Prof. Dr. Enno Ohlebusch
Abt. Theoretische Informatik

Universität Ulm

Fakultät für Informatik

Übungen zur Vorlesung
Algorithmen zur Sequenzanalyse
Sommersemester 2003

Aufgabe 3.1 Gegeben seien zwei Sequenzen u und v der Längen m und n. Geben Sie in Pseudocode einen Algorithmus mit der Laufzeit $O(mn)$ an, der eine $m \times n$-Matrix M berechnet, so dass für alle $i \in [1, m]$ und $j \in [1, n]$ gilt:

$$M_{i,j} = |lcp(u_i \dots u_m, v_j \dots v_n)|$$

Dabei bezeichnet $lcp(u', v')$ das längste gemeinsame Präfix zweier Sequenzen u' und v'. Für die Sequenzen $u = acaacacaacac$ und $v = aaacaccacaca$ sollte ihr Algorithmus die folgende Matrix liefern:

```
1 1 3 0 2 0 0 3 0 3 0 1
0 0 0 2 0 1 2 0 2 0 2 0
2 5 1 0 1 0 0 1 0 1 0 1
1 1 4 0 2 0 0 5 0 3 0 1
0 0 0 3 0 1 4 0 4 0 2 0
1 1 3 0 2 0 0 3 0 3 0 1
0 0 0 2 0 1 2 0 2 0 2 0
2 5 1 0 1 0 0 1 0 1 0 1
1 1 4 0 2 0 0 4 0 3 0 1
0 0 0 3 0 1 3 0 3 0 2 0
1 1 2 0 2 0 0 2 0 2 0 1
0 0 0 1 0 1 1 0 1 0 1 0
```

Aufgabe 3.2 Sei δ eine symmetrische Kostenfunktion, so dass für alle Edit-Operationen $a \to b$, $a \to c$ und $c \to b$ gilt:

$$\delta(a \to b) \leq \delta(a \to c) + \delta(c \to b) \tag{1}$$

Zeigen Sie, dass dann $edist_\delta$ eine Metrik auf $\mathcal{A}^*$ ist, d.h. für alle $u, v, w \in \mathcal{A}^*$ gilt:

$$\begin{aligned} edist_\delta(u,v) &= edist_\delta(v,u) \\ edist_\delta(u,v) &\leq edist_\delta(u,w) + edist_\delta(w,v) \\ edist_\delta(u,v) = 0 &\Leftrightarrow u = v \end{aligned}$$

Aufgabe 3.3 Sei A ein Alignment von $u \in \mathcal{A}^m$ und $v \in \mathcal{A}^n$. Sei r die Anzahl der Ersetzungsoperationen in A, i die Anzahl der Einfügeoperationen, und d die Anzahl der Löschoperationen. Zeigen Sie, daß dann $m + n = 2 \cdot r + i + d$ gilt.

Aufgabe 3.4 Eine Score-Funktion σ ist eine Funktion, die jeder Edit-Operation $a \to b$ einen Score $\sigma(a \to b) \in \mathbb{R}$ zuordnet. Seien $u, v \in \mathcal{A}^*$ und A ein Alignment von u und v. Der Score von A bzgl. σ ist

$$\sigma(A) := \sum_{a \to b \in A} \sigma(a \to b)$$

Wir sagen, dass A maximalen Score bzgl. σ hat, falls für alle Alignments A' von u und v gilt: $\sigma(A') \leq \sigma(A)$.

Sei δ eine Kostenfunktion. Aus δ erhält man eine Score-Funktion σ durch die Definition

$$\sigma(a \to b) = \begin{cases} x \cdot \delta(a \to b) + 2 \cdot y & \text{if } a, b \in \mathcal{A} \\ x \cdot \delta(a \to b) + y & \text{otherwise} \end{cases}$$

wobei $x, y \in \mathbb{R}$ Konstanten sind mit $x < 0$. Zeigen Sie, dass A genau dann maximalen Score bzgl. σ hat, wenn A minimale Kosten bzgl. δ hat.