UNIVERSITY *of* LIMERICK

OLLSCOIL LUIMNIGH

COLLEGE *of* INFORMATICS *and* ELECTRONICS

Department of Computer Science
and Information Systems

Final Assessment Paper

Academic Year:	2000/2001	Semester:	Autumn
Module Title:	Data Structures and Algorithms	Module Code:	CS4115
Duration of Exam:	$2\frac{1}{2}$ hours	Percent of Semester Marks:	75
Lecturer:	P. Healy	Paper marked out of:	100

Instructions to Candidates:

- There are three sections to the paper: Multiple Choice Questions, Short Questions and Long Questions
- The mark distribution is 40 marks for Multiple Choice Questions, 20 marks for Short Questions and 40 marks for the Long Questions
- Answer all questions in all sections
- **You *must* return this paper with your answer book and bubble sheet**

Section 1. Multiple Choice Answers (40 marks in total).

Use the machine-readable multiple-choice question grid that has been provided to answer these questions. Please completely mark in black exactly one circle on the grid for each answer.

1. The number of nodes in a *perfect* binary tree of height h is

   (a) $2^{h-1}-1$
   (b) 2^h-1
   (c) $2^{h+1}-1$
   (d) None of the above

2. Given the two statements below, which of them are true?
   S1: $a^{\log_c n} = n^{\log_c a}$, for a, c and n integer;
   S2: $log_i n \leq \log^i n$, for i and n integer.

   (a) Both statements are true
   (b) **S1** is true, but **S2** is false
   (c) **S1** is false, but **S2** is true
   (d) Both statements are false

3. Given the two statements below, which of them are true?
   S1: $p(n) = \Theta(q(n)) \Rightarrow q(n) = O(p(n))$;
   S2: $f(n) = O(g(n)) \Rightarrow g(n) = o(f(n))$.

   (a) Both statements are true
   (b) **S1** is true, but **S2** is false
   (c) **S1** is false, but **S2** is true
   (d) Both statements are false

4. Given the two statements below, which of them are true?
   S1: Heapsort's worst-case running time is worse than Quicksort's worst case running time;
   S2: Heapsort's average-case running time is poorer than Quicksort's average-case running time.

   (a) Both statements are true
   (b) **S1** is true, but **S2** is false
   (c) **S1** is false, but **S2** is true
   (d) Both statements are false

5. Finding the median of an array of n elements has running time best described by which of the following?

   (a) $O(n \log n)$
   (b) $O(\log n)$
   (c) $\Omega(n \log n)$
   (d) None of the above

6. When performing *external sorting*, the data structure used to efficiently do the k-way merge is

   (a) a linked list
   (b) an array
   (c) a binary tree
   (d) a heap

7. When performing *external sorting*, the number of passes required to sort n numbers m at a time using k tapes is

   (a) $\lceil \log_k \frac{n}{m} \rceil$
   (b) $\lceil \log^k \frac{n}{m} \rceil$
   (c) $\lceil k \log \frac{n}{m} \rceil$
   (d) $\lceil \log \frac{n}{mk} \rceil$

8. In order to reconstruct accurately a binary tree from its ordered traversals

   (a) all *three* traversals are necessary
   (b) any *two* are sufficient
   (c) the pre- and post-order traversals are necessary
   (d) some other combination *might* do the job

9. Given the two statements below, which of them are true?
   S1: In a hash table, a key insertion followed by a deletion of that key returns the hash table exactly to its state prior to the insertion;
   S2: In a hash table, an insertion of a non-existing key can be performed faster than a deletion of an existing key.

   (a) Both statements are true
   (b) **S1** is true, but **S2** is false
   (c) **S1** is false, but **S2** is true
   (d) Both statements are false

10. Given the two statements below, which of them are true?
    S1: The step of Heapsort that accounts for its asymptotic running time is the `buildheap()` step;
    S2: On pre-sorted input, Heapsort's running time is $O(n)$.

    (a) Both statements are true
    (b) **S1** is true, but **S2** is false
    (c) **S1** is false, but **S2** is true
    (d) Both statements are false

Section 2. Fill in the blank (5×4 marks).

- Please put your answers to these questions in the answer book provided to you, labelling your answers 2.1, 2.2, etc.

1. What is the analogue of a linear-time algorithm for a graph $G = (V, E)$? ________________.
2. Exponentiation can be done in ____________ time?
3. A *lower* bound for comparison-based sorting is ______________.
4. If memory is very plentiful in certain applications it may be profitable to consider as a sorting algorithm, ______________.
5. When multiplying together *huge* integers, new algorithms must be investigated. One possibility is the following.
   Suppose we want to multiply together 123,456 and 98,765,432. We can break the two numbers in two at their mid-point and then use the fact that $(a+b)(c+d) = ac+bc+ad+bd$. This is used in the following way: breaking 123,456 at the mid-point we get $a = 123 \times 10^3$ and $b = 456$; breaking 98,765,432 into two numbers we get $c = 9,876 \times 10^4$ and $d = 5,432$. So we now follow the formula above for $(a+b)(c+d)$. The powers of 10 can be treated separately since they amount to just adding on an appropriate number of 0s at the end of the number. For example, to compute $ac = 123 \times 10^3 \times 9,876 \times 10^4$ above, we multiply 123 and 9,876 and tack on at the end $3+4 = 7$ 0s.
   Since, in general, the numbers $a, b, c and d$ can be quite large themselves, we will need to do this recursively. Give a recurrence relation that quantifies the running time, $T(n)$, of multiplication of two n-digit long integers.

   ____

Section 3. Long Questions (40 marks).

- Please put your answers to these questions in the answer book provided to you
- Label your answers 3.1 and 3.2 in your answer books, please

1. Using a heap data structure, we can easily implement an $O(n \log n)$ algorithm for finding the *median* element of an array of numbers. A better alternative in the average case is the *quickselect* algorithm. (20 marks.)

   (a) Explain the working of the quickselect algorithm. (8 marks.)
   (b) Write down a recurrence relation for either
      i. the best or worst case running-time of quickselect (3 marks.)
      ii. the *average* case running-time of quickselect (6 marks.)
   (c) Now solve the recurrence relation you developed in part 1b above for
      i. the best or worst case running-time of quickselect (3 marks.)
      ii. the average case running-time of quickselect (6 marks.)

2. The Erdös Numbering Problem is to determine for a given researcher how *closely related* the researcher is to Paul Erdös, often considered to be the greatest mathematician / computer scientist of the 1900s.
   The Erdös Number (EN) of Erdös himself is 0; the Erdös an author is defined to be 1 if the author has co-written a paper in a mathematical journal with his majesty; otherwise, to determine an author's EN, find the lowest EN amongst all the people he has ever co-written a paper with and add 1 to that.
   So, if author "Joe Bloggs" has EN 7, then "Guisseppe Bloggini", a colleague and co-author of Joe Bloggs', will have EN *at most* 8. His may be lower if he has collaborated with more closely related colleagues of Erdös.
   Using data structures and algorithms that we have studied in class develop an algorithm that reads a database of co-authors and determines the EN of some specified author.
   What is the worst-case running time of your algorithm? What are its space requirements? (20 marks.)