TEL AVIV UNIVERSITY
Department of Computer Science
0368.4281 – Advanced topics in data structures
Fall Semester, 2005/2006

Homework 3, Dec 17, 2005

Due on Mon. Jan 2

1. Consider the following compression algorithm:

Given a string s of length n you run MTF on s to convert s to an integer sequence as follows. You initialize MTF's list to an empty list. When you first see a character, if it is not in the list yet you put it last. Then you replace the character which you serve by its index in the list (assume indices start from one), and move it to be the first in the list. Let $\overline{s} = \overline{s}_1, \ldots, \overline{s}_n$ be the resulting integer sequence.

Now we encode the integers $\overline{s}_1, \ldots, \overline{s}_n$ in a reversible manner. Let $b(\overline{s}_j)$ be the binary encoding of $\overline{s}_j$ (without any leading zeros). Let $|b(\overline{s}_j)|$ be the length in bits of this encoding. We encode $\overline{s}_j$ by $|b(\overline{s}_j)| - 1$ zeros followed by $b(\overline{s}_j)$.

Prove that the average length (per character) of this encoding is at most $\sum_\sigma 2\frac{n_\sigma}{n}\log(\frac{n}{n_\sigma})+1$, where n_σ is the number of occurrences of the character σ in s.

2. We define the following variation on the splay algorithm. This variation looks 3 steps (edges) towards the root from the node x and applies one of the rules in Figure 1 (or their mirror image) if possible. If it is not possible to apply one of the rules in Figure 1 we apply one of the regular zig-zig, zig-zag, or zig rules (Note that zig or zig-zig would apply only if x is at distance 1 or 2 from the root, respectively).

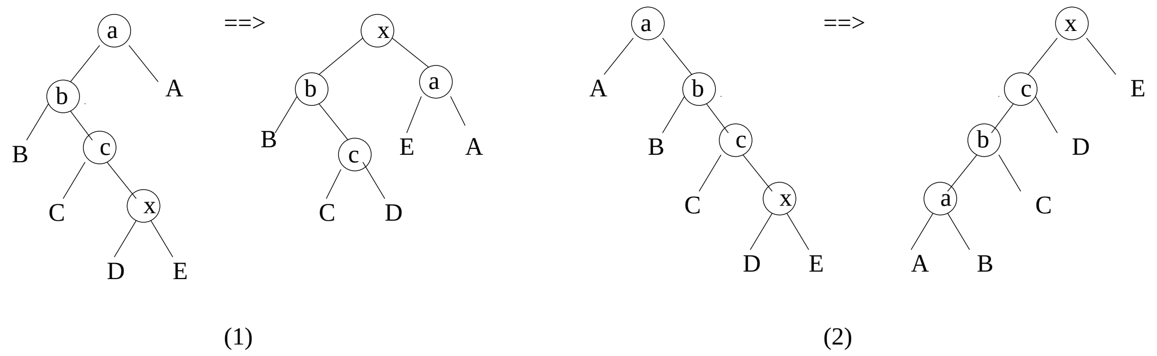

Figure 1: Splay cases in question (3)

Prove that the access lemma holds for this variation as well (with a different constant).

3. Assume that the dynamic optimality is true and prove that to splay the nodes of a binary tree one by one in symmetric order takes $O(n)$ time where n is the number of nodes in the tree.

4. Prove that the static optimality theorem follows from the working set theorem for a sufficiently long access sequences.

5. a) Consider the tree on the left in Fig 2 and prove that there is no access sequence on items $1, 2, 3$ that after serving it using the splay algorithm starting with that tree, we end up with the tree on the right.
b) Consider any binary tree on 4 items $1, 2, 3, 4$ (all these trees are shown in Fig 3). Prove that for any pair of such trees T_1 and T_2 there is an access sequence that if we start with T_1 and serve σ using the splay algorithm we end up with T_2.
c) Use induction to prove that for any $n \geq 4$ if T_1 and T_2 are binary trees on items $1, 2, \ldots, n$ then there is an access sequence σ such that if we start with T_1 and serve σ using the splay algorithm we end up with T_2.

Figure 2: Cannot reach the tree on the right from the tree on the left.

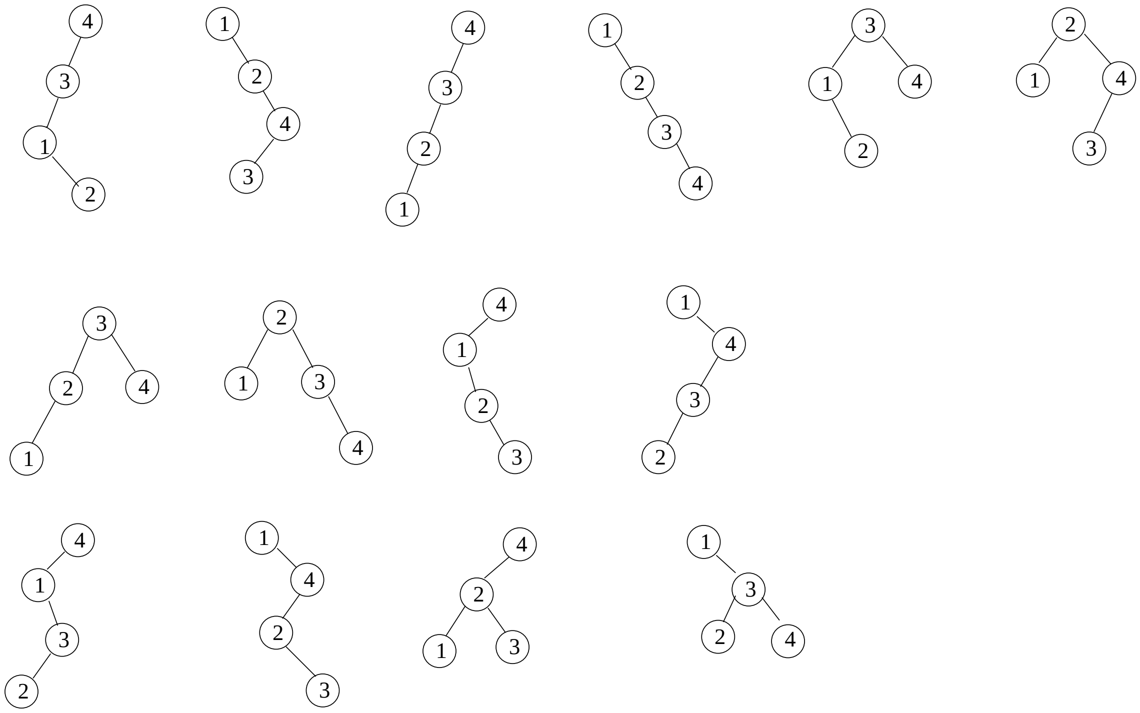

Figure 3: All binary trees with 1,2,3,4.