Exercise Sheet 8: Advanced SPARQL

Maximilian Marx, Markus Krötzsch

Knowledge Graphs, 2025-01-07, Winter Term 2025/2026

Exercise 8.1. Given a formula φ of propositional logic, show how to decide $\varphi \in \mathbf{SAT}$ using a SPARQL query that does not contain any BGPs.

Exercise 8.2. Which of the following QBF are true QBFs? Why?/Why not?

1. $\exists p_1.\ p_1$
2. $\forall p_1.\ p_1$
3. $\exists p_1.\ \bot$
4. $\forall p_1.\exists p_2.\ p_2 \to p_1$
5. $\forall p_1.\exists p_2.\forall p_3.\ (p_1 \vee p_2) \wedge p_3$
6. $\forall p_1.\forall p_2.\exists p_3.\forall p_4.\ (p_1 \wedge (p_2 \to p_4)) \vee \neg p_3$

Exercise 8.3. Can you write a SPARQL query for the Wikidata Query ServiceWikidata QLever endpoint[1] that finds all persons related to Q1339 (“Johann Sebastian Bach”) by a path going through P40 (“child”), P25 (“mother”), or P26 (“spouse”) edges, such that every person on this path has a statement for property P1303 (“instrument”) with value Q1444 (“Organ”)? How?/Why not?

Does anything change if you relax the restriction on all persons on the path and only require that they have a statement for property P1303 (“instrument”) with an arbitrary value?

Exercise 8.4. Find a family of SPARQL queries that produce solutions where a variable name is mapped to a value that requires an exponential number of characters to write down (measured in the size of the query and RDF graph). What can you say about the growth of the result’s size with respect to the size of the RDF graph when keeping the query fixed?

Exercise 8.5. Use the Wikidata QLever endpoint[1] to find answers to the following queries:

1. the top 25 universities that employed the most professors that have been educated there (Q16188175 (“Ingerid Dal”) is one such professor),
2. the top 10 musical instruments played by people who are composers by occupation or have composed something (Q1339 (“Johann Sebastian Bach”) is a good starting point to explore the schema),
3. the top 30 composers with the most musical works whose English label is longer than the average English label of musical works,
4. 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”)),

[1] https://qlever.dev/wikidata

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

* 6. 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.

[2] https://sqid.toolforge.org/