

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019

Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

Übungsblatt 4 vom 29.4.2019

Abgabe am 6.5.2019 bis 12:00 Uhr (mittags)

Aufgabe 1 (Maximal-Matches-Distanz als Filter)

(4 Punkte)

Entscheide mithilfe der Maximal-Matches-Distanz, ob die Sequenzen y_1 bis y_4 eine Edit-Distanz von max. 1 zur Sequenz x haben können. Warum?

$x = \text{AATCGAGGTAC}$

$y_1 = \text{AAGATCGGACC}$

$y_2 = \text{AATCGCGGTAC}$

$y_3 = \text{AATCGGGAGGT}$

$y_4 = \text{CTGAACGTCTG}$

Aufgabe 2 (Globales Alignment)

(4 Punkte)

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

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

1. Zeichne den globalen Alignmentgraphen für $x = \text{BC}$ und $y = \text{AB}$ 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 3 (Wiederholung zu Metriken auf Sequenzen)

(4 Punkte)

Beantworte kurz folgende Fragen zu Kapitel 3 des Skripts:

- Warum ist die Maximal-Matches-Distanz keine Metrik? Wie kann man aus ihr eine Metrik machen?
- Gegeben das q -Gramm Profil zu einem Wort x . Wie würde das q -Gramm Profil von \overleftarrow{x} aussehen?
- Was sind die asymptotischen Laufzeiten zur Berechnung der Edit-Distanz $d(x, y)$, der q -Gramm Distanz $d_q(x, y)$ und der Maximal Matches Distanz $\delta(x||y)$ zweier Sequenzen x und y und wie kommen sie zustande?

Aufgabe 4 (Längste gemeinsame Teilsequenz)

(3 Punkte)

Wie könnte man eine längste gemeinsame Teilsequenz zweier Sequenzen bestimmen?