

1.6. I file

1.6.1. Cosa è un file

I dati e i programmi sono archiviati nelle memorie di massa mediante strutture che si dicono *file*: (termine che, in italiano, significa “*archivio*”), cioè aggregati di dati espressi come sequenze di bit caratterizzate da un insieme di “proprietà”, delle quali le più importanti sono (nel corso dell’unità didattica alcune di queste saranno approfondite):

- Il *nome*: un file ha un nome liberamente attribuitogli dall’utente;
- Il *tipo*: indica essenzialmente il programma che lo ha creato;
- La *dimensione* (in byte o suoi multipli, kbyte o Mbyte), cioè lo spazio occupato in memoria;
- La *data di creazione*;
- La *data dell’ultima modifica*: dopo creato, un file può essere modificato;
- L’*utente che ha creato il file*;
- I *permessi di accesso*, se cioè il file può essere letto o modificato e da chi.

La sequenza di bit che costituisce un file è un aggregato di dati nei cosiddetti *file-dati*; è un codice in linguaggio macchina per i file che archiviano i programmi, detti *file-programmi*. Accedere alle informazioni memorizzate su un disco rigido significa, quindi, accedere ai suoi file, nelle due forme di file-dati e file-programmi.

1.6.2. Il file system e l'albero delle cartelle

Tra i moduli che compongono il sistema operativo vi è il *file system*; l’insieme dei programmi di sistema per l’accesso ai file, la loro allocazione in memoria e, più in generale, per la loro gestione secondo una strategia di archiviazione controllata dall’utente sono tutte funzioni gestite dal *file system*. Attraverso il file system l’utente crea, archivia, modifica, copia e trasporta file da un posto all’altro delle memorie di massa oppure da un computer ad un altro, facendo loro riferimento attraverso il *nome* e tralasciando tutte le complesse dinamiche di funzionamento dell’hardware.

La strategia di archiviazione dei file negli attuali sistemi operativi (Windows, il vecchio MS-DOS, MAC-OS, UNIX e derivati) è la metafora di un classificatore nel quale sono presenti più *cartelle (folder)* ove sono raggruppati e classificati i file.

In generale, in una cartella vi sono file oppure altre cartelle, che a loro volta contengono cartelle e file; in una *organizzazione gerarchica a più livelli* o, come si dice in gergo, *ad albero*; una cartella contiene file (foglie dell’albero) e cartelle (nodi) che a loro volta contengono file e cartelle. Il classificatore principale è in generale la *root* o *radice dell’albero*, diversa per i diversi sistemi operativi.

A ciascuna cartella è associata una *directory* (un elenco dei contenuti) che il file system espone a richiesta, presentando l’elenco di file e cartelle contenuti, ciascuna con le sue proprietà identificative (nome, tipo, dimensione, etc.).

Il disco rigido di un PC, ad esempio, potrebbe contenere una cartella per il sistema operativo in cui trovano posto file-programmi e file-dati di sistema, una cartella per ognuno dei programmi installati sul computer, una cartella per ciascuno dei suoi utenti e queste, a volontà dei singoli utenti, potrebbero essere organizzate in sub-cartelle per una classificazione personale del lavoro.

NOTA: Per una vecchia terminologia indotta da MS-DOS, talora si confonde il concetto di cartella con quello di directory: la cartella è un contenitore (logico) di file e cartelle, la directory ne è l'elenco dei contenuti.

1.6.3. Il percorso (pathname) per raggiungere un file o una cartella

L'insieme dei file di una determinata installazione è dunque organizzato, per i più diffusi sistemi operativi, in un *albero delle cartelle* che, partendo da una radice, si sviluppa per nodi e rami fino a pervenire alle foglie, che sono appunto i file.

Per individuare la collocazione di un file o di una cartella si usa allora indicare il *percorso (pathname)* che lo raggiunge a partire dalla radice e attraversando tutti i nodi (le cartelle) in cui è contenuto; per un file di nome *file* che sia posto in una cartella di nome *B*, a sua volta posta in altra cartella di nome *A*, che è infine posta in *RADICE*, si usano notazioni del tipo *RADICE/A/B* per individuare il percorso.

In alcuni Sistemi operativi, oltre al nome "semplice" del file, si usa anche un *nome completo* che premette al nome semplice il suo percorso: nell'esempio di cui sopra il nome completo sarebbe *RADICE/A/B/file*.

1.6.4. Tipi di file

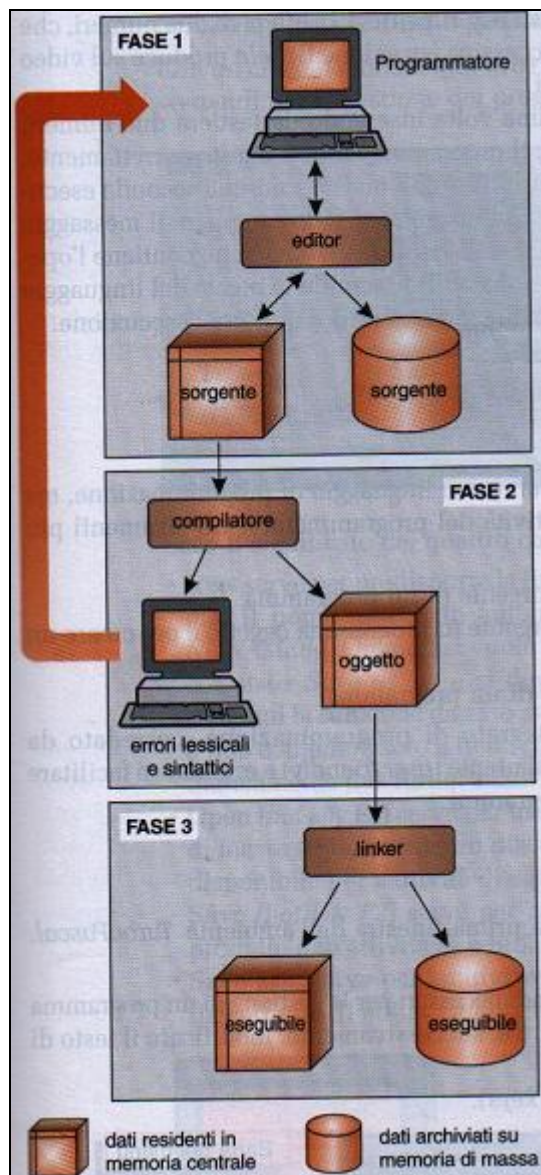
Il *tipo di file* (una delle proprietà del file, vedi RICHIAMO) identifica il programma che lo ha generato; invero esistono classi di programmi "compatibili" che generano lo stesso tipo di file ed alcuni tipi derivano anche da standard internazionali.

I file si possono in effetti suddividere in due categorie, che danno luogo a diversi tipi:

- *File-programmi*, cioè file "eseguibili" (di essi ne può essere lanciata l'esecuzione); si tratta di codici in linguaggio macchina generati da un traduttore di linguaggi (vedi figura, riportata dalla u.f. "Linguaggi di programmazione") il cui tipo è, in generale, *file eseguibile*. Esistono altri file eseguibili, ma qui non approfondiamo);
- *File-dati*, cioè file generati dai programmi stessi. Poiché ogni programma ha le sue regole per generare i dati; un file-dati necessita, per essere esaminato, dello stesso programma e quindi assume un *tipo associato al programma che lo ha generato*, ad esempio file di tipo "txt", per indicare che è un file di puro testo, file "doc" per indicare che è un documento generato dal programma *Microsoft Windows* oppure file di tipo "pdf" per indicare che è un file del programma *Adobe Reader*.

La tecnica per individuare il tipo cui appartiene un file è caratteristica del singolo sistema operativo; per MS-DOS e Windows si usa una "estensione" (vedi in seguito).

Va infine posto in evidenza che alcuni programmi sono in grado di "importare" anche file di tipo diverso dal proprio, previa una operazione di conversione del tipo, ad esempio, un editor di testi (*Microsoft Word*) può importare ed esportare il tipo di un file-dati di altro editor (*WordPerfect*).



Generazione di un file eseguibile

1.6.4.1. Struttura interna dei file: ASCII e binari (approfondimento)

Per quanto riguarda la struttura interna, i file possono essere *ASCII* oppure *binari*. I primi sono sequenze di caratteri codificati in ASCII (e quindi sequenze di bit che a 7 a 7, oppure 8 nel caso di ASCII esteso, rappresentano un carattere); fra i caratteri è compreso anche il carattere di "fine riga" (INVIO o "ritorno carrello"). Si tratta di file, dunque, che rappresentano testi organizzati in righe, come accade per i testi dei programmi in linguaggi ad alto livello non ancora tradotti (i cosiddetti *codici sorgente* dei programmi) o anche di semplici file-dati di editor elementari che non prevedono formattazione, cioè la possibilità di scegliere la grandezza e il tipo dei caratteri, il grassetto, il corsivo, etc, (i cosiddetti *file di testo*).

I *file binari*, invece, contengono informazioni codificate direttamente in bit, senza passare attraverso i caratteri: è ad esempio il caso dei file-programmi eseguibili, per i quali, appunto, le istruzioni sono codificate direttamente in binario, ma anche di file generati da programmi applicativi o di servizio; è ad esempio il caso di file che contengono dati compattati (in gergo, *zippati*) o file che contengono immagini, che sono interpretati solo dagli stessi programmi che li hanno generati o da programmi loro "compatibili".

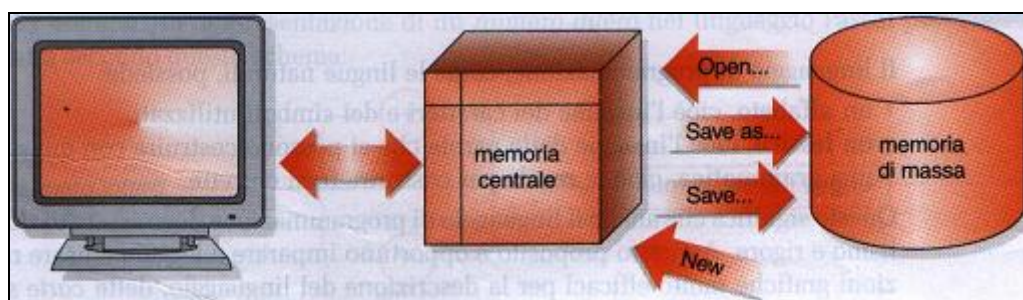
Sono anche diffusi, file ASCII arricchiti da *caratteri di controllo* che specificano il formato del testo o altre informazioni per l'interpretazione dei bit del file.

1.6.5. Il ciclo di vita di un file

Un file-dati “nasce” per esigenze dell'utente, che lo “crea” e lo archivia, poi lo può richiamare per rivederlo o per modificarlo, lo deve spostare da memoria di massa a memoria centrale per lavorarci sopra, lo può di nuovo archiviare o, se non serve più, lo può anche eliminare, e tutto ciò si può ripetere più volte: si può parlare di un “ciclo di vita” di un file. Per esso si definiscono le seguenti operazioni:

- **Creazione di un file (NEW).** Prima di iniziare a lavorare su un file, occorre “crearlo”, iniziando a definirne le proprietà ed, in primis, il tipo; poiché questo dipende dalla applicazione (vedi RICHIAMO), la creazione di un file avviene dall'interno del programma che dovrà usarlo: un file contenente un documento scritto è creato da un editor di testi, un file di disegni da un programma specializzato per fare disegni, un file eseguibile dal traduttore di linguaggio e così via. La creazione avviene provvisoriamente in memoria centrale, come evidenziato nella figura;
- **Primo salvataggio (Save as).** Con il “primo salvataggio (salva con nome) si completa la creazione del file: questo viene spostato dalla memoria centrale (volatile) a quella di massa (permanente) ed assume un nome;
- **Salvataggio (Save).** E' un'operazione analoga alla precedente, ma si effettua su un file che ha già un proprio nome;
- **Apertura (Open).** Quando un programma vuole lavorare su un file (per stamparlo, modificarlo, etc.) deve prima “aprirlo”. Con questa operazione, il file system trova il file nell'archivio e lo ricopia (almeno in parte) nella memoria centrale, dove il programma dovrà lavorare. In effetti, per ogni file aperto il file system costruisce apposite tabelle che gli servono per gestire il file;
- **Chiusura (Close).** Quando si è finito di lavorare con un file è necessario chiuderlo. La chiusura è un'operazione molto importante perché serve a liberare lo spazio di memoria occupato dalle tabelle per la gestione del file.

Anche le operazioni di salvataggio, apertura e chiusura avvengono all'interno del programma che lo ha generato (vedi RICHIAMO). Alcuni sistemi operativi (come Windows) consentono peraltro l'apertura di un file-dati anche direttamente dall'ambiente del sistema operativo.



Il ciclo di vita di un file

RICHIAMO: 1.6.4 Tipi di file

RICHIAMO: 1.1.8 Memoria di massa

RICHIAMO: 1.3.2 Ambiente in cui opera un computer

1.6.6. **Cancellazione logica e ripristino di un file**

Un concetto generale in informatica è quello della *eliminazione* (o cancellazione) *logica*: un oggetto (in senso generale, cioè un file, una lettera di posta, una "cartella", etc.) che non serva va eliminato, ma non definitivamente: si tratta di una *cancellazione logica* e non fisica; in altri termini, l'oggetto è "come se fosse cancellato", ma resta nel sistema, anche se su di esso non si possono effettuare operazioni. In molti casi, ciò comporta che l'oggetto successivamente può essere recuperato: è come un documento gettato nel cestino delle carte, da dove si può successivamente recuperare).

Questa storia, però, non può durare "all'infinito", perché conservare un oggetto richiede memoria e la memoria costa: i sistemi operativi oppure i programmi che gestiscono le cancellazioni logiche prevedono allora anche altre operazioni:

- recupero di uno o tutti gli oggetti cancellati;
- cancellazione definitiva (detta anche cancellazione fisica) di uno o tutti gli oggetti cancellati logicamente; soltanto quando ha subito una cancellazione definitiva, un oggetto non è più recuperabile.

In questo contesto si deve interpretare la cancellazione di un file, operazione tipicamente disponibile in ambiente sistema operativo:

- **Cancellazione (eliminazione) di un file.** Se la cancellazione è logica (com'è in genere), il file viene logicamente considerato inesistente, tranne che per le operazioni di recupero. Se la cancellazione è definitiva, sono eliminati i riferimenti al file all'interno del file system e lo spazio di memorizzazione si rende disponibile.

1.6.7. **Operazioni sui file**

Sui file, oltre alle operazioni generali di cui sopra ed a quelle specifiche dei singoli programmi, sono possibili alcune operazioni in ambiente sistema operativo, di cui si indicano le principali:

- **Copia.** Un file può essere copiato da una memoria di massa ad un'altra; per esempio da un disco rigido ad un floppy per trasportarlo su un altro computer. La copia di un file può avvenire anche all'interno dello stesso disco, da una cartella a un'altra. In ogni caso, questa operazione comporta la creazione di un nuovo file, eventualmente di nome diverso, ma nei contenuti identico a quello di partenza;
- **Spostamento.** Anziché replicare un file, è possibile spostarlo in un'altra posizione. L'operazione di spostamento che avviene all'interno di una stessa memoria di massa non comporta uno spostamento delle informazioni, ma solo una modifica ai riferimenti riguardanti la sua posizione;
- **Rinomina.** Ogni file è contraddistinto da un nome, che, peraltro, può essere modificato.

Le operazioni descritte riguardano sia i file programmi sia i file dati.

1.6.8. **Permessi di accesso ai file**

Un file può essere protetto da accessi che ne potrebbero violare la riservatezza oppure inavvertitamente danneggiare il contenuto, soprattutto in sistemi multi-utente (vedi RICHIAMO), laddove occorra proteggersi da utenti inesperti o in mala fede.

A tale scopo, il file è tipicamente caratterizzato da alcuni parametri: innanzitutto il "proprietario" e poi da alcuni *permesso di accesso* che sono fissati dal proprietario in una operazione che è tipicamente protetta con password. In particolare un file può essere:

- letto (in particolare, un file può essere a *sola lettura*);
- soprascritto o cancellato;

- eseguito (ovviamente, se è un file-programma).

I permessi possono essere assoluti (valgono anche per il proprietario) oppure selettivi e valere quindi per gli utenti non-proprietari oppure per classi di utenti. È il singolo il sistema operativo a fissare regole e procedure per la gestione di tali permessi.

RICHIAMO:1.4.4 Sistemi multi-

1.6.9. Allocazione di file sulle memorie di massa

Ciascuna installazione di PC contiene tipicamente un drive con disco rigido (HD, Hard Disk), un drive di floppy disk (o comunque almeno un disco rimovibile), altri eventuali dischi fissi o rimovibili.

Su un drive di disco fisso è collegato un disco, su uno di disco a supporto rimovibile possono essere inseriti supporti diversi (ad esempio, dischetti diversi in un floppy). I supporti sono in generale detti *volumi* ed hanno un nome simbolico che viene loro assegnato quando vengono "formattati" (vedi RICHIAMO) per il successivo uso.

Un file è tipicamente allocato su un supporto (disco, dischetto, nastro) detto genericamente *volume*, è articolato in record e distribuito su tracce e settori del disco (se il volume è un disco magnetico), secondo una strategia fissata e controllata dal sistema operativo.

RICHIAMO: 1.1.10 I dischi magnetici. La formattazione

1.6.9.1. Frammentazione di un file (approfondimento)

Quando un disco è vuoto, i file vengono memorizzati in settori adiacenti, in modo che i dati rimangano contigui; ma quando si cancella un file, resta uno spazio vuoto.

Per evitare sprechi, in tale spazio vengono memorizzati i dati di un altro file e, poiché è molto improbabile che il nuovo file abbia la stessa dimensione di quello che ha lasciato il vuoto, il nuovo file viene spezzato in più parti. Con l'inevitabile susseguirsi di copie e cancellazioni, aumentano i buchi e i file spezzettati: ci si ritrova con un disco molto frammentato.

Questo vuol dire che gli accessi ai file diventano più lenti, perché se i dati di un file sono memorizzati in settori distanti tra loro, la lettura ha bisogno di un maggior numero di posizionamenti della testina. Opportune utilità eseguono la *deframmentazione* dei dischi ricomponendo i dati dei file in settori contigui.

1.6.10. Copie di sicurezza: il backup

Fare un *backup* significa eseguire copie di sicurezza dei file o, più in generale, dei contenuti della memoria allo scopo di recuperare i dati memorizzati nel caso accidentale in cui accada un danneggiamento degli stessi.

In realtà, i dati sono esposti a vari pericoli: dall'uso incauto del computer da parte di un utente inesperto che li cancella o modifica inavvertitamente, all'ipotesi, non remota, di un guasto hardware che renda illeggibile il disco; quindi, quanto più si lavora su documenti importanti o che richiedono tempo per essere realizzati, tanto più è utile avere a disposizione copie recenti dei propri file.

Il backup si realizza su vari livelli di salvataggio di sicurezza:

- backup di tutto l'HD del computer, usando apposite unità a nastro o a disco rimovibile, dette appunto *unità di backup* (vedi RICHIAMO);

- backup selettivo di alcuni file, ricopiando gli stessi su dischi rimovibili di limitata capacità, come i floppy, le "memory pen" oppure speciali dischi rimovibili;
- backup di un singolo file, realizzandone una copia sul medesimo disco ove risiede il file originale.

RICHIAMO: 1.1.9 Le memorie di massa di un PC