

# Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WS 2011/2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/SequenzAnalyse>

## Übungsblatt 6 vom 09.12.2011

Abgabe in einer Woche vor Beginn der Vorlesung.

### Aufgabe 1 (Backtracing)

(4 Punkte)

Erweitere dein Programm von Zettel 4 um den Backtracing-Schritt. Das heißt, dein Programm soll nicht mehr nur die Edit-Distanz und die dazugehörige Edit-Matrix ausgeben, sondern auch eine optimale Edit-Sequenz. Als Alternative zu deinem eigenem Programm von Zettel 4 kannst du auch das Programm auf der Homepage zur Übung erweitern.

### Aufgabe 2 Maximal-Matches-Distanz, Eigenschaften einer Metrik

(6 Punkte)

Gegeben sind die Strings  $x = \text{AGTTTGACATTATG}$ ,  $y = \text{TTGCATT}$  und  $z = \text{TTAGTATTACATT}$ .

1. Berechne die links-rechts- und die rechts-links-Partition sowohl von  $\delta(x||y)$  als auch von  $\delta(y||x)$ . Was fällt dir dabei auf?
2. Zeige für  $x$ ,  $y$  und  $z$ , dass die Dreiecksungleichung für die Maximal-Matches-Distanz gilt.
3. Handelt es sich bei der Maximal-Matches-Distanz um eine Metrik? Begründe deine Antwort.
4. Berechne die Edit-Distanz mit Einheitskosten aller Paare aus  $x$ ,  $y$  und  $z$ . Hier ist wirklich nur die Edit-Distanz gefordert.
5. Welcher Zusammenhang besteht zwischen Maximal-Matches-Distanz und der Edit-Distanz mit Einheitskosten?

### Aufgabe 3 Filter

(3 Punkte)

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