

Kompression

FH Augsburg

Steve Moser

`smoser@anno74.de`

FH-Augsburg, DVA-Seminar III

Ein paar Worte vorweg

1. Was ist Datenkompression
2. Wer braucht Datenkompression
3. Internet & Co
4. LZW
5. Huffman
6. Lzip
7. Fazit

Was ist Datenkompression

- Algorithmen
 - verlustfrei (für Texte, wichtige Daten ...)
 - verlustbehaftet (Audio, Video, Grafik)
- Komprimierungsfaktor abhängig von
 1. Art der Information
 2. Struktur der Information
 3. Qualität der Struktur
 4. Aufwand für die Komprimierung und Dekomprimierung
 5. Bereitschaft Informationsverluste hinzunehmen

Wer braucht Datenkompression

- Privatanwender, Industrie und Wirtschaft
 1. Sicherung (Archivierung)
 2. reduzierte Speicherkosten
 3. schnellere Datenübertragung

Internet & Co -Grafik

- GIF
- JPEG
- JPEG2000
- PNG

- MPEG1 (VCD)

Internet & Co -Video

- MPEG1 (VCD)
- MPEG2 (SVCD)

Internet & Co -Video

- MPEG1 (VCD)
- MPEG2 (SVCD)
- MPEG3

Internet & Co -Video

- MPEG1 (VCD)
- MPEG2 (SVCD)
- MPEG3
- MPEG4 (Avatare)

Internet & Co -Audio

- MP3 (MPEG1 Audio Layer 3)
- MPEG2 AAC
- Ogg Vorbis

Lempel Ziv - Familie

- LZ77 (1977)

- Abraham Lempel, Jakob Ziv

| | | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|---|
| Stelle | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Zeichen | a | a | b | c | b | b | a | b | c |

| Stelle | Hit | Zeichen | Code |
|--------|-----|---------|--------|
| 1 | - | a | (0,0)a |
| 2 | a | b | (1,1)b |
| 3 | - | c | (0,0)c |
| 4 | b | b | (2,1)b |
| 5 | ab | c | (5,2)c |

Lempel Ziv - Familie

- LZ78 (hinzufügen eines Wörterbuches zur Laufzeit)

| | | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|---|
| Stelle | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Zeichen | a | b | b | c | b | c | a | b | a |

| Stelle | Wörterbuch | Code |
|--------|------------|-------|
| 1 | 1:a | (0,a) |
| 2 | 2:b | (0,b) |
| 3 | 3:bc | (2,c) |
| 4 | 4:bca | (3,a) |
| 5 | 5:ba | (2,a) |

Lempel Ziv - Familie LZW

- Lempel Ziv Welch
 - Weiterentwicklung des LZ78 durch Terry Welch
 - patentiert seit 1984
- Einsatzgebiete u.a in GIF, compress (Unix)

Lempel Ziv - Familie LZW

| | | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|---|
| Stelle | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Zeichen | a | b | b | a | b | a | b | a | c |

Im Wörterbuch bereits enthalten: 1:a, 2:b, 3:c

| Stelle | Wörterbuch | Code |
|--------|------------|------|
| 1 | 4:ab | (1) |
| 2 | 5:bb | (2) |
| 3 | 6:ba | (2) |
| 4 | 7:aba | (4) |
| 6 | 8:abac | (7) |
| 9 | - | (3) |

Huffman-Codierung

1. Ursprung im Morsealphabet
2. Wahrscheinlichkeitsverteilung
Oft auftretende Zeichen = kurze Codierung
Selten auftretende Zeichen = lange Codierung
3. Vorteile
 - einfacher Algorithmus
 - Kompression bis zu 66%
 - lässt sich auf beliebige Daten anwenden
4. Nachteile
 - Codierungstabelle kann Ergebnis verschlechtern

- Offizielles Projekt auf Sourceforge
- Komprimierung bis auf 0% der Originaldatei
- :-)

Noch Fragen?

Vielen Dank für Ihre Aufmerksamkeit