

Il computer a dieta

Cari amici lettori, compagni di queste premesse che hanno il solo scopo di avvisare di non prendere mai troppo sul serio quello che scrive il De Masi sulle pagine di MC, tanto, a leggerle o a non leggerle, il mondo non cambierà certo di un capello, sto scrivendo questo pezzo seduto sotto il porticato antistante la camera (definita, con somma modestia, miniappartamento - con quello che mi sono costate queste due settimane, potevo davvero dare un anticipo, ma per acquistarlo, un miniappartamento!), del villaggio dove le mie tre donne mi hanno ristretto, a passare le vacanze. Due giorni a ferragosto, nelle stradine assolate non passa nessuno, sono tutti in spiaggia a prendere pallonate in faccia dai giocatori di turno che si riscoprono campioni dopo un anno passato dietro la scrivania, o a farsi sgocciolare addosso dal ragazzino puntualmente orfano momentaneo di padre e madre, che per buona giunta non ho mai ben capito per quale motivo debba camminare nella sabbia arando con le zampe posteriori.

di Raffaello De Masi

Sono le dodici circa, tra poco tutti in sala a dare l'assalto alla diligenza del buffet, a ingozzarsi come oche, illudendosi poi di smaltire il tutto con tre minuti di jogging mattutino, puntualmente fatto con l'asciugamano gettato di traverso sul collo a dimostrazione di consumata arte e abitudine nell'esercizio fisico. Anche io, il primo giorno, mi carico di dodici primi e venti secondi, che forse neppure guarderò. Ma lo faccio ad arte, visto che quest'improvvisata saga dell'uso della forchetta scatterà, già lo so, la mia ulcera gastrica, così, stando poi male per un'intera notte, mi passerà completamente la voglia di mangiare porcherie e ritornerò al riso in bianco e al petto di pollo bollito.

E così, come ogni anno, le vacanze estive si trasformano in una perfetta occasione per ridurre il diametro cintola di uno o due centimetri. Insomma, una cura dimagrante in un momento in cui tutti



ingrassano; ingrassa perfino Jana, la mia bambina più piccola, un pulcino di sette anni che riesco a sollevare con una mano, anch'essa, come la prima, venuta dal freddo e dall'inferno di Chernòbyl ad illuminare di nuova luce casa mia e la sempiterna Atripalda. Io dimagrisco, niente da far gridare al miracolo, per l'amor di Dio, quel minimo che però dà una certa soddisfazione. Insomma, riduco il mio, ahimé sempre elevato, volume. E, proprio per questo motivo, quale migliore occasione per mettersi a parlare di dimagrimento dei nostri dati e dei

nostri file, che, beati loro, sono capaci di dimezzarsi nelle loro dimensioni?

Lo spazio vitale

Pur nella frenetica evoluzione che il computer stanno avendo negli anni, dove nulla è definitivo e nulla è destinato a durare a lungo, esistono pur sempre delle costanti, delle esigenze, dei punti di riferimento che continuano a rappresentare punti fissi di misura, verifica e confronto che, nel tempo, non hanno mai perso la loro importanza.

E ci sono dei punti fissi, delle chime-re irraggiungibili cui corre appresso qualunque utente e che difficilmente possono essere considerate raggiunte. Così nessuno, per quanto prodigo verso il suo PC, sarà mai soddisfatto della RAM presente, della velocità della propria connessione alla Rete, della capacità del suo disco rigido.

Fermiamoci a quest'ultima considerazione; oggi l'entry level degli HD ha almeno una dozzina di GB, gira a 7200 giri ed è spesso poco più di una fetta di pan carré, ma per poche decine di migliaia

Welcome to Xceed Software!

... (Detailed description of the website content follows, including product listings and navigation elements) ...

Welcome to ARJFolder Official Site

English Home

... (Detailed description of the ARJFolder website content, including file lists and download instructions) ...

sostituzione e le stringhe originali permette di ricostruire completamente un file, riportando alle sue esatte caratteristiche iniziali, neppure un bit in più o meno. Nel secondo caso la perdita di dati è definitiva; l'algoritmo cerca, infatti, dati che possono essere eliminati dal file senza danno. Nella maggior parte dei casi questa tecnica offre rapporti di compressione più elevati.

Sovente diversi programmi usano ambedue le tecniche descritte, vale a dire che viene prima applicata una tecnica con perdita di dati, seguita poi dall'altra. In questo caso appare evidente che i rapporti di compressione possono risultare elevati (per certi file grafici si può raggiungere anche il 90% e oltre); è altrettanto ovvio che le tecniche di compressione "lossy" non possono essere adottate per certi tipi di file, come i programmi, che non tollerano perdita alcuna di informazioni.

Monumento all'ingegno umano

Il più sperimentato dei metodi di compressione è legato indissolubilmente a un suffisso, .ZIP che è passato in breve tempo a sostantivo per antonomasia di file compresso, ed è divenuto lo standard de facto per la gestione dei file su Internet. Sebbene legato dall'abitudine e dal senso comune a un particolare programma, WinZip, lo standard è molto più universale, tanto che quasi tutti i programmi di compressione più usati (ricordiamo

le persone più sveglie e attente che abbia mai conosciuto, che a tal proposito mi chiese: "Ma che cosa fa, un programma di compressione? Toglie l'aria dai documenti? Un po' come le spugnette della Stanhome?"

La battuta può sembrare peregrina, ma non lo è; le spugnette formato sigaro, che bagnate moltiplicano il loro volume, praticamente sono realizzate eliminando dall'interno il componente inutile, l'aria. I file compressi mancano, da parte loro, delle parti di cui si può fare a meno,

o che in ogni caso, come l'aria, possono essere ripristinate, rimpiazzate in qualsiasi momento.

Sebbene esistano diversi metodi di compressione dei file, essi possono essere ricondotti a due tipi principali di procedura: senza perdita e con perdita di dati.

Nel primo caso viene utilizzato un algoritmo che ricerca stringhe che possono essere sostituite da altre più brevi. In caso di ricostruzione il possesso della tabella di corrispondenza tra le chiavi di

WinZip, <http://www.winzip.com> - ZipMagic, <http://www.ontrack.com>, e PKZip <http://www.pkware.com>) lo adottano come default.

Un file .ZIP è legato a un prototipo di compressione messo a punto, nel '77, da Abraham Lempel e Jacob Ziv; l'algoritmo LZ77 cerca, nel file, stringhe ridondanti, meglio se di dimensioni elevate, e le sostituisce con un puntatore a un indice che conserva, per così dire, tutte le abbreviazioni adottate. La differenza, mai comunque elevata, tra le va-

Mosaic Software Solutions
Home Download Order Contact The Admin FAQ

Welcome to the home of Mosaic Software Solutions, creators of AxMan, the industry leading file splitting utility!

AxMan 3.11 is now released! Download your free evaluation copy now!

AxMan has been inducted into the WUJHET Shareware Hall of Fame!

Copyright © 2000

made by Mike Hester (mohester@compuserve.com)

Mosaic Software Solutions
Home Download Order Contact The Admin FAQ

Welcome to the home of Mosaic Software Solutions, creators of AxMan, the industry leading file splitting utility!

AxMan 3.11 is now released! Download your free evaluation copy now!

AxMan has been inducted into the WUJHET Shareware Hall of Fame!

Copyright © 2000

made by Mike Hester (mohester@compuserve.com)

Check out the [screenshots](#) page to read what the experts are saying.

We are committed to bringing you only the finest quality shareware at reasonable prices. We also offer services to custom develop software for you - "You define it, we build it".

Please take a moment to browse around, and feel free to let us know if you have any questions by emailing muhammad@compuserve.com

HB ACTIVATED
Loading

Prepare for the outdoors with BIG SAVINGS!

Warrant Jarboe Timberland

per.com

CLICK HERE TO SHOP!

LinkOut.com

Batch Zip Toolkit
FREE DEMONSTRATION

CLICK HERE

Home Features Screenshots Download How to Buy Version History User Testimonials Technical Information Getting Support Frequently Asked Questions Contact Details Have a Suggestion?

Home Features Screenshots Download How to Buy Version History User Testimonials Technical Information Getting Support Frequently Asked Questions Contact Details Have a Suggestion?

Home Features Screenshots Download How to Buy Version History User Testimonials Technical Information Getting Support Frequently Asked Questions Contact Details Have a Suggestion?

Home Features Screenshots Download How to Buy Version History User Testimonials Technical Information Getting Support Frequently Asked Questions Contact Details Have a Suggestion?

Batch Zip Toolkit is an all in one file compression package that allows you to compress and decompress files individually, in batch mode with recursive batch. Our toolkit supports a wide variety of popular archive formats including Zip, Lha, Lzh, Ace and Rar.

Batch Zip Toolkit also allows you to test and fix Zip archives in batch, and also the facility to create self extracting archives. Click here for a more features!

Version 3.01 is now available. Click Here to Download

rie implementazioni, sta essenzialmente nella capacità dell'algoritmo usato di determinare il numero di byte, in una stringa, che il puntatore può convenientemente sostituire.

L'algoritmo LZ77 rappresenta lo zoccolo duro di tutta l'area, ma diverse migliorie sono state apportate, nel tempo, al prodotto iniziale. Oggi molti pacchetti di compressione applicano, alla forma iniziale, il metodo di codifica Shannon-Fano, che si basa su una brillante intuizione di questi due ricercatori. Il principio adottato si basa sulla frequenza di ripetizione di caratteri all'interno di un file. In altri termini ogni carattere ASCII viene ricodificato in funzione della sua frequenza, assegnando valori più bassi ai caratteri più frequenti. In questo modo è possibile assegnare a un nibble (gruppo di 4 bit), e non più a un byte (8 bit), la corrispondenza dei 16 caratteri più frequentemente ricorrenti in un file.

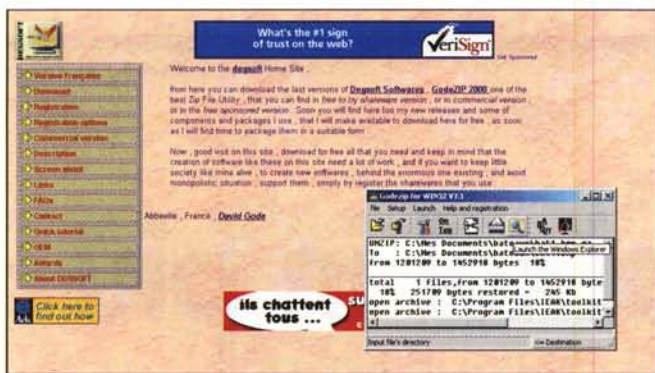
Operativamente la cosa funziona in questo modo; in base a precise verifiche statistiche si può essere ragionevolmente certi che il carattere più frequente in

uno scritto è lo spazio, seguito dalla vocale 'e' e dalla vocale 'a', e così via, mentre la palma della rarità, all'altro capo del principio, spetterà a caratteri poco frequenti, come %, &, \, o, nell'ambito delle lettere dell'alfabeto, alla z, q (magari anche maiuscole), alle accentate. L'algoritmo suddivide inizialmente tutto il contenuto del file in due gruppi di caratteri, quelli più frequenti e quelli meno, assegnando una prima codifica iniziale, a quelli e a questi, imponendo rispettivamente il primo (da sinistra) bit 0 o 1. Si continua così eseguendo una successiva verifica di frequenza tra quelli più usati, e di nuovo viene assegnato, con principio della maggiore o minore rappresentatività, il secondo bit con gli stessi valori. Ancora avanti con il terzo bit e poi con il quarto; da questo momento in poi tutti i caratteri rimasti avranno come primo nibble la sequenza 0000, con valori significativi solo nel secondo. Quindi sedici caratteri (quindici escludendo lo spazio) potranno

essere rappresentati da una sequenza di soli quattro bit.

Questo principio, pur teoricamente ineccepibile, ha avuto ancora spazio per un successivo affinamento. Infatti, e questo vale ovviamente solo per i file di testo, la tabella di distribuzione delle frequenze delle lettere dell'alfabeto non è identica in tutte le lingue. Così, mentre alcuni programmi e utility di compressione dispongono già di una tavola di riferimento fissa (basata sulla lingua inglese), altri pacchetti eseguono preliminarmente una scansione del file e costruiscono tabelle ad hoc, che verranno "agganciate" al file compresso stesso, a mo' di chiave di lettura.

Ovviamente questo tipo di analisi costa, in termini di tempo (i pacchetti di compressione di tal tipo vengono chiamati correntemente "a doppio passaggio"), ma l'eleganza applicativa e, so-



page, dove dati alfanumerici, non sempre suscettibili di trattamento con perdita, sono sovente frammisti a immagini, ideali per essere trattati in questo modo.

Inoltre i programmi di questa famiglia si sono sviluppati divenendo con il tempo degli "archiver", dei costruttori di archivi compressi. Gli esempi più noti sono PKZIP, WinZip e ZipMagic (ma basta sfogliare le pagine di shareware per trovarne infiniti altri altrettanto validi), che non solo comprimono dati, ma li "impacchettano" per così dire, in archivi compressi.

Un archivio è, per definizione, l'insieme di due o più file riuniti insieme in una unica identità fisica. L'utilità di questo tipo di prodotto è più che evidente; un archivio può contenere, ad esempio, un intero articolo di una rivista, comprensivo di testo e immagini, o un programma di installazione composto da file di programma, di dati, ancora immagini, .DLL ecc. Insomma, è la via maestra per distribuire prodotti composti da diversi file .PKZIP, e in genere, quasi tutte le altre utility di archiviazione usano come algoritmo di base lo LZSS, combinato con la codifica di Shannon-Fano nelle edizioni più vecchie e quella di Huffman nelle più recenti.

Non mancano analoghe applicazioni per altre piattaforme; ad esempio Stuffit (<http://www.aladdinsys.com>) è un sistema molto diffuso e noto in ambiente Macintosh, mentre in ambito UNIX è molto noto Gzip come compressore e Tar come archiver.

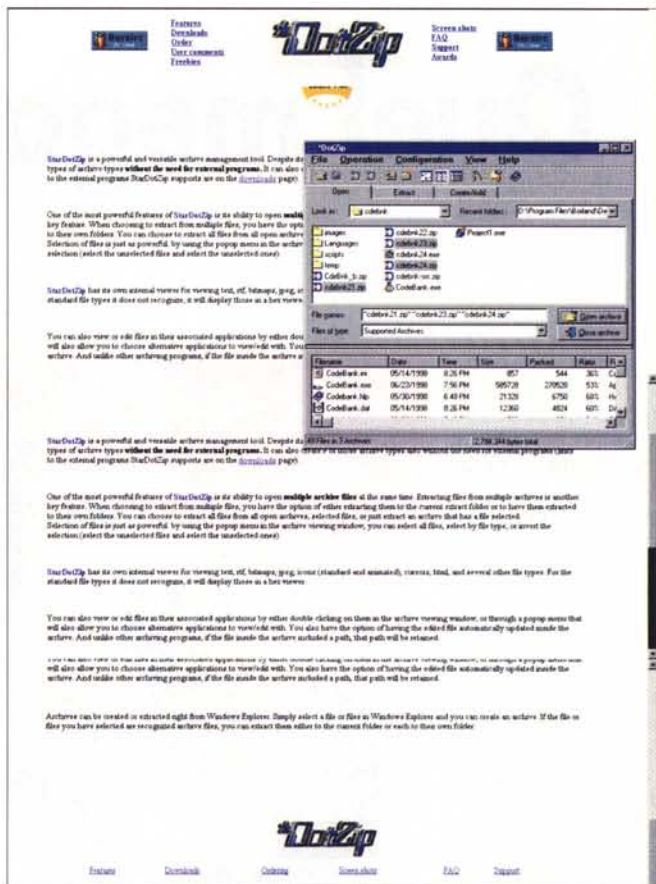
Da qui a passare all'idea di comprimere un intero disco rigido il passo è breve. L'idea non è recente; gli utenti Mac ricorderanno come DiskDoublers, elemento delle Now Utilities, potesse intervenire direttamente anche sul sistema operativo. Le normali tecniche utilizzabili attraverso PKZIP, WinZIP e soci sono applicabili in teoria anche a quest'uso, ma non sono facilmente maneggiabili. Così diversi produttori, compreso Microsoft, Cheyenne e Star hanno sviluppato, nel tempo, software destinato a maneggiare interi dischi rigidi; oggi Drive Space (Microsoft) è lo standard in questo campo applicativo.

DriveSpace 3, comparso originaria-

mente come accessorio di sistema operativo del W95, è presente anche in W98, è diretto discendente, anche nella tecnica di funzionamento, di DoubleSpace di MS-DOS 6. Il software rinomina l'hard disk con una lettera non utilizzata (generalmente H) e crea un archivio compresso del disco noto come CVF (Compressed Volume File). Il neonato drive H, noto come drive host, custodisce il CVF e una macro di boot dei file di sistema necessaria per lanciare Windows. L'età di DriveSpace è testimoniata dal fatto che può produrre solo CVF di 1 GB, che, a occhio e croce, rappresentano drive non compressi di 2-2,5 GB, valore ridicolo in questo nuovo mondo di dischi supervitaminizzati.

DriveSpace utilizza un algoritmo di compressione simile a WinZip per scoprire ed eliminare ridondanze di dati. Esso usa la codifica di Huffman per assicurare i riferimenti alle più comuni sequenze di dati usando il più ridotto numero di bit. In base a questi riferimenti, DSpace ricostruisce il file originario. Ovviamente, dati con elevata frequenza di ridondanza (immaginiamo un documento bit-mapped) saranno compressi meglio di altre forme di file. Così troveremo file di programmi applicativi in cui il livello di compressione sarà molto modesto, mentre documenti e, generalmente, immagini non già compresse attraverso altre forme potranno raggiungere rapporti elevati. Certi file, infine, come .GIF, .JPG, .MP3, non riceveranno alcun beneficio dalla "cura".

E non basta; a parte la perdita di prestazioni conseguente alle operazioni di compressione e decompressione, piut-



tosto sensibile su macchine di fascia bassa, l'operazione non era priva di rischi.

E' noto che Microsoft non incoraggiava l'uso di DoubleSpace, dichiarandolo, tout court, inaffidabile (al contrario DriveSpace, dalla versione 3, era molto più sicuro ed esente da bug). Ancora, DriveSpace impediva a ScanDisk di effettuare i suoi controlli, ed è completamente inutile su file system di tipo FAT 32, che, viceversa, offre notevoli migliorie nella distribuzione e nell'utilizzazione dello spazio su disco.

Conclusioni

Con questo termina la trattazione della compressione senza perdita di dati; i principi e le tecniche adottate e descritte sono quelle di base, anche se innumerevoli varianti sono state sviluppate dai diversi produttori di software.

Ma non è certo finita. La prossima volta discuteremo delle tecniche con perdita di dati. E vedremo più da vicino qualche pacchetto di compressione dalle prestazioni più interessanti.

A presto!

M&S