

Präsenzübungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WS 2011/2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/SequenzAnalyse>

Präsenzübungsblatt 4, Woche 50/2011

Aufgabe 1 (Balanced De-Bruijn-Subgraph)

Zeichne den balancierten De-Bruijn-Subgraph für die Sequenz $x = \text{AAGTTTATATTAGTATG}$ mit $q = 4$. Beantworte folgende Fragen zu den Eigenschaften des vorliegenden Graphen:

1. Wann nennt man einen Knoten in einem Graph balanciert?
2. Wann ist ein gerichteter Graph balanciert?
3. Was ist ein Euler-Pfad in einem Graph?
4. Was unterscheidet den De-Bruijn-Subgraph von einem De-Bruijn-Graph?
5. Welche anderen Worte haben das gleiche q-gram-Profil wie x ?

Aufgabe 2 (Edit-Matrix mit Scores)

Berechne die score-maximierende Edit-Distanz der Worte $x = \text{LEBKUCHEN}$ und $y = \text{LOBGESUCHT}$ mit Hilfe einer Edit-Matrix. Die Scores der Operationen sind dabei $I = D = -1, S = 0$ und $C = 1$. Gib alle optimalen Edit-Sequenzen an. Kannst du auch ein optimales Alignment ablesen?

Aufgabe 3 (Sensible Kosten- und Score-Funktionen)

Die Distanz zweier Sequenzen kann nicht nur mit Einheitskosten gemessen werden, sondern zum Beispiel auch mit mehr biologisch motivierten Kosten-Funktionen. Allerdings sollten diese Kosten-Funktionen gewisse Zusammenhänge beachten (sensible Kosten-Funktionen). Analog lassen sich sinnvolle Score-Funktionen definieren.

Ergänze die nachfolgenden Eigenschaften mit den entsprechenden Operationen $=, \leq, \geq, +$.

Diese Aufgabe sollte ohne Skript gelöst werden können!

$$\begin{aligned} 0 &\square \text{cost}(S_{a,a}) \square \text{cost}(D_a) \square \text{cost}(I_a) && \text{für alle } a \in \Sigma \\ \text{cost}(S_{a,a}) &\square \text{cost}(S_{a,c}) \square \text{cost}(S_{c,a}) && \text{für alle } a, c \in \Sigma \\ \text{cost}(S_{a,c}) &\square \text{cost}(S_{a,b}) \square \text{cost}(S_{b,c}) && \text{für alle } a, b, c \in \Sigma \\ \text{cost}(S_{a,c}) &\square \text{cost}(D_a) \square \text{cost}(I_c) && \text{für alle } a, c \in \Sigma \end{aligned}$$

$$\begin{aligned} \text{score}(S_{a,a}) &\square 0 \geq \text{score}(D_a) \square \text{score}(I_a) && \text{für alle } a \in \Sigma \\ \text{score}(S_{a,a}) &\square \text{score}(S_{a,c}) && \text{für alle } a, c \in \Sigma \\ \text{score}(S_{a,c}) &\square \text{score}(S_{c,a}) && \text{für alle } a, c \in \Sigma \\ \text{score}(S_{a,c}) &\square \text{score}(D_a) \square \text{score}(I_c) && \text{für alle } a, c \in \Sigma \end{aligned}$$