



Exercise Sheet 8: SQL I (until Thursday, 15.12.2016) (43 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 8.1 (5 points)

Answer the following questions:

1. What is the difference between correlated and uncorrelated subqueries? (2 points)
2. What is the difference between a set and a bag? How can you enforce the DBMS to return a set instead of a bag? (2 points)
3. What is the difference between the WHERE and the HAVING clause? (1 point)

Exercise 8.2 (15 points)

Consider the following relation schema.

Movie(id, title, year)

Person(id, name, gender, birthday)

Genre(name, description)

actor(person → Person, movie → Movie, role)

director(person → Person, movie → Movie)

reviewer(person → Person, movie → Movie, stars)

hasGenre(movie → Movie, genre → Genre)

The schema describes movie data, including the movie's title and the production year. Also actors, reviewers, and directors are described, by introducing the relationship types "actor", "reviewer", and "director", which connects persons with movies. The actor relationship type also includes the role played by the person in the respective movie. Persons are further described by their name, gender, and birthday. A movie can have several reviews. For example the movie "Aliens" can have a 5 star review given by a specific Person. The attribute "stars" of the "reviewer" has a value from 1 to 5 (lowest to highest) to basically indicate the opinion of a person on a movie. A movie can also have several genres, which are further described by a description attribute.

Based on this relation schema, please provide SQL statements for the following queries:

1. The names of all persons who were born in December. Assume the following format for dates DD.MM.YYYY, for instance '21.12.1981', means December the 21st of 1981. (2 points)
2. The title of all the movies that have been produced after 1979. (2 points)
3. The titles of all movies that have been reviewed by more than one person. (2 points)

4. The name of all persons who participated in a "drama" movie. (2 points)
5. The name of all persons who have reviewed the movie "The Angry Silence". (2 points)
6. The name of all persons who have reviewed movies with at least 3 stars. (2 points)
7. Considering the average number of movies persons played roles in. Return the names of all persons who played roles in more movies than the average. (3 points)

Exercise 8.3 (23 points)

Consider the following schema of a network of authors of scientific research papers

Author (id, name, numberOfPublications)

Collaborator (id1 → Author, id2 → Author)

Cites (id1 → Author, id2 → Author)

In words:

- An author in the table Author, has a unique id, a name and a number of publications.
- The table Collaborator states a mutual association between authors: if (22, 32) exist in the table, then also (32, 22) exist. This table has the information to find authors that have collaborated with each other to publish a paper (e.g. to find the network of collaborators of an author).
- The table Cites states an association between authors too. However, in this table the meaning of the association is different. An author with id1 cites author with id2 in a publication, but the opposite is not guaranteed. Thus, if (22, 32) appears in the table, that does not mean that (32, 22) is also present.

Using the previous schema, write SQL statements to answer the following queries.

1. Show the names of the collaborator network of Tilo. (3 points)
2. Find the names of the authors who have published less than the most successful author. Successful here is measured by the number of publications. Thus, the more publications an author has, the more successful is considered to be. (3 points)
3. What is the average number of collaborators per author? (3 points)
4. Show the name and the number of publications of the authors who only have collaborated with authors that have the same number of publications as themselves. (3 points)
5. For each pair of authors that have cited each other, show the name and the number of publications of each author. Please, include only once each pair and with the two names in alphabetical order. (3 points)
6. Find the name and the number of publications of the authors who are cited by more than one author. (3 points)
7. Consider the following situation: for an author A1 who has cited an author A2 and the two authors have not collaborated, find if they have a collaborator A3 in common. For all those cases, show the name and the number of publications of A1, A2, and A3 (5 points).