

ALGEBRA RELAZIONALE

Formalizzazione del modello logico Relazionale

Definizioni

Per introdurre il modello matematico relazionale è necessario ricordare alcune definizioni: insieme, prodotto cartesiano, relazione, tabella, n-uple, t-uple, attributo, dominio.

Insieme

una collezione di oggetti chiamati elementi dell'insieme.

Proprietà

- un elemento può appartenere o non appartenere a un determinato insieme, non ci sono vie di mezzo;
- un elemento non può comparire più di una volta in un insieme;
- gli elementi di un insieme non hanno un ordine di comparizione (l'insieme non è ordinato);
- gli elementi di un insieme lo caratterizzano univocamente: due insiemi coincidono se e solo se hanno gli stessi elementi.

Prodotto cartesiano

Il prodotto cartesiano di due insiemi A e B (si scrive $A \times B$) è l'insieme costituito dalle coppie ordinate (a,b), dove a è un elemento di A e b un elemento di B. Le coppie sono ordinate, cioè la coppia (a,b) è diversa dalla coppia (b,a), perciò il prodotto $A \times B$ è diverso dal prodotto $B \times A$.

Esempio: dati gli insiemi $A = \{\text{Rossi,Verdi}\}$ e $B = \{1980, 1990, 1991\}$

il loro Prodotto cartesiano vale $A \times B = \{(\text{Rossi},1980), (\text{Rossi},1990), (\text{Rossi},1991), (\text{Verdi},1980), (\text{Verdi},1990), (\text{Verdi},1991)\}$

Notare che l'insieme A è di 2 elementi, B di 3 elementi, $A \times B$ di 6 elementi. L'ordine con cui sono disposti gli elementi all'interno di un insieme non è importante, ma lo è invece, quello con cui sono disposti gli elementi all'interno delle parentesi tonde.

In generale il prodotto cartesiano può riguardare anche più di *due* insiemi (A, B, C, D,...), e al posto delle *coppie* si possono avere triple, quadruple, e in generale *n-uple* (pronunciato enn-uple).

Relazione (matematica)

Si può definire la Relazione matematica (o più brevemente Relazione) R su A e B indicata da R(A, B) un sottoinsieme dell'insieme prodotto cartesiano A x B.

Continuando il precedente esempio, una possibile relazione su A e B potrebbe essere costituita da

$$R(A, B) = \{(Rossi,1990), (Rossi,1991), (Verdi,1980) \}$$

Mentre in matematica per il Prodotto cartesiano, l'ordine degli elementi delle coppie (a, b) è importante, nelle relazioni studiate nei database, l'ordine (Rossi, 1990) è meno importante. Per questo motivo, al posto del termine *n-uple* viene usato anche il termine *t-uple*. Importante: vista la natura insiemistica della relazione, e visto anche il precedente esempio, si può dedurre che non ci possono essere due n-uple (o t-uple) identiche nella stessa relazione matematica; nonostante ciò, nella pratica dei DBMS, questa regola è lasciata libera, a discrezione del programmatore.

Tabella

È una delle possibili rappresentazioni in cui può essere espressa una Relazione.

La precedente Relazione R su A e B potrebbe essere espressa dalla seguente tabella:

R(A, B)

A	B
Rossi	1990
Rossi	1991
Verdi	1980

Purtroppo tutti i DBMS utilizzano il termine *tabella* in luogo del termine *relazione*; altri DMBS ad interfaccia grafica, addirittura, usano il termine *relazione* per indicare i collegamenti tra le chiavi primarie e le chiavi esterne: queste convenzioni sono terribili, ma conoscendole si evita almeno di fare confusione.

Attributi

Sono gli insiemi che vengono utilizzati per costruire la Relazione, come A e B , oppure Cognome, Nome, ecc.

Dominio

È l'insieme dei valori che può assumere un attributo: per esempio A è un insieme di nomi (stringhe), B è un insieme di numeri interi. Importante: ogni attributo ha un proprio dominio di valori, che non può essere modificato.

Vincoli intrarelazionali

I vincoli intrarelazionali sono regole che descrivono la struttura dei dati all'interno di una singola relazione (rappresentabile da una singola tabella). Si distinguono nei seguenti tipi:

- vincolo di unicità: impedisce che all'interno di un attributo sia inserito due volte lo stesso valore;
- vincolo di obbligatorio: impedisce che all'interno di un attributo sia lasciato un valore non inserito;
- vincolo di chiave primaria: oltre ad unire insieme i due precedenti vincoli (unicità ed obbligatorio) consente al DBMS di utilizzare l'attributo indicato come elemento chiave di accesso ai dati di ogni tupla.
- vincolo di dominio: indica se nel dominio dei valori di un attributo della relazione ce ne siano alcuni che non possono essere assunti; per esempio, si potrebbe desiderare che l'utente non possa inserire (anche per distrazione) valori negativi all'interno di una colonna chiamata "prezzo";
- vincolo di tupla: fissando una tupla, indica se nel dominio dei valori di un attributo della relazione ce ne siano alcuni che non possano essere assunti, a seconda del valore attualmente presente in un altro attributo; per esempio, si potrebbe desiderare che il prezzo di vendita di un prodotto sia sempre superiore al suo prezzo di acquisto;

L'unico vincolo intrarelazionale che viene espresso in forma grafica è quello di chiave primaria, utilizzando convenzionalmente la sottolineatura; gli altri tipi di vincolo devono essere espressi dal progettista a parole (legenda).

Vincoli extrarelazionali

Si tratta di vincoli che esprimono regole del modello logico relazionale che riguardano almeno due relazioni, in particolare studieremo solo un vincolo di questo tipo:

vincolo di **integrità referenziale**:

si DEVE applicare ad ogni chiave esterna di una relazione per impedire che quella chiave esterna assuma valori non presenti nella chiave primaria a cui essa fa riferimento.

Per rappresentare graficamente questo vincolo è possibile, oltre a indicare l'entità da cui deriva come esponente della chiave esterna, realizzare una freccia che collega la chiave esterna ad una chiave primaria.

Operazioni insiemistiche

Dopo aver compreso che cosa sia una relazione matematica, si può comprendere meglio quali siano le possibili operazioni che si possono operare tra di esse. Siccome le relazioni non sono altro che dei particolari insiemi, tra le operazioni relazionali, ricadono anche le classiche operazioni insiemistiche di X *prodotto cartesiano*, \cup *unione*, \cap *intersezione* e $/$ *differenza*.

Ogni operazione restituisce come risultato una relazione.

Unione

- L'unione di due relazioni R ed S, indicata con
$$R \cup S$$
è l'insieme delle tuple in R, in S o in entrambe
- L'unione di due relazioni può essere fatta *solo se hanno lo stesso grado*; inoltre il primo attributo di R deve essere *compatibile* con il primo attributo di S, il secondo attributo di R deve essere *compatibile* con il secondo attributo di S, e così via
- Se le relazioni hanno nomi di attributo diversi, nella relazione risultato per convenzione si usano i nomi della prima relazione (in questo caso R), a meno di opportune ridenominazioni
- Le tuple duplicate vengono eliminate
- Il grado della relazione risultato è uguale al grado delle relazioni operandi

A	B	C
a	b	c
d	a	f
c	b	d

R

D	E	F
b	g	a
d	a	f

S

A	B	C
a	b	c
d	a	f
c	b	d
b	g	a

$R \cup S$

Intersezione

- Il prodotto cartesiano di due relazioni R ed S, di grado k_1 e k_2 rispettivamente, indicato con

$$R \times S$$

è una relazione di grado k_1+k_2 le cui tuple sono tutte le tuple che hanno:

- come prime k_1 componenti le tuple di R
- come seconde k_2 componenti le tuple di S

- Nella relazione risultato i nomi dei primi k_1 attributi sono i nomi degli attributi della relazione R e i nomi degli ultimi k_2 attributi sono i nomi degli attributi della relazione S
- Se le due relazioni hanno attributi con lo stesso nome, è necessario ridenominare gli attributi in una delle due relazioni

A	B	C
a	b	c
d	a	f
c	b	d

R

D	E	F
b	g	a
d	a	f

S

A	B	C	D	E	F
a	b	c	b	g	a
a	b	c	d	a	f
d	a	f	b	g	a
d	a	f	d	a	f
c	b	d	b	g	a
c	b	d	d	a	f

$R \times S$

Differenza

- La differenza di due relazioni R ed S, indicata con

$$R - S$$

è l'insieme delle tuple che sono in R ma non in S

- La differenza (come l'unione) può essere eseguita solo se le relazioni hanno lo stesso grado e gli attributi sono compatibili

- In caso di attributi con nomi diversi, nella relazione risultato, per convenzione, si usano i nomi della prima relazione, a meno di opportune ridenominazioni
- Il grado della relazione risultato è uguale al grado delle relazioni operandi

A	B	C
a	b	c
d	a	f
c	b	d

R

D	E	F
b	g	a
d	a	f

S

A	B	C
a	b	c
c	b	d

R-S

Operazioni relazionali

Per descrivere le altre operazioni relazionali, si utilizzerà la soluzione dell'esercizio svolto su autoveicoli e cittadini:

CITTADINI

cf	cognome	nome
RSSMRA	Rossi	Mario
VRDGPP	Verdi	Giuseppe
BNCNNA	Bianchi	Anna
RSSMRC	Rossi	Marco
NRENRC	Neri	Enrico
GLLGNN	Gialli	Gianna

Selezione

Questa operazione si applica ad una sola relazione e produce come risultato una nuova relazione che ha lo stesso grado della relazione di partenza, ma solitamente cardinalità inferiore. In pratica si può ottenere una nuova tabella che contenga solo alcune delle righe contenute nella tabella di partenza. Questa operazione si indica anche con il simbolo della lettera greca minuscola *sigma*, seguito tra parentesi dal nome della relazione, e avente come pedice la condizione di selezione che deve essere verificata.

Ad esempio, per visualizzare dalla tabella cittadini, solo quelli di cognome 'Rossi' si scriverà:

$\sigma_{\text{cognome}='Rossi'}(\text{cittadini})$

cf	cognome	nome
RSSMRA	Rossi	Mario
RSSMRC	Rossi	Marco

Proiezione

Questa operazione si applica ad una sola relazione e produce come risultato una nuova relazione che ha la stessa cardinalità della relazione di partenza, ma solitamente grado inferiore. In pratica si può ottenere una nuova tabella che contenga solo alcune delle colonne contenute nella tabella di partenza. Questa operazione si indica anche con il simbolo della lettera greca minuscola *pi*, seguita, tra parentesi, dal nome della relazione, e avente come pedice la condizione di proiezione che deve essere verificata.

A	B	C
a	b	c
d	a	f
c	b	d

R

A	C
a	c
d	f
c	d

$\Pi_{A,C}(R)$

B	A
b	a
a	d
b	c

$\Pi_{B,A}(R)$

Ad esempio, per visualizzare dalla tabella cittadini, solo le informazioni contenute in nome, si scriverà:

$\pi_{\text{nome}}(\text{cittadini})$

nome
Mario
Giuseppe
Anna
Marco
Enrico
Gianna

Congiunzione

Questa operazione si applica a due relazioni e produce come risultato una nuova relazione che ha solitamente grado superiore a quello delle singole relazioni di partenza, e nulla si può affermare

sulla cardinalità. Le due tabelle di partenza DEVONO avere almeno una colonna contenente lo stesso tipo di dati. In pratica si può ottenere una nuova tabella possega le colonne di entrambe le tabelle di partenza e quelle righe che hanno un valore in comune ad entrambe le tabelle. Questa operazione si indica anche con il simbolo della lettera minuscola X, chiuso ai lati destro e sinistro. Il simbolo deve essere preceduto e seguito dai nomi delle due relazioni coinvolte e deve avere come pedice la condizione di congiunzione che deve essere verificata.

Ad esempio, visto che la tabella 'cittadini' e la tabella 'proprietà' hanno una colonna in comune (il CF del cittadino), potrebbero essere combinate insieme, scrivendo:

proprietà $\otimes_{cf=cf-cittadini}$ cittadini

cf-cittadino	targa-auto	data-acquisto	cognome	nome
VRDGPP	AZ12345	2005-12-31	Verdi	Giuseppe
RSSMRA	AZ12345	2005-12-31	Rossi	Mario
GLLGNN	OIL9876	2001-11-23	Gialli	Gianna
GLLGNN	EE33333	2004-02-29	Gialli	Gianna

NB: non si confonda grado con ordine.

Modello relazionale nei DBMS

Tenendo in mente il concetto di relazionale matematica e le definizioni di operazioni relazionali, va osservato che il modello logico utilizzabile in un DBMS non ha una corrispondenza esatta. Alcuni dei vincoli potrebbero essere presenti, altri no, mentre ne potrebbero esistere alcuni che qui non sono stati studiati e che quindi non appartengono a questo modello. Lo stesso discorso vale per le operazioni relazionali e diverrà più chiaro quando si studierà il linguaggio SQL.