Technische Universität München
Fakultät für Informatik
Lehrstuhl für Effiziente Algorithmen
Dr. Volker Heun

Wintersemester 2001/2002
Übungsblatt 7
5. Dezember 2001

Algorithmische Bioinformatik

Abgabetermin: Donnerstag, den 13. Dezember (15^{00} Uhr) in der Übung

Aufgabe 1

Sei Σ ein Alphabet mit $- \notin \Sigma$, $\overline{\Sigma} := \Sigma \cup \{-\}$ und $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. Sei $w : \overline{\Sigma} \times \overline{\Sigma} \to \mathbb{R}_+$ eine Kostenfunktion. Die Abstandsfunktion für ein Alignment $(\bar{x}, \bar{y})$ für $x, y \in \Sigma^*$ ist definiert durch:

$$\bar{d}_w(\bar{x}, \bar{y}) := \sum_{i=1}^{|\bar{x}|} w(\bar{x}_i, \bar{y}_i).$$

Ferner ist der Alignment-Abstand von $x, y \in \Sigma^*$ definiert durch:

$$d_w(x, y) := \min\{\bar{d}_w(\bar{x}, \bar{y}) \mid (\bar{x}, \bar{y}) \text{ ist ein Alignment für } x, y\}.$$

Beweise: Ist die Funktion w eine Metrik, so ist auch $\bar{d}_w$ und d_w eine Metrik.

Hinweis: *Die als freiwillig gekennzeichneten Aufgaben werden nicht bewertet und sind für die Klausurzulassung ohne Bedeutung.*

Freiwillige Aufgabe 1

a) Konstruiere eine Methode zur Konstruktion eines so genannten *verallgemeinerten Suffixbaumes* an, der die Suffixe aller Worte der Menge $M = \{w_1, w_2, \ldots, w_n\}$ von Zeichenreihen beinhaltet, wobei $w_i \in \Sigma^*$.

   Hinweis: Betrachte dazu das Wort $w' = w_1\$_1 w_2\$_2 \cdots w_n\$_n$, wobei $\$_1, \ldots, \$_n \notin \Sigma$ neue Symbole sind.

b) Dieser Baum kann jedoch größer als notwendig sein. Verfeinere ggf. die Methode zur Erzeugung eines verallgemeinerten Suffixbaumes, so dass der verallgemeinerte Suffixbaum nur noch alle Suffixe der Wörter von M enthält.

c) Konstruiere nach Deiner Methode den Suffixbaum für die Menge

$$\{ababbab, aabab, babab, abba\}.$$

— **Weiter siehe nächste Seite!** —

Freiwillige Aufgabe 2

Sei Σ ein Alphabet mit $- \notin \Sigma$, $\overline{\Sigma} := \Sigma \cup \{-\}$ und $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. Eine Funktion $w : \overline{\Sigma} \times \overline{\Sigma} \to \mathbb{R}_+$ heißt *Kostenfunktion für ein Abstandsmaß*, wenn folgende Bedingungen gelten:

- w ist symmetrisch, d.h. $\forall x, y \in \overline{\Sigma} : w(x,y) = w(y,x)$.
- w erfüllt die Dreiecksungleichung, d.h. $\forall x, y, z \in \overline{\Sigma} : w(x,y) \leq w(x,z) + w(z,y)$.
- Für $A := \max\{w(a,a) : a \in \Sigma\}$ und $B := \min\{w(a,b) : a \neq b \in \Sigma^- \vee a = b = -\}$ gilt $A \leq B$.

Die Funktion $w' : \overline{\Sigma} \times \overline{\Sigma} \to \mathbb{R}$ heißt *Kostenfunktion für ein Ähnlichkeitsmaß*, wenn folgende Bedingungen gelten:

- w' ist symmetrisch, d.h. $\forall x, y \in \overline{\Sigma} : w'(x,y) = w'(y,x)$.
- $w'(x,x) \geq 0$ für $x \in \Sigma$.
- $w'(x,y) \leq 0$ für $x \neq y \in \overline{\Sigma}$.
- $w'(-,-) \leq 0$.

a) Zeige, dass man zu jeder Kostenfunktion w für ein Abstandsmaß eine Kostenfunktion w' für ein Ähnlichkeitsmaß definieren kann, so dass $w'(x,y) = C - w(x,y)$ für ein geeignetes $C \in \mathbb{R}_+$.

b) Zeige, dass man zu jeder Kostenfunktion w' für ein Ähnlichkeitsmaß eine Kostenfunktion w für ein Abstandsmaß definieren kann, so dass $w(x,y) = D - w'(x,y)$ für ein geeignetes $D \in \mathbb{R}_+$.

c) Lassen sich die obigen Äquivalenzen auch ohne die Bedingung der Symmetrie bei der Definition einer Kostenfunktion eines Distanzmaßes bzw. Abstandsmaßes zeigen?

Hinweis:

Die Klausur findet am **4. Februar 2002** um 18:00 Uhr
im **Großen Physikhörsaal (1200)** statt!
Die Anwesenheit ist ab **17:45 Uhr** erforderlich!