Data integration: Problem set #1

Problem 1

Assume α is a selection condition on a relation schema R. A new relational algebra operator $A_\alpha(R)$ is defined in the following way:

$$A_\alpha(r) = \begin{cases} r & \text{if } \sigma_\alpha(r) = \emptyset \\ \sigma_\alpha(r) & \text{if } \sigma_\alpha(r) \neq \emptyset \end{cases}$$

for any instance r of R.

Express $A_\alpha(R)$ using the basic operators of the relational algebra.

Problem 2

You are given a relation with N columns of the same type. Write an SQL query that returns the tuples having the *maximum number of repetitions* of the same value. The query should have size polynomial in N.

Problem 3

Assume a tree is represented as a set of facts of the form `parent(`x`,`y`)` where x is a parent of y in the tree. Write a Datalog program that computes for every pair of nodes in the tree their *lowest common ancestor* (and no other common ancestors). Test this program using XSB Prolog.

You may use the built-in XSB Prolog predicate x `\==` y meaning that x is different from y. You may not use negation.