IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2014

BEng Honours Degree in Mathematics and Computer Science Part II
MEng Honours Degree in Mathematics and Computer Science Part II
MSc in Computing Science
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the Associateship of the City and Guilds of London Institute

PAPER MC202

SOFTWARE ENGINEERING - ALGORITHMS

Monday 28 April 2014, 14:30
Duration: 75 minutes
(Reading time 5 minutes)

Answer ALL TWO questions

Paper contains 2 questions
Calculators not required

Section A (Use a separate answer book for this Section)

1 a A *counting Bloom set* keeps track of the number of elements in a set that have marked a given location in the bit vector.

i) What problem does the counting Bloom set solve? How?

ii) Write the psuedo code for an algorithm to add an element to a counting Bloom set: `AddElement(b,x)`.

iii) Write the pseudo code for an algorithm to query a Bloom set: `IsElement(b,x)`.

iv) Write the pseudo code for an algorithm to delete an element from a counting Bloom set: `DeleteElement(b,x)`.

b i) What is a radix and how does it relate to search algorithms?

ii) Explain the essential differences between a binary search tree and a binary search trie.

iii) There is a fundamental tradeoff in the design of multiway search tries. Explain that tradeoff.

iv) Consider the following character strings.

`THE, THEN, THIN, THIS, TIN, SIN, SING`

Give a simple C++ class or struct for representing nodes in an existence trie designed for character strings such as those above. You only need to show the data member(s) and basic constructor for the class.

v) Draw a picture of the existence trie resulting from inserting the following character strings in the given sequence. (In the picture, you can represent a node in the trie as a simple black "dot".)

`THE, THEN, THIN, THIS, TIN, SIN, SING`

vi) If the insertion order were reversed, would the resulting existence trie be the same or different? Explain your answer.

The two parts carry, respectively, 30% and 70% of the marks.

Section B (Use a separate answer book for this Section)

2a The "simpler" version of the master method determines the big O solution of a recurrence as one of three cases relating the terms d and $\log_b a$.

i) Write down a general form for recurrences $T(n)$ using variables a, b, d.

Given the following recurrence $T(n) = 2T(n/2) + 1$.

ii) Which of the three cases of the master method gives the solution $T(n) = O(n)$? Write down the relation of d and $\log_b a$ and show that it holds by inserting the corresponding values from the recurrence.

iii) Show that a proof using the substitution method fails for the assumption $T(n) \leq cn$. Make the proof work by subtracting a lower-order term. Write down only the inductive step. You do not need to show the base case.

b Divide and Conquer

i) Write in pseudo-code a recursive function $f(x, n) = x^n$ for powering a number using a divide and conquer strategy.

Hint: $x^n = x^{n/2} \cdot x^{n/2}$ for even n; $x^n = x^{(n-1)/2} \cdot x^{(n-1)/2} \cdot x$ for odd n.

ii) Show that the running time complexity of this function is $O(\lg n)$.

Hint: Define the recurrences for the cases with even and odd n. Then use the master method or substituion method to show $T(n) = O(\lg n)$.

c Compute the Levenshtein distance between the two strings WEIRD and WIRED by constructing the corresponding 6x6 distance table. Remember that you need to initialise the first row and first column with the values 0 to 5.

i) Fill in the values of all entries in the table and highlight with arrows a path with minimum distance from entry $[5, 5]$ back to entry $[0, 0]$.

ii) Write down for each character of the string WIRED which operation of 'keep', 'delete', 'insert', or 'replace' is applied to obtain WEIRD.

The three parts carry, respectively, 35%, 35%, and 30% of the marks.