Computer Science C73 — September 21, 2022
Scarborough Campus — University of Toronto

Homework Assignment #3
Due: September 28, 2022, by 11:59 pm

• ***You must submit your assignment through the Crowdmark system.*** *You will receive by email an invitation through which you can submit your work in the form of separate PDF docucments with your answers to each question of the assignment. To work with a partner, you and your partner must form a group on Crowdmark. Crowdmark does not enforce a limit on the size of groups.* ***The course policy that limits the size of each group to at most two remains in effect:*** *submissions by groups of more than two persons will not be graded.*
• *It is your responsibility to ensure that the PDF files you submit are legible. To this end, I encourage you to learn and use the LaTex typesetting system, which is designed to produce high-quality documents that contain mathematical notation. You are not required to produce the PDF files you submit using LaTex; you may produce it any way you wish, as long as the resulting document is legible.*
• *By virtue of submitting this assignment you (and your partner, if you have one) acknowledge that you are aware of the policy on homework collaboration for this course.*[a]
• *For any question, you may use data structures and algorithms previously described in class, or in prerequisites of this course, without describing them. You may also use any result that we covered in class, or is in the assigned sections of the official course textbooks, by referring to it.*
• *Unless we explicitly state otherwise, you may describe algorithms in high-level pseudocode or point-form English, whichever leads to a clearer and simpler description.*
• *Unless we explicitly state otherwise, you should justify your answers. Your paper will be graded based on* ***the correctness and efficiency of your answers, and the clarity, precision, and conciseness of your presentation****.*

[a] "In each homework assignment you may collaborate with at most one other student who is currently taking CSCC73. If you collaborate with another student on an assignment, you and your partner must submit only one copy of your solution, with both of your names. The solution will be graded in the usual way and both partners will receive the same mark. Collaboration involving more than two students is not allowed. **For help with your homework you may consult only the instructor, TAs, your homework partner (if you have one), your textbook, and your class notes. You may not consult any other source.**"

Question 1. (10 marks) Consider Huffman's algorithm.

a. (2 marks) Give an example of a (small) set of symbols and their associated frequencies so that the maximum frequency of any symbol is equal to 2/5, and Huffman's algorithm could produce a tree in which no codeword has length 1. Show such a tree that could be produced by Huffman's algorithm in your example.

b. (8 marks) Prove that for any set of symbols, if some symbol has frequency (strictly) greater than 2/5, Huffman's algorithm will necessarily produce a codeword of length 1.

Question 2. (10 marks) The CSCC73 midterm test has been graded. Each of the k TAs has sorted his or her pile of exams alphabetically. Each pile contains exactly n papers. The TAs have all gone home, and Vassos is stuck with merging these k sorted piles into a single pile of kn papers, sorted alphabetically.

a. (5 marks) Here is one algorithm that Vassos can use to solve his problem: Merge the first two sorted piles into a (sorted) pile; then merge the resulting pile with the third sorted pile; then merge the resulting pile with the fourth sorted pile, etc. What is the running time of this algorithm in terms of k and n? (Merging two sorted piles of papers can be done in time proportional to the size of the resulting pile.)

b. (5 marks) Give a more efficient divide-and-conquer algorithm that Vassos can use to solve this problem. What is the running time of your algorithm? (You don't need to justify its correctness.)

Question 3. (15 marks) An ***interval*** is a pair (ℓ, r) of numbers such that $\ell \leq r$, or the ***empty interval*** denoted $\varnothing$. The ***intersection*** of two intervals (ℓ, r) and (ℓ', r') is the interval $\big(\max(\ell, \ell'), \min(r, r')\big)$ if $\max(\ell, \ell') \leq \min(r, r')$; otherwise, it is $\varnothing$. The ***length*** of interval (ℓ, r) is $r - \ell$, and the length of $\varnothing$ is 0.

We are given a list of $n \geq 2$ nonempty intervals $(\ell_1, r_1), (\ell_2, r_2), \ldots, (\ell_n, r_n)$. We want to find two distinct intervals in this list whose intersection is as long as possible. It is straightforward to do this in $\Theta(n^2)$ time. Describe a divide-and-conquer algorithm that solves this problem in $O(n \log n)$ time. Explain why your algorithm is correct and analyze its running time.