Logic, Data, and Incomplete Information

The Second Tsinghua Logic Summer School

Homework Assignment 1

(due Wednesday, June 29, 2022 - before class)

Problem 1: For each of the following operations on relations, give both a relational algebra expression and a relational calculus expression that defines it. For the relational algebra expressions, you may use derived relational algebra expressions discussed during the lectures.

1. 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.
2. The *symmetric difference* $R \triangle S$ of two ternary relations, where $R \triangle S$ consists of all triples that belong to exactly one of R and S.
3. 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 $(b, c) \in R_2$.

Problem 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 both a relational algebra expression and a relational calculus expression for the query 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.

Problem 3: This problem is about the *semijoin* $R \ltimes S$ of two relations R and S, which is the relation consisting of all tuples in R that "contribute" to the join $R \bowtie S$. Semijoins play a useful role in database query processing and optimization.

More precisely, the *semijoin* $R \ltimes S$ 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 relational calculus expression for $R \ltimes S$.

Problem 4: This problem is about the independence of relational algebra operations from other relational algebra operations.

1. Prove that the difference operation cannot be expressed in terms of the other four basic relational algebra operations union, cartesian product, projection, and selection.
2. 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 only. Can this new version of relational algebra express every query that is expressible in the standard relational algebra? Justify your answer.

Optional Challenge Problem: Prove that the union operation cannot be expressed in terms of the other four basic relational algebra operations difference, cartesian product, projection, and selection.