University of Waterloo
CS240, Fall 2015
Assignment 5

Due Date: Wednesday, December 2, at 5:00pm

Please read `http://www.student.cs.uwaterloo.ca/~cs240/f15/guidelines.pdf` for guidelines on submission. All problems are written problems; submit your solutions electronically as a PDF with file name `a05wp.pdf` using MarkUs. We will also accept individual question files named `a05q1w.pdf`, `a05q2w.pdf`, `a05q3.pdf`, `a05q4w.pdf` and `a05q5w.pdf` if you wish to submit questions as you complete them.

There are 63 possible marks available. The assignment will be marked out of 60.

Problem 1 Range Trees [5+5+5=15 marks]

a) Assume that we have a set of n numbers (not necessarily integers) and we are interested only in the number of points that lie in a range rather than in reporting all of them. Describe how a 1-dimensional range tree (i.e., a balanced binary search tree) can be modified such that a range counting query can be performed in $O(\log n)$ time (independent of k, the number of nodes). Provide the range counting query and justification of its runtime.

b) Now consider the 2-dimensional-case: We have a set of n 2-dimensional points. Given a query rectangle R, we want to find the number of points that lie in R. Preprocess the n points (by building an appropriate range-tree based data structure) such that you can answer any of these counting queries in time $O((\log n)^2)$. Provide the range counting query and justification of its runtime.

c) Suppose a two dimensional range tree data structure stores n points, and that the x-BST is perfect, i.e., every level is completely filled. Give an exact closed form formula in terms of n for the sum of the number of nodes in the x-BST plus the total number of nodes in all y-BSTs. y-BSTs.

Problem 2 Tries [2+2+2+2+2+2=12 marks]

a) Draw the trie on the following five strings (include edge labels for clarity): 0001, 1001, 1011, 010, 1000.

b) Draw the compressed trie on the following five string: 100, 0110, 01110011, 01110101, 01110100

c) Draw the result of inserting 10 into the compressed trie shown on Slide 10 of Module 8.

a	b	c	a	a	b	a	a	b	a	b	a	b	a	c	a	b	c	a	a

Table 1: Table for KMP problem.

d) Draw the result of deleting 01011 fron the compressed trie shown on Slide 10 of Module 8.

e) Draw the result of inserting the two strings `zog` and `b` into the compressed multi-way trie shown on Slide 15 of Module 8.

f) Draw the result of deleting `so` from from the compressed multi-way trie shown on Slide 15 of Module 8.

Problem 3 KMP [6+6+6=18 marks]

a) Compute the failure array for the pattern $P =$`ababac`.

b) Show how to search for pattern $P =$`ababac` in the text $T =$`abcaabaababababacabcaa` using the KMP algorithm. Indicate in a table such as Table 1 which characters of P were compared with which characters of T. Follow the example on slide 25 in module 8. Place each character of P in the column of the compared-to character of T. Put brackets around the character if an actual comparison was not performed. You may not need all space in the table.

c) Consider a pattern P and a text T. Assume that you are given the failure array for the string $P\Phi T$ (the concatenation of P, a character Φ that is not contained in P, and T). Explain how to use this array to find the first occurrence of P in T.

Problem 4 Boyer-Moore [3+7=10 marks]

a) Compute the last-occurence function L for the pattern $P =$`ratatat`. Give your answer as table as shown on Slide 32 of Module 8. Note: $\Sigma = \{a, r, t\}$.

b) Compute the suffix skip array S for the pattern $P =$`ratatat`. Give your answer as table as shown on Slide 33 of Module 8.

Problem 5 Suffix Trees [8 marks]

a) Draw the suffix tree corresponding to the text $T =$`abracadabra`. Use the recipe of Slide 37 of Module 8. Your suffix tree should look like the example on Slide 38. Children of a node should be ordered alphabetically.