ICS 261 Fall 2000
Data Structures
Midterm 1—Version A
October 26, 2000

This is a closed-book test. You may not use calculators, books, or notes during the exam.

Please write all of your answers on the answer sheet, and write your name and ID number, as well as the test version, on both sides of the answer sheet. This is version A. You may keep the exam.

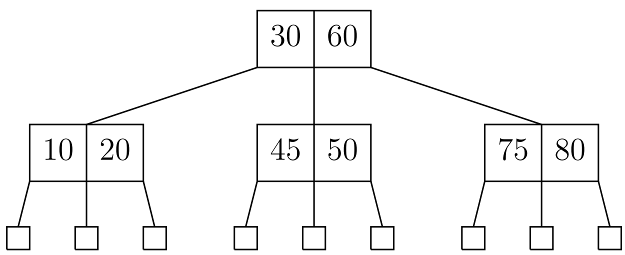
I suggest you look over the entire midterm before starting.

The maximum possible score is 55 on Part I and 45 on Part II, for a total of 100.

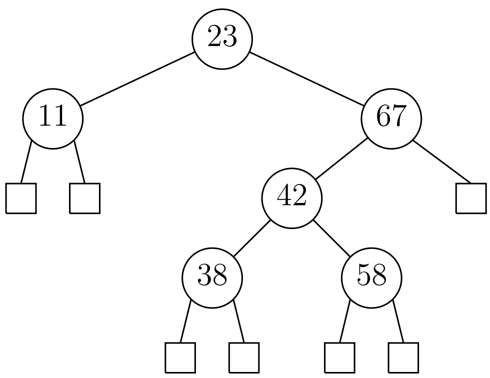
Part I. Written answers

Point values for each question in this part are given in square brackets after the question number. The notation "explanation optional" on some problems means that a completely correct final answer will get full credit, though an explanation could conceivably help you get partial credit.

1. [10—Explanation optional] Show the result of inserting a 47 into the B-tree of order 3 shown below. (In a B-tree of order 3 every internal node has has 2 or 3 children; these trees are also called (2,3) trees.)



2. [10—Explanation optional] Show the result of inserting a 35 into the splay tree shown below. (Don't forget to perform the appropriate splay.)



3. [10] Suppose we wish to use a B-tree of order m for data on a disk drive. Choosing m either too large or to small can worsen the performance.

   a) Why does the search time increase if m is chosen too small?

   b) Why does the search time increase if m is chosen too large?

Here is a reminder of one of the homework problems and its solution, in case it helps with the next problem. (But you don't necessarily have to use this to solve the next problem.) We asked "Suppose that we build a binary search tree by inserting n distinct items into an initially empty tree, and that all permutations of the input are equally likely. Let $p_{n,d}$ be the probability that the resulting tree has height at most d. (Measure height in the extended tree.) Give a recurrence that expresses $p_{n,d}$ in terms of values of p with smaller indices." The answer was

$$p_{n,d} = \frac{1}{n} \sum_{i=1}^{n} p_{i-1,d-1}\, p_{n-i,d-1}.$$

4. [15] Justify your answers clearly for this problem.

   a) Suppose we insert 7 keys into an initially empty binary search tree; the keys are all distinct and each of the 7! insertion orders is equally likely. What is the probability that the resulting extended binary search tree will have the minimum possible height (i.e., height 3)?

   b) Let q_k denote the probability that if we insert $2^k - 1$ distinct keys into an initially empty binary search tree, with each of the $(2^k - 1)!$ insertion orders equally likely, we obtain an extended binary search tree with the minimum possible height (i.e., height k). Assuming $k > 1$, give a formula for q_k in terms of q_{k-1}.

5. [10] In class we said that a family $\mathcal{H}$ of hash functions was *universal* if each hash function mapped keys into the range $\{0, 1, 2, \ldots, m-1\}$, and $\mathcal{H}$ had the property that for any two distinct keys x and y, only $m^{-1}|\mathcal{H}|$ of the hash functions h in $\mathcal{H}$ would result in the collision $h(x) = h(y)$. (Equivalently, for any $x \neq y$, if we select a hash function randomly from $\mathcal{H}$, the probability that $h(x) = h(y)$ would be only m^{-1}.)

   For an example, we let each key x be decomposed into $r + 1$ bytes as

$$x = (x_0, x_1, \ldots, x_r)$$

   where a byte could hold a number in the range $\{0, 1, \ldots, b\}$. Then let m be a prime number such that $m > b$. Construct $\mathcal{H}$ as follows: For $i \in \{0, 1, \ldots, r\}$, choose a_i uniformly and independently from $\{0, 1, \ldots, m-1\}$, and define

$$h_a(x) = \sum_{i=0}^{r} a_i x_i \bmod m.$$

   We showed that this would be universal.

a) [2] Suppose that we proceed as above but instead of picking the a_i separately, we just pick one random value from $\{0, 1, \ldots, m-1\}$ and set all of the a_i to that value. Is the resulting class of hash functions still guaranteed to be universal?

b) [8] Clearly justify your answer for part (a) by a proof or counterexample, as appropriate.

Part II. Multiple choice.

For each question, choose the *best* answer and write its letter clearly as a large plain capital on the answer form. Each question is worth 5 points, and there is no penalty for guessing.

It's a good idea to read each question carefully and look at all the choices. (Sometimes they may be in a different order than you would expect.)

6. Which of the functions shown below grows most quickly as n approaches infinity?

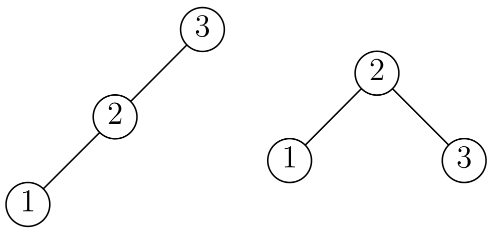
   A. $n3^n(\log n)^3$ B. $n^2 3^n$ C. $n^5 2^n$ D. n^{10} E. $(\log n)^{100}$

7. Which of the choices below correctly describes the growth rate of the function $(3n)!/(n!)^3$? (As a reminder, Stirling's formula states that $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.)

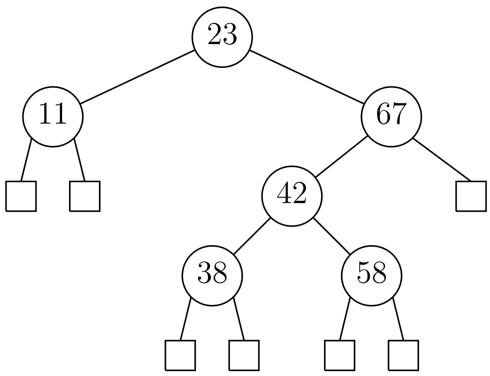
   A. $\Theta(8^n)$ B. $\Theta(n^{-3}\,27^n)$ C. $\Theta(n^{-1}\,27^n)$ D. $\Theta(27^n)$ E. $\Theta(n\,27^n)$

8. The two trees shown below are equivalent if we are considering them as

   A. Ordered trees
   B. Rooted trees
   C. Free trees
   D. None of the above



9. What is the internal path length of the binary search tree shown below?



   A. 6 B. 8 C. 10 D. 12 E. 15

10. Which of the following guarantees that only $O(1)$ rotations will occur when an insertion is done?

    A. AVL trees

    B. Red black trees

    C. Splay trees

    D. A and B

    E. A and C

11. Suppose that we start with an arbitrary valid red-black tree, and then perform the action given in one of the choices below. For which choice would the tree still be guaranteed to be a valid red-black tree?

    A. Change the root from red to black.

    B. Change the root from black to red.

    C. Select a path from the root to some leaf. Change all of the nodes on this path to black.

    D. Select a path from the root to some leaf. Change all of the nodes on this path to red.

    E. A and C

12. Suppose we are designing a division hashing function h. In particular, given a key K we will let z be the integer whose binary representation we obtain by concatenating 7-bit codes for the characters in K, and then let $h(K) = z \bmod m$. Which of the choices below would probably be the best choice for m?

    A. 127 B. 128 C. 137 D. 196 E. 256

13. For which of the collision resolution methods for hashing given below does the performance degrade the worst as the load factor α approaches 1?

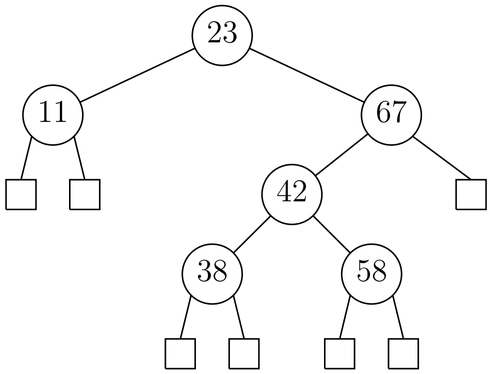
    A. double hashing

    B. linear probing

    C. open addressing with uniform hashing

    D. separate chaining

14. Suppose that we insert the key 35 into the splay tree shown below. By how much will the potential function Φ increase due to the insert *before* we do the appropriate splay? (Give the exact value of the increase, not a bound. Recall that the potential function is defined as the sum, over all nodes x, of $\lg |x|$, where in this context $|x|$ denotes the number of external nodes in the subtree rooted at x.)



A. $\lg \frac{7}{36}$ B. $\lg \frac{36}{7}$ C. $\lg \frac{9}{35}$ D. $\lg \frac{35}{9}$ E. 3