

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2014/2015

Dr. Roland Wittler · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/SequenzAnalyse>

Übungsblatt 10 vom 16.12.2014

Abgabe 06.01.2015

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?
3. Erkläre die Laufzeit des direkten multiplen Alignments.

Aufgabe 2 (Carrillo-Lipman-Heuristik)

(4 Punkte)

1. Wie viele Carrillo-Lipman-Schranken $U_{x,y}$ müssen berechnet werden, um k Sequenzen zu alignieren? Wie ist die asymptotische Laufzeit für die Berechnung der Schranken?
2. Für welche Sequenzen funktioniert die Carrillo-Lipman-Heuristik gut bzw. schlecht?

Aufgabe 3 (Center-Star-Approximation)

(5 Punkte)

Gegeben sind die Sequenzen $s_1 = \text{AACTG}$, $s_2 = \text{AGTG}$ und $s_3 = \text{CACG}$. 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. Gibt es ein multiples Alignment mit geringeren Sum-of-Pairs-Kosten?
4. Beschreibe in eigenen Worten die Laufzeit- und Speicherkomplexität der Center-Star-Approximation. Unterscheide dabei zuerst die einzelnen Phasen und erkläre dann das Gesamtergebnis.

(Bitte wenden.)

***** Freiwillige Weihnachtsbonusaufgaben *****
*** mit Zusatzpunkten ***

Aufgabe 4 (q-gram-Distanz)

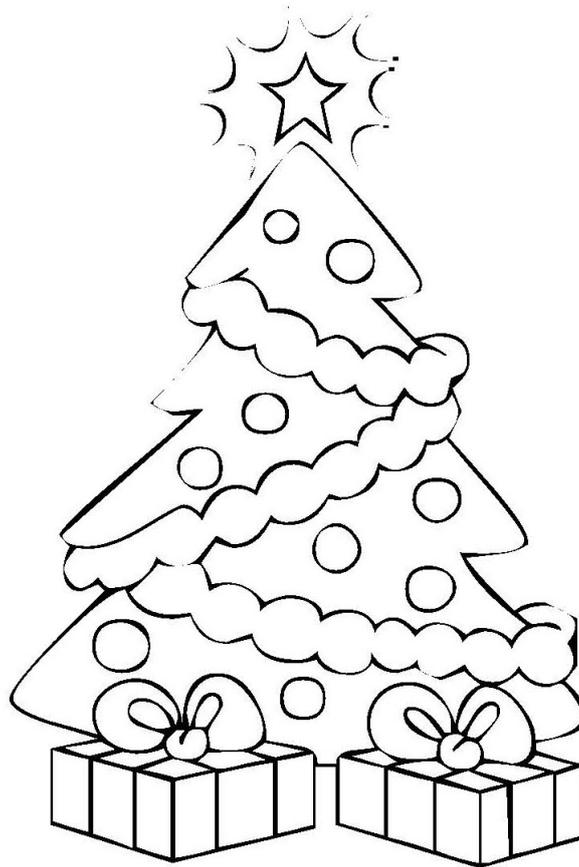
(***)

Gegeben seien der String $s = ACATGTCGTACATTGTCAT$ und $q = 4$. Zeige anhand des de-bruijn-Graphen für s , ob es weitere Strings s' gibt, deren q-gram-Distanz zu s gleich null ist. Gib einen String an oder begründe, warum es keinen gibt.

Aufgabe 5 (Hirschberg-Algorithmus)

(***)

Gegeben sind die Strings $x = AGCTTGAAG$ und $y = CAGTTTAGA$. Verwende den Hirschberg-Algorithmus und Einheitskosten, um ein Alignment der beiden Strings x und y in linearem Platz zu berechnen. Du kannst Alignments direkt berechnen, wenn einer der beiden Strings eine Länge von ≤ 1 hat. Gib das endgültige Alignment explizit an.



Wir wünschen euch schöne Weihnachten und einen guten Rutsch ins neue Jahr! :)