CS 317: DBIS: Final, Autumn 2002

Duration: 180 minutes | Total Marks: 100

Roll no: ______________________ **Name:** ______________________

PART A: Answer this part in the question paper itself. Use the back of the sheet if you need more space.

1. Short answers:

   (a) What is the difference between disjoint and overlapping generalization? ...2

   (b) Give a meaningful example of the need for role indicators in ER diagrams. (Don't give a diagram, just an example in words.) ...2

   (c) Draw the truth table for the AND operation on three-valued logic. ...3

   (d) Give an example of a simple selection view (in relational algebra) and an update on it that cannot be executed. ...3

   (e) Given a table r with a foreign key referencing s, what is the effect of adding the clause **on delete cascade** to the foreign key declaration? ...2

(f) Give an example (in English) of a commonly required access restriction that cannot be specified in SQL. ...3

(g) What is meant by recursion in a DTD. Explain with an example. ...3

(h) Given a set of dependencies $A \rightarrow B$, $B \rightarrow C$ and $C \rightarrow D$, and a relation $r(A, B, C, D)$, what is the best order for the BCNF decomposition to consider the dependencies, and why? ...3

(i) Give an example of a schedule that is serializable but not recoverable. ...3

(j) What information do undo log records contain, and what are they used for? ...3

(k) Why is persistence by reachability appropriate for persistent Java? ...3

(l) When inferring MVDs on a decomposition of R into R_1 and R_2, what do we need to do after computing $D+$, the closure of the given set of MVDs? ...3

2. For each of the following pairs of queries, indicate under what conditions are they equivalent (i.e. they return exactly the same answers),

(a) `select distinct r.A from r` and
`select distinct r.A from r, s where r.B = s.B` ...2

(b) `select distinct r.A from r` and
`select distinct r.A from r, s` ...3

3. Given the relation `person(name, sex, father, mother)`, with name as the primary key, and sex taking on values 'M' and 'F', write views in Datalog to find

(a) $brother(x, y)$: all x, y such that y is a brother of x (x may be female). ...3

(b) $anc(x, y, h)$: all x, y, h such that y is an ancestor of x at level h. E.g. parents are at level 1, grandparents at level 2, and so on. ...4

PART B: Answer this in your answer book. Answers to all subparts must be together.

4. Draw an ER diagram for recording cricket players statistics. Specifically, we want to track: players, matches, matches that a player played in, his performance in each innings: batting (scores, balls, how out) and bowling (O,M,R,W). ...7

5. Given relations $r(A, B)$, and $s(A, C)$:

   (a) Give an expression in SQL that is equivalent to ${}_B\mathcal{G}_{sum(C)}(\Pi_{BC}(r \bowtie s))$. Don't forget that Π eliminates duplicates. ...4

   (b) Give a relational algebra expression equivalent to
   ```
   select * from r where exists
        (select * from s where s.A = r.A)
   ```
   ...4

   (c) Give a QBE expression equivalent to $\{< a > \mid \exists\, c\ (< a, c > \in\ s\ \wedge\ \exists\, b_1, b_2\ (< a, b_1 > \in\ r \wedge\ < c, b_2 >\ \in\ r\ \wedge\ b_1\ >\ b_2))\}$ ...4

   (d) Suggest how to use extensions to aggregation to create nested sets in SQL. Specifically, write an extended SQL query on relation r, to nest all B values corresponding to each A value into a set. ...4

6. Dependency theory

   (a) Show using the schema $r(A, B, C, D, E)$ with functional dependencies $A \rightarrow B$ and $BC \rightarrow D$, that using the original set of dependencies is not enough to test if a decomposed relation is in BCNF. ...5

   (b) Consider the relations
   `registration(rollno, course)` and `student(rollno, name, year)`,
   with the following kind of dependency (constraint): students in year 1 should register for exactly 2 courses, students in year 2 should register for exactly 1 course. (No other year is permissible.)

   Suggest how to decompose the two relations in such a way that the union (NOT join) of the decomposed relations gives the original relations back, and such that superkey, foreign key and domain constraints on the decomposed relations (without any other constraints) are sufficient to ensure that the constraint above is satisfied.

   Give the decomposition and the super/foreign key declarations in pseudocode (no need for SQL). (NOTE: the solution may not be great if the number of courses is large, that's OK.) ...7

7. Given an XML dataset with DTD as follows

```
<!DOCTYPE marks [
  <!ELEMENT coursemarks(coursename, studentmark*) >
  <!ELEMENT studentmark(name, mark)>
  ...
]>
```

with elements coursename, name and mark all being of type PCDATA.

Write queries in XQuery/XPath to do the following

(a) Find names and marks of all students in the DBIS course. ...3

(b) Generate the "dense rank" of each student of DBIS. Dense rank is defined as follows: all students with the top mark get a rank of 1, all students with the next highest mark get a rank of 2, and so on. ...7

8. OLAP, Mining and IR

(a) Write a query in basic SQL equivalent to the following query in SQL with OLAP extensions (which generates a "data cube")

```
select r.A, r.B, sum(r.C)
from r
group by cube(r.A, r.B)
```

...3

(b) Consider the schema $r(A, B, C)$, where C is the class to which the tuple belongs, and A and B are attributes that can be used for classification. Assume all attribute values are single letters, and there are only 2 classes T and F.

Give an example with 4 "training instances" (tuples) and a partitioning condition that improves the "purity" (the exact definition of purity doesn't matter). ...4

(c) Web keyword search systems depend critically on relative ranking of pages.

Earlier IR approaches were based on number of occurrences of each query term in a particular document (term frequency), with query words weighted by their rarity (inverse document frequency). Explain why these earlier approaches were susceptible to spamming, that is, sites crafting pages to get an artificially high rank. ...3

Total Marks = 100