Data Structures and Algorithms Practice Exam

Introduction

The Data Structures and Algorithms portion of the Masters Comprehensive Exam (MCE) represents approximately 1/3 of the exam. The DS&A portion emphasizes basic knowledge and reasoning over in-depth problem analysis, since it has over 30 questions, and the entire comprehensive exam lasts for three hours.

Taking the exam

Since the MCE lasts for three hours, this leaves about one hour to complete the DS&A portion, which in turn leaves about two minutes to answer each problem. There is NO penalty for guessing, so make sure to answer each one. It is therefore recommended to take this practice test in a hour sitting, and without referring to the answers until the exam has been completed.

Grading of the exam

Exam grades are determined by the CECS Graduate Committee. Please see the CECS Graduate Advisor for more information.

Questions

1. Which of the following expressions surely supports the statement $f(n) = \Omega(g(n))$?

    a) $f(n) \leq 4g(n)$ for all $n \geq 1$

    b) $f(n) \geq 4g(n)$ for all $n \geq 136$

    c) $\lim\limits_{n \to \infty} \frac{f(n)}{g(n)} = 0$

    d) none of the above

2. Let k denote the degree of polynomial $p(n)$, and l the degree of polynomial $q(n)$. If $p(n) = o(q(n))$, then necessarily

    a) $k = l$.

    b) $k < l$.

    c) $k > l$.

    d) none of the above

3. If $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$, then it is always true that

    a) $f(n) = o(g(n))$.

    b) $f(n) = \Theta(g(n))$.

    c) $f(n) = \omega(g(n))$.

    d) both a) and b) are always true.

4. Which of the following expressions best describes the growth of the sum $1 + 1/2 + 1/3 + \cdots + 1/n$?

    a) $\Theta(1/n)$

    b) $\Theta(1)$

    c) $\Theta(\log n)$

    d) $\Theta(n \log n)$

5. Given positive functions $f(n)$ and $g(n)$, if we know that $\lim_{n \to \infty} (\log f(n) - \log g(n)) = 1$, then we also know that

    a) $f(n) = o(g(n))$.

    b) $f(n) = \Theta(g(n))$.

    c) $f(n) = \omega(g(n))$.

    d) more information is needed about f and g to reach a definite conclusion.

6. An algorithm takes as input an $n \times n$ Boolean matrix A. If the running time of the algorithm is $T(n) = O(n \log n)$ when n is used as the input size parameter, then which of the following expressions describes the big-O growth of $T(m)$, the running time of the algorithm when $m = n^2$ is used as the size parameter?

    a) $O(\sqrt{m} \log m)$

    b) $O(m^2 \log m)$

    c) $O(m \log m)$

    d) $O(m^2 \log^2 m)$

7. Which of the following is *not* a desirable property of a hash function $h(x)$?

a) It should be computable in O(1) time.

b) The range of $h(x)$ should stay within the desired hash-table size.

c) The range of $h(x)$ should include a wide range of integers.

d) If $x_1, \ldots, x_n$ are the items to be hashed, then the numbers $h(x_1), \ldots, h(x_n)$ should be uniformly distributed over the integers.

8. If a separate-chaining hash table has load factor $\lambda = 5$, then the average length of a chain equals

a) 5.

b) $\log 5$.

c) $5 \log 5$.

d) $\log(\log 5)$.

9. Perfect hashing

a) may be applied to static hash tables.

b) means there are no collisions when retrieving data from the hash table.

c) represents one of the most efficient ways of retrieving data.

d) all of the above

10. The worst case running time $T(n)$ for inserting n elements into an initially-empty heap is

a) $T(n) = O(n^2)$.

b) $T(n) = O(\sqrt{n} \log n)$.

c) $T(n) = O(n)$.

d) $T(n) = O(n \log n)$.

11. The worst case running time $T(n)$ for popping an element from a binary heap of size n is

a) $T(n) = O(1)$.

b) $T(n) = O(\log n)$.

c) $T(n) = O(n)$.

d) $T(n) = O(n \log n)$.

12. Which of the following algorithms does *not* require a heap for its efficient implementation?

a) Huffman's algorithm

b) Dijkstra's algorithm

c) Kruskal's algorithm

d) Prim's algorithm

13. Which of the following graph problems cannot be solved in time that is linear with respect to the sum of the number of vertices and edges in the graph (i.e. $O(m + n)$).

    a) determining if a simple graph is connected

    b) determining if a simple graph is bipartite

    c) determining a minimum spanning tree for a connected weighted graph

    d) determining a topological sort of the vertices for a directed acyclic graph

14. Suppose directed graph G is weakly but not strongly connected, and path $P = a, b, c, d$ provides a weak (but not strong) connection from a to d within G. Then we surely know that

    a) either (a, b), or (b, c), or (c, d) is an edge of G.

    b) (a, b), and (b, c), and (c, d) are edges of G.

    c) either (b, a), or (c, b), or (d, c) is an edge of G.

    d) (b, a), and (c, b), and (d, c) are edges of G.

15. Let $G = (V, E)$ be the undirected graph whose edges are $\{(a, b), (b, c), (c, d), (a, d)\}$. If T is the depth-first spanning tree of G rooted at vertex a, it follows that T has ____ branch(es).

    a) 1

    b) 2

    c) 3

    d) 4

16. The worst case running time $T(n)$ for inserting n elements into an initially-empty binary search tree is

    a) $T(n) = O(n^2)$.

    b) $T(n) = O(\sqrt{n} \log n)$.

    c) $T(n) = O(n)$.

    d) $T(n) = O(n \log n)$.

17. Binary search tree T is said to be balanced when, for every node n in T,

    a) n's left and right subtrees have equal height.

    b) n's left and right subtrees have equal size.

    c) n's left and right subtrees have heights that differ by at most one.

    d) n's left and right subtrees have sizes that differ by at most one.

18. If numbers from the set $\{1, \ldots, n\}$ are selected at random (without replacement) and inserted into an initially-empty binary search tree, then the big-O expression that best describes the average (taken over all possible resulting trees) height of the resulting tree is

    a) $O(n)$.

b) $O(\log n)$.

c) $O(\sqrt{n})$.

d) $O(n)$.

19. Which of the following sorting tasks would *not* be appropriate for Counting Sort?

    a) sorting millions of products by their 15-digit serial numbers

    b) sorting millions of employees by their ages

    c) sorting each day of the calendar year by the high temperature (in Celsius rounded to the nearest integer degree) recorded in the city of Long Beach on that given day

    d) sorting each vehicle registered in the state of California by the year it was manufactured

20. Which of the following best describes the worset-case running time of Insertion Sort when sorting n 32-bit unsigned integers?

    a) $T(n) = \Theta(\sqrt{n})$

    b) $T(n) = \Theta(n)$

    c) $T(n) = \Theta(n \log n)$

    d) $T(n) = \Theta(n^2)$

21. Quicksort is guaranteed to run in time $O(n \log n)$ so long as the pivot is

    a) randomly selected.

    b) set to the median of the first, middle, and last array element.

    c) set to the median of the array.

    d) none of the above

22. When following Kruskal's algorithm, the greedy choice is to

    a) remove the edge of greatest cost from the graph so long as its removal does not disconnect the graph.

    b) add the edge of least cost to the forest so long as its addition does not create a cycle.

    c) add the vertex having least connection cost to the current tree.

    d) remove the vertex having greatest connection cost from the tree.

23. When following Huffman's algorithm, the greedy choice is to

    a) select the two probabilities of least value for merging.

    b) select the two probabilities of greatest value for merging.

    c) add the codeword of shortest length to the code being constructed.

    d) remove the codeword of greatest length from the code being constructed.

24. Which of the following greedy algorithms has an implementation that most resembles that of Dijkstra's algorithm?

a) Kruskal's algorithm

b) Prim's algorithm

c) Huffman's algoirthm

d) Floyd-Warshall algorithm

25. If $T(n)$ satisfies $T(n) = 2T(n/3) + \sqrt{n}$, then

a) $T(n) = \Theta(\sqrt{n})$.

b) $T(n) = \Theta(n^2)$.

c) $T(n) = \Theta(n^{\log_3 2})$.

d) $T(n) = \Theta(n^{\log 3})$.

26. Which of the following recurrences *cannot* be solved directly by the Master Theorem?

a) $T(n) = 16T(n/4) + n$

b) $T(n) = T(n/5) + 20$

c) $T(n) = 3T(n/4) + n \log n$

d) $T(n) = 2T(n/2) + n \log n$

27. The running time of Strassen's algorithm is

a) $T(n) = \Theta(n^2)$.

b) $T(n) = \Theta(n^{2.5})$.

c) $T(n) = \Theta(n^{\log 7})$.

d) $T(n) = \Theta(n^3)$.

28. In dynamic-programming "memoization" refers to

a) storing the dynamic-programming recurrence in a static array so that it may be applied at each level of recursion.

b) storing solutions to smaller subproblems for future look up when solving larger subproblems.

c) the ability of a dynamic-programming algorithm to remember the current location in the recursion tree.

d) the ability to remember dynamic-programming recurrences for a comprehensive exam.

29. All of the following problems can be solved with a dynamic-programming algorithm that runs in worst-case polynomial time, *except* for

a) Matrix Chain Multiplication.

b) Optimal Binary Search Tree

c) Edit Distance Between Words

d) 0-1 Knapsack

30. The Floyd-Warshall dynamic-programming algorithm defines the matrix of numbers d_{ij}^k, where i and j are graph vertices satisfying $1 \leq i, j \leq n$, and $k \in \{0, 1, \ldots, n\}$. Moreover, d_{ij}^k represents

   a) the distance from i to j when restricted to paths that traverse at most k edges.

   b) the distance from i to j when restricted to paths that traverse at least k edges.

   c) the distance from i to j when restricted to paths whose intermediate vertices must lie within the set $\{1, \ldots, k\}$.

   d) the distance from i to j when restricted to paths whose intermediate vertices must lie within the set $\{k + 1, \ldots, n\}$.

31. The problem of determining if a bipartite graph has a perfect matching is most naturally reduced to the problem of

   a) finding a Hamilton cycle in a simple graph.

   b) finding a vertex cover in a simple graph.

   c) finding the maximum flow in a directed network.

   d) finding the longest path in a directed network.

32. The problem of determining the maximum flow in a directed network is most naturally reduced to the problem of

   a) determining if a 0-1 Knapsack has a profit that exceeds some threshold.

   b) finding the longest path in a directed network.

   c) determining if there is a path from one vertex to another in a graph.

   d) finding the longest path in an acyclic directed network.

33. Professor Jones has just proved that there is no polynomial-time algorithm for solving the 0-1 Knapsack optimization problem. Because of this we also know there is no polynomial-time algorithm for solving each of the following problems, except for

   a) determining if a Boolean formula has a satisfying assignment.

   b) determining if a graph has a Hamilton cycle.

   c) determining if a set of integers has a subset that sums to some value t.

   d) determining the longest path in a directed acyclic graph.

34. The Subset Sum decision problem is most closely associated with

   a) the Set Partition decision problem.

   b) the 3SAT decision problem.

   c) the Hamilton Cycle decision problem.

   d) the Vertex Cover decision problem.

Exam Solutions for Data Structures and Algorithms

1. b
2. b
3. b
4. c
5. b
6. a
7. b
8. a
9. d
10. d
11. b
12. c
13. c
14. c
15. a
16. a
17. c
18. b
19. a
20. d
21. c
22. b
23. a
24. b
25. c
26. d
27. c

28. b

29. d

30. c

31. c

32. c

33. d

34. a