Suggested Reading:

During the first two weeks of the course, we will cover material contained in Chapters 3, 4, and 5 of the “Foundations of Databases” book (but not necessarily in the order presented in that book). You are also encouraged to read the original papers by E.F. Codd posted at the course webpages, as well as the first two sections of the survey article “Elements of Relational Database Theory” by P.C. Kanellakis (also posted at the course webpages).

Problems:

1. For each of the following operations on relations, give both a relational algebra expression and a SQL expression that defines it. For the relational algebra expressions, you do not need to know the names of the attributes of the relations. For the SQL expressions, you may choose suitable names for the attributes of the relations (you are free to choose the names, but please spell out what names for the attributes you chose).
   
   (a) The symmetric part \( R^* \) of a binary relation \( R \), where \( R^* \) consists of all pairs \((a, b)\) such that both \((a, b)\) and \((b, a)\) are in \( R \).
   
   (b) The symmetric difference \( R \triangle S \) of two ternary relations, where \( R \triangle S \) consists of all tuples that belong to exactly one of \( R \) and \( S \).
   
   (c) The composition \( R_1 \circ R_2 \) of two binary relations \( R_1, R_2 \), where \( R_1 \circ R_2 \) consists of all pairs \((a, c)\) for which there is an element \( b \) such that \((a, b) \in R_1 \) and \((b, c) \in R_2 \).

2. An airline maintains a FLIGHTS database that includes a table called DIRECT with two attributes FROM and TO containing information about direct flights between two cities. Give a relational algebra expression for the relation AT-MOST-TWO consisting of all pairs \((c, d)\) of cities such that one can travel from city \( c \) to city \( d \) with at most two intermediate stops.

3. This problem is about the semijoin \( R \Join S \) of two relations \( R \) and \( S \), which is the relation consisting of all tuples in \( R \) that “contribute” to the join \( R \Join S \). Semijoins play a useful role in database query processing and optimization.

   More precisely, the semijoin \( R \Join S \) of two relations \( R \) and \( S \) is the relation consisting of all tuples \( t \) in \( R \) such that there is at least one tuple in \( S \) that agrees with \( t \) in all attributes that \( R \) and \( S \) have in common.

   Assume now that the attributes of \( R \) are \( A, B, C \) and the attributes of \( S \) are \( B, C, D \). Give both a relational algebra expression and a SQL expression for \( R \Join S \).

4. Suppose we have a database schema consisting of the relation schemas

   \[
   \begin{align*}
   &\text{FREQUENTS(DRINKER, BAR)} \\
   &\text{SERVES(BAR, BEER)} \\
   &\text{LIKES(DRINKER, BEER)}
   \end{align*}
   \]

   Give relational algebra expressions for the following queries:
(a) Find the drinkers who frequent at least one bar that serves a beer they like.
(b) Find the drinkers who frequent every bar that serves a beer Brad Pitt likes.

Optional Problem: Give SQL expressions for these two queries.

5. Prove or disprove each of the following relational algebra identities. This means that if they are true of all relations \( R \) and \( S \) of the same arity, then give a proof, otherwise provide a counterexample of two relations \( R \) and \( S \) of the same arity for which the identity fails.

(a) \( \pi_{1,2}(R) \cup \pi_{1,2}(S) = \pi_{1,2}(R \cup S) \).
(b) \( \pi_{1,2}(R) \cap \pi_{1,2}(S) = \pi_{1,2}(R \cap S) \).
(c) \( \pi_{1,2}(R - S) = \pi_{1,2}(R) - \pi_{1,2}(S) \).
(d) \( \sigma_{\theta}(R - S) = \sigma_{\theta}(R) - \sigma_{\theta}(S) \).

6. This problem is about the independence of relational algebra operations from other relational algebra operations.

(a) Prove that the difference operation cannot be expressed in terms of the other four basic relational algebra operations union, cartesian product, projection, and selection.
(b) Suppose that someone proposes a new version of the relational algebra in which the difference operation is replaced by the intersection operation. In other words, in this new version, we consider relational algebra expressions that are built using union, intersection, cartesian product, projection, and selection. Does this new version of the relational algebra relationally complete? Justify your answer as best as you can.
(c) Optional Challenge Problem A: Prove that the union operation cannot be expressed in terms of the other four basic relational algebra operations difference, cartesian product, projection, and selection.
(d) Optional Challenge Problem B: Prove that the selection operation cannot be expressed in terms of the other four basic relational algebra operations union, difference, cartesian product, and projection.

Note 1: All written assignments, term project report or term paper, and final examination are to be submitted in a typeset format (preferably in \LaTeX{}).

Note 2: The purpose of the homework assignments is to help you develop a better understanding of the material covered in class and also to expose you to some material that lack of time prevents us from covering in class. You should try to work out these problems on your own using your notes from the class. You may also want to use the “Foundations of Databases” book and the papers posted at the course webpages as references. If you discuss any of the problems with other students in the class, you are still expected to produce your own write-up of the assignment; moreover, you are expected to state the names of all other students with whom you have shared ideas about the problems. Finally, please state what published or online sources you used, if any, in the solutions of the problems.