

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2021

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2021summer/sa>

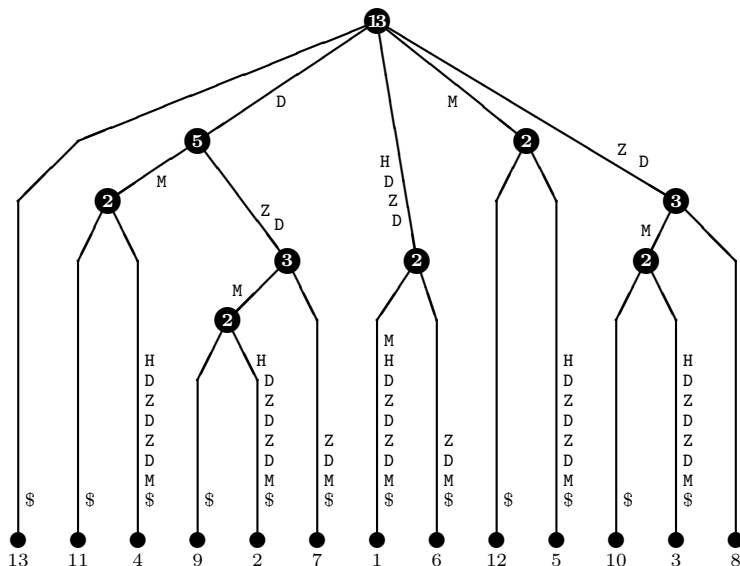
## Übungsblatt 10 vom 24.6.2021

Abgabe am 1.7.2021 bis 12:00 Uhr (mittags)

### Aufgabe 1 (Suffixbaum und lcp-Array)

(2 Punkte)

Gegeben seien die Sequenz  $s = \text{HDZDMHDZDZDM}$  und der Suffixbaum von  $s\$$  mit lexikographisch sortierten Kanten:



Berechne mit Hilfe dieses Baums das lcp-Array von  $s\$$ , ohne **rank** oder **pos** explizit auszurechnen.

### Aufgabe 2 (Manber-Myers-Algorithmus)

(9 Punkte)

- Gegeben sei die Sequenz  $s = \text{USRURTUSTSRUSRURURT}$ .
  - Erstelle das Suffix-Array  $\text{pos}(s\$)$  unter Verwendung des Manber-Myers-Algorithmus. Gib das Zwischenergebnis nach jeder Phase des Algorithmus' an.
  - Berechne die Arrays  $\text{rank}(s\$)$  und  $\text{lcp}(s\$)$ .
  - Verwende Binärsuche im Suffix-Array  $\text{pos}(s\$)$ , um alle Vorkommen des Musters  $p = \text{RUR}$  in  $s$  zu finden. Gib alle Zwischenschritte an.
- Wie viele Phasen benötigt (im *worst case*) der Manber-Myers-Algorithmus für einen String der Länge  $n$ ? Gib eine Beispielsequenz an, bei der diese Anzahl Phasen erreicht wird.

### Aufgabe 3 (Burrows-Wheeler-Transformation)

(9 Punkte)

- Gegeben sei der String  $t = \text{DDACCBDACBDDCCB}$ .
  - Berechne die Burrows-Wheeler-Transformierte  $s = \text{BWT}(t\$)$ . Gib alle Zwischenschritte an.
  - Demonstriere den String-Matching-Algorithmus unter Verwendung der Sequenz  $s = \text{BWT}(t\$)$  beispielhaft an der Suche des Musters  $p = \text{ACBDD}$  in  $t$ .
  - Schreibe  $r = \text{RLE}(s)$  als komprimierten String mit Hilfe von *run-length encoding* auf. Fasse dabei nur Buchstaben zusammen, die mindestens dreimal hintereinander vorkommen.
- Gegeben sei der String  $r = \text{BD4CDA\$CA3DBB}$ .
  - Dekomprimiere  $r$  in  $s = \text{RLE}^{-1}(r)$ .
  - Rekonstruiere den String  $t\$ = \text{BWT}^{-1}(s)$ . Gib alle Zwischenschritte an.