Assignment 3

CSE 332: Data Abstractions, Spring 2011
University of Washington
April 20, 2011
due: Monday, May 2, 2:15 p.m.

Instructions: Create a PDF representation of your answers and submit it via Catalyst CollectIt. You might wish to prepare the file using LaTeX. You may use the source file of this assignment as a starting point if you use LaTeX. The "pdflatex" command installed on most Linux systems is a convenient way to translate the LaTeX source file into a PDF document. You may prefer to draw the diagrams by hand and scan them to include in your document. The LaTeX source file shows how to include an image, such as a PNG, GIF, or JPG image, in your document.

Be sure that your name is clearly visible at or near the top of the first page. The LaTeX source file also shows where you can put that. The due time for this assignment is 2:15 PM, which means that if you turn it in at the last minute (but this is NOT recommended) from somewhere on campus, you'll still have time to get to class.

1. (25 points) A "minimal AVL tree" $T_{\mathrm{m}h}$ for a particular value of h is an AVL tree of height h that has as few nodes as possible. From the point of view of runtimes for FIND, INSERT, and DELETE, it's a worst-case tree – as unbalanced as possible.

   (a) (10 points) Draw the minimal AVL tree $T_{\mathrm{m}5}$ of height 5 in which each internal node is right-heavy.

   (b) (1 points) How many nodes S_5 does this tree contain?

   (c) (1 point) How many nodes P_5 does the "maximal" AVL tree $T_{\mathrm{M}5}$ of height 5 contain?

   (d) (4 points) What is the ratio of S_5 to P_5? (a simple fraction is fine.) How would you expect the ratio of S_n to P_n to change as h increases? (Justify your answer.)

   (e) (4 points) What is the average node depth $a5$ in $T_{\mathrm{m}5}$?

   (f) (3 points) What is the average node depth $b5$ in $T_{\mathrm{M}5}$?

   (g) (2 points) What is the ratio of $a5$ to $b5$? (Give the answer to 4 decimal places.)

2. (35 points) Consider the following sequence of (key, value) pairs, where each word is a string of characters (letters).

```
(cat, Garfield), (dog, Fido), (elephant, Dumbo), (bird, Tweety),
(mouse, Mickey), (cow, Elsie), (pig, Porky), (lion, Leo),
(rabbit, Peter), (bear, Pooh)
```

(a) (15 points) With an initially empty B-tree having parameters $M = 3$, and $L = 3$, insert the keys of the above pairs (ignore the values in this exercise). Show the tree immediately after each restructuring (change in the number of nodes).

(b) (10 points) Create an empty hash table of size $M = 11$. Define the function f on lower-case letters of the alphabet as follows: $f(\text{'a'}) = 0, f(\text{'b'}) = 1, \ldots, f(\text{'z'}) = 25$. Define a hash function h on strings as follows:

$$h(c_1 c_2 \cdots c_k) = (f(c_1) + f(c_k)) \bmod M$$

Assume that collisions will be resolved using quadratic probing. Insert the (key, value) pairs into the hash table. Give a drawing or other illustration of the final table, showing both the keys and the values. What is the load factor for this table at the end of the insertion sequence?

(c) (10 points) Create an empty hash table of size $M = 4$ that uses the same hashing function as in the previous hash table, but it resolves collisions by maintaining a separate linked list to represent all the items for a bucket of the hash table. Insert the (key, value) pairs into this hash table. What is the load factor for this table at the end of the insertion sequence?

3. (40 points) Consider the (out-of-order) sequence of numbers that follows. It is to be sorted into nondecreasing order.

```
31, 41, 59, 26, 53, 58, 97, 93, 23, 84, 62, 64
```

(a) (15 points) Illustrate the sorting of the number sequence using Heapsort. Show the heap right after BuildHeap finishes, after 6 DELETEMIN operations, and at the end of the sequence. (Show the binary heap as a tree – not as an array.)

(b) (10 points) Illustrate the sorting of the number sequence using Merge Sort. Show the state of the data after the merging at each level of recursion. For a sequence of length 3 or less, the data should be sorted directly, without more recursion. You should have 3 diagrams: after the length-3 sequences have been sorted, after the length-6 sequences have been sorted and after the whole operation is finished.

(c) (15 points) Illustrate the sorting of the number sequence using Quicksort. Use the median-of-three method in the book for selecting the pivot. Show the array after each of the first three partitioning steps and after the first recursive call returns.