Exercise 5.1. 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

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

Exercise 5.2. 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 5.3. 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 5.4. 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)
)
```

   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 5.5. 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.

---

[1]https://query.wikidata.org
[2]https://tools.wmflabs.org/sqid/#/