

## Master Big Data Management and Analytics

## Training Linked Open Data with SPARQL

**Emmanuel Jamin**

Open Data and Semantic Web Consultant in Everis

[emmanuel.jamin@gmail.com](mailto:emmanuel.jamin@gmail.com)

**Jovan Varga**

Postdoc Researcher at Universitat Politècnica de Catalunya

[jvarga@essi.upc.edu](mailto:jvarga@essi.upc.edu)

**Linked Open Data with SPARQL**

**02 – Laboratory session**

**REQUIRED KNOWLEDGE FOR THE LAB**

The laboratory session will subsume the entire course but, specially, the **“Open Data”** lecture by Anna Queralt and the “**Linked Data and Semantic modelling**” lecture by Oscar Romero. For the Virtuoso laboratory, you should be familiar with the SPARQL endpoint of DBpedia (<http://dbpedia.org/snorql/>).

For that, you had to prepare the laboratory session with the following document:

“***160115\_Linked-Open-Data\_SPARQL\_before\_class.doc***”.

In this document, you learnt:

* How to browse the knowledge base with simple SPARQL queries
* How to find different kinds of entities
* How to design basic ontologies

**NEEDED TOOLS**

This lab only requires the web browser to access to the SPARQL endpoints. During the lab, you will have to realize the exercises directly with the provided SPARQL endpoints through any web browser in front of the lecturer.

Here is the SPARQL endpoints that you should use to answer the practice:

* <http://dbpedia.org/snorql/>

**DELIVERABLES**

During the Lab session, you will have to solve some exercises (the mandatory exercises PART B) that the lecturer will provide. At the end of the practice, you will have to deliver your solution in one document with the different SPARQL you have to create.

Also, if you want to solve more complex exercises, you will be able to realise some homework to produce optional queries (PART C).

For both parts, a solution template will be provided to you and you just have to fill it with your solution:

* 160122\_LOD-SPARQL\_mandatory\_<Names>.doc
* 160122\_LOD-SPARQL\_optional\_<Names>.doc

This document will have to be uploaded to the Moodle system or sent by email.

Exercises:

**Part A: Questions (30 minutes)**

The first 30 minutes of the session will be spent on answering all the possible questions you might have come up with during the week by preparing the tasks before the session (see above).

**Part B: In class practice (2h 30)**

Statement:

Solve the following queries with SPARQL. You are suggested to validate your answers with the SPARQL Explorer (<http://dbpedia.org/snorql/>). If you create any PREFIX include it in your answer. You can also check the DBpedia ontology schema at: <http://dbpedia.org/ontology/> or <http://mappings.dbpedia.org/server/ontology/classes/>.

During the practice, you have to create different SPARQL queries:

1. **Basic SPARQL queries (15 min)**

During the introduction, the very simple query to find a Person called Tim was presented:

SELECT ?person

WHERE {

?person rdf:type <http://dbpedia.org/ontology/Person> .

?person foaf:name ?name .

FILTER regex(str(?name), "Tim") .

}

For the section, you can use the following web page which list all the concept in a hierarchical way: <http://mappings.dbpedia.org/server/ontology/classes/>.

* 1. Create the SPARQL query to find all the locations which are a country:

You have to find all the values of the locations which are countries. For this, you will first have to find URI of the Country concept in the knowledge base.

* 1. Create the SPARQL query to find all the politicians:

You have to find all the values of the people who are politicians. For this, you will first have to find the URI of the Politician concept in the knowledge base.

* 1. Create the SPARQL query to find all the companies:

You have to find all the values of the organizations which are companies. For this, you will first have to find URI of the Company concept in the knowledge base.

1. **Discovering the main entities and the semantic relations (30 min)**
   1. Create the SPARQL query to find all the sub-concepts of the concept Person in the knowledge base:

You have to find all the concepts that are sub-concepts of the concept Person. For this, you will have to use the specific property that exists between the concepts.

This query should be generic and will be used as a template to find other kind of concepts.

* 1. Create the SPARQL query to find all the properties that can be used to describe a person:

You have to find all the properties which belongs to the domain of the concept Person. All the retrieved properties can be used to describe a person.

This query should be generic and will be used as a template to find other kind of concepts.

* 1. Create the SPARQL query to find all the properties that can be used to describe a company:

You have to find all the properties which belongs to the domain of the concept Company. All the retrieved properties can be used to describe a company. For this, you can just adapt the previous query (b) to the concept Company.

* 1. Create the SPARQL query to find a specific Property in the knowledge base

You have to find the property of the concept “Company” that refers to the industry that the company belongs to.

This query should be generic and will be used as a template to find other properties.

* 1. Create the SPARQL query to find a specific value in the knowledge base

You have to find the Company that is called “IBM”. Also based on regular expression, you can search for the specific instance that contains the word “IBM”.

This query should be generic and will be used as a template to find other values.

1. **Creating SPARQL queries (45 min)**

With the previous queries, you should be able to find any kind of concepts and their semantic relations with other concepts. For the mandatory queries, you have to create at least 3 of the following queries:

1. Create the SPARQL query to find all the companies that have the United States listed as location and that work in the Media industry . In the results, you should display the Company name, the industry and the location.
2. Create the SPARQL query to find all the companies in the Automotive industry ordered by number of employees (in descending order). In the results, you should display the Company name, the industrial sector and the number of employees.
3. Create the SPARQL query to find all the Politicians born in United States. You should display in the results the name of the politician, the city where he was born and his function. The results should be order by cities.
4. Create the SPARQL query to find the Leaders of all the countries with a population higher than 10 million. In the results, you should display the country name, the leader name and the population. The list should be ordered by the country's population (descending order).
5. Create the SPARQL query to find all the movies directed by Scorsese and published after 1990. In the results, you should display the name of the director, the name of the movie and the date of the release.

For each query, you have to search for the relevant concepts and semantic relations using the different templates created in the previous sections.

**Part C: Optional practice (homework)**

1. **Creating SPARQL queries**
   1. If not yet finished, create the missing queries of the Part B.

In the part B.2, you have 5 exercises to create the SPARQL queries but only 3 are mandatories. If you didn’t finished all the exercises of this part, they should be realized as the optional exercises.

* 1. Create the query to find the actor of movies who were produced by Indian companies during the period between 1960 and 1990.

In the results, you should display the name of the actor, the movie name, the date of the release, and the name of the director, the company name and the location of the company. You should distinct the results by movies and order them by ascending date.

1. **Designing basic ontologies**

When the previous query is realised, you have to design the schema of the ontologies used to find this information in the knowledge base. Concepts and properties should be represented as a semantic graph.

1. **Build a federated query**

By using the federated mechanism of SPARQL, you should create a query that permits to collected knowledge from two different SPARQL endpoints, for example:

* <http://dbpedia.org/snorql/>
* http://lod2.openlinksw.com/sparql

First, you have to find a common concept between the two knowledge base.

Then, you have to find instances of this concepts from both repositories.