

OLLSCOIL NA hÉIREANN
The National University Of Ireland

NATIONAL UNIVERSITY OF IRELAND, GALWAY

SUMMER EXAMINATIONS 1999

SECOND YEAR EXAMINATION IN COMPUTING STUDIES

[CS201]

Second Paper

METHODOLOGY and INFORMATION SYSTEMS

Prof. D. Bell
Dr. G. Lyons
Ms. J. Griffith

Candidates are required to answer **FIVE** questions.
Two questions must be answered from each section

All questions carry equal marks.

Time allowed: **Three hours**

SECTION A

- Q. 1** i) Distinguish between the preorder, inorder and postorder traversal of a binary tree. A binary tree T has 9 nodes. The inorder and preorder traversals of T yield the following sequences of nodes:

Inorder: E A C K F H D B G

Preorder: F A E K C D H G B

Draw the tree T, explaining the steps taken in your answer.

- ii) What data structure would you choose if you were told that the elements in a list must be kept in sorted order and that the majority of operations on the data structure will be insertions. Explain your answer.
- iii) Outline the steps required in an insertion algorithm for the following three data structures: arrays, linked lists and binary search trees.

- Q. 2** i) Develop an algorithm to find an element in a sorted list of numbers where:
- a) the elements are stored in an array.
 - b) the elements are stored in a linked list.
- ii) Define a binary search tree and develop a searching algorithm using a binary search tree.
- iii) Compare the algorithm you developed in i) a) with the algorithm you developed in ii). What is the worst case situation when using binary search trees?
- Q. 3** i) Insertion sort and selection sort are two sorting techniques. Write pseudo-code to explain either algorithm.
- ii) Divide and conquer approaches to sorting represent an improvement on Insertion and Selection Sort.
- a) Explain what is meant by a *Divide and Conquer* technique.
 - b) Outline a divide and conquer algorithm for sorting.
- iii) Compare the performance of the divide and conquer algorithm you develop in ii) with the performance of bubble, insertion and selection sort.
- Q. 4** i) Develop an algorithm to find the N^{th} term in the sequence:
0, 1, 1, 2, 3, 5, 8, 13, 21,
- ii) Develop and write pseudo-code for an algorithm which merges two sorted arrays of length N and M into another array of length $N + M$ such that the sorted order is maintained.
- iii) Describe Huffman's algorithm for compression. The following 8 tree weights are given:

22 5 11 19 2 11 25 5

Construct a binary-tree with minimum-weighted path length using the above data and Huffman's algorithm.

SECTION B

- Q. 5**
- i) Outline the advantages and disadvantages of sequential file organisation. Detail some methods used when inserting new records.
 - ii) What is meant by the term “clustering”? Clearly distinguish between “inter” and “intra” file clustering in the course of your answer. Explain what is meant by a dense and a non-dense index and give an example of one of these.
 - iii) A company assigns a 4-digit employee number to each of its 68 employees. The number is used as a primary key in the company’s employee file. Given the following five employee numbers:

9614 5882 6713 4409 1825

Find a 2-digit hash address for each of the employee numbers using:

- a) division method with modulo 97.
- b) mid-square method.

Why is a collision resolution technique required? Describe one collision resolution procedure.

- Q. 6**
- i) Detail the main concepts associated with the relational model.
 - ii) Given the following relational schema and interpretation with keys underlined:

SUPPLIER: S_No, S_Name, Status, City
PART: P_No, P_Name, Size, Material, Colour
JOB: J_No, J_Name, City, Company, Start_Date
SHIPMENT: S_No, P_No, J_No, Quantity, Unit_Price

A supplier (with associated number, name, status and city location) supplies parts (with associated number, name, size, material type and colour) to jobs (with associated numbers, names, city location) in certain companies and with certain start dates. These parts have an associated unit price and are shipped in a certain quantity.

Develop SQL queries to satisfy the following information needs:

- a) List the name and city of all suppliers who have a status greater than 20.
- b) Find the total price of all parts for all jobs shipped to a company with company name *Dell*
- c) List the supplier number and supplier name of all suppliers who supply a part with part number *P4* or supply to a job with job number *J3*
- d) List the name of all parts supplied to a job located in *Athens*.
- e) List the supplier number and supplier name of all suppliers who supply a part with a blue colour.

iii) Discuss the factors to consider when normalising data.

Q. 7 i) List and describe the desirable properties of a transaction.

- ii) What is the system log? Describe the main entries used in a system log. Describe and outline recovery algorithms for a system operating under:
- a) immediate update protocol.
 - b) deferred update protocol.

iii) Discuss the problem that occurs with the following schedule:

| T1 | T2 |
|----------------|----------------|
| read_item(X); | |
| X := X - N; | |
| | read_item(X) |
| | X := X + M; |
| write_item(X); | |
| | write_item(X); |

Outline a procedure to prevent this type of problem from arising.

Q. 8 i) Why is a database approach not adequate for all information management needs? What is meant by the term *information overload*?

ii) Describe the main components of an IR system. How is the accuracy of an IR system measured? Give formulas for the metrics used.

iii) Describe the Vector Space Model.

iv) Discuss the use and a possible implementation of a stop list and a stemming algorithm.