

Informatica di Base¹ — Linea 1

Jianyi Lin

Dip. di Matematica e Scienze dell'Informazione
Università degli Studi di Milano, Italia

jianyi.lin@unimi.it

a.a. 2011/12

¹© 2011 J.Lin, M. Monga. Creative Commons Attribuzione-Condividi allo stesso modo 2.5 Italia License.
<http://creativecommons.org/licenses/by-sa/2.5/it/>. Rielaborazione del materiale © 2009/10 © S. Mascetti.

Lezione XIII: Basi di dati

Dati in forma tabellare

Conservare i dati in forma tabellare è molto comune ed intuitivo: ciò spiega il successo dei fogli di calcolo.

Nei fogli di calcolo, però:

- Ogni cella è trattata singolarmente
- Il formato dei dati è descritto “per cella”

Per organizzare grandi insiemi di dati servono forme più strutturate.

Le basi di dati

Le **basi di dati** (o **database**, DB) servono per organizzare, memorizzare e facilitare l'elaborazione automatica dell'informazione.

Il tipo di basi di dati piú comune è quello basato su dati in forma tabellare (**basi di dati relazionali**, proposte per la prima volta da E. Codd negli anni '70).

Un DB è gestito tramite un "Database management system" (DBMS).

Basi di dati relazionali

Nei DB relazionali si conservano:

Entità

Le **entità** denotano i dati da memorizzare: una persona, un'automobile, una lettera a Babbo Natale, ecc.

Attributi

Ogni entità è caratterizzata dagli **attributi** che si vogliono conservare: per una persona potrebbero essere il nome, il cognome, il codice fiscale, ecc.

Inoltre le entità possono essere messe in **relazione** fra loro (p.es. Una persona possiede un'automobile): per capire come dobbiamo esaminare meglio come vengono memorizzate le entità.

Esempio

Entità automobile:

modello "Fiat Panda NP"

colore "Giallo"

targa "AB768BA"

immatricolazione "1/4/2003"

passengeri 4

Ogni automobile registrata nel database è detta **record** del DB. In forma tabellare gli attributi sarebbero probabilmente le colonne, e ogni record una riga.

Gli attributi sono *atomici*: se volessi considerare il mese e il giorno di immatricolazione separatamente dovrei avere altri attributi.

Dati e metadati

La struttura di ogni record è predefinita e fissa: è decisa dal progettista del DB in funzione dell'uso dei dati previsto.

- Ogni record ha un insieme fisso (tupla) di attributi:
automobile \rightsquigarrow \langle modello, colore, targa, immatricolazione, passeggeri \rangle
- Per ogni attributo è definito il *dominio*:
modello \rightsquigarrow stringa di max 30 caratteri, targa \rightsquigarrow stringa di esattamente 7 caratteri, passeggeri \rightsquigarrow numero intero positivo

Queste sono *informazioni sui dati*: **metadati**.

Esemplari di dati

I metadati definiscono il *tipo* di dato che verrà conservato nel database. Ciascun record dovrà essere un **esemplare** conforme al tipo definito dai metadati.

In gergo si dice **istanza** (*instance*) del tipo definito dall'entità.
Es: il record della Panda è un'*istanza* dell'entità automobile.

Chiave

Esiste un insieme di attributi in grado di identificare univocamente un record?

Se c'è viene detto **chiave candidata** e normalmente sono più d'uno.

Se ne sceglie uno: **chiave primaria**. Generalmente è opportuno scegliere insiemi di attributi il più piccoli possibile, spesso con un solo attributo.

Nel caso dell'automobile potrebbe essere la targa.

Relazione

Una **relazione** può essere memorizzata con una tabella che associ le chiavi primarie delle entità coinvolte:

- Entità automobile (chiave: targa)
- Entità persona (chiave: codice-fiscale)
- Relazione appartiene-a \langle automobile.targa, persona.codice-fiscale \rangle

Generalmente si hanno relazioni *binarie* e possono esserci relazioni 1 a 1, 1 a molti, molti a molti.

Operazioni su dati tabellari

Un insieme di tre semplici operazioni è sufficiente per estrarre i dati da un DB senza limitazioni ai criteri di estrazione.

selezione Un sottoinsieme di righe che rispettano un determinato criterio

proiezione Un sottoinsieme di colonne

prodotto cartesiano La tabella ottenuta combinando tutte le righe della prima con tutte le righe della seconda

Selezione

Superficie	Piano	Stanze	Indirizzo	Box
100	4	4	Via Rossi, 5	No
50	4	2	Via Verdi, 4	Sì
150	8	7	Via Bianchi, 9	Sì

criterio di selezione: Piano = 4

Superficie	Piano	Stanze	Indirizzo	Box
100	4	4	Via Rossi, 5	No
50	4	2	Via Verdi, 4	Sì

I criteri di selezione possono essere combinati con le solite funzioni logiche

not trasforma una condizione vera in una falsa, e viceversa

and vera solo se tutte le condizioni sono vere

or vera se almeno una condizione è vera

not (Piano = 8) and ((stanze > 2) or (superficie \geq 75))

Proiezione

Superficie	Piano	Stanze	Indirizzo	Box
100	4	4	Via Rossi, 5	No
50	4	2	Via Verdi, 4	Sì
150	8	7	Via Bianchi, 9	Sì

proiezione: indirizzo, box

Indirizzo	Box
Via Rossi, 5	No
Via Verdi, 4	Sì
Via Bianchi, 9	Sì

Prodotto cartesiano

Si applica a due tabelle ed è analogo al prodotto cartesiano fra insiemi

$$\{a,b\} \times \{c,d,e\} = \{\{a,c\},\{a,d\},\{a,e\},\{b,c\},\{b,d\},\{b,e\}\}$$

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Bambola	rosa	Femmina
Anna	4	Femmina	Macchinina	blu	Maschio

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Andrea	2	Maschio	Bambola	rosa	Femmina
Luca	3	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Bambola	rosa	Femmina
Anna	4	Femmina	Macchinina	blu	Maschio
Anna	4	Femmina	Bambola	rosa	Femmina

Join

È frequentissimo fare il prodotto e poi una selezione.

join naturale

Si moltiplicano due tabelle A e B e si selezionano solo le righe per le quali un attributo (specificato) di A ha valore uguale ad un attributo (specificato) di B

Join (cont.)

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Andrea	2	Maschio	Bambola	rosa	Femmina
Luca	3	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Bambola	rosa	Femmina
Anna	4	Femmina	Macchinina	blu	Maschio
Anna	4	Femmina	Bambola	rosa	Femmina

join con Bambini.genere = Regalo.GiocoDa

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Macchinina	blu	Maschio
Anna	4	Femmina	Bambola	rosa	Femmina

Informatica di base – Linea 1