

Text Indexing and Information Retrieval

Letztes Übungsblatt

Besprechung: 26.1.2015

Aufgabe 1 (Theorie)

Zeigen Sie alle Datenstrukturen, die für die $O(m \log \sigma)$ -Rückwärtssuche auf dem Text

$$T = \text{missmississippi\$}$$

benötigt werden. Führen Sie beispielhaft die Rückwärtssuche für das Pattern $P = \text{is}$ durch.

Aufgabe 2 (Theorie)

Sei $D[1, n]$ ein Array der Länge n mit Werten aus $[1, r]$. Zeigen Sie eine Datenstruktur der Größe $O(n \log r)$ Bits, die es ermöglicht, folgende Anfragen in $O(\log r)$ Zeit zu beantworten: für gegebene Indizes $1 \leq \ell \leq r \leq n$ und einen Wert $k \in [1, r - \ell + 1]$, finde das k -t-größte Element in $D[\ell, r]$. Hinweis: Wavelet-Bäume.

Aufgabe 3 (Theorie)

Sei S eine Menge von n Punkten auf einem $(n \times n)$ -Gitter gegeben, so dass keine 2 Punkte die gleiche x-Koordinate haben. Entwerfen Sie eine Datenstruktur der Größe $O(n \log n)$ Bits, mit der Sie *4-seitige Bereichsanfragen* beantworten können. Solche Anfragen sollen für 4 der Anfrage übergebene Koordinaten x_l, x_r, y_t und y_b alle k Punkte aus S ausgeben, die in $[x_l, x_r] \times [y_b, y_t]$ liegen. Die Anfragezeit soll $O(k \log n)$ sein. Hinweis: Wavelet-Bäume.