



Exercise Sheet 6: Relational Algebra (until Thursday, 30.11.2017) (40 Points)

Please note: you need **50%** of all exercise points to receive the *Studienleistung* for this lecture. In order to pass the RDB I Module, you need both the *Studienleistung* **and** you need to pass the exam. Exercises have to be turned in until **Thursday before the lecture** either in the lecture hall or into our mailbox at the IFIS floor (Mühlenpfordtstraße 23, 2nd floor). Please do not forget your **Matrikelnummer** and your **tutorial group number** on your solutions. **If you forget** to write your Matrikelnummer and/or your tutorial group number, you get **automatically 0 points**. Your solutions may be in German or English. Unless otherwise specified: **Always use your own words!**

Exercise 6.1 (5 Points)

The extended and advanced relational algebra operators can also be expressed by the 6 basic operations. Try to find a relational algebra expressions using the 6 basic operations to express the maximum aggregation function $\mathcal{F}_{\max}(R)$, for relation $R(a)$.

Exercise 6.2 (17 Points)

Consider the following schema:

City (name, country_code → Country, city_population)

Country (code, name, continent, surface_area, indep_year, gov_form, capital → City)

Language (lang_id, code → Country, name, percent, official)

The schema describes a miniworld of countries, cities and the languages spoken in those countries. Most of the attributes should be easy to understand. However, here are some clarifications:

- **City:** the name of the city is unique and each city belongs to a country given by country_code. The city population is an integer.
- **Country:** for each country there is a unique code (primary key); surface_area is a float value in square meters. The attribute indep_year is an integer that specifies the year of independence of the corresponding country. Attribute gov_form is a string with the specific form of government e.g. "Republic", "Constitutional Monarchy", "Federal Republic", "Monarchy", etc.
- **Language:** in each country more than one language could be spoken in different proportions (percent attribute) and more than one language for a given country could be recognized as the official language (official is a Boolean attribute). Moreover, the primary key is just a consecutive number. Thus, for queries involving the name of the language one must consider the attribute "name".

Using the described schema, answer the following queries using Relational Algebra:

1. **(2 P)** Show the average population in Europe.
2. **(4 P)** Show the name and the continent of the countries that have at least the same official spoken languages as Canada.
3. **(2 P)** List the name of the capital with the highest population.
4. **(2 P)** Find the most frequently form of government in Africa.

5. **(3 P)** Show the names of the countries that have at least two official languages and a surface area greater than 65000 square meters.
6. **(4 P)** Show the names of the countries with the least percentage of population that can speak English. (Do not use aggregation operator)

Exercise 6.3 (9 Points)

Consider the following schema:

Student(ID, name, semester, faculty)

Team(ID1→Student, ID2→Student, homeworkID→Homework, score)

Homework(ID, description, max_score)

Likes(ID1→Student, ID2→Student)

The schema is a miniworld of students at TU Braunschweig. They have been organized in a social network using databases. We have collected information about students of all the bachelor semesters of three different faculties: Informatik, Mathematics, and Physics. A brief clarification of the schema follows:

Student: there is a student with a unique ID and a given first name in a certain semester in a certain faculty.

Team: the student with ID1 works in a team with the student with ID2 in a homework assignment with homeworkID. Team is a mutual relationship: if (666,007) is in the Team table, so is (007,666). The score attribute is the grade given to the students.

Homework: each homework assignment has a unique ID, a maximum number of points to give (max_score) and a description.

Likes: The student with ID1 likes the student with ID2. Liking someone is not necessarily mutual, so if (007,666) is in the Likes table, there is no guarantee that (666,007) is also present.

Answer the following queries using Relational Algebra.

1. **(2 P)** Find the names of all students who work in a team with someone whose name is Codd.
2. **(2 P)** Which faculties have students that do not work in teams?
3. **(2 P)** Find the name and the semester of all students who are liked by more than one other student.
4. **(3 P)** Find the name, semester and faculty of all the students who have teammates of the same semester.

Exercise 6.4 (9 Points)

Consider the following schema:

Answer(original →Tweet, answer →Tweet)

Tweet(id, text, date, time, user →User)

User(id, name, age, country)

Hashtag(id, name)

Hashtag_used(tweet → Tweet, hashtag →Hashtag)

- Date has the format JJJJMMDD, for example 20171123 for the 23. November 2017
- Time is in the format HHMM, for example 1500, for 15:00.
- The text of a tweet (text) has a maximum length of 280 characters.

1. **(4 P)** Find the user name of the user with the most answers to his tweets.
2. **(5 P)** Find the IDs of the tweets that used a hashtag for the very first time.