Birla Institute of Technology & Science - Pilani, Hyderabad Campus
Second Semester 2014-2015
CS F211 / IS F211: Data Structures and Algorithms
Comprehensive Examination

Type: Closed **Time: 180 mins** **Max Marks: 100** **Date: 05.05.2015**

1.a. Assume you have an array A[1..n] of n elements. A *majority element* of A is any element occurring in more than n/2 positions (so if n = 6 or n = 7, any majority element will occur in at least 4 positions). Assume that elements cannot be ordered or sorted, but can be compared for equality. (You might think of the elements as chips, and there is a tester that can be used to determine whether or not two chips are identical.) Design an efficient divide and conquer algorithm to find a majority element in A (or determine that no majority element exists). Aim for an algorithm that does O(n log n) equality comparisons between the elements. [10 Marks]
Note: Any algorithm that takes more than O(n log n) will not be considered for evaluation.

1.b. Consider the following variation on Mergesort for large values of n. Instead of recursing until n is sufficiently small, recur at most a constant r times, and then use insertion sort to solve the 2^r resulting subproblems. What is the (asymptotic) running time of this variation as a function of n? [7 Marks]

2.a. In each of the following situations, indicate whether f = O(g), or f =Ω(g), or both (in which case f = Θ(g)).

(i). $f(n) = n^{1.01}$ $g(n) = n \log^2 n$
(ii). $f(n) = 2^n$ $g(n) = 2^{n+1}$
(iii). $f(n) = 1^k + 2^k + \ldots + n^k$ $g(n) = n^{k+1}$
(iv). $f(n) = \log(3n)$ $g(n) = \log(2n)$ [6 Marks]

2.b. For the following function f(n), find a simple function g(n) such that f(n) = Θ(g(n))

$$f(n) = (0.99)^n + n^{100}$$ [4 Marks]

2.c. Given two sets S1 and S2 (each of size n), and a number x, describe an O(nlogn) algorithm for finding whether there exists a pair of elements, one from S1 and one from S2, that add up to x. [6 Marks]

3.a. Consider the problem of deciding whether two multisets of integers (that is, sets with repeated elements allowed), each of size n, are identical in the sense that each integer occurs the same number of times in both sets. Assume that all integers i in the multisets are in the range $0 \leq i \leq m$. Give a deterministic algorithm for testing whether two multisets are identical that runs in O(n log n) time. [7 Marks]

3.b. A program, Prog1, written by one of the programmer in an IT organization uses an implementation of the sequence ADT (data structure like an array, linked list) as its main component. It performs atRank, insertAtRank and remove operations in some unspecified order. It is known that Prog1 performs n2 atRank operations, 2n insertAtRank operations, and n remove operations. Which implementation of the sequence ADT should the programmer use in the interest of efficiency: the array-based one or the one that uses a doubly-linked list? Support your answer with the help of asymptotic notation and analysis. [8 Marks]

4.a. Suppose that we are given a sequence of n values x1, x2, ..., xn and seek to quickly answer repeated queries of the form: given i and j, find the smallest value in xi, . . . , xj . Design a data structure that uses O(n2) space and answers queries in O(1) time. [6 Marks]

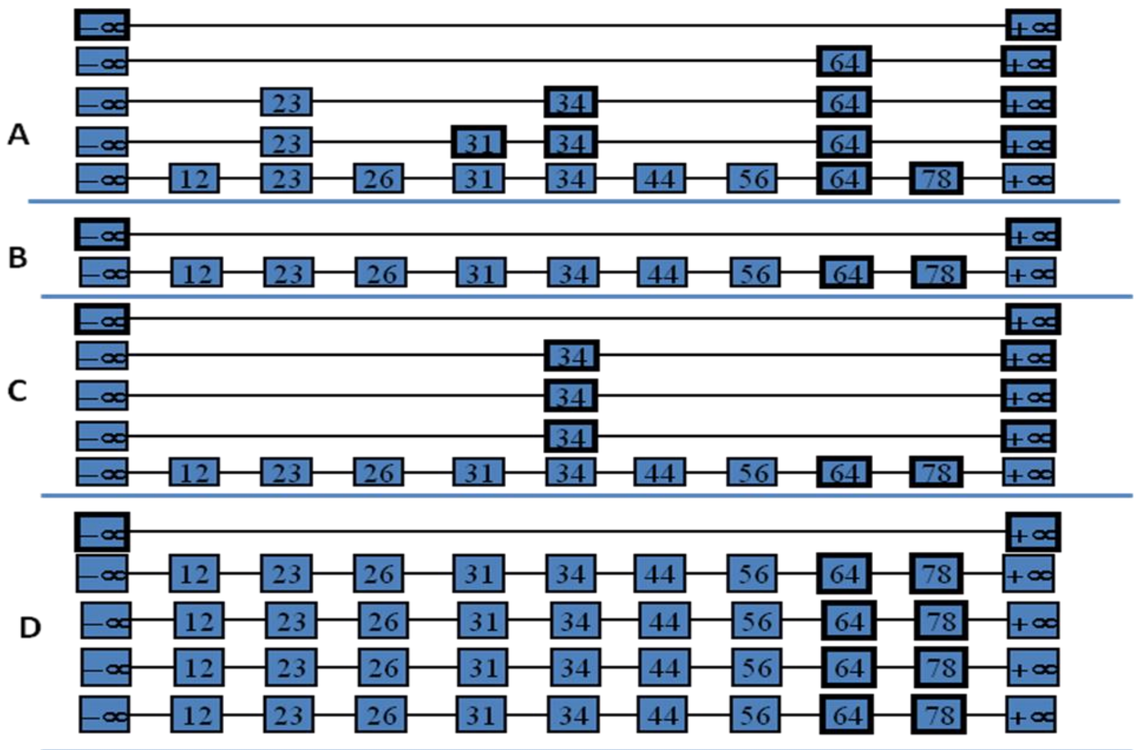
4.b. Prove that s successful search in a hash table takes expected time Θ(1+α). [8 Marks]

4.c. **P**rove that a red-black tree with n internal nodes has height at most 2 lg(n+1). [7 Marks]

5.a. Suppose you are given two sets A and B, each containing n positive integers. You can choose to reorder each set however you like. After reordering, let a_i be the ith element of set A, and let b_i be the ith element of set B. You then receive a payoff of $\prod_{i=1}^{n} (a_i b_i)$. Give an algorithm that will maximize your payoff. And also prove that your algorithm maximizes the payoff. [6 Marks]

5.b. Edit Distance - Given two text strings A of length n and B of length m, you want to transform A into B with a minimum number of operations of the following types: delete a character from A, insert a character into A, or change some character in A into a new character. The minimal number of such operations required to transform A into B is called the edit distance between A and B. [6 Marks]

5.c. Which of the following will not represent the outcome of a randomized skip list data structure? [3 Marks]



6.a. Prove that the height of a skip list on a set of n elements is more than $1 + t \log n$ with probability less than $1/n^{t-1}$ [6 Marks]

6.b. A forensic lab receives a delivery of n samples. They look identical, but in fact, some of them have different chemical composition. There is a device that can be applied to two samples and tells whether they are different or not. It is known in advance that most of the samples (more than 50%) are identical. Find one of those identical samples making no more than n comparisons. (Beware: it is possible that two samples are identical but do not belong to the majority of identical samples.) [10 Marks]