

# Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, Wintersemester 2010/2011  
Dipl.-Inform Peter Husemann · Dr. Roland Wittler

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010winter/SequenzAnalyse>

**Blatt 4 vom 05.11.2010**

**Abgabe in einer Woche vor Beginn der Vorlesung.**

## Aufgabe 1 Edit-Matrix

(3 Punkte)

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

1. Berechne von Hand die Edit-Matrix  $D(x, y)$  für die Sequenzen  $x = \text{ANANAS}$  und  $y = \text{ANNASHASE}$ .
  - Gib die Matrix, sowie die Edit-Distanz  $d(x, y)$  der beiden Strings an. (Zur einfacheren Korrektur schreibe  $x$  links vertikal neben die Matrix und  $y$  oben horizontal an die Matrix.)
  - Erstelle während der Berechnung von  $D(x, y)$  auch (separat) 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 Programmieraufgabe

(5 Punkte)

Schreibe eine Funktion in Java, welche zwei Strings  $x$  und  $y$  als Eingabe akzeptiert und *eine* der optimalen Edit-Sequenzen ausgibt, die  $x$  in  $y$  überführt.

(Bei dieser Aufgabe wird auch der abgegebene Code bewertet. Ist die Funktion sauber und verständlich geschrieben? Sind wichtige Schritte dokumentiert? Werden Sonderfälle oder falsche Eingaben passend behandelt?)

## Aufgabe 3 Edit-Distanzen

(3 Punkte)

Die Rekurrenz zur Berechnung der Standard-Edit-Distanz mit Einheitskosten lautet

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + \mathbb{1}_{\{x[i] \neq y[j]\}} & \text{die Rekursionsbasis ist gegeben durch:} \\ D(i-1, j) + 1 & D(0, 0) = 0 \\ D(i, j-1) + 1 & D(0, j) = j \quad \text{für } 1 \leq j \leq |y| \quad \text{und} \\ & D(i, 0) = i \quad \text{für } 1 \leq i \leq |x| \end{cases}$$

für  $1 \leq i \leq |x|, 1 \leq j \leq |y|$ ,

1. Wie müssen die Rekurrenz und die Basisfälle geändert werden, falls die *Hamming-Distanz* berechnet werden soll?
2. Wie würden die Rekurrenz und die Basisfälle bei der *LCS-Distanz* aussehen?
3. Ist es möglich, die *Editflip-Distanz* rekursiv zu berechnen? Welche Schwierigkeiten können dabei auftreten?