THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

COMPSCI.220.C.S1 – Algorithms and Data Structures

Assignment 2 – Data sorting efficiency

Out: Monday, 24$^{\mathbf{th}}$ of March, 2016

Due: Friday, 8$^{\mathbf{th}}$ of April, 2016

Assignment 2 (Lectures 7 – 12) is worth **70 marks** representing **7%** of your total course grade.

Objectives. Learning how to analyse efficiency of data sorting and selection techniques.

Requirements. You should give clear and detailed answers to the following questions:

1. (*10 marks*) Given a list of seven single-digit integers of your UID (if the UID is longer, keep its first seven digits), describe ordering the list with `insertion sort` in a table, similar by layout to Table 2.1 (Textbook). Your table should detail all six steps of sorting, including numbers of comparisons, C_i, and backward moves, M_i at each step i of sorting, as well as show the list after each step, $i = 1, 2, \ldots, 6$.

2. (*10 marks*) Using the math induction, prove that the maximum and average numbers of inversions in a list $[a_1, \ldots, a_n]$ of integers a_i are $I_{\max:n} = \frac{1}{2}(n-1)n$ and $I_{\text{ave}:n} = \frac{1}{4}(n-1)n$, respectively. The base case $I_{\max:2} = 1$ and $I_{\text{ave}:2} = 0.5$ is for $n = 2$.

3. (*10 marks*) Determine the number of inversions in your list of seven single-digit integers of Question 1 and compare this number with the average number $I_{\text{ave}:7}$ in Question 2 and with the total number of data comparisons, $C = \sum_{i=1}^{6} C_i$, and moves, $M = \sum_{i=1}^{6} M_i$, performed by `insertion sort` in Question 1.

4. (*10 marks*) You know two strategies for selecting k top-rank achievers (by GPA) in an unordered list of GPA records for n university students:

   (a) Run k times `quickselect` with linear processing time, $T_{\text{qselect}}(n) = cn$, in order to sequentially fetch the students of the rank n, $n-1$, and so on up to $n-k+1$, or

   (b) Run once `quicksort` with linearithmic processing time, $T_{\text{qsort}}(n) = cn \log_2 n$, to sort the entire list in ascending GPA order and fetch the k top-rank students.

   Let both the algorithms have the same factor c and let data fetching take negligibly small time, comparing to sorting. Find out, in terms of k and n, when Option 4a is faster than Option 4b, and determine, which option should be used if $k = 10$ and $n = 32,768 \equiv 2^{15}$.

5. (*10 marks*) The first stage of `quicksort` of a list of size n places the selected pivot to its proper sorted position i; $0 \le i \le n-1$. How many inversions between the pivot in its initial position and all other $n-1$ list elements may exist in the average case and in the worst case and how many of them will be eliminated after this stage in each case?

6. (*10 marks*) Show successive steps of `heapsort` for your list of seven single-digit integers of Question 1 in a table, similar by layout to Table 2.6 (Textbook). Note that `percolateDown` algorithm of Lemma 2.31 is used at the initial stage of `heapsort` to build the max-heap.

7. (*10 marks*) A heap of size n can be built either in linearithmic time, $\mathrm{O}(n \log n)$, by inserting n items one-by-one into an initially empty heap and restoring the heap property after each insertion; see Lemma 2.25 (Textbook), or faster – in linear time, $\Theta(n)$, – by heapifying directly an entire initial list of n items; see Lemma 2.31 (Textbook). Explain, which unneeded by `heapsort` property makes the former process slower than the latter.

Submission: Your report should answer in detail to Questions 1–7. The report should be submitted as a single Adobe PDF file `CS220assign2.pdf` (only this file will be marked, so check that it can be read by PCs in the departmental Computer Labs). **Scanned handwritten documents are forbidden** (even as images in a pdf file) and will not be accepted for marking.

Submit your file electronically to Assignment Drop Box (`https://adb.auckland.ac.nz`) (**not to Canvas!**) before Friday, 8$^{\text{th}}$ of April 2016, 09:00 pm (ADB time). If submitted after this due date, the penalty of 10% will be before 9$^{\text{th}}$ of April 2016, 09:00 p.m.; then the penalty of 50% will be before 10$^{\text{th}}$ of April 2016, 09:00 p.m., and no submission afterwards.

Marking scheme	% of marks
Clear structure of your report with detailed explanations	up to 20
Correctness of your final answers	up to 20
Correctness of your intermediate steps in deriving these answers	up to 20
Detailed explanations of all steps with references, if necessary, to the textbook	up to 40
	Total: up to 100