

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2020

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga · Leonie R. Brockmann · Rebecca K. Pfeil

<https://gi.cebitec.uni-bielefeld.de/teaching/2020summer/sa>

Übungsblatt 7 vom 4.6.2020

Abgabe am 11.6.2020 bis 12:00 Uhr (mittags)

Aufgabe 1 (Dot-Plot) (3 Punkte)

Gegeben seien die beiden Sequenzen $x = \text{GATGCCTGGACTTGC}$ und $y = \text{AGGACTTGGATGCCTT}$. Erstelle die folgenden *dot plots* für x und y :

- ungefiltert;
- gefiltert mit $q = 3$;
- gefiltert mit $q = 6$.

Aufgabe 2 (BLAST) (6 Punkte)

Gegeben seien die Query $x = \text{RLMLLRML}$ und die Datenbanksequenz $y = \text{RRMLMRRLRLMLMLR}$. Das Alphabet ist $\Sigma = \{\text{L}, \text{M}, \text{R}\}$. Der Score für einen Match sei $+3$ und für einen Mismatch -1 . Indels sollen nicht möglich sein, haben also Score $-\infty$. Benutze $q = 4$ und $k = 8$.

- Gib die Nachbarschaft von x an.
- Gib die BLAST-Hits von x in y an.

Aufgabe 3 (Σ -Baum) (4 Punkte)

Gegeben sei die Sequenz $s = \text{REFEREE}$.

- Zeichne den kleinsten Σ -Baum T und den kleinsten kompakten Σ^+ -Baum T' , die alle Suffixe von s darstellen.
- Gib für T und T' jeweils die Menge der Worte $x \in \Sigma^*$ an, für die $\text{node}(x)$ definiert ist.
- Welche Menge $\text{words}(T)$ von Worten wird durch T dargestellt? Gibt es einen Unterschied zu $\text{words}(T')$?

Aufgabe 4 (Suffixbäume) (5 Punkte)

Gegeben sei die Sequenz $s = \text{TATTLETALE}$.

- Führe den WOTD-Algorithmus schrittweise zur Erstellung des Suffixbaums für die Sequenz $s\$$ aus. Sortiere dabei die Kanten lexikographisch (wobei $\$ < \text{A} < \text{E} < \text{L} < \text{T}$).
- 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.

Aufgabe 5 (Suffixbäume) (2 Punkte)

- Analysiere die *worst-case*- und *average-case*-Laufzeit des WOTD-Algorithmus in eigenen Worten.
- Zeige, dass der Speicherverbrauch eines Suffixbaums linear bezüglich der Eingabe ist.