Datenkompression

Sommersemester 2016 — **Übungsblatt 4** — Prof. Dr. E. Ohlebusch

Institut für Theoretische Informatik — J. Lorenz

Ausgegeben am 31.05.2016 — Besprechung am 07.06.2016

Aufgabe 4.1

Die Zeichenfolge S := `MISS MISSION MISSISSIPPI` über dem Quellenalphabet Σ := {`M`, `I`, `S`, `O`, `P`, `N`} soll mit dem LZ78 Verfahren komprimiert werden. Die Leerzeichen dienen dabei nur der Formatierung und sollen nicht mitkodiert werden. Wieviel Bits werden für den komprimierten String benötigt?

Aufgabe 4.2

Wir betrachten einen LZW-Decodierer mit folgendem Anfangswörterbuch:

Index	Eintrag
1	`a`
2	`b`
3	`c`

a) Decodieren Sie die Folge 1,2,3,6,4,8 mit diesem LZW-Decodierer.

b) Werden in dem Trie eines LZW-Decodierers Zeiger auf Kindknoten benötigt? Werden Zeiger auf Elternknoten benötigt?

c) Beschreiben Sie einen LZW-Decodierer, dessen Wörterbuch mit Hilfe einer Tabelle anstelle eines Tries implementiert ist. Welche Vorteile sehen Sie in solch einem Verfahren?

Tipp: Für die Folge 1,2,2,1,3,4,9,3 könnte die Tabelle wie folgt aussehen:

1	2	3	4	5	6	7	8	9	10
0	0	0	1	2	2	1	3	4	9
`a`	`b`	`c`	`b`	`b`	`a`	`c`	`a`	`a`	`c`

Aufgabe 4.3

Sei BWT = `ttc$attaa`.

a) Decodieren Sie BWT unter Verwendung des LF-Arrays.

b) Implementieren Sie die inverse Burrows-Wheeler-Transformation.

c) Erweitern Sie das Verfahren, so dass zusätzlich zum decodierten String das Suffix-Array berechnet wird.

Aufgabe 4.4

a) Wenden Sie die Move-To-Front Transformation auf den String S = `seeaaassnnnne` an.

b) Wenden Sie die inverse Move-To-Front Transformation auf den String $S_M = 21303231113$ an. Verwenden Sie dabei (`a, e, n, s`) als initiale Liste.

Aufgabe 4.5

Sei S ein String der Länge n, der kein Abschlusszeichen \$ enthält. Wir bezeichnen mit L die letzte Spalte der Matrix, die alle n Rotationen von S in lexikografischer Reihenfolge enthält, und mit SA das Suffix-Array von S. Außerdem definieren wir den String $\mathsf{BWT}[1..n]$ durch

$$\mathsf{BWT}[i] := \begin{cases} S[\mathsf{SA}[i]-1] & \text{falls } \mathsf{SA}[i] > 1 \\ S[n] & \text{falls } \mathsf{SA}[i] = 1 \end{cases} .$$

a) Finden Sie einen String S, für den $L \neq \mathsf{BWT}$ gilt.

b) Die Funktion $f_{\mathrm{L}} \colon \Sigma^n \to \Sigma^n$ ordne jedem String S das zugehörige L zu. Ist f_{L} bijektiv?

c) Die Funktion $f_{\mathrm{BWT}} \colon \Sigma^n \to \Sigma^n$ ordne jedem String S den String BWT zu. Ist f_{BWT} bijektiv?

d) Sei D_n die Menge aller Strings der Länge n, die das Abschlusszeichen \$ genau einmal enthalten, nämlich am Ende. W_n sei die Menge aller Strings der Länge n, die das Abschlusszeichen \$ genau einmal enthalten, an einer beliebigen Position. Die Funktion $g_{\mathrm{BWT}} \colon D_n \to W_n$ ordne jedem String mit Abschlusszeichen \$ den String BWT zu. Ist g_{BWT} bijektiv?