

Text Indexing and Information Retrieval

Übungsblatt 5

Besprechung: 19.12.2013

Aufgabe 1 (Theorie)

Entwerfen Sie eine möglichst einfache Datenstruktur linearer Größe, die Bereichsminimumsanfragen auf einem Array der Größe n in $O(\log n)$ Zeit beantworten kann.

Aufgabe 2 (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 linearer Größe, mit der Sie *3-seitige Bereichsanfragen* in optimaler Zeit beantworten können. Solche Anfragen sollen für 3 der Anfrage übergebene Koordinaten x_l, x_r, y_t alle k Punkte aus S ausgeben, die in $[x_l, x_r] \times [-\infty, y_t]$ liegen. Die Anfragezeit soll $O(k)$ sein. Hinweis: Bereichsminimumsanfragen.

Aufgabe 3 (Theorie)

Das in der Vorlesung vorgestellte Verfahren zur Repräsentation von Kartesischen Bäumen hat jedem Kartesischen Baum der Größe s einen Bit-Vektor der Länge $2s + 1$ zugewiesen. Eine andere Möglichkeit zur Repräsentation von Bäumen in $2s$ Bits sind *balancierte Klammersequenzen*: führe eine Tiefensuche durch den Baum durch, und schreibe dabei eine '(', wenn ein Knoten das erste Mal besucht wird, und eine ')', wenn er das letzte Mal besucht wird. Warum ist diese Repräsentation für die Datenstruktur für Bereichsminimumsanfragen *nicht* geeignet? Geben Sie mindestens ein Beispiel an, das zeigt, warum dies nicht funktioniert.

Aufgabe 4 (Theorie)

Entwerfen Sie einen $O(n \log n)$ -Zeit Algorithmus zur Berechnung der LZ-Zerlegung eines Textes $T_{1,\dots,n}$, der neben dem Suffix Array A für T nur 2 Arrays für previous- und next-smaller-values auf A benötigt.