Marburg University

Department of Computer Science
Prof. Dr. Sebastian Wild

Date: 2025-01-16
Version: 2026-01-15 22:08

Exercise Sheet 11 for Algorithms of Bioinformatics (Winter 2025/26)

Hand In: *Until 2026-01-23 18:00, on ILIAS.*

Problem 1

10 points

Fix any string S, and let $\mathcal{T}$ be the suffix tree associated with S. If v is any internal node of $\mathcal{T}$, and $S_v[0..k)$ is the string found by traversing $\mathcal{T}$ from the root to v, let $S'_v = S_v[1..k)$. (In other words, S'_v is S_v but without its first character.) Prove or disprove the following: if we traverse T from the root following string S'_v, we stop at an internal node of the suffix tree.

Problem 2

20 points

Given a string $S[0..n)$, count the number of *distinct* nonempty substrings of S in $O(n)$ time.

Problem 3

30 points

Solve the longest common substring problem using suffix arrays. In other words, given a family of nonempty strings $S_1, \ldots, S_k$, compute the longest substring which appears in all of $S_1, \ldots, S_k$, in linear time with respect to $N = \sum_i |S_i|$, using suffix arrays.

For full marks, the running time should not depend on k (other than via N). Note that the solution based on generalized suffix trees from class does not achieve this.