

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WS 2011/2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/SequenzAnalyse>

Übungsblatt 4 vom 25.11.2011

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Implementierung der Edit-Distanz)

(6 Punkte)

Implementiere in Java ein Programm zur Berechnung der Edit-Distanz zweier Sequenzen der Länge m und n .

- Schreibe eine Funktion, welche die Edit-Distanz iterativ berechnet, so wie du es auf dem letzten Zettel mit der Edit-Matrix gemacht hast. Diese Variante nennt man auch *backward dynamic programming*. Deine Funktion soll neben der Edit-Distanz auch die Edit-Matrix ausgeben.
- Schreibe eine weitere Funktion, welche die Edit-Distanz rekursiv berechnet. Gebe neben der Edit-Distanz auch die Anzahl an Rekursionsschritten an.
- Messe für beide Funktionen die Laufzeit bei unterschiedlichen Eingabesequenzen der Länge m und n . Stelle deine gemessenen Werte in Abhängigkeiten von $m + n$ graphisch dar.

Achtung! Beginne mit dem Vergleichen von kurzen Sequenzen und beobachte die Laufzeit deines Programms.

Aufgabe 2 (Rekurrenzen für Distanz-Modelle)

(4 Punkte)

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

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

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 ja, gib die Rekurrenz und die Basisfälle an. Welche Schwierigkeiten treten hierbei auf?

Aufgabe 3 (Backtracing und Backtracking)

(2 Punkte)

Erkläre den Unterschied zwischen backtracing und backtracking bei der Rekonstruktion der optimalen Edit-Sequenz.