

Der LZ78 Algorithmus

Kompression

Zu Beginn der Kompression ist das *Wörterbuch* und die *Präfix*-Zeichenkette leer. Der Kompressor liest nun ein Zeichen "z" aus dem *Eingabedatenstrom*. Ist die Zeichenkette "*Präfix+z*" im Wörterbuch enthalten, so wird der *Präfix* um z erweitert. Dieses Einlesen und Erweitern wird solange fortgesetzt, bis "*Präfix+z*" nicht im Wörterbuch enthalten ist. In diesem Fall werden dann der den *Präfix* repräsentierende *Kode* und das momentane Zeichen z ausgegeben. Schließlich wird "*Präfix+z*" noch im *Wörterbuch* eingetragen und der Kompressor beginnt mit einem neuen *Präfix*. Ein Spezialfall liegt vor, wenn der *Präfix* leer ist und z nicht im *Wörterbuch* vorkommt. Dann wird ein spezieller *Kode*, der einen leeren String repräsentiert, und z ausgegeben. Schließlich wird z in das *Wörterbuch* eingetragen.

Der Algorithmus in Einzelschritten

1. Zu Beginn sind *Präfix* und das *Wörterbuch* leer.
2. $z :=$ nächstes Zeichen aus dem *Eingabedatenstrom*
3. Ist "*Präfix+z*" im *Wörterbuch* ?
 - o JA: *Präfix* := "*Präfix+z*"
 - o NEIN:
 1. Gebe den *Kode* für "*Präfix*" und z aus
 2. Trage "*Präfix+z*" im *Wörterbuch* ein.
 3. Leere *Präfix*
4. Ist das Ende des *Eingabedatenstroms* erreicht ?
 - o NEIN: Gehe zu Schritt 2.
 - o JA: Wenn *Präfix* nicht leer ist, gebe seinen korrespondierenden *Kode* aus.

Beispiel

Kleinbuchstaben: Schon verarbeitete Zeichen
 Großbuchstaben: Noch zu verarbeitende Zeichen
 Ausgabe: Kode, Zeichen

Position	0	1	2	3	4	5	6	7	8
Eingabedatenstrom	A	A	B	C	A	A	A	B	C

Kodierungsvorgang

Position	verarbeitet:nicht verarbeitet	Kode:Wörterbucheintrag	Ausgabe
0	:AABCAAABC	1:A	0,A
1	a:ABCAAABC	2:AB	1,B
3	aab:CAAABC	3:C	0,C
4	aabc:AAABC	4:AA	1,A
6	aabcaa:ABC	5:ABC	2,C