

Medical Informatics: Tutorial 4

RDF Modelling

- Please attempt all questions on this worksheet in advance of the tutorial, and bring with you all work. Tutorials cannot function properly unless you do the work in advance.
- You are welcome to bring along any questions you may have, e.g. from the lectures.
- Assessment is formative, meaning that tutorials do not contribute to your final grade.

Introduction

In this tutorial you will get to practise with the RDF data model. In particular, you will familiarise yourself with DBpedia and you will have a look at a medical ontology. More importantly, you'll get to write your first statements in RDF.

Exercise 1: Exploring DBpedia

In this exercise, you will familiarise yourself with DBpedia, one of the most prominent RDF knowledge bases. Go through the following tasks:

1. Access the DBpedia page about the Royal Infirmary of Edinburgh and have a look at its content.
2. Pick 5 entries that you find interesting and write down how you would express them in English.
3. Each of these statements is essentially an RDF triple. Can you identify the subject, the predicate and the object of each triple?
4. If you haven't already done so for the previous task, identify the URIs used for each subject, predicate and object (if applicable) in each case.
5. Write down the 5 statements in the Turtle serialisation. Remember to specify prefixes if you're using QNames instead of full URIs.
6. From the menu on the top (within the DBpedia page for the Royal Infirmary of Edinburgh), select Formats → Turtle. Open the file with a text editor and identify 2 triples in which the full URI is used for a resource. What do these triples mean?

Exercise 2: Browsing a medical ontology

The Human Disease Ontology is an ontology for describing the classification of human diseases organised by aetiology. You can access it here: <http://disease-ontology.org/>

1. Use the navigation panel on the left to explore the disease hierarchy. What are the subclasses of “disease of cellular proliferation”? Is melanoma a cell type cancer or an organ system cancer?
2. Study the metadata for “melanoma”. How would you explain the different entries in simple English?
3. Get a visualisation for “melanoma” by clicking on the visualise button. What does this graph tell us?
4. Search for a disease of interest and have a look at its metadata. What other terms are there for that disease? What subclasses does it have, if any? What is its superclass?
5. How does the Human Disease Ontology relate to the ontology components that we discussed in Lecture 8?
6. What potential uses could you imagine for the Human Disease Ontology?

Exercise 3: Making statements in RDF

In this exercise, you will create your own RDF data. The RDF dataset should be captured in the Turtle serialisation format (and thus saved as a .ttl file), and it should consist of at least 6 triples. You can choose among two options for the topic of the data:

- Describe a book / film / music album that you like, using the Dublin Code vocabulary.
- Specify your FOAF profile, in other words describe yourself, using the FOAF vocabulary.

Before writing down your triples, you should familiarise yourself with the **vocabulary** that you’re going to be using. You can find a description of the different terms in the following webpages:

- For the FOAF vocabulary: <http://xmlns.com/foaf/spec/>
- For the Dublin Core vocabulary: <http://dublincore.org/documents/dcmi-terms/>

If you need to use terms from other vocabularies too, you may find the following websites useful:

- <https://lov.okfn.org/dataset/lov/>
- <http://vocab.cc/>

If you need to create your own vocabulary (do you really need to?), then you can create a URI using a namespace that you own or using <http://usher.ed.ac.uk/medinf/vocab/>.

You should also choose appropriate **URIs for the entities** in your dataset. DBpedia is very likely to have existing URIs for the things you want to talk about. Alternatively, a Google search might reveal more options.

Validate your RDF data to check it is syntactically correct, and fix any errors that may occur. There are many validators available, you could try this one (note: choose to 'translate' from N3 to N3): <http://rdf-translator.appspot.com/>