LZW-Kompression

Die LZW-Kompression ist ein **Wörterbuchverfahren** nach Lempel-Ziv-Welch.



Wörterbuchverfahren hinterlegen wiederkehrende Zeichenfolgen in einem Wörterbuch. Kommen diese Zeichenfolgen dann im zu komprimierenden Text erneut vor, reicht ein Verweis auf diesen Eintrag. Das LZW-Verfahren arbeitet dabei mit einem dynamischen Wörterbuch, welches direkt während der Kompression selbst erzeugt wird und damit keinen zusätzlichen Speicherplatz benötigt.

Um Platz für das Wörterbuch neben den normalen (ASCII-)Zeichen zu schaffen, reichen 8 Bit nicht aus. Für gewöhnlich werden 12 Bit für jedes Zeichen bzw. jeden Wörterbucheintrag verwendet. Das Wörterbuch kann also maximal $2^{12} = 4096$ Zeichen und Zeichenkombinationen beinhalten, wovon die ersten 256 Einträge bei Texten fest mit den ASCII-Zeichen vorbelegt sind.

Die Codierung verläuft nach folgendem Algorithmus:



- 1. Lies eine **möglichst lange** Zeichenkette ein, die bereits im Wörterbuch steht. Zu Beginn ist das jeweils nur ein einzelnes Zeichen!
- 2. Schreibe den Code des gefundenen Eintrags in die Ausgabe.
- 3. Lege aus der eben gefundenen Zeichenkette und dem **nachfolgenden** Zeichen einen neuen Wörterbucheintrag mit der nächst möglichen Codierung an.
- 4. Wenn nötig wird das letzte Byte der Ausgabe mit 0 aufgefüllt

Beispiel

Codierung

Die Zeichenkette BABAABBAA soll mit LZW **codiert** werden. Das Wörterbuch ist zu Beginn des Vorgangs im Bereich von 000_{16} bis $0FF_{16}$ mit den ASCII-Zeichen befüllt¹⁾. Zum besseren Verständnis des weiteren Ablaufs sollte man im Hinterkopfhaben, dass der ASCII Code des großen A 65_{10} =41₁₆ ist, der des großen B 66_{10} =42₁₆

ı	Dez	Hex	Sym	
	64	40	0	1
	65	41	Α	1
	66	42	В	1
	67	43	С	1

Noch zu bearbeitende Zeichenkette	Gefundener Eintrag	Ausgabe (12Bit)	Neuer Wörterbucheintrag
B ABAABBAA	B ← 042 ₁₆	042 ₁₆	BA → 100 ₁₆
A BAABBAA	A ← 041 ₁₆	041 ₁₆	AB → 101 ₁₆
BAABBAA	BA ← 100 ₁₆	10016	BAA → 102 ₁₆
AB BAA	AB ← 101 ₁₆	101 ₁₆	ABB → 103 ₁₆

Noch zu bearbeitende Zeichenkette	Gefundener Eintrag	Ausgabe (12Bit)	Neuer Wörterbucheintrag
BAA	BAA ←102 ₁₆	102 ₁₆	

Die Zeichenfolge wird also folgendermaßen codiert: **042041100101102**₁₆. Das sind 7,5 Bytes ²⁾. Die Kompression ist also bei solch kurzen Zeichenketten noch nicht drastisch - wenn man sich jedoch vorstellt, dass das Wörterbuch stets längere Zeichenketten mit einem einzigen 12Bit Code zugreifbar macht, kann die Kompression unter Umständen bei längeren Texten deutlich stärker ins Gewicht fallen.

Decodierung

Bei der **Decodierung** werden 12-Bit-Blöcke eingelesen. Das Wörterbuch wird während des Vorgangs mit Einträgen befüllt die aus dem ersten Zeichen des aktuellen Eintrag und dem vorangehenden Eintrag bestehen. Wir nehmen den codierten String von oben: **042041100101102**₁₆.

Aktueller 12Bit-Block (Hexadezimal)	Gefundener Eintrag (erster Buchstabe)	Neuer Wörterbucheintrag	Ausgabe
042	B (B)		В
041	A (A)	BA = 100 ₁₆	А
100	BA (B)	AB = 101 ₁₆	ВА
101	AB (A)	BAA = 102 ₁₆	AB
102	BAA (B)	ABB = 103 ₁₆	BAA

Decodiert lautet der Text also BABAABBAA.

Anmerkung: Das Wörterbuch musste zur Decodierung **nicht** gesondert übertragen werden - es "ensteht" während des Vorgangs.

Material

01_lzw-vorlage-codierung.odt	479.1 KiB 03.10.2022 16:53
01_lzw-vorlage-codierung.pdf	74.6 KiB 03.10.2022 16:53
01_lzw-vorlage-decodierung.odt	478.4 KiB 03.10.2022 16:53
01_lzw-vorlage-decodierung.pdf	73.4 KiB 03.10.2022 16:53
06-kompression-lzw.odp	43.0 KiB 29.09.2022 07:20
06-kompression-lzw.pdf	157.1 KiB 29.09.2022 07:20
ab.png	9.9 KiB 03.10.2022 15:58
lzw-codierung-beispiel.png	130.2 KiB 14.01.2025 09:15
lzw-decodierung-beispiel.png	94.9 KiB 14.01.2025 09:06
lzw_a2_cod_dec.odp	78.4 KiB 05.12.2023 14:01
lzw_a2_cod_dec.pdf	81.5 KiB 05.12.2023 14:01

https://info-bw.de/ Printed on 26.07.2025 18:16

26.07.2025 18:16 3/3 LZW-Kompression

pixel.png

2.0 KiB 03.10.2022 17:01

I)

Die ersten 256 Zeichen des 12 Bit Raums, der für die Codierung zur Verfügung steht wenn man mit ganzen Bytes arbeiten möchte, wird das mit Nullen zu 8 Byte aufgefüllt

From:

https://info-bw.de/ -

Permanent link:

https://info-bw.de/faecher:informatik:oberstufe:codierung:lzw:start?rev=1664815010



