Datenkompression

Sommersemester 2015 — **Übungsblatt 4** — Prof. Dr. E. Ohlebusch

Institut für Theoretische Informatik — T. Beller

Ausgegeben am 09.06.2015 — Besprechung am 16.06.2015

Aufgabe 4.1

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

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

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?

Aufgabe 4.2

Sei $\mathsf{BWT} = \texttt{ttc\$attaa}$.

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

b) Implementieren Sie die inverse Burrows-Wheeler-Transformation unter Verwendung des LF-Arrays.

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

Aufgabe 4.3

a) Wenden Sie die Move-To-Front Transformation auf den String $S = \texttt{seeaaassnnnne}$ 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 4.4

Sei S ein String und S_M die Move-To-Front-Transformierte von S. Seien Σ bzw. Σ_M die Mengen aller in S bzw. S_M vorkommenden Zeichen. Zeigen oder widerlegen Sie folgende Aussagen:

a) Es gilt immer $H(S) \leq H(S_M)$.

b) Es gilt immer $H(S) \geq H(S_M)$.

c) Es gilt immer $|\Sigma| \leq |\Sigma_M|$.

d) Es gilt immer $|\Sigma| \geq |\Sigma_M|$.