Assignment 3

José Cutileiro - Group 70

May 5, 2024

1 Introduction

In this assignment, the goal is to answer some questions posed by the course instructors using SPARQL queries. The answers take into account the data provided in the first assignment (the csv files).

2 My approach

To solve this task, I took into consideration both the diagram developed in the first assignment and the ontology elaborated for the second assignment. However, while following my approach, I encountered some irregularities in the data initially provided. At certain points, the data did not align with my approach.

Not only occasionally were the data not aligned with my approach, but in two specific cases, the primary keys used in the data did not correspond to mine, which becomes problematic when answering questions using SPARQL. This happens because my approach represents the object exclusively through its primary key.

For example, in the courses.csv file, the "Owned By" column contains the names of the programs, however, the primary key of the programs is actually the program code. This makes navigating between these data more complex because it's necessary to deal with all combinations between the programs and the courses. To solve this type of problem, I developed a Python script that corrects these columns by assigning the desired values (**"fixcsv.py"**).

In addition to these corrections, it is necessary to place the data into GraphDB so that queries can be executed and the desired results obtained. For this purpose, I developed a new Python script that iterates through the data in the original and corrected CSV files (**csv2rdf.py**), converting the information into the ".ttl" format. This approach greatly simplifies the data import process. It is important to note that the names of classes, relationships, and attributes strictly adhere to the ontology created in Assignment 2, ensuring consistency during the import of both files into GraphDB.

After this step, it's just a matter of crafting the SPARQL queries to obtain the desired results.

3 Decisions and assumptions

In addition to following the previous assumptions, such as the decision about which attributes are the primary keys and which relationships are present in each entity, I needed to make an additional assumption. This assumption is that individuals who are both students and teachers are automatically considered Teaching Assistants. This assumption is necessary because the concept of Teaching Assistant is omitted both in the diagram and in the ontology. Thus, to answer queries related to Teaching Assistants, it is necessary to select a person who is both a teacher and a student, and automatically they will be considered a Teaching Assistant.

4 Queries and results

4.1 Find the name, director and department of all programmes.

=== QUERY ===

```
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?name ?department_name ?director_id where {
    ?programme rdf:type :Programme .
    ?director rdf:type :Senior_Teacher .
    ?programme :programmeName ?name .
    ?programme :programmeDept ?department .
    ?director :oversees ?programme .

    ?department :deptName ?department_name .
    ?director :teacherId ?director_id .

}

=== RESULTS ===
name,       department_name,        director_id
P-51,           D6,             19580515-0017
P-52,           D6,             19611219-0014
P-53,           D6,             19600905-0003
P-54,           D6,             19630126-0001
P-21,           D3,             19570615-0011
P-41,           D5,             19580218-0007
P-42,           D5,             19620831-0024
P-71,           D8,             19610620-0006
P-72,           D8,             19660630-0020
P-73,           D8,             19600814-0002
P-74,           D8,             19601021-0018
P-11,           D2,             19620424-0026
P-12,           D2,             19610623-0005
P-13,           D2,             19690408-0009
P-14,           D2,             19560812-0016
P-01,           D1,             19620522-0023
P-31,           D4,             19650303-0019
P-32,           D4,             19570826-0012
P-33,           D4,             19570828-0008
P-34,           D4,             19610918-0027
P-61,           D7,             19680712-0028
```

4.2 Find the names of all students who worked as teaching assistants in courses given by the D3-2 division in study period 2 in academic year 2023/2024.

```
=== QUERY ===
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?name ?academicYear ?studyPeriod ?divisionName where {
    ?TA rdf:type :Teacher .
    ?TA rdf:type :Student .
    ?TA :worksIn ?CI .

    ?course rdf:type :Course .
    ?course :hasExecution ?CI .
    ?course :courseDivision ?division .

    ?CI :academicYearCourse ?academicYear .
```

```
    ?CI :studyPeriod ?studyPeriod .

    ?division :divisionName ?divisionName .

    FILTER(?studyPeriod = "2.0") .
    FILTER(?academicYear = "2023-2024") .
    FILTER(?divisionName = "D3-2") .

    ?TA :teacherName ?name .
}
```

```
=== RESULTS ===
name,     academicYear,    studyPeriod,    divisionName
TA 36,    2023-2024,            2.0,            D3-2
TA 138,   2023-2024,            2.0,            D3-2
TA 38,    2023-2024,            2.0,            D3-2
TA 74,    2023-2024,            2.0,            D3-2
TA 60,    2023-2024,            2.0,            D3-2
```

4.3 Find all teachers who are assigned more than 120 hours in course 1015 in study period 1 in academic year 2018/2019.

```
=== QUERY ===
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?name ?id ?assignedHours ?courseCode ?studyPeriod ?academicYear where {
    ?hours rdf:type :Hour_Information .
    ?teacher rdf:type :Teacher .
    ?CI rdf:type :Course_Instance .
    ?course rdf:type :Course .

    ?hours :associatedWith ?teacher .
    ?hours :assignedHours ?assignedHours .
    ?hours :given ?CI .
    ?course :hasExecution ?CI .
    ?course :courseCode ?courseCode .
    ?CI :studyPeriod ?studyPeriod .
    ?CI :academicYearCourse ?academicYear .

    FILTER(?assignedHours > "120") .
    FILTER(?courseCode = "1015") .
    FILTER(?studyPeriod = "1.0") .
    FILTER(?academicYear = "2018-2019") .

    ?teacher :teacherName ?name .
    ?teacher :teacherId ?id .
}
```

```
=== RESULTS ===
name,id,assignedHours,courseCode,studyPeriod ,academicYear
Teacher 19,19650303-0019,240.0,1015,1.0,2018-2019
TA 59,19750102-0059,140.0,1015,1.0,2018-2019
Teacher 7,19580218-0007,280.0,1015,1.0,2018-2019
TA 38,19790702-0038,140.0,1015,1.0,2018-2019
Teacher 20,19660630-0020,240.0,1015,1.0,2018-2019
```

```
TA 57,19780424-0057,120.0,1015,1.0,2018-2019
```

4.4 Find all students registered for course instance I-910 that were not registered for course instance I-911.

```
=== QUERY ===
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?id ?name ?graduated where {
?student rdf:type :Student .
?ci rdf:type :Course_Instance .
?student :register ?regis .
?ci :hasRegistration ?regis .

    ?ci :instanceId "I-910" .

    FILTER NOT EXISTS {
        ?ci :instanceId "I-911" .
    }

    ?student :studentId ?id .
    ?student :studentName ?name .
    ?student :graduated ?graduated .
}

=== RESULT ===
id,              name,      graduated
19921201-0094,   TA 94,      False
```

4.5 Find all programmes along with the total number of owned courses. List the results in descending order of number of owned courses.

```
=== QUERY ===
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?code ?name (COUNT(?course) AS ?courseCount) WHERE {
    ?programme rdf:type :Programme .
    ?programme :programmeCode ?code .
    ?programme :programmeName ?name .
    ?programme :owns ?course .
} GROUP BY ?code ?name
ORDER BY DESC(?courseCount)

=== RESULTS ===
code,name,courseCount
10061,P-61,45
10021,P-21,33
10001,P-01,32
10041,P-41,21
10042,P-42,20
10032,P-32,14
10071,P-71,14
10052,P-52,13
10012,P-12,12
```

```
10033,P-33,12
10051,P-51,11
10054,P-54,11
10011,P-11,10
10072,P-72,10
10074,P-74,9
10013,P-13,8
10014,P-14,8
10034,P-34,6
10073,P-73,6
10053,P-53,3
10031,P-31,2
```

4.6 Find the number of: senior teachers

```
=== QUERY ===
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT (COUNT(?senior_teachers) AS ?count) WHERE {
    ?senior_teachers rdf:type :Senior_Teacher
}

=== RESULT ===
count
30
```

4.7 Find the number of: people

```
=== QUERY ===
PREFIX : <http://www.semanticweb.org/35196/ontologies/2024/3/assignment2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT (?senior_teacher_count + ?student_count AS ?total_count) WHERE {
    {
        SELECT (COUNT(?senior_teacher) AS ?senior_teacher_count) WHERE {
            ?senior_teacher rdf:type :Senior_Teacher .
        }
    }
    {
        SELECT (COUNT(?student) AS ?student_count) WHERE {
            ?student rdf:type :Student .
        }
    }
}

=== RESULT ===
total_count
440
```

5 How to obtain the same results?

```
To achieve the same results, simply follow the instructions below:
```

1. Create a folder containing all the CSV files provided in Assignment 1.
2. Run the fixcsv.py program to add the two missing CSV files.
3. Run the csv2rdf.py program to convert the data to Turtle format.
4. Load the files statments.ttl and out.ttl into GraphDB.
5. Test each of the queries to verify if the obtained result is as expected.

Note: In my submission, I included the CSV files with the original
results and a file with the original queries (queries.txt).