

Exercise Sheet 6

Submit until Tuesday, **December 5 at 12:00 pm (noon)**

Exercise 1 (10 points)

Write a program that reads a SPARQL query from a given file, translates it to a corresponding SQL query, and executes that query on a given SQLite3 database. The idea of the translation and the (simple) Python interface to SQLite3 were explained in the lecture. Your program should support all SPARQL queries of the form

```
SELECT <variables> WHERE { <triples> }
```

where <variables> is a whitespace-separated list of variables, <triples> is a dot-separated list of triples, where the three components of each triple (subject, predicate, object) and the dots are separated by whitespaces, and each component can be either a variable (e.g. *?x*), an identifier (e.g. *Marie_Curie* or *wd:Q937*), or a string literal (e.g. "Canis lupus familiaris") for objects. Also support *ORDER BY ASC*(<var>), *ORDER BY DESC*(<var>), and *LIMIT* <*k*> clauses.

Start from the code template given on the Wiki. In particular, that template provides the full implementation of a function *parse_sparql* that parses a given SPARQL query into the components relevant for this exercise.

Make sure that the test cases given in the code template on the Wiki are fulfilled, but feel free to adapt syntactic details of the test cases to your particular implementation. This is especially relevant for this exercise, since there are many ways to express the same SQL query (in particular, with different variable names or with a different order of the conditions in the *WHERE* clause).

Exercise 2 (5 points)

The ten youngest CEOs born in the US, with their gender and birth date

For the query above, formulate the appropriate SPARQL query on the dataset *wikidata-simple.tsv* linked on the Wiki (use *grep* to figure out the names of the entities and relations needed for the query). If you find several possibilities for formulating the query, prefer the one that gives the most results. Save your SPARQL query in a file *ceos.sparql*.

```
SELECT ?pto WHERE { wd:Q7204254 rdfs:label ?pto }
```

Build a SQLite3 database from the dataset, with a single table with columns *subject*, *predicate*, *object*, and create an index for each column, as discussed in Lecture 5. Run the SPARQL query using your program from Exercise 1, and report the number of query results and the query time in your *experiences.md*.

Exercise 3 (5 points)

All British monarchs and their reigning periods, most recent first

Repeat Exercise 2 for this query, but on the dataset *wikidata-complex.tsv* linked on the Wiki. In that dataset, the IRIs are the original Wikidata identifiers (instead of human-readable names like in Exercise 2), and there are statement nodes as discussed in Lecture 6. To figure out the appropriate identifiers for your SPARQL query, use Wikidata or, if you want, <https://qllever.cs.uni-freiburg.de/wikidata> (but beware that the latter searches the complete Wikidata, not just the dataset for this exercise sheet).

Like for Exercise 2, build a SQLite3 database from the dataset, with a single table with columns *subject*, *predicate*, *object*, and create an index for each column. You find the ground truth for the query in https://en.wikipedia.org/wiki/List_of_British_monarchs (which is a manually compiled Wikipedia article). It's possible, but not easy to get all these with a single query. If you find a single query, call your query file *monarchs.sparql*. If you want, you can formulate multiple queries, which give the complete result when concatenating their individual results. In that case, call your query files *monarchs.1.sparql*, *monarchs.2.sparql*, etc.

Commit your code and query files to our SVN, in a new subfolder *sheet-06*. The dataset files should not be committed.

As usual, in your *experiences.md*, provide a brief account of your experience with this sheet and the corresponding lecture. Make sure to add a statement asking for feedback. In this statement specify to which degree and on which parts of the sheet you want feedback. In addition, say how much time you invested and if you had major problems, and if yes, where.

What happened in 1707?