BufferedOutputStream(
        new FileOutputStream(Config.OUTPUT_PATH+"volume.nt")), true), "NT");
```

The complete code consist of 11 functions that create ABOX for concepts in our TBOX. We decided to use ntriples (nt) as the format of our ABOX. After its execution, the code will produce 11 output files: company.nt, conference.nt, edition.nt, journal.nt, keyword.nt, paper.nt, person.nt, proceeding.nt, reviews.nt university.nt and volume.nt.

## C.3 Linking ABOX to TBOX

---

### Uploading Data on Virtuoso

1. First we need to upload our TBOX and ABOX to the virtuoso server.
2. Upload TBOX,
  - a. Save the TBOX created in Protégé as a “*research\_publication.owl*” file.
  - b. We also create “*research\_publication.owl.graph*” file with following content, <http://localhost:8890/research>
  - c. Then, put these 2 files at the

/usr/local/virtuoso-opensource/share/virtuoso/vad  
location in the cloud server machine provided to us.

- d. Finally, we run the rdf loader inside virtuoso server under the Database/InteractiveSQL section,

```
ld_dir('/usr/local/virtuoso-opensource/share/virtuoso/vad',  
'research_publication.owl','http://localhost:8890/research');  
rdf_loader_run();
```

3. Upload ABOX,

- a. Save all the 11 files created using Apache Jena from section C.2 as “.nt” files.  
b. Create their respective 11 “.nt.graph” files as we did for TBOX with the following content, <http://localhost:8890/research>

C. Then put all these 22 files at the

/usr/local/virtuoso-opensource/share/virtuoso/vad  
location in the cloud server machine provided to us.

- D. Finally, we run the rdf loader inside virtuoso server under the Database/InteractiveSQL section,

```
ld_dir ('/usr/local/virtuoso-opensource/share/virtuoso/vad/data',  
'*.nt','http://localhost:8890/research') ;  
rdf_loader_run();
```

4. We can confirm that our data has been loaded by going to Virtuoso dashboard, then go to Linked Data, Graph, and click on Graph Tab. The newly created database “research” should show on the list. We can also try to query our loaded database from the sparql endpoint. **Go to localhost:8890/sparql**, then put **localhost:8890/research** on the database and run the following query:

```
SELECT COUNT (*) WHERE { ?a ?b ?c . }
```

## Linking TBOX and ABOX

The TBOX and ABOX are linked using the **INSERT** statement of Sparql for all the below cases:

1. To link ABOX and TBOX of Author, run the following SPARQL query:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>  
PREFIX base: <http://www.semanticweb.org/>  
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
  
Insert {  
  ?a rdf:type base:Author  
}
```

```
WHERE {  
  ?a foaf:lastName ?b  
}
```

2. To link ABOX and TBOX of Paper, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>  
PREFIX pro: <http://www.semanticweb.org/property/>  
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
  
INSERT {  
  ?a rdf:type base:Paper  
}  
WHERE {  
  ?a pro:isbn ?b  
}
```

3. To link ABOX and TBOX of University, run the following SPARQL query:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>  
PREFIX base: <http://www.semanticweb.org/>  
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
  
INSERT {  
  ?a rdf:type base:University  
}  
WHERE {  
  ?a foaf:homepage ?b  
}
```

4. To link ABOX and TBOX of Company, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>  
PREFIX pro: <http://www.semanticweb.org/property/>  
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
  
INSERT {  
  ?a rdf:type base:Company  
}  
WHERE {  
  ?a pro:industry ?b  
}
```

5. To link ABOX and TBOX of Keyword, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Keyword
}
WHERE {
  ?a pro:keyword ?b
}
```

6. To link ABOX and TBOX of Edition, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Edition
}
WHERE {
  ?a pro:edition_number ?b
}
```

7. To link ABOX and TBOX of Proceedings, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Proceedings
}
WHERE {
  ?a pro:linked_to ?b
}
```



8. To link ABOX and TBOX of Volume, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Volume
}
WHERE {
  ?a pro:volume_number ?b
}
```

9. To link ABOX and TBOX of Review, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Review
}
WHERE {
  ?a pro:decision ?b
}
```

10. To link ABOX and TBOX of Conference, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Conference
}
WHERE {
  ?a pro:confName ?b
}
```

11. To link ABOX and TBOX of Journal, run the following SPARQL query:

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

INSERT {
  ?a rdf:type base:Journal
}
WHERE {
  ?a pro:jourName ?b
}
```

## Summary Statistic

Below are summary statistic of the graph obtained after linking the ABOX with TBOX:

| Measure                    | Value  |
|----------------------------|--------|
| Total number of classes    | 15     |
| Total number of properties | 38     |
| Total number of instances  | 114359 |
| Total number of triples    | 807157 |

Percentage of number of instance by class:

| Class      | Percentage of Total Instance |
|------------|------------------------------|
| Author     | 14,12%                       |
| Paper      | 18,04%                       |
| Volume     | 0,10%                        |
| Company    | 0,09%                        |
| University | 0,26%                        |
| Conference | 9,05%                        |

|   |        |
|---|--------|
| Edition   | 34,86% |
| Keyword   | 13,60% |
| Review  | 0,80%  |
| Proceedings   | 9,05%  |
| Others (owl#class,<br>owl#DatatypeProperty,<br>owl#ObjectProperty, etc) | 0,05%  |

## C.4 Queries on top of the Ontology

---

### 1. Find all the Authors (With TBOX)

```

PREFIX base: <http://www.semanticweb.org/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

Select ?a as ?Author, ?c as ?FirstName, ?d as ?LastName
WHERE {
    ?a rdf:type base:Author .
    ?a foaf:firstName ?c .
    ?a foaf:lastName ?d
}
```

### 1. Find all the Authors (Without TBOX)

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX pro: <http://www.semanticweb.org/property/>

SELECT ?a as ?Author, ?c as ?FirstName, ?d as ?LastName
WHERE {
    ?a pro:writes ?b .
    ?a foaf:firstName ?c .
    ?a foaf:lastName ?d
}
```

## 2. Find all the properties whose domain is Author (With TBOX)

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

Select *
WHERE {
    ?a rdfs:domain base:Author
}
```

## 2. Find all the properties whose domain is Author (Without TBOX)

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT distinct(?d)
WHERE {
    ?a foaf:lastName ?b .
    ?a ?d ?f
}
```

## 3. Find all the properties whose domain is either Conference or Journal (With TBOX)

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

Select *
WHERE {{
    ?a rdfs:domain base:Journal
}
UNION
{
    ?a rdfs:domain base:Conference
}
}
```

### 3. Find all the properties whose domain is either Conference or Journal (Without TBOX)

```
PREFIX prop: <http://www.semanticweb.org/property/>

SELECT distinct(?c)
WHERE {
  {
    ?a prop:jourName ?b .
    ?a ?c ?d
  }
  UNION
  {
    ?a prop:confName ?b .
    ?a ?c ?d
  }
}
```

### 4. Find all the things that Authors have created (either Reviews or Papers) (With TBOX)

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

Select ?b as ?Author, ?a as ?AuthorCreated, ?c as ?ReviewOrPaper
WHERE {
  {
    ?a rdfs:domain base:Author .
    ?a rdfs:range base:Review .
    ?b ?a ?c
  }
  UNION
  {
    ?a rdfs:domain base:Author .
    ?a rdfs:range base:Paper .
    ?b ?a ?c
  }
}
```

**4. Find all the things that Authors have created (either Reviews or Papers) (Without TBOX)**

```
PREFIX base: <http://www.semanticweb.org/>
PREFIX pro: <http://www.semanticweb.org/property/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

Select ?a as ?Author, ?b as ?Review, ?c as ?PaperWrote
WHERE {
{
    ?a pro:reviews ?b .
}
UNION
{
    ?a pro:writes ?c
}
}
```

## References:

---

- [1]: <https://protegewiki.stanford.edu/wiki/Ontology101>
- [2]: [https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf)
- [3]: <https://protege.stanford.edu/>
- [4]: <https://www.eea.europa.eu/themes/air/air-quality-standards>
- [5]: <http://jena.apache.org/index.html>