Technische Universität München
Department of Informatics
Chair for Efficient Algorithms
Prof. Dr. Ernst W. Mayr/Dr. Jens Ernst
Johannes Nowak

Summer Term 2005
Exercise Sheet 6
June 6, 2005

Selected Topics in Computational Biology

Due: 13.06.2005 **after** *the lecture*

Exercise 1 (10 points)

Consider the algorithm for finding the longest common substring of two strings described in the lecture. Give a detailed description of steps II)-IV) in pseudocode. Show that only one linear-time traversal of the generalized suffix tree is necessary.

Exercise 2 (10 points)

Consider again the problem of finding the longest common substring of two strings s^1 and s^2. Suppose we have a suffix tree T for one of the two strings, say s^1. Give an algorithm for finding the longest common substring of s^1 and s^2 in time $O(|s^2|)$ with the help of T.
Hint: Adding s^2 to T does not help!

Exercise 3 (10 points)

The *matching statistics* of a text t with respect to a pattern p are defined as follows. For every position i, $1 \leq i \leq |t|$, $ms(i)$ is the length of the longest substring of t starting at position i that is also a substring of p. (For example, let $t = abbabbaabbaba$ and $p = aaabb$, then $ms(1) = 3$ and $ms(7) = 4$.) Devlop an efficient algorithm that computes the matching statistics of given t and p for all $1 \leq i \leq |t|$.
Hint: Suffix trees may help, aim for $O(\max\{|t|, |p|\})$.

Exercise 4 (10 points)

For $s, t \in \Sigma^+$ and a weight function $w : (\Sigma \times \Sigma) \longrightarrow \mathbb{R}_0^+$, show that the alignment distance between s and t satisfies the equality $\bar{d}_w(s^R, t^R) = \bar{d}_w(s, t)$, where s^R, t^R are the strings obtained from s, t by reversal.