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 4
May 24, 2005

Selected Topics in Computational Biology

Due: 31.05.2005 **after** *the lecture*

Exercise 1 (10 points)

Consider the tree T_i which is constructed in the i-th step of McCreight's algorithm. Proof that

a) All nodes with the only possible exception of the node corresponding to the locus of $head_i(t)$ have a valid suffix link.

b) In step i, the algorithm visits the contracted locus of $head_i(t)$ in T_{i-1}.

Exercise 2 (10 points)

Compare Ukkonen's and McCreight's algorithm for suffix tree construction. Describe the basic idea of both algorithms. Are there similarities between the algorithms? Where are the differences?

Exercise 3 (10 points)

Let $s, t \in \Sigma^*$ and $|s| = |t| = n$. Show that the number of all possible different alignments of s and t is exponential in the length of the sequences. To achieve this, you can proceed as follows.

a) Show that $\binom{2n}{n}$ is a lower bound for the total number of alignments.

b) Use Stirling's formula ($n! \approx \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$ as $n \to \infty$) to show that this number grows exponentially in n.

Exercise 4 (10 points)

Consider the method of Amir et al. for *Approximate Text Indexing With One Error*. Why is it necessary to modify the algorithm if exact matches exist?