CS345 Midterm Examination

Monday, October 28, 1996, 3:15 – 4:30 pm

Directions

The exam is *open book*; any written materials may be used. Answer all 8 questions on the exam paper itself. The total number of points is 100. You may make additional comments at the bottoms of pages or at the end if you feel the need to clarify or justify your answers. Do not forget to **sign the pledge** below.

I acknowledge and accept the honor code.

____

Print your name here: ____

1	2	3
4	5	6
7	8	

Problem 1: (10 points) Show how to evaluate the rules

```
p(X,Y) :- a(X,Y)
p(X,Y) :- a(X,Z) & a(W,Y) & p(X,W) & p(Z,Y) & p(Z,W)
```

using seminaive evaluation.

a) What are the basis steps that initialize P, the relation for IDB predicate p, and ΔP, the "new" facts of P?

b) Give the inductive rules that compute the new values of P and ΔP at each round.

Problem 2: (10 points) Suppose we are given EDB relations/predicates $m(X,Y)$ meaning that Y is the mother of X and $f(X,Y)$ meaning that Y is the father of X. Write a Datalog program to compute $f\_a(X)$, all the **female** ancestors of X. Note that a female ancestor need not have only females in the line of descent between her and X.

Problem 3: (10 points) Consider the following rules

```
p(X,Y) :- a(X,Y) & NOT b(X)
q(X,Y) :- a(X,Y) & NOT p(Y,X)
r(X,Y) :- q(X,Y) & NOT b(Y)
```

and the EDB consisting of the three facts $a(1,2)$, $a(2,3)$, and $b(2)$.

a) In the space below, draw the dependency graph for the predicates. Indicate the stratum of each predicate.

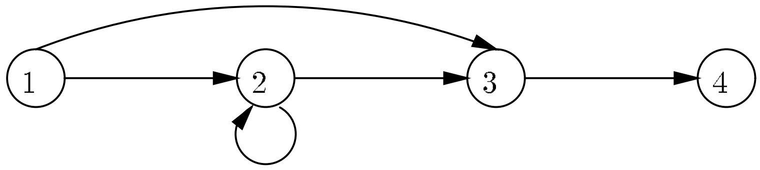
b) Compute the stratified model for these rules and EDB.

c) Give another minimal model for the rules.

Problem 4: (15 points) Consider the "win" rule:

```
win(X) :- move(X,Y) & NOT win(Y)
```

with the *move* relation $move(1,2)$, $move(1,3)$, $move(2,2)$, $move(2,3)$, and $move(3,4)$. The figure below suggests *move*.



a) In the space below, use the alternating fixedpoint method to find the well-founded model.

b) What is the well-founded model?

c) If we apply the Gelfond-Lifschitz transform for these rules and EDB to the model $M = \{win(1), win(3)\}$, what is GL(M)?

Problem 5: (20 points) Consider the following rules, in which a is an EDB "arc" relation describing a directed graph, and p is its "path" IDB relation, the transitive closure of a.

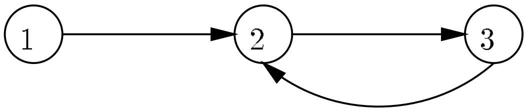
```
(1)    p(X,Y) :- a(X,Y)
(2)    p(X,Y) :- a(X,Z) & p(Z,Y)

(3)    q(X,Y) :- p(X,Y) & NOT q(Y,X)
```

a) In the space below, draw the dependency diagram for the predicates.

b) Are these rules stratified? Why or why not?

---

c) Suppose we wish to consider whether these rules are modularly stratified (with some particular EDB relation A). It is sufficient to put a and p together in one module (the "lower module") and q by itself in another (the "upper module"). The lower module can never present a problem, but the upper module might or might not violate the modularly-stratified condition. If the graph consists of the arcs $a(1,2)$, $a(2,3)$, and $a(3,2)$, representing the graph below,



what are the instantiations of rule (3) for this EDB? (Remember you must compute p first.)

---

d) Draw below the dependency graph for the ground atoms of q.

e) Are the rules modularly stratified for this EDB? Why or why not?

f) Under what circumstances will the graph represented by EDB relation A make the rules modularly stratified? Explain briefly.

Problem 6: (10 points) Below is a Datalog program P:

```
p(X) :- b(X)
p(X) :- a(X,Y) & b(Y)
```

Let Q be the conjunctive query

```
p(X) :- b(X) & a(X,Y)
```

a) To test if $Q \subseteq P$, we need to use the canonical database D constructed from Q. What is D?

b) What is $P(D)$?

c) Is $Q \subseteq P$? Explain.

Problem 7: (10 points) For each of the two conjunctive queries Q below, briefly describe a **polynomial-time** algorithm that takes an arbitrary conjunctive query R and tests whether $Q \subseteq R$. That is, test whether there is a containment mapping from the input query R to the specific query Q, in time that is polynomial in the size of R. You may assume that R is "plausible," i.e., its head is $p(A)$ for some variable A, and each subgoal has predicate a and two arguments. In each case the test is *very* simple. You do not need more than a few lines to describe it, and simplicity counts.

a) Q is `p(X) :- a(X,Y).`

b) Q is `p(X) :- a(X,X).`

Problem 8: (15 points) For each $n > 1$, consider the pair of conjunctive queries:

C_n: `s(X) :- p(X,X`$_1$`) & q(X`$_1$`,X`$_2$`) & p(X,X`$_2$`) & q(X`$_2$`,X`$_3$`)`
`& ... & q(X`$_{n-2}$`,X`$_{n-1}$`) & p(X,X`$_{n-1}$`) & q(X`$_{n-1}$`,X`$_n$`)`
`& p(X,X`$_n$`) & q(X`$_n$`,X`$_1$`)`

D_n: `s(X) :- p(X,Y) & p(X,Z) & q(Y,Z) & q(Z,Y)`

a) For what values of $n > 1$ is $D_n \subseteq C_n$?

---

b) For those n for which containment holds, describe a containment mapping.

How many different containment mappings are there for these n? ______________.

c) For the smallest $n > 1$ for which containment does not hold, give a counterexample database to show $D_n \subseteq C_n$ is false.
