

Algorithmen auf Sequenzen

Übungsblatt 7

Abgabetermin: Dienstag, 13.12.2016, 10 Uhr
(Via email an sophie.friedl@bio.ifl.lmu.de oder persönlich)

1. Aufgabe (Dekorierung der Tandem-Repeats)

Zeigen Sie, wie man mit Hilfe des mit dem Vokabular von $t \in \Sigma^*$ dekorierten Suffix-Baumes $T(t\$)$ einen Algorithmus zur Ausgabe aller Tandem-Repeat-Paare bzw. aller rechtsverzweigender Tandem-Repeat-Paare entwickeln kann und analysieren Sie seine Laufzeit.

Welchen Vorteil hat dieser Algorithmus im Vergleich zu den anderen in der Vorlesung für diesen Zweck vorgestellten Algorithmen?

2. Aufgabe (Suche von Wörtern, *nicht abzugeben*)

Sei $t = _w_1_ \cdots _w_m \in (\Sigma \cup \{_\})^*$, wobei $_ \notin \Sigma$, $w_i \in \Sigma^+$, und $n = |t|$.

Beschreiben Sie einen Algorithmus, der t in Zeit $O(n)$ so vorverarbeitet, dass anschließend für jedes Wort p in Zeit $O(|p|)$ entschieden werden kann, ob p eines der w_i ist, wobei der zusätzlich zum Text t benötigte Platz der resultierenden Datenstruktur $O(m)$ sein soll (wobei die Zeichenreihe t selbst auch zur Verfügung steht).

Wie groß ist der (maximale) zusätzliche Platzbedarf der vorgeschlagenen Datenstruktur während der Konstruktion?

3. Aufgabe (LCA)

Wenden Sie auf den rechten Baum die lineare Vorverarbeitung aus der Vorlesung an (mit $k = 5$) und beantworten Sie die LCA-Anfragen $\text{lca}(N, T)$, $\text{lca}(P, L)$ und $\text{lca}(E, K)$ gemäß dem Algorithmus aus der Vorlesung. Dabei müssen die Bit-Vektoren $V^{B,j}$ angegeben werden, aber nicht deren Herleitung.

