WK5_Graph_Data_Task

January 24, 2020

1 Graph Data Task

1.1 Importing rdflib

rdflib is one of the most widely used Python libraries for working with RDF data. We first need to tell Python that we're going to be using rdflib, with the use of the import command.

```
In [1]: # rdflib - one of the most widely used Python libraries for working with RDF data.

        import rdflib
```

1.2 Familiarising ourselves with the data

In this programming task we are going to be using a small RDF dataset about hospitals. The dataset includes information about 4 different hospitals: * Glasgow Royal Infirmary * Royal Infirmary of Edinburgh * Aberdeen Royal Infirmary * University Hospital Heidelberg

The information about each hospital may include: URI, name, number of beds, country, the fact that it is a hospital.

Here is an extract of the dataset:

<http://dbpedia.org/resource/Glasgow_Royal_Infirmary> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://schema.org/Hospital> . <http://dbpedia.org/resource/Glasgow_Royal_Infirmary <http://xmlns.com/foaf/0.1/name> "Glasgow Royal Infirmary" . <http://dbpedia.org/resource/Glasgow_Royal_Infirmary> <http://dbpedia.org/ontology/country> <http://dbpedia.org/resource/Scotland> . <http://dbpedia.org/resource/Glasgow_Royal_Infirmary> <http://dbpedia.org/ontology/bedCount> 1077 .

```
In [ ]: # Will use small RDF dataset about hospitals, including information about 4 different
        # - Glasgow Royal Infirmary
        # - Royal Infirmary of Edinburgh
        # - Aberdeen Royal Infirmary
        # - University Hospital Heidelberg
```

1.2.1 Parsing the data

The dataset is called *hospitals.ttl* and it can be found in the *readonly* directory.

Let's parse the dataset with the use of the *parse* function. You can ignore anything printed.

```
In [2]: # Dataset: called hospitals.ttl - can be found in readonly directory.

        # USE:
        # .parse() function to parse dataset.

        from rdflib import Graph
        g = Graph()
        g.parse("./readonly/hospitals.ttl", format="ttl")

Out[2]: <Graph identifier=N35fd358c761f4d8f81c7dcbae0fc0a3d (<class 'rdflib.graph.Graph'>)>

In [ ]: # Creates new graph object called g.
```

This creates a new Graph object called *g*.

1.2.2 Getting a sense of the data

We can get the number of triples in our dataset with the use of *len*.

Run the code below (the result should be 14).

```
In [3]: # Number of triples in dataset

        len(g)

Out[3]: 14
```

To print out all the triples in the dataset in the form of *subject predicate object*, run the code below.

Note that it is not recommended to print an entire dataset, as it may be very large! For this example it is ok, though.

```
In [4]: # Print out all triples in dataset:
        # s - subject
        # p - predicate
        # o - object

        for s, p, o in g:
            print(s, p, o)
```

```
http://dbpedia.org/resource/Glasgow_Royal_Infirmary http://xmlns.com/foaf/0.1/name Glasgow Roya
http://dbpedia.org/resource/University_Hospital_Heidelberg http://www.w3.org/1999/02/22-rdf-syn
http://dbpedia.org/resource/Royal_Infirmary_of_Edinburgh http://dbpedia.org/ontology/country ht
http://dbpedia.org/resource/Glasgow_Royal_Infirmary http://www.w3.org/1999/02/22-rdf-syntax-ns#
http://dbpedia.org/resource/Royal_Infirmary_of_Edinburgh http://xmlns.com/foaf/0.1/name Royal I
http://dbpedia.org/resource/Aberdeen_Royal_Infirmary http://www.w3.org/1999/02/22-rdf-syntax-ns
http://dbpedia.org/resource/Glasgow_Royal_Infirmary http://dbpedia.org/ontology/country http://
http://dbpedia.org/resource/Aberdeen_Royal_Infirmary http://xmlns.com/foaf/0.1/name Aberdeen Ro
http://dbpedia.org/resource/Aberdeen_Royal_Infirmary http://dbpedia.org/ontology/bedCount 922
http://dbpedia.org/resource/University_Hospital_Heidelberg http://dbpedia.org/ontology/country
```