Exercise Sheet 7: More Advanced SPARQL

Maximilian Marx, Markus Krötzsch

Knowledge Graphs, 2024-12-17, Winter Term 2024/2025

Exercise 7.1. (6.1)
Use the Wikidata query service[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")),
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.

Exercise 7.2. Consider the RDF graph G:

```
eg:x eg:edge eg:x ;
     eg:value 1 .
eg:y eg:edge eg:x, eg:y ;
     eg:value 2 .
eg:z eg:edge eg:x, eg:y, eg:z ;
     eg:value 3 .
```

Evaluate the following expression of the SPARQL algebra over G:

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

```
Group(⟨ ?s ⟩,
  LeftJoin(
    BGP_G(?s eg:value ?v),
    BGP_G(?s eg:edge ?o),
    ?s != ?o)
)
```

The semantics of grouping are as follows: Consider some list of expressions $\Phi = \langle \varphi_1, \ldots, \varphi_n \rangle$. For a solution mapping μ, define $\Phi(\mu)$ as the list $\langle \varphi_1(\mu), \ldots, \varphi_n(\mu) \rangle$ of values obtained by evaluation these expressions for the bindings of μ. Then

$$\mathsf{Group}(\Phi, M) = \left\{ \Phi(\mu) \mapsto \{\mu' \in M \mid \Phi(\mu') = \Phi(\mu)\} \mid \mu \in M \right\}$$

Exercise 7.3. Wikidata also contains lexicographic information: *Lexemes* are entities that have a language, a *Lemma* (the actual character sequence), and support claims the same way that other Wikidata entities do. In the Wikidata query service, lexemes are encoded using an `rdf:type` of `ontolex:LexicalEntry`. The language is identified by `dct:language`, and the lemma by `wikibase:lemma`. A lexeme can also have *senses* specifying their meaning, they can be reached by the `ontolex:sense` property, from which `wdt:P5137` connects to the corresponding Wikidata item.

Using this, write a query that uses the Wikidata query service[3] to find the top 10 languages by the number of lexemes that have at least one meaning corresponding to some kind of snow.

Exercise 7.4. 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 7.5. 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. $\forall p_1 . \forall p_2 . \exists p_3 . \forall p_4.\ (p_1 \wedge (p_2 \to p_4)) \vee \neg p_3$

[3] https://query.wikidata.org