

Exercise Sheet 9 for Algorithms of Bioinformatics (Winter 2025/26)

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

Problem 1

50 points

In this exercise, we consider again the “gold standard” for inexact string matching, where we would like to find all approximate occurrences of a pattern $P \in \Sigma^m$ in a text $T \in \Sigma^n$ ($n > m$)

Specifically, we consider here the *k-Difference Inexact String Matching Problem*: There, a subword $T[i..i + m']$ of T is considered an occurrence of search string $P[0..m]$ iff $T[i..i + m']$ and P have *edit distance* at most k , i.e., $\text{dist}_{\text{edit}}(T[i..i + m'], P)$, abbreviated $T[i..i + m'] \approx_k P$ here.

For full marks, your algorithm is expected to return the set of *all* indices i , such that there is a length m' with $T[i..i + m'] \approx_k P$.

Design an algorithm for the k -difference inexact string matching problem that is more efficient than the $\Theta(nm)$ baseline from class when $k \ll m \leq n$ and determine its running time.

Hint: Use LCE queries.