

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

Übungsblatt 7 vom 24.05.2018

Abgabe bis spätestens Freitag, 31.05.2018 23:59 GMT+2 via Email oder Postfach
(U10-151)

Aufgabe 1 (Suffixbäume)

(5 Punkte)

- Gegeben ist der String $s\$ = \text{TTACTATCCA\$}$, wobei $\$ < A < C < T$.
 - Zeichne den Suffixbaum für $s\$$, sortiere dabei die Blätter lexikographisch.
 - Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in $s\$$. Die Indizierung beginnt bei 1.
 - Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.
- Analysiere die *worst-case*- und *average-case*-Laufzeit des WOTD-Algorithmus in eigenen Worten.

Aufgabe 2 (Maximale Repeats)

3+3* Punkte)

Lesen Sie Abschnitt 7.7.3 im Skript über das effiziente Auffinden von maximalen Repeats in einem Text s mit Hilfe des Suffixbaums von s .

- Stelle den Suffixbaum von $s = \text{AGCTCTACTCGCTA}$ auf.
- Satz:** In jedem String der Länge n gibt es höchstens n Teilworte, die maximale Repeats sind.
Argumentiere unter Berücksichtigung des Suffixbaums, warum diese Aussage korrekt ist. Bedenke: Es stimmt nicht, dass an jeder Position nur ein maximales Repeat beginnen oder enden kann.
- Für 3* Punkte: Finde alle maximalen Repeats in s unter Verwendung des im Skript geschilderten Algorithmus. Beschreibe dein Vorgehen beim Annotieren des Suffixbaums aus Aufgabenteil 1.

Aufgabe 3 (Links-Rechts-Partition)

(3 Punkte)

Die Links-Rechts-Partition $P_{lr}(s, t)$ eines Strings s bezüglich eines Strings t (siehe Abschnitt 3.8 im Skript) kann mit Hilfe eines Suffixbaumes effizient berechnet werden.

- Überlege dir einen Algorithmus, der die Links-Rechts-Partition $P_{lr}(s, t)$ in linearer Zeit berechnet. Gib die einzelnen Schritte deines Algorithmus explizit und verständlich an.
- Verwende diesen Algorithmus, um $P_{lr}(s, t)$ für $s = \text{TATCTCTTCT}$ und $t = \text{AATCATAT}$ zu berechnen.