Homework 3

Team Members: Tushar Yadav, Pratistha Soni, Dhruti Dilipbhai Patel

1. Consider the following table in a relational database storing the assignment of courses to classrooms in a university.

Course number (primary key)	Room	Department
CPSC-583	CS-110B	ComputerScience
CPSC-597	CS-110B	ComputerScience
CPSC-473	CS-406	ComputerScience

Convert the relational data into a Semantic Web representation. Use the following two properties:

http://example.org/is-located-in
http://example.org/is-offered-by

i. CPSC-583 is located in CS-110B and is offered by ComputerScience:

<http://example.org/CPSC-583>
<http://example.org/is-located-in>
<http://example.org/CS-110B> ;
<http://example.org/is-offered-by>
<http://example.org/ComputerScience> .

ii. CPSC-597 is located in CS-110B and is offered by ComputerScience:

<http://example.org/CPSC-597>
<http://example.org/is-located-in>
<http://example.org/CS-110B> ;
<http://example.org/is-offered-by>
<http://example.org/ComputerScience> .

iii. CPSC-473 is located in CS-406 and is offered by ComputerScience:

<http://example.org/CPSC-473>
<http://example.org/is-located-in>
<http://example.org/CS-406> ;
<http://example.org/is-offered-by>
<http://example.org/ComputerScience> ..

(a) Give the **triple** representation of the Semantic Web data. The first triple is already given (ignoring the full URI for subject and object).

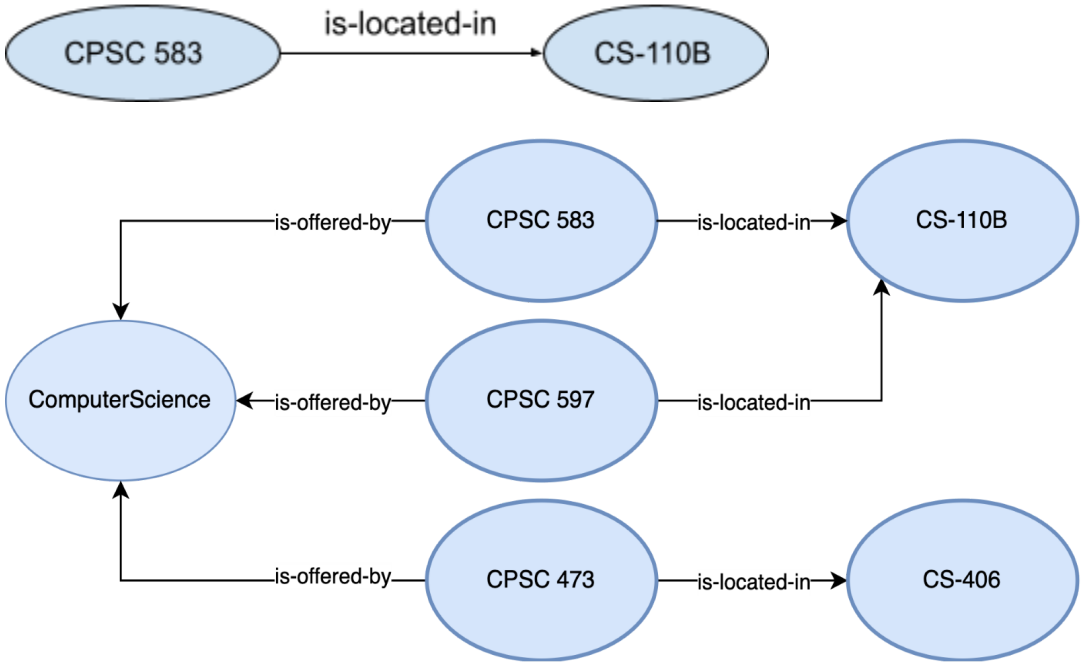
PREFIX ex: http://example.org/

CPSC-583 ex:is-located-in CS-110B .

CPSC-597 ex:is-located-in CS-110B

CPSC-473 ex:is-located-in CS-406

(b) Give the **graph** representation of the Semantic Web data. The first edge is already given.



2. Read the attached paper "Data Integration through Ontology-Based Data Access to Support Integrative Data Analysis: A Case Study of Cancer Survival" by Hansi Zhang et al. (presented at the 2017 IEEE International Conference on Bioinformatics and Biomedicine).

   a. Write a short essay (approximately 200 words) that describes

      i. what is the problem the authors are trying to solve,

      ii. how Semantic Web technology provides a solution, and

      iii. what, in your opinion, are the advantages and disadvantages of their Semantic Web-based approach.

In their paper titled "Data Integration through Ontology-Based Data Access to Support Integrative Data Analysis: A Case Study of Cancer Survival," the authors tackle the complex problem of integrating and analyzing data from various sources to better understand the factors influencing cancer survival. The paper understands that heterogeneity of data creates multiple issues in terms of schema, design or semantics when dealing with factors affecting cancer survivals.

Authors of the paper uses Ontology Based Data Access (OBDA) to create a global ontology which consists of relations between different data sources. Semantic mappings are established between the ontology and these sources, ensuring data elements and relationships are harmonized. Now authors use query tool like SPARKQL to access the data and perform analysis over the data. This approach helps in maintain quality of results, data remains consistent throughout the system because of the global ontology and every type of data values can be used comprehensively to make analysis.

Advantages of Semantic Web-based approach is that the data remains consistent throughout, quality of data is maintained, data integration between multiple data sources is easier and querying data from different data sources gets easier. Data analysis gets more comprehensive as we can incorporate many other data sources of varying types(heterogeneous). On the other hand, disadvantages of Semantic Web-based approach include complexity and difficulty in creating an ontology using varying data sources and that to create Web based ontologies implementing this semantic web-based approach we need experts in the field. It may also require additional computational resource depending upon the size and computational logic of the whole system.

b. Figure 2 in the paper is a fragment of the ontology they developed. Using only this information, write a SPARQL query to retrieve a list of all patients who have lung cancer and who live in a county with a high rate of smoking (ocrv:avg_smoke > 0.2). Hint: you can look at some of the example queries in Tables 4-7.

```
PREFIX :<http://www.semanticweb.org/ontologies/OCRV#>
SELECT ?p where {
    ?p a ocrv:patient.
    ?p ocrv;:has_therapy ocrv:radiation_therapy.
    ?p ocrv:has_disease ocrv:lung_cancer
    ?p ocrv:lives_in ocrv:avg_smoke }.
```

3. Write SPARQL queries to list:

a. All English language films that starred *Irrfan Khan*.

```
SELECT ? film
WHERE {
?film a dbo:Film ;
        dbpedia2:language ?language ;
        dbpedia2:starring ?actor .
FILTER (LANG(?language) = 'en' && REGEX(?actor, "Irrfan Khan", "i")) }
```

b. All English language films that starred *Irrfan Khan* and *Angelina Jolie*.

```
SELECT ?film
WHERE { ?film a dbo:Film ;
        dbpedia2:language ?language ;
        dbpedia2:starring ?actorf ;
        dbpedia2:starring ?actorm .
FILTER (regex(?actorf, "Angelina Jolie", "i") && regex(?actorm, "Irrfan Khan", "i"))}
```

SPARQL Explorer for https://dbpedia.org/sparql

SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
SELECT ?film
WHERE {
  ?film a dbo:Film ;
        dbpedia2:language ?language ;
        dbpedia2:starring ?actorf ;
        dbpedia2:starring ?actorm .
  FILTER (regex(?actorf, "Angelina Jolie", "i") && regex(?actorm, "Irrfan Khan", "i"))
}
```

Results: Browse Go! Reset

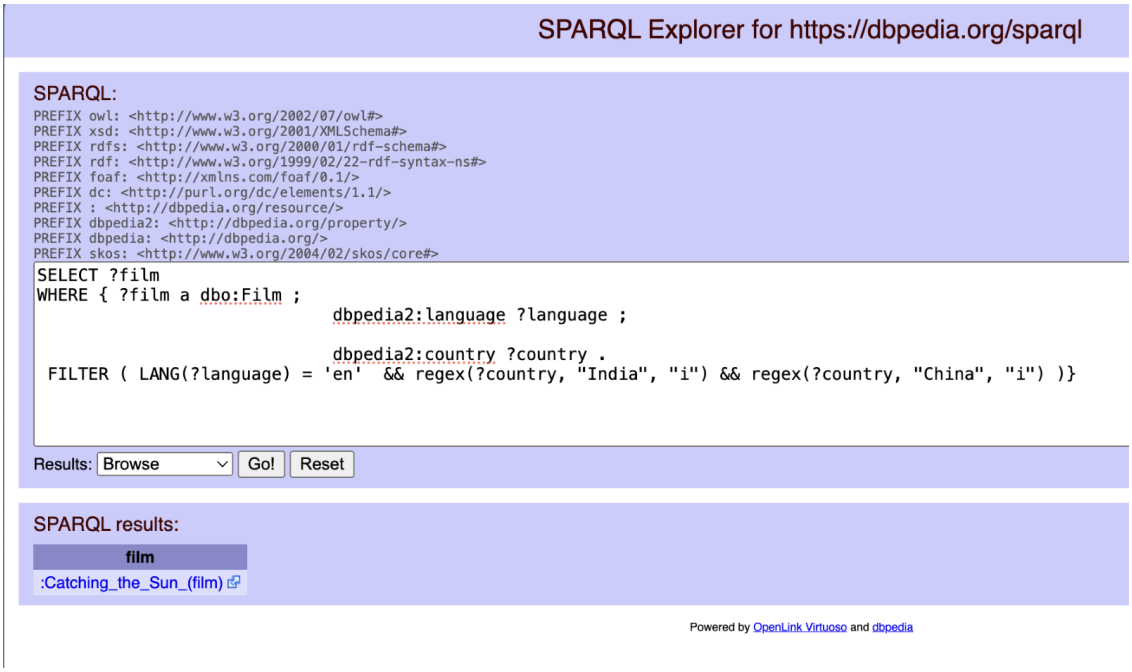
SPARQL results:

film
:A_Mighty_Heart_(film)

Powered by OpenLink Virtuoso and dbpedia

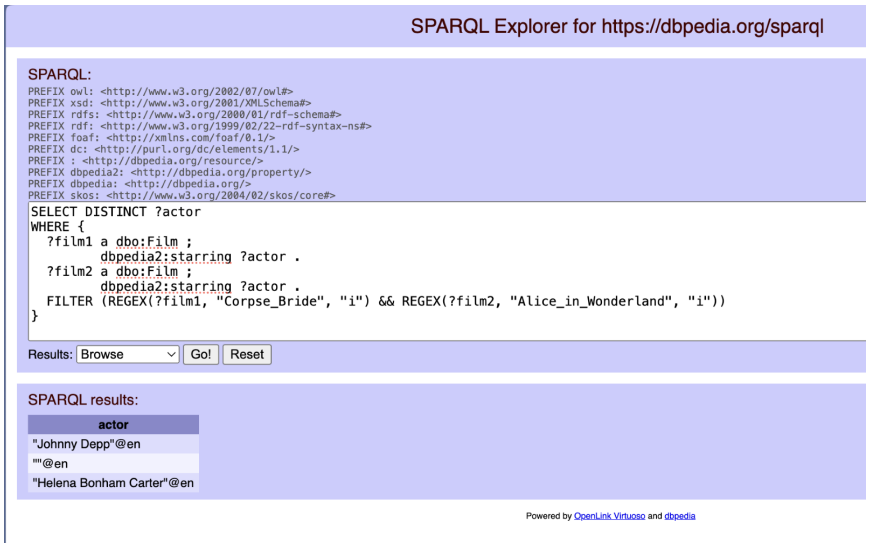
c. All English language films set in both *India* and *China*

```
SELECT ?film
WHERE { ?film a dbo:Film ;
        dbpedia2:language ?language ;
        dbpedia2:country ?country .
FILTER ( LANG(?language) = 'en'  && regex(?country, "India", "i") && regex(?country, "China",
"i") )}
```



d. All actors who have acted in both "*Corpse Bride*" and "*Alice in Wonderland*" (2010 film)

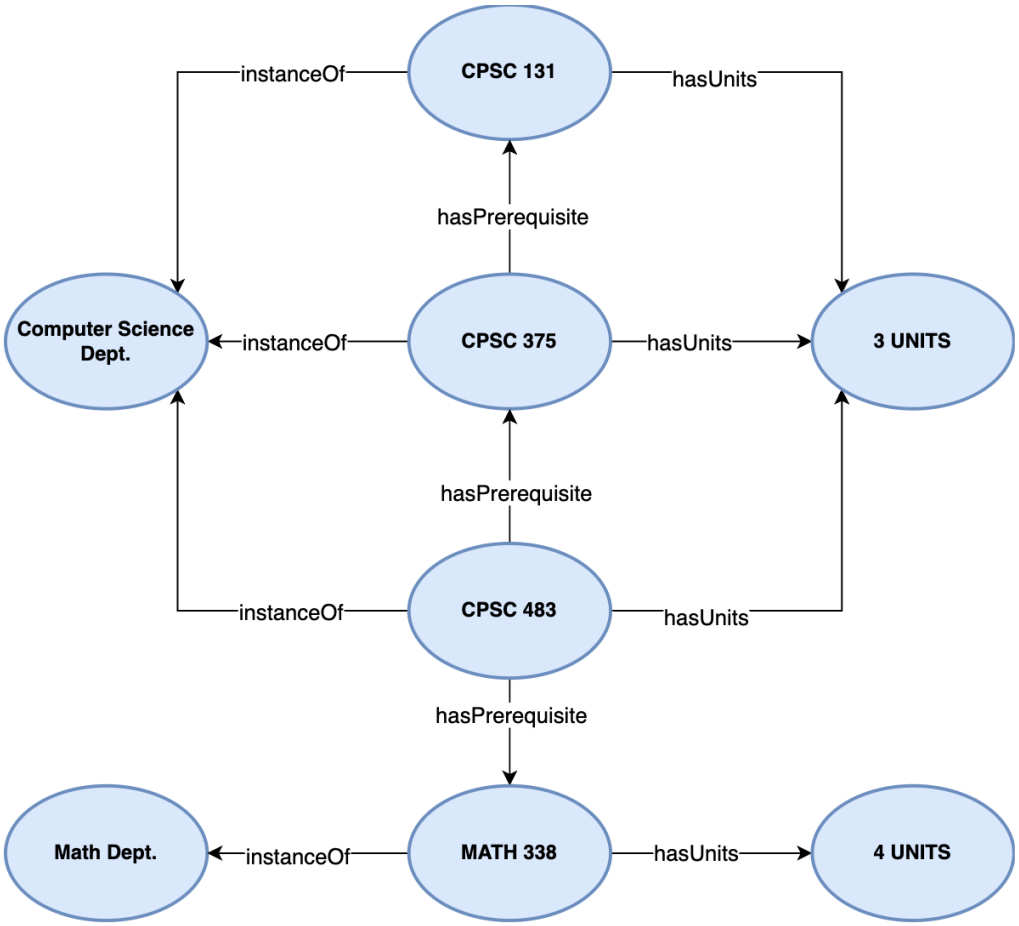
```
SELECT DISTINCT ?actor
        WHERE {
                ?film1 a dbo:Film ;
                        dbpedia2:starring ?actor .
                ?film2 a dbo:Film ;
                        dbpedia2:starring ?actor .
FILTER (REGEX(?film1, "Corpse_Bride", "i") && REGEX(?film2, "Alice_in_Wonderland", "i"))}
```



4. Consider the following facts.

   CPSC131, CPSC375, CPSC483 are Computer Science courses. MATH338 is a Math course. CPSC483 has CPSC375 and MATH338 as prerequisites. CPSC375 has CPSC131 as a prerequisite. All courses are 3 units except Math courses which are 4 units. Every course belongs to a department. Computer Science Department and Mathematics Department are two departments.

   a. Represent the above facts as a semantic network (draw the network). Use the following properties/relationships: instanceOf, subsetOf, hasPrerequisite, belongsTo, hasUnits.



b. Using default reasoning, answer the following queries or state that they cannot be answered.

i. How many units is CPSC131?

Since all courses are 3 units by default, and there is no specific information given about CPSC131 having a different number of units, we can default to the standard number of units for courses, which is 3 units. Therefore, CPSC131 has 3 units.

ii. How many units is MATH338?

MATH338 is a Math course, and all Math courses are 4 units. Therefore, MATH338 is 4 units.

iii. Is CPSC375 a prerequisite of CPSC483?

Yes, CPSC375 is one of the prerequisites for CPSC483.

iv. Is CPSC131 a prerequisite of CPSC483?

No, CPSC131 is not a direct prerequisite for CPSC483. However, it is a prerequisite for CPSC375, which is a prerequisite for CPSC483.

v. What are the prerequisites for CPSC483?

The prerequisites for CPSC483 are CPSC375 and MATH338.

vi. Does CPSC131 belong to the Computer Science Department?

Yes, CPSC131 is a Computer Science course, and therefore it belongs to the Computer Science Department.