58093 String Processing Algorithms (Autumn 2013)
Practice problems

Please give feedback by filling the feedback form at `https://ilmo.cs.helsinki.fi/kurssit/servlet/Valinta?kieli=en`

The problems below are intended for self-practice on the last topics of the course. You are not expected to return answers to the problems in any form. Solutions will be posted to the course home page on Tuesday, December 10.

1. Let $T =$ `lallilla$`.
   (a) Give the suffix tree of T including suffix links.
   (b) Give the suffix array of T together with the LCP array.
2. The reverse of a string $S[0..m)$ is the string $S^R = S[m-1]S[m-2]..S[0]$. Describe an algorithm for finding the longest factor S of $T[0..n)$ such that the reverse S^R is a factor of T too. The algorithm should work in linear time on an integer alphabet of size $\sigma \leq n$.
3. Let $\mathcal{R} = \{S_1, S_2, \ldots, S_k\}$ be a set of strings over a constant size alphabet such that no string in $\mathcal{R}$ is a factor of another string in $\mathcal{R}$. The *shortest distinguishing factor* of S_i is the shortest string that occurs in S_i but not in any other string in $\mathcal{R}$. Describe an algorithm for finding the shortest distinguishing factor for all strings in $\mathcal{R}$. The time complexity should be $\mathcal{O}(||\mathcal{R}||)$, where $||\mathcal{R}||$ is the total length of the strings in $\mathcal{R}$.
4. Prove Lemma 4.11. *Hint:* Generalize Lemma 1.25(b) (Lecture 3, slide 47) from three strings to many strings.
5. Let $L =$ `rttrraa$ii` be the Burrows–Wheeler transform of a text T.
   (a) What is T?
   (b) Simulate backward search on T for the pattern $P =$ `ari`.
6. Let $T =$ `senselessness$`.
   (a) Give the $C = C_1 \cup C_2$ suffixes of the DC3 algorithm for T.
   (b) Give the $C = C^*$ suffixes of the SAIS algorithm for T.