

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2025/26

Prof. Dr. Jens Stoye

Leonard Bohnenkämper

Tutorien: Lennart Finke, Sofie Jans

<https://gi.cebitec.uni-bielefeld.de/teaching/2025winter/sa1>

Übungsblatt 8 vom 18.12.2025

Abgabe bis 08.01.2026 bis 9:30 Uhr per Mail an den Tutor/die Tutorin

Aufgabe 1 (Σ -Baum)

(6 Punkte)

Gegeben sei die Menge der Worte $W = \{star, tars, sad, salad, art, card, at, scar, cars, cat\}$.

1. Zeichne den kleinsten Σ -Baum T und den kleinsten kompakten Σ^+ -Baum T' , welcher alle Worte aus W darstellen.
2. Gib für T und T' jeweils die Menge der Worte $x \in \Sigma^*$ an, für die $node(x)$ definiert ist.
3. Welche Menge $words(T)$ von Worten wird durch T dargestellt? Gibt es einen Unterschied zu $words(T')$?

Aufgabe 2 (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 $\$ < \text{A} < \text{B}$).
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.

(Bitte wenden!)

Aufgabe 3 (Suffixbäume und Substrings)

(6 Punkte)

Betrachte für diese Aufgabe den String $s = \overset{123456789}{\text{RABARBARA}}$. Unten ist der Suffixbaum für $s\$$ gegeben, für den Kanten lexikographisch sortiert sind.

Tipp: Jeder Substring v eines Strings $x = uvw$ ist der Präfix (v) eines Suffixes (vw) von x !

1. Markiere alle einzigartigen Substrings von s im Suffixbaum.
2. Markiere den lexikographisch kleinsten einzigartigen Substring von s im Suffixbaum.
3. Entwickle einen Algorithmus, um den lexikographisch kleinsten einzigartigen Substring zu finden. Beantworte dazu zunächst folgende Fragen:
 - (a) Wo findet man einzigartige Substrings im Suffixbaum?
 - (b) Wenn die Kanten lexikographisch sortiert sind, wo finden sich die kleinsten, wo die größten Suffixe im Suffixbaum?

