

Datenkompression

Sommersemester 2018

Übungsblatt 3

Prof. Ohlebusch

Institut für Theoretische Informatik

Uwe Baier

Ausgegeben am 31.05.2018

Besprechung am 07.06.2018

Aufgabe 3.1

Die Zeichenfolge $S := \text{MISS MISSION MISSISSIPPI}$ über dem Quellenalphabet $\Sigma := \{\text{M, I, S, O, P, N}\}$ soll komprimiert werden. Die Leerzeichen dienen dabei nur der Formatierung und sollen nicht mitkodiert werden.

- Verwenden Sie das LZ77-Verfahren (Suchpuffer 7 Zeichen und Codierpuffer 3 Zeichen lang). Welche Ausgabesymbole könnte man weglassen, um die Kompressionsrate zu verbessern?
- Verwenden Sie das LZSS-Verfahren (Suchpuffer 7 Zeichen und Codierpuffer 3 Zeichen lang). Ab welcher Länge sollte man hier (anstatt die Zeichen selbst zu codieren) das Format mit Offset und Länge verwenden?
- Verwenden Sie das LZ78-Verfahren.
- Welches der Verfahren komprimiert S am besten, welches am schlechtesten?

Aufgabe 3.2

Wir betrachten einen LZW-Dekodierer mit folgendem Anfangswörterbuch:

| Index | Eintrag |
|-------|---------|
| 1 | a |
| 2 | b |
| 3 | c |

Dekodieren Sie die Folge 1,2,3,6,4,8 mit dem angegebenen LZW-Dekodierer.

Aufgabe 3.3

In dieser Aufgabe sollen die Kodierverfahren von Lempel Ziv parallelisiert werden. Dazu bezeichne stets p die Anzahl der verfügbaren Prozessoren, S den zu kodierenden Text und n die Länge des Textes. Beschreiben Sie sowohl für das LZ77-Verfahren (mit Gleitfenster), als auch für das LZ78-Verfahren eine Möglichkeit, die Verfahren zu parallelisieren. Erörtern Sie zusätzlich die Anzahl an zusätzlichen Paaren (Tripeln), die aufgrund der Parallelisierung im Vergleich zu den nicht-parallelisierten Varianten entstehen können.

Aufgabe 3.4

Leiten Sie den ursprünglichen String S der Burrows-Wheeler-Transformation $L = \text{yee}\$yaass$ her, d.h. invertieren Sie die Transformation.

Hinweis: S endet mit dem lexikographisch kleinsten Zeichen \$.