

Text Indexing and Information Retrieval

Übungsblatt 6

Besprechung: 16.1.2014

Aufgabe 1 (Theorie+Praxis)

- a) Welches ist der Originaltext T zur Burrows-Wheeler-Transformation $L = 0YDBBBBAAAA0D\$?$
Führen Sie den Vorwärts- oder Rückwärtsalgorithmus zur Dekodierung aus der Vorlesung aus!
- b) Implementieren Sie den Algorithmus zur Dekodierung der BWT (vorwärts oder rückwärts).

Aufgabe 2 (Theorie)

- a) Komprimieren Sie die BWT aus Aufgabe 1 mit dem Unärcode (nach vorausgehendem Move-To-Front).
- b) Komprimieren Sie die BWT aus Aufgabe 1 mit dem Elias- γ Code (wieder nach vorausgehendem Move-To-Front).

Aufgabe 3 (Theorie)

Eine andere, nicht in der Vorlesung behandelte Möglichkeit zur Kompression der BWT ist die *Huffman-Kodierung* (wieder nach vorausgehendem Move-To-Front). Recherchieren Sie hierzu und wenden Sie das Verfahren auf die BWT aus Aufgabe 1 an.

Aufgabe 4 (Theorie)

Die (unkomprimierte) BWT hat den Nachteil, dass man keinen *random access* mehr auf dem Text hat: möchte man das Zeichen $T[i]$ lesen, muss man im schlechtesten Fall zunächst den ganzen String dekodieren. Überlegen Sie sich eine Datenstruktur, mit der man bei $O(k)$ Platz ein beliebiges Zeichen in $O(n/k)$ Zeit lesen kann.