

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2015/2016

Prof. Dr. Jens Stoye · M.Sc. Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015winter/SequenzAnalyse>

Übungsblatt 7 vom 15.12.2015

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Das Erfüllbarkeitsproblem der Aussagenlogik)

(4 Punkte)

1. Was ist das Erfüllbarkeitsproblem der Aussagenlogik und wie nennt man es noch?
2. Gegeben die Formel $\overline{(a \vee b)} \vee c$ und die Belegungen $a = 1, b = 0$ und $c = 1$, was ist das Ergebnis?
3. Gegeben die Formel $a \wedge (b \vee c)$. Finde alle möglichen Belegungen für a, b und c , sodass die Formel 1 ergibt.
4. Wie hoch ist die Laufzeit, wenn du alle Belegungen einer Eingabe mit k verschiedenen Variablen prüfen willst?

Aufgabe 2 (Laufzeit-Abschätzung)

(4 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 Rechenschritte, die konstante Zeit benötigen, braucht man, um ein paarweises Alignment dieser Sequenzen zu berechnen? Beachte diesmal auch die Anzahl an Vorgängern, die für jedes Feld in der Edit-Matrix beachtet werden müssen. Wie lange braucht dann genau ein Rechenschritt?
3. Wie viele solcher Sequenzen kann man mit einem direkten multiplen Alignment mit $(2^k - 1) \cdot n^k$ Rechenschritten in dieser Zeit alignieren?
4. Wie lautet die asymptotische Laufzeit des direkten multiplen Alignments? Erkläre, wie sie zustande kommt.

Aufgabe 3 (Sum-of-Pairs Score)

(4 Punkte)

Gegeben sind die vier Sequenzen $s_1 = \text{AGATC}$, $s_2 = \text{AAAC}$, $s_3 = \text{GGT}$, $s_4 = \text{GGGTC}$, sowie lineare Gapkosten von 2 und die folgende Substitutions-Scorematrix:

	A	C	G	T
A	3	-2	-3	-1
C	-2	5	-1	-2
G	-3	-1	3	-3
T	-1	-2	-3	5

Berechne den Sum-of-Pairs Score des folgenden multiplen Alignments:

A	G	A	T	C
A	A	A	C	-
-	-	G	G	T
G	G	G	T	C

Finde ein multiples Alignment mit besserem Score und gib sowohl das Alignment als auch seinen Score an.

Aufgabe 4 (Carrillo-Lipman-Heuristik)

(3 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?