Xiheng He

Lisanne Friedrich

Exercises for Algorithmic Bioinformatics II

Assignment 1

Xiheng He

November 2021

Exercise 1 (Chang-Lawler proof, 10P):

Prove the following lemma (also given in the lecture slides):

Lemma. When the CL search excludes a region R of T, then there is no k-approximate match of P in T that completely contains region R.

Is it also guaranteed that a region is excluded if it contains $\geq k + 1$ mismatches? Prove it or give a counter-example.

When count $> k$ loop for searching terminated too early. That leads to more than k mismatches in the region R, which means that the region should be excluded. That contradicts the lemma.

counter-example:

pattern: baaaccaaaabacaba

k = 2

text: aaaabaaacacbabbbabaaaccaaaabacabacccbcaaabcacb

region included: acccbcaa

mismatch = 3