

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SoSe 2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2012summer/SequenzAnalyse>

Übungsblatt 4 vom 04.05.2012

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (§ – Die Wache)

(1 Punkt)

Bei der Erstellung eines Suffixbaums wird am Ende des Strings ein eindeutiges Zeichen, z.B. $\$ \in \Sigma$, angehängt. Warum?

Aufgabe 2 (Verallgemeinerter Suffixbaum)

(5 Punkte)

Gegeben seien die Strings $s=GAAG$ und $t=AAGG$.

1. Zeichne den generalisierten Suffixbaum von s und t (mit $\# < \$ < A < G$).
2. Finde den längsten gemeinsamen Substring von s und t und gib seine Vorkommen an.
3. Formuliere allgemein, wie man längste gemeinsame Substrings zweier Strings im generalisierten Suffixbaum finden kann.
4. Nutze deine Überlegungen, um das längste palindromische Teilwort in $x = BANANAS$ zu finden.

Aufgabe 3 (Speicherung von Suffixbäumen)

(6 Punkte)

Gegeben ist der String $s\$ = ABAAB\$$, wobei $\$ < a < b$.

1. Zeichne den Suffixbaum für s , sortiere dabei die Blätter lexikographisch, nenne die inneren Knoten x, y und z .
2. Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in s . Die Indizierung beginnt bei 1.
3. Beschrifte jede Kante mit dem Start- und End-Index des dazugehörenden Strings. Wenn es mehrere Möglichkeiten gibt, nehme die größten Indizes.
4. Ist es notwendig, sowohl Start- als auch End-Index eines Substrings im Suffixbaum anzugeben? Erkläre!
5. Gib das Array zur Speicherung des Suffixbaums analog zu Seite 90 unten im Skript an.