

# Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WiSe 2009/2010

Prof. Dr. Jens Stoye · Dipl.-Inform. Nils Hoffmann

<http://wiki.techfak.uni-bielefeld.de/gi/GILectures/2009winter/SequenzAnalyse>

**Blatt 13 vom 28.01.2010**

**Abgabe in einer Woche vor Beginn der Vorlesung.**

Für die folgenden Aufgaben solltest Du im Skript das Kapitel über Suffixbäume (Kapitel 8) lesen, insbesondere die Abschnitte 8.1-8.3 und 8.5.

## Aufgabe 1 Suffixbaum

(6 Punkte)

Erzeuge für das Wort  $y = ACABACABACAB$  den Suffixbaum. Verwende zur Konstruktion den in der Vorlesung verwendeten, einfachen WOTD Algorithmus. Sortiere die Blätter lexikographisch ( $\$ < A < B < C$ ) und beschrifte jedes Blatt mit dem Startindex des zu ihm gehörenden Suffixes in  $y$ . Verwende  $\$$  als Endmarkierung.

1. Überlege dir eine Methode, um die Anzahl der Vorkommen des Substrings  $CAB$  in  $y$  mit Hilfe des Suffixbaums zu ermitteln. Wie gehst Du dabei vor?
2. Bestimme alle Anfangspositionen des Teilwortes  $ACA$  in  $y$  mit Hilfe des Suffixbaums.
3. Nutze den Suffixbaum von  $y$ , um das längste Teilwort zu finden, welches mindestens zweimal in  $y$  vorkommt.

## Aufgabe 2 Generalisierter Suffixbaum

(4 Punkte)

Erzeuge für die Worte  $x = CABBBAB$  mit der Endmarkierung  $\#$  und  $y = BBACAB$  mit der Endmarkierung  $\$$  den generalisierten Suffixbaum von  $x\#y\$$ . Die Blätter sollen gemäß der Ordnung  $\# < \$ < A < B < C$  sortiert werden.

1. Finde ein längstes gemeinsames Teilwort von  $x$  und  $y$  mit Hilfe des generalisierten Suffixbaums. Beschreibe dein Vorgehen!
2. Finde alle kleinsten gemeinsamen Teilworte der Länge  $\ell = 2$  von  $x$  und  $y$ , die in  $x$  und  $y$  jeweils höchstens einmal auftreten.