

**Foundations of Sequence Analysis**  
**Winter semester 2003/2004**

**Exercises**

**Exercise 6, Discussion: 01/12/2004.**

1. (Suffix Array – Suffix Tree)

Given an  $n$ -character string  $S$ , a suffix array for  $S$  is an array of length  $n$ , containing all integers in the range 1 to  $n$  referring to starting positions of all suffixes in  $S$ . These numbers specify the lexicographic order of the referred suffixes.

Example:

Let  $S = abab$  be the given string, then

3 – ab  
1 – abab  
4 – b  
2 – bab

represents the suffix array for  $S$ .

- (a) Implement some algorithm to construct the suffix array for a given string.
- (b) Implement an appropriate representation for suffix trees, assuming an alphabet  $\mathbf{A} = \{0, \dots, n - 1\}$  consisting of numbers between 0 and  $n - 1$ . For each node, the childlist can either be implemented as an array or as a linked list.
- (c) Given a string and its suffix array, implement an algorithm to construct the suffix tree by using the suffix array.
- (d) Given a suffix tree for a given string, construct the suffix array for the string by a traversal of the suffix tree.
- (e) Implement methods checking the correct behavior of the previously implemented algorithms.

Remark:

For each node in the suffix tree, the childlist is supposed to be ordered such that the list of children is ordered by the first character of the edges pointing from the node to its children.