CS145 Written Assignment #3

Due Thursday April 22*

IMPORTANT: *Some problems in this assignment may require more than average thought—start early (remember there is also Programming Assignment #1), and don't hesitate to ask for help if you really get stuck.*

1. Consider a relation $R(A, B, C, D, E)$ with FD's $AB \to C$, $C \to D$, and $D \to AE$.

   (a) Find all keys of R.

   (b) $D \to AE$ is a BCNF violation for R. Suppose we decide to decompose R into $R_1(A, D, E)$ and $R_2(B, C, D)$. Give a minimal basis for the FD's that hold in R_1.

   (c) Find all keys of R_1.

   (d) Is there any FD in R_1 that violates BCNF? If so, decompose R_1 using this FD.

   (e) Give a minimal basis for the FD's that hold in R_2.

   (f) Find all keys of R_2.

   (g) Is there any FD in R_2 that violates BCNF? If so, decompose R_2 using this FD.

   (h) Are your results for (d) and (g) in BCNF? Explain briefly.

2. Consider a relation $R(A, B, C)$ with a given set of FD's $\mathcal{F} = \{AB \to C, A \to B\}$.

   (a) Prove that $\mathcal{F}$ is a *minimal* basis by showing that neither of the two FD's logically implies the other. (*Hint:* To show that fd_1 does not imply fd_2, we simply need to construct a relation instance that satisfies fd_1 but violates fd_2.)

   (b) Find another minimal basis $\mathcal{G}$ in which the left-hand side of every FD has only one attribute. (*Hint:* Try (d) first.)

   (c) To support your answer for (b), show that every FD in $\mathcal{F}$ follows from $\mathcal{G}$, and every FD in $\mathcal{G}$ follows from $\mathcal{F}$. For the proof, you can use either Armstrong's Axioms or attribute closure computation.

   (d) When (b) was given on an exam at a large western university, more than half the class answered $\mathcal{H} = \{A \to B, B \to C\}$. Show that answer is wrong by giving a relation instance that satisfies $\mathcal{F}$ but violates $\mathcal{H}$.

3. Here is a movie database with three relations:

   ```
   MovieDirector(title, director, year)
   MovieCast(title, actor, salary)
   MovieReview(title, reviewer, score)
   ```

   All attributes are strings except `year`, `salary`, and `score`, which are integers. Write the following queries in relational algebra. To save you some writing, you may use the abbreviated schema `MD(T,D,Y)`, `MC(T,A,S)`, and `MR(T,R,S)`. For complicated queries, you can use expression trees or assignments of intermediate results.

*Please refer to CS145 Course Information Page (`http://www.stanford.class/cs145/info.html`) for submission instructions and late policy.

(a) Find the directors who have also acted.

(b) Find the actors who have never directed.

(c) Find the director and the cast of *Casablanca*.

(d) Find all movies starring Vivien Leigh in the 1930's.

(e) Find the directors who reviewed their own films and gave themselves a score of 10.

(f) Find the actors who starred in at least one movie directed by Kurosawa.

(g) Find the actors who starred only in movies directed by Kurosawa.

(h) (Complete this part if you want to get a "+" for this assignment.) Find the actors who starred in every movie directed by Kurosawa.

(i) (Complete this part if you want to get a "+" for this assignment.) Find the actor with the highest salary in *Star Wars*.

4. **Personal Database Application (PDA)**

   Remind us of your final schema from Problem #4 of Written Assignment #2. Reconsider your schema in the light of the theory of FD's and BCNF. For each of your relations, tell whether its schema is in BCNF. If not, either redesign your schema so the relation is in BCNF, or give a rationale for leaving in in non-BCNF form (e.g., the amount of redundancy introduced is minimal, and splitting the relation would cause some reasonable queries to become multirelational). Indicate your final choice of design, whether or not you choose to decompose one or more relations.

   Is there anything you still don't like about the schema? Feel free to consult with one of the course staff in case you have any doubts. This is your last chance to fiddle with it on paper before committing it to bits in Oracle.