DA 2032 - Database Theory

Assignment 2

This second assignment consists of two parts: A theoretical and a practical part. Deadline is *October 5, 2009*.

1   Theoretical Part

Consider the following two schemata:

> Schema 1: R(A,B,C,D)

> Schema 2: R1(A,B,C), R2(B,D)

1. Consider Schema 1 and assume that the only functional dependencies are $A \rightarrow B$, $C \rightarrow D$, and those possibly derived from these. Is Schema 1 in BCNF?

2. Consider Schema 2 and assume that the only functional dependencies are $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow A$, and $A \rightarrow D$, and those derived from these. Is Schema 2 in BCNF?

3. Take away the FD $A \rightarrow D$ from Question 2). Is Schema 2 then in BCNF?

4. Consider Schema 1 and assume that the only functional dependencies and multivalued dependencies are $A \rightarrow BC$, $B \rightarrow D$, $B \twoheadrightarrow CD$, and those derived from these. Is Schema 1 in NF4?

5. Consider Schema 1 and assume that the only functional dependencies and multivalued dependencies are $A \rightarrow BD$, $D \rightarrow C$, $C \twoheadrightarrow AB$, and $B \twoheadrightarrow D$, and those derived from these. Is Schema 1 in NF4 ?

2   Practical Part - Movie Database

Read in the movie data from the previous assignment. For this, write a program in a programming language of your choice. One idea is to write a program that produces a textfile containing a number of INSERT INTO statements and then use mysql in batch-mode:

```
mysql --user=root < batchfile.sql
```

If you are already running mysql, you can execute an SQL script file using:

```
mysql> source  file_name
```

If you run into performance problems, try grouping INSERT INTO commands together. How this works, you find in the documentation (one example is given in Chapter 3.6 `http://dev.mysql.com/doc/refman/5.0/en/examples.html`). If you run into other problems – e.g., timeouts – you need to experiment a little, just like in real life.

Another idea is to use the `Load Data` command, confer the MySQL documentation.

When reading in the data, follow these rules:

- A movie is defined uniquely by its *name* and *year*. This allows for, e.g., remakes.

- If you find a movie which is not contained in movies.txt, then ignore this movie, i.e., ignore datasets that reference movies not contained in movies.txt.

- In times.txt, ignore *episodes* and other non-integer information (e.g., "fps") and set the time of the corresponding movies to 0 (for unknown runtime). If a movie has more than one runtime specified, take the first one that you find in the file. For example, "Boy Who Turned Yellow, The 1972" has two times specified, so take "17" here.

Create SQL statements for the following queries:

1. What year did *Kasiping* air?

2. In how many movies has *Abel, Alfred* starred?

3. In which movies that *Aaronson, Evan* has produced did *Beers, Thom* star in?

4. In how many horror movies have people with last name *Clark* acted?

5. Which movies have a play time of 3 1/2 hours or longer?

6. How many movies have been produced, directed, and written by one and the same person?

7. What year was issued the maximum number of movies?

8. In which movie acted the minimum number of actors?

Hand in your program that populates the database, and the SQL statements together with the answers to the questions.