LU04a - Komprimierung: Einführung

Siehe auch http://de.wikipedia.org/wiki/Datenkompression

Einleitung

Bei der Komprimierung wird die Datenmenge reduziert, indem überflüssige Informationen entfernt werden. Dabei werden die Daten so verkleinert, dass sich alle (verlustfrei) oder die wichtigsten (verlustbehaftet) Daten speichern lassen. Dadurch wird der benötigte Speicherplatz und die Übertragungszeit der Daten über ein Netzwerk reduziert.

Kompressionsfaktor

Der Kompressionsfaktor berechnet sich aus dem Speicherbedarf der komprimierten Daten gegenüber den ursprünglichen Daten. Je grösser der Faktor, desto grösser ist die Kompression.

```
                              Speicherbedarf der komprimierten Daten * 100
Kompressionsfaktor = 100 - ( ---------------------------------------------- )
                                Speicherbedarf der ursprünglichen Daten
```

Beispiel

- Unkomprimierte Textdatei: 578 Kilobyte
- Komprimierte Datei: 346 Kilobyte
- Kompressionsfaktor = `100 - (346 * 100 / 578)` = 40.138 %

Verlustfreie Datenkompression

Bei der verlustfreien Datenkompression können die ursprünglichen Daten wieder hergestellt werden. Dabei werden Redundanzen innerhalb der Daten erkannt und zusammen gefasst. Es werden also nur Informationen entfernt, die später wieder rekonstruiert werden können.

Anwendung

Textdokumente

Wenn Sie ein Textdokument komprimieren, können nur verlustfreie Verfahren zum Einsatz kommen. Schliesslich wollen Sie den Text später vollständig wiederherstellen können. Um einen Text zu komprimieren, können verschiedene Verfahren eingesetzt werden:

- Wiederholungen zusammenfassen (siehe Run length encoding).

- Wörterbuch (siehe Wörterbuch)
- Entropiecodierung (siehe Entropie)

Ausführbare Programme

Wie bei den Textdokumenten müssen Programme mit verlustfreien Methoden komprimiert werden. Dabei kommen die gleichen Verfahren wie bei Textdokumenten zum Einsatz.

Verlustbehaftete Datenkompression

Bei diesen Verfahren können die ursprünglichen Daten nicht mehr vollständig rekonstruiert werden.

Anwendung

Bildformate

Bilder benötigen sehr viel Speicherplatz. Besonders im Internet möchte man die Datenmenge reduzieren, damit die Seiten schneller laden. Das JIFF-Format (Dateiendung .jpg) komprimiert die Daten, indem ein Teil der Farbinformationen entfernt werden. Details finden Sie unter JFIF.

Musikdateien

Beim digitalen Speichern einer Musikdatei geht immer ein Teil der Information verloren. Ein Ton entsteht durch Schwingungen der Luft, welche wir als Kurven darstellen. Beim digitalisieren der Töne (Sampling) können nicht alle Feinheiten der Kurven abgebildet werden. Einige Beispiele finden Sie unter

- https://praxistipps.chip.de/was-ist-mp3-einfach-und-verstaendlich-erklaert_41919
- https://de.wikipedia.org/wiki/Abtastrate

Werkzeuge

Jeder Benutzer eines Computers verwendet in seinem Alltag viele komprimierte Dateien, ohne es bewusst zu merken. Fast alle Dateiformate die wir täglich benutzen verwenden eine Form der Kompression. Hin und wieder setzen wir aber auch bewusst Werkzeuge ein, um Daten zu komprimieren.

Die meisten Werkzeuge kombinieren verschiedene Verfahren um die Datenmenge möglichst stark zu reduzieren.

ZIP

Das ZIP-Dateiformat besteht zunächst aus einem Datencontainer, der mehrere Dateien enthält. Innerhalb dieses Contrainers werden die einzelnen Dateien komprimiert. Daher ist es möglich, einzelne Dateien aus einem ZIP-Archiv zu entpacken.

Unter Microsoft Windows ist ZIP das am häufigsten eingesetzte Werkzeug zum Komprimieren der Daten. Das liegt unter anderen daran, dass neuere Versionen von Windows direkt mit Dateien in ZIP-Archiven arbeiten können. Es ist also in vielen Anwendungsfällen gar nicht nötig, die Dateien zu entpacken.

7-Zip

siehe auch http://de.wikipedia.org/wiki/7-Zip

7-Zip ist ein freies Werkzeug zum Komprimieren und Entpacken von Dateien. Es verwendet standardmässig Archive im 7z-Format (Dateiendung .7z), kann aber viele andere Formate lesen und schreiben. Damit ist es ein nützliches Werkzeug um unterschiedliche komprimierte Dateiformate zu verarbeiten.

Zusätzlich bietet 7-Zip die Möglichkeit, die komprimierten Daten zu verschlüsseln.

tar, gz und bz

Diese Formate bzw. Programme sind unter Linux und bei Software-Entwicklern weit verbreitet.

Das Programm „tar“ dient dazu, mehrere Dateien zu einer einzigen Datei zusammen zu fassen. Eine solche Datei hat die Endung „.tar“ und wird als „Tarball“ (dt. Teerkugel) bezeichnet. Dies hat den Vorteil, dass ein Benutzer nur eine einzige Datei herunterladen muss.

Weil „tar“ die Daten nicht komprimiert, wird in der Regel der Tarball anschliessend komprimiert. Dazu kommen Komprimierungsprogramme wie „gzip“ (Dateiendung .gz) oder „bzip2“ (Dateiendung .bz2) zum Einsatz. Die komprimierte Datei hat eine Dateiendung wie

- .tar.gz bzw. .tgz (Tarball mit gzip komprimiert)
- .tar.bz2 bzw. .tbz (Tarball mit bzip2 komprimiert)

---

m114-A0G

Marcel Suter

From:
https://wiki.bzz.ch/ - **BZZ - Modulwiki**

Permanent link:
https://wiki.bzz.ch/de/modul/m114/learningunits/lu04/komprimierungeinfuehrung

Last update: **2026/01/28 21:12**