Advanced Algorithms in Bioinformatics (P4)

Sequence and Structure Analysis

Freie Universität Berlin, Institut für Informatik
Knut Reinert, Sandro Andreotti
Sommersemester 2012

1. Exercise sheet, 10. April 2012
Discussion: 18. April 2012

Exercise 1.

Subscribe to the mailing list **https://lists.fu-berlin.de/listinfo/bioinf-P4-2012**

Exercise 2. Exact DNA matching with Horspool

- Find all occurrences of the pattern `ACAACC` in the string `ACACTCCCCCGACAACC` 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?)

- Prove that the safe shift computed by the Horspool is indeed safe (no occurence of the pattern in the text is missed).

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

- Assume the DNA alphabet and a block size $B = 2$. For which patterns would we observe the shift worst-case independent of the text?
- Does this worst-case occur if the block size is chosen as proposed in the algorithm pseudocode? Why?

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.