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

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 8

Besprechung: 12.12.2016

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 mindestens zwei dieser Ideen einfache rank-Datenstrukturen, 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). Vergleichen Sie die Laufzeit der daraus resultierenden Verfahren.

Aufgabe 2 (Theorie)

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

$$T = \texttt{abaababbababbabaababbaba}$$

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

Aufgabe 3 (Theorie)

Die zu rank inverse Operation ist select: $\mathrm{select}_1(B, i)$ liefert für ein $1 \leq i \leq n$ die Position der i-ten 1 in einem Bitvektor $B[1, n]$.

a) Zeigen Sie, wie Sie select-Anfragen in $O(\log n)$ Zeit mit einer Datenstruktur der Größe $o(n)$ Bits beantworten können.

b) Nehmen Sie an, dass der Bitvektor B dünn besetzt ist, nämlich nur $o(\frac{n}{\lg n})$ Einsen hat. Geben Sie für diesen Fall eine Datenstruktur der Größe $o(n)$ Bits an, die select-Anfragen in $O(1)$ Zeit beantworten kann.