LINGUAGGI RELAZIONALI DI MANIPOLAZIONE DATI

Data una base di dati relazionale BD = $\{r_1(R_1), r_2(R_2),..., r_m(R_m)\}$ ($r_j(R_j)$ = relazione definita sullo schema R_j)

libro

Cod-libro	Autore	Titolo	Prezzo
1	Umberto Eco	In nome della Rosa	37.000
4	Herman Hesse	Il giuoco delle perle di vetro	15.000
3	Carmelo Bene	Sono apparso alla Madonna	23.700

vendita

Venditore	Libro	Quantità
Rossi	1	5.000
Bianchi	3	1.000
Rossi	4	3.000

***AGGIORNAMENTO : funzione che produce un una nuova istanza di base di dati BD' = $\{r_1'(R_1), r_2'(R_2),..., r_m'(R_m)\}$ sullo stesso schema $\{R_1, R_2,..., R_m\}$**

llibro

Cod-libro	Autore	Titolo	Prezzo
1	Umberto Eco	In nome della Rosa	37.000
4	Herman Hesse	Il giuoco delle perle di vetro	15.000
3	Carmelo Bene	Sono apparso alla Madonna	23.700
5	Daniel Pennac	La prosivendola	20.000

vendita

Venditore	Libro	Quantità
Rossi	1	5.000
Bianchi	3	10
Rossi	4	3.000

*** INTERROGAZIONE : funzione che produce una relazione su un dato schema**

Quali sono i nomi e le quantità dei libri venduti da Rossi?

Titolo	Quantità
In nome della Rosa	5.000
Il giuoco delle perle di vetro	3.000

LINGUAGGI RELAZIONALI DI INTERROGAZIONE

PROCEDURALI

ALGEBRICI : ogni interrogazione è una espressione composta a partire da operatori che, applicati nell'ordine fissato dalla composizione, calcolano la relazione risultante

*** Algebra relazionale**

DICHIARATIVI

BASATI SUL CALCOLO DEI PREDICATI : ogni interrogazione è una espressione composta a partire da predicati che descrivono la relazione risultante

*** Calcolo relazionale su domini**
*** Calcolo relazionale su tuple**

ALGEBRA RELAZIONALE

* OPERATORI SU INSIEMI

R(A, B, C), S(A, B, C), Dom(A) = {a1,a2}, Dom(B) = {b1,b2}, Dom(C) = {c1,c2}
r(R), s(S)

r

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2

s

A	B	C
a1	b1	c1
a2	b2	c2

UNIONE: $r \cup s = \{t \mid t \in r \text{ or } t \in s\}$

$r \cup s$

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2
a2	b2	c2

INTERSEZIONE: $r \cap s = \{t \mid t \in r \text{ AND } t \in s\}$

$r \cap s$

A	B	C
a1	b1	c1

DIFFERENZA: $r - s = \{t \mid t \in r \text{ AND } t \notin s\}$

$r - s$

A	B	C
a1	b2	c1
a2	b1	c2

ALGEBRA RELAZIONALE

* OPERATORI SU INSIEMI

R(A, B, C), Q(D, E), Dom(A) = {a1,a2}, Dom(B) = {b1,b2}, Dom(C) = {c1,c2}, Dom(D) = {d1}, Dom(E) = {e1,e2}, r(R), q(Q)

r

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2

q

D	E
d1	e1
d1	e2

PRODOTTO CARTESIANO: $r \times s = \{t_r t_s \mid t_r \in r \text{ AND } t_s \in s\}$

$r \times s$

A	B	C	D	E
a1	b1	c1	d1	e1
a1	b1	c1	d1	e2
a1	b2	c1	d1	e1
a1	b2	c1	d1	e2
a2	b1	c2	d1	e1
a2	b1	c2	d1	e2

COMPLEMENTO: $\sim r \{t \mid t \in T - r\}$ ($T = A \times B \times C$)

$\sim r$

A	B	C
a1	b1	c2
a1	b2	c2
a2	b1	c1
a2	b2	c1
a2	b2	c2

ALGEBRA RELAZIONALE

SELEZIONE: un sottoinsieme delle n-ple di una relazione r,

La "condizione di selezione" del sottoinsieme è una proposizione che coinvolge attributi

Def. una *formula proposizionale* F è

- **una formula atomica del tipo Aθ B o Aθ c, dove A e B sono nomi di variabili definite sullo stesso dominio, c è un valore costante del dominio di A e θ è un operatore di confronto ($=, \neq, \leq, <, >, \geq$);**
- **una combinazione di formule atomiche mediante i connettivi logici (AND, OR, NOT)**

Def. Dato un insieme di valori s su un dominio

- **s soddisfa una formula proposizionale atomica F sse F(s) = TRUE**
- **s soddisfa NOT F sse s non soddisfa F**
- **s soddisfa F OR V sse s soddisfa F o s soddisfa V**
- **s soddisfa F AND V sse s soddisfa F e s soddisfa V**

SELEZIONE su r(R) RISPETTO A UNA PROPOSIZIONE F:

$$\sigma_F(r) = \{t \in r \mid t \text{ soddisfa } F\}$$

R(A, B, C), Dom(A) = {a1,a2}, Dom(B) = {b1,b2}, Dom(C) = {c1,c2}, r(R)

r

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2

$\sigma_{A = a2}(r)$

A	B	C
a2	b1	c2

Esempi di selezione

impiegati

Cognome	Nome	Età	Stipendio
Rossi	Mario	25	2.000.000
Neri	Luca	40	3.000.000
Verdi	Nicola	36	4.500.000
Rossi	Marco	40	3.900.000

$\sigma_{(\text{Età} < 30) \text{ OR } (\text{Stipendio} > 4.000.000)}$ (impiegati)

Cognome	Nome	Età	Stipendio
Rossi	Mario	25	2.000.000
Verdi	Nicola	36	4.500.000

cittadini

Cognome	Nome	CittàDi Nascita	Residenza
Rossi	Mario	Roma	Milano
Neri	Luca	Roma	Roma
Verdi	Nicola	Firenze	Firenze
Rossi	Marco	Napoli	Firenze

$\sigma_{\text{CittàDiNascita} = \text{Residenza}}$ (cittadini)

Cognome	Nome	CittàDi Nascita	Residenza
Neri	Luca	Roma	Roma
Verdi	Nicola	Firenze	Firenze

ALGEBRA RELAZIONALE

PROIEZIONE: relazione definita su un sottoinsieme di attributi della relazione originaria.

$$\pi_Y(r) = \{t[Y] \mid t \in r\},$$

R(A, B, C), Dom(A) = {a1,a2}, Dom(B) = {b1,b2}, Dom(C) = {c1,c2}, r(R)

r

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2

$\pi_{A,B}$ (r)

A	B
a1	b1
a1	b2
a2	b1

cittadini

Cognome	Nome	CittàDi Nascita	Residenza
Rossi	Mario	Roma	Milano
Neri	Luca	Roma	Roma
Verdi	Nicola	Firenze	Firenze
Rossi	Marco	Napoli	Firenze

$\pi_{\text{Cognome, Nome}}$ (cittadini)

Cognome	Nome
Rossi	Mario
Neri	Luca
Verdi	Nicola
Rossi	Marco

π_{Cognome}(cittadini)

Cognome
Rossi
Neri
Verdi

In generale: $|\pi_Y(r)| = |r|$ sse Y è superchiave per r.

ALGEBRA RELAZIONALE

JOIN NATURALE: correla dati in relazioni diverse, in base a valori comuni.

X1, X2 insiemi di attributi, R(X1), S(X2) schemi di relazione, r(R), s(S) relazioni

$$r \bowtie s = p \text{ su } P(X1 \cup X2) = \{t \text{ su } X1 \cup X2 \mid t[X1] \in r \text{ e } t[X2] \in s\},$$

R(A, B, C), S(C,D, E), r(R), s(S), Dom(A) = {a1,a2,a3}, Dom(B) = {b1,b2,b3}, Dom(C) = {c1,c2,c3}, Dom(D) = {d1,d2}, Dom(E) = {e1,e2},

r

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3

s

C	D	E
c1	d1	e1
c2	d2	e2

$r \bowtie s$

A	B	C	D	E
a1	b1	c1	d1	e1
a2	b2	c2	d2	e2

Proprietà del join naturale: r, s, p relazioni sugli schemi R(X1), S(X2), P(X3)

* **proprietà commutativa:** $r \bowtie s = s \bowtie r$
* **proprietà associativa:** $(r \bowtie s) \bowtie p = r \bowtie (s \bowtie p) = r \bowtie s \bowtie p$
* **grado di** $r \bowtie s \leq$ **grado di r + grado di s**
* $| r \bowtie s | \leq | r | * | s|$;

RELAZIONI TRA JOIN NATURALE E ALTRI OPERATORI

* JOIN E PRODOTTO CARTESIANO

r(R), s(S), R(X1), S(X2)

r ⋈ s = r × s sse X1 ∩ X2 = ∅

r

A	B	C
a1	b1	c1
a2	b2	c2

s

D	E
d1	e1
d1	e2

r ⋈ s

A	B	C	D	E
a1	b1	c1	d1	e1
a1	b1	c1	d1	e2
a2	b2	c2	d1	e1
a2	b2	c2	d1	e2

In questo caso:
*** grado di r ⋈ s = grado di r + grado di s**
*** | r ⋈ s | = | r | * | s|;**

* JOIN E INTERSEZIONE

r(R), r'(R), R(X1), (due relazioni sullo stesso insieme di attributi)

r ⋈ r' = r ∩ r'

r

A	B	C
a1	b1	c1
a2	b2	c2

r'

A	B	C
a1	b1	c1
a2	b1	c1

r ⋈ r'

A	B	C
a1	b1	c1

RELAZIONI TRA JOIN NATURALE E PROIEZIONE

r(R), s(S), q(Q), R(X1), S(X2), Q(X1$\cup$X2), X = X1 $\cap$ X2

*** Detta p = r $\bowtie$ s, allora in generale $\pi_{X1}(p) \subseteq r$, $\pi_{X2}(p) \subseteq s$.**

$\pi_{X1}(p) = r$ sse per ogni t $\in$ r esiste t' $\in$ s tale che t[X] = t'[X]

$\pi_{X1}(p) = s$ sse per ogni t' $\in$ s esiste t $\in$ r tale che t[X] = t'[X]

Se $\pi_{X1}(p) = r$ e $\pi_{X1}(p) = s$ il join si dice *completo*

*** Dette r = $\pi_{X1}(q)$, s = $\pi_{X2}(q)$ allora in generale q $\subseteq$ r $\bowtie$ s.**
Se q = r $\bowtie$ s si dice che r ed s sono una *decomposizione senza perdita* di q

r

A	B	C
a1	b1	c1
a2	b1	c2

s

C	E
c1	e1
c1	e2

r $\bowtie$ s

A	B	C	E
a1	b1	c1	e1
a1	b1	c1	e2

$\pi_{A,B}$ (r)

A	B
a1	b1
a2	b1

$\pi_{B,C}$ (r)

B	C
b1	c1
b1	c2

$\pi_{A,B}$ (r) $\bowtie$ $\pi_{B,C}$ (r)

A	B	C
a1	b1	c1
a1	b1	c2
a2	b1	c1
a2	b1	c2

ϑ-JOIN

impiegati

Cognome	Progetto
Rossi	A
Neri	A
Neri	B

progetti

Codice	Nome
A	Venere
B	Marte

Gli attributi Progetto e codice portano la stessa informazione (hanno la stessa semantica), ma hanno nomi diversi: se faccio il join naturale

impiegati $\bowtie$ progetti

Impiegato	Progetto	Codice	Nome
Rossi	A	A	Venere
Neri	A	A	Venere
Neri	B	A	Venere
Rossi	A	B	Marte
Neri	A	B	Marte
Neri	B	B	Marte

Non ottengo il risultato voluto

SOLUZIONE: Estensione del join naturale a confronti tra attributi con nome diverso (ϑ = formula proposizionale)

impiegati $\bowtie_{Progetto=Codice}$ progetti

Impiegato	Progetto	Codice	Nome
Rossi	A	A	Venere
Neri	A	A	Venere
Neri	B	B	Marte

ϑ-join come operatore derivato:

$$r \bowtie_{\vartheta} s = \sigma_{\vartheta} (r \times s)$$

Caso particolare: equi-join: ϑ è l'operatore di uguaglianza (=)

DEFINIZIONE DI ALGEBRA RELAZIONALE

L'algebra relazionale è una 6-pla

$$\mathbf{AR = (A, D, dom, \vartheta, O)}$$

- A = {A_1, A_2, ..., A_n} insieme di attributi;

- D insieme di domini;

- dom : A $\rightarrow$ D funzione totale da A a D;

- s = {r_1, r_2, ..., r_k} insieme delle relazioni su tutti i possibili schemi;

- θ = {=,$\neq$, $\leq$, <, >, $\geq$} insieme degli operatori di confronto;

-O = {$\cup$, $\cap$, ~, -, $\times$, σ, π, ρ, $\bowtie$, $\bowtie_\vartheta$} insieme degli operatori relazionali.

Una *espressione algebrica* E è ogni espressione formata <u>correttamente</u> a partire da
- operatori appartenenti ad O,
- relazioni appartenenti a s,
- relazioni costanti

Il *valore* di una espressione algebrica E, indicato con E(u), u $\in$ s è la <u>relazione</u> risultante dall'applicazione degli operatori in E alle relazioni u considerate.

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

E1, E2 espressioni dell'algebra relazionale

*** EQUIVALENZA DIPENDENTE DALLO SCHEMA**

$E1 \approx_R E2$ se $E1(r) = E2(r)$ per ogni istanza r di R;

ES. $\pi_{A,B,C}(r \bowtie s) \approx_R \pi_{A,B}(r) \bowtie \pi_{A,C}(s)$

r

A	B	D
a1	b1	d1
a2	b2	d2
a1	b2	d3

$\pi_{A,B}(r)$

A	B
a1	b1
a2	b2
a1	b2

s

A	C	E
a2	c1	e1
a2	c2	e2
a1	c3	e3

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

A	C
a2	c1
a2	c2
a1	c3

$\pi_{A,B,C}(r \bowtie s)$

A	B	C
a1	b1	c3
a2	b2	c1
a2	b2	c2
a1	b2	c3

$\pi_{A,B}(r) \bowtie \pi_{A,C}(s)$

A	B	C
a1	b1	c3
a2	b2	c1
a2	b2	c2
a1	b2	c3

s2

A	C	E	D
a2	c1	e1	d2
a2	c2	e2	d4
a1	c3	e3	d5

$\pi_{A,B,C}(r \bowtie s2)$

A	B	C
a2	b2	c1

EQUIVALENZA DI ESPRESSIONI ALGEBRICHE

E1, E2 espressioni dell'algebra relazionale

*** EQUIVALENZA ASSOLUTA**

$$E1 \approx E2 \text{ se } E1 \approx_R E2 \text{ \underline{per ogni schema} } R;$$

Es.

$$\sigma_{\vartheta}(r \times s) \approx r \bowtie_{\vartheta} s$$

$$\pi_{A,B}(\sigma_{A<0}(r)) \approx \sigma_{A<0}(\pi_{A,B}(r))$$

L'equivalenza è semantica, non computazionale

$$\Downarrow$$

Trasformazioni di espressioni in espressioni equivalenti sono usate per ottimizzare il costo delle interrogazioni (riducendo le dimensioni dei risultati intermedi)

TRASFORMAZIONI DI INTERESSE

*** Atomizzazione delle selezioni:**

$$\sigma_{F1 \text{ AND } F2}(E) \approx \sigma_{F1}(\sigma_{F2}(E))$$

*** Idempotenza delle proiezioni:**

$$\pi_A(E) \approx \pi_A(\pi_{A,B}(r))$$

*** Anticipazione della selezione rispetto al join:**

$$\sigma_F(E1 \bowtie_F E2) \approx E1 \bowtie_F \sigma_F(E2)$$

(se F contiene solo attributi che compaiono in E2)

*** Anticipazione della proiezione rispetto al join:**

$$\pi_A(E1 \bowtie_F E2) \approx E1 \bowtie_F \pi_A(E2)$$

(se A è un insieme di attributi che compaiono solo in E2 e gli altri attributi di E2 non sono coinvolti nel join).

In generale:

$$\pi_Y(E1 \bowtie_F E2) \approx \pi_Y(\pi_{Y1}(E1) \bowtie_F \pi_{Y2}(E2))$$

dove

- *** X1 = insieme di attributi in E1;**
- *** X2 = insieme di attributi in E2;**
- *** $J1 \subseteq X1$, $J2 \subseteq X2$ attributi presenti in F;**
- *** $Y1 = (X1 \cap Y) \cup J1$;**
- *** $Y2 = (X2 \cap Y) \cup J2$.**

*** Combinazione di selezione e prodotto cartesiano a formare un join:**

$$\sigma_{\vartheta}(E1 \times E2) \approx E1 \bowtie_{\vartheta} E2$$

ALTRE TRASFORMAZIONI

*** Distributività della selezione rispetto all'unione:**

$$\sigma_F (E1 \cup E2) \approx \sigma_F (E1) \cup \sigma_F (E2)$$

*** Distributività della selezione rispetto alla differenza:**

$$\sigma_F (E1 - E2) \approx \sigma_F (E1) - \sigma_F (E2)$$

*** Distributività della proiezione rispetto all'unione:**

$$\pi_F (E1 \cup E2) \approx \pi_F (E1) \cup \pi_F (E2)$$

*** Relazioni fra operatori insiemistici e selezioni composte:**

$$\sigma_{F1 \text{ AND } F2} (E) \approx \sigma_{F1} (E) \cap \sigma_{F2} (E)$$

$$\sigma_{F1 \text{ OR } F2} (E) \approx \sigma_{F1} (E) \cup \sigma_{F2} (E)$$

$$\sigma_{F1 \text{ AND } (\text{NOT } F2)} (E) \approx \sigma_{F1} (E) - \sigma_{F2} (E)$$

*** Proprietà commutativa e associativa degli operatori binari (meno la differenza)**

*** Proprietà distributiva del join rispetto all'unione:**

$$E \bowtie (E1 \cup E2) \approx (E \bowtie E1) \cup (E \bowtie E2)$$