

Homework 7

Due at 11:59pm 03/06/2015

Question 1 (100 points)

In this homework, you need to work with Sesame Triple Store.
To finish the homework, you should write a **program** (any language you like) to do the following tasks in order.

1 Get data from DBpedia (20 points)

You can use a **POST** request or **Remote Data repository** to retrieve the data. The program should collect the following information of each university: (1) university URIs and (2) their names.

2 Decide on a model for the data use Schema.org (0 points)

Your program should model your data using CollegeOrUniveristy in schema.org. You can use the **name** property for the university names.

3 Use a SPARQL Construct query create the RDF and load the data into the local triple store (10 points)

Based on the previous schema, your program should create a SPARQL Construct query to add the tuples you get from DBpedia to the local repository.

4 Use select query to find all Organizations (10 points)

Now, your program should generate a Select SPARQL query to find all the Organizations. Organization is a class defined in schema.org. Save the query and the output of your SPARQL query in a file called q4.txt.

5 Load Schema.org schema file (10 points)

Download the schema file from Schema.org and load the file into your local triple store. The schema file can be downloaded at link <http://schema.rdfs.org>. You can choose your own format.

6 Use select query to find all Organizations (10 points)

Now, run the same query as Q4 and save the output into a file called q6.txt.

7 Enable the RDFS inference (10 points)

Enable ForwardChaining inference of your local triple store.

8 Use the select query to find all Organization (10 points)

Now, run the same query as Q4 and save the output into a file called q8.txt

9 Insert a triple saying that USC has an alumni C. L. Max Nikias (10 points)

Use the alumni property of CollegeOrUniversity. Your program should insert a triple that says USC has an alumni C.L.Max Nikias. The subject is the URI of USC and the object is the URI of C.L.Max Nikias. You can find the URIs of USC and C.L.Max Nikias in the USC DBpedia page.

10 Use select query to find all person from the triple store (10 points)

Your program should generate a Select query to find all Person from the triple store. Person is also defined in Schema.org. Run the program with ForwardChaining Inference turned off and then turned on. Save the output of your query both before and after into a file call q10.txt.

What to hand in:

1. A document that includes
 - queries for question 3, 4 and 10
 - description of how your program works for each task. Explain why it returned (or didn't return) the data that it did (no more than 3 sentences for each task)
2. q4.txt, q6.txt, q8.txt and q10.txt. These can be triples or a result set from a SPARQL select query.
3. The source code of your program

Reference:

[1] Sesame Tutorial

<http://openrdf.callimachus.net/sesame/2.7/docs/users.docbook?view>

[2] Sesame download <http://sourceforge.net/projects/sesame/files/Sesame%202/>

Submission Instructions:

(1) Compress and zip ALL your homework files into one .zip file. Note that only .zip file extensions are allowed. Other compression extensions such as .tar, rar, or 7z will NOT be accepted.

(2) Please make sure ALL source files are included in your zip file when submitted. Only your FINAL submission will be graded.