CS589 Principles of Database Systems
Fall 2008 Homework 4
Assigned: Wednesday, November 19, 2008
Due: (NOTE – MONDAY!) Monday, December 1, 2008
100 points possible

1. Let r be a relation over relation schema R with schema(R) = XY and schema (S) = Y, where X and Y are both sets of attributes and all of the attributes of S (Y) are contained in the set of attributes of R.

   Then the relational algebra divide operator can be expressed as follows:

   $R \div S = \pi_X(R) - \pi_X((\pi_X(R) \times S) - R)$

   For the following schema, where s-id is the student id and c-id is the course id. The relation Core-course lists all of the required classes and the relation Completed represents that a particular student has completed a particular class.

   Completed(s-id, c-id)
   Core-course(c-id)

   a. (5 points) Write a relational algebra query WITHOUT using the divide operator that finds all student ids for students that have complete all of the core courses. That is, write a relational algebra query WITHOUT using the divide operator that is equivalent to Completed ÷ Core-course.

   b. (5 points) Write a query in Datalog (with negation, without recursion) that is equivalent to Completed ÷ Core-course.

   c. (5 points) Write a query in (tuple or domain) calculus that is equivalent to Completed ÷ Core-course.

2. Using the following schema:

   Person(<u>id</u>, name, age, gender)
   Car(<u>vin</u>, make, model, year, color)
   Owns(<u>id, vin</u>)

   (39 points) Write example, well-formed formulas (that could appear in a domain calculus query) to demonstrate each of the following positive and negative occurrences of variables. Note, you don't need to write 13 different formulas. One of your formulas might allow you to provide an example of more than one of the positive or negative occurrences. Be sure to be clear about which variable, in which formula, you are talking about for each answer.

Positive occurrences of a variable x	**Negative occurrences of a variable x**
$R(y_1, y_2, \dots, x, \dots, y_k)$	
x=v where v is a constant	x=y where y is a variable, y occurs negatively
$\neg F$ if x occurs negatively in F	$\neg F$ if x occurs positively in F
$F_1 \wedge F_2$ if x occurs positively in F_1 or positively in F_2	$F_1 \wedge F_2$ if x occurs negatively in F_1 and negatively in F_2
$F_1 \vee F_2$ if x occurs positively in F_1 and positively in F_2	$F_1 \vee F_2$ if x occurs negatively in F_1 or negatively in F_2
$F_1 \rightarrow F_2$ if x occurs negatively in F_1 and positively in F_2	$F_1 \rightarrow F_2$ if x is positive in F_1 and negative in F_2
$\exists y{:}A$ (F) if $x \neq y$ and x occurs pos. in F	$\exists y{:}A$ (F) if $x \neq y$ and x occurs neg. in F

3. Consider the following query language:

   Positive relational algebra consisting of: $\cap$, $\cup$, $\bowtie$, σ, Π
   (Note: there is NO set difference in positive relational algebra.)

   Prove that positive relational algebra is contained in allowed (tuple or domain, your choice) relational calculus. In particular:

   a. (5 points) state and prove the base case

   b. (5 points) state the inductive hypothesis

   c. (16 points) prove each case of the induction step

4. (20 points) In the Ramakrishnan/Gehrke book, 3rd Edition (the 386/586 book), for problem 24.1, write queries 1, 3, 4, 7 in Datalog (with recursion and negation, as needed) and in SQL:1999 syntax.