

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WiSe 2009/2010

Prof. Dr. Jens Stoye · Dipl.-Inform. Nils Hoffmann

<http://wiki.techfak.uni-bielefeld.de/gi/GILectures/2009winter/SequenzAnalyse>

Blatt 6 vom 26.11.2009

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 Partitionen

(3 Punkte)

Gib jeweils die links-nach-rechts und rechts-nach-links Partition von x bezüglich y und deren Länge an.

1. $x = \text{babbabbabaaa}$ und $y = \text{bab}$
2. $x = \text{bb}$ und $y = \text{cbbba}$
3. $x = \text{dca}$ und $y = \text{baab}$

Aufgabe 2 Maximal-Matches Distanz

(3 Punkte)

Berechne δ und gib jeweils die links-nach-rechts und rechts-nach-links Partitionen an:

1. $\delta(\text{xylophon} \parallel \text{colophonon})$
2. $\delta(\epsilon \parallel \text{bbbb})$
3. $\delta(\text{bbbb} \parallel \epsilon)$

Aufgabe 3 Maximal-Matches Distanz, Eigenschaften einer Metrik

(4 Punkte)

Gegeben sind die Strings $x = \text{babbabbabaaa}$, $y = \text{bab}$ und $z = \text{ababbabbabaaa}$.

1. Zeige für x , y und z , dass die Dreiecksungleichung für die Maximal-Matches Distanz gilt.
2. Berechne die Edit-Distanz mit Einheitskosten aller Paare aus x , y und z .
3. Welcher Zusammenhang besteht zwischen Maximal-Matches Distanz und der Edit-Distanz mit Einheitskosten?

Aufgabe 4 Filter

(3 Punkte)

1. Was ist mit *Filtern* im Zusammenhang mit Sequenzvergleichen gemeint? Warum verwendet man Filter?
2. Welche Distanzmodelle kann man beim Filtern benutzen?
3. Worin unterscheidet sich eine Heuristik von einem Filter?