

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2014/2015

Dr. Roland Wittler · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/SequenzAnalyse>

Übungsblatt 7 vom 25.11.2014

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Suffixbäume)

(4 Punkte)

1. Beschreibe zwei Anwendungen aus der bioinformatischen Praxis, für die man einen Suffixbaum verwenden kann.
2. Gegeben ist der String $s\$ = \text{BABBABAB\$}$, wobei $\$ < a < b$.
 - (a) Zeichne den Suffixbaum für s , sortiere dabei alle ausgehenden Kanten lexikographisch.
 - (b) Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in s . Die Indizierung beginnt bei 1.
 - (c) Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.

Aufgabe 2 (Kürzester eindeutiger Substring)

(4 Punkte)

1. Gib eine Definition des Problems des kürzesten eindeutigen Substrings in eigenen Worten an.
2. Beschreibe eine mögliche Anwendung des Problems.
3. Gib einen Linearzeit-Algorithmus an, mit dem man einen kürzesten eindeutigen Substring eines Strings s finden kann, wenn der Suffixbaum von $s\$$ gegeben ist.

Aufgabe 3 (Maximale Repeats)

(5 Punkte)

1. Finde alle maximalen Repeats in $s = \text{TATGTACCGTATAC}$ unter Verwendung des im Skript geschilderten Algorithmus (Abschnitt 7.6.3). Beschreibe dein Vorgehen beim Annotieren des Suffixbaumes.
2. **Satz:** In jedem String der Länge n gibt es höchstens n maximale Repeats.

Argumentiere unter Berücksichtigung des Suffixbaumes, warum die Aussage korrekt ist. Bedenke: Es stimmt nicht, dass an jeder Position nur ein maximales Repeat beginnen oder enden kann.

Aufgabe 4 (Maximal-Matches Distanz)

(2 Punkte)

Berechne die Links-Rechts-Partition von $t = \text{BABAABABB}$ bezüglich s aus Aufgabe 1. Inwiefern ist der Suffixbaum von s dazu nützlich? Welche Maximal-Matches Distanz $\delta(t||s)$ ergibt sich und warum?