

Foundations of Sequence Analysis  
Winter 2005/2006

Exercises

Übung 4, Besprechung am 21.11.2004 (14-16h in E0-160)  
bzw. 24.11.2004 (14-16h in C01-148 und E01-108).

1. **Edit-Distanz – Optimale Alignments.** Seien  $\mathbf{u} = \text{AGTGC}$  und  $\mathbf{t} = \text{ATCAC}$  zwei Sequenzen. Berechnen Sie (mit Stift und Papier) die Edit-Distanz und alle optimalen Alignments bezüglich der verschiedenen Kostenfunktionen  $\delta$ .

- (a) Einheitskosten:

$$\delta(\alpha \rightarrow \beta) = \begin{cases} 0 & : \alpha, \beta \in A \wedge \alpha = \beta \\ 1 & : \text{sonst} \end{cases}$$

- (b) Hamming-Kosten:

$$\delta(\alpha \rightarrow \beta) = \begin{cases} 0 & : \alpha, \beta \in A \wedge \alpha = \beta \\ 1 & : \alpha, \beta \in A \wedge \alpha \neq \beta \\ \infty & : \text{sonst} \end{cases}$$

- (c)

$$\delta(\alpha \rightarrow \beta) = \begin{cases} 0 & : \alpha, \beta \in A \wedge \alpha = \beta \\ 3 & : \alpha, \beta \in A \wedge \alpha \neq \beta \\ 1 & : \text{sonst} \end{cases}$$

2. **Eigenschaften einer Kostenfunktion.** Ein Alignment von zwei Sequenzen hängt im wesentlichen von der Kostenfunktion ab.

- (a) Was ist die spezielle Eigenschaft der Kostenfunktionen aus 1(b) und 1(c).  
(b) Welche Eigenschaften sollte eine sinnvolle Kostenfunktion erfüllen.

3. **DP-Algorithmus.** Implementieren Sie den Dynamische Programmieralgorithmus (in Java, Haskell oder C) mit folgende Ausgaben:

- (a) Edit-Distanz,  
(b) ein optimales Alignment,  
(c) alle optimalen Alignments.

**Sternchen-Aufgabe.** Stellen Sie die minimierenden Kanten und Pfade des Edit-Graphen bezüglich der Sequenzen  $\mathbf{u}$  und  $\mathbf{v}$  und der drei verschiedenen Kostenfunktionen aus Aufgabe 1 zeichnerisch dar.