

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2017

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

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

Präsenzübungsblatt 3, Woche 19/2017

Aufgabe 1 (Edit-Matrix 1)

1. Berechne die Edit-Matrix $D(x, y)$ für die Sequenzen $x = \text{TAKTIK}$ und $y = \text{STATISTIK}$ und gib die Matrix, sowie die Edit-Distanz $d(x, y)$ der beiden Strings an.
(Schreibe x vertikal links neben die Matrix und y oben horizontal an die Matrix.)
Erstelle während der Berechnung von $D(x, y)$ auch die Backtracing-Matrix $E(x, y)$ und gib diese ebenfalls an.
2. Verwende $E(x, y)$, um *alle* optimalen Edit-Sequenzen zu finden und gib sie an.

Aufgabe 2 (Edit-Matrix 2)

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 gefunden werden. Wie das funktioniert, kann dir dein Tutor erklären.

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 3 (Rank und Unrank)

Gegeben sind das Alphabet $\Sigma = \{\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}$ 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?