CSE 636: Test #2 (due May 7, 2013)

Submit all the answers by `submit_cse636` as a **single pdf file**. There will not be any deadline extensions.

This is **individual work**. Duplicate solutions will be considered violations of academic integrity. Please write your name and the text **"The submitted solutions are my individual work."** at the beginning of the submitted file.

You need to complete **four problems out of five**. If you complete 5 problems, the fifth will count as **extra credit.**

Problem 1 (25 pts)

You are to implement an airline database using RDF/RDFS. For each airport, store its unique name and unique abbreviation (e.g., "BUF"). For each connection, store the airlines that provide it and the connecting airports.

To do:

- Define an RDFS schema for this database. Can all the relevant integrity constraints be captured?
- Show an example RDF instance (at least 5 tuples),
- Write the following queries in SPARQL 1.1:
  - *Q1.1: Find all the airlines connecting to Buffalo.*
  - *Q1.2: For each airport, find the number of connecting airlines.*
  - *Q1.3: Find the cities connected to Buffalo (directly or indirectly) through Delta.*

Problem 2 (25 pts)

Consider a data exchange scenario, given by the following source-to-target dependencies:

$$\forall y.\ U(y) \Rightarrow \exists z.\ T(y,z)$$

$$\forall x,y.\ W(x,y) \Rightarrow T(y,x).$$

and the target dependency:

$$\forall x,y,z.\ T(x,y) \wedge T(x,z) \Rightarrow y = z.$$

You are given the following source instance

$$r_2 = \{W(b,a), W(a,c), U(a), U(d)\}.$$

To do:

1. Construct a universal solution for r_2, explaining the construction.
2. Compute the certain answers to the queries $Q_{3.1} \equiv T(x,y)$ and $Q_{3.2} \equiv \exists y.\ T(x,y)$, given the source instance r_2.
3. Show a source instance r_0 for which no universal solution exists, assuming the above dependencies.

Problem 3 (25 pts)

Consider two relations R and S with the templates r^{bu} and $s^{o[\{1,2\}]f}$ respectively,

$$q(X, Y) :- r(X, Y).$$

$$q(X, Y) :- s(X, Y).$$

To do:

1. Show a template under which the query is feasible *without postprocessing.*
2. Show a template under which the query is feasible *with postprocessing* but is not feasible without postprocessing.

In each case explain your answer and show how the adornments are computed. Try to obtain the most general template possible.

Problem 4 (25 pts)

You are given two relations $P(A, B)$ and $Q(A, B)$, and the following integrity constraints:

1. A is a key of P;
2. P and Q are disjoint.

To do:

1. Write down first-order logic formulas expressing the constraints.
2. Rewrite the query $Q_{4.1} \equiv$ `SELECT * FROM P`. using the residue approach. The result should be a SQL query.
3. Given an instance $r_4 = \{P(a, b), P(a, c), Q(a, c)\}$, compute all the repairs of r_4.
4. Compute the consistent answers to $Q_{4.1}$ in r_4 with respect to the given integrity constraints.

Problem 5 (25 pts)

You will study a different form of provenance in the semiring model. The CL-provenance of a tuple will be its *confidentiality level* – an integer c such that $1 \leq c \leq 10$.
To do:

1. What is the appropriate $\mathcal{K}$-semiring for A-provenance? Assume that the confidentiality level in a tuple in the join result is the highest among the input tuples; for the union result, the lowest.
2. Given
   - the database schema consisting of relations $R(A, B)$ and $S(B, C)$,
   - the annotated database instance (the tuples are annotated with their confidentiality levels):
     $$r_5 = \{R(a, b)\{7\}, R(d, b)\{3\}, S(b, c)\{9\}, S(b, d)\{2\}, S(c, d)\{5\}\},$$
   - the query $Q_5 \equiv \sigma_{C=d}(\pi_B(R(A, B)) \bowtie S(B, C))$.

   show the tuples in the result of Q_5 together with their CL-provenance. Show also all the steps of the derivation.