UNIVERSITY OF EDINBURGH

FACULTY OF SCIENCE AND ENGINEERING

COMPUTER SCIENCE THREE

DATABASE SYSTEMS

Friday 6 September 2002

09:30 to 11:00

Examiners:

Prof. K. Turner (External)

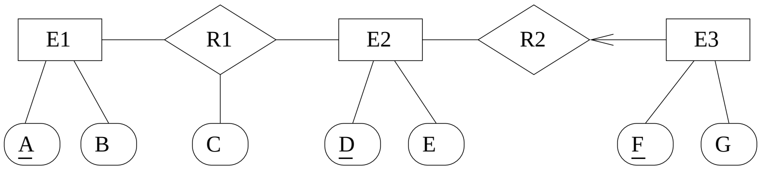
Prof. C. Stirling (Chairman)

INSTRUCTIONS TO CANDIDATES

Answer any TWO questions.

All questions carry equal weight.

1. (a) [15 points] The following entity-relationship diagram has attributes $ABCDEFG$.



i. Write down a minimal set of functional dependencies on the attributes given by this diagram. [4 *points*]

ii. Write down an appropriate SQL DDL (assume any type, e.g. `INTEGER` for all attributes) [7 *points*]

iii. How would you modify your answers to the previous two questions if E3 were a *weak* entity dependent on E2 through the connecting relationship R2? [4 *points*]

(b) Consider the two transactions

T_1: read (X_1); C_1; write (Y_1)

T_2: read (X_2); C_2; write (Y_2)

in which C_1 and C_2 are computations that do not involve a read or write. That is, each transaction reads from just one data item, does some computation and then writes to just one data item.

For each of the following scenarios say whether or not *all possible* schedules are seralizable. If there is a schedule that is not serializable, write it down and draw its conflict graph.

i. X_1, X_2, Y_1 and Y_2 are all different items. [2 *points*]

ii. X_1, X_2, Y_1 and Y_2 are all the same item. [2 *points*]

iii. The $X_1 = X_2 = X$, $Y_1 = Y_2 = Y$, but $X \neq Y$. [2 *points*]

iv. $X_1 = X_2 = Y_1$, but Y_2 is a distinct item. [2 *points*]

v. $X_1 = Y_2$ and $X_2 = Y_1$, but $X_1 \neq X_2$ [2 *points*]

2. (a) We are designing a database for scheduling classes and enrollments. We have the attributes Student, Class, Time-slot, Room, Lecturer, Marks, or $SCTRLM$

A Class determines a Time-slot, so we have the functional dependency $C \to T$

i. Similarly write down the functional dependency associated with each of the collowing conditions:

A. A Class determines a Room [1 *points*]

B. A Class determines a Lecturer [1 *points*]

C. A Room cannot have two different Classes in the same Time-slot [1 *points*]

D. A Student cannot attend two different Classes in the same Time-slot [1 *points*]

E. A Student cannot be in two different Rooms in the same Time-slot [1 *points*]

F. A Lecturer cannot teach two different Classes in the same Time-slot [1 *points*]

G. Marks are awarded for a Student in a Class [1 *points*]

ii. Write down a minimal cover for the functional dependencies you have listed above (include $C \to T$) [3 *points*]

iii. Write down what you think is a sensible decomposition of $SCTRLM$. [2 *points*]

iv. Is your decomposition in Boyce-Codd normal form? If not say why. [3 *points*]

(b) The *selectivity* $S(A, R)$ of an attribute A in a relation R is the ratio of the number of distinct values of A to the number of tuples in R. If R has n tuples ($n > 0$) then

- $1/n \leq S(A, R) \leq 1$, and
- if the values of A are all distinct then $S(A, R) = 1$

Similarly comment on the following selectivities (assume n tuples in R)

i. $S(A, R)$ when A is a boolean attribute. [2 *points*]

ii. $S(A, R)$ when A is a key for R. [2 *points*]

iii. $S(A, R)$ and $S(B, R)$ when R satisfies $A \to B$. [2 *points*]

iv. $S(A, \pi_A(R))$. [2 *points*]

v. $S(A, \mathsf{SELECT}\ A\ \mathsf{FROM}\ R)$. . [2 *points*]

3. The following four DTDs are proposed as an XML representation for a relational database whose scheme is R(A,B), S(A,C). Undefined elements are to be taken as PCDATA.

(i)
```
⟨!ELEMENT DB   (R,S)⟩
⟨!ELEMENT R    (RT*)⟩
⟨!ELEMENT S    (ST*)⟩
⟨!ELEMENT RT   (A,B)⟩
⟨!ELEMENT ST   (A,C)⟩
```

(ii)
```
⟨!ELEMENT DB   ((RT|ST)*)⟩
⟨!ELEMENT RT   (A,B)⟩
⟨!ELEMENT ST   (A,C)⟩
```

(iii)
```
⟨!ELEMENT DB   ((RT,ST)*)⟩
⟨!ELEMENT RT   (A,B)⟩
⟨!ELEMENT ST   (A,C)⟩
```

(iv)
```
⟨!ELEMENT DB   (((A,B)|(A,C))*)⟩
```

(v)
```
⟨!ELEMENT DB   (((A,(B|C))*)⟩
```

(a) For each DTD say briefly whether or not it is a good representation of the relational data. If there are problems, describe them. [5 *points*]

(b) Given the instances R =

A	B
a1	b1
a2	b2

S =

A	C
a3	c3

provide the XML corresponding to the DTDs (i), (ii) and (iv) above. [6 *points*]

(c) What would be an appropriate DDT if the XML representation of the instances above were

```
⟨DB⟩
  ⟨ROW⟩ ⟨A⟩ a1 ⟨/A⟩ ⟨B⟩ b1 ⟨/B⟩ ⟨/ROW⟩
  ⟨ROW⟩ ⟨A⟩ a2 ⟨/A⟩ ⟨B⟩ b2 ⟨/B⟩ ⟨/ROW⟩
  ⟨ROW⟩ ⟨A⟩ a3 ⟨/A⟩ ⟨C⟩ c3 ⟨/C⟩ ⟨/ROW⟩
⟨/DB⟩
```

[4 *points*]

(d) The XPath expression document("db.xml")/R/RT/A is the equivalent of $\pi_A(R)$ in representation (i). Write an XPath expression for $\pi_C(S)$ in representation (ii) [3 *points*]

(e) Write an XPath expression for $\pi_A(R)$ in the representation of your answer to question 3c above. [3 *points*]

(f) $\pi_A(\sigma_{A=B}(R))$ may be expressed in XQuery, using representation (i) as

```
⟨answer⟩ {
  for $X in document("db.xml")/R/RT
    where $X/A = $X/B
    return ⟨AT⟩ {$X/A} ⟨/AT⟩
} ⟨/answer⟩
```

Similarly, using representation (i), express $R \bowtie S$ in XQuery. [4 *points*]