

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2024

Prof. Dr. Jens Stoye · Leonard Bohnenkämper

<https://gi.cebitec.uni-bielefeld.de/teaching/2024winter/sa>

Übungsblatt 2 vom 24.10.2024

Abgabe am 31.10.2024 bis 10:00 Uhr (morgens)

Aufgabe 1 (*Edit-Operationen*)

(2 Punkte)

Berechne den String

$$S = E(\text{EDITDISTANCE}, \text{DDCDCS}_E\text{I}_A\text{CDDDDDD}).$$

Gib die ersten 8 Zwischenschritte an. Wie hoch sind die (Einheits-)Kosten der *Edit*-Sequenz?

Aufgabe 2 (*Edit-Sequenz und -Distanz*)

(5 Punkte)

Überführe den String $x = \text{ALLHALLOWSEVE}$ in den String $y = \text{HALLOWEEN}$ durch Anwendung der *Edit*-Funktion.

1. Gib eine *Edit*-Sequenz mit minimalen Kosten an, die x in y überführt. Wie viele *Edit*-Sequenzen mit minimalen Kosten gibt es? Gibt es immer eine *Edit*-Sequenz, die sowohl die geringsten Kosten als auch die kürzeste Länge unter allen *Edit*-Sequenzen hat?
2. Gib eine *Edit*-Sequenz mit *maximalen* Kosten an, die x in y überführt. Sind die *Edit*-Sequenzen mit maximalen Kosten immer die längsten? Wie viele *Edit*-Sequenzen mit maximalen Kosten gibt es?

Aufgabe 3 (*Berechnung der Edit-Distanz*)

(5 Punkte)

Gegeben seien die Strings $x = \text{BANANE}$ und $y = \text{TANNE}$. 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.

Aufgabe 4 (*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:

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

1. Wie sehen die Rekurrenz und die Basisfälle für die *Hamming-Distanz* aus?
2. Wie sehen die Rekurrenz und die Basisfälle für die *LCS-Distanz* aus?
3. 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?