

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2017

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

<https://gi.cebitec.uni-bielefeld.de/teaching/2017summer/sa>

**Übungsblatt 3 vom 11.05.2017**

**Abgabe 23.05.2017**

## Aufgabe 1 (Metriken auf Sequenzen)

(12 Punkte)

Gegeben sei ein Alphabet aus Zahlen  $\Sigma = \{1, 2, 3\}$  der Größe  $\sigma = 3$  und die beiden Sequenzen  $s_1 = 321321232$  und  $s_2 = 132132323$ . Berechne von  $s_1$  und  $s_2$ :

1. Die Manhattan Distanz.
2. Die euklidische Distanz.
3. Die Maximum Metrik.
4. Die Hamming Distanz.
5. Die Edit-Distanz, gib die Edit-Matrix  $D$  und alle optimalen Edit-Sequenzen mit an.
6. Die  $q$ -gram Distanz für  $q = 3$ , gib außerdem die  $q$ -gram Profile von  $s_1$  und  $s_2$  an. Gibt es einen String  $s_3 \neq s_2$ , der für  $q = 3$  eine Distanz von 0 zu  $s_2$  hat? Wenn ja, gib ihn an, wenn nein, begründe deine Antwort.
7. Die Maximal Matches Distanz von  $s_1$  zu  $s_2$ , gib die Partitionierung mit an.

## Aufgabe 2 (Filter)

(3 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? Und wie verhalten sich diese zur Edit-Distanz?
3. Worin unterscheidet sich eine Heuristik von einem Filter?