

## 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**

**01 – Preparation of the Laboratory session**

**OBJECTIVES**

This is a hands-on lab to get familiar with the triple stores (RDF repositories). The main goals are as follows:

Given a SPARQL endpoint, you are supposed to:

* Query the knowledge base in order to answer each of the queries,
* Understand the semantic relations between the entities,
* Design the main ontologies used for this practice.

Try not to think relational-wise and also grasp the subtle details. That is, get familiar with the pros and cons of such systems so that in the future you can decide when to use them.

**REQUIRED KNOWLEDGE FOR THE LAB**

The laboratory session will subsume the entire course but, in particular the Open Data lecture (with Anna) and the Semantic Web lecture (with Oscar). For the Linked Open Data with SPARQL laboratory, you should be familiar with the SPARQL endpoint of DBpedia (<http://dbpedia.org/snorql/>). For that, you have to follow all the instructions of the present document.

**NEEDED TOOLS**

This lab requires the web browser to access to the SPARQL endpoints. During the lab, you will have to realize the exercises directly in the provided SPARQL endpoints through any web browser in front of the lecturer. Then, to prepare the lab correctly, you have to learn how to create SPARQL queries.

**TRAINING (ACTIVITIES TO DO DURING THE WEEK)**

To prepare the practice in class, you have to realise the following tasks:

* First, you learnt the SPARQL syntax during the Semantic Web lecture with Oscar Romero. Some exercises have been already done during this session.
* To go further, you have to follow the SPARQL tutorial to learn the basic concepts: <http://www.slideshare.net/LeeFeigenbaum/sparql-cheat-sheet>
* Then, to get familiar with the basics, you can also practice by creating more SPARQL queries. For this, you can try the different examples of SPARQL queries proposed in this tutorial: <http://www.w3.org/2009/Talks/0615-qbe/>. You should especially try the examples where it is mentioned “*Try it with*[*a DBPedia-specific SPARQL endpoint*](http://dbpedia.org/sparql)”. There you will be able to learn how to:
  + How to browse the knowledge base with simple SPARQL queries
  + How to find different kinds of entities
  + How to query several RDF repositories with Federated queries
* With the previous tutorial, you already tried to use the SPARQL endpoint of DBpedia: <http://dbpedia.org/sparql> or <http://dbpedia.org/snorql/> (this is the same knowledge base but with different user interface). Now, you should be able to customize the proposed SPARQL queries to get more complex results.

**LABORATORY ORGANIZATION**

We will have one lab session on Friday, February 3rd that is an application of the theoretical part “**Linked Data and Semantic modelling**” lecture by Oscar Romero and the **“Open Data”** lecture by Anna Queralt.

During the lab session, the lecturer will be present at the practice and will help you out. Thus, it is a session to solve your doubts and help you to prepare the final delivery.

The lab session has a 3 hours duration and is organised as following:

**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)**

The specific questions will be provided at class. You will work in:

1. Creating SPARQL queries
2. Discovering main entities and the semantic relations
3. Designing basic ontologies

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

The specific practice will be provided at class. You will work with the SPARQL endpoints to query several knowledge repositories.

At the end of the lab session, you will have to provide your solution to solve the mandatory queries. And if you want to go further more optional queries can be done during or after the lab session. The solution of this optional queries can be delivered one week after.

**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

These documents will have to be uploaded to the Moodle system.