

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2023

Prof. Dr. Jens Stoye · Tizian Schulz

<https://gi.cebitec.uni-bielefeld.de/teaching/2023summer/sa>

**Übungsblatt 7 vom 18.5.2023**

**Abgabe am 25.5.2023 bis 12:00 Uhr (mittags)**

## Aufgabe 1 (Links-Rechts-Partition)

(8 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.
2. Verwende diesen Algorithmus, um  $P_{lr}(s, t)$  für  $s = \text{ACGGTCAACA}$  und  $t = \text{CAACG}$  zu berechnen. Gib alle Zwischenschritte an.

## Aufgabe 2 (Suffixbaum-Anwendungen)

(10 Punkte)

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

1. Berechne den Suffixbaum von  $s$ . Sortiere dabei die von einem Knoten ausgehenden Kanten lexikographisch (mit  $\$ < A < B < C < D$ ).
2. Kürzeste eindeutige Substrings:
  - (a) Finde alle eindeutigen Substrings in  $s$  mit Hilfe des Suffixbaums von  $s$ . Welche sind die kürzesten eindeutigen Substrings?
  - (b) Beschreibe eine mögliche Anwendung für kürzeste eindeutige Substrings.