Exercise Sheet 7: SPARQL Semantics and Wikidata

Maximilian Marx, Markus Krötzsch

Knowledge Graphs, 2025-12-17, Winter Term 2025/2026

Exercise 7.1. Show that Theorem 6.6 from the lecture fails in the presence of blank nodes: find disjoint BGPs P_1 and P_2 such that

$$\mathsf{BGP}_G(P_1 \cup P_2) \neq \mathsf{Join}(\mathsf{BGP}_G(P_1), \mathsf{BGP}_G(P_2)).$$

Exercise 7.2. Show that there are sets of solution mappings M_1 and M_2 such that

- each solution in M_1 is compatible with each solution in M_2,
- M_1 and M_2 together contain more than two solutions, and
- $\mathsf{Join}(M_1, M_2)$ contains just one solution.

Note: for simplicity, we only consider sets here instead of multisets, and ignore multiplicities of solutions.

Exercise 7.3. Transform the following SPARQL group graph pattern into an expression of the SPARQL algebra. List all intermediate results.

```
{ ?person rdfs:label ?personLabel . FILTER (LANG(?personLabel) = "en")
  { ?person wdt:P166 wd:Q185667 } UNION
  { ?person wdt:P166 wd:Q1417143 }
  OPTIONAL { ?person wdt:P800 ?notableWork }
}
```

Exercise 7.4. 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[1] to find the top 10 languages by the number of lexemes that have at least one meaning corresponding to some kind of snow.

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