Exercise Sheet 4: More SPARQL and Wikidata
Maximilian Marx, Markus Krötzsch
Knowledge Graphs, 2023-11-26, Winter Term 2024/2025

Exercise 4.1. (Exercise 2.5)
Write a program that reads a graph in N-Triples format and checks whether the graph is bipartite. Use it to decide whether `authorship.nt.gz`[1] and `coauthors.nt.gz`[1] are bipartite.

Hint: each of the uncompressed graphs is roughly 4 GiB in size. In Python, you can use `gzip.GzipFile`[2] to process the compressed file without decompressing it first. There is also `authorship-snippet.nt.gz`[1], a small part of the graph that you can use during development.

Please note: In order to get the correct data files, please install git-lfs[3] on your system, and then activate it in your local repository (git lfs install).

Exercise 4.2. (Exercise 2.6)
The bibliographic database DBLP[4] offers individual data records as RDF in N-Triples format. This data can be downloaded from the URL obtained by appending `.nt` to the URI. Use this interface to find all publications that have https://dblp.org/pid/s/RudiStuder as their only author.

- Download some RDF files in your browser to find out how this information is encoded.

- Write a program that crawls a small part of the data to answer the query.

Note: If your program sends too many requests in a short time, the server will deny the request and cancel the connection. Dirty trick: use `time.sleep(1)` before executing a request.

Hint: `requests`[5] provides a high-level API for making HTTP requests in Python, but you may need to install it, e.g., using `pip`.[6] A built-in alternative that provides a lower-level interface is `urllib.requests`.[7]

Exercise 4.3. (Exercise 3.4)
Using the Wikidata query service,[3] find the people that have received more than one Nobel Prize. Does your query find all five persons that have won two Nobel prizes? Why/Why not?

Exercise 4.4. (Exercise 3.5)
Is it possible to count the number of triangles in the P47 ("shares border with") property using the Wikidata query service[3]? Why? Why not? Is it possible to count the number of triangles of federal states of Germany?

Exercise 4.5. Solve Exercise 4.2 (i.e., Exercise 2.6) by querying the DBLP SPARQL endpoint https://sparql.dblp.org: find all publications having only https://dblp.org/pid/s/RudiStuder as their sole author.

---

[1]https://github.com/knowsys/Course-Knowledge-Graphs/tree/main/data/dblp
[2]https://docs.python.org/3/library/gzip.html
[3]https://git-lfs.github.com/
[4]https://dblp.org
[5]https://requests.readthedocs.io/en/latest/
[6]https://pypi.org/project/pip/
[7]https://docs.python.org/3/library/urllib.request.html
[3]https://query.wikidata.org/

Exercise 4.6. Use the Wikidata query service (WDQS)[1] to find all people that have returned from two spaceflights operated by organisations from different countries and the amount of time they have spent in space. You can expect Q255764 ("Yelena Kondakova") in the results.

Hint: For all the exercises using the WDQS, you can use the SQID browser[2] to explore the schema.

---

[1] https://query.wikidata.org
[2] https://sqid.toolforge.org/