

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2015/2016

Prof. Dr. Jens Stoye · M.Sc. Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015winter/SequenzAnalyse>

Übungsblatt 3 vom 10.11.2015

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Rank und Unrank)

(3 Punkte)

Gegeben seien das Alphabet $\Sigma = \{A, C, G, T\}$, die Ranking-Funktion r_Σ für Zeichen über dem Alphabet Σ

$$r_\Sigma(A) = 0, r_\Sigma(C) = 1, r_\Sigma(G) = 2, r_\Sigma(T) = 3,$$

sowie die aufsteigenden Ranking-Funktion

$$r_{up}(z) = \sum_{i=1}^q r_\Sigma(z[i]) \cdot \sigma^{i-1}.$$

Die Update-Funktion für die Ranking-Funktion findest du im Skript auf Seite 23 und 24.

Gib für die folgenden Berechnungen jeweils alle Zwischenschritte mit an.

1. Gegeben seien das Wort $x = GGC$ und $q = 3$. Berechne den Rank von x nach der aufsteigenden Ranking-Funktion.
2. Benutze nun die entsprechenden Update-Funktionen der Ranking-Funktionen, um den Rank des Wortes $y = GCA$ zu berechnen.
3. Gegeben seien der Rank 101 und $q = 4$. Berechne das Wort a der aufsteigenden Funktion, die diesem rank entsprechen.

Aufgabe 2 (Worte mit gleichem q -gram-Profil)

(3 Punkte)

Gegeben sei der String $x = \text{TACGTACGTGTA}$ und $q = 4$. Wie kannst du andere Strings finden, die zu x unterschiedlich sind, aber das gleiche q -gram-Profil haben? Gib alle Strings an, die diese Eigenschaften erfüllen.

Aufgabe 3 (Maximal-Matches-Distanz)

(4 Punkte)

Berechne die Maximal-Matches-Distanz δ und gib jeweils die links-nach-rechts und die rechts-nach-links Partitionen an.

1. $\delta(\text{vancouver} \parallel \text{thevaniscovered})$
2. $\delta(\epsilon \parallel \text{ababcc})$
3. $\delta(\text{ababcc} \parallel \epsilon)$

Welche Beobachtung machst du bei 2. und 3.? Was sagt das über die Maximal-Matches-Distanz aus?

Bitte wenden.

Aufgabe 4 (Globales Alignment)

(4 Punkte)

Wir definieren die folgenden Scores für die Editoperationen Insertion und Deletion: $\mathcal{I} = \mathcal{D} = -1$. Für Copy (\mathcal{C}) und Substitution (\mathcal{S}) gilt die Scorematrix $\mathcal{S}_{a,c}$:

	A	G	C
A	+2	-1	-1
G	-1	+2	-1
C	-1	-1	+2

1. Zeichne den globalen Alignmentgraphen für $x = \text{AG}$ und $y = \text{AC}$ mit *allen* Kanten und Knoten.
2. Füge deiner Zeichnung die Kantenbeschriftungen λ und die Kantengewichte w hinzu.
3. Berechne die Knotenwerte $S(v)$.
4. Gib ein optimales globales Alignment an und markiere dessen Pfad im Graphen.