

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2013/2014

Prof. Dr. Jens Stoye · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/SequenzAnalyse>

Übungsblatt 2 vom 29.10.2013

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Edit-Matrix)

(5 Punkte)

Die Edit-Distanz kann effizient mit Hilfe der Edit-Matrix berechnet werden.

1. Berechne die Edit-Matrix $D(x, y)$ für die Sequenzen $x = \text{FEIIER}$ und $y = \text{EIMER}$ 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 (Dynamische Programmierung)

(5 Punkte)

1. Schreibe ein Programm, das als Eingabe zwei Strings x und y bekommt und deren Edit-Distanz ausgibt. Die Edit-Distanz sollst du (a) einmal rekursiv und (b) einmal mittels dynamischer Programmierung bestimmen. Gib außerdem die Zeit aus, die jeweils für (a) und (b) benötigt wird.
2. Teste dein Programm nun mit Strings der Länge 5, 10 und 15. Trage die Zeiten, die benötigt werden, grafisch gegeneinander auf. (Das kannst du per Hand machen.)
3. Warum ist es ineffizient, die Edit-Distanz rekursiv zu berechnen?

Aufgabe 3 (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?