

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 9 vom 10.12.2010

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 Globales Alignment

(4 Punkte)

Wir definieren die folgenden Scores für die Editoperationen Insertion und Deletion: $\mathcal{I}_c = \mathcal{D}_c = -1$. Für Copy (\mathcal{C}) und Substitution (\mathcal{S}) gilt die Scorematrix $S_{a,c}$:

	A	B	C
A	+2	-1	-1
B	-1	+2	-1
C	-1	-1	+2

1. Zeichne den globalen Alignmentgraphen für $x = BC$ und $y = AC$ mit *allen* Kanten und Knoten.
2. Füge deiner Zeichnung die Kantenbeschriftungen λ und die Kantengewichte w hinzu.
3. Berechne die Knotenwerte $S(v)$.
4. Gib ein optimales globales Alignment an und markiere dessen Pfad im Graphen.

Aufgabe 2 Lokales Alignment

(5 Punkte)

Gegeben sind die Sequenzen $x = TCAATGCA$ und $y = TACCAACGTG$, und folgende Scorefunktion: +2 Punkte für ein Match, -3 für ein Mismatch, -2 für ein Indel.

1. Beschreibe (in Worten und mithilfe einer groben Skizze) den Alignmentgraphen von x und y für die Berechnung eines lokalen Alignments mithilfe des Smith-Waterman Algorithmus.
2. Berechne den Alignment-Score und gib ein optimales lokales Alignment an. Verwende dazu nicht den Graphen, sondern stattdessen eine Matrix, ähnlich der Editmatrix.

Aufgabe 3 Alignmentgraph formal

(5 Punkte)

Gegeben sei eine (kurze) Sequenz x , und eine andere (lange) Sequenz y . Gesucht ist ein bezüglich Edit-Distanz (mit Einheitskosten) optimales Vorkommen von ganz x irgendwo in y .

1. Wie sieht der Alignment-Graph zu diesem Problem aus? Beschreibe genau, welche Kanten der Graph enthält und welche Kosten diese haben. (Gib die Kanten und Kosten formal an. Eine Skizze ist hier nicht ausreichend und auch nicht gefordert.)
2. Gib eine explizite Rekursionsvorschrift für $S(v)$ an, wie sie für die anderen Alignment-Typen in der Vorlesung angegeben wurden (vgl. Tabellen 5.1–5.3 im Skript). Achtung: Beachte, dass hier anstelle einer Scorefunktion eine Kostenfunktion (Edit-Distanz) verwendet werden soll.