Eberhard Karls Universität Tübingen
Wilhelm-Schickard-Institut für Informatik
Prof. Dr. Daniel Huson
Dr. Gunnar Rätsch, Dr. Johannes Fischer

Wintersemester 2008/09
Exercise set no. 4

Advanced Methods for Sequence Analysis

due: Monday, November 10th, before the lecture

Exercise 1

Let T = antananarivo\$.

a) Show step-by-step how the $O(n)$-algorithm presented in the lecture constructs the suffix array for T. Assume that the recursive call gives the result immediately.

b) Show step-by-step how the $O(n)$-algorithm presented in the lecture constructs the corresponding LCP-array. Mark all (un)successful character comparisons!

Exercise 2

In this exercise, we look at two small modifications of the algorithm from Sect. 3.5.

a) Assume that instead of sorting the suffixes starting at positions $i \not\equiv 0(3)$ by a recursive call (with result A_{12}), we only sort the suffixes starting at positions $i \equiv 1(2)$ recursively (with result A_1), and try to derive A_0 from A_1. Is this a good idea? If so, why? If not, what goes wrong?

b) Now assume that instead of deriving A_0 from A_{12}, we also call the suffix sorting routine recursively for the suffixes starting at positions $i \equiv 0(3)$. Give the resulting recurrence relation for $\mathcal{T}'(n)$, the running time of this modification. Then solve this recurrence!

Exercise 3

Let $T \in \Sigma^*$ be a text of length n and $k > 0$ be an integer. Find an algorithm that finds in $O(n)$ time the number of *different* factors in T of length exactly k.

Hint: What does $H[i] \geq k$ imply?

Exercise 4

Let W be a set of m words $w_1, \ldots, w_m$, $w_i \in \Sigma^*$, with total length $n = \sum_{i=1}^{m} |w_i|$. Give an efficient algorithm that preprocesses W into a data structure of final size $O(m)$ (in addition to the space needed for storing the strings in W) that allows to answer online queries of the form "is $P \in \Sigma^*$ one of the w_i's?" in $O(|P|)$ time.