

Text Indexing and Information Retrieval

Übungsblatt 7

Besprechung: 5.12.2016

Aufgabe 1 (Praxis)

Implementieren Sie den BWT-Algorithmus. Benutzen Sie hierzu bestehende Suffixarray-Konstruktionsalgorithmen (etwa aus vorherigen Übungsblättern oder öffentlichen Repositories).

Experimentieren Sie mit verschiedenen Kodierungen (z.B. γ -, Fibonacci- oder Huffman-Codes) und vergleichen Sie die Größen der so komprimierten Texte. Implementieren und vergleichen Sie auch eine einfache Lauflängen-Kodierung von L (s. Skript).

Aufgabe 2 (Theorie)

Beschreiben Sie einen Linearzeit-Algorithmus, der aus dem BW-transformierten Text den Originaltext *vorwärts* rekonstruiert.

Aufgabe 3 (Theorie)

Erweitern Sie den Algorithmus für die Rücktransformation der BWT so, dass er auch das Suffix-Array für den Originaltext T berechnet. Kann dabei das LF-Array mit dem Suffix-Array überschrieben werden?