F20BD/F21BD Big Data Management

SPARQL Lab

In this lab you will use the SPARQL end point provided for access to the DBpedia version of Wikipedia. For information on DBpedia see http://wiki.dbpedia.org.

You will use the Virtuoso SPARQL Query Editor for DBpedia. Use the web browser of your choice and open the following page: http://dbpedia.org/sparql.

You should now be presented with the following page:

1. Write a SPARQL query to answer the following question: *Who are the authors born in New York*? (Hint: Browse the results from the above queries to identify the relevant predicates and URIs for the graph patterns in the where clause.)

2. Consider the following example of a SPARQL query (taken from the "A Semantic Web Primer" book by Antoniou et al.

```
SELECT ?apartment
WHERE {
        ?apartment  swp:hasNumberOfBedrooms ?bedrooms.
         FILTER (?bedrooms > 2).
}
```

Note the "FILTER" keyword, which allows you to specify a condition on a literal value. A filter may make use of a regular expression, e.g.

```
FILTER regex(?address, "^4 Baron Way").
```

DBpedia supports a predicate called distanceToEdinburgh (`dbo:distanceToEdinburgh`).

- Write a SPARQL query that will retrieve all places with a distance greater than 20km from Edinburgh (sorted by distance).
- Now write a SPARQL query that will give us a list of all places which have a postal code that starts with 'EH' and how far from Edinburgh they are. Again, listed in order of distance.

4. SPARQL 1.1 supports aggregate functions, e.g. COUNT, SUM, MIN, MAX,and AVG, as well as GROUP BY and HAVING in much the same way as SQL does.
    - Write a SPARQL query that gives us the average distance of places to Edinburgh.
    - Write a SPARQL query that gives us the average distance of places per postcode area, e.g. EH, FF, etc. Order the output according to the average distance.
    - Repeat the query from above, but only show post code areas with an average distance above 200km.