

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

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

Übungsblatt 10 vom 21.06.2018

Abgabe 28.06.2018

Aufgabe 1 (Laufzeit-Abschätzung)

(5 Punkte)

Gegeben sind k Sequenzen der Länge $n = 1000$. Die Berechnung eines exakten paarweisen Alignments mittels *dynamic programming* dauert ca. zwei Sekunden. Ein Tag hat 86400 Sekunden und ein Jahr im Durchschnitt 365,2425 Tage.

1. Wie viele paarweise Alignments dieser Form kann man berechnen, bevor in 5 Milliarden Jahren die Sonne zu einem roten Riesen wird?
2. Wie viele solcher Sequenzen kann man mit einem direkten multiplen Alignment ($\mathcal{O}(2^k \cdot n^k)$) in dieser Zeit alignieren, wenn man dabei annimmt, dass ein multiples Alignment der Länge $n = 1000$ für $k = 2$ zwei Sekunden benötigt?
3. Erkläre die Laufzeit des direkten multiplen Alignments.

Aufgabe 2 (Center-Star-Approximation)

(5 Punkte)

Gegeben sind die Sequenzen $s_1 = \text{GACCC}$, $s_2 = \text{GACT}$ und $s_3 = \text{GCTC}$. Benutze für deine folgenden Berechnungen Einheitskosten.

1. Berechne die *Center-Sequenz* s_c .
2. Erstelle das multiple Alignment A_c und gib seine Sum-of-Pairs-Kosten an.
3. Was kannst du über die Optimalität des gefundenen Alignments sagen?
4. Beschreibe in eigenen Worten die Laufzeit- und Speicherkomplexität der Center-Star-Approximation. Unterscheide dabei zuerst die einzelnen Phasen und erkläre dann das Gesamtergebnis.

Aufgabe 3 (Distanzen und Metriken)

(4 Punkte)

Gegeben sind die zwei Strings $x = \text{ATTT}$ und $y = \text{ATTA}$. Berechne

- die Hamming-Distanz $d_H(x, y)$,
- die q -gram-Distanz $d_2(x, y)$ für $q = 2$,
- die Maximal-Matches-Distanzen $\delta(x||y)$ und $\delta(y||x)$,
- die Distanz mit der Metrik, die aus der Maximal-Matches-Distanz hervorgeht, $d_{||}(x, y)$ (siehe Theorem 3.25, Seite 29).