

# Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2022

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

<https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa>

Präsenzübungsblatt 7, vom 23./24.5.2022

## Aufgabe 1 (Suffixbäumer)

Gegeben ist der String  $s\$ = \text{ATAATTATTTCCGCCGGGG\$}$ , wobei  $\$ < A < C < G < T$ .

1. Zeichne den Suffixbaum für  $s$ , sortiere dabei alle ausgehenden Kanten lexikographisch.
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 (Suche im Suffixbaum)

Suche die folgenden *Patterns* in  $s$  mit Hilfe des Suffixbaums von  $s\$$  (aus Aufgabe 1).

AATA    AATTA    ATTAC    CGGG    GCCGC    GGCG    TAATT    TAC    TTA    TTCGC

## Aufgabe 3 (Suffixbaum-Anwendungen)

Finde alle kürzesten eindeutigen Substrings in  $s$  mit Hilfe des Suffixbaums von  $s\$$  (aus Aufgabe 1).