

Präsenzübungen 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>

Präsenzübungsblatt 2, Woche 44/2016

Aufgabe 1 (Edit-Matrix 1)

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

Aufgabe 3 (Laufzeit der q -gram-Distanz)

Begründe, warum die Berechnung der q -gram-Distanz von zwei Strings in linearer Zeit berechnet werden kann. Gehe dafür die Schritte durch, die notwendig sind, um die Distanz zu berechnen.