

Dr. Jürg M. Stettbacher

Neugutstrasse 54
CH-8600 Dübendorf

Telefon: +41 43 299 57 23

Email: dsp@stettbacher.ch

Übung

LZ Codierung



Eine diskrete gedächtnislose Quelle (DMS¹) erzeugt zu jedem diskreten Zeitpunkt $n = 0, 1, 2, \dots$ eine Zufallsvariable $X[n]$, deren Wert 8 Bit umfasst. Das heisst, es entsteht mit der Zeit eine Sequenz $X[.]$ wie oben dargestellt.

1. Codieren Sie den Ausschnitt $X[.]$ mit dem LZ-77 Verfahren.
Der Suchbuffer sei 15 Zeichen lang, der Vorschabuffer 8 Zeichen.
2. Wieviele Bit umfasst ein LZ-77 Token?
3. Im weiteren Verlauf der Codierung mit LZ-77 bilden wir den Token (14, 8, A).
Was ist das Problem dabei?
4. Codieren Sie den Ausschnitt $X[.]$ mit dem LZ-78 Verfahren.
Das Wörterbuch sei für 63 Einträge ausgelegt.
5. Wieviele Bit umfasst ein LZ-78 Token?
6. Codieren Sie den Ausschnitt $X[.]$ mit dem LZW Verfahren.
Das Wörterbuch sei für 511 Einträge ausgelegt.
7. Wieviele Bit umfasst ein LZW Token?
8. Welches Verfahren erzielt die beste Kompressionsrate² R ?
(Zählen Sie nur vollständige Token und die damit übertragenen Quellsymbole.)
9. Verifizieren Sie Ihre Token aller drei Verfahren, indem sie diese wieder decodieren.

¹ Englisch: Discrete Memoryless Source.

² Die Kompressionsrate R ist definiert als die Anzahl Bits am Ausgang des Encoders dividiert durch Anzahl Bits am Eingang.

Antworten

1. LZ-77 Codierung:

| Nr. | Text | Token |
|-----|-------|------------|
| 1 | C | (0, 0, C) |
| 2 | CA | (1, 1, A) |
| 3 | CB | (2, 1, B) |
| 4 | CCC | (5, 2, C) |
| 5 | BB | (4, 1, B) |
| 6 | CCBC | (4, 3, C) |
| 7 | CBCCA | (11, 4, A) |
| 8 | B... | |

2. LZ-77 Token: $4 + 3 + 8 = 15$ Bit

3. Da der Vorschabuffer nur 8 Zeichen umfasst, ist eine maximale Länge der Übereinstimmung von 7 möglich und das achte Zeichen wird als zusätzliches Zeichen im Token vermerkt.

4. LZ-78 Codierung:

| Index | Eintrag | Token |
|-------|---------|--------|
| 0 | (null) | - |
| 1 | C | (0, C) |
| 2 | CA | (1, A) |
| 3 | CB | (1, B) |
| 4 | CC | (1, C) |
| 5 | CBB | (3, B) |
| 6 | CCB | (4, B) |
| 7 | CCBC | (6, C) |
| 8 | CAB | (2, B) |
| 9 | ... | |

5. LZ-78 Token: $6 + 8 = 14$ Bit

6. LZW Codierung:

| Index | Eintrag | Token |
|-------|---------|-------|
| ... | ... | |
| 65 | A | |
| 66 | B | |
| 67 | C | |
| ... | ... | |
| 256 | CC | (67) |
| 257 | CA | (67) |
| 258 | AC | (65) |
| 259 | CB | (67) |

| Index | Eintrag | Token |
|-------|---------|-------|
| 260 | BC | (66) |
| 261 | CCC | (256) |
| 262 | CBB | (259) |
| 263 | BCC | (260) |
| 264 | CBC | (259) |
| 265 | CCB | (256) |
| 266 | BCCA | (263) |
| 267 | AB | (65) |
| | ... | |

7. LZ-78 Token: 9 Bit

8. Kompression:

| Verfahren | #Token | #Bit/Token | #Zeichen | R |
|-----------|--------|------------|----------|--|
| LZ-77 | 7 | 15 | 19 | $\frac{7 \cdot 15}{19 \cdot 8} = 0.69$ |
| LZ-78 | 8 | 14 | 20 | $\frac{8 \cdot 14}{20 \cdot 8} = 0.70$ |
| LZW | 12 | 9 | 19 | $\frac{12 \cdot 9}{19 \cdot 8} = 0.71$ |

Bei diesem (kurzen) Beispiel komprimiert das LZ-77 Verfahren am besten.

9. LZ-77: (C C A C B C C C B B C C B C C B C C A ...)
 LZ-78: (C C A C B C C C B B C C B C C B C C A B ...)
 LZW: (C C A C B C C C B B C C B C C B C C A ...)