**CS345 Midterm Examination**
Wednesday May 5, 1999, 12:50–2:05 PM

## Directions

The exam is *open book*; any written materials may be used. Answer all 4 questions on the exam paper itself. The total number of points is 100. Do not forget to **sign the pledge** below.

I acknowledge and accept the honor code.

______________________________

Print your name here: ________________________________________

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |

**Problem 1:** (48 points) In this problem, you will be given some Datalog programs, plus EDB, and some propositional logic programs. For each, you must classify what is given by the simplest semantics that are defined, and give the model according to that semantics (or in case F, below, two alternative models). *Note*: in propositional logic, every "locally" stratified model is stratified, since there is no EDB, and no need to instantiate predicates. The classifications are:

A) There is a unique least fixpoint obtainable by seminaive evaluation. In this case, give the least fixpoint.

B) There is no unique least fixpoint, but there is a stratified semantics. In this case, give the stratified semantics.

C) There is no stratified semantics, but there is a locally stratified semantics. In this case, give the locally stratified semantics.

D) There is no locally stratified semantics, but there is a modularly stratified semantics. In this case, give the modularly stratified semantics.

E) There is no modularly stratified semantics, but there is a two-valued, well-founded semantics. In this case, give the well-founded semantics.

F) There is not a two-valued, well-founded semantics. In this case, give *both* the (three-valued) well-founded semantics, and the unique stable model if there is one.

a) Datalog program:

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

EDB: $\{b(1),\ b(2), a(1,2,3),\ a(3,2,1)\}$.

Classification (A–F): ________

Model(s):

b) Datalog program:

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

EDB: $\{b(1),\ a(1,2,3),\ a(2,1,3),\ a(2,3,3)\}$.

Classification (A–F): ________

Model(s):

c) Datalog program:

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

EDB: $\{b(1),\ a(1,2,3),\ a(2,3,1),\ a(3,1,3)\}$.

Classification (A–F): ________

Model(s):

d) Datalog program:

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

EDB: $\{b(1),\ b(2),\ a(1,2),\ a(2,3),\ a(3,4)\}$.

Classification (A–F): ________

Model(s):

e) Program:

```
p :- NOT q
q :- NOT r
r :- q & NOT p
```

Classification (A–F): ________

Model(s):

f) Program:

```
p :- q
q :- p
r :- NOT q
s :- r & NOT p
```

Classification (A–F): ________

Model(s):

g) Program:

```
p :- NOT q & NOT r
p :- NOT p
q :- NOT p & NOT r
r :- NOT p & NOT q
```

Classification (A–F): ________

Model(s):

---

---

---

h) Program:

```
p :- q & r
q :-
p :- r & s
r :- s & t
r :- t & q
t :- q
```

Classification (A–F): ________

Model(s):

---

---

---

**Problem 2:** (15 points) Consider the conjunctive queries:

$Q_1$: `p(X,Y) :- a(X,A) & a(A,A) & a(A,Y) & a(X,B) & a(B,C) & a(C,Y)`

$Q_2$: `p(X,Y) :- a(X,U) & a(U,Y) & a(X,V) & a(V,V) & a(V,W) & a(W,Y)`

a) Is $Q_1 \subseteq Q_2$? Give either a containment mapping or a counterexample database.

b) Is $Q_2 \subseteq Q_1$? Give either a containment mapping or a counterexample database.

c) Although we did not prove it in class, it is not hard to show that if $C_1$ is the "smallest" conjunctive query equivalent to $C_2$ (i.e., $C_1$ has as few subgoals as any CQ equivalent to $C_2$) then after renaming variables and reordering subgoals, $C_1$ is identical to $C_2$ with some subgoals of $C_2$ removed. Thus, we can search for a smallest equivalent to a given CQ by considering whether it is equivalent to what results by removing one or more subgoals. Find the smallest equivalent to $Q_1$ above.

**Problem 3:** (28 points) This question concerns the following two conjunctive queries with arithmetic comparisons:

$P_1$: `panic :- a(X,Z) & a(Z,Y) & a(Y,X) & X<Z & Z<Y`
$P_2$: `panic :- a(A,B) & A<B`

a) Suppose we want to test $P_2 \subseteq P_1$, using Klug's method. If we use as constants for our canonical databases the integers $0, 1, \ldots$ (up to as many as we need), what are all the canonical databases for $P_2$ that we need to consider? (*Hint*: remember that these are the "basic canonical databases," with order of blocks in the partition specified).

b) Which, if any, of the basic canonical databases produce the head (`panic`) of $P_2$?

c) Do those canonical databases that produce the head of $P_2$, also produce the head of $P_1$ when that CQ is applied to the canonical database? Either demonstrate that they all do, or give a counterexample.

d) Now, suppose we want to apply the G/Z/O test to see whether $P_1 \subseteq P_2$. First, rewrite $P_1$ and $P_2$ so they are "rectified"; i.e., no variable appears twice among the relational subgoals of one CQ.

e) What are all the containment mappings from (the rectified) $P_2$ to (the rectified) $P_1$? You may describe a containment mapping by statements of the form $S \to T$, meaning that variable $S$ of $P_2$ is mapped to variable $T$ of $P_1$.

f) What is the condition involving inequalities of numbers (and logical connectives such as "OR") that you must show to verify that $P_1 \subseteq P_2$?

g) Do you believe that your statement from (f) is true or false? Explain briefly.

**Problem 4:** (9 points) An information-integration system similar to "Information Manifold" has a global predicate $par(X, Y)$, meaning that $Y$ is a parent of $X$. There are two views:

$v_1$`(X,Y) :- par(X,Z) & par(Z,Y)`
$v_2$`(X,Y) :- par(X,U) & par(U,V) & par(V,Y)`

providing some grandparent and great-grandparent facts, respectively. The query $Q$ asks for sixth-generation ancestry:

```
Q: ans(X,Y) :- par(X,A) & par(A,B) & par(B,C) &
               par(C,D) & par(D,E) & par(E,Y)
```

List all the solutions (in terms of the views) whose expansions are contained in the query, and that are minimal (i.e., if you delete one or more subgoals, then the solution is no longer contained in the query).