Marburg University

Department of Computer Science
Prof. Dr. Sebastian Wild

Date: 2025-12-12
Version: 2025-12-19 17:38

Exercise Sheet 10 for Algorithms of Bioinformatics (Winter 2025/26)

Hand In: *Until 2025-12-19 18:00, on ILIAS.*

Problem 1

30 + 30 points

In the lecture, we considered an algorithm for computing all tandem repeats.

Formally, "all tandom repeats of T" means the following set

$$R \;=\; \{(i,\ell) : T[i..i+\ell) = T[i+\ell, i+2\ell)\}.$$

a) Describe a method based on this algorithm to compute all *triple repeats*, i.e., all subwords of shape xxx in a text T.

b) Generalize your method to *higher order repeats*, i.e., subwords of the form x^k for arbitrary $k \geq 2$.

Problem 2

20 + 20 points

Assume we are given a binary string $S \in \Sigma^n$ of length n.

a) Find the shortest string over Σ that *does not* appear as a substring of S.

b) Find the shortest string that appears *exactly once* as a substring of S.

For full marks your solution should run in $O(n)$ time.