CS145 Midterm Examination

Tuesday, November 6, 2001, 9:30–10:45 AM

Directions

The exam is *open book/notes*; any written materials may be used.

For each of the 15 questions, circle the letter (a), (b), (c), or (d) of your chosen answer. Do not circle more than one answer. If you wish to change your answer, please indicate clearly what your "final answer" is.

Score = 55 plus 3 times number-right minus number-wrong, so random guessing nets you nothing on the average, and 100 is a perfect score.

If you wish to explain or demonstrate your solution to a problem for partial credit, you may use page bottoms or the backs of the pages (but warn us on the front). Please use this option sparingly, e.g., if you think the question is flawed or open to multiple interpretations, because we shall only be awarding partial credit in rare situations.

You have 5 minutes per question. Use your time wisely, and do not spend too much time on any one question.

Do not forget to **sign the pledge** below.

I acknowledge and accept the honor code.

____

Print your name here: ____

In each of the first 6 questions, you are asked to compare two queries Q_1 and Q_2. You must tell whether the queries are:

1. The same [choice (a)], meaning that for every database the answers to the two queries are the same. That is, the same tuples are produced by each query, and a tuple is produced the same number of times by each query. The order in which tuples are produced is not to be considered.
2. Completely different [choice (d)], meaning that there are databases where Q_1 produces more of some particular tuple, and other databases where Q_2 produces more of some particular tuple. Note that the query producing the smaller number of copies of a tuple may produce zero copies of that tuple.
3. One is contained in the other but they are not the same [choice (b) or (c)]. For instance, Q_1 is contained in Q_2 if on every database, Q_2 produces at least as many copies of each tuple as Q_1 does. Note that it is possible Q_2 produces one or more copies of a tuple, while Q_1 produces none of that tuple.

General advice:

- Do not assume a query has a trivial syntactic error and therefore produces nothing.
- Relations mentioned in the queries may have attributes not mentioned, but their existence should not affect the answer.
- Relations may have `NULL`'s.
- SQL queries should be assumed to be in standard SQL unless stated otherwise.
- In SQL, it is possible that there may be duplicate tuples, but in relational algebra assume the relations are sets.
- Although we did not cover it explicitly in class, SQL duplicate elimination and grouping treat `NULL` like any other value.

In the following three problems, assume the schemas are $R(A, B)$ and $S(B, C)$:

Question 1:

Q_1: $R \bowtie S$

Q_2: $R \underset{R.B=S.B}{\bowtie} S$

(a) Q_1 and Q_2 produce the same answer.
(b) The answer to Q_1 is always contained in the answer to Q_2.
(c) The answer to Q_2 is always contained in the answer to Q_1.
(d) Q_1 and Q_2 produce different answers.

Question 2:

Q_1: $\sigma_{A=1}(R \bowtie S)$
Q_2: $(\sigma_{A=1}(R)) \bowtie S$

(a) Q_1 and Q_2 produce the same answer.
(b) The answer to Q_1 is always contained in the answer to Q_2.
(c) The answer to Q_2 is always contained in the answer to Q_1.
(d) Q_1 and Q_2 produce different answers.

Question 3:

Q_1: `SELECT R.A FROM R, S WHERE R.B = S.B;`
Q_2:
```
SELECT R.A FROM R WHERE
    R.B IN (SELECT B FROM S);
```

(a) Q_1 and Q_2 produce the same answer.
(b) The answer to Q_1 is always contained in the answer to Q_2.
(c) The answer to Q_2 is always contained in the answer to Q_1.
(d) Q_1 and Q_2 produce different answers.

Question 4: In this question, the schema is $R(A, B, C)$.

Q_1: `SELECT DISTINCT * FROM R;`
Q_2: `SELECT * FROM R GROUP BY A, B, C;`

(a) Q_1 and Q_2 produce the same answer.
(b) The answer to Q_1 is always contained in the answer to Q_2.
(c) The answer to Q_2 is always contained in the answer to Q_1.
(d) Q_1 and Q_2 produce different answers.

Question 5:

Q_1: `SELECT A, COUNT(DISTINCT B) FROM R WHERE C > 10 GROUP BY A;`
Q_2: `SELECT A, COUNT(B) FROM R WHERE C > 10 GROUP BY A;`

(a) Q_1 and Q_2 produce the same answer.
(b) The answer to Q_1 is always contained in the answer to Q_2.
(c) The answer to Q_2 is always contained in the answer to Q_1.
(d) Q_1 and Q_2 produce different answers.

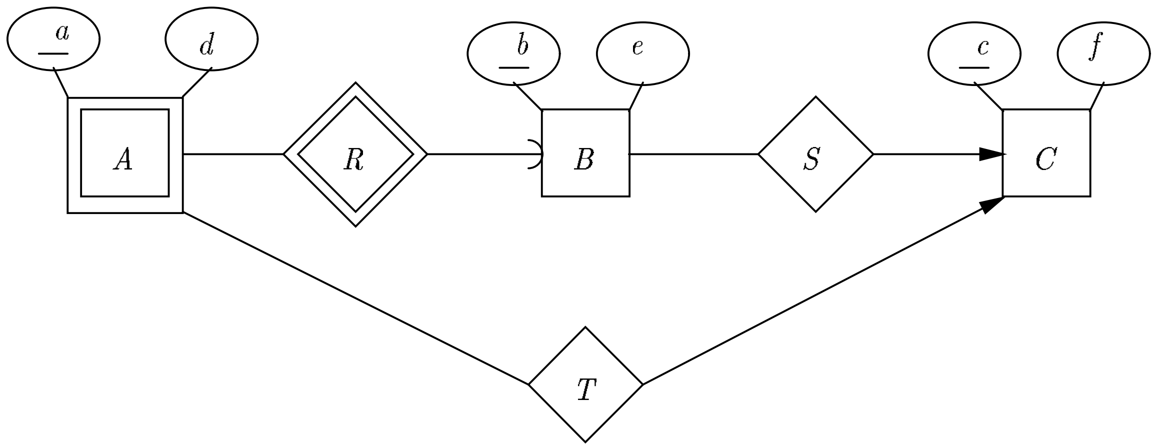
Question 6:

Q_1: `SELECT DISTINCT * FROM R;`
Q_2: `(SELECT * FROM R) INTERSECT (SELECT * FROM R);`

(a) Q_1 and Q_2 produce the same answer.
(b) The answer to Q_1 is always contained in the answer to Q_2.
(c) The answer to Q_2 is always contained in the answer to Q_1.
(d) Q_1 and Q_2 produce different answers.

In the remainder of the exam, the answers have no particular form.

The following two questions are based on the E/R diagram below:



Question 7: Which of the relationships is necessarily redundant?
(a) R (b) S (c) T (d) None are redundant

Question 8: If we use the strategy outlined in the text and class for converting this E/R diagram to a database schema, which of the following relations would we create?
(a) $T(a, c)$ (b) $B(b, e)$ (c) $R(a, b)$ (d) $R(a, b1, b2)$

Question 9: Relation $R(A, B, C)$ satisfies the multivalued dependency $A \twoheadrightarrow B$, and has (possibly among others) the following tuples in its current instance: $(0, 1, 2)$, $(0, 3, 4)$, and $(1, 5, 6)$. Which of the following tuples is *not necessarily* in the current instance of R?
(a) $(0, 1, 4)$ (b) $(0, 3, 4)$ (c) $(0, 5, 2)$ (d) None of the above.

The following three questions refer to a relation $R(A, B, C, D, E)$ with functional dependencies $A \to B$, $BC \to D$, and $E \to C$.

Question 10: The number of keys of R is:
(a) 1 (b) 2 (c) 8 (d) 11

Question 11: Which of the following functional dependencies does *not necessarily* hold in R?
(a) $AC \to D$ (b) $AE \to C$ (c) $BC \to B$ (d) $CE \to D$

Question 12: If we project R onto $S(B, C, D, E)$, which of he following functional dependencies holds in S and *also* does not violate the BCNF condition for S?
(a) $BC \to D$ (b) $BE \to D$ (c) $B \to E$ (d) $E \to C$

Question 13: The following is *PSM code* (not PL/SQL). It's intent is to insert into relation S (which has a single attribute) all those integers in the range 1 through n that are not already in S. However, several statements have syntax errors in them.

```
(1)  CREATE PROCEDURE Foo(n INTEGER)
(2)  DECLARE x INTEGER;
(3)  BEGIN
(4)      x := n;
(5)      WHILE x >= 1 DO
(6)          IF x NOT IN (SELECT * FROM S) THEN
(7)              INSERT INTO S VALUES(x);
(8)          END IF;
(9)          x := x - 1;
(10)     END LOOP;
(11) END;
```

Which of the following lines *does not* contain a significant syntax error?
(a) The procedure header of line (1)
(b) The while-statement of line (5)
(c) The assignment statement of line (9)
(d) The loop-ender of line (10)

The following two questions are based on the declarations:

```
CREATE TABLE T (
    a INT PRIMARY KEY,
    b INT
);

CREATE TABLE S (
    c INT PRIMARY KEY,
    d INT REFERENCES T(a) CHECK(d < ANY
        (SELECT b FROM T))
);
```

The current relation T is:

a	b
1	0
3	6
5	3

and the current S is:

c	d
1	1
2	1
3	5

Question 14: Which of the following modifications will *not* be rejected by the constraint checker?

(a) `INSERT INTO S VALUES(1,5);`
(b) `DELETE FROM T WHERE b = 3;`
(c) `DELETE FROM T WHERE b = 6;`
(d) None of the above; i.e., they will all be rejected.

Question 15: *Note*: *This question is very hard. Do not spend too much time on it unless you have nothing else to do.* In how many orders can the six tuples be deleted from their respective relations without causing any constraint violations (regardless of whether the constraint checker will reject the deletions)?

(a) 60 (b) 64 (c) 72 (d) 80
