Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye
Dr. Lucas Robidou

Sequence Analysis 3

Summer 2024

Exercises

Number 4, Discussion: 2024-May-16

1. Which data structures could one use to store the branching vertices in a suffix tree?

   What is the resulting memory usage per branching vertex and what is the search time for a word of length m?

2. Given the suffix tree $T(s)$, explain how all positions of the k occurrences of a substring p of s can be enumerated in $O(|p| \cdot \log |\Sigma| + k)$ time.

3. Given the suffix tree $T(s)$, explain how the number of different substrings of s can be computed in $O(|s|)$ time.

   Explain why there is no factor $\log |\Sigma|$ in the asymptotic complexity, other than in the previous exercise.

4. The relationship between the suffix tree for a string s and for the reverse string $\tilde{s}$ is not obvious. However, there is a significant relationship between the two trees. Find it, state it, and prove it.

   Hint: It has to do with suffix links.

5. Show all intermediate steps when constructing the suffix tree $T(s)$ of $s =$ `abbbabbbaa` using Ukkonen's algorithm.