

DISPENSA
SUGLI
ARCHIVI

A cura del
Prof. Enea Ferri

Introduzione

L'argomento che verrà illustrato si riferisce ad una struttura che in qualche modo fa riferimento alle tabelle di record.

Ricordiamo che un record:

- E' un insieme di informazioni, non necessariamente dello stesso tipo, che si riferiscono ad un elemento di un certo insieme;
- ogni singola informazione costituisce un **CAMPO** del record;
- ad ogni campo viene assegnato un **NOME**, che attribuisce un significato all'informazione (es. **COGNOME**, **NOME**...);
- la successione dei nomi dei campi costituisce il FORMATO del record.

Una tabella di record è un Insieme finito di record aventi lo stesso formato. E' quindi fondamentale definire come per gli array il numero degli elementi.

Un record in una tabella può essere individuato:

- 1) per chiave
- 2) per posizione

la chiave è quella parte del record che serve per distinguerlo da tutti gli altri. Può essere costituita da uno o più campi del record.

Individuare un record per chiave significa ricercare all'interno della tabella il record avente la chiave specificata;

Individuare un record per posizione implica la conoscenza a priori della posizione del record all'interno della tabella

Terminata l'esecuzione del programma che fa uso della tabella, vengono perse tutte le informazioni che conteneva: questo accade perché la tabella di record viene costruita e gestita in memoria centrale.

1 Generalità sugli archivi

1.1 Concetti di archivio, attributo, chiave

Un archivio è un insieme di informazioni tali che:

- tra esse esista un nesso logico (ovvero si riferiscano ad uno stesso argomento);
- siano rappresentate secondo un formato che ne renda possibile l'interpretazione (ad esempio elenco telefonico);
- siano registrate su un supporto su cui è possibile scrivere e rileggere informazioni anche a distanza di tempo;
- la loro organizzazione ne permetta una facile consultazione (ad esempio elenco telefonico ordinato per città e per ogni città in ordine alfabetico).

Quindi un archivio è un insieme di registrazioni (o records) ciascuna delle quali è costituita da un insieme di informazioni dette attributi (o campi). Tali informazioni descrivono le proprietà caratteristiche di un oggetto o ENTITÀ della realtà che si sta studiando e consentono di distinguere i vari esemplari delle entità in esame. Ogni campo è caratterizzato da un nome, che permette di fare riferimento ad una proprietà comune a tutte le entità (es. campo COGNOME) e da un valore che specifica la proprietà del singolo oggetto (es. ROSSI).

In figura 1 è riportato un semplice esempio di archivio relativo ad un insieme di persone.

<i>Cognome</i>	<i>Nome</i>	<i>Data Di Nascita</i>	<i>Professione</i>	<i>Sesso</i>
Annese	Gemma	5/12/64	Impiegato	"femmina"
Bianchi	Anna	25/6/86	Studente	"femmina"
Rossi	Carlo	1/11/61	Avvocato	"maschio"
Rossi	Giuseppe	15/3/63	Insegnante	"maschio"
Vergaro	Alfredo	3/12/85	Studente	"maschio"

Figura 1 Archivio *PERSONE*.

Per creare un archivio occorre specificare:

- Nome dell'archivio, di solito con riferimento al suo contenuto (es. FORNITORI, CLIENTI, MAGAZZINO...);
- Tracciato record, cioè quali informazioni compongono il record;
- Supporto da usare per archiviare i dati (es. fogli di carta, nastri magnetici, dischi magnetici...);
- Eventuale dimensione massima dell'archivio

Uno degli attributi (o un gruppo di attributi) di un record viene detto **CHIAVE**, e serve per individuare un record all'interno dell'Archivio.

Una **CHIAVE PRIMARIA** è una chiave che individua in modo univoco una registrazione fra quelle contenute nell'archivio: il suo valore può essere posseduto da un solo record. La chiave può essere composta da uno o più campi, in base al tipo di applicazione e di uso che se ne vuole fare.

Nel caso dell'archivio *PERSONE*, una chiave primaria potrebbe essere: {*Cognome, Nome, DataDiNascita*}. Una chiave primaria esiste sempre: essa è al limite costituita da tutti gli attributi. Infatti:

- deve individuare in modo univoco il record, cioè non devono esserci due record nell'archivio con uguale chiave (UNIVOCITA')
- deve essere tale che, tralasciato un campo, non deve più esserci l'univocità (NON RIDONDANZA)
- si deve scegliere come chiave la combinazione con minore numero di campi.

I requisiti di una chiave sono:

- la **precisione**, per eliminare ambiguità nell'individuazione di un record;
- la **compattezza**, per risparmiare spazio, ma soprattutto tempo di elaborazione;
- la **chiarezza**, per contenere la maggior parte di informazioni, nel minore spazio possibile.

Si può scegliere come chiave un CODICE FORMALIZZATO, ovvero che non sia un campo informativo o una combinazione di essi. In questo caso si parla di:

- a) Codice progressivo numerico : 01, 02, 03, ...
- b) Codice progressivo alfanumerico : A01, A02, ...A10, B01,

Una chiave non primaria è detta **chiave secondaria**: il suo valore individua, in generale, più di una registrazione. Ad esempio, con riferimento all'archivio *PERSONE*, si può osservare che {*Nome*} e {*Professione, Sesso*} sono due possibili chiavi secondarie.

Una chiave secondaria *selettiva* è una chiave cui è associato un numero relativamente basso di registrazioni (ad esempio {*Nome*} nell'archivio *PERSONE* è una chiave secondaria selettiva).

Possono esistere più chiavi primarie (disgiunte o meno), ma nel seguito faremo spesso l'ipotesi che esista un particolare attributo con funzione di *chiave primaria*.

1.2 Operazioni ed interrogazioni su archivi

Su un archivio si possono compiere le seguenti operazioni:

- a) **Creazione** di un archivio vuoto.
- b) **Inserimento** di un record con chiave *K*.
- c) **Reperimento** di un record di chiave *K*.
- d) **Aggiornamento** (di campi non chiave) di un record di chiave *K*.
- e) **Cancellazione** di un record di chiave *K*.
- f) **Visita** a tutti i records dell'archivio.

L'operazione di Creazione consiste nel decidere la struttura, il tipo di dati che dovrà contenere e le relazioni tra essi.

Le operazioni di Inserimento, Aggiornamento e Cancellazione vengono generalmente classificate come operazioni di modifica dell'archivio e richiedono di fatto una ricerca preventiva. Ad esempio, l'operazione di inserimento è preceduta da una ricerca: il record viene inserito solo se non esiste già il record con la stessa chiave.

Il reperimento di un record consente di sapere se esiste un certo record, ricerca da effettuare sul campo chiave primaria e, in caso affermativo, di conoscerne tutti gli altri attributi.

La modifica di un campo chiave può essere effettuata solo cancellando prima il vecchio record e inserendo poi quello nuovo.

L'operazione di visita è essenziale in tutti quei casi in cui occorre applicare una certa elaborazione a tutti o parte dei records dell'archivio (si pensi ad una stampa o ad un aggiornamento). A livello elementare, l'operazione di visita viene effettuata in base a due operazioni: una di accesso al «primo» record dell'archivio e una di accesso al record «successivo» all'ultimo cui si è fatto accesso.

Si distinguono due tipi di ordini di visita: la *visita in ordine qualsiasi* e la *visita ordinata* per valori crescenti (o decrescenti) di un attributo, che deve essere di un tipo per il quale sia definita una relazione di ordinamento totale (ad esempio *integer*, *string* ecc.). Più in generale la visita ordinata può essere definita per una chiave qualsiasi, cioè formata da più di un attributo. In questi casi la relazione di ordinamento è definita dall'ordinamento sugli attributi che compongono la chiave.

Ad esempio considerati due records, per il primo sono letti gli attributi (a_1, a_2) , per il secondo gli attributi (b_1, b_2) ; il primo record precede il secondo se $(a_1, a_2) \leq (b_1, b_2)$ ovvero se:

$a_1 < b_1$ oppure $a_1 = b_1$ e $a_2 \leq b_2$.

La visita ordinata comporta un'operazione di ordinamento, a meno che l'archivio non sia stato organizzato in modo da poter essere visitato in modo ordinato, il che significa che può essere attraversato secondo l'ordinamento della chiave primaria.

Parlando genericamente di archivi ordinati, si intende che è possibile eseguire la visita ordinata secondo valori crescenti (o decrescenti) della chiave senza ricorrere ad una procedura di ordinamento.

L'operazione di reperimento di un record con chiave K è solo un caso particolare di interrogazione (*query*) dell'archivio. Una sua generalizzazione presa frequentemente in considerazione è la seguente:

g) *Ricerca* tutti i records con chiave compresa fra K_1 e K_2 .

Questa operazione è detta **interrogazione per campo di variabilità** (range query). Si osservi che se l'archivio non è ordinato è necessario compiere la visita completa.

Ad esempio, con riferimento ad un ipotetico archivio *STUDENTI*, una possibile interrogazione per campo di variabilità su chiave secondaria potrebbe essere:

«trovare tutti gli studenti di età compresa fra i 15 e i 18 anni»

Nella loro forma più generale le interrogazioni possono assumere il seguente aspetto:

h) *Ricerca* tutti i records i cui attributi soddisfano un predicato p .

Il predicato p può essere *semplice* o *composto*. Un predicato semplice è del tipo:

$\langle \text{nome attributo} \rangle (\text{oprel}) \langle \text{valore attributo} \rangle$

dove (**oprel**) è uno dei sei operatori relazionali = , ≠ , ≤ , ≥ , < , > . Esempi:

Residenza= 'ROMA' , *Altezza*> 1.70, ecc.

Un predicato composto si ottiene da quelli semplici o composti mediante applicazione degli operatori booleani OR, AND, NOT.

Esempio:

$(\text{Residenza} = \text{'ROMA'}) \text{ OR } ((\text{Residenza} \neq \text{'ROMA'}) \text{ AND } (\text{Anni} \geq 15) \text{ AND } (\text{Anni} \leq 20))$

Si richiedono cioè tutti gli studenti residenti a Roma o, se residenti fuori Roma, quelli di età compresa fra i 15 e i 20 anni.

i) Ordinamento dei record secondo una determinata regola.

Questa operazione risponde all'esigenza di un'organizzazione degli archivi tale da consentire una facile consultazione.

l) Fusione (MERGE) tra due o più archivi.

Consente la formazione di un unico archivio utilizzando i dati contenuti in due o più archivi.

1.3 Fattori che influenzano la scelta dell'organizzazione di un Archivio

Un termine usato frequentemente per indicare l'implementazione di un *archivio* è quello di **organizzazione**. Con ciò si intende sia il *tipo di rappresentazione* (struttura dati) su memoria di massa, detta anche Organizzazione FISICA, sia le *procedure di base* per l'accesso ad un

record, detta anche Organizzazione LOGICA, che definisce le modalità di gestione.

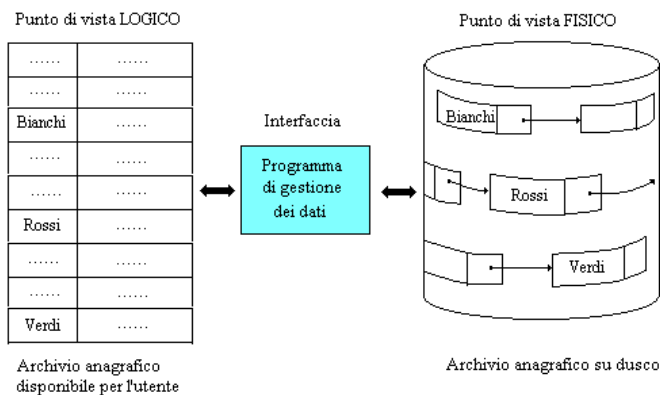
I fattori più rilevanti che possono influenzare la scelta di un'organizzazione piuttosto di un'altra sono:

1) *Tipi di operazioni previste e loro frequenza d'uso.*

Non tutte le operazioni vengono richieste con la stessa frequenza. Si parla di archivio **dinamico** se è

soggetto a frequenti inserimenti/cancellazioni e *archivio statico* in caso contrario.

In alcune organizzazioni le operazioni di inserimento e di cancellazione possono degradare l'efficienza della ricerca, al punto da rendere necessaria una periodica riorganizzazione dell'archivio.



Organizzazione Logica e organizzazione fisica.

Se esiste l'esigenza di reperire frequentemente i records in base ad un particolare attributo può essere conveniente tenerne conto nell'organizzazione dell'archivio gestendo **strutture di accesso per chiave secondaria**.

Se sono previste interrogazioni per campo di variabilità di una chiave può essere conveniente ricorrere ad un'organizzazione ordinata rispetto a quella chiave.

2) *Tempi e metodi di elaborazione.*

Se tutte le operazioni avvengono in modo interattivo, il tempo medio di risposta deve essere molto breve (dell'ordine di pochi secondi). A seconda delle applicazioni potranno essere imposti anche dei vincoli sul tempo di risposta e ciò potrà influire sulla scelta dell'organizzazione. Questa sarà generalmente diversa da quella che conviene adottare nei casi in cui non esistono interrogazioni interattive e, più in generale, quando esiste un'ampia tolleranza sul tempo di risposta. In quest'ultimo caso le richieste di modifica o di ricerca possono essere raccolte in un apposito file e applicate periodicamente tutte in una volta (*elaborazione non-interattiva* od *off-line*). L'elaborazione off-line è conveniente quando è elevato il **rapporto di attività**, definito come il *rapporto fra numero di records da elaborare e numero di records presenti nell'archivio* (con riferimento ad una certa unità di tempo).

Il **tempo di accesso in memoria secondaria** è importante per valutare il tipo di organizzazione. Salvo rare eccezioni si possono infatti ritenere trascurabili i tempi di elaborazione in memoria centrale.

3) *Tipo e dimensione del supporto di memoria secondaria.*

Se il supporto di memoria disponibile è ad accesso sequenziale (nastri magnetici) vengono necessariamente escluse tutte le organizzazioni basate **sull'accesso diretto**. Se la dimensione del supporto è critica può non essere possibile adottare alcune organizzazioni veloci che richiedono dello spazio aggiuntivo rispetto a quello strettamente necessario per memorizzare i records dell'archivio.

4) *Integrità e ripristino dell'archivio.*

Occorre valutare le conseguenze ed i costi dovuti alla distruzione di una parte delle informazioni dell'archivio in seguito a cause accidentali o per un errore di programmazione. Se l'archivio fa uso di puntatori, l'alterazione di un puntatore può rendere inaccessibile una parte consistente di informazioni.

Nell'organizzare l'archivio bisognerà consentire la *ricostruzione dell'archivio a partire dall'ultima copia salvata* su nastro o su un qualsiasi supporto rimovibile. La tecnica generalmente adottata consiste nel memorizzare su un apposito file (sequenziale) la *storia delle modifiche* avvenute dopo l'ultimo salvataggio. La ricostruzione può così avvenire applicando successivamente tali modifiche all'ultima copia salvata.

5) *Vincoli imposti da riferimenti.*

Può accadere che i records dell'archivio siano «puntati», cioè che il loro indirizzo sia utilizzato da qualche struttura dati presente nello stesso ambiente. Ciò impone di fatto la necessità che il record non possa essere rimosso dalla sua posizione almeno fino a quando non

c'è la certezza che non sia più puntato. Questo vincolo introduce qualche problema in più rispetto al caso in cui i records non sono puntati ed influenza il tipo di organizzazione che è possibile adottare. Una soluzione a questo problema consiste nel sostituire i puntatori fisici con dei puntatori simbolici costituiti dalle chiavi primarie dei records. Ciò richiede degli accessi supplementari ma si è liberi di scegliere l'organizzazione che si ritiene più opportuna.

2 Organizzazione degli Archivi

L'utente o il programmatore opera sugli archivi identificandoli con un nome (**nome logico**, interno al programma), che è associato al nome con il quale viene ritrovato all'interno del supporto magnetico (**nome fisico**). Quando si fa riferimento ad un archivio su memoria di massa, di usa spesso il termine FILE.

La registrazione dei dati su un supporto di memoria di massa deve essere fatta in modo da organizzarli nel migliore modo possibile, cercando di ottimizzare l'uso dello spazio disponibile sul supporto, e rendere bassi i tempi di accesso e di ritrovamento dei dati contenuti nella memoria di massa. I file inoltre li distinguiamo anche in base all'uso che ne viene fatto. In particolare:

- File Principali, che contengono le informazioni utilizzate per gestire le entità che caratterizzano la realtà rappresentata con il modello informatico (Es. Magazzino, Clienti, Fornitori, Ordini, Contabilità...);
- File movimenti o file transazioni, che contengono informazioni che saranno utilizzate per aggiornare i file principali (Movimenti di magazzino, Giornale contabile, Fatture giornaliera, Bolle di accompagnamento, ...);
- File Storici, importanti per conservare nel tempo informazioni sulle entità, necessarie per motivi fiscali, statistici o comunque di sicurezza (Buste paga, Modello 740, Modello DM10, Riepilogo fatture emesse, ...).

Un'altra distinzione riguarda il modo con cui un File interagisce con la procedura che lo utilizza:

- chiameremo File di Input un file i cui dati vengono utilizzati dal programma
- chiameremo File di Output un file su cui verranno registrate le informazioni;
- chiameremo File di Input/Output un file da cui vengono lette le informazioni con eventuali aggiornamenti dei dati o registrazione di nuovi record.

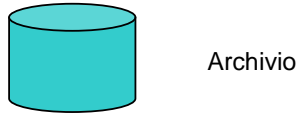
Per rappresentare in modo grafico le operazioni sugli archivi si utilizza il **Diagramma di flusso dei dati**: è uno schema costituito da simboli opportunamente collegati tra loro, ciascuno dei quali rappresenta un tipo di elemento che compone il sistema ed evidenziano in modo grafico le operazioni da svolgere sui dati contenuti negli archivi per risolvere un problema informatico.

Sono previsti quattro tipi di oggetti:

- 1) il **processo**, ovvero un programma che opera sui dati, elaborandoli;



- 2) l'**Archivio**, un serbatoio all'interno del quale vengono memorizzati e conservati i dati;



- 3) la **periferica**, dispositivo esterno con cui il sistema comunica;



Terminale



Stampa

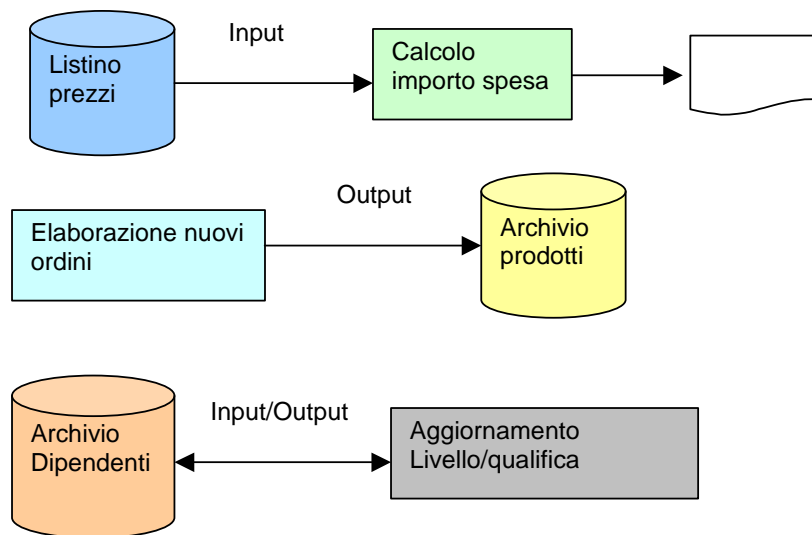


Nastro

- 4) il **flusso**, che rappresenta il percorso dei dati in entrata da un archivio o una periferica ad un processo o in uscita da un processo ad un archivio o ad una periferica.



Ad esempio, per rappresentare le modalità di utilizzo di file descritti precedentemente, si può schematizzare nel modo seguente:

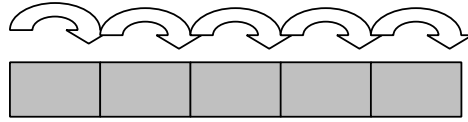


Le organizzazioni più utilizzate sono:

- SEQUENZIALE
- SEQUENZIALE CON INDICE
- DIRETTA

2.1 Organizzazione sequenziale.

Consiste nel registrare i record uno di seguito all'altro, in modo sequenziale.



Negli archivi sequenziali è sempre possibile definire un **ordine fisico** di registrazione. Se su uno o più campi è possibile definire un criterio di ordinamento si parlerà di **ordine logico** tra le registrazioni. Quando l'ordine logico coincide con l'ordine fisico, l'archivio verrà detto Sequenziale ordinato, altrimenti Sequenziale non ordinato o seriale.

2.1.1 Operazioni sugli archivi sequenziali.

CREAZIONE E CARICAMENTO.

Consistono nella generazione di un archivio vuoto e in un ciclo di caricamento in cui i singoli record vengono registrati in modo sequenziale. Nel caso si voglia lavorare con archivi sequenziali ordinati, sarà necessario far seguire questa fase da una di ordinamento.

RICERCA

Se si effettuano interrogazioni e aggiornamenti su di un archivio, una delle operazioni più frequenti è quella di ricerca di un record. La tecnica varierà in funzione del tipo di supporto utilizzato (nastri o dischi magnetici) e al fatto che esista o meno un ordinamento tra i record.

- Archivio non ordinato: la ricerca avviene effettuando la scansione dell'intero archivio fino al raggiungimento del record cercato.
- Archivio ordinato: in questo caso la ricerca partendo dal primo record, procede in modo sequenziale interrompendosi quando il record viene trovato, oppure quando il record esaminato ha una chiave superiore a quella cercata, secondo il metodo di ordinamento utilizzato.

INSERIMENTO

La tecnica varierà a seconda che esista o meno un ordinamento tra i record.

- Archivio non ordinato: in questo caso l'inserimento non presenta grossi problemi, in quanto è sufficiente aggiungere il nuovo elemento in coda al file;
- Archivio ordinato: in questo caso sarà necessario individuare la posizione corretta in cui porre la nuova registrazione. Si procede nel modo seguente:
 - Si crea un nuovo File;
 - Si ricopiano tutti i record che precedono quello che deve essere inserito;
 - Si inserisce il nuovo record nel nuovo file;
 - Si ricopiano i rimanenti record nel nuovo file;
 - Si cancella il vecchio file;
 - Si rinomina il nuovo file con il nome vecchio.

Una soluzione alternativa potrebbe essere quella di porre tutte le registrazioni nuove, acquisite in un certo periodo di tempo, in un archivio temporaneo, procedendo poi periodicamente ad una ristrutturazione dell'archivio principale.

AGGIORNAMENTO

Con questo termine si fa riferimento a quell'operazione con cui si modificano i valori di uno o più campi (eccetto la chiave) di un record già esistente. Nel caso di organizzazione sequenziale l'operazione si compone di due fasi:

- ricerca del record di chiave K e la sua modifica in memoria centrale;
- riscrittura dell'intero archivio con la sostituzione del record modificato, seguendo la procedura seguente:
 - si crea un nuovo file;
 - si ricopiano tutti i record che precedono quello da modificare;
 - si scrive il record modificato sul nuovo file;
 - si ricopia il resto del vecchio file nel nuovo;
 - si cancella quindi il vecchio file
 - si rinomina il nuovo file assegnandogli il nome di quello originario.

CANCELLAZIONE.

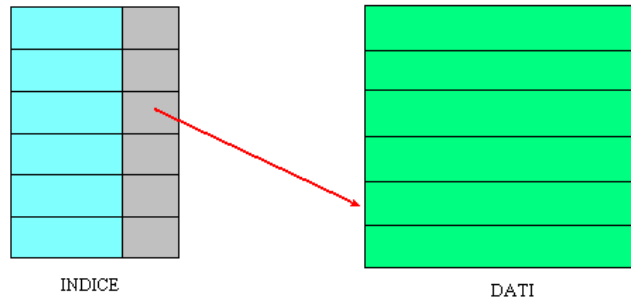
La cancellazione di un record dovrebbe avvenire nel modo seguente:

- creare un nuovo file
- ricopiare tutti i record che precedono quello da cancellare
- ignorare il record da cancellare
- ricopiare il resto del vecchio file nel nuovo
- cancellare il vecchio file
- rinominare il nuovo file assegnandogli il nome di quello originario.

2.2 Organizzazione sequenziale con indice.

Nelle applicazioni gestionali molto spesso le entità che devono essere rappresentate con record vengono identificate attraverso un elemento caratteristico (chiave) che non necessariamente è un numero, ma una variabile alfanumerica, ad esempio la matricola di un dipendente, il codice di un articolo, il codice di un cliente, il codice fiscale, ecc.

L'organizzazione dell'archivio in questo caso, per facilitare il ritrovamento dei dati, fa uso di una particolare archivio, detto **Archivio Indice**, in cui sono registrate secondo un certo ordinamento, le chiavi relative a ciascun record, a ciascuna delle quali poi è associato un **PUNTATORE** al record relativo, ovvero la posizione che il record occupa all'interno dell'**Archivio Principale**. L'organizzazione è detta Sequenziale con indice (Index sequential, o ISAM= Index Sequential Access Method).



In questo modo l'utente può accedere al record specificandone solo la chiave. E' possibile comunque accedere a tutti i record del file leggendoli in modo sequenziale a partire dal primo, sempre seguendo l'ordine delle chiavi.

L'indice può essere relativo alla chiave primaria oppure ad una delle chiavi secondarie previste dalla gestione dell'archivio.

Poiché l'archivio indice contiene soltanto la chiave e il puntatore al record dell'archivio dati corrispondente, conviene gestire l'archivio indice in una tabella di record in memoria, per rendere veloce la ricerca e l'inserimento ordinato di una nuova chiave. L'inserimento nell'archivio dei dati avverrà in modo sequenziale in fondo all'archivio. Ci penserà l'Indice a conservare l'ordinamento di lettura.

La procedura di gestione di un archivio ISAM si può sintetizzare:

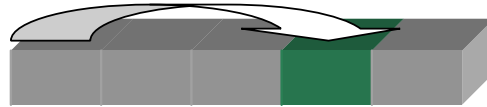
- L'archivio Indice verrà letto sequenzialmente e copiato nella tabella in memoria, avente lo stesso tracciato record.
- A seconda dell'operazione di inserimento, cancellazione logica, modifica di un record, la ricerca avverrà nella tabella indice secondo le tecniche di ricerca note.
- Il file dei dati resterà sulla memoria di massa e si accederà ad esso per numero record in modo diretto dopo avere operato sull'archivio indice. Una volta letto il record e visualizzati i dati, possono essere modificati e aggiornati facendo riferimento al numero del record che indica la sua posizione su memoria di massa.
- L'inserimento di un nuovo record avverrà in modo ordinato nella tabella e in fondo nell'archivio dei dati.
- Alla fine dell'elaborazione verrà ricopiato in memoria di massa la tabella di record per una successiva elaborazione.

2.3 Organizzazione diretta

Le organizzazioni Sequenziale e Sequenziale con Indice, rispondono in modo adeguato a buona parte delle problematiche legate alla gestione di grandi masse di dati, ma non risultano efficienti nei casi in cui gli archivi siano dinamici e di frequente consultazione, ovvero quando si presentano con una certa frequenza le operazioni di inserimento e di cancellazione. L'organizzazione in grado di rispondere con maggiore efficienza a questo problema viene detta ad **accesso diretto**, poiché si basa sull'idea di ricavare direttamente dal valore della chiave l'indirizzo della registrazione ad essa associata.

Questo tipo di organizzazione si chiama anche RANDOM, volendo mettere in evidenza il contrario di sequenziale, cioè l'utente può accedere al record che interessa senza leggere tutti quelli che lo precedono.

Ad ogni record viene associato un valore numerico che corrisponde alla posizione occupata dal record nel file (il primo record si trova in posizione relativa 1). In questo modo è possibile accedere ad un record facendo riferimento alla posizione che esso occupa nel file.



Questo può essere fatto in due maniere:

- a) Utilizzando come chiave la posizione relativa del record all'interno del File;
- b) Facendo uso di una Funzione, detta funzione di trasformazione, che trasformi la chiave in un indirizzo all'interno dell'archivio. La corrispondenza Chiave \leftrightarrow Record è garantita da un algoritmo, detto Algoritmo di randomizzazione, e il metodo di accesso viene spesso indicato come indirizzamento Hash.

Sul file random possono essere eseguite le seguenti operazioni:

- a) Apertura del file, per mettere a disposizione di un'applicazione i dati contenuti in un archivio.
- b) Lettura di un record che occupa una certa posizione all'interno del file, per portare in memoria centrale i campi registrati nel file.
- c) Scrittura di un nuovo record in coda al file, o nella posizione specificata dalla funzione di trasformazione della chiave, con risoluzione dell'eventuale problema dei sinonimi (trattato più avanti), per registrare sul file i campi del record presenti in memoria centrale.
- d) Posizionamento sul record da leggere o scrivere oppure per determinare a partire da quale record debba iniziare una lettura o scrittura sequenziale.
- e) Riscrittura, per modificare il contenuto di alcuni campi di un record che è stato letto nel file.
- f) Cancellazione di un record, per eliminare le informazioni che non servono più nelle applicazioni utilizzate. Di solito si imposta in modo opportuno un determinato campo del record (cancellazione logica), per escludere tale record dalle elaborazioni. In un secondo momento si procederà alla RIORGANIZZAZIONE dell'archivio per eliminare i record cancellati.
- g) Chiusura, per mettere fine al collegamento tra memoria centrale e memoria di massa attivato con l'apertura del file.