

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 10 vom 17.12.2010
Abgabe 14.01.2011 vor der Vorlesung

Aufgabe 1 Affine Gapkosten

(5 Punkte)

1. Warum sollten die Kosten für eine *gap extension* nicht höher als die Kosten für ein *gap open* gewählt werden, also $e \leq d$?
2. Zeige, dass affine Gapkosten subadditiv sind.
3. Berechne ein optimales globales Alignment mit affinen Gapkosten von den Sequenzen $x = \text{AAGTTGCC}$ und $y = \text{AACC}$ effizient mit Hilfe des Gotoh-Algorithmus (berechne die Matrizen S , H und V) und gib dessen Gesamtscore an. Verwende dabei: Score für Match = 2, Score für Mismatch = 0, Kosten für Gap-open $d = 1$, sowie Kosten für Gap-extension $e = 0.5$.

Aufgabe 2 Suboptimale Alignments

(4 Punkte)

1. Beschreibe in eigenen Worten, was man unter *überlappenden* Alignments versteht.
2. Warum ist man bei der Bestimmung von suboptimalen Alignments in erster Linie an *nichtüberlappenden* Alignments interessiert?
3. Beschreibe kurz die Funktionsweise des Waterman-Eggert-Algorithmus. Durch welchen Trick kann man in der Praxis die Laufzeit verkürzen?

Aufgabe 3 Approximatives Stringmatching

(5 Punkte)

1. Finde die Endpositionen aller Vorkommen des Patterns $x = \text{TATA}$ im Text $y = \text{ACTATAGCATA}$ mit maximal $k = 1$ Fehlern. Verwende die Cutoff-Variante von *Sellers' Algorithmus* mit Einheitskosten und markiere die *last essential indices*.
2. Gib zu jeder gefundenen Endposition alle zugehörigen Alignments an.
3. Warum macht es Sinn, in der Praxis nicht *alle* gefundenen Endpositionen auszugeben? (Stichwort: *Runs*.) Welche Endpositionen würde man im obigen Beispiel nicht ausgeben?