**Exercise Sheet 7: More Advanced SPARQL**
Maximilian Marx, Markus Krötzsch
Knowledge Graphs, 2019-12-03, Winter Term 2019/2020

**Exercise 7.1.** 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 7.2.** 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 $G$:

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

**Exercise 7.3.** Can you write a SPARQL query for the Wikidata Query Service[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 7.4.** 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.

---