

# Übungen zur Vorlesung Grundlagen der Sequenzanalyse

Universität Bielefeld, WS 2006/07

Dr. Sven Rahmann · Dipl.-Bioinf. Katharina Jahn

<http://gi.cebitec.uni-bielefeld.de/teaching/2006winter/sequenzanalyse/>

**Blatt 5 vom 16.11.2006**

**Abgabe am 23.11.2006 vor der Vorlesung um 8:30 in H3**

**Aufgabe 1** Gegeben sei eine (kurze) Sequenz  $P \in \Sigma^m$ , das Pattern, und eine andere (lange) Sequenz  $T \in \Sigma^n$ , der Text. Gesucht ist ein bezüglich Edit-Distanz optimales Vorkommen von ganz  $P$  irgendwo in  $T$ . Wie sieht der Alignment-Graph zu diesem Problem aus? Beschreibe genau, welche Knoten und Kanten es gibt und welche Kosten die Kanten haben.

**Aufgabe 2** Schreibe explizit den dynamic programming Algorithmus zum Problem aus Aufgabe 1 auf.

**Aufgabe 3** Warum sollte eine Gapkostenfunktion  $g : \mathbb{N}_0 \rightarrow \mathbb{R}_0^+$  stets subadditiv sein, d.h., die Bedingung  $g(i + j) \leq g(i) + g(j)$  für alle Gap-längen  $i$  und  $j$  erfüllen?

**Aufgabe 4** Zeichne den (globalen) Alignment-Graphen  $G(x, y)$  für  $x = \text{BCACD}$  und  $y = \text{DBADAD}$ . Beschrifte alle Kanten mit ihren Kosten (Standard-Einheitskosten für Edit-Distanz) und markiere die minimierenden Kanten. Ermittle alle optimalen Alignments durch Rückverfolgen aller minimierenden Pfade.