

# Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

## Präsenzübungsblatt 2, Woche 17/2018

### Aufgabe 1 (Edit-Matrix)

Durch Einzeichnen von Strichen in die Edit-Matrix  $D(x, y)$ , können auch ohne die Backtracing-Matrix  $E(x, y)$  leicht alle optimalen Edit-Sequenzen mittels Backtracking gefunden werden.

1. Berechne die Edit-Matrix  $D(x, y)$  für die Sequenzen  $x = \text{GTGCCTAC}$  und  $y = \text{CTGCCCTTTAC}$  und gib die Matrix, sowie die Edit-Distanz  $d(x, y)$  der beiden Strings an.
2. Gib *alle* optimalen Edit-Sequenzen an.

### Aufgabe 2 (Rank und Unrank)

Gegeben sind das Alphabet  $\Sigma = \{\text{A, C, G, T}\}$  und die Wortlänge  $q = 4$ .

1. Informiere dich zunächst im Skript über *rank* und *unrank* (S. 24 ab Definition 3.14 und S. 25).
2. Berechne den Rang des Wortes **GACC** und danach (ohne vollständige Neuberechnung, sondern durch ein Update in konstanter Zeit) den Rang des Wortes **ACCA**. Verwende die absteigende Variante der Codierung von  $q - 1$  nach 0. Gib alle Zwischenschritte an.
3. Welches Wort hat den Rang 137 unter Verwendung der absteigenden Codierung?