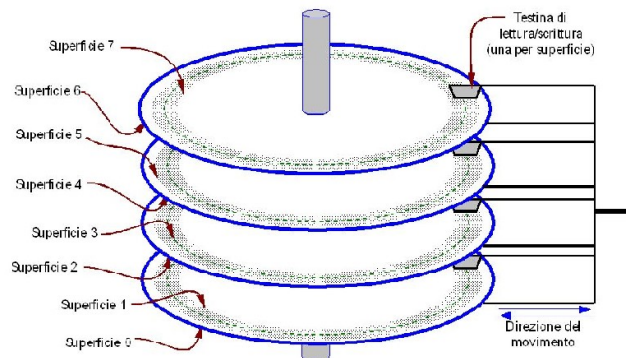
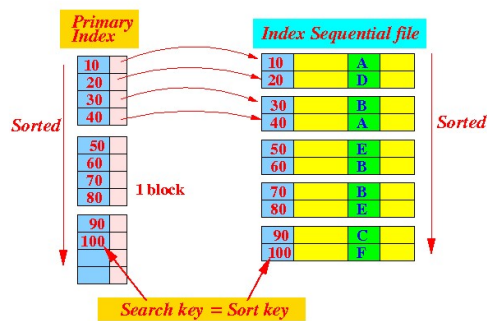
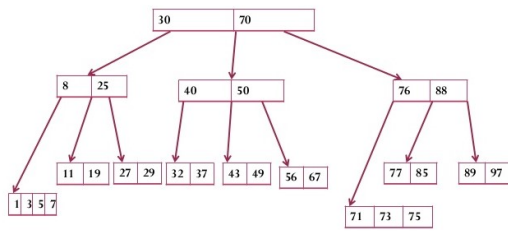
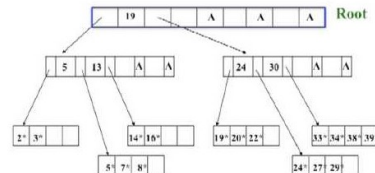


**Archiviazione a basso livello**

Lezioni del prof. Ennio Ranucci  
Appunti raccolti dagli alunni della V B INF. 2008/2009  
specializzazione Informatica Abacus



**Le tracce in grigio formano un "cilindro"**



# Archiviazione a basso livello:



I.T.I.S. "Francesco Giordani" Caserta  
Specializzazione informatica ABACUS  
prof. Ennio Ranucci

## Archivi cartacei e archivi elettronici

Schedario  $\leftrightarrow$  File(o archivio)

Schedone  $\leftrightarrow$  Record

Parte significativa di informazione  $\leftrightarrow$  Campo

## Criteri per la scelta della chiave di una informazione:

Un insieme di campi usato come identificatore di record di un file è detto chiave.  
Identificare un record significa distinguerlo da tutti gli altri.

### 1. Metodo della chiave implicita :

come chiave si usa una parte della informazione stessa

*Esempio: il numero di telefono di un elenco telefonico.*

### 2. Metodo della chiave mediante tabella :

all'informazione viene associato un codice mediante una tabella.

*Esempio : Ad ogni dipendente di una azienda viene assegnata una matricola.*

### 3. Metodo della chiave mediante regole logiche:

la chiave è determinata mediante l'applicazione di certe regole predefinite, ad una parte dell'informazione.

*Esempio : Codice fiscale.*

## Caratteristiche dei records

I records di un file sono detti **omogenei** se contengono tutti le stesse informazioni nello stesso ordine. Sono detti non omogenei se record diversi possono contenere informazioni di "tipo" diverso.

La **lunghezza di un record** è il numero di bytes utilizzati per rappresentare l'informazione e può essere:

1. **Fissa** : se la sua lunghezza è prestabilita e non dipende dal contenuto dei campi.
2. **Variable** : se il record ha lunghezza che si adatta al contenuto dei campi.

## Operazioni sugli archivi

- Creazione di un file;
- Operazioni di aggiornamento;
- Ricerca di un record di un file;
- Ordinamento di un file;
- Scansione;
- Fusione.

Le operazioni di aggiornamento sono:

- a) Inserzione di un record;
- b) Cancellazione di un record;
- c) Modifica di un record.

La ricerca può avvenire per confronto della chiave oppure per trasformazione della chiave.

## Organizzazione archivi

- a) Sequenziale;
- b) Casuale;
- c) Mista.

L'**organizzazione sequenziale** accede ad un dato record solo dopo che è avvenuto un accesso a tutti quelli che nel file lo precedono.

L'**organizzazione casuale** accede ad un record direttamente.

**Mista** utilizza entrambe le tecniche sopra indicate.

Esempio: Si effettua una ricerca sequenziale nell'indice del libro e poi, trovata la pagina, si accede direttamente ad essa.

## Criteri per l'organizzazione dell'informazione

Metodo della suddivisione delle informazioni	Metodo del collegamento fra informazioni
Si suddividono le informazioni in tanti files quante sono le informazioni possibili	Si crea il file principale che contiene le informazioni comuni a tutte le elaborazioni e i riferimenti ai files secondari per ciascuna elaborazione.
<p style="text-align: center;">Esempio</p> File1 : Matricola, Titolo, Nome, Cognome, Residenza  File 2 : Matricola, Stipendio, Contingenza, Scatti	<p style="text-align: center;">Esempio</p> File principale: Matricola, Titolo, Nome, Cognome, Residenza, Riferimento Stipendio, Riferimento CUD.  File secondario 1: Livello, Ruolo, Numero ore straordinario, anzianità di servizio.  File 2: Totale reddito imponibile, detrazioni.

## Raccolta, Immissione e Codifiche dei dati

4 fasi :

- a) Acquisizione
- b) Rappresentazione
- c) Immissione e controllo
- d) Validazione

Nella fase di immissione e controllo possono verificarsi 3 tipi di errori :

**a) Errori impliciti :**

L'errore è nell'informazione stessa.

Esempio : "L'altezza di un monte non è corretta"

**b) Errore di rappresentazione:**

Errori che si verificano durante la rappresentazione.

Esempio : "Tabella di corrispondenza tra i valori e la loro rappresentazione" non corretta.

**c) Errore di trascrizione:**

Errori che si presentano durante l'immissione dei dati.

Esempio : "Inversione di due caratteri"

### Controlli di ridondanza e quadratura

Un controllo di ridondanza (informazione non necessaria) consiste nell'uso di informazioni ridondanti che possono essere utilizzate per controllare l'esattezza dell'informazione stessa.

In pratica ogni dato è diviso in due parti:

- a) Informazione
- b) Codice autocontrollante

Esempio : "Codice del conto corrente: 3456/23"

Fasi da seguire per il controllo di ridondanza:

- 1) Acquisire l'informazione costituita dall'informazione vera e propria e dal codice autocontrollante.
- 2) Applicare l'algoritmo al codice ottenendo un codice autocontrollante.
- 3) Confrontare il codice con il codice autocontrollante.

#### Algoritmo pari dispari

Num è l'informazione vera e propria

Cod = codice autocontrollante da calcolare

K = numero di cifre utilizzate per rappresentare il codice autocontrollante.

Algoritmo:

- 1) Si moltiplicano le cifre di "Num" di posto dispari \* 1 e quelle di posto pari \* 2;
- 2) Si calcola la somma dei prodotti ottenuti;
- 3) Se la somma calcolata è costituita da k cifre termina l'algoritmo, altrimenti si applica di nuovo lo stesso algoritmo.

Esempio 1:

$k=1$

Num = 13452

1)I prodotti sono :  $1*1,3*2,4*1,5*2,2*1$

2)La somma dei prodotti è uguale a :  $1+6+4+10+2 = 23$

Essendo  $k$  uguale a 2 si ripete l'algoritmo

1)I prodotti sono :  $2*2,3*1$

2)La somma dei prodotti è uguale a :  $4+3 = 7$

Quindi si scrive  $13452 / 7$

Esempio 2:

Num = 15234

$K = 2$

1)I prodotti sono :  $1*1,5*2,2*1,3*2,4*1$

2)La somma dei prodotti è uguale a :  $1+10+2+6+4 = 23$

Quindi si scrive :  $15234 / 23$

Conclusione :

Applicando l'algoritmo pari/dispari alle due coppie di chiavi 13452 e 13425, 15234 e 13452, si evince che l'algoritmo pari/dispari evidenzia gli errori di scambio di posizione di due cifre ma non evidenzia errori nello scambio di gruppi di due cifre.

### L'algoritmo dei numeri primi dispari

1)Moltiplicare la prima cifra a destra di Num\*3, la seconda cifra a destra di Num\*5, la terza cifra a destra di Num\*7, e così via, seguendo la successione dei numeri primi dispari.

2)Calcolare la somma dei prodotti precedenti e dividerla per l'  $h+2$  esimo numero primo dispari, (con  $h$  uguale al numero delle cifre di Num ).

3)Se  $R$  (resto della divisione precedente) è un numero di  $k$  cifre allora è il codice autocontrollante altrimenti applicare di nuovo l'algoritmo ad  $R$ .

Esempi :

Algoritmo(1) :

Chiave = 15234

$K = 2$

1) $1*13,5*11,2*7,3*5,4*3$

2) $13+55+14+15+12 = 109$

$$109 / 19 = 5$$

3) Il resto è 14

Algoritmo(2) :

Chiave = 13452

$K = 2$

$H = 5 \rightarrow$  (H+2esimo numero primo = 19)

1) $1*13,3*11,4*7,5*5,2*3$

2) $13+33+28+25+6 = 105$

$$105 / 19 = 5$$

3)Il resto è 10

Conclusione :

Con questo algoritmo si evidenziano anche errori nell'immissione del codice dovuti a scambi di gruppi di due cifre.

### **Validazione**

Qualunque sia la modalità d'immissione dei dati e qualunque siano i controlli effettuati durante l'immissione è opportuno sottoporre i dati ad una ulteriore verifica attraverso opportuni programmi. Questa fase è detta **validazione** e riguarda soprattutto il significato del dato.

Esempio(1)

Età = 200;

Un programma può verificare che il dato non è valido.

Esempio(2)

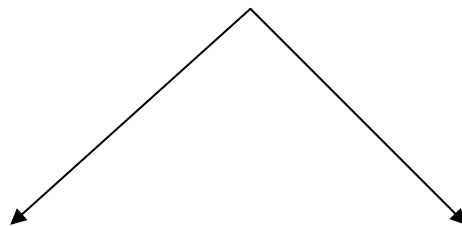
Un dipendente può essere assunto con una determinata qualifica solo se possiede il titolo di studio richiesto.

## Aspetti fondamentali nell'organizzazione di un archivio

1)Economia del tempo legato al tempo di ricerca.

2)Economia dello spazio di memoria legato alla rappresentazione dell'informazione.

### Criteria di economia dello spazio



#### **Compattamento dell'informazione**

1)Campi a lunghezza variabile(risparmio spazio in quanto adeguo la lunghezza)

- Anteporre lunghezza attuale
- Anteporre tabella delle lunghezze dei campi a lunghezza variabile
- Usare carattere speciale di fine gambo

2)Campi opzionali

- Tabella dei bits(0 campo presente)  
(1 campo assente)
- Tabella di codici(mettiamo un codice per ogni campo opzionale)  
Rinvio ad una tabella per l'interpretazione del dato

3)Criteria misti

- Tabella Lunghezza
- Tabella Codici/Lunghezza

#### **Collegamento a dati comuni**

1)Collegamento tramite rinvio(tramite tabella il rinvio)

2)Collegamento tramite gerarchia  
(Ex : tutti quelli che abitano a via Laviano)

# Classificazione dei metodi di accesso agli archivi

Get uno solo : --> ricerca per chiave primaria

Get alcuni : --> ricerca per chiave secondaria

Get tutti (o molti): --> ricerca per chiave secondaria

Nota : detto "A" il predicato di ricerca, i records del file che soddisfano il predicato "A" sono detti "records obiettivo"

## Organizzazione sequenziale

- Organizzazione sequenziale non ordinata

Definizione :

Un file sequenziale consiste in un insieme di blocchi contigui, ciascuno dei quali contiene un numero prefissato di records.

Questa organizzazione viene detta "flat file".

Organizzazione adatta per le ricerche di tipo get tutti e get molti.

La ricerca può essere solo lineare.

Aggiornamento:

1)Inserimento: può avvenire in coda per sovrapposizione dei records cancellati logicamente.

2)Modifica: Presupposto essenziale(ricerca con successo).

- Lettura del record interessata alla modifica
- Modifica del campo/i indicato/i
- Appoggio sul record temporaneo
- Riposizionamento della testina
- Scrittura del record temporaneo nel file.

3)Cancellazione:

- Cancellazione logica: Operazioni previste per la modifica, modificando o il tag di cancellazione o una parte della chiave.
- Cancellazione fisica

\*Nel caso della cancellazione logica si ha una maggiore occupazione di spazio.

## Organizzazione sequenziale ordinata

Definizione : Un insieme di blocchi contigui ciascuno dei quali contiene un numero prefissato di records ordinati rispetto al campo chiave.

Organizzazione adatta alla ricerca get tutti o get molti.

Tipo di ricerca può essere: ricerca lineare, ricerca lineare ordinata, ricerca binaria/dicotomica/logaritmica.

Aggiornamento:

1) Inserimento: L'inserimento viene gestito con le aree di Overflow:

- Si determina il blocco nel quale il record dovrebbe trovarsi in accordo con l'ordinamento logico
- Se c'è spazio nel blocco il record viene memorizzato altrimenti si memorizza il record nell'area di Overflow, provvedendo all'aggiornamento dei puntatori.

L'inserimento può avvenire anche con il **metodo delle riorganizzazioni periodiche**:

- Si inseriscono i nuovi records in un file secondario;
- Successivamente si procede alla fusione del file primario con il file secondario ed il nuovo file diventa primario.

Questo tipo di metodo può essere utilizzato quando le nuove inserzioni vengono utilizzate solo dopo un precisato intervallo di tempo.

2) Modifica: Presupposto essenziale (ricerca con successo).

- Lettura del record interessata alla modifica
- Modifica del campo/i indicato/i
- Appoggio sul record temporaneo
- Riposizionamento della testina
- Scrittura del record temporaneo nel file.

In pratica come nella organizzazione sequenziale non ordinata.

3) Cancellazione:

- Cancellazione logica: modificando il tag di cancellazione o una parte della chiave.
- Cancellazione fisica

I file sequenziali ordinati possono essere periodicamente riorganizzati al fine di compattare ed ottimizzare l'archivio.

## Ordinamento esterno

L'ordinamento esterno presenta il seguente problema:

In genere i records del file non possono essere tutti contenuti in memoria centrale per ovvie limitazioni di spazio.

Si procede quindi alla suddivisione del file in sottoarchivi tali che ciascun sottoarchivio contenga un numero di records adeguato per essere trasferiti in memoria centrale.

Ogni sottoarchivio può essere ordinato in memoria centrale e salvato in memoria permanente.

Operando, poi, la fusione dei sottoarchivi direttamente sul supporto di memoria permanente si ottiene l'ordinamento esterno.



## Organizzazione sequenziale con indice

Definizione:

L'organizzazione sequenziale con indice è costituita da un file sequenziale con eventuali aree di Overflow ed una gerarchia di indice di blocco.

Vantaggi:

Migliora la velocità di accesso.

Organizzazione adatta al caso get tutti e get molti e anche get uno solo.

- Get tutti : trascura gli indici
- Get uno solo : Usa gli indici

La ricerca avviene nel seguente modo:

a) Ricerca lineare sul file

b) Accesso diretto al blocco nel quale dovrebbe trovarsi il record da ricercare; Il blocco viene individuato attraverso gli indici.

Ad esempio : "Supponiamo di avere un solo indice, si cerca in modo sequenziale nell'indice la prima chiave che sia maggiore od uguale alla chiave di ricerca.

Mediante il puntatore associato si accede al blocco che dovrebbe contenere la chiave di ricerca.

Aggiornamento:

Presenta gli stessi problemi dell'organizzazione sequenziale non ordinata provvedendo quando necessario all'aggiornamento degli indici.

## ISAM (Indexed Sequential Access Method)

L'ISAM è una organizzazione dipendente dall'hardware, in esso i records e gli indici sono organizzati in maniera tale da riempire perfettamente le tracce dei cilindri e da minimizzare gli spostamenti delle testine.

I livelli di indici sono 3 :

- Indici di Volume
- Indici di Cilindro
- Indici di Traccia

Gli indici di Traccia sono dispersi tra i dati. Per ciascuna traccia è rappresentato il valore più grande delle chiavi in quella traccia.

L'indice di cilindro è memorizzato su un cilindro diverso dai cilindri contenenti i dati.

L'indice di Volume o Unità è allocato permanentemente in memoria centrale.

## VSAM (Virtual Storage Address Method)

È una evoluzione dell'ISAM indipendente dall'hardware. Un file VSAM è organizzato in aree di controllo per contenere i dati e ciascuna area di controllo è suddivisa in intervalli di controllo.

Le aree di controllo sono l'analogo dei cilindri.

Gli intervalli di controllo sono l'analogo delle tracce.

## Alberi – B (B-Trees)

Definizione:

Un albero- B di ordine M è un albero che soddisfa le seguenti proprietà:

- Le foglie dell'albero sono tutte allo stesso livello.
- Ogni nodo dell'albero, inclusa la radice, ha un numero "D" di figli tale che D è minore uguale a  $2M+1$ .
- Ogni nodo dell'albero, esclusa la radice ha un numero D di figli tale che D sia maggiore uguale a  $M+1$ .
- La radice è una foglia oppure ha almeno 2 figli.

L'organizzazione è adatta alla ricerca get uno solo.

### Algoritmo di ricerca in un albero B

- Si trasferisce in memoria centrale la radice.
- Si ricerca il record di chiave K tra i records memorizzati nella pagina.

Nel caso di ricerca senza successo si possono verificare 3 casi:

- La chiave K è minore di  $K_1$  (la prima chiave della pagina corrente)
- La chiave K è compresa tra una qualunque coppia di chiavi consecutive.
- La chiave k è maggiore di  $K_d$  (l'ultima chiave della pagina corrente)

Nel caso della ricerca senza successo si trasferisce in memoria centrale la pagina così determinata:

- Se  $K < K_1$  si trasferisce la pagina alla quale fa riferimento il puntatore  $P_0$ .
- Se K è compreso tra  $K_i$  e  $K_{i+1}$  si trasferisce la pagina alla quale fa riferimento il puntatore  $P_i$ .
- Se  $K > K_d$  si trasferisce la pagina alla quale fa riferimento il puntatore  $P_d$ .
- La ricerca termina senza successo quando il puntatore scelto è nullo.

### Inserzione

L'inserzione è possibile se la ricerca termina senza successo.

Se nella ultima pagina visitata c'è spazio libero allora inserisco il nuovo record altrimenti si procede all'operazione di suddivisione(Split).

Quindi nel caso che nell'ultima pagina visitata siano già memorizzati  $2M$  records viene effettuata la seguente operazione detta suddivisione:

Si considera l'insieme ordinato dei records memorizzati nella pagina e del record da inserire.

Questo insieme viene suddiviso in due sottoinsiemi:

Il primo "A" costituito da tutti i records con chiave minore della chiave del record che occupa la posizione mediana dell'insieme di partenza.

Il secondo sottoinsieme "B" costituito da tutti i records con chiave maggiore della chiave del record che occupa posizione mediana dell'insieme di partenza.

I records di A restano memorizzati nella pagina corrente.

I records di B vengono memorizzati in una nuova pagina o un nodo allo stesso livello dell'albero.

Il record che occupa la posizione mediana viene memorizzato nella pagina madre(iniziale).

### Cancellazione

La cancellazione di un record richiede la ricerca con successo.

Distinguiamo due casi :

- Se la ricerca con successo termina in una foglia è sufficiente cancellare il record dalla pagina(livello h-1).

B) Se la ricerca termina in una pagina ad un livello compreso tra 0 e  $h-2$ , allora la chiave  $k$  svolge la funzione di separatore e deve essere pertanto sostituito da una chiave di un altro record che abbia lo stesso significato.

Scegliamo, quindi, come chiave separatore da sostituire alla chiave da cancellare la prima della foglia che si raggiunge seguendo il ramo dei maggioranti della chiave da sostituire, scegliendo sempre il primo puntatore a sinistra.

## Concatenazione e Underflow

Quindi la cancellazione in ogni caso elimina un record da una foglia. Se nella foglia dopo l'eliminazione della chiave rimangono meno di  $M$  records allora il nodo non soddisfa uno dei requisiti degli alberi B di ordine  $M$ . Diventa necessario effettuare delle operazioni per ovviare all'inconveniente.

Queste operazioni sono chiamate **UNDERFLOW** e **CONCATENAZIONE**.

- Se il numero di records della pagina in cui avviene la cancellazione sommato al numero di records della pagina adiacente è maggiore o uguale a  $2M$  allora si esegue l'operazione di **UNDERFLOW**
- Se il numero di records della pagina in cui avviene la cancellazione sommato al numero di records della pagina adiacente è uguale a  $2M-1$  allora si esegue l'operazione di **CONCATENAZIONE**.

## Underflow

Nell'operazione di underflow si prende in considerazione l'insieme costituito dalle chiavi dei records rimasti nella pagina in cui avviene la cancellazione, quelli del nodo adiacente e dalla chiave del record che opera come separatore tra le pagine prima indicate. Questo insieme viene suddiviso in due sottoinsiemi, prendendo come separatore il record mediano.

Quest'ultimo viene allocato come nuovo separatore tra le 2 pagine prese in esame.

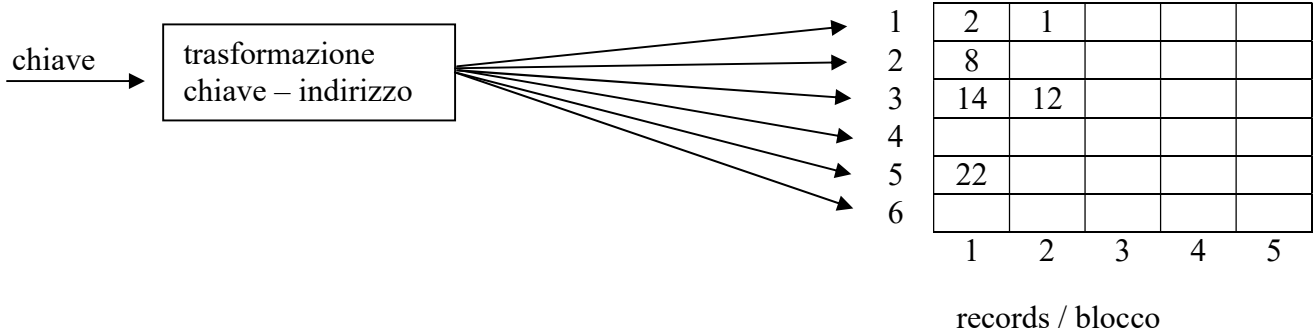
- I records con chiave minore della chiave del record mediano sono memorizzati dalla pagina in cui avviene la cancellazione.
- I records con chiave maggiore della chiave del record mediano sono memorizzati nella pagina adiacente.

## Concatenazione

L'operazione di concatenazione è l'operazione inversa della suddivisione (SPLIT), con essa, infatti i  $2M-1$  records memorizzati nelle pagine adiacenti e il corrispondente record separatore vengono memorizzati nella pagina in cui avviene la cancellazione.

## Accesso Mirato

Accedere in modo mirato ad un record di un file significa poter reperire quel record leggendo dalla memoria di massa un solo blocco, quello che contiene il record.



### Esempio di trasformazione delle chiavi in indirizzo:

Le 2000 chiavi possibili dei record sono le seguenti:

- AT2000 a AT2999
- PC500 a PC1499

- 1) Se i primi due caratteri della chiave sono AT, sottrae 1999 al valore numerico.
- 2) Se i primi due caratteri della chiave sono PC, somma 501 al valore numerico.

Si supponga che le 2000 chiavi siano:

- AT2000 a AT2999
- PC500 a PC1499

Gli indirizzi corrispondenti saranno:

- Da 1 a 1000
- Da 1001 a 2000

## Organizzazione Hash

**Definizione:** L'organizzazione hash è costituita da un'area primaria, da un'area di overflow e dalla funzione hash.

L'area primaria è il file hash che è suddiviso in buckets. Un bucket è un conveniente raggruppamento di records fisici, normalmente si considera per bucket una traccia di disco.

Logicamente un bucket è suddiviso in celle o slot, una cella è lo spazio riservato alla memorizzazione di un solo record. Ad ogni bucket, è associato un indirizzo relativo all'interno del file hash.

Organizzazione adatta alla ricerca get uno solo.

### Inserimento:

Tramite la funzione hash alla chiave viene associato l'indirizzo di casa(home address) del record.

Nota: Detto N il numero di buckets che costituiscono il file hash, l'intervallo 1..N è detto spazio degli indirizzi.

Due o più chiavi per le quali la funzione hash genera uno stesso indirizzo di casa sono dette sinonimi. Una collisione avviene quando per effettuare l'inserzione di un nuovo record la funzione hash determina per la chiave del record da inserire un indirizzo già occupato da un record con una chiave diversa.

La risoluzione di una collisione richiede un intervento in due passi:

- 1) Il primo passo consiste nello scandire il bucket al fine di determinare se esiste nel suo interno una cella libera dove memorizzare il record.
- 2) Se il bucket è già pieno, si determina un overflow che deve essere opportunamente risolto con una tecnica di trattamento dell'overflow.

### Tecniche di trattamento dell'overflow

- a) Metodo dell'overflow aperto.

Questo metodo memorizza il nuovo record nella prima cella libera del primo bucket non pieno successivo al bucket pieno.

Pseudocodifica:

*N = Massimo intervallo dello spazio degli indirizzi*

*I = CHIAVE(MOD 10) + 1*

*IF OVERFLOW THEN*

*REPEAT*

*I = I + 1*

*UNTIL il bucket I è non pieno*

*END REPEAT*

*END IF*

*Inserisci il record in I*

- b) Metodo della ricerca non lineare.

Allo scopo di evitare il fenomeno dell'agglomerazione(clustering) cioè il raggruppamento di records sinonimi in buckets successivi, sono stati predisposti altri metodi basati su una ricerca non sequenziale.

Il metodo della ricerca quadrati utilizza una successione di indirizzi(REHASH) generata per mezzo di una funzione di trasformazione quadratica.

La forma della funzione quadrata è  $A \cdot i^2 + B \cdot i + C$ .

"C" è l'indirizzo di casa per il record da inserire.

"A" e "B" sono due costanti arbitrarie.

Pseudocodifica

*N = Massimo intervallo dello spazio degli indirizzi*

*X=0*

*Indcasa = CHIAVE(MOD N)+1*

*IF OVERFLOW THEN*

*REPEAT*

*X=X+1*

*I = (X<sup>2</sup>+3X+ Indcasa)*

*(MOD N) + 1*

*UNTIL I non è pieno*

*END REPEAT*

*Indcasa = I*

*END IF*  
*Inserisci il record in Indcasa*

c) Metodo delle catene mescolate

Il metodo utilizza una catena costituita dai buckets del file principale che non sono pieni. L'Overflow di un qualsiasi buckets viene memorizzato in testa alla suddetta catena.

	2					
1	7	10	51	71	21	NIL
2	5	69	105	204		4
3	3	4	1	12	17	NIL
4						6
5	1	21	31	41	52	NIL
6						NIL

Pseudocodifica

*N = Massimo dell'intervallo dello spazio degli indirizzi*  
*A = CHIAVE(MOD N) + 1*  
*IF A non è pieno*  
    *THEN inserisci record in A*  
    *ELSE B= indirizzo bucket in testa catena di overflow*  
        *Inserisci record in B*  
    *IF B risulta pieno*  
        *THEN cancella B dalla catena di overflow*  
    *END IF*  
*END IF*

d) Metodo delle catene separate

Questo metodo utilizza un'area di Overflow fisicamente diversa dall'area primaria dell'archivio. Con questo metodo si mantengono delle catene di buckets di Overflow, ciascuna costituita da records di Overflow tutti provenienti dallo stesso indirizzo di casa.

Pseudocodifica

*N = Massimo dell'intervallo dello spazio degli indirizzi*  
*A = CHIAVE(MOD N) + 1*  
*IF A non è pieno*  
    *THEN inserisci record in A*  
    *ELSE accede alla catena dei buckets*  
        *IF esiste un bucket B non pieno*  
            *THEN inserisci record in B*  
            *ELSE aggiungi il bucket C alla catena*  
                *togli C dalla parte non utilizzata dall'area di Overflow*  
                *inserisci record in C*  
    *END IF*  
*END IF*

## Metodo di ricerca per chiave secondaria

Definizione: Un metodo di accesso secondario è una organizzazione fisica di un file che permette di accedere in modo efficiente ai records obiettivo determinati per mezzo di un insieme di valori di

chiavi secondarie specificato in una ricerca su chiavi secondarie. Una ricerca per chiave secondaria è detta anche domanda e permette di accedere ad una certa percentuale di records che costituiscono un file.

### **Definizione di condizione atomica**

Le condizioni atomiche sono quelle attorno alle quali è costruita una domanda. La forma è Nome di campo {<=>} valore di campo.

Esempio :

Citta = 'Caserta'

### **Definizione di condizioni su campo**

Una condizione di campo, CC, è una disgiunzione di condizioni atomiche CA1 or CA2 or ... or CAN.

Esempio:

eta = 21 or eta = 22 or eta = 30

### **Definizione di condizione di record**

Una condizione di record, CR, è una congiunzione di condizioni di campo, CC1 And CC2 And.. And CCN.

Esempio:

(eta = 21 or eta = 22) And (citta = 'Caserta' or citta = 'Napoli')

### **Definizione di condizione di domanda**

Una condizione di domanda, CD, è una disgiunzione di condizioni di record, CR1 or CR2 or ... or CRN.

Esempio:

Get From Impiegati Where

(eta = 18-21 And (citta='Caserta' or citta = 'Napoli') And codicelavoro = 'Programmatore') or (eta>50 And codicelavoro = 'Analista')

## **Metodo delle multicatene**

Definizione: Ad ogni record vengono associati, uno per ogni chiave secondaria tanti puntatori quante sono le chiavi secondarie che lo costituiscono. Ogni puntatore determina, indipendentemente dagli altri, quella particolare sequenza di accessi ai records che permette di ottenere un particolare raggruppamento logico determinato da un valore nella chiave secondaria corrispondente.

Esempio:

Indirizzo	Matricola	Cognome	CodiceLavoro		Città		Età	
0	1	Rossi	P	2	CE	1	22	3
1	2	Bianchi	A	4	CE	3	26	2
2	3	Verdi	P	3	NA	4	26	NIL
3	4	Viola	P	NIL	CE	NIL	22	NIL
4	5	Gialli	A	NIL	NA	NIL	35	NIL

Organizzazione adatta alla ricerca get alcuni.

## **Ricerca**

Esempio 1:

Get From Impiegati Where codicelavoro='Programmatore'

E' sufficiente ricercare il valore programmatore nel campo codicelavoro e poi seguire la corrispondente catena.

L'organizzazione a multicatene è particolarmente efficace nel caso di condizioni di domanda costituite da una sola condizione atomica.

Esempio 2:

Get From Impiegati Where eta<35

E' necessario ricercare i valori minori di 35, accedere alle catene corrispondenti e poi fare l'unione di queste.

Esempio 3:

Get From Impiegati Where Eta = 18-21

Come esempio 2.

La strategia di ricerca in una organizzazione a multicatene nel caso di condizioni del record costituite da più di una condizione di campo consiste:

- 1) Si crea un indice di livello 1 contenente le chiavi secondarie e un sottoindice di livello 2 contenente i diversi valori di ciascuna chiave secondaria.
- 2) Si individuano negli indici al secondo livello di ingressi relative alle condizioni di campo specificate.
- 3) Si determina la catena più corta per mezzo dei contatori associati agli ingressi individuati.
- 4) Si accede solo ai records della catena più corta; Per ciascun record di questa catena si controlla se la congiunzione delle altre condizioni di campo è verificata.

Esempio 4:

Get From Impiegati Where eta = 26 And (citta='Caserta' or citta='Napoli' And codicelavoro='Programmatore').

## **Le operazioni di aggiornamento**

### **Inserzione**

L'inserzione di un nuovo record in un file organizzato a multicatene comporta l'inserzione di quel record in tutte le catene delle quali deve fare parte.

Se la multicatena è organizzata con una gerarchia di indice nel caso che il record appena inserito sia il primo record che contiene un certo valore in un dato campo allora deve essere inserito un nuovo ingresso nel sottoindice corrispondente.

### **Cancellazione**



In modo analogo la cancellazione di un record comporta la cancellazione di quel record da tutte le catene delle quali fa parte.

Una cancellazione può interessare l'unico record del file che contiene un certo valore in un dato campo, allora deve essere cancellato dal sottoindice l'ingresso relativo a quel valore.

### **Modifica**

La modifica di una chiave secondaria richiede la cancellazione del vecchio record dalla catena cui apparteneva in virtù del valore da modificare e la inserzione del nuovo record nella catena alla quale deve appartenere in virtù del valore modificato.

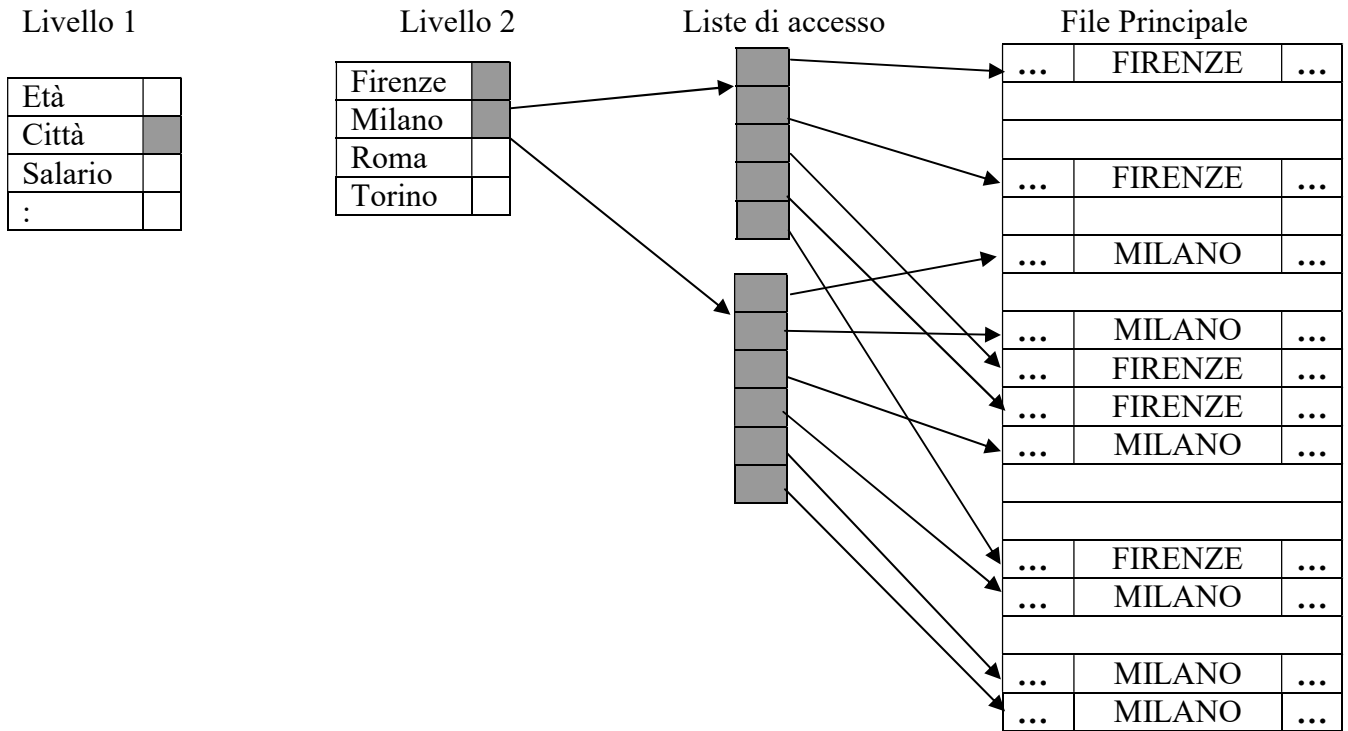
L'organizzazione a multicatene di un file F perde via via di efficienza con l'aumentare della complessità della domanda rivolta su F .

# L'organizzazione a files invertiti

Definizione: In una organizzazione a file invertiti per collegare tra di loro quei record che hanno lo stesso valore nello stesso campo si utilizzano le liste di accesso che sono insieme di puntatori a records.

Organizzazione adatta alla ricerca get alcuni.

Esempio:



## Ricerca

La ricerca viene effettuata visitando tutti i record della lista di accesso interessata.

## Operazioni di aggiornamento

Le operazioni di aggiornamento in modo analogo alle organizzazioni a multicatene richiedono l'inserimento o la cancellazione dei puntatori nelle liste di accesso ed eventualmente nei sottoindici corrispondenti.