

COMPUTER SCIENCE 112 - Fall 2013 FINAL EXAM

Name : _____

CIRCLE your lecture: VENUGOPAL TJANG

- Be sure your exam has 6 questions. DO NOT TEAR OFF THE SCRATCH PAGES OR REMOVE THE STAPLE.
- Be sure to fill your name and circle your lecture above, and your name in all subsequent pages where indicated.
- This is a CLOSED TEXT and CLOSED NOTES exam. You MAY NOT use calculators, cellphones, or any other electronic device during the exam.

Do not write below this line

Problem -----	Value -----	Score -----
1 Medley	10	_____
2 Heap Update Priority	20	_____
3 Graph Connected Components	20	_____
4 Dijkstra's Algorithm	25	_____
5 Sorting	25	_____
6 Movie Streaming Website	25	_____
TOTAL	125	_____

1. Medley (10 pts, 3+3+4)

a) Consider sequential (linear) search on the following list:

10, 5, 8, -9, 16, 25, 88

Suppose the first 4 items are searched with a probability of 0.1 each, and the last three items are searched with a probability of 0.2 each. What is the average number of target-to-item comparisons for successful search? Circle the correct choice. (1 pt)

- A. 4.0
- B. 4.6
- C. 3.5
- D. None of the above

Show your derivation (2 pts):

b) Starting with an empty AVL tree, items are inserted one at a time. Which of the following sequence of insertions would result in rebalancing with rotation(s)? (1 pt)

- A. 5, 3, 7, 2, 6, 8
- B. 7, 5, 8, 6, 3, 2
- C. 20, 15, 35, 45, 10, 25
- D. 17, 13, 21, 19, 7, 25

Show the final AVL tree with balance factors for all nodes (2 pts):

198:112 Fall 2013 Final Exam; Name: _____

c) A max heap has 17 nodes.

i) Show the binary tree heap structure with integer keys at the nodes (2 pts)

ii) An item is inserted into this max heap and sifted up. Construct a worst-case scenario so that the maximum possible amount of work is done during sift up: show the key that is inserted, where it's inserted, and show the final heap structure after sift up is completed. (2 pts)

2. Heap Update Priority (20 pts, 14+6)

Suppose you are given the following heap class (assume that the constructor and all methods are already implemented, only public members are shown):

```
public class Heap <T extends Comparable<T>> {
    ...
    public Heap() { ... }
    public insert(T item) { ... }
    public T deleteMax() throws NoSuchElementException { ... }
    public boolean isEmpty() { ... }
    public int size() { ... }
    ...
}
```

Suppose an application (client) uses the Heap class to store HeapItem objects:

```
public class HeapItem implements Comparable<HeapItem> {
    public String id;    // unique
    public int priority;
    public HeapItem(String id, int priority) {
        this.id = id; this.priority = priority;
    }
    public int compareTo(HeapItem h) {
        return priority - h.priority;
    }
}
```

a) Implement a client method (i.e. in the application) that will update the priority of an item in a heap. Note: your code may **ONLY** use the given public constructor and methods of the Heap class. You may use, additionally, an ArrayList if you need, but no other data structure. You are free to implement helper methods as necessary.

```
// update the priority of the item in the heap with the given id,
// to the new priority - throws exception if item is not found
public static void updatePriority(Heap<HeapItem> heap, String id, int newPriority)
throws NoSuchElementException {
```

198:112 Fall 2013 Final Exam; Name: _____

b) Analyze the worst case big O running time of your implementation (including helper methods if any). Show your derivation. Count number of operations at the level of the method calls on the data structures, and multiply by the running times of those methods. Assume that all listed methods of the `Heap` class run in worst case $O(1)$ time.

3. Graph Connected Components (20 pts)

You are given an undirected, unweighted graph that may be disconnected i.e. some vertices may not be reachable from other vertices. Every group of mutually reachable vertices forms an island, called a *connected component*. There is no path between vertices in different connected components. If a graph is not disconnected, then its vertices are in a single connected component, which is the entire graph itself.

Implement a method using depth-first search that will number each vertex with the connected component that it belongs to. If there is a single connected component, then all vertices will be numbered 0 (zero). If there are two islands, then vertices in one island will all be numbered 0, and those in the other will all be numbered 1. And so forth.

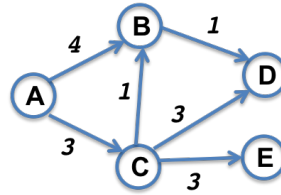
Write your implementation in the `components` method in the following `Graph` class. Assume that when this method is called, the `adjLists` structure is already populated. You can implement helper methods as necessary.

```
class Neighbor {
    int vnum;
    Neighbor next;
}

public class Graph { // undirected, unweighted
    Neighbor[] adjLists;
    ...
    // returns an array with connected component numbers for all vertices
    // i.e. returned[i] is the component number for vertex i
    public int[] components() {
        /* COMPLETE THIS METHOD */
    }
}
```

198:112 Fall 2013 Final Exam; Name: _____

4. Dijkstra's Algorithm (25 pts;18+7)



a) Dijkstra's shortest paths algorithm is executed on the graph above, starting at A. Assume that vertices A,B,C,D,E are given numbers 0,1,2,3,4 respectively in the implementation. Also assume that neighbors of a vertex are stored in alphabetical order of their names. Trace the execution of the algorithm as follows: at the end of every step, show the distance array and the fringe in the table below. The fringe is stored in a min-heap, in which distance updates can be made, apart from delete min, and insert. To show the fringe, draw the binary tree heap structure, with (vertex name,distance) information at each node. Every time there is a change to the heap, show the number of item-to-item comparisons needed to make that change, and also what operation (insert/delete/update) resulted in that change. (Ignore the time needed to *locate* an item in the heap for a distance update.)

Step	Distance array	Fringe heap	Comparisons (number, for what operation)

198:112 Fall 2013 Final Exam; Name: _____

b) The worst case running time of Dijkstra's algorithm is $O((n + e) \log n)$, on a graph with n vertices and e edges, provided the fringe is stored in a min-heap with $O(\log n)$ insert, delete min, and update operations, and the graph is stored in adjacency linked lists. Assume that the graph is stored in an adjacency matrix instead, without changing any other data structure. State clearly which step of the algorithm would be impacted by this change, describe how it would be impacted, and derive the new running time. You don't have to derive the time for the other, non-impacted steps, but you must state all the steps, and write down their big O running times, then sum to single big O .

5. Sorting (25 pts, 4+5+8+8)

a) You are given an array with **100,000** entries to sort. Under what circumstances would you use each of the following: (i) insertion sort, (ii) mergesort, and (iii) quicksort?

b) A list of five characters, 'a', 'b', 'c', 'd', and 'e', is stored in an array. Give a starting arrangement of these characters that would result in mergesort making the *most* number of comparisons to sort the list. Show the mergesort recursion tree, as well number of comparisons for each merge, otherwise you will not get any credit. Also, if your answer is not the worst case, you will NOT get ANY credit, even if the process is correctly described. (Assume that a split on an odd number of entries results in one extra entry in the left half.)

198:112 Fall 2013 Final Exam; Name: _____

Parts (c) and (d) are based on the following:

This year, thousands of students took the SAT (Scholastic Aptitude Test) in New Jersey in preparation to enter college. The total score on the SAT is out of 2400—scores are in multiples of 10. You are asked by the SAT office to write a program that will sort the students in descending values of their scores. If two students have the same score, it does not matter in which order they appear in the sorted result. Assume that the computer on which you will run your program has enough memory to store the names and scores of all the students, either in an array or in a linked list. Also, for this question, you can assume the number of students who took the SAT this year is exactly 65,536, which is 2^{16} .

c) Assuming you were to use mergesort with linked lists, and counting **ONLY** every score-to-score comparison as a unit time operation, how many such unit time operations (not big O) would it take in the worst case to do the sort? Your answer must be simplified to the least possible number of terms.

d) Come up with an algorithm that is better than mergesort above, i.e. it makes fewer unit time operations in the worst case. State what data structure(s) you will use, and clearly describe your algorithm. Then specify what unit time operations you are counting, and determine the total number of such units. You will NOT get any credit if your algorithm is not better than mergesort for this application, even if it works, and you have analyzed it correctly.

198:112 Fall 2013 Final Exam; Name: _____

SCRATCH PAGE

6. Movie Streaming Website (25 pts, 5+8+12)

Credit is for correctness as well as efficiency.

A website streams movies to customers' TVs or other devices. Movies are in one of several genres such as action, drama, mystery, etc. Every movie is in exactly one genre (so that if a movie is an action movie as well as a comedy, it is in a genre called "action-comedy"). The site has around 10 million customers, and around 25,000 movies, but both are growing rapidly.

The site wants to keep track of the most popular movies streamed. You have been hired as the lead engineer to develop a tracking program.

i) Every time a movie is streamed to a customer, its name (e.g. "A Very Harold & Kumar 3D Christmas") and genre ("Comedy") is sent to your program, which then updates the data structures it maintains. (Assume your program can get the current year with a call to an appropriate Java class, in $O(1)$ time.)

ii) Customers want to know the top k most streamed movies in genre g in year y . (If y is the current year, then accounting is done up to the current date of the year.) For example, what were the top 10 most streamed comedy movies in 2013? Here $k=10$, $g="comedy"$ and $y=2013$. This query is sent to your program which should output the top k movie names, in descending order of number of times streamed.

a) Describe the data structures used to implement the above.

b) For (i), describe the algorithm, step by step, to update the data structures. Analyze its big O running time. Show your work. Specify whether the running time is worst case or something else.

198:112 Fall 2013 Final Exam; Name: _____

c) For (ii), describe the algorithm, step by step, to output the top k streamed movies in genre g in year y . Describe additional data structures, if any, used in this process. Analyze its big O running time. Show your work. Specify whether the running time is worst case or something else.

SCRATCH PAGE

198:112 Fall 2013 Final Exam; Name: _____

SCRATCH PAGE

SCRATCH PAGE