

CSE 636: Test #1 (due March 5, 2013)

Submit all the answers by `submit_cse636` as a **single pdf file**.

Problem 1 (30 pts)

Assume a directed graph is represented as a set of facts of the form $node(x)$ for a node x , and $edge(x,y)$ for an edge (x,y) . A node y in a directed graph is *reachable* from a node x if there is a path from x to y in the graph.

1. Write a Datalog program P_1 that returns pairs (x,y) such that both x and y are reachable from some node z .
2. Write a Stratified Datalog \neg program P_2 that returns **true** if there is a node unreachable from any other node in the graph.
3. Run the program P_2 using `xsb` on a positive example (returns **true**) and a negative example (returns **no**), with at least 4 nodes each. Document the runs.
4. Explain why the program P_2 is stratified.
5. Does there exist a Datalog program without negation which is equivalent to P_2 ? Explain.

Problem 2 (20 pts)

We call a DTD *conforming* if there is a document that conforms to it.

1. Show an example DTD which is not conforming.
2. For any DTD without attributes, show how construct a Datalog program that returns **true** if and only if the DTD is conforming.
3. Using your approach, determine if the following DTD is conforming:

```
<!DOCTYPE p [  
  <!ELEMENT p (a,b)>  
  <!ELEMENT a (e | c)>  
  <!ELEMENT c (d?,p*)>  
  <!ELEMENT b (#PCDATA)>  
>
```

Problem 3 (30 pts)

You are given two databases containing information about movies.

The database A consists of `Movie` elements, with subelements `Title` (single, required), `Director` (single, optional), and `Actor` (zero or more).

The database B consists of `Actor` elements, with an attribute `Name`(required) and subelements `Movie` that have attributes `Title` (required) and `Director` (optional).

1. Define the schemas of A and B using XML Schema.
2. Define in XQuery the mapping M between B and A, assuming that a movie is identified by its title.
3. Assume that A already contains some local data conforming to the above schema. Modify the above mapping to resolve conflicts in A between the local data and the data obtained from B through the mapping:

- If the director for a movie is different in A and B or missing, do not return it.
 - The set of actors for a movie should be the union of the corresponding sets from A and B (duplicates eliminated).
4. Run the last two queries on example databases A and B (at least 5 movies and 10 actors in each) using an XQuery system, and report the results. Document the runs.

Problem 4 (20 pts)

Definitions:

- A knowledge base KB (TBox + ABox) is *satisfiable* if there is an interpretation that satisfies it.
- A knowledge KB *implies* an assertion A if every interpretation that satisfies KB also satisfies A.

You have the following ontology:

There are two kinds of entities: movies and persons, mutually disjoint. Each movie has a (single) director (a person) and stars zero or more actors (also persons). Each director directs one or more movies. Each actor stars in one or more movies.

1. Express this ontology as a satisfiable description logic knowledge-base KB. Explain why it is satisfiable.
2. Which axioms of KB can be expressed in (a) RDF/RDFS, (b) DL-Lite?
3. Define the following notions in description logics:
 - *actor, director*;
 - *actor that starred in at least two movies directed by Hitchcock*;
 - *movie in which Hitchcock was both actor and director*.
4. Express KB using first-order logic.
5. Add axioms and assertions to KB to make the resulting knowledge-base unsatisfiable.

Extra credit: Consider the knowledge base consisting of the following Description Logic ABox:

Inherits(eve, mark).
Inherits(mark, jean).
Inherits(eve, jean).
Inherits(jean, marie).
French(mark).
 \neg *French(marie).*

Does this knowledge base imply the following fact:

$eve \in (\exists Inherits.(French \sqcap \exists Inherits.\neg French))?$

Explain your answers in detail using the formal semantics of description logics.