

Introduzione

- Elenco telefonico
- Conti correnti
- Catalogo libri di una biblioteca
- Orario dei treni – aerei
-

ESEMPI DI INSIEMI DI DATI DA
ORGANIZZARE ED USARE IN MANIERA
EFFICIENTE

Introduzione

- Più utenti con ognuno il proprio archivio da gestire.
- Riportare le modifiche (aggiunte, cancellazioni, cambiamenti,...) in ogni archivio
- Rischio di:
 - Dimenticanze
 - Controlli incrociati per evitare inconsistenze
 - Sovraccarico (duplicazione) inutile di lavoro
 - Difficoltà nel condurre ricerche sui dati
- La soluzione è un archivio (**base di dati**) centralizzato e condiviso

Che cosa è una Base di dati?

Base di Dati (*data base*): una collezione di informazioni, generalmente non omogenee, organizzate in una struttura che ne rende efficiente la ricerca e l'accesso automatici.

Caratteristiche di un DB:

- la persistenza della memorizzazione dei dati, usualmente suddivisi su più *file* e memorizzati su dispositivi di memoria secondaria di grandi capacità
- funzioni per il *recupero* delle informazioni (*data retrieval*), per l'*introduzione* di nuove informazioni (*data entry*) e per l'*aggiornamento* o modifica di quelle già presenti (*data editing*).

Il DBMS ed i suoi linguaggi

Le funzionalità di un DB sono realizzate da un sistema software chiamato *data base management system* (DBMS), che integra al suo interno più moduli: alcuni destinati ad interagire con gli utenti; altri alla definizione ed al mantenimento della struttura, nonché alla elaborazione delle informazioni; altri ancora al coordinamento con il sistema di gestione dei file ovvero col sistema operativo che ospita il DBMS.

- DDL (*data definition language*), istruzioni che specificano come l'organizzazione concettuale dei dati debba essere fisicamente realizzata
- Il processore delle interrogazioni, *queries*, i.e. il modulo sw che trasforma le interrogazioni dell'utente, espresse in un *linguaggio di interrogazione*, in istruzioni per il recupero dei dati e per l'elaborazione delle risposte

I modelli dei dati

Ve ne sono di diversi tipi:

- *gerarchico*: molto usato per i DB di tipo testuale
- *reticolare*: si interrega navigando; oggi riscoperto come “orientato agli oggetti”
- *relazionale*: basato sul concetto di relazione (**tabella**).

Il modello relazionale

Dati due insiemi A e B (esempio: A=Nomi, B=numeri telefonici) il prodotto cartesiano $A \times B$ di A e B è l'insieme:

$$A \times B = \{ \langle a, b \rangle \mid a \in A, b \in B \}$$

es. $\text{Nomi} \times \text{NumTel} = \{ \langle \text{Filippo}, 011\ 56 \dots \rangle, \langle \text{Caterina}, 095\ 78 \dots \rangle, \dots \}$

Una relazione insiemistica R su A e B è un sottoinsieme di $A \times B$

Un data-base di tipo relazionale è una collezione di relazioni insiemistiche. Queste relazioni si rappresentano come tabelle.

Una tabella

schema

campi

COGNOME	NOME	TEL.
Tanzi	Rosario	095 89 ...
Federici	Susanna	011 55 ...
De Bernardi	Silvio	06 44 ...

record

Esempio di tabella

Esami sostenuti

Materia	Nome	Corso	Data	Voto	Matricola
Inorganica	Filippo	Chimica	23/02/00	24/30	9723456
Lett. Greca	Caterina	Lettere	15/12/99	28/30	8745666
Organica	Filippo	Chimica	08/07/99	22/30	9723456
Lett. Latina	Caterina	Lettere	09/02/00	29/30	8745666
Semiotica	Caterina	Lettere	11/04/00	27/30	8745666

Correlazione tra tabelle

Diverse tabelle possono essere messe in correlazione tra loro mediante la ripetizione di campi in comune

Studente			
Nome	Corso	Matricola	
Filippo	Chimica	9723456	←
Caterina	Lettere	8745666	
...	...		
Esame			
Materia	Data	Voto	Matricola
Inorganica	23/02/00	24/30	9723456 ←
Lett. Greca	15/12/99	28/30	8745666
Organica	08/07/99	22/30	9723456 ←

Vincoli di integrità

I vincoli di integrità sono condizioni che debbono essere soddisfatte **dagli elementi di una tabella** per evitare incoerenze e prevenire errori.

Iscrizioni

Studente	Matricola	Età
Verdi	10345	25
Prodi	76890	230
Salemi	76890	21

valore fuori dominio

replica scorretta

Le chiavi

Una chiave è un insieme di attributi (campi) i cui valori identificano univocamente i record **(le righe di una tabella)**.

Cognome	Nome	Tel
Rossi	Andrea	02 78 ...
Rossi	Paolo	0347 17 ...
Sorbi	Andrea	081 43 ...
Valentini	Paolo	06 85 ...

Una relazione (tabella) può avere più chiavi; una chiave i cui valori non siano mai nulli e quindi individuino tutti i record si dice *primaria*.

Operazioni sulle relazioni (tabelle)

Il recupero dei dati e le operazioni di incrocio con cui un linguaggio di interrogazione permette di rispondere alle richieste dell'utente si basano su operazioni sulle relazioni (tabelle), tra le quali consideriamo:

- la proiezione (π)
- la selezione (σ)
- il join naturale (incrocio)

Proiezione

La proiezione di una tabella è una tabella con un sottoinsieme delle colonne della prima.

Proiezione su A, C

A	B	C
a	b	c
a'	b'	c'

T



A	C
a	c
a'	c'

$\pi_{A,C}(T)$

Selezione

La selezione $\sigma_P(T)$ costruisce una nuova tabella scegliendo i record della tabella T che soddisfano il criterio (predicato) P .

A	B	N
a	b	3
a'	b'	4
a''	b''	9

T



A	B	N
a	b	3
a''	b''	9

$\sigma_P(T)$

$P = N$ multiplo di 3

Criteri non atomici

Un criterio è un predicato sui valori degli attributi (campi):
questo predicato può essere “atomico”:

$$\text{Nome} = \text{“Pippo”}$$

oppure essere composto di parti atomiche collegate con
operatori booleani:

$$\text{Nome} = \text{“Pippo”} \text{ OR } \text{Cognome} = \text{“Rossi”}$$
$$\text{Nome} = \text{“Pippo”} \text{ AND } \text{Cognome} = \text{“Rossi”}$$

Connettivi booleani: tavole di verità

Negazione

A	Not A
Vero	Falso
Falso	Vero

Congiunzione

A	B	A And B
Vero	Vero	Vero
Vero	Falso	Falso
Falso	Vero	Falso
Falso	Falso	Falso

Disgiunzione

A	B	A Or B
Vero	Vero	Vero
Vero	Falso	Vero
Falso	Vero	Vero
Falso	Falso	Falso

Join naturale

Il join naturale di due relazioni (tabelle) che abbiano coordinate (nomi dei campi) in comune è una relazione (tabella) le cui coordinate sono l'unione delle coordinate delle relazioni (tabelle) date, e i cui record sono quelli i cui valori nelle coordinate comuni sono eguali

A	B
a	b
a'	b'

B	C
b	c
b''	c''



A	B	C
a	b	c

Le interrogazioni (query)

Le interrogazioni sono espresse in una forma opportuna:

- attraverso un linguaggio di interrogazione, ad es. SQL:

```
SELECT COGNOME,ANNO  
FROM GIOCATO IN  
WHERE GOL>30 AND SQUADRA = JUVENTUS
```

- attraverso schemi di risposta (*query by example*), come ad esempio in Access.

In entrambi i casi il modulo di esecuzione della *query* compone opportunamente le operazioni proprie del DB su cui lavora (es. proiezione, selezione, join,...) per calcolare la risposta.

Interrogazioni su DB relazionali

Studente			
Nome	Corso	Matricola	
Filippo	Chimica	9723456	
Caterina	Lettere	8745666	
...	...		

Esame			
Materia	Data	Voto	Matricola
Inorganica	23/02/00	24/30	9723456
Lett. Greca	15/12/99	28/30	8745666
Organica	08/07/99	22/30	9723456

SELECT Materia FROM Esame WHERE Matricola IN
SELECT Matricola FROM Studente WHERE Nome = Filippo

↓

Materia
Inorganica
Organica

tabelle

Interrogazione

risultato