Information Integration on the Web: Homework 3

Due on February 6, 2015

Prof. Ambite & Knoblock

Tushar Tiwari

Problem 1

Get the student populations of USC and UCLA respectively using one query. [20 points]

Solution

```
BASE <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>

SELECT * WHERE
{
    { <University_of_Southern_California> dbpedia-owl:numberOfStudents ?num } UNION { <
        University_of_California,_Los_Angeles> dbpedia-owl:numberOfStudents ?num } .
}
```

num
39958
42163

Figure 1: Output of query1.sparql

Problem 2

Get all the private universities in Los Angeles, and sort the universities in descending order of their student populations. [20 points]

Solution

```
BASE <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT DISTINCT ?university, ?num_of_students WHERE
{
    ?university dbpprop:type <Private_university> .
    { ?university dbpprop:location <Los_Angeles> } UNION { ?university dbpprop:city <Los_Angeles> }
         UNION { ?university dbpedia-owl:city <Los_Angeles> } .
    OPTIONAL { ?university dbpedia-owl:numberOfStudents ?num_of_students } .
}
ORDER BY DESC(?num_of_students)
```

university	num_of_students
http://dbpedia.org/resource/University_of_Southern_California	39958
http://dbpedia.org/resource/Loyola_Marymount_University	9369
http://dbpedia.org/resource/Mount_St._Mary's_College	3299
http://dbpedia.org/resource/Occidental_College	2123
http://dbpedia.org/resource/Los_Angeles_Recording_School	1500
http://dbpedia.org/resource/Woodbury_University	1500
http://dbpedia.org/resource/Loyola_Law_School	1297
http://dbpedia.org/resource/Southwestern_Law_School	964
http://dbpedia.org/resource/USC_School_of_Architecture	580
http://dbpedia.org/resource/American_Academy_of_Dramatic_Arts	220

Figure 2: Output of query2.sparql

Problem 3

Get the campus types and their corresponding counts of all the public universities in Los Angeles.[20 points]

Solution

```
BASE <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT ?campus, count(?campus) AS ?count WHERE
{
    {
        SELECT DISTINCT ?university ?campus
        {
            ?university dbpprop:type <Public_university> .
            ?university dbpprop:city <Los_Angeles> .
            { ?university dbpprop:campus ?campus } UNION { ?university dbpedia-owl:campus ?campus }
                .
        }
    }
}
```

campus	count
"419.0"^^<http://dbpedia.org/datatype/acre>	1
"Urban"@en	3
http://dbpedia.org/resource/Urban_area	5

Figure 3: Output of query3.sparql

Problem 4

Get the universities among all the private and public universities located in Los Angeles, which have been established for more than 100 years. Sort the universities in ascending order of established years.[20 points]

Solution

```
BASE <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT distinct ?university, ?established_years WHERE
{
    { ?university dbpprop:type <Public_university> } UNION { ?university dbpprop:type <
        Private_university> } .
    { ?university dbpprop:location <Los_Angeles> } UNION { ?university dbpprop:city <Los_Angeles> }
         UNION { ?university dbpedia-owl:city <Los_Angeles> } .
    ?university dbpprop:established ?established_raw .
    BIND(IF(DATATYPE(?established_raw) = xsd:date, YEAR(?established_raw), ?established_raw) AS ?
        established) .
    BIND(YEAR(NOW()) - ?established AS ?established_years) .
    FILTER(?established_years > 100) .
}
ORDER BY ASC(?established_years)
```

university	established_years
http://dbpedia.org/resource/Southwestern_Law_School	104
http://dbpedia.org/resource/Loyola_Marymount_University	104
http://dbpedia.org/resource/Occidental_College	128
http://dbpedia.org/resource/Keck_School_of_Medicine_of_USC	130
http://dbpedia.org/resource/USC_Viterbi_School_of_Engineering	130
http://dbpedia.org/resource/American_Academy_of_Dramatic_Arts	131
http://dbpedia.org/resource/Woodbury_University	131
http://dbpedia.org/resource/University_of_California,_Los_Angeles	133
http://dbpedia.org/resource/UCLA_Graduate_School_of_Education_and_Information_Studies	134
http://dbpedia.org/resource/University_of_Southern_California	135

Figure 4: Output of query4.sparql

Problem 5

Provide another query you are interested in and list the returned results. This query should be substantially different from the previous four. [20 points]

Solution

In this query the net worth of all the billionaires whose alma mater is either USC or UCLA is summed up and stacked up against each other by grouping over the alma mater.

```
BASE <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT ?almamater, CONCAT("US$ ", CONCAT(STR(SUM(?networth)), " billion")) AS ?total_worth WHERE {
    { ?person dbpedia-owl:almaMater <University_of_Southern_California> } UNION { ?person dbpedia-
        owl:almaMater <University_of_California,_Los_Angeles> } .
    ?person dbpprop:networth ?networth_raw .
    ?person dbpedia-owl:almaMater ?almamater .
    FILTER(?almamater = <University_of_Southern_California> OR ?almamater = <
        University_of_California,_Los_Angeles>) .
    FILTER(REGEX(?networth_raw, "billion", "i")) .
    BIND(REPLACE(?networth_raw, "billion\\s?[USD]*$", "", "i") AS ?networth2) .
    BIND(REPLACE(?networth2, "^[US]*\\s?\\$", "", "i") AS ?networth3) .
    BIND(IF(REGEX(?networth3, ","), xsd:decimal(REPLACE(?networth3, ",", ".")), xsd:decimal(?
        networth3)) AS ?networth) .
}
GROUP BY ?almamater
```

almamater	total_worth
http://dbpedia.org/resource/University_of_Southern_California	US$ 11.3 billion
http://dbpedia.org/resource/University_of_California,_Los_Angeles	US$ 4.75 billion

Figure 5: Output of query5.sparql