

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

Übungsblatt 6 vom 17.05.2018

Abgabe 24.05.2018

Aufgabe 1 (Approximative Textsuche)

(5 Punkte)

1. Finde die Endpositionen aller Vorkommen des Musters $x = \text{TGGC}$ im Text $y = \text{TTGACTGCC}$ mit maximal $k = 1$ Fehlern, Gapkosten $\lambda = 0.5$, Substitutionskosten $\mathcal{S} = 1$ und $\mathcal{C} = 0$. Verwende die Cutoff-Variante von Sellers Algorithmus und markiere die *letzten essentiellen Indizes*.
2. Gib zu jeder gefundenen Endposition, die den Schwellwert k erfüllt, alle zugehörigen Alignments an.
3. Warum macht es Sinn, in der Praxis nicht *alle* gefundenen Endpositionen auszugeben? Welche Endpositionen würde man im obigen Beispiel nicht ausgeben?

Aufgabe 2 (Forward-Backward Technik)

(4 Punkte)

Gegeben sind die Sequenzen $s = \text{TATGACGTC}$ und $t = \text{TGAGATGC}$. Berechne, jeweils mit Einheitskosten,

- die Edit-Matrix D ,
- die Edit-Matrix D^{rev} der reversen Strings,
- die Gesamtkostenmatrix T und
- die Zusatzkostenmatrix C .

Gib ein optimales globales Alignment von s und t an und markiere es in D und C .

Aufgabe 3 (Paarweises Alignment mit linearem Speicherbedarf)

(4 Punkte)

Zeige anhand der Sequenzen aus Aufgabe 2, wie das optimale Alignment berechnet werden kann, während nur linearer Speicherplatz benötigt wird. Simuliere dazu die Schritte des in der Vorlesung besprochenen Divide-and-Conquer Verfahrens auf Kopien der Edit-Matrix. (Es kann teilweise auf die in Aufgabe 2 berechneten Werte zurückgegriffen werden.)

1. Markiere in einer neuen, leeren Matrix der entsprechenden Größe, für welche Bereiche D und D^{rev} berechnet werden würden bzw. berechne die Werte wenn nötig neu und trage sie ein.
2. Trage die Werte von T für die mittlere Zeile $m' = \lceil m/2 \rceil$ ein.
3. Markiere ein Minimum in T und die entsprechende Spalte, an der die Matrix geteilt wird.
4. Führe rekursiv die oberen Schritte für die beiden Teilmatrizen aus bis jeweils nur noch eine Zeile übrig ist.

Aufgabe 4 (BLAST)

(3 Punkte)

Informiere Dich genauer über BLAST (Skript S. 60/61).

1. Welche Rolle spielt der *X-Drop-Algorithmus* und wie funktioniert er?
2. Was sind die Unterschiede zwischen BLAST Version 1.4 und den Versionen ab 2.0?
3. Was bedeutet der Begriff *Sensitivität* im Kontext eines Matches?