

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2017

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

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

Übungsblatt 2 vom 02.05.2017

Abgabe am 11.05.2017 vor Beginn der Vorlesung

Aufgabe 1 (Alphabete)

(4 Punkte)

Gegeben sei das Alphabet $\Sigma = \{b, g\}$.

1. Wie ist Σ^6 definiert?
2. Schreibe alle Worte $w \in \Sigma^4$ auf.
3. Wie sind Σ^* und Σ^+ definiert und was ist der Unterschied zwischen ihnen?
4. Was ist der Unterschied zwischen ε , $\{\varepsilon\}$ und $\{\}$?

Aufgabe 2 (Edit-Operationen)

(4 Punkte)

Berechne den String $S = E(ITNZKAOLM, \mathcal{I}_S \mathcal{D} \mathcal{S}_T \mathcal{I}_O \mathcal{S}_C \mathcal{D} \mathcal{C} \mathcal{I}_H \mathcal{S}_O \mathcal{D} \mathcal{C} \mathcal{C})$, gib alle Zwischenschritte an. Wie hoch sind die Kosten der Edit-Sequenz? Gib eine weitere Edit-Sequenz mit maximalen Kosten an, die den gleichen String S ergibt.

Aufgabe 3 (Edit-Sequenz und -Distanz)

(5 Punkte)

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

1. Gib eine möglichst kurze *Edit*-Sequenz an, die x in y überführt.
2. Gib drei der längsten *Edit*-Sequenzen an, die x in y überführen.
3. Wie hoch ist die *Edit*-Distanz bei deinen Lösungen?

Aufgabe 4 (Edit-Distanzen)

(3 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?

Aufgabe 5 (Rekursive Berechnung der Edit-Distanz)

(5 Punkte)

Implementiere eine Funktion zur rekursiven Berechnung der Edit-Distanz in Haskell. Benutze hierfür die Rekurrenzformel aus Aufgabe 4. Teste deine Implementation mit verschiedenen Eingabegrößen (Sequenzlängen). Wie verändert sich dabei die Laufzeit? Wie lässt sich dies erklären?