

Übungen zur Algorithmischen Bioinformatik I

Blatt 9

Abgabetermin: Montag, 3.7.2017, 10 s.t.

1. Aufgabe (Bonus-Aufgabe):

Gegeben sei eine Menge $S = \{s_1, \dots, s_\ell\}$ von Zeichenreihen mit $n = \sum_{i=1}^{\ell} |s_i|$. Konstruieren Sie einen Algorithmus, mit dem man in Zeit $O(n)$ alle Zeichenreihen $s_i \in S$ finden kann, die Teilwörter einer anderen Zeichenreihe $s_j \in S$ sind.

Hinweis: Ein Suffix-Baum kann helfen.

2. Aufgabe:

Sei $w : \bar{\Sigma}_0^2 \rightarrow \mathbb{R}_+$ eine Metrik. Zeigen Sie, dass die zugehörige Alignment-Distanz \bar{d}_w ebenfalls eine Metrik ist.

Die Eigenschaften einer Metrik sollen hier direkt nachgewiesen werden, d.h. die Kenntnis, dass in diesem Fall die Edit-Distanz mit der Alignment-Distanz identisch ist, darf nicht verwendet werden.

3. Aufgabe:

In der Vorlesung wurde für eine metrische Kostenfunktion $w : \bar{\Sigma}_0^2 \rightarrow \mathbb{R}_+$ gezeigt, dass die zugehörige Alignment-Distanz identisch mit der zugehörigen Edit-Distanz ist.

Begründen oder widerlegen Sie die Aussage, dass diese Gleichheit auch gilt, wenn w alle Eigenschaften einer Metrik

- (a) **bis auf** die Dreiecksungleichung erfüllt,
- (b) **bis auf** die Symmetriebedingung erfüllt,
- (c) **bis auf** die Definitheitsbedingung erfüllt.

4. Aufgabe (Bonus-Aufgabe):

Gegen sei folgende sinnvolle Kostenfunktion w' für ein Ähnlichkeitsmaß auf $\Sigma = \{a, b, c\}$:

w'	a	b	c
a	6	3	-4
b	3	4	-1
c	-4	-1	3

mit $w'(x, -) = -4 \forall x \in \Sigma$.

- (a) Bestimmen Sie eine Konstante $D \in \mathbb{R}_+$, so dass w mit $w(a, b) = D - w'(a, b) \forall a, b \in \Sigma$ und $w(a, -) = \frac{D}{2} - w'(a, -) \forall a \in \Sigma$ eine sinnvolle Kostenfunktion für ein Distanzmaß ist.
- (b) Ist dies die einzige mögliche Wahl für D , so dass w noch sinnvoll ist?
- (c) Gegeben sei nun $s_{w'}(a, b)$ für zwei Sequenzen a und b . Wie läßt sich daraus \bar{d}_w berechnen, vorausgesetzt $|a|$ und $|b|$ sind bekannt.

Hinweis: Begründungen nicht vergessen.