Exercise Sheet 5: Wikidata
Maximilian Marx, Markus Krötzsch
Knowledge Graphs, 2018-11-20, Winter Term 2018/2019

Exercise 5.1. Inspect the Wikidata page of item Q323270 ("Saxon State and University Library"), and download its RDF-encoded data in Turtle format from https://www.wikidata.org/wiki/Special:EntityData/Q323270.ttl. Find out how the statement for the property "coordinate location" is encoded in the RDF graph, and draw the subgraph containing all triples in the representation, including all auxiliary nodes introduced for this statement (you may abbreviate URIs and labels).

Hint: The mapping from Wikidata statements to RDF is described in detail in the online documentation "RDF Dump Format".[1]

Exercise 5.2. Use the Wikidata query service[2] to find all characters that appear in two distinct works that are instances of a subclass of Q7725634 ("literary work").

Hint: You can use the SQID browser[3] to explore how literary works are modelled on Wikidata. Start from a simpler query and gradually refine it. For simple queries, it is good to use LIMIT to avoid overly large results while testing.

Exercise 5.3. Use the Wikidata query service[2] 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.

Hint: Q255764 ("Yelena Kondakova") should be among the results.

Exercise 5.4. Using the Wikidata query service,[2] find all people that have received more than one Nobel Prize. Does your query find all four persons that have won two Nobel prizes? Why/Why not?

Hint: Use the SQID browser. Start from Q7191 ("Nobel Prize") and look for *direct instances*, *own statements*, and *statements from related entities*.

Exercise 5.5. Write a program that takes as input `wikidata-p47.nt.gz`,[4] a subset of Wikidata in N-Triples format, and counts the number of triangles in the P47 ("shares border with") property. Make sure that you do not count other edges that may be present in the input.

Exercise 5.6. Using the Wikidata query service,[2] count the number of triangles in the P47 ("shares border with") property.
   Compare your results to those of Exercise 5.5.

Exercise 5.7. An instance of 4×4-Sudoku is a partially-filled table as illustrated below. The goal is to fill the remaining cells with values 1, 2, 3, and 4 such that no value occurs twice in a row, in a column, or in one of the four 2×2 blocks.

1			
	2		
		4	
			3

Use SPARQL to solve this problem: find a query that returns all admissible ways of filling the grid as its answers.

1. First, define a suitable SPARQL query *and* underlying RDF graph to solve the problem.

2. Then show that your query can be modified to work using the Wikidata query service[2] over the RDF data of this system.