Hand-in 2: ER model, SQL queries, and relational algebra

The hand-in should be sent to Andrea Campagna, acam@itu.dk at the latest

Wednesday October 1, 23.59.

Parts of the hand-in can be written by hand and scanned, but it has to be handed in electronically.

Hand-ins are done in groups of 3-4 students.

The course has 4 hand-ins that all have to be approved to enter the exam. All hand-ins will be corrected and you will get feedback from Andrea. The feedback may contain comments on what to improve to get the hand-in approved. Hand-in 2 can be improved and handed in again together with Hand-in 3, hence you will have two chances to get a hand-in approved. (The same will hold for the following 2 hand-ins.)

Hand-in 3 and 4 have to be handed in at the latest October 20 and November 10, respectively.

In this second part of the project you are asked to make an E-R model for the database and write queries on the data in the database, using both relational algebra and SQL.

In the first part of the project you defined a schema for your database. The database is a movie database storing information about actors/actresses, directors, writers, movies, awards, rating, etc. Now you are asked to make an E-R model of your database. (Some of you used your knowledge of E-R modeling already in the first part of the project, but it was not handed in.) The E-R model for your database can correspond to your design in Hand-in 1, or it can be an updated version of the design from Hand-in 1. If you have noticed problems with the first version of the design (by yourself or when you got feedback from Andrea on Hand-in 1) those problems have to be addressed in the updated design. Also, have a look at the queries below to make sure that all information necessary to answer the questions is available in your database.

Update the schema in SQL to correspond to the design in the E-R model. In case the schema is unchanged from Hand-in 1, include it anyway.

For the following queries formulate 1, 2, 4, and 7 in relational algebra and formulate all 8 queries in SQL. Note that the difficulty may vary depending on your data model. If you are not able to formulate a query in SQL you should include an
explanation of why the query is hard to formulate given your data model. If some data required is missing in your model, include it.

Queries:
1. Which movies have John Travolta and Uma Thurman starred in together?
2. Which actors and directors have a name starting with Q?
3. What is the name and birth year for all actors in Pulp Fiction? Your query should list the actors in increasing order of birth year.
4. What are the titles and years of all movies from the 90s where John Travolta starred?
5. How many Oscars has Quentin Tarantino won?
6. In 1994, what was the average length of a movie for each language?
7. Which actors in *Pulp Fiction* have never, before or after, starred in the same movie as one of the other actors in *Pulp Fiction*.
8. Which movie starring John Travolta has the highest user rating?

You are recommended to define the schema and to test the SQL queries in MySQL, however if you want to use standard SQL features not supported in MySQL this is allowed. If you have problems with foreign key constraints in MySQL, just leave them out of your schema and make a note of it in your hand-in.

All names of entities, attributes, relationships etc. have to be well chosen and in English.

To be handed in
- Name of all group members, including e-mail addresses
- An introduction describing the work
- A graphical presentation of the E-R model, including entities, relationships, attributes, cardinality constraints, and keys.
- A description of all (non-trivial) assumptions.
- A schema, written in SQL, corresponding to the E-R model.
- The 4 queries in relational algebra.
- The 8 queries in SQL.
- In case you were not able to solve some part of the hand-in, a description of the problem. (However, try to contact Andrea before the deadline if you have problems.) If you are not satisfied with some part of the hand-in, e.g. if you have discussed alternative solutions and don’t know which is the best choice, you can also add some comments on it. It can be valuable at your feedback meeting.

In case you have some issues from the first hand-in that have to be corrected to get it approved you should hand in:
- Your original hand-in 1.
- A description of the problems with the first solution and how they are solved in this hand-in.

You are not required to correct the first hand-in separately if your new design solves the problems.
Course goals covered by this hand-in

After the course the students should be able to:

- define a database design by E-R modelling, using the concepts entity, attribute, key, cardinality, and relationship.
- express simple relational expressions using the relational algebra operators select, project, join, intersection, union, set difference, and cartesian product.
- write SQL queries, involving multiple relations, compound conditions, grouping, aggregation, and subqueries.