ECS735P: The Semantic Web - Final Coursework

Topic - Higher Education Ontology

By: Satyam Sharma

Student Number: 220760793

Introduction to the Ontology

Step 1 – Ontology Design – Creating the T-Box

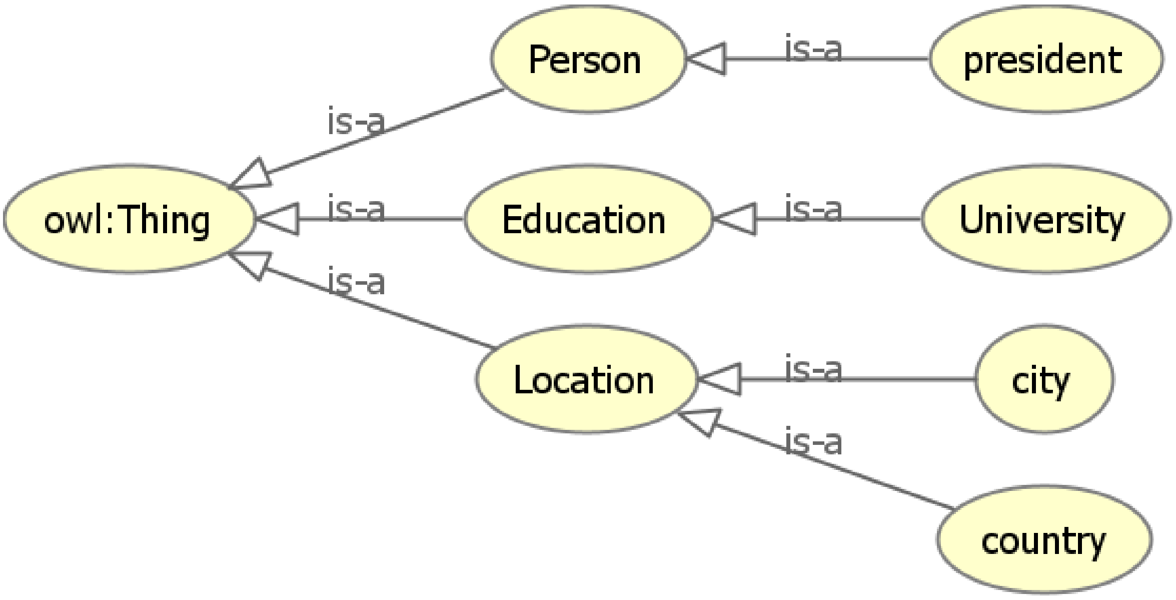
- Defining the Classes – I created a basic ontology which had 7 classes, Education, university, Location, city, country, Person and president. The university class is defined as a subclass of Education class and the city and country classes are defined as subclasses of Location class. Similarly, president class is defined as a subclass of Person class.
- Defining the Object Properties – My ontology has 3 object properties defined using the owl:ObjectProperty property. The three properties are – countryLocated, hasPresident and locatedIn. The countryLocated property has the University class as its domain and the Country class as its range. The hasPresident property has the University class as its domain and the Person class as its range. The locatedIn property has the University class as its domain and the City class as its range.
- Defining the Data Properties – My ontology has 3 data properties defined using the owl:DataTypeProperty property. These properties are founded, motto, and numberOfStudents. The founded property has the University class as its domain and xsd:integer as its range. The motto property has the University class as its domain and xsd:string as its range. The numberOfStudents property has the University class as its domain and xsd:integer as its range.
- The design decisions I made in this ontology follow the basic principles of ontology design, which include:
    - Indentifying and defining classes and their relationships.
    - Identifying and defining properties and their domains and ranges
    - Maintaining the consistency of the ontology.

Step 2 – Why I choose the specific domains and range restrictions for the object properties

- The above-mentioned domains and range restrictions for the object properties helps to ensure that the ontology is consistent and that only valid values can be used for each property.

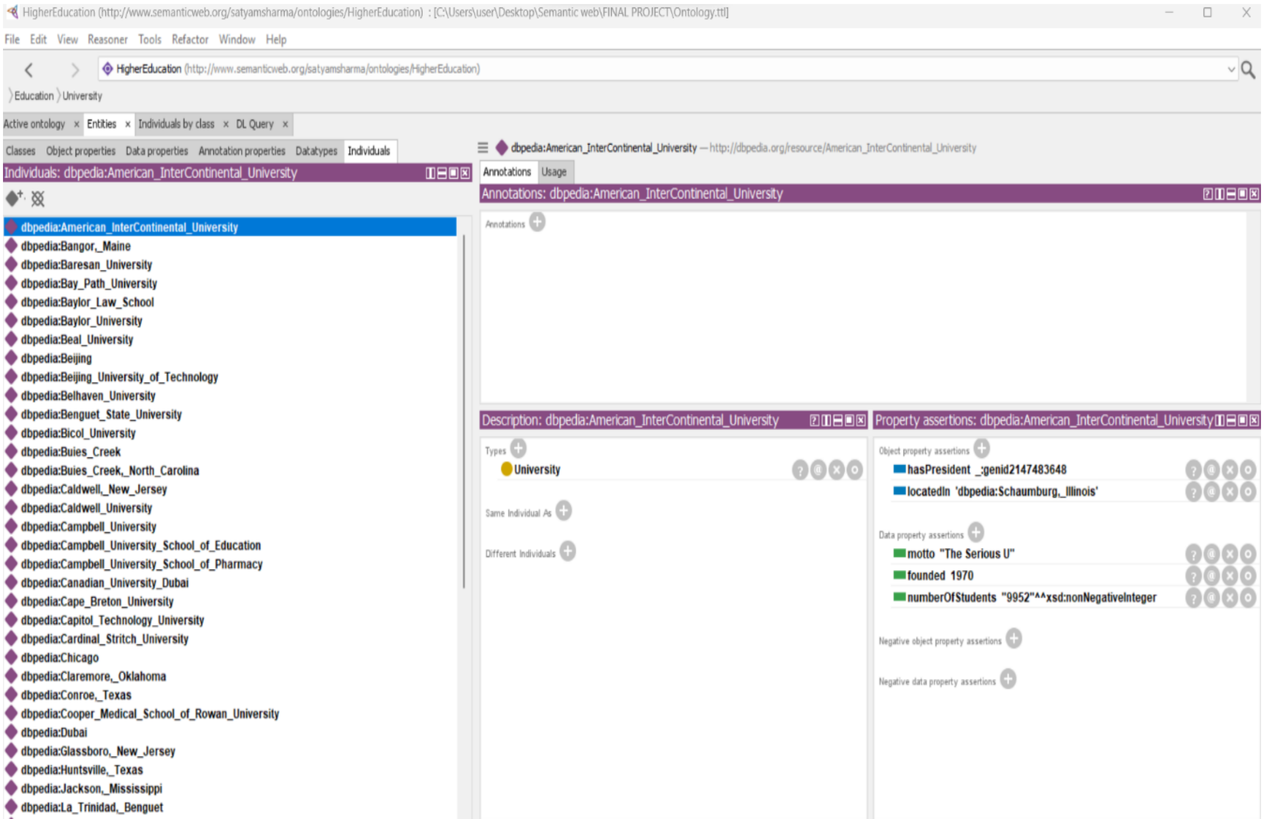
Step 3 – Why I choose the specific domains and range restrictions for the data properties

- The above-mentioned domains and range restrictions for the data properties are appropriate for accurately capturing the attributes of the University class, and help to ensure consistency and validity in the ontology.


owl:Thing
Person
Education
Location
president
University
city
country
is-a

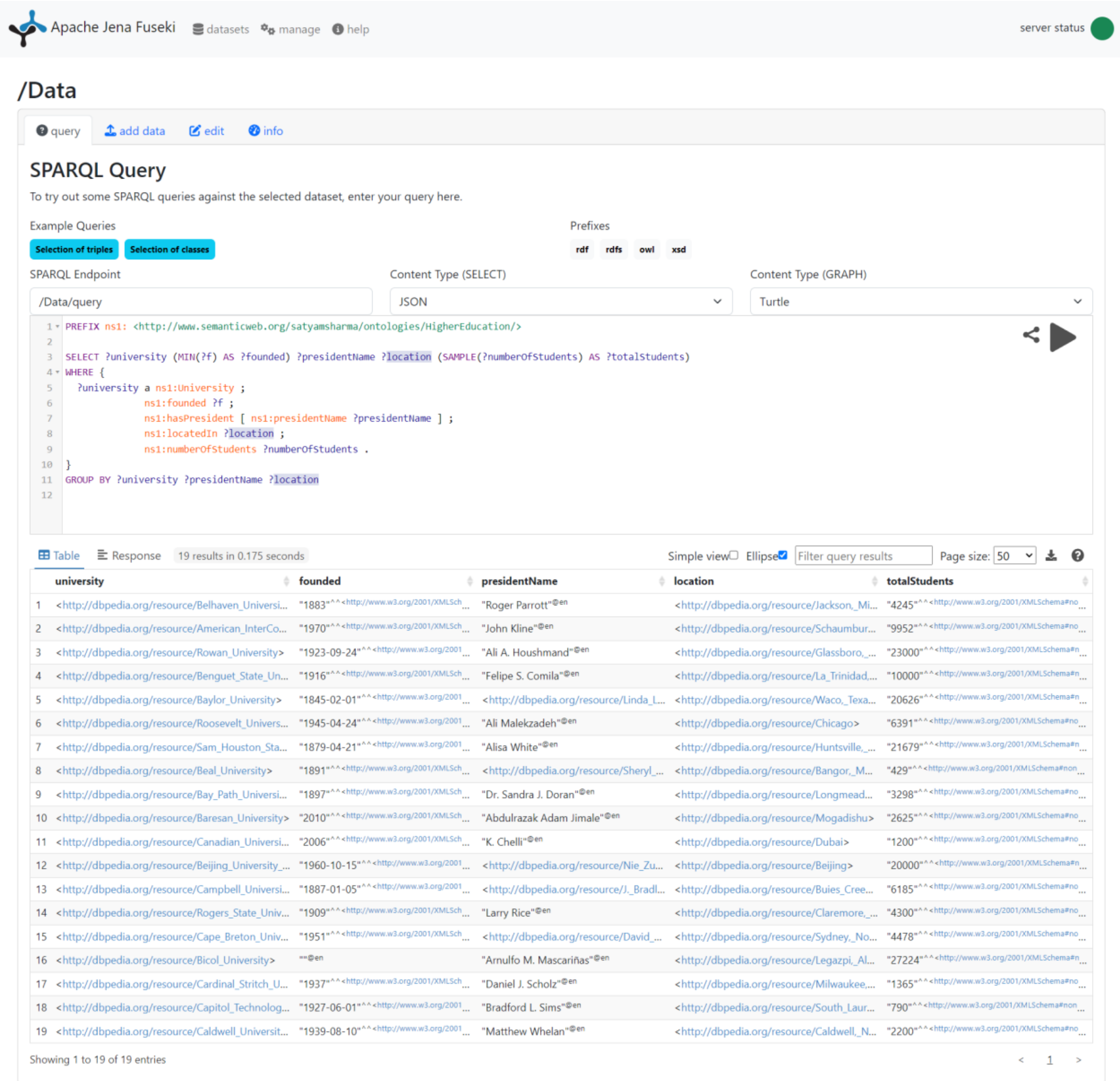
Basic Task

Step 1 – In this task, I have used a Python script which contains a SPARQL query to populate data in an ontology with DBPedia data. In the given Python code added alongside the report, the rdflib and SPARQLWrapper packages are used to load the original ontology, defined a SPARQL query, execute the query, and merge the results into the original graph. The input Turtle file in Task 4 is "file.ttl", which contains the original ontology. In this, the SPARQL query selects 19 universities from the DBpedia endpoint based on their names, and constructs statements for each university that specify its type, country, president, location, number of students, foundation year, and motto. Once the query is executed, the results are merged into the original graph using the += operator, and the updated graph is saved to a new file in Turtle format "Ontology.ttl". The final output of the Python script is a message confirming that the ontology has been populated and saved to the new file.



<u>Step 2 -</u> Now the original data uploaded to the Ontology.ttl file is added to the Fuseki server since I was not able to query via the SPARQL Query tab in Protege because of a Protege issue. The following query is executed on the Fuseki server UI and I have attached the Fuseki query and the query results below to show the obtained results after running the query.

This SPARQL query retrieves information on universities added to the ontology, including their name, founding year, president name, location, and total number of students enrolled. It filters universities that are of type ns1:University and have the necessary properties.



Apache Jena Fuseki   datasets   manage   help   server status

/Data

query   add data   edit   info

SPARQL Query

To try out some SPARQL queries against the selected dataset, enter your query here.

Example Queries: Selection of triples | Selection of classes

Prefixes: rdf | rdfs | owl | xsd

SPARQL Endpoint: /Data/query

Content Type (SELECT): JSON

Content Type (GRAPH): Turtle

```
PREFIX ns1: <http://www.semanticweb.org/satyamsharma/ontologies/HigherEducation/>

SELECT ?university (MIN(?f) AS ?founded) ?presidentName ?location (SAMPLE(?numberOfStudents) AS ?totalStudents)
WHERE {
  ?university a ns1:University ;
              ns1:founded ?f ;
              ns1:hasPresident [ ns1:presidentName ?presidentName ] ;
              ns1:locatedIn ?location ;
              ns1:numberOfStudents ?numberOfStudents .
}
GROUP BY ?university ?presidentName ?location
```

Table   Response   19 results in 0.175 seconds   Simple view   Ellipse   Filter query results   Page size: 50

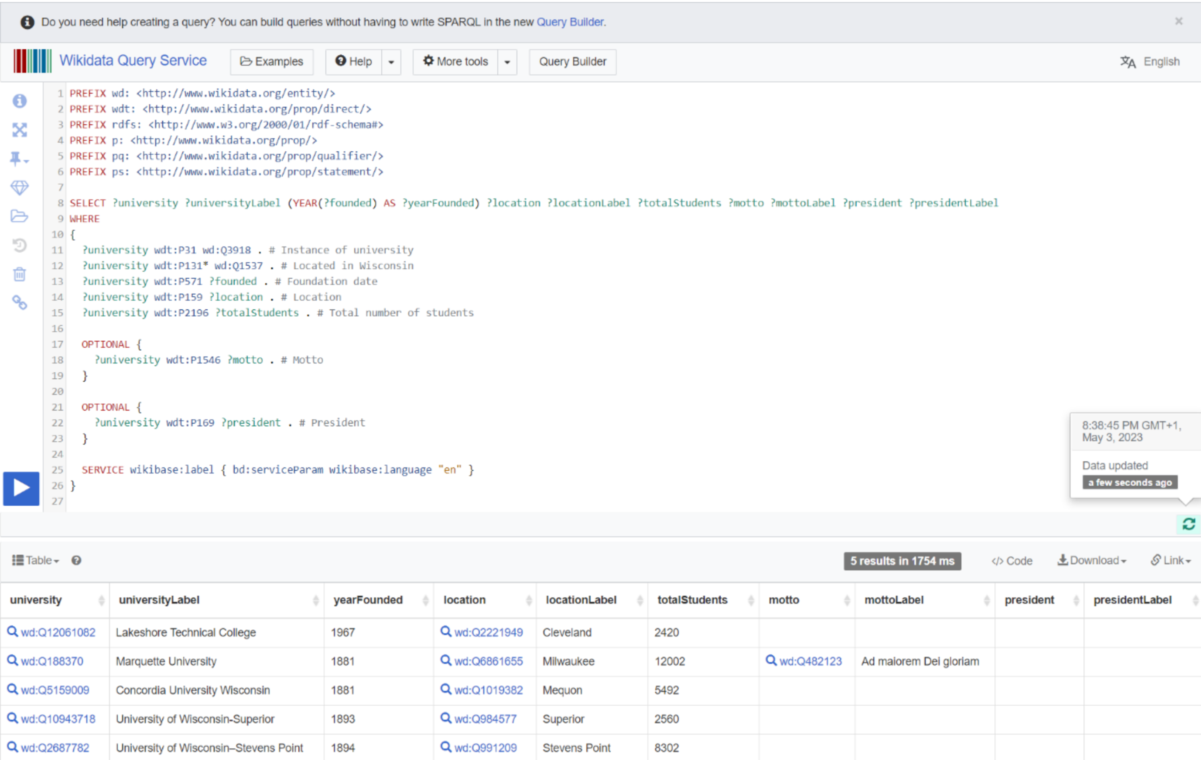
	university	founded	presidentName	location	totalStudents
1	<http://dbpedia.org/resource/Belhaven_Universi...	"1883"^^<http://www.w3.org/2001/XMLSch...	"Roger Parrott"@en	<http://dbpedia.org/resource/Jackson,_Mi...	"4245"^^<http://www.w3.org/2001/XMLSchema#no...
2	<http://dbpedia.org/resource/American_InterCo...	"1970"^^<http://www.w3.org/2001/XMLSch...	"John Kline"@en	<http://dbpedia.org/resource/Schaumbur...	"9952"^^<http://www.w3.org/2001/XMLSchema#no...
3	<http://dbpedia.org/resource/Rowan_University>	"1923-09-24"^^<http://www.w3.org/2001...	"Ali A. Houshmand"@en	<http://dbpedia.org/resource/Glassboro,_...	"23000"^^<http://www.w3.org/2001/XMLSchema#n...
4	<http://dbpedia.org/resource/Benguet_State_Un...	"1916"^^<http://www.w3.org/2001/XMLSch...	"Felipe S. Comila"@en	<http://dbpedia.org/resource/La_Trinidad,...	"10000"^^<http://www.w3.org/2001/XMLSchema#n...
5	<http://dbpedia.org/resource/Baylor_University>	"1845-02-01"^^<http://www.w3.org/2001...	<http://dbpedia.org/resource/Linda_L...	<http://dbpedia.org/resource/Waco,_Texa...	"20626"^^<http://www.w3.org/2001/XMLSchema#n...
6	<http://dbpedia.org/resource/Roosevelt_Univers...	"1945-04-24"^^<http://www.w3.org/2001...	"Ali Malekzadeh"@en	<http://dbpedia.org/resource/Chicago>	"6391"^^<http://www.w3.org/2001/XMLSchema#no...
7	<http://dbpedia.org/resource/Sam_Houston_Sta...	"1879-04-21"^^<http://www.w3.org/2001...	"Alisa White"@en	<http://dbpedia.org/resource/Huntsville,_...	"21679"^^<http://www.w3.org/2001/XMLSchema#n...
8	<http://dbpedia.org/resource/Beal_University>	"1891"^^<http://www.w3.org/2001/XMLSch...	<http://dbpedia.org/resource/Sheryl_...	<http://dbpedia.org/resource/Bangor,_M...	"429"^^<http://www.w3.org/2001/XMLSchema#non...
9	<http://dbpedia.org/resource/Bay_Path_Universi...	"1897"^^<http://www.w3.org/2001/XMLSch...	"Dr. Sandra J. Doran"@en	<http://dbpedia.org/resource/Longmead...	"3298"^^<http://www.w3.org/2001/XMLSchema#no...
10	<http://dbpedia.org/resource/Baresan_University>	"2010"^^<http://www.w3.org/2001/XMLSch...	"Abdulrazak Adam Jimale"@en	<http://dbpedia.org/resource/Mogadishu>	"2625"^^<http://www.w3.org/2001/XMLSchema#no...
11	<http://dbpedia.org/resource/Canadian_Universi...	"2006"^^<http://www.w3.org/2001/XMLSch...	"K. Chelli"@en	<http://dbpedia.org/resource/Dubai>	"1200"^^<http://www.w3.org/2001/XMLSchema#no...
12	<http://dbpedia.org/resource/Beijing_University_...	"1960-10-15"^^<http://www.w3.org/2001...	<http://dbpedia.org/resource/Nie_Zu...	<http://dbpedia.org/resource/Beijing>	"20000"^^<http://www.w3.org/2001/XMLSchema#n...
13	<http://dbpedia.org/resource/Campbell_Universi...	"1887-01-05"^^<http://www.w3.org/2001...	<http://dbpedia.org/resource/J._Bradl...	<http://dbpedia.org/resource/Buies_Cree...	"6185"^^<http://www.w3.org/2001/XMLSchema#no...
14	<http://dbpedia.org/resource/Rogers_State_Univ...	"1909"^^<http://www.w3.org/2001/XMLSch...	"Larry Rice"@en	<http://dbpedia.org/resource/Claremore,_...	"4300"^^<http://www.w3.org/2001/XMLSchema#no...
15	<http://dbpedia.org/resource/Cape_Breton_Univ...	"1951"^^<http://www.w3.org/2001/XMLSch...	<http://dbpedia.org/resource/David_...	<http://dbpedia.org/resource/Sydney,_No...	"4478"^^<http://www.w3.org/2001/XMLSchema#no...
16	<http://dbpedia.org/resource/Bicol_University>	""@en	"Arnulfo M. Mascariñas"@en	<http://dbpedia.org/resource/Legazpi,_Al...	"27224"^^<http://www.w3.org/2001/XMLSchema#n...
17	<http://dbpedia.org/resource/Cardinal_Stritch_U...	"1937"^^<http://www.w3.org/2001/XMLSch...	"Daniel J. Scholz"@en	<http://dbpedia.org/resource/Milwaukee,...	"1365"^^<http://www.w3.org/2001/XMLSchema#no...
18	<http://dbpedia.org/resource/Capitol_Technolog...	"1927-06-01"^^<http://www.w3.org/2001...	"Bradford L. Sims"@en	<http://dbpedia.org/resource/South_Laur...	"790"^^<http://www.w3.org/2001/XMLSchema#non...
19	<http://dbpedia.org/resource/Caldwell_Universit...	"1939-08-10"^^<http://www.w3.org/2001...	"Matthew Whelan"@en	<http://dbpedia.org/resource/Caldwell,_N...	"2200"^^<http://www.w3.org/2001/XMLSchema#no...

Showing 1 to 19 of 19 entries   < 1 >

Bonus Task 1

Step 1 - Data Source Selection and Acquisition

For Step 1, the appropriate and diverse data sources were selected by querying data from Wikidata using SPARQL. The SPARQL query selected universities that are instances of "university" and located in Wisconsin. The query also retrieved information such as the university's founding date, location, total number of students and also the motto, and president if available.

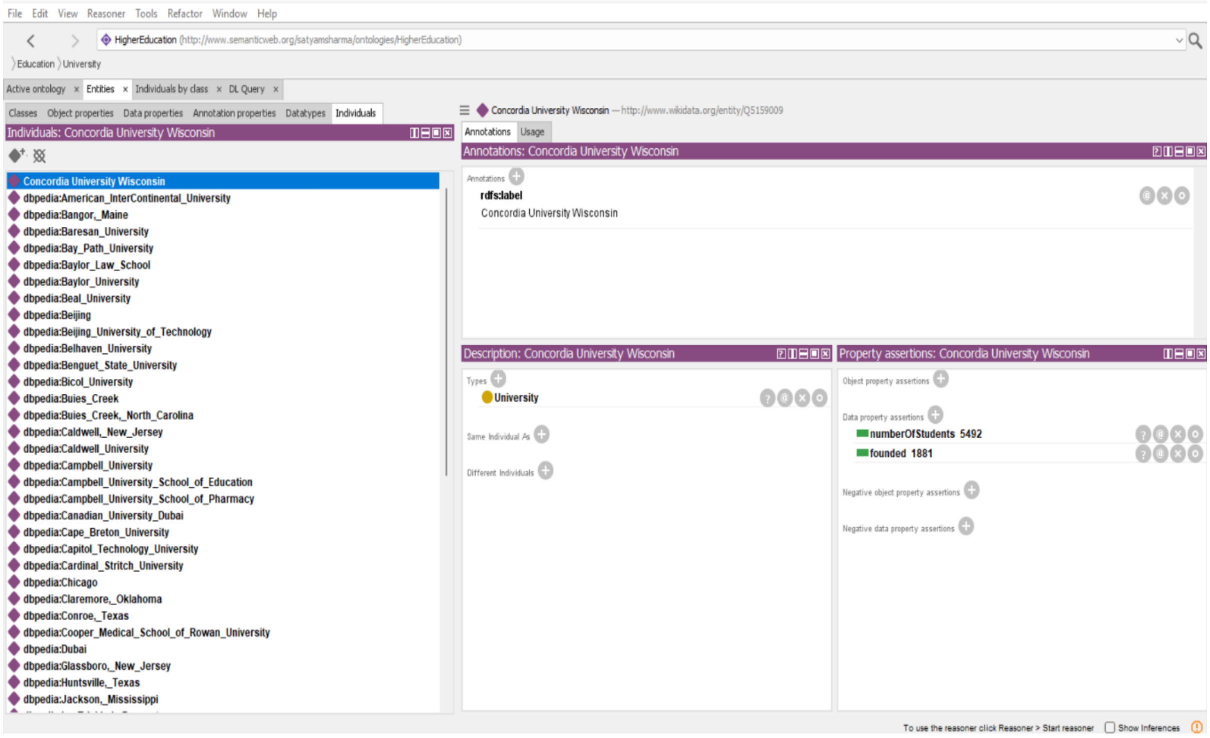



```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>

SELECT ?university ?universityLabel (YEAR(?founded) AS ?yearFounded) ?location ?locationLabel ?totalStudents ?motto ?mottoLabel ?president ?presidentLabel
WHERE
{
  ?university wdt:P31 wd:Q3918 . # Instance of university
  ?university wdt:P131* wd:Q1537 . # Located in Wisconsin
  ?university wdt:P571 ?founded . # Foundation date
  ?university wdt:P159 ?location . # Location
  ?university wdt:P2196 ?totalStudents . # Total number of students

  OPTIONAL {
    ?university wdt:P1546 ?motto . # Motto
  }

  OPTIONAL {
    ?university wdt:P169 ?president . # President
  }

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

university	universityLabel	yearFounded	location	locationLabel	totalStudents	motto	mottoLabel	president	presidentLabel
wd:Q12061082	Lakeshore Technical College	1967	wd:Q2221949	Cleveland	2420				
wd:Q188370	Marquette University	1881	wd:Q6861655	Milwaukee	12002	wd:Q482123	Ad maiorem Dei gloriam		
wd:Q5159009	Concordia University Wisconsin	1881	wd:Q1019382	Mequon	5492				
wd:Q10943718	University of Wisconsin-Superior	1893	wd:Q984577	Superior	2560				
wd:Q2687782	University of Wisconsin–Stevens Point	1894	wd:Q991209	Stevens Point	8302				

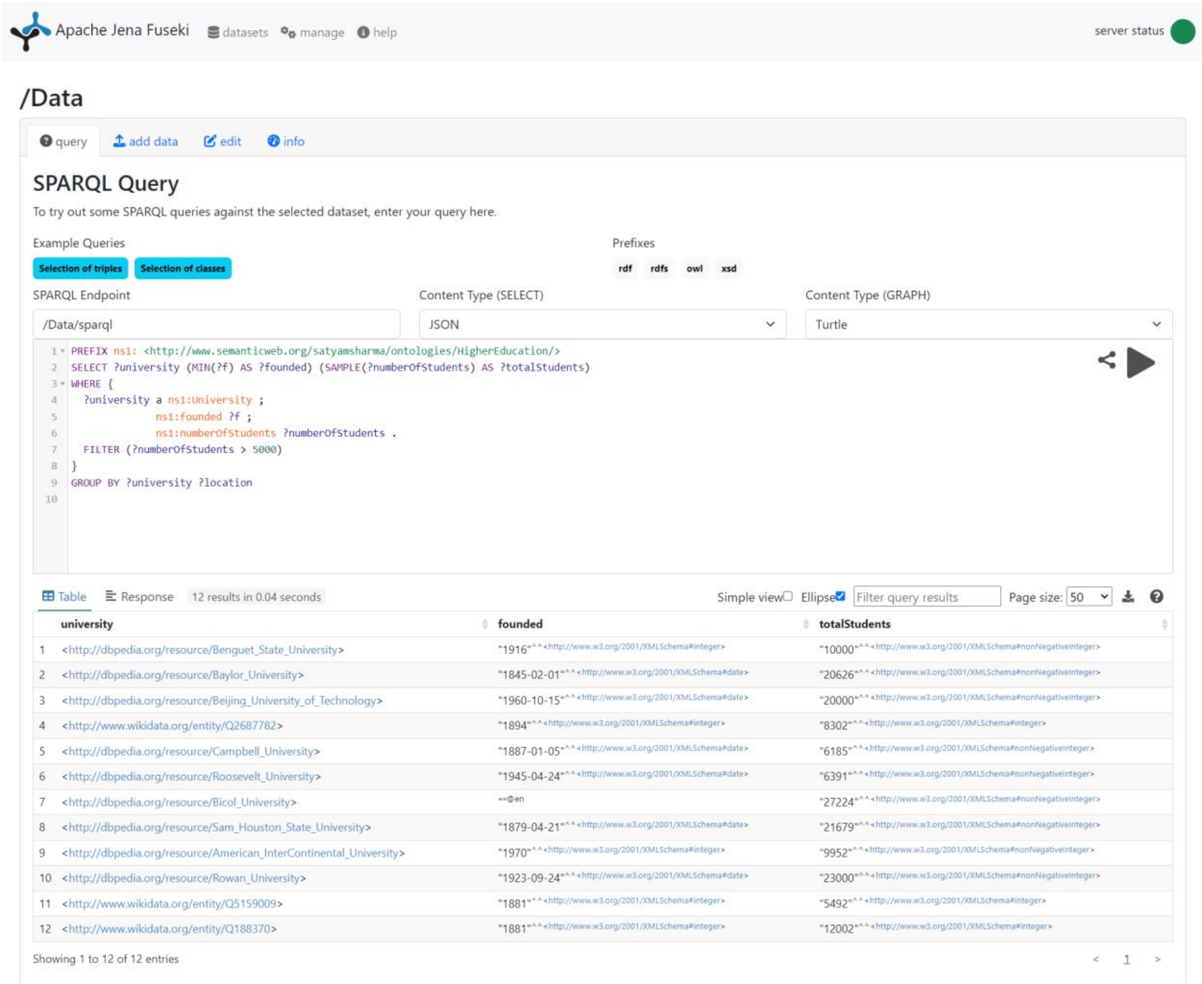
Step 2 - Data Transformation to Ontology-compatible Format.

For Step 2, the appropriate mechanism to retrieve and transform the data was to use pandas to read the CSV file “query.csv” and then add the data to the existing ontology Ontology.ttl using rdflib. The Python script iterated through the CSV rows and added the university entity and its properties to the ontology. The updated ontology was then serialized and saved as a updated_ontology.ttl. The python file has been submitted along with this report. The updated_ontology.ttl file is then uploaded to Protégé and the screenshot of the new University added through the csv file is given below –



Step 3 – Querying the new ontology "updated_ontology.ttl" using the Fuseki Server

For Step 3, the updated query is run on the Fuseki server to retrieve information about universities with more than 5000 students enrolled. The result is attached below –




```
PREFIX ns1: <http://www.semanticweb.org/satyamsharma/ontologies/HigherEducation/>
SELECT ?university (MIN(?f) AS ?founded) (SAMPLE(?numberOfStudents) AS ?totalStudents)
WHERE {
  ?university a ns1:University ;
              ns1:founded ?f ;
              ns1:numberOfStudents ?numberOfStudents .
  FILTER (?numberOfStudents > 5000)
}
GROUP BY ?university ?location
```

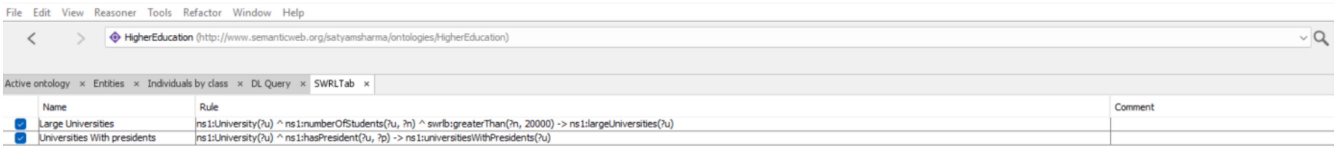
	university	founded	totalStudents
1	<http://dbpedia.org/resource/Benguet_State_University>	"1916"^^<http://www.w3.org/2001/XMLSchema#integer>	"10000"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
2	<http://dbpedia.org/resource/Baylor_University>	"1845-02-01"^^<http://www.w3.org/2001/XMLSchema#date>	"20626"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
3	<http://dbpedia.org/resource/Beijing_University_of_Technology>	"1960-10-15"^^<http://www.w3.org/2001/XMLSchema#date>	"20000"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
4	<http://www.wikidata.org/entity/Q2687782>	"1894"^^<http://www.w3.org/2001/XMLSchema#integer>	"8302"^^<http://www.w3.org/2001/XMLSchema#integer>
5	<http://dbpedia.org/resource/Campbell_University>	"1887-01-05"^^<http://www.w3.org/2001/XMLSchema#date>	"6185"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
6	<http://dbpedia.org/resource/Roosevelt_University>	"1945-04-24"^^<http://www.w3.org/2001/XMLSchema#date>	"6391"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
7	<http://dbpedia.org/resource/Bicol_University>	""@en	"27224"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
8	<http://dbpedia.org/resource/Sam_Houston_State_University>	"1879-04-21"^^<http://www.w3.org/2001/XMLSchema#date>	"21679"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
9	<http://dbpedia.org/resource/American_InterContinental_University>	"1970"^^<http://www.w3.org/2001/XMLSchema#integer>	"9952"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
10	<http://dbpedia.org/resource/Rowan_University>	"1923-09-24"^^<http://www.w3.org/2001/XMLSchema#date>	"23000"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>
11	<http://www.wikidata.org/entity/Q5159009>	"1881"^^<http://www.w3.org/2001/XMLSchema#integer>	"5492"^^<http://www.w3.org/2001/XMLSchema#integer>
12	<http://www.wikidata.org/entity/Q188370>	"1881"^^<http://www.w3.org/2001/XMLSchema#integer>	"12002"^^<http://www.w3.org/2001/XMLSchema#integer>

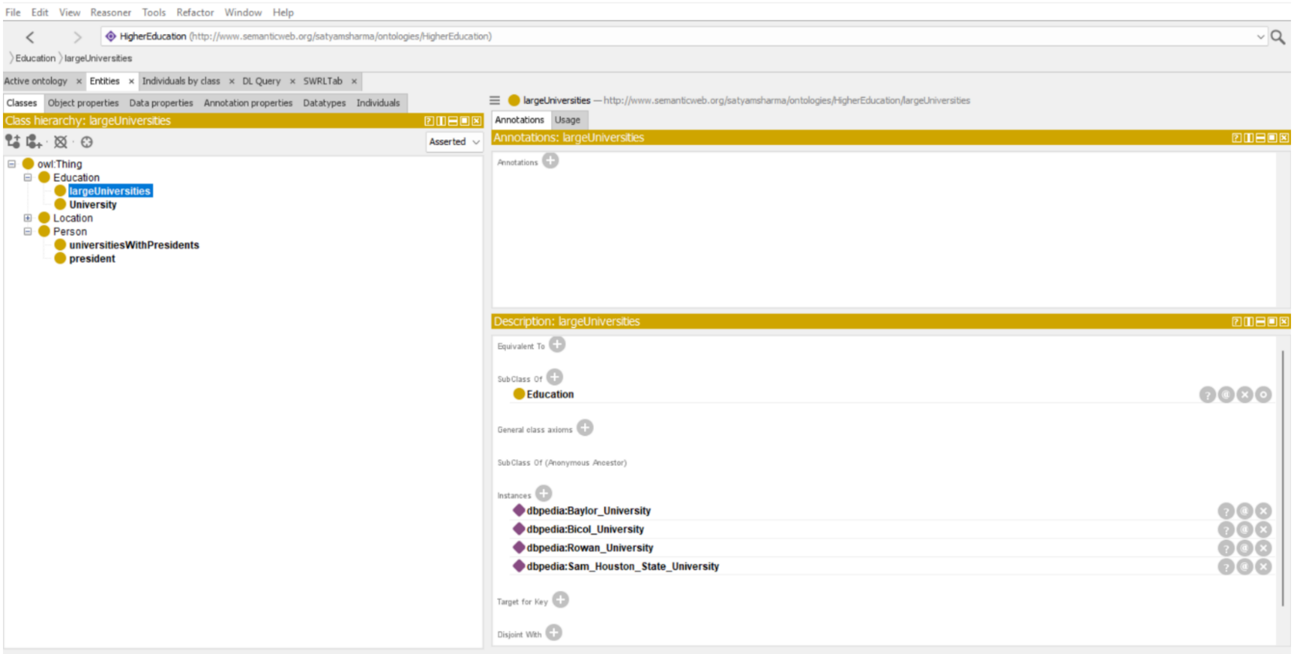
Out of the 12 universities retrieved using the query, 3 universities (result 4, 11, and 12) belong to the data collected from the CSV file queried through Wikidata, while the remaining universities are from the original ontology populated in Basic Task. Result 4 belongs to the University of Wisconsin-Stevens Point, result 11 belongs to Concordia University Wisconsin, and result 12 belongs to Marquette University. Finally, the universities with more than 5000 students enrolled are printed, and they can be classified as big universities based on this criterion.

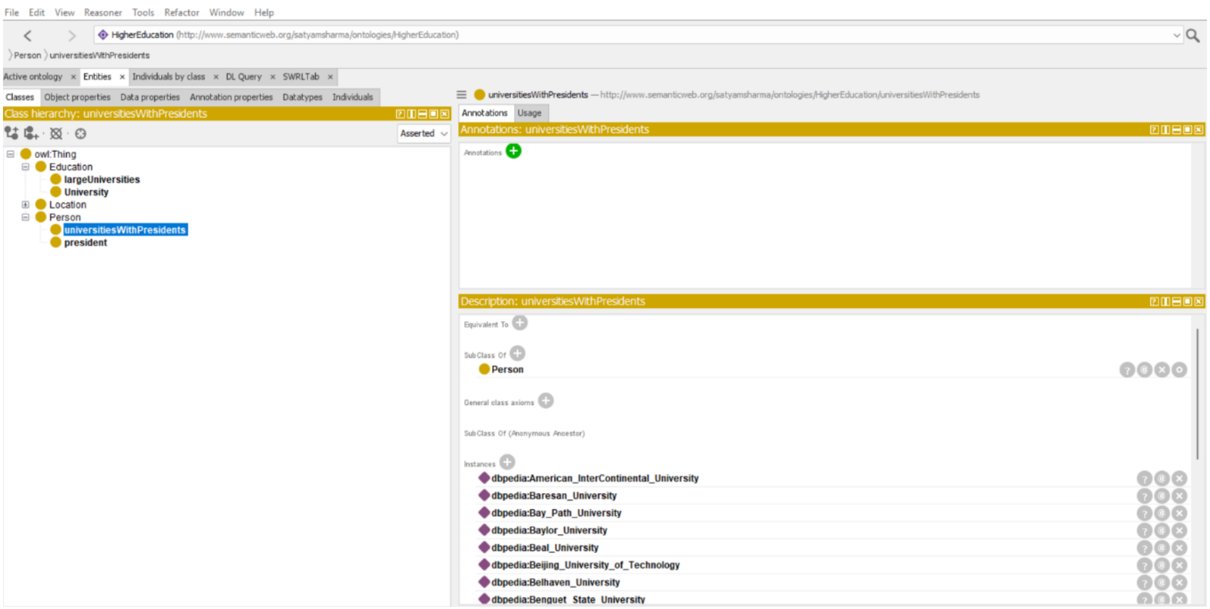
Bonus Task 2

Step 1 –

- I have added two SWRL rules to my Ontology created in Bonus Task 1. The SWRL rules use built-in functions from the SWRLBuiltins library.
- Rule 1 - ns1:University(?u) ^ ns1:numberOfStudents(?u, ?n) ^ swrlb:greaterThan(?n, 20000) -> ns1:largeUniversities(?u). This first SWRL rule gives the Large Universities which are saved under Education Class as a subclass,
- Rule 2 - ns1:University(?u) ^ ns1:hasPresident(?u, ?p) -> ns1:universitiesWithPresidents(?u). The second SWRL rule gives those universities which has a President, it is saved a s class under Person.
- Also, two new sub classes has been created, largeUniversities under Education class and universitiesWithPresidents under Person class.





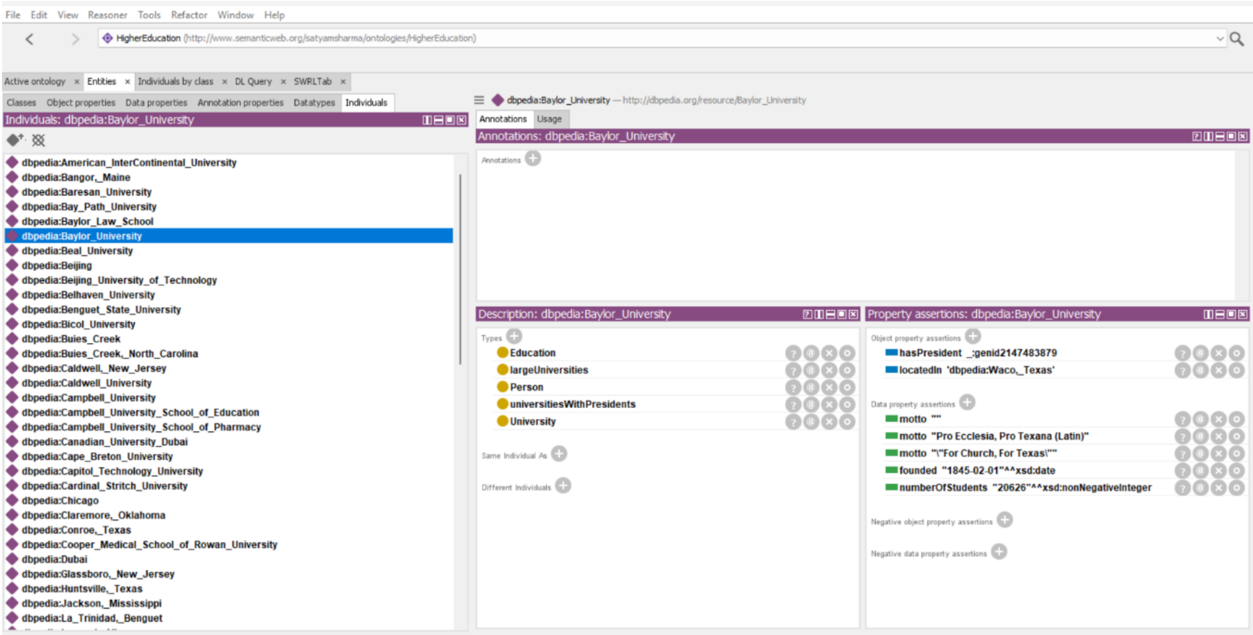


Step 2 –

- Rule 1 defines the conditions for classifying a university as a large university based on the number of students it has. Specifically, if a university has more than 20000 students, it is classified as a large university.
- Rule 2 defines the conditions for classifying a university as having a president. Specifically, if a university has a president, it is classified as a university with presidents.

Step 3 –

- After inspection of the ontology, I have derived that all the four universities classified as Large Universities – Baylor University, Bicol University, Rowan University and Sam Houston State University have Presidents, which shows the importance that big universities give to leadership and it is also be taken as one of the factors for a university to do well and grow if they are under proper leadership.
- Below is an image of the Baylor University which is a Large University as it has 20626 students and has a President as well.



Conclusions –

In this project, I developed an ontology for university data using the Protégé-OWL editor. This comprehensive semantic data model leveraged my skills in Description Logic (DL) and SWRL, leading to an engaging and educational experience.

To build the ontology, I sourced data from DBPedia (primary source) and Wikidata (secondary source), ensuring proper relationships between classes and the appropriate addition of data and object properties. While working on the project, I encountered a few challenges, such as using SPARQL Queries on DBPedia and Wikidata to obtain accurate values for populating my data, as well as incorporating SWRL rules into my ontology without impacting the T Box and A Box.

Thankfully, the professor's lab sessions, lecture slides, and example ontologies, along with various online resources, provided valuable guidance to help me overcome these obstacles.

Overall, this project not only expanded my knowledge in the field, but also demonstrated the practical application of Description Logic and SWRL skills.

In the zip file, I have included the following files –

- Ontology Files –
    1. file.ttl – This contains the basic ontology (T-Box)
    2. Ontology.ttl – This contains the populated ontology file where data has been populated from DBPedia
    3. Updated_ontology.ttl- This contains the new ontology file which has data from the original ontology file – Ontology.ttl and the csv file which has been created via query from the wikidata query service.
    4. SWRL_ontology.ttl – This file contains the populated ontology with the SWRL rules.
- CSV file - Query.csv – This contains the query generated from wikidata query service containing universities from Wisconsin State of US.
- Python files –
    1. pythonfile.py – This file is used to populate file.ttl with data from DBPedia with the help of SPARQL Query and python.
    2. pythonfile(BonusTask).py – This file is used to merge data from query.csv file to the ontology, which shows that external data can be added to the ontology.

Steps to run the code –

1. To simply view the T-box, upload the file.ttl file into Protege.
2. For the basic task, upload Ontology.ttl file to display both the T-box and A-box.
3. For Bonus Task 1, upload Updated_ontology.ttl file to examine the DBPedia data, which now contains data from WikiData as well.
4. For Bonus Task 2, upload 'SWRL_ontology.ttl' file which contains the SWRL rules.