

Datenkompression

Sommersemester 2014

Übungsblatt 3

Prof. Dr. E. Ohlebusch

Institut für Theoretische Informatik

S. Arnold

Ausgegeben am 05.06.2014

Besprechung in der Übung am 12.06.2014

Aufgabe 3.1

Wir betrachten einen PPMC-Kodierer, der Kontexte der Länge ≤ 2 berücksichtigt.

- Geben Sie die Häufigkeitstabelle an – sortiert nach Kontext und Folgezeichen –, nachdem die Zeichenkette $S := \text{acbbcbaacbbac}$ kodiert wurde.
- Welche Folge von Escape-Symbolen $\$$ und Buchstaben wird verwendet, um ein c zu kodieren, das auf S folgt? Geben Sie auch die jeweiligen Wahrscheinlichkeiten zu diesen Symbolen an.
- Nach einem Escape-Sprung von einem Kontext xy zu einem kürzeren Kontext y lassen sich sofort alle bereits nach y aufgetretenen Zeichen ausschließen, die auch schon auf xy gefolgt sind. Wie könnte damit die Kodierung des auf S folgenden c verkürzt werden?

Aufgabe 3.2

Die Zeichenfolge $S := \text{annasnasseananas}$ über dem Quellenalphabet $\Sigma := \{a, e, n, s\}$ soll komprimiert werden.

- Verwenden Sie das LZ77-Verfahren (Such- und Kodierpuffer jeweils 7 Zeichen lang). Welche Ausgabebits könnte man weglassen, um die Kompressionsrate zu verbessern?
- Verwenden Sie das LZSS-Verfahren (Such- und Kodierpuffer jeweils 7 Zeichen lang). Warum sollte man hier bei Übereinstimmungen einer Länge größer als 2 das Format mit Offset und Länge verwenden, bei kürzeren Längen aber die Zeichen selbst kodieren?
- Verwenden Sie das LZ78-Verfahren.
- Welches der drei Verfahren komprimiert S am besten, welches am schlechtesten?

Aufgabe 3.3

Konstruieren Sie eine Zeichenkette, die sich mit LZ77 besser komprimieren lässt als mit LZ78.

Aufgabe 3.4

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 diesem LZW-Dekodierer.
- Werden in dem Trie eines LZW-Dekodierers Zeiger auf Kindknoten benötigt? Werden Zeiger auf Elternknoten benötigt?
- Beschreiben Sie einen LZW-Dekodierer, 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