Final Exam **Comp Sci 2C03** - *Data structures and Algorithms* *Winter 2021*

Name __________________________________ Student No.__________________________________

For your exam you are ONLY allowed to refer to the lecture notes, text book, and tutorials.

Time allotted: 3 hours (write and submit PDF) **Total Marks: 33**

Multiple choice questions [10 marks]

Circle your answer. Each question has just one correct answer. Therefore multiple selections will not get a mark.

1. Suppose `x` is a linked-list node and not the last node/tail in the list. What is the effect of the following code fragment: `x.prev.next = x.next`?

   (a) Deletes from the list the node immediately following `x`.

   (b) Deletes from the list the node immediately preceding `x`.

   (c) Deletes from the list the node `x`.

   (d) None of the above.

2. Insertion sort is the sorting algorithm of choice for small (< 10) arrays.

   (a) True

   (b) False

3. Suppose that your application will have a huge number of find the maximum operations, but a relatively small number of insert and remove the maximum operations. Which priority-queue implementation do you think would be most effective:

   (a) heap

   (b) unordered array

   (c) ordered array

   (d) None of the above

4. What is the sequence of nodes printed during a preorder traversal of the tree given in Figure 1?

   (a) A C E H I N R S

   (b) E A S C I H R N

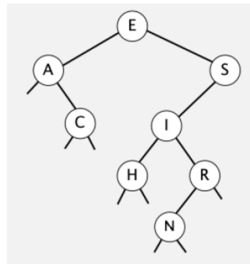

Figure 1: Trees for multiple choice Question 4.

(c) <u>E A C S I H R N</u>

(d) S E A R C H I N

5. Which insertion order for the keys S E A R C H I T leads to a 2-3 tree of height 1.

   (a) A I R E H C T S

   (b) A E T C H S R I

   (c) A E T C R S H I

   (d) <u>None of the above</u>

6. Hashing/hash table is not suited for ordered symbol-table operations.

   (a) <u>True</u>

   (b) False

7. An edge weight undirected graph with all distinct weights has a unique MST.

   (a) <u>True</u>

   (b) False

8. Adding a constant to every edge weight does not change the solution to the single-source shortest-paths problem.

   (a) True

   (b) <u>False</u>

9. To sort one million 32-bit integers. Which sorting method is the best to use?

   (a) Insertion sort

   (b) Meargesort

   (c) Quicksort

   (d) <u>LSD string sort</u>

10. Consider the `text = c c c c c c c c c`, and the pattern `pattern = c c c`. Let BMC = no.of comparisons done by Boyer-Moore (with only bad character rule), RKC= no.of comparisons done by Rabin-Karp, and KMPC = no.of comparisons done by KMP. Which of the below statement is correct?

    (a) KMP > BMC and KMP < RKC

    (b) KMP < BMC and KMP > RKC

    (c) <u>KMP < BMC and KMP < RKC</u>

    (d) KMP > BMC and KMP > RKC

Provide detailed answers to the 6 questions below

1. (a) Using ONLY the definition of $\cancel{O(f(n))}$ $\Theta(f(n))$ prove that for $T(n) = n^2 \log_2 n + 2n + 1$, $T(n) \in \Theta(n^2 \log_2 n)$. Your proofs using Limits will not get a mark. [2 marks]
   Answer: choose $c_1 = 1, c_2 = 4, n_0 = 2$.

   (b) Using ONLY the definition of $\Theta(f(n))$ prove that for $T(n) = n^3 + 5n + 10$, $T(n) \notin \Theta(n^2)$. Your proofs using Limits will not get a mark. [2 marks]
   Answer: The proof is by contradiction. Since for any function $f(n) \in \Theta(g(n))$; $f(n) \in O(g(n))$ and $f(n) \in \Omega(g(n))$. Then, by assumption $T(n)$ must be in $O(n^2)$. Let c, n_0 be the least constants, such that the below inequality holds for all $n \geq n_0$

$$n^3 + 5n + 10 \leq c \cdot n^2.$$

   However, for all $n \geq c + 1$, $n^3 + 5n + 10 > c \cdot n^2$ – a contradiction. Therefore, $T(n) \notin \Theta(n^2)$.

2. About how many compares will Quick.sort() make when sorting an array of N items with just two distinct keys. Your solution in big-Oh notation will suffice? [3 marks]
   Answer: Let the array consists of two distinct keys $a, b > a$. In the first call to the partition function, the input array is divided into two subarray such the first array consists of all as and the second subarray consists of all bs. Since the sub arrays have equal elements, quicksort has the worst case performance. Hence resulting in at most $O(n^2)$ compares.

3. Compute the MST of the undirected edge-weighted graph shown in the Figure 2 using Kruskal's Algorithm. You solution must contain the order in which edges are added to the MST. [3 marks]

   Answer: We order the edges by the ascending order of their weights. Below are the edges in an ascending order:
   0-1 1
   0-3 2
   1-3 2

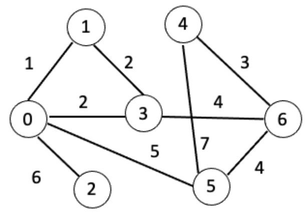

Figure 2: Undirected weighted edge graph

4-6 3
3-6 4
5-6 4
0-5 5
0-2 6
4-5 7

We add edges, 0-1, 0-3. We don't add the 1-3 as it creates the cycle 0-1-3-0. Then we add 4-6, 3-6, 5-6. We don't add 0-5 as it creates the cycle 0-3-6-5-0. Then we add the edge 0-2. Since, we have added v-1 (7-1=6) edges, we stop. Therefore, 4-5 is not evaluated.

4. How can we find shortest paths in undirected positive edge-weighted graphs? Just provide the outline for your solution. [3 marks]

   Answer:

   (a) Build an edge weighted digraph corresponding to the given edge weighted graph (undirected), by adding two directed edges corresponding to each undirected edge, one in each direction with the same weight.

   (b) Then run Dijkstra's algorithm for each vertex as the source to find shortest path from every vertex to every other vertex.

5. Let w be a nonempty string of length n. An integer p such that $w[i] = w[i + p]$ for all $i = 0, 1, ..., n - p - 1.$ is called a **period** of w.

   (a) What is the relationship between a period p and a border b of a string of length n (just the formula with suffice). [2 marks]
   Answer: $n - |b| = p$.

   (b) How would you use the border array (discussed on slides# 7, 8 of C5P3.pdf) of a string of length n to compute the minimum period of the input string w. The minimum period of a string is its smallest period. For example, the minimum period of the string ABCABCABCABCAB is 3. [2 marks]

Answer: Let p^* be the minimum period of the string. The we can compute it using the border array as follows: $p^* = n - \beta[n]$.

6. Given the input string, `text = a b a a b a a c b c b a a a b`. You may assume that the input string is given in 7-bit ASCII.

   (a) Give the encoded string obtained by applying Huffman encoding. You solution must contain the prefix-free code trie created by the Huffman encoding. [3 marks]
   Answer: $f_a = 8, f_b = 5, f_c = 2$. The prefix-code trie created by the Huffman encoding is given in Figure 3. The prefix code for $a = 1, b = 01$, and $c = 00$. The compressed string is `= 1 01 1 1 01 1 1 00 01 00 01 1 1 1 01`. The encoding of the trie is `= 0 0 1 1100011 1 1100010 1 1100001`. The encoded string is `= 0 0 1 1100011 1 1100010 1 1100001 1 01 1 1 01 1 1 00 01 00 01 1 1 1 01`.

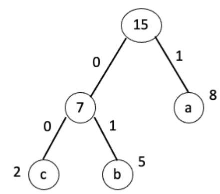

Figure 3: Prefix-code trie created by the Huffman encoding.

   (b) Give the encoded string obtained by applying the LZW compression algorithm. You may assume that the output is in 8-bit codewords in hexadecimal. Your solution must contain the symbol table and the entries added to it during the encoding process. [3 marks]

   Answer: The encoded string and the symbol table generated during the LZW compression algorithm is given in Figure 4.

Text:	a	b	a	ab	aa	c	b	cb	aa	ab	
Encoded String:	61	62	61	81	83	63	62	86	83	81	80

61	**a**
62	b
63	c
....	
81	ab
82	ba
83	aa
84	aba
85	aac
86	cb
87	bc
88	cba
89	aaa

Figure 4: Encoded string and the symbol table generated during the LZW compression algorithm