

Data preparation per CD-ROM

Nel corso dell'appuntamento con la rubrica di memorie ottiche abbiamo spesso analizzato il media CD-ROM, verificandone le caratteristiche tecniche hardware e software, introducendone il file system e gli standard logici di formattazione. In questo articolo parliamo invece delle metodologie di lavoro necessarie per la preparazione dei dati per un'applicazione da distribuire su CD-ROM

Il termine «preparazione dati» è genericamente usato per riferirsi alla creazione di file contenenti le informazioni volute ed alla scelta dei metodi di retrieval di suddette informazioni. Naturalmente anche un media in grado di ospitare 550 Megabyte è al centro di una simile problematica, soprattutto considerandone le sue caratteristiche peculiari, ossia la sola lettura e la lentezza di seek del CD.

In realtà per parlare di data preparation è necessario suddividere in almeno quattro fasi distinte la procedura di lavoro: acquisizione, strutturazione dei file, editing delle informazioni e riversamen-

to finale. Ovviamente le prime tre sono finalizzate alla creazione del database ed alla preparazione, o meglio alla formattazione, per il successivo trasferimento sul media ottico.

Prima della fase finale è però necessaria un'accurata procedura di test dell'applicativo realizzato, che spesso rivela errori di acquisizione o di logica di retrieval che necessitano quindi di ritornare indietro di uno o due step per la relativa correzione; è normale che esista un certo numero di queste sequenze test, scoperta errore, correzione, nuovo test prima di arrivare ad un disegno definitivo.

La fase finale, il riversamento o pre-mastering, porta tutti i dati in formato definitivo su un nastro o altro media (WORM, videotape), pronti per essere dati alle fabbriche che stampano fisicamente il CD. La versatilità del CD-ROM nell'accogliere diversi tipi di dati fa sì che un applicativo per tale media possa tranquillamente richiedere:

— file dati normali, come quelli sui floppy;

— immagini digitalizzate, da scanner o telecamera;

— sequenze audio;

— sequenze video.

Il primo passo della preparazione dati è quindi quello di convertire tutti questi tipi di dati in una struttura file comune, fatta però in modo che ciascuna informazione nel file sia riconoscibile da un drive CD-ROM.

Il passaggio successivo è quindi

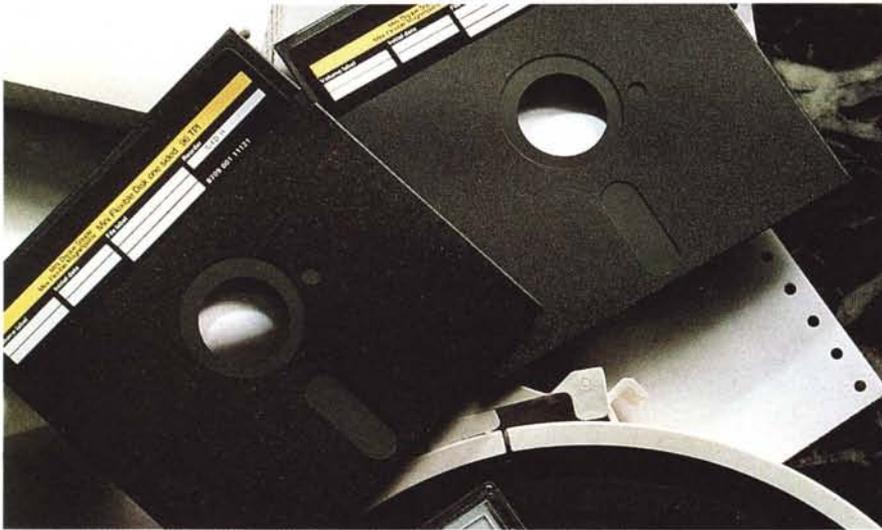
quello di verificare ed editare le informazioni, cosa non legata specificatamente all'ambiente delle memorie ottiche, ma comune a tutto il settore del data retrieval. Non ci sono quindi particolari procedure specifiche, se non quelle consuete di editing.

Creazione della struttura

L'organizzazione dei dati all'interno del CD-ROM deve essere quindi creata dopo la modifica definitiva dei file che ne comporranno le informazioni. Questo consente di verificare, quindi testare, il disco prima che sia effettivamente prodotto; questo viene fatto mediante quattro fasi distinte.

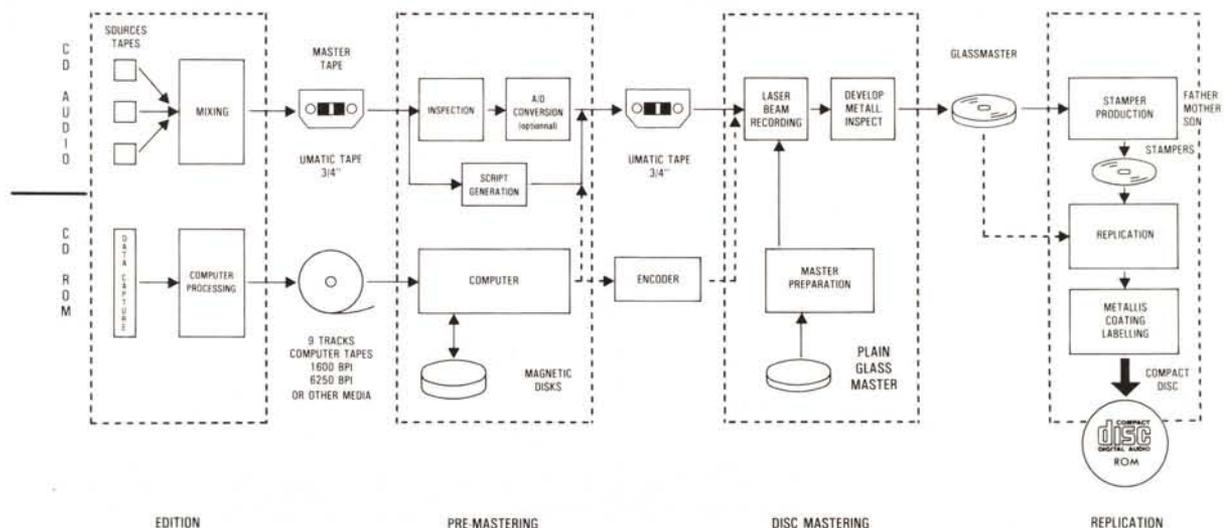
— Indicizzazione: cioè la creazione di una lista delle locazioni ove i dati possono essere reperiti sul CD-ROM. L'indice è infine memorizzato sul CD-ROM assieme ai dati, in modo da consentirne una veloce ricerca.

La selezione di un metodo di indicizzazione (B-tree, Hashing, eccetera) è strettamente correlata al software di retrieval che si intende utilizzare, ed entrambi sono funzione del tipo di applicazione. Un disco contenente dati statistici non ha bisogno di un sistema di retrieval alla Hypercard, così come un CD contenente le pagine gialle ha invece bisogno di un B-tree per le ricerche a chiave parziale.



Floppy, nastri e pagine: tutto deve essere convertito in un formato adatto per il CD-ROM.

COMPACT DISC PRODUCTION FLOW CHART



La quantità di spazio richiesto da un indice non è cosa facile a determinarsi in modo aprioristico, poiché è funzione delle dimensioni del database e del metodo di indicizzazione.

Un indice inverso di ogni parola di un file può facilmente essere più grosso del file stesso. Mentre una tabella Hashing di alcune keyword può essere piccolissima.

— Compressione dati: questo è un termine generico che indica qualunque tipo di operazione atta a riscrivere i dati in maniera che occupino meno spazio. Le tecniche più comuni sono quelle di eliminazione degli spazi vuoti o di zone ripetitive.

La compressione serve ovviamente a salvare spazio su disco, ma non per carenza di capacità sul CD-ROM, ma per velocizzare l'applicativo. Infatti il CD-ROM permette di trasferire 2 kilobyte di dati utente per ciascun settore, alla velocità di 75 blocchi al secondo, quindi il transfer rate è di 150 KB/sec; il problema è che il CD è lento nelle seek, quindi la compressione dei dati riduce la distanza fisica tra le informazioni, consentendo così una maggiore velocità di accesso grazie alla minor distanza fra i dati da recuperare.

— Definizione geografica del CD: ossia la disposizione fisica dei dati nel CD-ROM determina in modo fondamentale la velocità con cui l'applicativo può accedere alle informazioni memorizzate. Vi sono tre metodi base per la mappatura fisica dei dati:

— File sequenziale contiguo, ossia il file inizia ad un blocco e termina senza interruzioni ad un altro, tutto di seguito, a blocchi adiacenti.

— File mappati esattamente come nell'hard disk. Il processo di mappatura consente di avere un file in splittato in differenti locazioni fisiche, senza che la struttura logica ne sia influenzata. Naturalmente servono degli array di puntatori per gestire simili file. Il metodo consente di tenere raggruppati in blocchi contigui quelle parti di file che statisticamente hanno il maggiore accesso, in modo che non sia necessario fare delle seek molto lunghe.

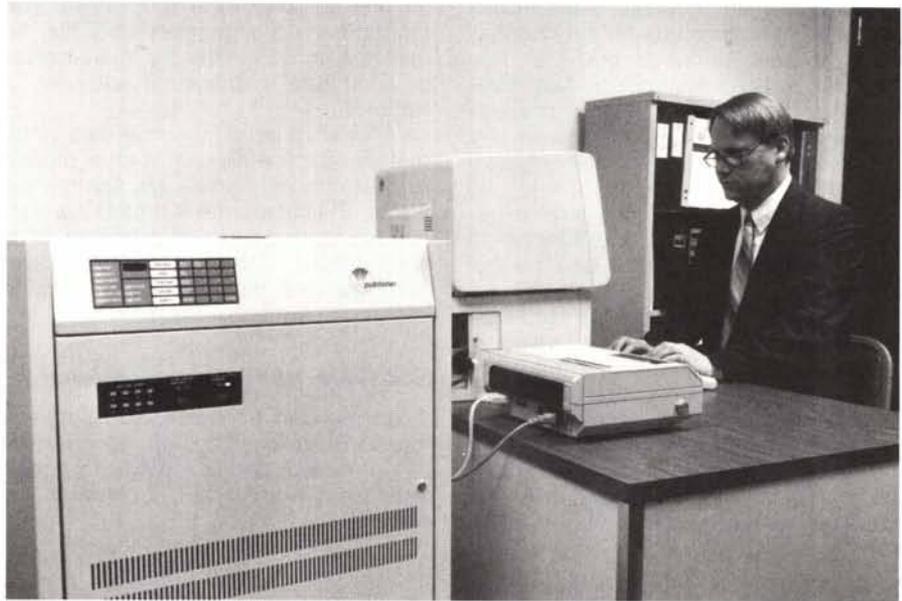
Ad esempio in un CD multimediale nel file di testo vi sarebbe la parola «Chip», poi nel blocco contiguo del CD potrebbe esserci il file immagine del «chip», poi nei blocchi successivi il resto del testo. In questo modo il tempo di accesso sarebbe inferiore a quello richiesto da due file separati ma tutti su blocchi contigui.

— File interallacciati: cioè i file sono suddivisi in blocchi da 2k ciascuno e interallacciati fra loro. Questa tecnica è utilissima per minimizzare i tempi di accesso fra file correlati, come un indice ed il database. La vicinanza dei blocchi annulla il tempo di seek del drive.

Non esiste un metodo ideale per la preparazione fisica del CD, è funzione dell'applicativo, ma certo è che tale metodo deve essere accuratamente

calibrato per ottimizzare gli accessi veloci alle informazioni correlate.

— Scrambling dei dati: qualora la sicurezza dei dati debba essere preservata, la codifica delle informazioni crittografandole, rappresenta un ottimo sistema di protezione. In realtà è necessario avere il software che decodifichi i dati prima di passarli all'applicativo, ma ciò si può rivelare un collo di bottiglia se l'algoritmo di decriptazione è troppo



Il CD-Publisher della Meridian-Data è uno dei sistemi più completi per l'in-house publishing di CD-ROM.



Un esempio di CD-ROM in house publishing è il Laser-ROM della HP, totalmente realizzato e testato col CD-Publisher.

complesso, ma del resto se non lo fosse perderebbe la sua efficacia proiettiva.

Test dell'applicativo

Una volta messa a punto la struttura del CD, è necessario procedere alle verifiche della sua completa funzionalità. Una delle tecniche più in voga negli States è quella di dare ad alcuni clienti selezionati un beta test dell'applicativo per ricevere un feedback sulla facilità d'uso, sull'interfaccia utente, su eventuali errori di acquisizione e così via. Tale metodo è però adatto solo alla fase finale del progetto (in quanto prevede la tiratura di qualche copia) dopo che sono stati applicati uno dei seguenti metodi.

— *High speed testing*: con questo termine ci si riferisce alla metodologia di test che adopera gli hard disk per testare l'applicativo.

C'è però da notare che gli hard disk sono almeno cinque volte più veloci di un lettore CD-ROM, quindi il test risulta falsato. Ovviamente se si deve debuggare un problema nel software di retrieval o verificare l'ortografia dei testi, questo è il metodo più consigliabile, ma nel caso si voglia verificare l'applicativo definitivo, bisogna considerare

che il software di retrieval non lavorerà così veloce...

— *CD-ROM simulation*: un apposito software rende l'hard disk lento come il drive per CD, in modo che lo sviluppatore possa verificare se l'applicazione è accettabile o meno. Tempi di accesso che sembrano buoni con l'hard disk possono diventare insostenibili sul CD-ROM, con la simulazione software del drive CD, chi sviluppa l'applicativo può determinare se sia il caso di pensare ad una compressione dati, ad un riarrangiamento della geografia dei file, se utilizzare la tecnica dei file interallacciati o, addirittura, cambiare il software di retrieval.

— *Master di vetro*: il nome dice tutto, testare l'applicativo sul vero e proprio master che poi servirà alla stampa dei dischi. È sicuramente il metodo di verifica più accurato, ma da applicare all'ultimo stadio di lavorazione (dopo aver fatto già uno dei precedenti test), in quanto è piuttosto costoso.

Software per la preparazione

Sul mercato attuale esistono diversi tool software che consentono lo sviluppo in house dei CD-ROM, fornendo tutto ciò che serve per la preparazione dei dati.

Una stazione di lavoro «personale» per la produzione di un CD-ROM deve prevedere almeno un PC-AT, un drive per nastri a 9 tracce con densità 1600 e 6250 BPI, hard disk con la massima capacità che le proprie tasche consentono: naturalmente più ci si avvicina ai 550 Mega del CD-ROM, più possibilità di lavoro «comodo» vi sono. I più comuni programmi per simili sistemi (Topix, CD-Publisher, CD-Master e simili) sono di produzione statunitense e permettono i classici step necessari alla preparazione dei dati: determinazione della sequenza dei file, in modo da ottimizzare gli accessi, criptazione dei dati se desiderata, compressione, creazione della struttura della directory del CD-ROM, comprendente la creazione degli indici, ed infine il file transfer da disco magnetico al nastro da premaster della directory CD-ROM e dei file nell'ordine definito.

Alcuni programmi hanno molte feature in più, quali la conversione di diversi formati file in un unico compatibile con lo standard High Sierra (ISO 9660), la gestione del disco rigido come immagine virtuale del CD-ROM, così da gestire nello stesso modo i blocchi contigui: i file che sono splittati sull'hard disk lo saranno, nello stesso modo, sul CD-ROM, in maniera che si possano provare differenti combinazioni di interallacciamento file. Naturalmente è importante anche che vi sia l'opzione di simulazione del CD-ROM, in maniera di poter testare l'applicativo sia col metodo High-Speed che con la reale simulazione di risposta di un drive per CD-ROM.

Conclusioni

Appare chiaro che preparare i dati per una trasformazione su CD-ROM non è una cosa molto difficile, infatti oggi già molte aziende decidono di attrezzarsi per l'in-house publishing dei loro dati, affidandosi a service esterni solo per la stampa vera e propria del disco, che continua a diminuire di prezzo. Non vorrei apparire un visionario, ma sono convinto che nel giro di pochi anni appariranno negozi che offriranno la data preparation, così come adesso vi sono i centri di desk top publishing. L'ambiente universitario potrà essere un eccellente mercato per una simile iniziativa, ma presto la necessità di preparare i propri dati per la pubblicazione (su CD, non più su cartaceo) diventerà molto comune.



Registratore a raggio laser per la produzione delle matrici dei CD-ROM.

Lo standard nel modulo continuo

Predisposta per l'utilizzazione dei software in commercio la linea COMPUTER FORM consente, all'azienda come al professionista, di eliminare completamente i tempi morti, guadagnando in professionalità ed efficienza. Anche per quantità limitate è possibile perso-

nalizzare i moduli con il logo dell'azienda. Questo servizio viene realizzato tramite il punto vendita.

Tracciati a modulo continuo in una, due, tre o quattro copie, su carta autoricalcante in vari formati a secondo del tipo di modello.

**DISPONIBILI
MODULI FAC-SIMILE
PER PROVE**



MODULI CONTINUI COMPUTER FORM



FISCALI

- Bolle di accompagnamento serie AB
- Fatture accompagnatorie serie D
- Dichiarazioni IVA
- Dichiarazioni dei redditi 740
- Dichiarazioni dei redditi 750
- Dichiarazioni dei redditi 760
- Dichiarazioni dei sostituti d'imposta 770
- Elenchi clienti
- Elenchi fornitori
- Dichiarazioni d'intento per esportatori
- Registri IVA acquisti e fatture

CONDOMINIO

- Avvisi - ricevute di pagamento
- Convocazione d'assemblea

CORRISPONDENZA

- Lettere a una e più copie
- Computer mailer
- Buste

GESTIONE DEL PERSONALE

- Mod. 01/M-INPS
- Mod. D.M. 10 1-2/89 INPS
- Mod. D.M. 10 3/89 INPS
- Cedolini retribuzioni
- Modd. 101 e 102
- Moduli di versamento IRPEF dipendenti e lavoro autonomo
- c.c.p. IRPEF dipendenti e lavoro autonomo
- Certificazione ritenute d'acconto

CONTABILITÀ

- Libri giornale e inventari
- Fatture differite
- Parcelle per professionisti
- Ricevute - tratte
- Rad. 2
- Prospetti generici per partitari, estratti conto, ecc.

GENERICI

- Moduli in bianco
- Moduli a lettura facilitata
- Tracciati uso bollo per notai
- Schede per schedari UNI A 6
- Schede per schedari rotativi