

Algorithmen und Datenstrukturen in der Bioinformatik

Drittes Übungsblatt WS 13/14

Abgabe Montag, 11.11., 12:00 Uhr

Name:
Matrikelnummer:

Niveau I

Aufgabe 1: Needleman-Wunsch Algorithmus für DNA-Sequenzen

Berechnen Sie für die Sequenzen $S = \text{tgca}$ (horizontal) und $T = \text{ctagc}$ (vertikal) ein globales Alignment nach dem Algorithmus von Needleman-Wunsch.

Ein Match trägt 3 zum Score des Alignments bei, ein Mismatch -5 und ein Space -2 .

- a) Berechnen Sie die *DP-Matrix*, und
 - b) tragen Sie dabei jeweils *alle* Traceback-Kanten ein (also eventuell mehrere pro Zelle).
 - c) Heben Sie *die beim Traceback benutzten* Kanten besonders hervor, und
 - d) notieren Sie das resultierende *globale Alignment*.
-

Aufgabe 2: Needleman-Wunsch Algorithmus für Protein-Sequenzen

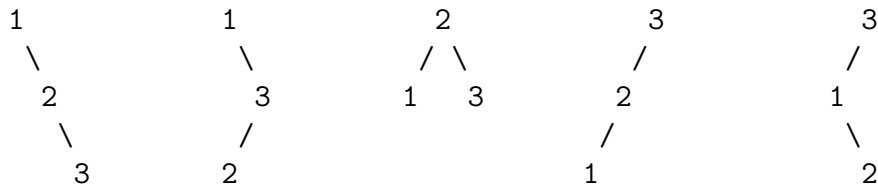
Berechnen Sie für die Sequenzen $S = \text{TINA}$ (horizontal) und $T = \text{HILTA}$ (vertikal) ein globales Alignment nach dem Algorithmus von Needleman-Wunsch.

- a) Suchen Sie im Internet nach der *PAM120-Matrix* (z.B. <http://www.behav.org/gene/bio-info/pam120.jpg>) und nehmen Sie zusätzlich eine Gap Penalty von -4 an.
- b) Führen Sie die Schritte a) bis d) wie bei Aufgabe 1 aus.

Niveau II

Aufgabe 3: Dynamische Programmierung

Wie viele verschiedene binäre Suchbäume mit den Schlüsseln $1 \dots n$ gibt es? Für $n = 3$ gibt es beispielsweise 5 verschiedene solcher Suchbäume:



- Geben Sie eine rekursive Berechnungsvorschrift der Zahl $T(n)$ an. Nutzen Sie $T(0) = T(1) = 1$ als Rekursionsanker.
- Skizzieren Sie einen Algorithmus, der dynamische Programmierung nutzt um $T(n)$ zu bestimmen.
- Was ist dessen Laufzeit?

Zur Erinnerung: Ein binärer Suchbaum ist ein echter Binärbaum (jeder Knoten hat genau zwei Kinder oder ist ein Blatt) mit der Eigenschaft, dass alle Elemente im linken Teilbaum eines Knotens x kleiner oder gleich x sind und alle Elemente im rechten Teilbaum größer oder gleich x sind.

Tipp zu (a): Jedes der n Elemente kann die Wurzel des Suchbaums sein und durch die Sucheigenschaft werden die restlichen Elemente auf die beiden Teilbäume darunter verteilt.