

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019
Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

Übungsblatt 8 vom 27.5.2019
Abgabe am 3.6.2019 bis 12:00 Uhr (mittags)

Aufgabe 1 (Σ -Baum)

(4 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 (WOTD-Algorithmus)

(4 Punkte)

1. Führe den WOTD-Algorithmus schrittweise zur Erstellung des Suffixbaums für den String $t\$ = \text{BABBA\$}$ aus (so wie es im Skript auf Seite 68 gezeigt ist).
2. Informiere dich im Skript über das average case Laufzeitverhalten von WOTD und erkläre es mit deinen eigenen Worten.
3. Welche Eigenschaften machen den WOTD-Algorithmus in der praktischen Durchführung interessant?

Aufgabe 3 (Maximal-Matches-Distanz)

(2 Punkte)

Berechne die Links-Rechts-Partition von $s = \text{BABAABABB}$ bezüglich t aus Aufgabe 2. Inwiefern ist der Suffixbaum von t dazu nützlich?

Aufgabe 4 (Laufzeit- und Speichereffizienz von Suffixbäumen)

(3 Punkte)

In den inneren Knoten eines Suffixbaumes müssen auf irgendeine Weise die Kinder gespeichert werden. Wie viel Platz insgesamt (bei Textlänge n und Alphabetgröße σ) und wie viel Zeit zum Überprüfen, ob der Baum ein Muster der Länge m darstellt, benötigt man jeweils bei Verwendung

1. eines Arrays fester Größe σ ?
2. eines sortierten Arrays aller Kinder?
3. einer verketteten Liste aller Kinder?