CSCI 230 -- PA 4

Sorting II

Feel free to discuss and help each other out but does not imply that you can give away your code or your answers! Make sure to read all instructions before attempting this lab.

You can work with a lab partner and each one must submit the same PDF file (include both names in the submission file). Each person must include a brief statement about your contribution to this assignment.

You must use an appropriate provided template from Canvas and output "Author: Your Name(s)" for all your programs. If you are modifying an existing program, use "Modified by: Your Name(s)".

Exercise 1 – Implement/use Quick Sort from the book and test it on a small list of 10 random integer values first. Collect data regarding number of key comparisons and data moves for a small list of 10 integer values. You must try a sorted list, a descending list, and a random list. In addition, collect data for a random list of 100 values.

Exercise 2 – Download a small data file ***small1k.txt*** containing 1,000 values in random order. In addition, download the large data file ***large100k.txt*** containing 100,000 values in random order. Sort the two data files using Quick-Sort-3 ("median of three" as pivot so you need to modify quick sort from exercise 1). For each value in the file, you must create a <key, value> pair and then sort a list of <key, value> pairs. You will have to sort two different lists of <integer, string> pairs and <string, integer> pairs. For instance, you need to create <1234, "1234"> pair and <"1234", 1234> pair for an input value of 1234 from the file. Output all relevant information below for each input data file. There should be 4 sets of output for the two data files.

1. Input file name
2. Number of values sorted and key data type
3. Number of Key Compares and Data Moves
4. The time in milliseconds
5. The first 5 entries and last 5 entries (<key, value> pair)

Question 1: Do the counts of key comparisons between 10 random values and 100 random values seem reasonable for both quick sort? Explain why or why not.

Question 2: Do the collected run times between small data file and large data file seem

reasonable? Explain why or why not. Compare the collected run times between integer key and string key for large data file.

Extra Credit Option A: Perform exercise 2 and collect data for either merge sort, heap sort, or Shell sort (pick one).

or

Extra Credit Option B: Implement bucket sort algorithm in the book and test it on the large data file of 100k values (integer keys in the range [0, 99999]). Output the time in milliseconds, the first 5 values, and last values. You might want to compare to the collected run time for quick sort in exercise 2 for large data file with integer key.

Fill out and turn in the PA submission file for this assignment (save as PDF format)