

Datenkomprimierung

Klaus Gloner

Das explosionsartige Anwachsen der Programmgrößen stellt wohl jedermann eines Tages vor die Entscheidung: Entweder eine größere Festplatte anschaffen oder einen Onlinekomprimierer einsetzen. Im nachfolgenden wollen wir die Arbeitsweise einiger bekannter Vertreter dieser Programmspezies besprechen.

Der Gedanke der Datenkomprimierung hat seine Anfänge in den 40er Jahren, wo die Prinzipien als Teil der Informationstheorie formuliert wurden. In den 70er und 80er Jahren wurden diese theoretischen Ansätze verbessert und weiter entwickelt. Die verwendeten Algorithmen basieren auf den Arbeiten von Lempel, Ziv und Welch, weshalb man auch von LZW oder LZ - Verfahren spricht.

Starke Impulse für die Entwicklung von Packprogrammen gingen in den 70er Jahren von den Jüngern der damals entstehenden Datenfernübertragung aus, die ihre Dateien mit der für heutige Zeiten unvorstellbaren Geschwindigkeit von zunächst 300 Bit/s übertrugen. Die Datenfernübertragung ist im Übrigen auch der (beinahe) einzige Bereich, wo Hardware-Onlinekomprimierer in großem Umfang eingesetzt werden, ja zum Standard wurden: Jedes Modem, das V.42bis bzw. MNP 5 "versteht", ist mit einem Online - Komprimierer, der bis maximal 4 : 1 bzw. 2 : 1 komprimiert, ausgerüstet. Beachten Sie, daß diese Kompressionsraten für eine Standard-Textdatei gelten und in der Praxis kaum zu erreichen sind!

Als die PC's noch jung waren, in der Ära der 8086er und frühen 80286er, gab es - von den Anwendern allerdings kaum bemerkt - eine Reihe von I/O - Karten, die Daten auf ihrem Weg von und zur Festplatte komprimierten. Softwarekomprimierer konnten für solche Echtzeitkomprimierung wegen der mangelnden Leistungsfähigkeit der Prozessoren nicht eingesetzt werden.

Kompressionsverfahren unterteilen sich in 2 Arten: Kompression mit Verlusten und Kompression ohne Verluste. Die Kompression mit Verlusten wird bei der Kompression von Bilddateien angewendet, wo es zu Qualitätsverlusten kommt. Bekannte Vertreter dieser Gattung sind Programme, die nach dem JPEG-, MPEG- und Indeo-Verfahren arbeiten. Die erreichten Kompressionsraten liegen bei 100 : 1.

Für Programm- und Datendateien ist ein solches Verfahren natürlich nicht anwendbar, hier arbeitet man mit der Kompression ohne Verluste, die es, stark abhängig von der Dateiart, auf Kompressionsraten von 2 : 1 (*.EXE - Dateien) bis 12 : 1 (z. B. dBase - Dateien) bringt. Die 3 bekanntesten und leistungsfähigsten Packprogramme sind derzeit PKZip, Arj und Lha.

Eine Untergattung dieser Programmart, die die Fähigkeit hatte, Programmdateien zu komprimieren, die trotzdem lauffähig bleiben, führte und führt ein Schattendasein. PKLite und Lzexe sind 2 Vertreter dieser Gattung, ein anderer Vertreter, Diet, dürfte der erste Online - Komprimierer gewesen sein.

Wie arbeitet ein Komprimierer? Nun, im Prinzip ganz einfach: Er untersucht eine Datei nach identischen Zeichenfolgen - denken Sie z. B. an große, gleichfarbige Flächen in Grafikdateien. Ein anderes Beispiel für solche Redundanzen sind dBase - Dateien, die, bedingt durch die gleiche Datenfeldlänge, große Menge an Leerzeichen enthalten.

Als Beispiel werden immer wieder gerne 2 Zeilen aus einem Musical zitiert:

"The rain in Spain falls mainly on the plains"

Diese Textpassage enthält mehrmals die Buchstabenfolge "ain" und "the" und ist 45 Byte lang. Ein LZW - Pac??ker macht daraus:

"The rain in Sp[9,4]falls m[11,3]ly on [34,4]pl[5,3]s."

Die erste Zahl in der eckigen Klammer ist der "Pointer", die zweite Zahl ist der "Token". [9,4] und bedeuten in unserem Beispiel: "Gehe in der Datei 9 Zeichen zurück und füge die folgenden 4 Zeichen an dieser Stelle ein".

Mit dieser, natürlich sehr primitiven Methode würde die Passage aus "My Fair Lady" immerhin schon rund 20 % weniger Speicherplatz benötigen.

Packprogramme wie PKZip, Arj oder Lha untersuchen nach dieser Methode eine Datei komplett und kommen dadurch zu sehr guten Kompressionsergebnissen.

Onlinekomprimierer müssen dagegen große Datenmengen in Echtzeit packen und gehen deshalb einen Kompromiß ein: Dateien werden nur durch ein "Fenster" betrachtet und komprimiert werden nur Bytefolgen, die in einem "sliding dictionary" (die deutsche Übersetzung "Wörterbuch" gibt m. E. nicht genau das wieder, was gemeint ist) enthalten sind. Dadurch verringert sich die Kompressionsrate beispielsweise bei dBase - Dateien von 12 : 1 auf etwa 7 : 1. Da, wo es noch schneller gehen muß, bei Modems oder Streamern sinkt die Kompressionsrate noch weiter ab.

Die bekanntesten Online-Komprimierer sind Dblspace (Wird ab MS-DOS 6.0 als Bestandteil des Betriebssystems geliefert), SuperStor (in DR-DOS bis Version 6.0

enthalten), Stacker und XtraDrive und werden in einem eigenen Artikel besprochen.

Packprogramme

Als derzeitigen "Standard" kann man die als Shareware bzw. Freeware vertriebenen Programme PKZip, ARJ und LHA bezeichnen. Die Möglichkeiten, die die Programme bieten, sind weitgehend identisch, bei der Packgeschwindigkeit und der Kompressionsrate bestehen deutliche Unterschiede: Hier hat PKZip ganz klar die Nase vorn, gefolgt von ARJ. Die beiden Programme bieten auch die für größere Backups interessante Möglichkeit des Packens über mehrere Disketten. LHA dagegen ist ein Freeware-Programm, das damit also auch ohne weitere Lizenzzahlungen im kommerziellen Bereich eingesetzt werden kann.

Die Kompression von *.EXE- oder *.COM - Dateien, die dabei ausführbar bleiben, bieten Programme wie PKLite, LZEXE, Diet und Exepack an. PKLite, vom gleichen Autor wie PKZip ist in diesem Quartett derzeit das einzige "vernünftige" Programm, womit gemeint ist, daß es beide Dateitypen "beherrscht" und auch Dateien mit internen Overlays komprimieren kann. Diet dürfte der erste Online-Komprimierer gewesen sein, ist allerdings hoffnungslos veraltet und wird, genau wie LZEXE, das mit internen Overlays (Diese Overlays kommen z. B. in Clipper-Programmen, aber auch in vielen anderen *.EXE-Dateien sehr häufig vor) große Probleme hat, nicht mehr weiterentwickelt. Exepack ist ein Microsoft-Produkt, das ich aber nicht ernsthaft als Packprogramm bezeichnen möchte, die Leistungen sind, gelinde ausgedrückt, unbefriedigend.

Ein Manko haben PKLite und die übrigen, im vorigen Absatz behandelten Programme: Sie können nichts mit der Struktur von Windows - *.EXE-Dateien anfangen. Diese EXE - Dateien verfügen über einen erweiterten Dateikopf: Die ersten 256 Byte eines solchen Programmes bestehen aus dem üblichen Dateivorspann für EXE - Dateien, der mit der Meldung "Dieses Programm benötigt Microsoft Windows" endet. Darauf folgt dann der Windows-spezifische Dateivorspann, der von einem herkömmlichen Packer komprimiert und damit nicht mehr ausführbar gemacht werden würde.

Ein in Amerika entwickeltes Programm, Winlite, verspricht hier Abhilfe. Es ist speziell zum Komprimieren von WINDOWS - Programmen gedacht. Tests mit diversen Windows-Programmen haben allerdings keine guten Erfahrungen gebracht: Der Großteil der mit Winlite behandelten EXE - Dateien waren nach der Komprimierung regelrecht verstümmelt.

Eine uneingeschränkte Empfehlung verdient sich dagegen ein anderes, relativ unbekanntes Programm: dLite, das einen sehr interessanten Weg geht: Es besteht aus einem Kompressionsprogramm, das alle Arten von Dateien, also sowohl Daten- wie auch Programmdateien komprimiert. Auch mit Windows - Programmdateien, seien es nun EXE-, DLL- oder andere Dateien gibt es nach unseren Beobachtungen keine Probleme. Das Entpacken zur Laufzeit besorgt ein kleiner, speicherresidenter Treiber, der gerade mal 10 kB Hauptspeicher belegt und sich mit Quemm z. B. hochladen läßt - nicht aber mit LOADHIGH von DOS!

Die Kompressionsraten können sich mit vergleichbaren Packprogrammen durchaus messen. Hier zwei Werte aus der Praxis:

Windows-Verzeichnis

- vor der Komprimierung 5.015.908 Bytes
- nach der Komprimierung 3.491.810 Bytes

Foxpro - Verzeichnis

- vor der Komprimierung 6.546.576 Bytes
- nach der Komprimierung 4.352.973 Bytes

Die vorgestellten Programme sind - mit einigen Utilities - auf 3 1/2" - Disketten zum Preis von S 180.-- (einschl. Porto und Verpackung) beim Autor erhältlich. Sie können sie unter dem Stichwort "Packer" anfordern.

Hier die Adresse: Klaus Gloner, Dr. Hainstrasse 8, 5620 Schwarzach - Tel. 06415/7215 - Fax 06415/7218 □