

# Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2010

Prof. Dr. Jens Stoye · Dr. Inke Herms

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010summer/SequenzAnalyse>

**Blatt 2 vom 23.04.2010**

**Abgabe in einer Woche vor Beginn der Vorlesung.  
Bitte gib auch den Namen deines Tutors an.**

## **Aufgabe 1 (Kürzester eindeutiger Teilstring)**

(3 Punkte)

Gib einen Linearzeit-Algorithmus an, mit dem man einen kürzesten eindeutigen Substring eines Strings  $s$  finden kann, wenn der Suffixbaum von  $s\$$  gegeben ist (die Substrings, die mit  $\$$  enden, zählen hierbei nicht, sondern nur Substrings von  $s$  selbst).

## **Aufgabe 2 (Links-Rechts-Partition)**

(4 Punkte)

Die Links-Rechts-Partition  $P_{lr}(s, t)$  eines Strings  $s$  bezüglich eines Strings  $t$  (siehe Abschnitt 3.8 im Skript) kann mit Hilfe eines Suffixbaumes effizient berechnet werden.

- Überlege dir einen Algorithmus, der die Links-Rechts-Partition  $P_{lr}(s, t)$  in linearer Zeit berechnet.
- Verwende diesen Algorithmus, um  $P_{lr}(s, t)$  für  $s = \text{BABAABABB}$  und  $t = \text{BABBABA}$  zu berechnen.

## **Aufgabe 3 (Maximale Repeats)**

(5 Punkte)

Lies für diese Aufgabe Abschnitt 8.7.4 im Skript, insbesondere den Algorithmus zum Finden maximaler Repeats.

- Finde alle maximalen Repeats in  $s = \text{ABCDAABCABCD}$  unter Verwendung des im Skript geschilderten Algorithmus. Beschreibe dein Vorgehen beim Annotieren des Suffixbaumes.
- Satz:** In jedem String der Länge  $n$  gibt es höchstens  $n$  maximale Repeats.

Argumentiere unter Berücksichtigung des Suffixbaumes, warum die Aussage korrekt ist. Bedenke: Es stimmt nicht, dass an jeder Position nur ein maximaler Repeat beginnen oder enden kann.