

Archivi

Sul libro...



1

L'organizzazione degli archivi e le basi di dati

- | | | |
|----------|---|----|
| 1 | Gli archivi | 6 |
| 2 | I file e le memorie di massa | 8 |
| 3 | L'organizzazione degli archivi | 11 |
| 4 | Le applicazioni gestionali | 14 |
| 5 | Le basi di dati | 15 |
| 6 | I limiti dell'organizzazione convenzionale degli archivi | 17 |

Definizione di archivio

- Un **archivio** è un insieme organizzato di informazioni caratterizzate da alcune proprietà fondamentali:
 - Esiste un **nesso logico** tra di esse (cioè sono in qualche modo inerenti a un medesimo argomento);
 - sono rappresentate secondo un **formato** che ne rende possibile l'interpretazione;
 - sono registrate su un **supporto** di memoria **permanente** su cui è possibile leggere e scrivere;
 - sono **organizzate** in modo che siano facilmente consultabili.

Archivio e record

- Un **archivio** è una struttura dati astratta costituita da un insieme di **record**
 - Ogni record è identificato dalla sua posizione all'interno dell'archivio: **indirizzo logico**
 - Esempio: anagrafiche degli studenti

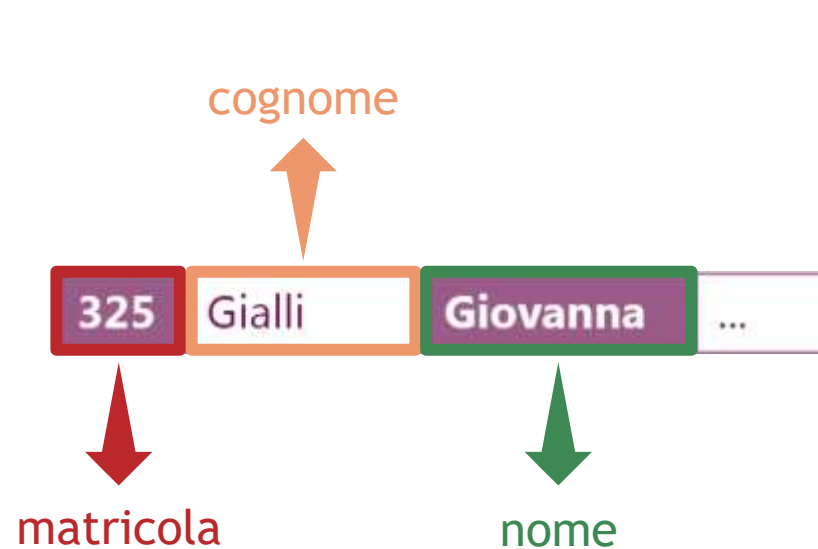
225	Bianchi	Giovanni	...
236	Carminati	Alfredo	...
325	Gialli	Giovanna	...
425	Verdi	Enrico	...
456	Negri	Francesca	...
538	Viola	Annalisa	...
585	Rossi	Giuliano	...
624	Magnani	Pietro	...
788	Bruni	Alessandra	...



Record: insieme di dati logicamente omogenei che descrivono i singoli elementi di una realtà considerata

Record e campi

- Un **record** è una struttura di dati composta da un insieme finito di elementi eterogenei logicamente connessi detti **campi**



↓

I **campi** contengono i valori assunti dalle caratteristiche scelte per descrivere gli elementi della realtà considerata (attributi)

- L'elenco dei campi costituisce il **tracciato del record**



FILE

- Il **file** è una struttura fisica di memoria in cui è possibile memorizzare informazioni sotto forma di sequenze di byte o sequenza di record.
 - È la struttura adatta a implementare la struttura dati astratta archivio
 - Ogni record all'interno del file è individuato tramite **indirizzo fisico**
- Un archivio di dati è sempre implementato mediante uno o più file.



Creazione di un archivio

- La **creazione di un archivio** richiede la definizione preliminare (durante la fase di analisi del problema) delle seguenti specifiche:
 - il **nome** dell'archivio, che lo identifica e serve a ricordarne il contenuto (esempio "archivio fornitori");
 - il **tracciato del record**;
 - il **supporto** da usare per archiviare i dati;
 - la **dimensione massima** dell'archivio;
 - il modo con cui i dati sono strutturati e collegati tra loro, cioè l'**organizzazione** dell'archivio.

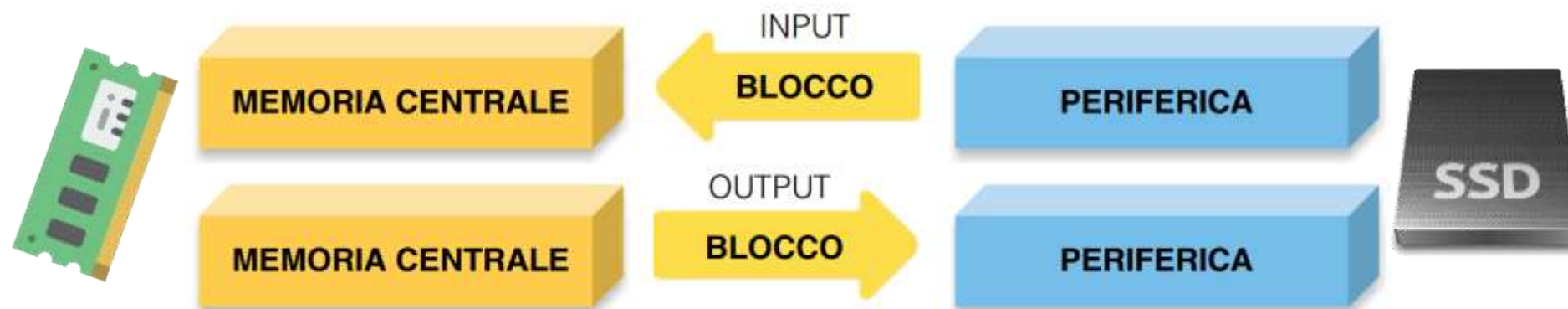
Creazione di un archivio

- Dopo aver creato l'archivio, su di esso si possono effettuare operazioni di:
 - **manipolazione**, cioè inserimento di nuovi dati o variazione dei dati registrati;
 - **interrogazione**, cioè reperimento all'interno dell'archivio delle informazioni necessarie.



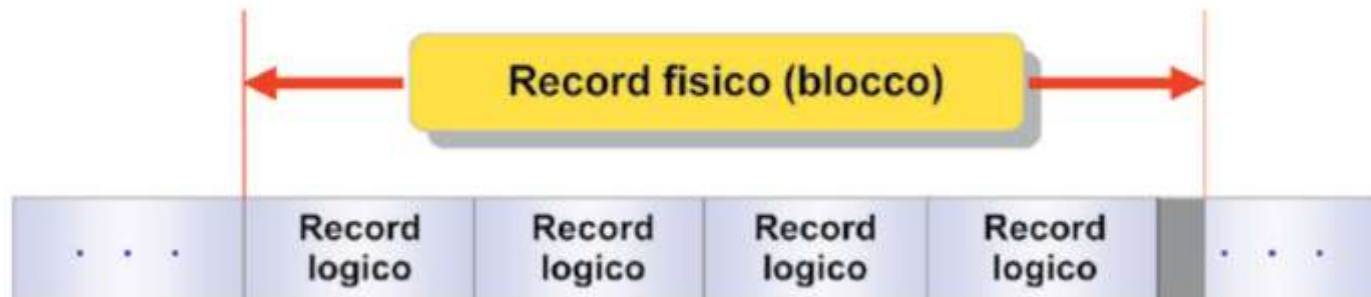
Blocco fisico

- I dati tra periferica e memoria centrale vengono trasferiti in blocchi (**record fisico**)
 - La dimensione del blocco varia da sistema a sistema e assume valori dell'ordine di alcuni KB: 2KB, 4KB, 8KB
 - I dati provenienti dal disco vengono copiati in una zona di memoria centrale, detta **buffer** di I/O.



Blocco fisico e record logico

- Un blocco fisico può contenere più record logici.
- Si definisce **fattore di blocco** di un file il numero di record logici contenuti in un blocco.



Organizzazione degli archivi



- La registrazione dei dati su memoria di massa deve essere fatta in modo da organizzare i dati nel migliore dei modi possibile, tenendo conto dei parametri di spazio e tempo.
- Organizzazioni possibili:
 - Organizzazione sequenziale
 - Organizzazione ad accesso diretto
 - Organizzazione a indici

Organizzazione sequenziale

- L'organizzazione sequenziale consiste nel registrare i record uno di seguito all'altro, in modo sequenziale, intervallati da sequenze di caratteri che indicano la fine del record (EOR).
- Consente l'uso di record a lunghezza variabile.



Organizzazione sequenziale

- L'organizzazione sequenziale di un archivio rende possibile accedere a un record solo dopo aver visitato tutti i record che lo precedono. Si parla in tal caso di accesso sequenziale.

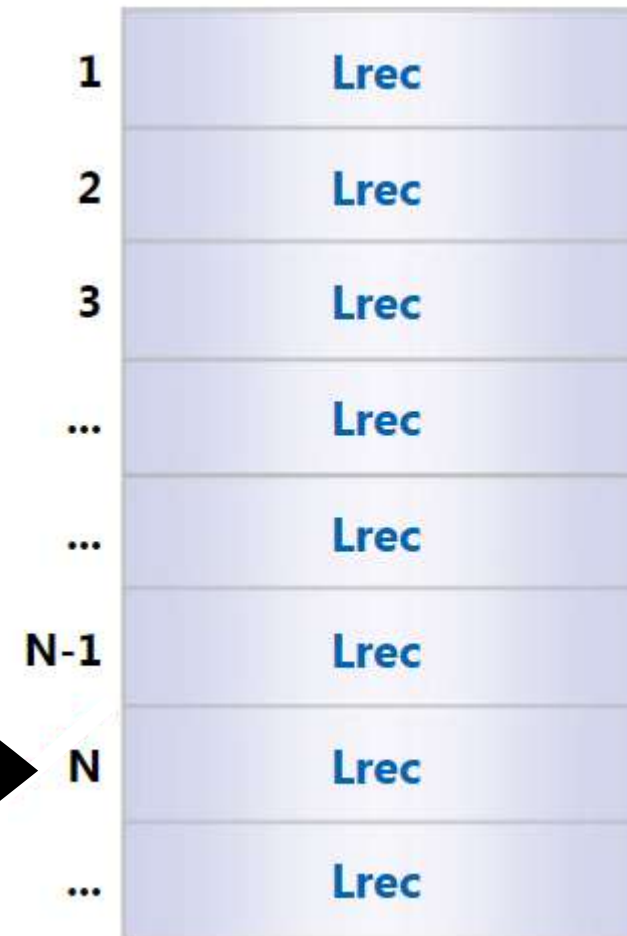


Organizzazione sequenziale

- Organizzazione sequenziale:
 - Semplice da gestire
 - Record di lunghezza diversa
 - Efficace per file di piccole dimensioni
 - Scrittura di un nuovo record solo in coda (si mantiene l'ordine di immissione)
 - Adatto per operazioni batch, meno per quelle interattive

Organizzazione ad accesso diretto

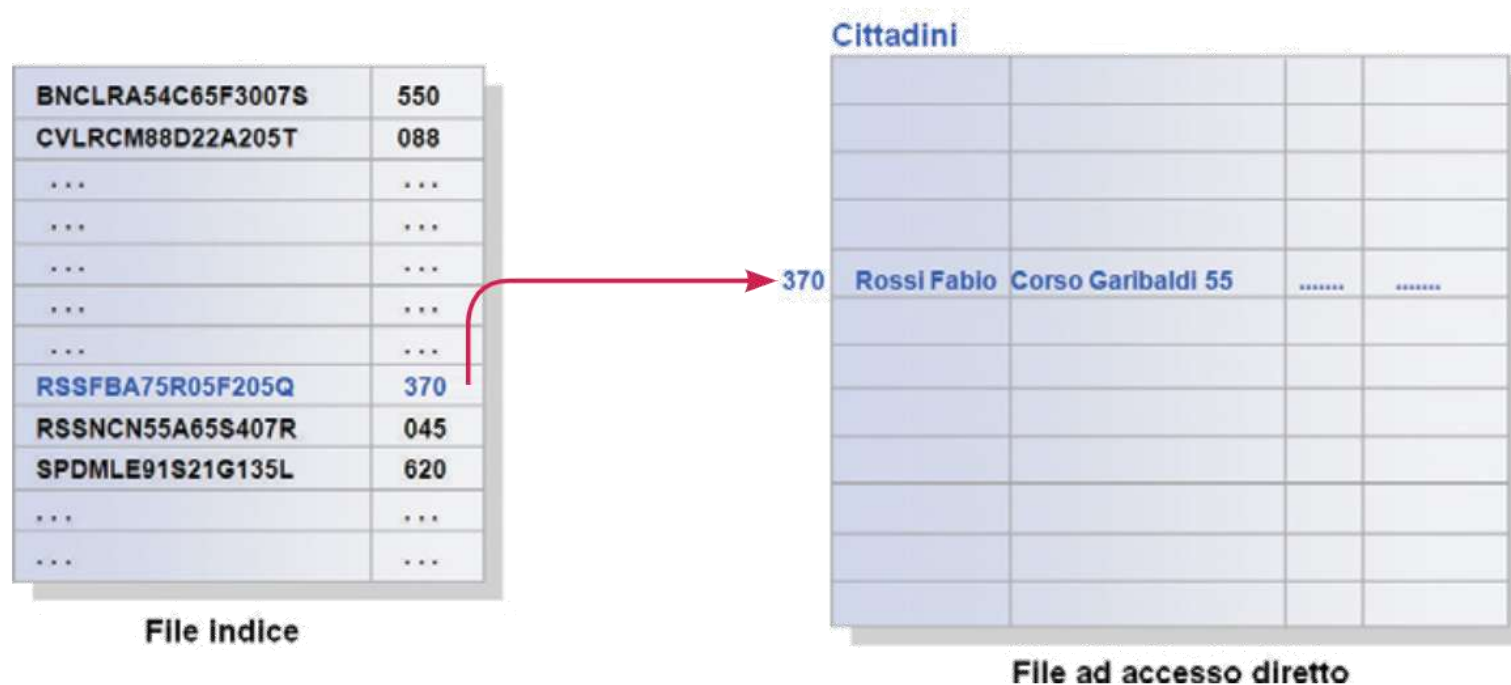
- Con l'organizzazione ad **accesso diretto** i record, registrati uno di seguito all'altro, sono di lunghezza fissa.
 - Si parla di **accesso diretto** o **random**



$$(N - 1) \cdot L_{record} + 1$$

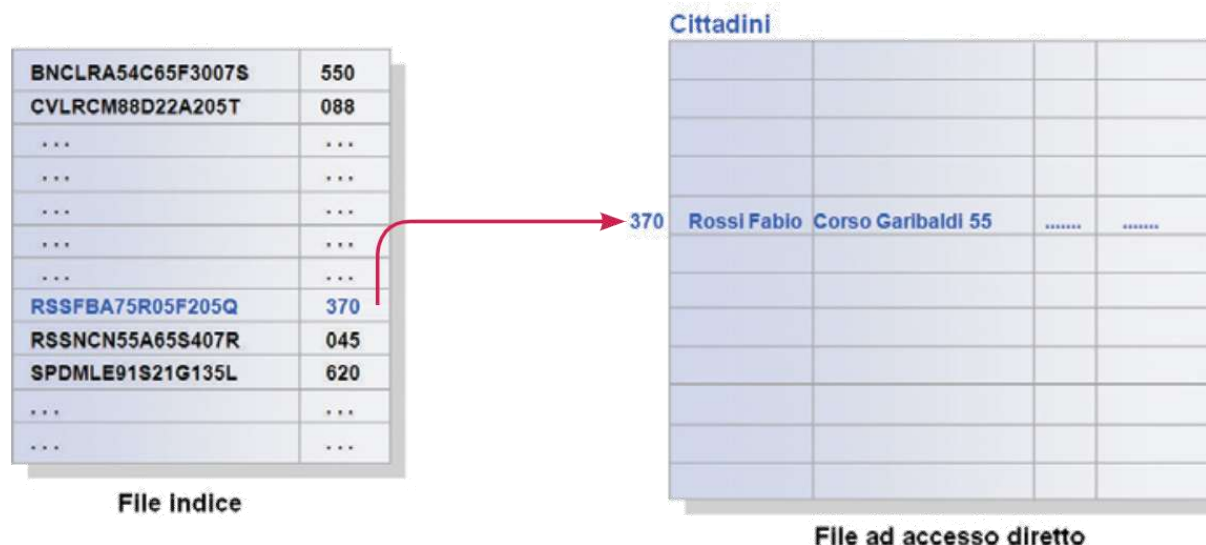
Organizzazione a indici

- L'organizzazione a indici nasce per identificare record non in base alla posizione occupata nel file ma in base a informazioni contenute nel record stesso.



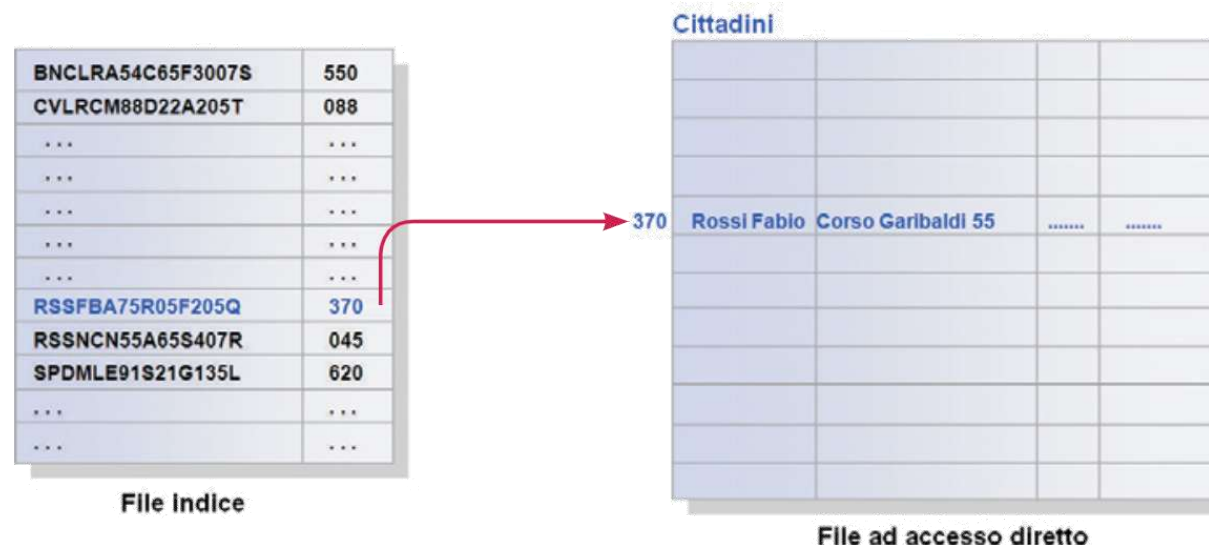
Organizzazione a indici

- Nell'organizzazione sequenziale a indici, sono presenti due elementi:
 - un **archivio primario**, organizzato ad accesso diretto.
 - un **indice** (o dizionario) contenente l'elenco ordinato delle chiavi, a ciascuna chiave è associato un puntatore al record dell'archivio primario.



Organizzazione a indici

- La ricerca di un record avviene con una **ricerca binaria** (o dicotomica) nella tabella delle chiavi per recuperare la posizione del record cercato nel file ad accesso diretto.
- È possibile costruire un indice sia sul campo che identifica univocamente un record (**chiave primaria**) sia su altri campi.



Organizzazione a indici



➤ Svantaggi

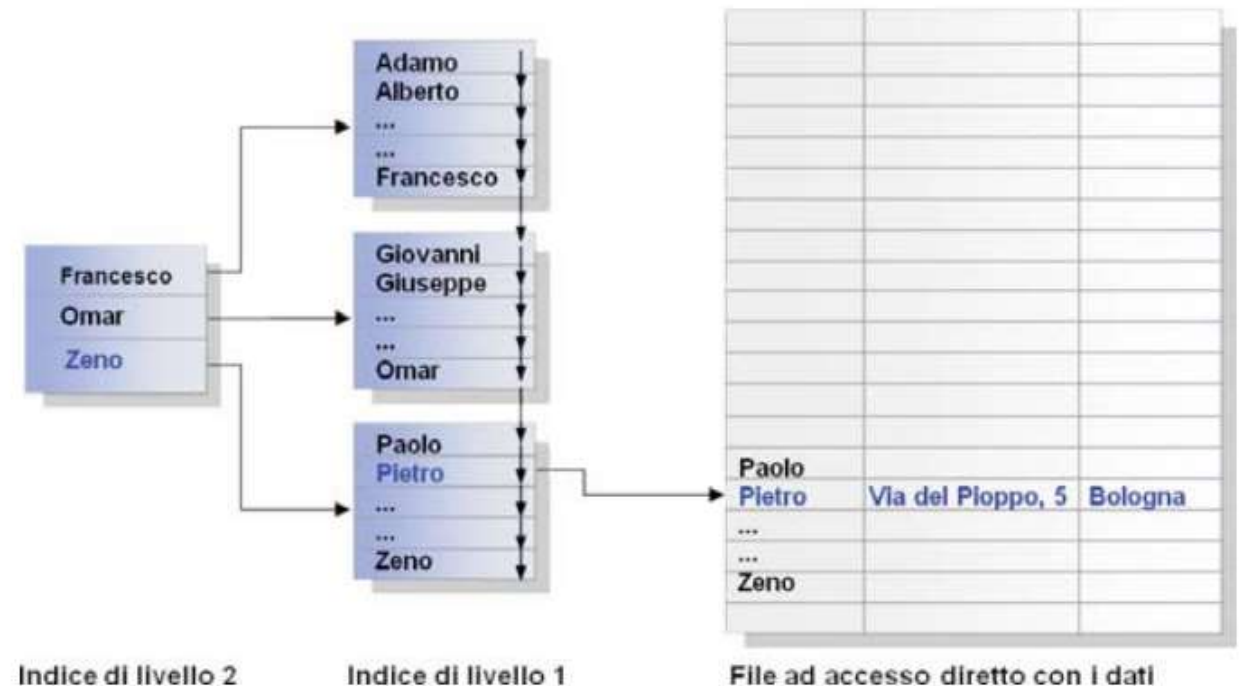
- I dati occupano maggiore **spazio** su disco perché, oltre allo spazio occupato dai record con i dati, bisogna considerare anche lo spazio occupato dal file indice;
- L'inserimento di un nuovo record richiede più **tempo** perché il sistema deve aggiornare anche la struttura dati con i valori dell'indice.

➤ Vantaggi:

- **Miglioramento** complessivo delle **prestazioni** in quanto, in genere, un dato è scritto una sola volta, ma è letto molte volte.

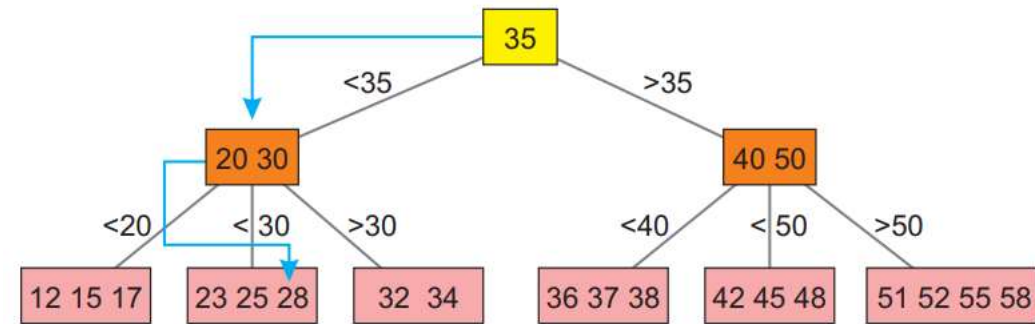
Organizzazione a indici su più livelli

- Nel caso di archivi di grandi dimensioni è possibile migliorare l'efficienza mediante un'organizzazione a indici su più livelli.
- Il file degli indici viene a sua volta indicizzato costruendo un indice (indice di livello 2) per accedere al file degli indici vero e proprio (indice di livello 1).
- Se anche l'indice di livello 2 è troppo grande si può costruire un indice di livello 3 per accedere all'indice di livello 2...



Altri metodi di indicizzazione

► Alberi bilanciati (B-tree)



► Hashing

