

Übungen 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>

Übungsblatt 7 vom 19.5.2022

Abgabe am 26.5.2022 bis 12:00 Uhr (mittags)

Aufgabe 1 (Σ -Baum)

(4 Punkte)

Gegeben sei die Sequenz $s = \text{C A B A C A A}$.

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

Aufgabe 2 (Suffixbäume)

(6 Punkte)

Gegeben sei die Sequenz $s = 3033213021$.

1. Führe den WOTD-Algorithmus schrittweise zur Erstellung des Suffixbaums für die Sequenz $s\$$ aus. Sortiere dabei die Kanten lexikographisch (wobei $\$ < 0 < 1 < 2 < 3$).
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 3 (Suffixbäume)

(4 Punkte)

1. Analysiere die *worst-case*- und *average-case*-Laufzeit des WOTD-Algorithmus in eigenen Worten.
2. Zeige, dass der Speicherverbrauch eines Suffixbaums linear bezüglich der Eingabe ist.

Aufgabe 4 (Links-Rechts-Partition)

(6 Punkte)

Die Links-Rechts-Partition $P_{\text{LR}}(s, t)$ 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 $P_{\text{LR}}(s, t)$ 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 $P_{\text{LR}}(s, t)$ für $s = \text{ACGGTCAACA}$ und $t = \text{CAACG}$ zu berechnen. Gib alle Zwischenschritte an.