Algorithmische Grundlagen und Vermittlung der Informatik
Lehrstuhl 11 (Algorithm Engineering)
Prof. Dr. Johannes Fischer

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 12

Besprechung: 19.1.2015

Aufgabe 1 (Praxis)

Informieren Sie sich unter `https://graphics.stanford.edu/~seander/bithacks.html` über die Möglichkeiten, in einem (32- oder 64-Bit) Wort die Anzahl der Einsen zu zählen. Implementieren Sie mit einer dieser Ideen eine einfache rank-Datenstruktur, z.B. nur mit dem Array M und Block-Größe $s = 32$ oder $s = 64$ (je nachdem, wie Sie die Einsen in einem Wort zählen).

Aufgabe 2 (Theorie)

Zeigen Sie alle Datenstrukturen, die für die $O(m)$-Mustersuche auf dem Text

$$T = \texttt{abaababbababbabaaababbaba}$$

benötigt werden. Für die rank-Datenstruktur auf dem BW-transformierten Text können Sie $s = 4$ und $s' = 8$ annehmen.

Aufgabe 3 (Theorie)

In Blatt 9, Aufgabe 2(a) hatten wir eine Möglichkeit zur platzeffizienten Kodierung von Bäumen kennengelernt. (Obwohl diese Kodierung damals nicht für Kartesische (binäre) Bäume geeignet war, ist sie doch für beliebige *geordnete* Bäume geeignet.)

Ausgehend von einer solchen Klammersequenz (der Länge $2n$ für einen Baum auf n Knoten) sollen Sie in dieser Aufgabe eine Datenstruktur sublinearer Größe (d.h. $o(n)$ Bits) entwerfen, die für die Positionen x und y zweier öffnender Klammern (die den Knoten u und v entsprechen) in $O(1)$ Zeit die Position z der öffnenden Klammer findet, die dem LCA von u und v entspricht. *Hinweis:* Orientieren Sie sich an den in der VL vorgestellten RMQ- und rank-Datenstrukturen.