School of Computer Science
University of Guelph

CIS*3490 The Analysis and Design of Algorithms

Winter 2022
Instructor: Fangju Wang

Assignment 4 (100%)

Question 1 (70%)

File *data_A4_Q1.txt* contains 2045 word instances. Develop two programs to read the file, create optimal binary search trees (BSTs) for the words, and search the trees. In a tree, create a node for a word, even the word may appear more than once in the file, i.e. may have multiple instances. For example, *undergraduate* appears nine times in the file, but there should be only one node for it. For each word, calculate its probability from its frequency (repetitions) in the file.

1.1 (40%)
Develop a program to create an optimal BST by using the technique of dynamic programming (described in section 8.3 in the textbook). You are required to calculate the "average number table" and "root table". After creating the tree, the program prompts the user to enter a word (key) and searches the tree for the word. Whenever the program compares user input K with word W at node V, it displays word W, the minimum average number of comparisons of the subtree rooted at V (three digits after the decimal point), and the subtree to go (left or right). If the program fails to find the K in the tree, it displays "not found" after comparisons. Please see the guide for suggested output format.

1.2 (30%)
Develop a program to create an optimal BST by using the greedy technique. After creating the tree, the program prompts the user to enter a word (key) and searches the tree for the word. Whenever the program compares user input K with word W at node V, it displays word W, the probability of W (three digits after the decimal point), and the subtree to go (left or right). If the program fails to find the K in the tree, it displays "not found" after comparisons. Please see the guide for suggested output format.

Question 2 (30%)

Develop a program to implement the Gale-Shapley stable marriage algorithm. The program reads in a file storing preference lists, and outputs the solution. File *data_A4_Q2_1.txt* and *data_A4_Q2_2.txt* can be used to develop and test your program. In a data file, the first value is the n, followed by two sets of preference lists. When your program is executed, it prompts a file name, reads in the file, and displays the result as an $n \times n$ matrix. The data file used to grade your program will have the same format, and the n value may be 3 or 4. Please see the guide for the formats of input file and suggested output.

Note: Write your own code for this assignment. NO code from any source is allowed.

Due time: 08:00am, Monday, March 28, 2022. Submit your work as a tar file to Moodle.