

# 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 2 vom 21.10.2014

Abgabe in einer Woche vor Beginn der Vorlesung.

### Aufgabe 1 (Rank und Unrank)

(6 Punkte)

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

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

sowie die absteigenden und die aufsteigenden Ranking-Funktionen

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

Die Update-Funktionen für beide Ranking-Funktionen findest du im Skript auf Seite 23 und 24.

Gib für die folgenden Berechnungen jeweils alle Zwischenschritte mit an.

1. Gegeben seien das Wort  $x = ATG$  und  $q = 3$ . Berechne den Rank von  $x$  nach der absteigenden und der aufsteigenden Ranking-Funktion.
2. Benutze nun die entsprechenden Update-Funktionen der Ranking-Funktionen, um den Rank des Wortes  $y = TGC$  zu berechnen.
3. Gegeben seien der Rank 100 und  $q = 4$ . Berechne das Wort  $a$  und das Wort  $b$  nach der absteigenden und der aufsteigenden Funktion, die diesem rank entsprechen.

### Aufgabe 2 ( $q$ -gram-Profil und de Bruijn-Graph)

(5 Punkte)

Gegeben seien das Wort  $x = AATTAAATAA$  und  $q = 4$ .

1. Erstelle das vollständige  $q$ -gram-Profil für  $x$ .
2. Wann ist ein  $q$ -gram-Profil *dicht*? Handelt es sich bei dem  $q$ -gram-Profil von  $x$  um ein dichtes Profil?
3. Zeichne den de Bruijn-Graphen für  $x$ .
4. Gibt es ein Wort  $x' \neq x$ , das eine Distanz von 0 zu  $x$  hat? Wenn ja, nenne es. Wenn nein, begründe deine Antwort.
5. Diskutiere, ob es sich bei der  $q$ -gram-Distanz um eine Metrik handelt.

### Aufgabe 3 (Maximal-Matches Distanz)

(4 Punkte)

Berechne  $\delta$  und gib jeweils die links-nach-rechts und rechts-nach-links Partitionen an.

1.  $\delta(\text{manhattan} \parallel \text{hatharlemen})$
2.  $\delta(\epsilon \parallel \text{aaaa})$
3.  $\delta(\text{aaaa} \parallel \epsilon)$

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

### Aufgabe 4 (Filter)

(4 Punkte)

1. Was ist mit *Filtern* im Zusammenhang mit Sequenzvergleichen gemeint? Warum verwendet man Filter?
2. Welche Distanzmodelle kann man beim Filtern der Edit-Distanz benutzen?
3. Worin unterscheidet sich eine Heuristik von einem Filter?