Assignment 2 CSC263
Due: Oct 14, 2008

1. Consider a dictionary over the integers extended by the following two operations:

   (a) CLOSEST-NON-ELEMENT(S, x) which given a finite subset $S \subseteq \mathbf{Z}$ and an integer $x \in \mathbf{Z}$, returns an integer $y \in \mathbf{Z} - S$ (i.e., an integer y *not* in S) that is closest to x. (Note: x need not be in S.)

   (b) CLOSEST-PAIR(S) which returns two integers in S which are closest together in value. In other words, if CLOSEST-PAIR(S) returns the integers a and b, then they must satisfy the condition

   $$\forall x, y \in S \quad (x \neq y \rightarrow |a - b| \leq |x - y|).$$

   It is an error if S contains fewer than two elements.

   Explain how to augment a red-black tree so that CLOSEST-NON-ELEMENT(S, x) and CLOSET-PAIR(S), as well as the standard operations SEARCH(S, x), INSERT(S, x), and DELETE(S, x) can all be performed in $O(\log |S|)$ time. Justify why your algorithms are correct and run within the required time bound.

2. Consider two hash tables, T_1 and T_2, with the same number of buckets. With T_1, we use chaining to resolve collisions, where each chain is a doubly-linked list and insertions are done at the front of the list. With T_2, we use linear probing to resolve collisions, and deletions are done by replacing the item with a special "deleted" item. For both tables, we use the same hash function.

   (a) Let $S = S_1, S_2, ..., S_n$ be a sequence of INSERT and DELETE operations. Suppose that we perform S on both T_1 and T_2. For $1 \leq i \leq n$, let $C_{1,i}$ be the number of item comparisons made when performing S_i on T_1, and let $C_{2,i}$ be the number of item comparisons made when performing S_i on T_2. Show that for $1 \leq i \leq n$, $C_{1,i} \leq C_{2,i}$.

   (b) Explain why the claim in part (a) is no longer true if the sequence S also contains SEARCH operations.

   (c) Let S be as in part (a). Suppose we perform S on both T_1 and T_2, then do a SEARCH operation where each item in the table is equally likely to be searched for. Let E_1 be the expected number of item comparisons made when performing the SEARCH operation on T_1, and let E_2 be the expected number of item comparisons made when performing the SEARCH operation on T_2. Prove that $E_1 \leq E_2$.