Algorithmic Bioinformatics I: Exercises
Assignment 11
Deadline: Tuesday, 21.07.2009, 10 ct

Exercise 1 (String matching):
State all shifts and comparisons of the naive algorithm, the Knuth-Morris-Pratt and the Boyer-Moore algorithms on the following example: $P = \text{ALOA}, T = \text{ALGORITHMIK}$.

Exercise 2 (KMP modification):
Modify the algorithm of Knuth-Morris-Pratt such that it finds all occurrences of a pattern $s \in \Sigma^m$ in a text $t \in \Sigma^n$. Describe the necessary modifications and analyze the runtime.

Exercise 3 (Z boxes):
State all $Z_i$ values for the given binary string:
$00101001000101000101$

Exercise 4 (Repeats):
A repeat is a substring $S$ of a string $T$ of the form $S = \alpha^k$ ($k$-fold concatenation of the string $\alpha$), where $\alpha$ is called the basis of the repeat. A maximal repeat to the basis $\alpha$ is a repeat which is not a substring of a longer repeat to this basis. Develop an algorithm which finds for a given basis all maximal repeats in a given string. Analyze the runtime of your algorithm. Full points are given for a linear algorithm. \textit{Hint:} Use e.g. the Z-algorithm.