CMPSCI 645: Homework 2

Due: Monday, April 6 2015, 11:59pm

Name: ______________________________

Question	Points	Score
1	30	
2	35	
3	35	
Total:	100	

Please turn in this homework electronically, as a PDF, through Moodle. You may handwrite your solutions, and then scan the document, or type directly into the PDF form.

1. **Normalization** [30 points]

   Consider the following relational schema and set of functional dependencies.

   $$R(A,B,C,D,E) \qquad \begin{array}{l} CD \rightarrow E \\ A \rightarrow B \end{array}$$

   (a) [10 points] List **all** superkey(s) for this relation. Justify your answer in terms of functional dependencies and closures.

   (b) [5 points] Which of these superkeys form a key (i.e., a minimal superkey) for this relation?

   (c) [15 points] Decompose R into BCNF. Show your work for partial credit. Your answer should consist of a list of table names and attributes and an indication of the keys in each table (either underline the corresponding attributes, or explicitly state the keys).

2. **Datalog** [35 points]

   (a) [15 points] Consider the following two datalog programs computing the transitive closure:

```
P1:
    T(x,y) :- R(x,y)
    T(x,y) :- T(x,z),R(z,y)

P2:
    T(x,y) :- R(x,y)
    T(x,y) :- T(x,z),T(z,y)
```

   Suppose R is a graph that consists of a single path: $R(a_0, a_1), R(a_1, a_2), \ldots, R(a_{n-1}, a_n)$. Thus, the transitive closure T computed by both programs consists of all $\binom{n}{2}$ ground facts of the form $T(a_i, a_j)$, for $1 \leq i < j \leq n$. Assume that we evaluate both programs using the semi-naive evaluation algorithm.

   i. For a fixed $m = 1, \ldots n-1$, how many times will the fact $T(a_1, a_{m+1})$ be discovered by `P1`?

   i. ____________

   Explain your answer:

   ii. How many times will the fact $T(a_1, a_{m+1})$ be discovered by `P2`?

   ii. ____________

   Explain your answer:

(b) [20 points] Consider a graph where each node x is either a leaf, or has two outgoing edges $(x, y), (x, z)$. In the former case, we store x in a relation $L(x)$; in the latter case we store the triple (x, y, z) in a relation T. (Thus, x is a key in $T(x, y, z)$.) Consider the following game with two players. Players take turns in moving a pebble on the graph. If the pebble is on a node x, then the player whose turn it is may move it to one of the two children, y or z. A player wins if it is her turn to move and the pebble is on a leaf. Write a datalog program that computes the set of starting nodes from which player 1 has a winning strategy. That is, your program should compute a relation $P1(x)$ that returns all nodes x such that, if player 1 starts the game on x (and plays smartly!) then she is guaranteed to win the game.

3. **Conjunctive Queries** [35 points]

   (a) [5 points] Find a full semi-join reduction for the query below.

$$q(x) :- R(x, y), S(y, z), T(y, u)$$

(b) i. [20 points] Indicate for each pair of queries q, q' below, whether $q \subseteq q'$. If the answer is yes, provide a proof; if the answer is no, give a database instance I on which $q(I) \not\subseteq q'(I)$.

α)

$$\begin{aligned} q(x) &:- \quad R(x,y), R(y,z), R(z,x) \\ q'(x) &:- \quad R(x,y), R(y,z), R(z,u), R(u,v), R(v,z) \end{aligned}$$

α) ______________

Explain your answer:

β)

$$\begin{aligned} q(x,y) &:- \quad R(x,u,u), R(u,v,w), R(w,w,y) \\ q'(x,y) &:- \quad R(x,u,v), R(v,v,v), R(v,w,y) \end{aligned}$$

β) ______________

Explain your answer:

ii. [10 points] Consider the two conjunctive queries below, and notice that $q_1 \subset q_2$.

$$q_1(x) = R(x, y), R(y, z)$$
$$q_2(x) = R(x, y)$$

1. Find a conjunctive query $r(x)$ s.t. $q_1 \subset r \subset q_2$