

11. Übung „Algorithmen der Bioinformatik I“

1. Ermitteln Sie alle optimalen globalen Alignments für die Strings „KUNST“ und „KITSCH“, wobei die Kostenfunktion $\delta(a, b)$ für zwei Buchstaben a und b des Alphabets oder Leerstrings λ wie folgt definiert ist:

$$\delta(a, b) = \begin{cases} 0, & a = b \\ 2, & a \neq b, \quad a, b \neq \lambda \\ 1, & a = \lambda \text{ oder } b = \lambda \end{cases}$$

Geben Sie die Edit-Transkripte und die Edit-Distanzen an. (2 Punkte)

2. Diese Aufgabe beschäftigt sich mit dem Sequenzalignment und Edit-Transkripten für zwei Strings S_1 und S_2 . Zeigen Sie, dass jedes Alignment von S_1 und S_2 durch genau ein Edit-Transkript erzeugt werden kann. (3 Punkte)
3. In der Vorlesung wurde der Satz eingeführt, dass zu jedem Edit-Transkript $T = (t_1, \dots, t_l)$ für zwei Strings S_1 und S_2 genau ein Pfad von $(0, 0)$ nach (n, m) im Edit-Graphen $G(S_1, S_2)$ existiert. Zudem wurde der Beweis für die eine Richtung des Satzes geführt, d.h. es wurde gezeigt, dass zu jedem Transkript ein Pfad existiert. Zeigen Sie die Rückrichtung, d.h. dass jeder Pfad im Graphen genau einem Edit-Transkript entspricht. (3 Punkte)
4. Zu zwei Strings S_1, S_2 betrachten wir den Edit-Graphen mit Knoten (i, j) , $0 \leq i \leq m$, $0 \leq j \leq n$, zur Bestimmung der minimalen Edit-Distanz D_{ij} der Teilstrings $S_1[1..i]$ und $S_2[1..j]$. Uns interessieren Abschätzungen für die Anzahl der möglichen Pfade in diesem Graphen.

- a) Für welche Knoten (i, j) im Graphen existiert genau ein möglicher Pfad? Leiten Sie für alle anderen Knoten (i, j) eine rekursive Darstellung der Anzahl der möglichen Pfade $p(i, j)$ her! (2 Punkte)
- b) Schätzen Sie die Anzahl $p(m, n)$ der möglichen Pfade zum Knoten (m, n) nach unten ab durch einen Ausdruck der Form

$$p(m, n) \geq C\alpha^{\min(m, n)}.$$

Dabei sollen C, α möglichst groß sein. (3 Punkte)