Computer Science 477/677 Spring 2008 Practice Final Examination

Name:_____

No books, notes, scratch paper, or calculators. Use pen or pencil, any color. Use the rest of this page and the backs of the pages for scratch paper. If you need more scratch paper, it will be provided.

The entire test is 495 points.

1. True or False. [5 points each]

   (a) _____ Huffman's algorithm is a greedy algorithm.

   (b) _____ There are situations, which can arise in practice, where the most efficient search structure is unordered list with linear search.

   (c) _____ A heap is a particular form of a search structure.

   (d) _____ A stack is a particular form of a priority queue.

   (e) _____ If I visit the nodes of a binary tree in preorder, it is just as if I started at the root and used depth-first search and always chose the leftmost path when there was a choice.

   (f) _____ If a 2-3 tree has n data items, we can be sure that the height of the tree is $O(\log n)$.

   (g) _____ There are practical applications in which *unordered list*, together with *linear search*, is the most efficient implementation of a search structure.

2. Fill in the blanks. [5 points each blank.]

   (a) _____ and _____ are two examples of divide-and-conquer sorting algorithms we have discussed in class.

   (b) The *head* of a linked list is the first node, which contains a link to the second node, *etc.* Let's call the other end of the linked list the "foot." When a stack is implemented as a linked list, the top of the stack is located at the _____ of the linked list. When a queue is implemented as a linked list (not a circular linked list) the front of the queue is located at the _____ of the linked list and the rear of the queue is located at the _____ of the linked list.

   (c) The time to execute polyphase mergesort on n items is $O($_____$)$ and is also $\Omega($_____$)$. These are different since we don't know how many runs there are.

   (d) A connected planar graph of 10 nodes must have at least _____ edges. However, it cannot have more than _____ edges. (Exact answers please. No credit for asymtotic answers.)

   (e) You should use a stack to implement depth-first search in a graph, but you should use a _____ to implement breadth-first search.

   (f) Any comparison-based algorithm for sorting n items must take _____ steps to execute in the worst case. (The answer should be in terms of "Ω.")

3. [5 points] What is the **only** difference between the abstract data types *queue* and *stack*?

4. Solve the recurrences. Give asymptotic answers in terms of n, using either O, Ω, or Θ, whichever is most appropriate. [5 points each.]

   (a) $f(n) = 2f(n/2) + n$.

   (b) $f(n) \geq f(n/2) + 2n^2$.

   (c) $f(n) \leq f(n-1) + n^2$.

   (d) $F(n) = 4F\left(\frac{n}{2}\right) + 5n^2$

5. [15 points] Consider the following procedure:

```
void george(int n)
 {
   int m = n;
   while (m > 1)
    {
      for (int i = 1; i < m; i++)
       cout << "I cannot tell a lie.  I chopped down the cherry tree." << endl;
      m = m/2;
    }
 }
```

   Consider the question of how many lines of output execution of `george(n)` would produce. Write down an appropriate recurrence for this question, and give an asymptotic solution in terms of `n`, using either O, Ω, or Θ, whichever is most appropriate.

6. [10 points] What is a loop invariant?

7. [15 points] A stack is implemented using an array `A` and a variable `n` which indicates the number of items currently in the stack. The items are stored in positions `A[0]` ... `A[n-1]`. Write code, or pseudo-code, which pushes a new item, `x`, onto the stack. Do not check for overflow. You should have only **two** lines of pseudo-code.

8. [15 points] Give a **mathematically correct** definition of the statement, "$f(n) = O(n^2)$." (If you write more than 15 words, your answer is probably wrong. I will take off points if you give an example, or write anything else that is unnecessary.)

9. [20 points] Sort the functions given below in order of increasing asymptotic complexity. If two or more functions have the same asymptotic complexity, indicate that.

   n, $5n + \sqrt{n}$, n^{100}, $\log_2 n$, $n^{0.001}$, 1.001^n, $\log_2 \binom{2n}{n}$

   We use the notation $\binom{i}{j}$ to denote the combinatorial $C(i,j)$ (" i choose j"), defined to be $\frac{i!}{j!(i-j)!}$.

10. [20 points] Describe an efficient implementation of a data structure which represents a set of real numbers, with the following operators:

    (a) Make the set empty, *i.e.*, delete all numbers from the set.

    (b) Test to see whether the set is empty.

    (c) Test to see whether the set has at least two elements.

    (d) Insert any given real number into the set.

    (e) Return the minimum number in the set; do not delete it.

    (f) Return the second smallest number in the set; do not delete it.

11. Fill in each blank. 5 points each blank. In each case, the correct answer is either *stack*, *queue*, or *heap*.

    (a) An algorithm which evaluates arithmetic expressions which involved digits, operators, and parentheses, such as "$2 + 2$" or "$3 * (-5 + 14 * 7)$," and which is allowed to read the input stream only once, should use a _____.

    (b) You would use a _____ for an algorithm to find the shortest distance (least number of edges) from a node to a given other node in a graph.

    (c) An operating system, running a program with subprogram calls, including recursive calls, holds the uncompleted subprograms on a _____.

12. [20 points] What is the worst-case asymptotic expected time complexity of randomized quicksort, *i.e.*, where each pivot item is randomly selected from the subarray currently being sorted?

13. [20 points] Insert the following items into a **treap**, in the order shown. Assume that your random number generator assigns the value shown in the table to each item. Show the treap after each insertion. (Thus, you should have 6 figures, although you may draw intermediate figures to do your work.)

A	0.5
G	0.1
M	0.6
Q	0.3
S	0.4
T	0.2

14. Consider the following array representing a weighted directed graph G.

$$\begin{bmatrix} \infty & 2 & \infty & \infty \\ 5 & \infty & 3 & \infty \\ \infty & \infty & 1 & 0 \\ 1 & \infty & \infty & \infty \end{bmatrix}$$

(a) Draw a picture of G. [5 points]

(b) Use matrix operations to compute the reflexive transitive closure G^*, showing your work. [15 points]

15. [20 points] The purpose of the following code fragment is to print out a number which is equal to twice the value of n. Write the loop invariant for the loop. Assume that $n \geq 0$. (Hint: the correct answer is a very short equation.)

```
int m = n;
int p = 0;
while (m > 0){
  m = m - 1;
  p = p + 2;
 }
cout << p << endl;
```

16. [20 points] Use dynamic programming to solve the single source least cost path problem for the acyclic directed graph shown in Figure 1, where the source is s. Your answer should consist of two arrays: one showing the cost of each least cost path, the other the back pointers.

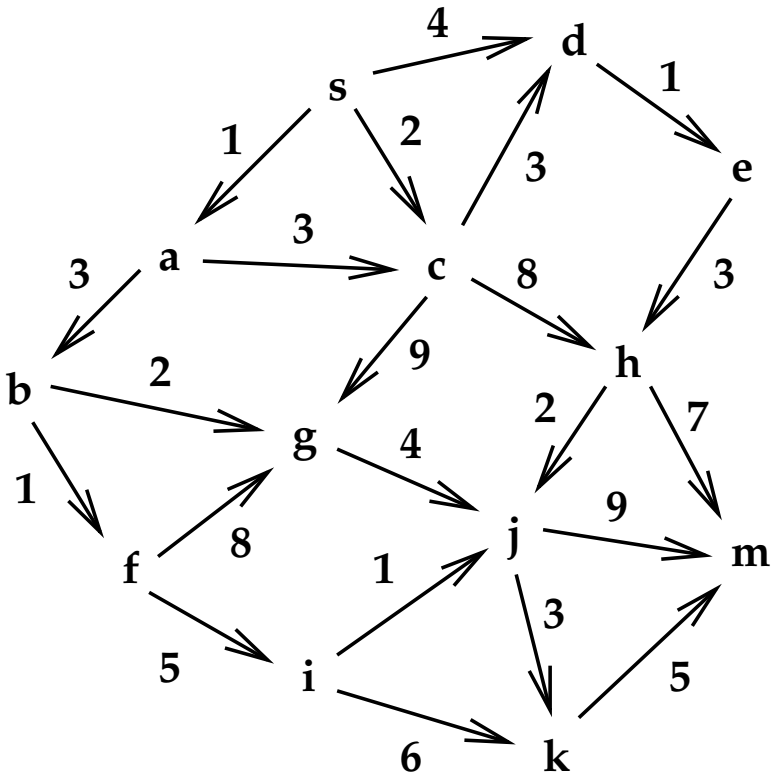


Figure 1: A Weighted Directed Acyclic Graph

17. [5 points each] Each part below describes a sorting algorithm that we studied this semester. Name each algorithm.

    (a) _____ An algorithm that sorts n items in an array, using $O(1)$ work space, which is comparison-based and takes $O(n \log n)$ time. It is not a divide-and-conquer algorithm.

    (b) _____ A divide-and-conquer comparison-based algorithm that requires $O(\log n)$ work space, and has very fast expected time complexity for randomly sorted data, but takes quadratic time in the worst case.

    (c) _____ A comparison-based algorithm that requires linear work space. If the data are already largely sorted, it will take less time. I recommend this algorithm for sorting external files that are too large to store in RAM.

    (d) _____ A comparison-based algorithm that requires $O(1)$ work space and runs in quadratic time. This algorithm has been independently invented by thousands, perhaps even millions, of programmers. It takes $O(n^2)$ time despite the fact that the naive think it's efficient. Its only major advantage is that it requires comparison only between adjacent items, which makes it suitable for certain distributed applications.

    (e) _____ A sorting algorithm that is not comparison-based. It works well primarily in cases where the range of data values is known in advance. It requires large work space (the actual amount depending on the implementation). This algorithm would be efficient in sorting a large set of personnel records using social security numbers as the key.

18. [20 points] Explain how you would use a search structure to implement a sparse array. Your explanation should include the following:

    (a) How array entries are represented.

    (b) How you would set all items in the array to the default value. assume that the default value is zero.)

    (c) How you would implement Store.

    (d) How you would implement Fetch.

    If you write more than 20 words for any of the above three explanations, you are probably not answering the question correctly.

19. [20 points] Explain the difference between a hash table that uses chaining and a hash table that uses probing.

20. [10 points] Draw a sorting network that sorts three items, using three comparators.

21. [20 points] The bit-matrix below represents a directed graph G. How many strong components does G have? Show your work. (You do not need to use the algorithm for finding strong components that we discussed in class.)

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

22. [20 points] Starting with an empty tree, the following six items are inserted into an AVL tree, in this order: `B,A,D,C,F,E`. show the tree at each step, and show the rebalancing steps, if needed.

23. [20 points] Use Dijkstra's algorithm to find solve the single-source minimum path problem on a weighted directed graph with cycles.

24. [15 points] Fill in each blank with the best possible answer from this list: $O(1)$, $O(\log n)$, $O(\log^2 n)$, $O(n)$, $O(n \log n)$.

    You are given an array of n numbers. Using only comparison-based algorithms:

    (a) You can find the median value in _____ time.

    (b) If the array is sorted, you can find the median value in _____ time.

    (c) Using a randomized algorithm, you can find the median value in _____ expected time.

25. [20 points] (This is the hard problem of the test.) A three dimensional $10 \times 8 \times 6$ array `X` is stored in row-major order. The indices range from 0 to 9, 0 to 7, and 0 to 5, respectively. Each item in the array occupies two addressable memory locations. The first item in the array, `X[0,0,0]`, starts at address 1024. Suppose our program does not have bounds-checking, and we execute the following:

```
for i from 0 to 9
 for j from 0 to 7
  for k from 0 to 5
   X[i,j,k] = i+j+k;
write X[4,9,8];
```

What will be the output? Show your work. (You may use the next page for work.)