



Exercises for *Foundations in Data Engineering*, WiSe 23/24

Alexander Beischl, Maximilian Reif (i3fde@in.tum.de)

<http://db.in.tum.de/teaching/ws2324/foundationsde>

Sheet Nr. 14

Exercise 1 RDF Query Optimization

- a) Given the following RDF triples, determine the set \mathcal{S}_C of characteristic sets $S_C(s)$ for subjects s .

S	P	O
Catch Me If You Can	publication date	2002
Catch Me If You Can	IMDb ID	tt0264464
Leonardo DiCaprio	starred in	Catch Me If You Can
Leonardo DiCaprio	award received	Golden Globe Award for Best Actor
Grace Hopper	award received	Presidential Medal of Freedom
Tom Hanks	starred in	Catch Me If You Can
Tom Hanks	award received	Presidential Medal of Freedom
Tom Hanks	member of political party	Democratic Party
Nathalie Baye	starred in	Catch Me If You Can
Nathalie Baye	award received	César Award for Best Actress

- b) Use the characteristic set to estimate the result cardinality of this query:

```
select distinct ?a
where {
  ?a <starred in> ?m.
  ?a <award received> ?n.
}
```

Exercise 2 Solve the following tasks by using XQuery. You can test your queries at: <http://xquery.db.in.tum.de> by using the doc('uni3') dataset.

- Return all professors who give one or more lectures.
- Find all students who attend all lectures.
- Return students with the highest number of semesters, use aggregate functions.
- Calculate the sum of semester periods per week (sws) for each professor. Include professors without any lecture in your results.
- Find all students attending all four-hour lectures.
- Return the name of all students who never had a grade better than 3.0 in any exam.
- Calculate the examination material's extent for each student. Therefore, return the name of the student and the sum of semester periods per week of all lectures in which the student wrote an exam.
- Some of the students' names contain the name of a professor. Return all such students.

- i) Determine the level of awareness of each professor among students. We assume students only get to know professors by attending their lectures or writing an exam.

Exercise 3 The queries in this exercise work on the mondial dataset

```
http://dbis.informatik.uni-goettingen.de/Mondial/mondial.dtd
http://dbis.informatik.uni-goettingen.de/Mondial/mondial.xml
```

For the execution of XQuery requests, you may use xqilla for example:

```
apt-get install xqilla
```

The example in exercise 1 assumes, that the `mondial.*` files are in the current working directory.

1. Set up an XQuery engine and output the whole mondial document.

Example:

```
let $mondial := doc("mondial.xml")/mondial
return $mondial
```

2. Find the names of all continents.
3. For every country, compute the average population of its cities.

```
<country name="Albania">79166.66666666667</country>
...
<country name="Finland">100842.11764705883</country>
...
```

4. For every country, compute the percentage of population living in cities to the whole population.

```
<country name="Albania">14.619271092376557</country>
<country name="Greece">18.632551932449434</country>
...
```

5. Find all continents with their names and countries in the continents.

```
<continent name="Europe">
  <country name="Albania"/>
  <country name="Greece"/>...
...
</continent>
...
```