

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2024

Prof. Dr. Jens Stoye · Leonard Bohnenkämper

<https://gi.cebitec.uni-bielefeld.de/teaching/2024winter/sa>

Übungsblatt 8 vom 05.12.2024

Abgabe am 12.12.2024 bis 10:00 Uhr (morgens)

## Aufgabe 1 (Suffixbäume)

(8 Punkte)

Gegeben sei die Sequenz  $s = \text{BABABBABA}$ .

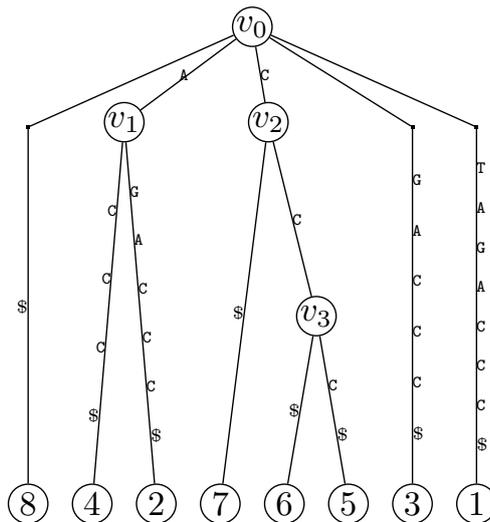
1. Führe den WOTD-Algorithmus schrittweise zur Erstellung des Suffixbaums für die Sequenz  $s\$$  aus. Sortiere dabei die Kanten lexikographisch (wobei  $\$ < A < B$ ).
2. Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in  $s$ . Die Indizierung beginnt bei 1.
3. Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.

## Aufgabe 2 (Links-Rechts-Partition)

(6 Punkte)

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

1. Überlege dir einen Algorithmus, der die Links-Rechts-Partition  $P_{lr}(s, t)$  in linearer Zeit berechnet. (Hinweis: Verwende den Suffixbaum von  $t\$$ .) Gib die einzelnen Schritte deines Algorithmus' explizit und verständlich an. Optimalerweise nutzt du hierzu Pseudocode.
2. Verwende diesen Algorithmus, um  $P_{lr}(s, t)$  für  $s = \text{ACCTAGC}$  und  $t = \text{TAGACCC}$  zu berechnen. Du findest den Suffixbaum von  $t$  unten. Gib deine Zwischenschritte an.



(Bitte wenden!)

**Aufgabe 3 (Shortest Unique Substring)**

(6 Punkte)

Betrachte für diese Aufgabe den String  $s\$ = 123456789 \text{ RABARBARA}$  Auf der nächsten Seite ist der Suffixbaum für  $s\$$  gegeben.

1. Nutze den Algorithmus aus dem Skript, um einen kürzesten einzigartigen Substring zu finden.
2. Modifiziere den Algorithmus so, dass er die  $n$  kürzesten einzigartigen Substrings findet. Gib die einzelnen Schritte deines Algorithmus' explizit und verständlich an. Optimalerweise nutzt du hier ebenfalls Pseudocode.
3. Nutze diesen Algorithmus, um die 5 kürzesten Substrings zu finden und markiere sie im Suffixbaum.

