



Understanding Compression

Author: Debbie Grignani, 2BrightSparks Pte Ltd.

Website: www.2brightsparks.com

Many of us have experienced difficulties in sending data files like text documents and digital photos via email because the file size is just too great. That is where file compression comes in useful.

What is File Compression?

File compression enables a computer user to reduce the size of a data file so that it is more convenient for storage or sharing.

During file compression, excess information is eliminated when some kind of pattern in the data is recognized. There are different algorithms used to search for a pattern in order to reduce the size of the file, and most compression software programs use a variation of Lempel-Ziv (1977) Compression, also known as LZ77. LZ77 is an adaptive dictionary-based compression algorithm that builds a dictionary based on text that has been previously encountered.

A Simple Example of File Compression

To better understand this process we will look at a simple example. Take the quote by Aristotle: "One swallow does not make a summer, neither does one fine day; similarly one day or brief time of happiness does not make a person entirely happy."

Let us say that each character takes up 1 bit, therefore the size of the quote is 146 bytes. Ignoring capital letters, there are words that are repeated which we can review in the table that is shown on the next page.

Index Number	Repeated Word	Number of repetitive times
1	one	3
2	does	3
3	not	2
4	make	2
5	day	2

By substituting the index number in the place of the word, this is the result:

“1 swallow 2 3 4 a summer, neither 2 1 fine 5; similarly 1 5 or brief time of happiness 2 3 4 a person entirely happy.”

From a size of 146 bits to a 118 bit size, which converts to only a 19% compression rate. This example does not significantly compress the data, but you get the idea. More complex compression algorithms seek out patterns that include eliminating punctuation and spaces and therefore compressing files a lot more.

There are two forms of compression – **Lossless** and **Lossy** Compression.

Lossless Compression

What we have seen so far is known as lossless compression. That is when a compressed file is decompressed it looks identical to the original. No information is lost in the process of lossless compression. When compressing data and programs, lossless compression must be used.

Lossy Compression

In order to achieve greater compression, lossy compression is used. However, only certain types of data can accept lossy compression. They include graphics, audio and video files. In this compression method, some degree of data lost is inevitable since redundant and unnecessary information are literally eliminated and lost forever. The good side of this type of compression is that it reduces the size of the file tremendously.

Backup and Compression

The two tasks of backup and compression go hand in hand, since compression can help storage sizes of backups to be kept to a minimum. You can either manually compress your files using programs like WinZip before each backup is run, or simply make use of backup software that has file compression as an option.

[2BrightSparks Pte Ltd](#) has two options of backup and synchronization software that offer compression capabilities. A freeware version called SyncBackFree and commercial versions called [SyncBackSE](#) and [SyncBackPro](#). SyncBackSE and SyncBackPro come packed with more options and its compression capabilities a significantly more advanced as compared to SyncBackFree. Both programs provide ease of use and allow you to manage your storage space better with compressed backups.

2BrightSparks Pte Ltd.

www.2brightsparks.com

Power with Ease - Indispensable Utility Software