

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2022

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa>

## Übungsblatt 2 vom 14.4.2022

Abgabe am 21.4.2022 bis 12:00 Uhr (mittags)

### Aufgabe 1 (*Edit-Operationen*)

(2 Punkte)

Berechne den String

$$S = E(\text{KMPEMKL}, \text{CD S}_R \text{C S}_N \text{I}_A \text{C D}).$$

Gib alle Zwischenschritte an. Wie hoch sind die (Einheits-)Kosten der *Edit*-Sequenz?

### Aufgabe 2 (*Edit-Sequenzen und Berechnung der Edit-Distanz*)

(10 Punkte)

Gegeben seien die Strings  $x = \text{ACGGGC}$  und  $y = \text{TCGGT}$ . Betrachte Einheitskosten.

1. *Edit*-Sequenz mit maximalen Kosten:

- Gib eine *Edit*-Sequenz mit maximalen Kosten an, die  $x$  in  $y$  überführt:
- Wie viele *Edit*-Sequenzen mit maximalen Kosten gibt es?
- Sind die *Edit*-Sequenzen mit maximalen Kosten immer die längsten?

2. *Edit*-Distanz und optimale *Edit*-Sequenz (mit minimalen Kosten):

- Benutze zur Berechnung der *Edit*-Distanz der beiden Strings eine *Edit*-Matrix und gib die Distanz an. Benutze außerdem eine weitere Matrix, in der die optimalen *Edit*-Operationen gespeichert werden, um alle optimalen *Edit*-Sequenzen bestimmen zu können.
- Schreibe eine optimale *Edit*-Sequenz explizit auf.
- Wie viele optimale *Edit*-Sequenzen gibt es?
- Sind die optimalen *Edit*-Sequenzen immer die kürzesten?

### Aufgabe 3 (*Edit-Distanzen*)

(8 Punkte)

Die Rekurrenz zur Berechnung der *Standard-Edit*-Distanz mit Einheitskosten lautet für  $1 \leq i \leq |x|, 1 \leq j \leq |y|$ :

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + \mathbb{1}_{\{x[i] \neq y[j]\}} \\ D(i-1, j) + 1 \\ D(i, j-1) + 1 \end{cases}$$

Die Rekursionsbasis ist gegeben durch:

$$D(0, j) = j \text{ für } 0 \leq j \leq |y| \text{ und } D(i, 0) = i \text{ für } 0 \leq i \leq |x|$$

- Wie sehen die Rekurrenz und die Basisfälle für die *Hamming-Distanz* aus?
- Wie sehen die Rekurrenz und die Basisfälle für die *LCS-Distanz* aus?
- Ist es möglich, die *Edit+Flip-Distanz* rekursiv zu berechnen? Wenn nein, warum nicht? Wenn ja, gib eine Formel und eine Rekursionsbasis an. Welche Schwierigkeiten könnten zudem hierbei auftreten?