

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2016/2017

Prof. Dr. Jens Stoye · M.Sc. Tizian Schulz

<https://gi.cebitec.uni-bielefeld.de/teaching/2016winter/sa>

**Übungsblatt 2 vom 03.11.2016**

**Abgabe am 15.11.2016 vor Beginn der Vorlesung**

## Aufgabe 1 (Berechnung der Edit-Distanz)

(5 Punkte)

Gegeben seien die Strings  $x = \text{ERBSE}$  und  $y = \text{ESCHE}$ . Benutze zur Berechnung der Edit-Distanz der beiden Strings eine Edit-Matrix und gib die Distanz an. Benutze außerdem eine Backtracing-Matrix, um alle optimalen Edit-Sequenzen bestimmen zu können. Schreibe eine optimale Edit-Sequenz explizit auf.

## Aufgabe 2 ( $q$ -gram-Profil)

(4 Punkte)

Gegeben seien die Wörter  $x = 0101010$  und  $y = 101011010$ , sowie  $q = 3$ .

1. Erstelle das vollständige  $q$ -gram-Profil für  $x$  und  $y$ .
2. Gib die  $q$ -gram-Distanz von  $x$  und  $y$  an.
3. Diskutiere, ob es sich bei der  $q$ -gram-Distanz um eine Metrik handelt.

## Aufgabe 3 (Worte mit gleichem $q$ -gram-Profil)

(3 Punkte)

Gegeben sei der String  $x = \text{CATATGCATGCA}$  und  $q = 3$ . 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 4 (Maximal-Matches Distanz)

(4 Punkte)

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

1.  $\delta(\text{montreal} \parallel \text{monmarelen})$
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 5 (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?