

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 6 vom 18.05.2012

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Maximale Repeats)

(5 Punkte)

1. Finde alle maximalen Repeats in $s = \text{AGTGCAATGTGCAT}$ unter Verwendung des im Skript geschilderten Algorithmus (Abschnitt 9.7.4, auf der Homepage zur Übung befindet sich eine überarbeitete Version). 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 maximaler Repeat beginnen oder enden kann.

Aufgabe 2 (Suffixarray)

(3 Punkte)

Gegeben ist der String $s = \text{ATGTCACTGTACA}$. Gib das Suffixarray `pos`, sein Inverses `rank` und das `lcp`-Array von s an. Beachte $\$ < A < C < G < T$ und Indizierung beginnend mit 0.

Aufgabe 3 (Suffixarray Interpretation)

(2 Punkte)

Inwiefern ist das Suffixarray `pos` ein q -gram Index für alle $q \geq 1$ gleichzeitig?

Aufgabe 4 (Manber-Myers Algorithmus)

(4 Punkte)

Betrachte das Beispiel im Skript auf Seite 106, 107.

1. Wie viele Phasen braucht man für einen String der Länge 24?
2. Führe den Manber-Myer Algorithmus schrittweise für den String `GCGTGGCTCTCGC` aus.