

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2024

Prof. Dr. Jens Stoye · Leonard Bohnenkämper

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

**Übungsblatt 10 vom 19.12.2024**

**Abgabe am 09.01.2025 bis 10:00 Uhr (morgens)**

## Aufgabe 1 (Manber-Myers-Algorithmus)

(8 Punkte)

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

- 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  $p = \text{AT}$  in  $s$  zu finden. Gib alle Zwischenschritte an.

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

(3 Punkte)

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)

1. Gegeben sei der String  $t = \text{BABBABA}$ .

- Berechne die Burrows-Wheeler-Transformierte  $s = \text{BWT}(t\$)$ . Gib alle Zwischenschritte an.
- 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.

2. Gegeben sei die BWT eines (anderen) Strings  $u$  mit  $\text{BWT}(u\$) = \text{XZZYYYYYWZWWZZX\$X}$ . Demonstrieren Sie den String-Matching-Algorithmus unter Verwendung der BTW beispielhaft an der Suche des Musters  $p = \text{WYXZZ}$  in  $u$ .

3. Gegeben sei der String  $r = \text{XZ4YZW\$YW3ZXX}$ .

- Dekomprimiere  $r$  in  $s = \text{RLE}^{-1}(r)$ .
- Rekonstruiere den String  $t\$ = \text{BWT}^{-1}(s)$ . Gib alle Zwischenschritte an.