

ARCHIVI CLASSICI

Concetti di base

Per svolgere una qualsiasi attività gestionale, amministrativa, o statistica è necessario utilizzare grandi quantità di dati e scegliere per essi una opportuna organizzazione, in modo da renderne agevole la consultazione e l'elaborazione.

Se esaminiamo un sistema di archiviazione manuale, tutte le informazioni vengono fisicamente registrate su diversi supporti cartacei (registri, quaderni, classificatori,...) e organizzate in liste, elenchi, tabelle in modo che le più comuni operazioni di ricerca, aggiornamento e consultazione possano essere effettuate facilmente.

In un sistema automatizzato ci troviamo di fronte ad una situazione analoga; infatti i dati possono essere registrati sul supporto magnetico utilizzando le diverse strutture di base messe a disposizione dal linguaggio (file sequenziali, file testo, file ad accesso relativo, ...) e organizzati in quelle strutture astratte che meglio si adattano alle esigenze di progettazione e sviluppo dei programmi applicativi. Inoltre, se guardiamo attentamente come sono organizzati i dati in un certo contesto applicativo, ci accorgiamo che essi non sono registrati tutti su un'unica struttura, ma in generale risiedono su svariati supporti, tra loro indipendenti.

Ovviamente le diverse strutture non contengono informazioni tra loro completamente indipendenti ed è quindi necessario creare delle correlazioni che consentano di muoversi agevolmente tra le varie strutture e di prelevare da ciascun elenco solo la parte di informazione voluta.

Definiamo archivio l'insieme di tutte le strutture che contengono le informazioni necessarie alle elaborazioni in un certo contesto.

Un archivio può essere composto da diversi sottoarchivi, ciascuno dei quali è organizzato in modo da rispondere adeguatamente a delle specifiche esigenze, anche molto diverse tra loro.

In un archivio è possibile suddividere i vari sottoarchivi in diverse categorie, in base all'uso:

- sottoarchivi permanenti, contenenti le informazioni primarie di un sistema informativo;
- sottoarchivi temporanei, contenenti i risultati di elaborazioni momentanee o intermedie;
- sottoarchivi storici, contenenti informazioni non più attuali ma a cui può essere necessario fare riferimento occasionalmente;
- sottoarchivi di sicurezza o *di backup*, necessari a garantirsi da accidentali cancellazioni.

In informatica un archivio può essere considerato come un insieme di file, ciascuno dei quali contiene informazioni tra loro omogenee e relative ad una certa classe di oggetti (ad esempio nell'archivio scuola sarà presente il file STUDENTI contenente i dati anagrafici di tutti gli alunni).

Un file è quindi un insieme di registrazioni o record, ciascuno dei quali descrive le proprietà specifiche del singolo oggetto (nel nostro esempio un record conterrà tutti i dati anagrafici di un certo studente) ed è individuato attraverso un indirizzo logico che corrisponde alla posizione che la registrazione occupa nel file

Ciascuna registrazione a sua volta è costituita da un insieme prefissato di informazioni elementari, dette campi, ciascuna delle quali individua un preciso attributo dell'oggetto (ad esempio nel file STUDENTI il record conterrà, tra gli altri, i campi destinati a contenere nome, cognome, data e luogo di nascita dello studente).

Ogni campo sarà caratterizzato da un nome, che permette di fare riferimento ad una proprietà comune agli oggetti in questione (campo INDIRIZZO) e da un valore che specifica la proprietà del singolo oggetto (Via P. Rossi 5, MILANO).

Tipicamente in una registrazione l'insieme dei campi si divide in due parti:

- una, detta chiave, necessaria all'individuazione della registrazione nell'intero archivio;
- l'altra contenente tutte le informazioni supplementari relative al singolo oggetto.

In particolare se ogni valore della chiave individua al più una registrazione, la chiave viene detta chiave primaria.

Nel caso invece in cui ad un valore di chiave corrispondano più registrazioni si parla di chiave secondaria.

Operazioni sugli archivi

Abbiamo visto come un archivio possa essere considerato come una particolare struttura dati astratta sulla quale, ovviamente, è possibile definire un insieme di operazioni.

Anche se non esiste una precisa standardizzazione sul tipo e sul numero di tali operazioni, è importante avere almeno una visione generale di quelle che più frequentemente compaiono nei programmi applicativi, poiché spesso rappresentano un vincolo nella progettazione, definizione e realizzazione dell'organizzazione da dare alla struttura informativa.

- **Creazione dell'archivio vuoto**

Con questa operazione, sulla base dell'organizzazione scelta in fase di progettazione, vengono fisicamente generati tutti i sottoarchivi previsti dai programmi applicativi, ciascuno con la propria struttura e il proprio tracciato record.

- **Inserimento di una registrazione in un archivio già esistente**

Questa operazione, che permette di aggiungere nuovi dati in un archivio, può assumere una forma articolata e complessa, poiché ciò può voler dire inserire dati in più sottoarchivi, generando anche le dovute correlazioni.

Naturalmente, l'operazione di inserimento risulterà ancora più complessa nel caso in cui i singoli sottoarchivi posseggano una qualche forma di ordinamento.

- **Cancellazione di una registrazione di chiave K**

Si tratta dell'operazione inversa a quella precedente e spesso ne possiede tutte le caratteristiche.

- **Aggiornamento di una registrazione di chiave K**

Individuata la registrazione di chiave K, il problema consiste nel modificare uno o più campi non chiave.

Anche per questa operazione valgono considerazioni analoghe a quelle fatte in precedenza.

- **Ricerca di una registrazione di chiave K**

Il problema consiste nell'individuare nella struttura informativa tutti i dati relativi a una o più registrazioni associate al valore K della chiave.

Come abbiamo visto negli esempi, la ricerca può comportare la scansione di più sottoarchivi, ciascuno dei quali concorrerà a fornire l'intera informazione.

Questa operazione assume un'importanza fondamentale nella gestione dell'archivio, poiché, oltre ad essere una delle operazioni più richieste dall'utente (interrogazione dell'archivio), rappresenta un passo obbligato per compiere altre operazioni, quali ad esempio quelle di cancellazione e modifica.

- **Visita di un archivio**

Con questa operazione si esaminano tutte le registrazioni di un archivio per compiere su ciascuna di esse una certa elaborazione.

A livello elementare una visita significa, a partire da una "prima" registrazione, l'individuare la registrazione "successiva" a quella appena esaminataoli sottoarchivi posseggano una qualche forma di ordinamento.

Ciò implica che una visita può essere compiuta sia in un ordine qualsiasi (basta esaminare tutte le registrazioni una e una sola volta) che secondo una certa logica (secondo un ordine di chiave).

- **Ordinamento di un archivio**

Spesso per ottimizzare le operazioni di ricerca o di visita, si rende necessario procedere all'ordinamento delle registrazioni che compongono un archivio.

A questo proposito vale la pena di sottolineare la differenza esistente tra l'ordine fisico di registrazione e l'ordine logico con cui si dovrebbero susseguire i vari record.

Si definisce ordine fisico la sequenza in cui sono disposte le singole registrazioni sul supporto di memorizzazione, mentre si parla di ordine logico facendo riferimento alla sequenza dei valori crescenti o decrescenti della chiave.

Quando in un sottoarchivio l'ordine fisico coincide con quello logico, si parlerà di sottoarchivio ordinato.

Dato che, specie per archivi molto grandi e complessi, l'ordinamento dei dati può richiedere

tempi anche molto lunghi, risulta particolarmente importante scegliere di volta in volta la tecnica più idonea.

- **Copia di un archivio**

Questa operazione in generale viene utilizzata per costruire le copie di sicurezza delle parti fondamentali di un archivio.

- **Fusione e separazione di archivi**

A volte si rende necessario generare, a partire da più sottoarchivi tra loro omogenei, un unico archivio che risulta la combinazione dei precedenti.

Questa operazione risulta particolarmente significativa nel caso di sottoarchivi, ordinati, in quanto occorre mantenere l'ordine delle chiavi.

L'operazione inversa, che permette di suddividere i dati di un archivio in più sottoarchivi tra loro fisicamente indipendenti, viene ad esempio effettuata in fase di ristrutturazione dell'archivio o quando le sue dimensioni superano la capacità massima del supporto magnetico a disposizione.

Tipi di Organizzazione

Tra le operazioni fondamentali che si eseguono su un archivio, abbiamo visto che riveste particolare importanza la ricerca di una registrazione nota la chiave.

E' ovvio che un'operazione di questo tipo può essere sempre compiuta o ricavando dalla chiave le informazioni che permettono di calcolare l'indirizzo del blocco di memoria che contiene la registrazione voluta o scandendo, completamente o in parte, l'archivio finché non si raggiunge una registrazione avente il valore della chiave uguale a quella cercata.

In entrambi i casi, ogni risposta ottimizzata a questo problema dipende in maniera sostanziale dal modo in cui sono stati memorizzati i dati nell'archivio e da come questi sono stati fra loro correlati. In particolare, l'esigenza di evitare o calcoli complessi dell'indirizzo o scansioni troppo lunghe, ha favorito lo sviluppo di svariate tecniche per l'organizzazione degli archivi al fine di ottimizzarne l'uso.

Per fornire una visione complessiva di queste problematiche in questo paragrafo daremo una breve descrizione delle principali strutture di archivio, che verranno poi analizzate più approfonditamente nei prossimi capitoli.

- **Gli archivi sequenziali**

In questo tipo di archivio le registrazioni sono memorizzate in blocchi consecutivi senza particolari accorgimenti per permetterne il ritrovamento.

Un archivio sequenziale si dice ordinato quando l'ordine di successione dei record rispetta l'ordine ascendente o discendente dei valori della chiave, mentre negli altri casi si parla di archivi sequenziali disordinati.

Tra questi ultimi meritano un cenno a parte gli archivi seriali, in cui non ha particolare significato parlare di chiave d'accesso, in quanto i record non sono organizzati in alcun modo.

In questo caso le registrazioni possono essere elaborate logicamente solo in ordine seriale, cioè nell'ordine in cui sono state memorizzate sul dispositivo, come accade ad esempio nei file testo e nei file di backup.

L'organizzazione sequenziale si presta in particolare modo a quelle applicazioni che richiedono frequentemente lo stesso tipo di elaborazione su buona parte delle registrazioni.

- **Gli archivi sequenziali con indice**

Questa struttura costituisce una variante di quella sequenziale in cui all'archivio principale vengono associate una o più tabelle, dette tabelle indice, attraverso le quali è possibile velocizzare le operazioni di ricerca.

Il programma di elaborazione, potendo disporre di un indice, lo può consultare e saltare di conseguenza tutte le registrazioni che non devono essere elaborate, accedendo direttamente al blocco interessato.

- **Gli archivi ad accesso diretto**

Questo tipo di organizzazione si basa sul principio di ricavare direttamente dal valore della chiave il massimo numero di informazioni sull'indirizzo del record ad essa associato.

L'organizzazione dell'archivio in questo caso è chiamata ad accesso diretto o casuale (random) per indicare che è possibile accedere ad una registrazione direttamente usando la chiave, senza considerare la posizione che il record occupa all'interno della struttura.

In particolare la parola casuale sta ad indicare che i record generalmente non sono memorizzati in sequenza ma disposti casualmente in singole posizioni di memoria, individuabili con una funzione di trasformazione della chiave.

Gli archivi casuali vengono utilizzati in particolare modo nei casi in cui le informazioni non vengono elaborate sequenzialmente, ma si intendono privilegiare le operazioni di ricerca e interrogazione.

- **Archivi a liste**

In questi archivi l'indirizzo di una registrazione è contenuto in un campo della registrazione che logicamente la precede, come accade per le liste di record disposte in memoria centrale. Anche in questo caso la memorizzazione fisica può avvenire in un ordine totalmente diverso da quello con cui devono essere scanditi i record in ordine logico; ciò significa che due elementi logicamente contigui possono occupare sul supporto magnetico posizioni anche

distanti tra loro.

Si costituisce così all'interno dell'archivio una struttura di lista definita come una successione di elementi che occupano in memoria posizioni qualsiasi, ma in cui ogni elemento è legato al successivo tramite un puntatore.

Questa struttura può rilevarsi particolarmente interessante nei casi in cui le operazioni di inserimento o estrazione sono piuttosto frequenti oppure quando occorre riorganizzare l'archivio secondo diversi ordini.

Negli archivi a liste, infatti, vengono privilegiate le operazioni di aggiornamento ed ordinamento a scapito di quelle di ricerca.

- **Gli archivi a liste invertite**

Questa struttura estende il concetto di indice a tutti i campi del record che vengono considerati tutti o in parte come chiavi secondarie di ricerca.

In pratica all'archivio principale vengono associati tanti indici quante sono le chiavi secondarie, costruendo così delle tabelle in cui ad ogni valore di una chiave secondaria viene affiancata la lista dei riferimenti ai record che possiedono detto attributo.

Ciò che si ottiene è un archivio che viene detto archivio parzialmente invertito.

Invece, nel caso in cui si considerino chiavi secondarie tutti i campi del record, l'archivio viene detto archivio totalmente invertito.

Questi archivi, organizzati per chiavi secondarie, sono utili quando non è possibile utilizzare una singola chiave per rintracciare una registrazione ma una combinazione di chiavi.

- **Altri tipi di organizzazione**

Partendo dalle configurazioni base appena viste, è possibile costruire nuovi tipi di organizzazioni più complesse che cercano di privilegiare uno o più aspetti particolari, adattandoli alle specifiche esigenze dell'utente.

Sarà così possibile costruire, ad esempio, archivi multilista, archivi ad albero, archivi reticolari e così via.

Metodi di elaborazione

Un elemento di fondamentale importanza da valutare per decidere l'organizzazione più idonea con cui strutturare un archivio è il metodo di elaborazione più frequentemente usato in fase applicativa, ove con metodo di elaborazione si intende la modalità con cui l'archivio viene interrogato, aggiornato e visitato.

I metodi di elaborazione fondamentali sono:

- **seriale;**
- **sequenziale;**
- **sequenziale selettivo;**
- **casuale.**

Metodo di elaborazione seriale

Consiste nell'accedere ai record rispettando lo stesso ordine usato per la registrazione, indipendentemente dal fatto che l'archivio sia ordinato o no.

Questo metodo risulta particolarmente efficace nei casi in cui l'ordine logico di registrazione non riveste alcuna importanza.

Un esempio può essere trovato nella costruzione di copie di sicurezza o nella stampa del contenuto dell'intero archivio.

Metodo di elaborazione sequenziale

Consiste nell'accedere alle registrazioni seguendo l'ordine logico delle chiavi.

Se l'archivio è ordinato basterà seguire l'ordine fisico di registrazione, altrimenti sarà necessario ricercare ogni volta il record successivo a quello appena esaminato.

Questo metodo si adatta a tutti quei casi in cui lo stesso tipo di elaborazione si deve applicare a gran parte dei record, considerati però in un ordine ben preciso.

Metodo di elaborazione sequenziale selettivo

Deriva direttamente dal metodo precedente e consiste nell'elaborazione di un numero limitato di registrazioni, saltando quelle che non interessano ma procedendo sempre secondo l'ordine logico.

Metodo di elaborazione casuale

L'accesso ai record non dipende da alcun ordine e le operazioni di aggiornamento o interrogazione possono essere effettuate di volta in volta su uno qualsiasi dei record.

Questo metodo di elaborazione, ove è possibile applicarlo, consente all'utente di avere risposte molto rapide, ma per contro rende piuttosto complesse le operazioni sequenziali su tutto l'archivio.

Scelta dell'organizzazione

Nel decidere le modalità secondo cui strutturare un determinato archivio, il progettista deve prestare attenzione a tutta una serie di fattori, spesso antagonisti, quali il rispetto delle esigenze dell'utente, l'integrazione reale delle procedure operative nell'ambito dell'ambiente di lavoro, i vincoli spesso imposti dalle risorse fisiche a disposizione e, non ultimi, i costi di realizzazione e di gestione.

In ultima analisi quindi organizzare un archivio significa decidere, nel rispetto dei vincoli imposti dall'utenza, sia la struttura che si intende dare ai dati su memoria di massa sia l'insieme delle procedure di gestione e dei protocolli di accesso alle informazioni.

Tasso di elaborazione (hit rate)

Con questo termine, riferito ad una particolare esecuzione, si intende il rapporto percentuale tra il numero di record a cui si accede e il numero totale di registrazioni presenti nell'archivio.

La conoscenza del valore di questo parametro risulta particolarmente importante per la progettazione dell'archivio a patto di avere informazioni precise sul numero delle elaborazioni complessive a cui l'archivio verrà sottoposto in un certo periodo di tempo.

In ogni caso appare evidente che un alto tasso di elaborazione tende a favorire le organizzazioni sequenziali o sequenziali con indice.

Modalità di lavoro

Un altro fattore che può condizionare le scelte del progettista è costituito dal modo in cui avviene l'interazione tra l'utente e il sistema.

Diverso è infatti costruire un archivio per un utente che chiede solo elaborazioni a lotti, rispetto ad uno che richiede modalità di tipo interattivo.

Nel primo caso, infatti, saranno da privilegiare organizzazioni di tipo sequenziale, mentre nel secondo ci si orienterà verso metodi con tempi rapidi di accesso.

Crescita e volatilità dell'archivio

In un archivio le aggiunte e cancellazioni possono provocare delle variazioni apparenti nelle dimensioni fisiche dell'archivio stesso, in quanto molto spesso i record cancellati logicamente permangono fisicamente nell'archivio ma non sempre possono essere recuperati immediatamente. Per questo motivo, quando le aggiunte e cancellazioni sono molto frequenti, conviene privilegiare delle strutture che consentono un recupero dinamico dello spazio inutilizzato.

Inoltre, nella scelta dell'organizzazione occorre tener presente anche la possibilità di crescita reale dell'archivio, in quanto una crescita sproporzionata potrà, in alcune organizzazioni, dar luogo ad un aumento inaccettabile dei tempi di ricerca o alla necessità di ristrutturare addirittura l'intero archivio.

Nella progettazione di strutture di questo tipo, quindi, occorrerà orientarsi verso organizzazioni piuttosto flessibili che consentano delle facili e rapide ristrutturazioni.

Sicurezza

La necessità di garantire la sicurezza di un archivio (integrità dei dati registrati) può far pendere la bilancia per un tipo di scelta piuttosto che per un altro.

Infatti, maggiore è la complessità della struttura di un archivio, tanto più esso è vulnerabile (si pensi alla perdita di un puntatore in una struttura a lista) e il procedimento di ricostruzione più difficile.

Da tutto ciò appare evidente, quindi, che per gli archivi molto complessi occorre prevedere o delle duplicazioni dei dati fondamentali o delle tecniche di backup opportune.

Tipologia dell'hardware a disposizione

Nel caso in cui si sia vincolati da un determinato tipo di risorse fisiche a disposizione, occorre strutturare l'archivio tenendo conto delle eventuali limitazioni che il sistema di elaborazione (CPU, Sistema Operativo, linguaggio, ...) e i supporti di memorizzazione possono imporre.

Costi

Il progetto di una certa organizzazione per un archivio comporta inevitabilmente una serie di costi molto spesso niente affatto trascurabili e quindi occorre tener conto anche dei vincoli finanziari imposti dalla situazione.

In particolare occorrerà trovare un giusto equilibrio tra la qualità delle prestazioni offerte dal sistema di archiviazione e i costi di progettazione e implementazione.

Tra i fattori di costo che possono influenzare la scelta, ricordiamo:

- costi per l'acquisto dell'hardware necessario;
- costi della progettazione e dello sviluppo del software applicativo;
- costi per l'installazione e costituzione dell'archivio;
- costi di gestione e manutenzione.

Anche la suddivisione dei costi tra le varie voci rappresenta un problema non trascurabile per il progettista, poiché a volte conviene sostituire il sistema di elaborazione a disposizione piuttosto che sviluppare il software necessario con strumenti inadeguati (S.O. e linguaggi).