Exercise Sheet 4: More SPARQL and Wikidata
Maximilian Marx, Markus Krötzsch
Knowledge Graphs, 2023-11-14, Winter Term 2023/2024

Exercise 4.1. 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.

Exercise 4.2. 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 WDQS over the RDF data of this system.

Exercise 4.3. A *k-clique* in a simple graph $G = \langle V, E \rangle$ is a set $C = \{ v_1, v_2, \ldots, v_k \}$ of k vertices, where any two vertices $v, w \in C$ are adjacent, i.e., $\{ \{ v, w \} \mid v, w \in C \} \subseteq E$. Recall that a *simple path* from vertex s to vertex t is a sequence of vertices $p_0, p_1, \ldots, p_\ell$ with $\ell > 0$ and $s = p_0 \xrightarrow{e_1} p_1 \xrightarrow{e_2} \cdots \xrightarrow{e_\ell} p_\ell = t$ such that if $p_i = p_j$ for some $i \neq j$, then $\{ i, j \} = \{ 0, \ell \}$.

Compute the function $f : \mathbb{N} \to \mathbb{N}$ that maps a number k to the number of distinct simple paths $f(k)$ in a k-clique. What is $f(5)$?

Exercise 4.4. Use the WDQS to check for the existence of a 5-clique in the P3373 ("sibling") property.

Exercise 4.5. Consider a simple bipartite graph $G = \langle V, E \rangle$ with $V \neq \varnothing$. Show that the following are equivalent:

1. G has exactly two distinct 2-colourings

2. G is connected

Exercise 4.6. Use the Wikidata query service[1] to find answers to the following queries:

1. the top 30 composers with the most musical works whose English label is longer than the average English label of musical works,

2. the top 20 bands by the number of former members who are still alive (band members are modelled using P527 ("has part") and P463 ("member of")),

3. the top 42 music genres by the number of bands and musicians, and

* 4. for every sovereign state (Q3624078), the music genre(s) with the most bands or musicians from this state.

Hints:

- Use the SQID browser[2] to explore the schema.

- Some of the queries can be written in different ways. While this should not have any impact on the results, it might lead to different query execution plans. Hence, when your query times out, try a different approach.

- GROUP_CONCAT does not work on labels injected via the label service, you will need to retrieve these from the RDF data, or explicitly list them as parameters to the label service. They are represented using the rdfs:label predicate; you can use FILTER(LANG(?label) = "en") to restrict the results to English labels.