Data Integration: Test #1 (due April 23, 2007)

Submit your solutions by email as a single PDF file. This is individual work. Duplicate solutions may receive reduced or no credit.

Problem 1 (20 pts)
Assume an undirected graph is represented as a set of facts of the form node($x$) for a node $x$, and edge($x,y$) and edge($y,x$) for an edge{$x,y$}. A graph is connected if for every two different nodes in the graph there is a path between them. A node $x$ is an articulation point of a graph if: (1) the graph is connected, and (2) the graph with $x$ and its incident edges removed is no longer connected.

1. Write a Datalog ¬ program $P_1$ that checks whether a given graph is connected.
2. Write a Datalog ¬ program $P_2$ that returns the set of articulation points of a given graph.
3. Explain why the program $P_2$ is stratified.
4. Can the program $P_2$ be written in Datalog without using negation?

Problem 2 (20 pts)
Assume you are given two different databases, A and B, representing the same information about company sales broken by product and year. The first database has a separate relation for every product and each of those relations has two attributes Year and Amount. The second database has one relation Sales with an attribute Product and a separate attribute for every year that contains the year’s sales amount.
Define in FISQL the mappings between the databases A and B (in both directions). Do not assume any fixed set of products or years.

Problem 3 (20 pts)
The source database has 3 relations:
- Faculty(EmpSSN,EmpName,Rank);
- Staff(EmpSSN,EmpName,Level);
- Dependent(EmpSSN,DepSSN,DepName,Status) where Status is equal to 1 for the spouse and 2 for a child.

The target database has 2 relations:
- EmpSpouse(EmpSSN,EmpName,SpouseName);
- EmpChildren(EmpSSN,ChildName,Age),

and 2 integrity constraints:
- EmpSSN → EmpName SpouseName;
• EmpChildren(EmpSSN) is a foreign key referencing EmpSpouse(SSN).

1. Define the appropriate source-to-target dependencies and write down the target constraints.

2. You are given a source instance consisting of the following facts: Faculty(123, mark, full), Staff(456, frank, 11), Dependent(123, 999, julie, 1), Dependent(333, 321, bill, 2). Compute a corresponding universal target instance. What are the certain answers for the following queries: (A) EmpSpouse(x, y, z), (B) ∃y, z. EmpSpouse(x, y, z)?

3. Show a source instance for which there is no corresponding universal target instance.

Problem 4 (20 pts)

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

1. A is a key of Q;
2. P is a subset of Q.

Write down first-order logic formulas expressing the constraints.

Rewrite the following queries using the residue approach:

• Q1: SELECT * FROM P
• Q2: SELECT A FROM Q.

Do the rewritten queries compute exactly the consistent answers to the given queries? Explain your answer.

Problem 5 (20 pts)

For each basic operation τ of the relational algebra, check whether the following property holds:

Given any database D, each consistent query answer to τ in D w.r.t. a set of FDs F is also an answer to τ in D.