

Advanced Algorithms in Bioinformatics (P4)

Sequence and Structure Analysis

Freie Universität Berlin, Institut für Informatik
Prof. Dr. Knut Reinert, Sandro Andreotti
Sommersemester 2009

1. Exercise sheet, 22. April 2009

Discussion: 22. April 2009

Exercise 1.

Subscribe yourself to the **mailing lists** as described in the lecture notes.

Exercise 2. Exact DNA matching with Horspool

Find all occurrences of the pattern ACAACC in the string ACACTCCCCGACAACC using the Horspool algorithm.

Why does the algorithm show such a poor performance in this example?

How can (for this example) the number of comparisons be reduced? (Algorithm?)

Exercise 3. Exact multiple DNA matching with Wu-Manber

Find all occurrences of the patterns ATATATA, TATAT, TAGACG in the string AGATAGACGATATATACG using the Wu-Manber Algorithm. Use a block size of 2. You may use the identity as the hash function, so you have no collisions

*Exercise 4. ** Exact multiple DNA matching with Horspool

As mentioned in the lecture, the Horspool algorithm can also be used for multiple string matching. You need a trie of the reversed patterns for search.

Use the same patterns and string as in Exercise 3.