
Algorithmische Bioinformatik I

Abgabetermin: Mittwoch, den 12. Juni, vor der Vorlesung

Dies ist ein Bonus-Blatt, d.h. dass die erzielten Punkte für den Notenbonus berücksichtigt werden, die zu erzielenden Punkte der Notenbonus-Aufgaben jedoch nicht.

Aufgabe (Notenbonus) 1

Beweise oder widerlege für alle Funktionen $f, g : \mathbb{N} \rightarrow \mathbb{R}$:

- a) $f + g \in O(f) \Rightarrow g \in O(f)$.
- b) $g \in o(f) \Rightarrow f + g \in \Theta(f)$.

Erinnerung: Betrachtet man Funktionen $f, g : \mathbb{N} \rightarrow \mathbb{R}$, so schreibt man $g = O(f)$, wenn $|g| = O(|f|)$ gilt, wobei hier $|f| : \mathbb{N} \rightarrow \mathbb{R}_+ : x \mapsto |f(x)|$ für $f : \mathbb{N} \rightarrow \mathbb{R}$ ist (analog für Ω , Θ , o , ω).

Aufgabe (Notenbonus) 2

Betrachte die folgende Suchwortmenge $S = \{ein, eis, kreisen, reise, steiles, teil\}$.

- a) Konstruiere einen Suchwort-Baum nach Aho-Corasick;
- b) Konstruiere die Failure-Links;
- c) Markiere nach dem in der Vorlesung angegebenen Algorithmus alle Knoten darin, die auf einen Treffer hinweisen (inklusive der Hit-Links);
- d) Wende den Algorithmus von Aho-Corasick mit dem konstruierten Suchwort-Baum auf das folgende Wort an: *eieinsteilesteilinkreisen*

Hinweis: Verwende verschiedene Farben, aus denen ersichtlich wird, welche Teile des Baumes (bzw. welche Annotationen) zu welchem Aufgabenteil gehören (zeichne ggf. den Baum mehrmals).

Aufgabe 3

Betrachte den Boyer-Moore-Algorithmus, der für die Bestimmung des Shifts bei einem Mismatch nur die Weak-Good-Suffix-Rule berücksichtigt:

$$S'[j] = \min \left\{ \sigma : \begin{array}{l} (s_{j+1} \cdots s_{m-1} \in \mathcal{R}(s_{j+1-\sigma} \cdots s_{m-1}) \wedge \sigma \leq j) \vee \\ (s_0 \cdots s_{m-1-\sigma} \in \mathcal{R}(s_0 \cdots s_{m-1}) \wedge \sigma > j) \vee (\sigma = m) \end{array} \right\}$$

Gib eine Konstante $c > 0$ sowie eine unendliche Familie $\mathcal{F} = \{(s, t) : s, t \in \Sigma^*\}$ an, die für jedes Paar $(m, n) \in \mathbb{N}^2$ ein Paar $(s, t) \in \mathcal{F}$ mit $|s| \geq m$, $|t| \geq n$ und $|s| \leq |t|$ enthält und bei dem diese Variante für eine erfolglose Suche mindestens $c \cdot (|s| \cdot |t|)$ Zeichenvergleiche ausführt.