UNIVERSITY OF DUBLIN CS4B1A1

TRINITY COLLEGE

FACULTY OF ENGINEERING & SYSTEMS SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

B.A. (Mod.) Computer Science **Trinity Term 2001**
Degree Examination

4BA1 INFORMATION SYSTEMS

Monday, 28th May 2001 The Mansion House 9.30–12.30

Prof. J. Grimson and Dr. M. Mac an Airchinnigh

Attempt **five** questions, at least **two** from each section.
Please use separate answer books for each section.
Students may avail of the HANDBOOK OF MATHEMATICS of Computer Science

SECTION A

1. Most large information-intensive organizations today purchase a standard Database Management System package which they then customize (general-purpose approach), rather than developing their own information system (tailor-made approach). Discuss the advantages and disadvantages of the general-purpose approach versus a tailor-made approach.

2. An employee database is to hold information about employees, the department they are in, and the skills which they hold. The attributes to be stored are **emp-id, emp-name, emp-phone, dept-name, dept-phone, dept-mgrid, skill-id, skill-name, skill-date, skill-level**.

\ . . . continued over.

An employee may have many skills, such as word processing, typing, librarian, web design, etc. The date on which the skill was last tested and the level displayed at that test are recorded for the purposes of assigning work and determining salary. An employee is attached to one department and each department has a unique manager.

(a) Draw a dependency (determinancy) diagram for the above database, stating clearly any assumptions that you make.

(b) Derive a set of well normalized (Boyce-Codd Normal Form) relations, indicating the primary key of each relation and identifying all foreign keys.

(c) Write an SQL query to determine the names and phone numbers of employees who can both file and type to a level of 6 or more.

3.

(a) What are the ACID properties of a transaction?

(b) Why do most Database Management Systems not allow nesting of transactions i.e.

begin transaction T_1
...
...
...
  begin transaction T_2
  ...
  ...
  commit T_2
...
...
commit T_1

(c) Explain, briefly, the role of serialization in determining the correctness of the concurrency control scheduler.

(d) Which concurrency control method would be best suited to a high-contention environment. (A high contention environment is one where many of the transactions are accessing the same data). Explain your answer.

\ ... continued over.

(e) Under Two-phase locking, why is it not possible for a transaction which has down-graded a lock on a data object from a **write** lock to a **read** lock to then upgrade it again to a **write** lock?

(f) Three transactions T_1, T_2, and T_3, execute concurrently according to the following schedule:

$$S = \{r_1(a), r_3(c), r_2(b), w_1(a), r_3(d), w_3(c), r_3(e), w_2(b), r_1(b), \\ r_2(a), w_1(b), w_2(a), r_3(e)\}$$

where $r_i(x)$ and $w_i(x)$ denote a read and write operation, respectively, on data object x by transaction T_i.

Is this schedule S serialisable and, if so, what is the equivalent serial schedule?

4.

(a) Define the following terms:

(i) checkpoint record,

(ii) restart file,

(iii) write ahead protocol.

(b) What actions must the Recovery Manager take at an asynchronous checkpoint?

(c) Outline the restart procedure for the undo-redo recovery protocol assuming asynchronous checkpointing is used.

(d) Modify the algorithm for synchronous checkpointing.

(e) Which of the four principal recovery protocols involves the

(i) lowest overhead,

(ii) highest overhead,

during normal operation i.e. in the absence of failure?

SECTION B

5. We may consider the trees of Trinity College to be physically distributed among a set of disjoint physical regions on the College Green site (i.e., the island site). Each tree is uniquely tagged with a numerical label. To capture these notions let us introduce the models

$$\lambda \in \mathit{TREES} = \mathit{LABEL} \longrightarrow \mathit{TREE} \tag{1}$$

$$\kappa \in \mathit{HABITATS} = \mathit{REGION} \longrightarrow (\mathit{LABEL} \longrightarrow \mathit{TREE}) \tag{2}$$

where $X \longrightarrow Y$ denotes the entire collection of total and partial maps from X to Y. Now that we have labelled trees it seems appropriate to introduce an information base. Specifically, for each labelled tree there is information giving its common name, its scientific name, the date of planting, a brief description of its leaves and fruit, an image perhaps, etc. We model this quite simply and abstractly by

$$\gamma \in \mathit{INFOBASE} = \mathit{LABEL} \longrightarrow \mathit{INFO} \tag{3}$$

(a) Write precise mathematical expressions for each of the following:

(i) "the labels of the *TREES* are the same as those of the *HABITATS* as well as those of the *INFOBASE*";

(ii) "a new tree t labelled l is planted in the region r and the associated information i recorded".

(b) Before the construction of the New Library, currently underway, it was necessary to move our trees from one region of the College to another. In the specification of such a move operation the concept of "system of trees" was introduced and modelled by the Cartesian product $\mathit{SYSTEM} = \mathit{TREES} \times \mathit{HABITATS} \times \mathit{INFOBASE}$. Using this definition the moving of a tree labelled l from region r to region s may be specified by an operation of the form

$$\mathit{Move}\colon \mathit{LABEL} \times \mathit{REGION} \times \mathit{REGION} \times \mathit{INFO} \longrightarrow (\mathit{SYSTEM} \longrightarrow \mathit{SYSTEM}) \tag{4}$$

$$\mathit{Move}[l, r, s, i]\langle \lambda, \kappa, \gamma \rangle := \langle \ldots \rangle \tag{5}$$

where i denotes the new information. Fill in the details. Assume any pre-conditions that you feel are necessary.

(c) Ignoring the change to the *INFOBASE* prove that a simplied version of your *Move* operation satisfies the property that $(\mathit{Move}[l, s, r] \circ \mathit{Move}[l, r, s])\kappa = \kappa$.

6. A garden γ is deemed to consist of a collection of areas or regions in each of which some plants share their space with artifacts ($x \in X$), such as paths, fountains, seats, etc., and non-plant life forms ($y \in Y$), such as earthworms, frogs, etc. The space of such gardens may be modelled by:

$$\gamma \in GARDEN = AREA \longrightarrow (PLANTS \times Xs \times Ys) \tag{6}$$

$$PLANTS = \mathcal{P}PLANT, \quad Xs = \mathcal{P}X, \quad Ys = \mathcal{P}Y \tag{7}$$

subject, of course, to an appropriate invariant.

One of the primary goals of the process of garden design is considered to be the particular harmonious arrangement of the areas of a garden in which plants are suitably matched with each other, with soil and other conditions, and with certain artifacts and life forms. To assist the designer in this work, a catalogue κ of plants is provided:

$$\kappa \in CATALOGUE = PLANT \longrightarrow (\mathcal{P}AREA \times Xs) \tag{8}$$

whereby one may check whether one's favourite plant, such as Tamarisk (*Tamarix gallica*), is appropriate for certain areas in a garden and what sorts of artifacts might be deemed suitable to accompany it.

(a) Write simple English interpretations for each of the following:

(i) $\cup/ \circ \mathsf{rng} \circ (1_{AREA} \rightarrow \pi_1)\gamma = \mathsf{dom}\,\kappa$, where 1_{AREA} is the identity map on *AREA*.

(ii) $\{a\} \subseteq \cup/ \circ \mathsf{rng} \circ (1_{PLANT} \rightarrow \pi_1) \lhd_{\pi_1\gamma(a)} \kappa$.

(b) In a brainstorming session on the user requirements for the development of software for computer-aided garden design, some doubt was expressed as to the usefulness or even naturalness of the above catalogue model. Indeed, an alternative was proposed:

$$\kappa_1 \in CATALOGUE_1 = PLANT \longrightarrow (AREA \longrightarrow Xs) \tag{9}$$

Obtain expressions for the constraints that the associations between (i) plants and areas, and (ii) plants and artifacts, should be the same in both models.

(c) The retrieve map $\mathcal{R}$ which takes κ_1 to κ in a natural way is given by

$$\mathcal{R}(\kappa_1) = (1_{PLANT} \rightarrow \mathsf{dom})\kappa_1 \bowtie (1_{PLANT} \rightarrow \cup/\,\mathsf{rng})\kappa_1$$

Let the adding of a new plant p to the catalogue κ_1 be specified by $\kappa_1 \sqcup [p \mapsto \mu]$. If this refines $\kappa \sqcup [p \mapsto \langle A, X \rangle]$ then determine the exact relationship between μ and $\langle A, X \rangle$.

7. Consider the relationship between master and disciple. In modelling this relationship we may wish to choose between *total* maps such as

$$MAST \xrightarrow{\alpha} DISC \tag{10}$$

$$DISC \xrightarrow{\beta} MAST$$

$$MAST \xrightarrow{\gamma} \mathcal{P}DISC$$

Due to the nature of the master-disciple relation we note that there can not be a master without the corresponding disciple. Nor can there be a disciple without the corresponding master. Furthermore, it is an absolute requirement that any model adequately capture the notion that "no one can be the disciple of two masters at the same time". Let us call this **the constraint.** In addition, by the very nature of this reality, it will always be the case that the number of disciples is greater than or equal to the number of masters.

(a) Give your opinion on each of the maps α, β, and γ, saying whether or not it is injective (one-to-one) or surjective (onto) and write mathematical expressions for **the constraint** in each case, if appropriate. In particular declare your opinion whether the map is a good or bad model, giving reasons.

(b) From your analysis of part (a) you will have concluded that $DISC \xrightarrow{\beta} MAST$ should be considered the canonical model for the master-disciple relation. Specifically, to each master m in the codomain there is a corresponding non-empty fiber of disciples $\beta^{-1}(m)$. With the aid of a suitable example, or otherwise, explain the idea of taking sections $MAST \xrightarrow{\sigma} DISC$ through the fibers. Show that $\beta \circ \sigma = 1_{MAST}$.

(c) Every canonical map $DISC \xrightarrow{\beta} MAST$ will give rise to n sections, $n \geq 1$,

$$MAST \underset{\sigma_n}{\overset{\sigma_1}{\underset{\longrightarrow}{\overset{\longrightarrow}{\cdots}}}} DISC \xrightarrow{\beta} MAST \tag{11}$$

Show how one may "sum" these sections to obtain a map

$$MAST \xrightarrow{\gamma} \mathcal{P}DISC \tag{12}$$

[**Hint:** one might consider the use of an operator $(1_{MAST} \to \mathsf{j})$ where $\mathsf{j}(d) = \{d\}$ and then application of an indexed union operator $\circledcup$.]

8. The $\Sigma^\star$-morphisms ψ from the free monoid over an alphabet $\Sigma = \{a_1, a_2, \ldots, a_n\}$, denoted $(\Sigma^\star, \cdot, 1)$, into a monoid $(M, +, e)$ form a collection of fundamental word processing algorithms and, by extension, list processing algorithms.

(a) Write down a naïve recursive algorithm to compute the length of a list and show how to obtain its corresponding tail-recursive form.

(b) There is a sense in which the length map len is analogous to the (natural) logarithm map log. Explain. Justify the analogy by developing the notion of inverse words (or inverse lists) and show that, for example, $\mathsf{len}(\sigma^{-1}) = -\mathsf{len}(\sigma)$.

(c) The $\Sigma^\star$-morphisms ψ have corresponding tail-recursive forms ψ_w, $w \in \Sigma^\star$, where

$$\begin{aligned} \psi_{uv}(m) &= \psi_v \psi_u(m) \\ \psi_{av}(m) &= \psi_v(m + F(a)) \\ \psi_1(m) &= m \end{aligned}$$

with the connection $\psi_w(m) = m + \psi(w)$, $\psi(w) = \psi_w(e)$ and $F\colon \Sigma \longrightarrow M$. Show that ψ_w has the closed form:

$$\begin{aligned} \psi_w(m) &= m + {}^{+}/ \circ F^\star w \\ \psi_1(m) &= m \end{aligned}$$

[Note that we often use the more elegant form ψ_w in preference to the 'linear' syntactic form $\psi[w]$.]