

Algorithmen auf Sequenzen

Übungsblatt 11

Abgabetermin: Dienstag, 24.01.2017, 10 Uhr

(Via email an sophie.friedl@bio.ifi.lmu.de oder persönlich)

1. Aufgabe (Minimal eindeutiges Teilwort)

Gegeben sei eine Zeichenreihe $t \in \Sigma^*$. Ein Wort $w \in \Sigma^*$ heißt *minimal eindeutiges Teilwort* von t , wenn w genau einmal in t auftritt und wenn jedes Teilwort von w mindestens zweimal in t auftritt.

Entwerfen Sie einen Algorithmus basierend auf Extended-Suffix-Arrays, der alle minimal eindeutigen Teilwörter von t erkennt. Zeigen Sie die Korrektheit des Algorithmus und analysieren Sie dessen Laufzeit. Die Laufzeit soll linear in $|t|$ sein.

2. Aufgabe (Suffix-Array und Suffix-Baum)

Beschreiben Sie, wie man für einen Text t aus dessen Suffix- und LCP-Array den Suffix-Baum für t in Zeit $O(|t|)$ konstruieren kann.

3. Aufgabe (Burrows-Wheeler-Transformierte)

Betrachten Sie das Wort $t = \text{ABANANAISANANANAS}$.

- (a) Konstruieren Sie die Burrows-Wheeler-Transformierte $\hat{t}$ zu t .
- (b) Geben Sie die zugehörige LF-Funktion für $\hat{t}$ an
- (c) Bestimmen Sie die Werte $C(\cdot)$ und $Occ(\cdot, \cdot)$ für b).
- (d) Suchen Sie nach $s = \text{NANA}$ im FM-Index für t mit Hilfe des in der Vorlesung angegebenen Algorithmus.