ALGEBRA RELAZIONALE (CAPITOLO 3)

R. Basili

a.a. 2021-2022

Algebra Relazionale

- **Linguaggio di interrogazione o *query language* è un linguaggio tramite il quale è possibile richiedere informazioni contenute in una base di dati attraverso l'uso di vocabolari dei dati definiti dall'utente e requisiti logici espressi in un algebra di tipo booleano.**

- **Algebra relazionale**: orientata alle procedure, ha natura teorica
- **Calcolo relazionale**: orientato alla descrizione dei dati (dichiarativo), solo teorico
- **Datalog**: orientato alla descrizione dei dati (dichiarativo), aumenta la espressività, poco diffuso

Operatori su relazioni

- Il risultato dell'applicazione di un operatore ad una relazione è una relazione.
- Posso quindi costruire espressioni e query complesse

Operatori

- Unione (U), Differenza (-) e Intersezione ($\cap$)
- Ridenominazione (ρ), Selezione (σ) e Proiezione (Π)
- Join (equijoin, naturale o prodotto cartesiano)

Operatori insiemistici

- E’ possibile applicare gli operatori insiemistici solamente a coppie di relazioni OMOGENEE, cioè definite sugli stessi attributi.
- Date le relazioni A e B definite sullo stesso schema:
- UNIONE ($A \cup B$) = relazione contenente le tuple presenti in A ed in B
- DIFFERENZA ($A – B$) = relazione contenente le tuple presenti in A ma non in B
- INTERSEZIONE ($A \cap B$) = relazione contenente le tuple presenti sia in A che in B

(Nota: $A \cap B = A-(A-B)$)

UNION COMPATIBILITY

- Due relazioni sono *Union compatible* sse:
  - Esse hanno lo stesso numero di campi
  - Campi corrispondenti secondo l'ordine sx->dx hanno lo stesso dominio

- Solo in tal caso quindi sono applicabili gli operatori insiemistici di unione, differenza ed intersezione

Operatore di Ridenominazione (ρ)

- Sia r una relazione definita sull'insieme di attributi X e sia Y un (altro) insieme di attributi con la stessa cardinalità. Siano $A_1\ A_2 \ldots\ A_k$ e $B_1\ B_2 \ldots\ B_k$ rispettivamente un ordinamento per gli attributi in X ed Y allora la RIDENOMINAZIONE

$$\rho_{B_1 B_2 .. B_k \rightarrow A_1 A_2 .. A_k}(r)$$

- contiene una tupla t' per ciascuna tupla $t \in r$ tale che:

$$t'[B_i]=t[A_i] \qquad i=1,\ldots,k$$

- La RIDENOMINAZIONE **modifica solamente il nome degli attributi** agendo sullo schema della relazione e **lasciando inalterata la sua istanza**

OPERATORE DI SELEZIONE (σ)

$\sigma_F(r)$

Formula proposizionale

Formula proposizionale F su X è una formula booleana (¬,∧,∨) di *termini* T.

Dove T e' definito come:

attr op attr o ***attr op c***

op è un operatore di confronto (=,≠,≤,≥,<,>);

S

sid	nome	login	età	gpa
0012	Rossi	ro@ec	18	3.4
0072	Verdi	bi@ec	19	3.2
0033	Bianchi	bi@tt	18	3.8

$\sigma_{eta'=18}(S)$

sid	nome	login	età	gpa
0012	Rossi	ro@ec	18	3.4
0033	Bianchi	bi@tt	18	3.8

Produce una relazione avente lo stesso schema di r e contenente le tuple di r per le quali F e' **vera**

Operatore di Proiezione (π)

$\pi_{C}(r)$

Condizione

Condizione di proiezione *C* su r è l'elenco degli attributi dello schema X di r che debbono venire 'proiettati':

$attr_1, attr_2, \ldots, attr_n$

$attr_i \in X$ per $1 \leq i \leq n$

S

sid	nome	login	età	gpa
0012	Rossi	ro@ec	18	3.4
0072	Verdi	bi@ec	19	3.2
0033	Bianchi	bi@tt	18	3.8

$\pi_{età}(S)$

età
18
19

Distinte

Tutte le tuple della relazione r contribuiscono al risultato di una proiezione, ma soltanto con i valori corrispondenti agli attributi specificati dalla condizione C.

Esempi di Istanze:

Sailors (1)

S1

sid	sname	rating	age
22	Davide	7	45.0
31	Luca	8	55.5
58	Remo	10	35.0

Sailors (2)

S2

sid	sname	rating	age
28	Eolo	9	35.0
31	Luca	8	55.5
44	Giulio	5	35.0
58	Remo	10	35.0

Reserves

R1

sid	bid	day
22	101	10/10/03
58	103	11/12/02

Boats

B1

bid	bname	color
101	Futura	verde
58	Onda	rossa

IL PRODOTTO CARTESIANO

R X S

Ogni tupla di S viene associata ad ogni tupla di R. Lo schema risultante ha tutti gli attributi di R seguiti da quelli di S.

S1 X R1

(sid)	sname	rating	age	(sid)	bid	day
22	Davide	7	45.0	22	101	10/10/03
22	Davide	7	45.0	58	103	11/12/02
31	Luca	8	55.5	22	101	10/10/03
31	Luca	8	55.5	58	103	11/12/02
58	Remo	10	35.0	22	101	10/10/03
58	Remo	10	35.0	58	103	11/12/02

In caso di *conflitto* sui nomi degli attributi di R ed S applico l'operatore di ***ridenominazione*** :

$\rho(C(1 \rightarrow sid1, 5 \rightarrow sid2), S1XR1)$

Operazioni di join

$$R \bowtie_C S = \sigma_C(R \times S)$$

Condizione di Join

La relazione risultante avrà lo **schema** del prodotto cartesiano di R ed S e **l'istanza** composta dalle tuple del prodotto cartesiano che soddisfano la condizione C

$$S \bowtie_{S1.sid<R1.sid} R$$

(sid)	sname	rating	age	(sid)	bid	day
22	Davide	7	45.0	58	103	11/12/02
31	Luca	8	55.5	58	103	11/12/02

Operazioni di Equi-Join e Natural-Join

$$R \bowtie_{C} S$$

Condizione di Join

- ***Equi-Join*** è un join in cui la condizione ***C*** e' composta solamente di operazioni di uguaglianza.
- ***Natural Join*** e'un Equi-Join su ***<u>tutti</u>*** gli attributi in comune

$$S \bowtie_{sid} R$$

(sid)	sname	rating	age	(sid)	bid	day
22	Davide	7	45.0	~~22~~	101	11/10/03
58	Remo	10	35.0	~~58~~	103	11/12/02

Operazione di Divisione tra Relazioni

M/N

Divisione: prologo

- In analogia con la divisione intera (cioè tra numer interi) tra due numeri *m* ed *n*:
- Di fatto il quoziente q ed il resto r hanno le seguenti proprietà

$m \quad n \qquad q, \; r$

$13 : 4 = 3 \qquad m : n = q$

$r = 1$

$\longrightarrow \; n \cdot q + r = m$

$\longrightarrow \; q$ intero max t.c. $q \cdot n \leq m$

DIVISIONE: PROLOGO (2)

- La definizione di divisione tra relazioni M ed N richiede delle proprietà analoghe per il quoziente Q $\triangleq$ M/N:

$$n \cdot q + r = m$$

q intero max t.c. $q \cdot n \leq m$

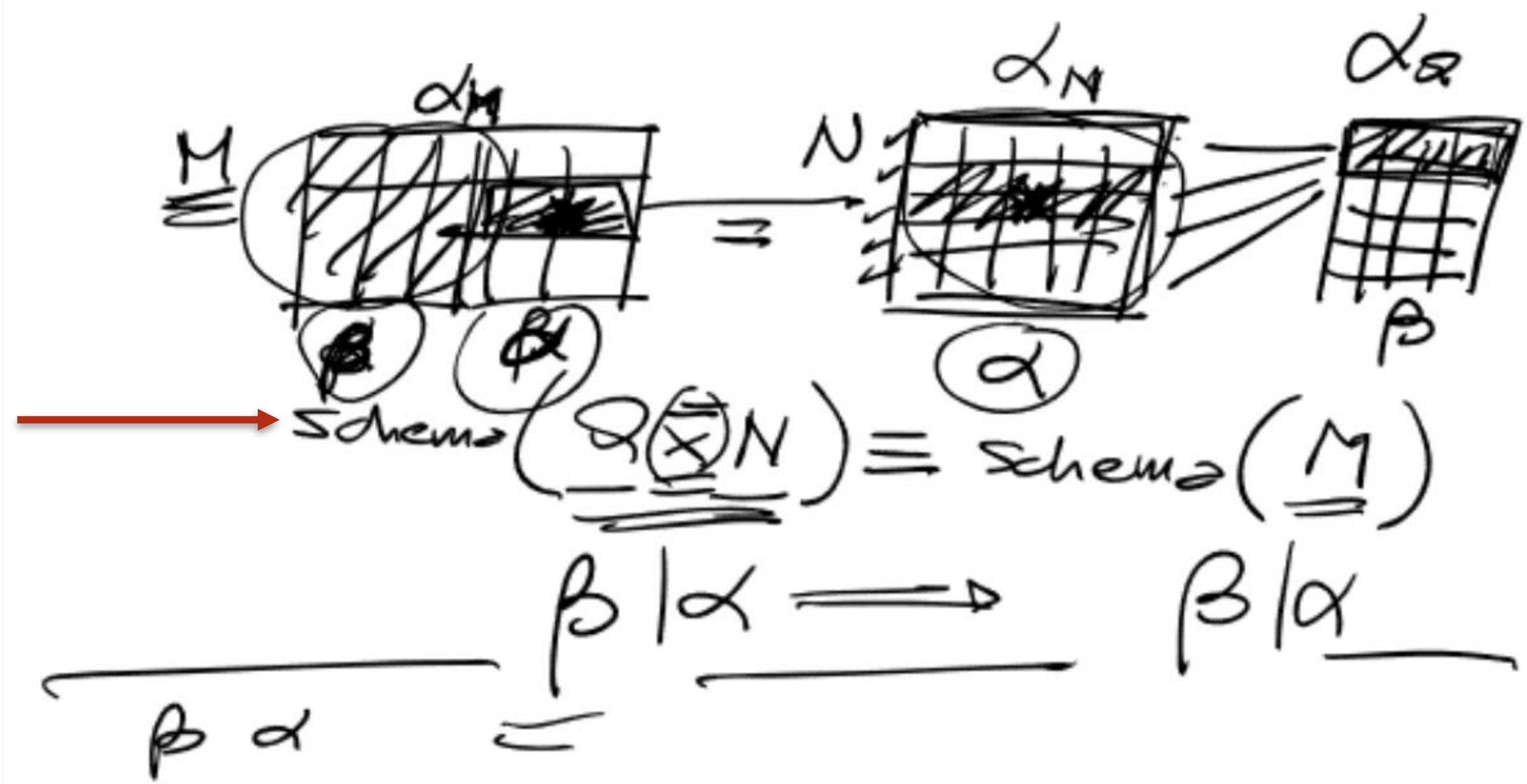
$M, N \qquad M \text{ div } N \qquad Q$

$$\longrightarrow Q \times N \subseteq M$$

$$M \times N \triangleq Q$$

DIVISIONE: PROLOGO (3)

- Le conseguenze di queste proprietà desiderabili coinvolgono gli schemi logici di M, N e Q:



DIVISIONE: PROLOGO (3)

- Esempio:

$M = \{\langle 1,3\rangle, \langle 2,3\rangle, \langle 1,2\rangle\}$

$N = \{\langle 2\rangle, \langle 3\rangle\}$

$Q = \{\langle 1\rangle\}$ $\qquad$ $\underline{Q \times N} = \{\langle 1,2\rangle, \langle 1,3\rangle\}$

$Q \times N \subset M$ $\quad (*)$

Se agg. $\langle 2\rangle$ a Q

$Q' = Q \cup \{\langle 2\rangle\}$

$\Longrightarrow$ $Q' \times N \supset M$ $\quad (*)$ viene violata

Operazione di Divisione

$$M/N = \pi_{\beta}(M) - \pi_{\beta}((\pi_{\beta}(M) \times N) - M)$$

$M^{(\beta\alpha)} = \{ <1,3> <2,3>, <1,2> \}$

$N = \{ <2>, <3> \}$

$Q = \{ <1> \}$ $\qquad$ $Q \times N = \{ <1,2>, <1,3> \}$

$\underbrace{Q \times N}_{\beta\ \alpha} \subset M$ $\quad$ ($*$)

Se agg $<2>$ a Q

$Q' = Q \cup \{ <2> \}$

$\Rightarrow$ $Q' \times N \supset M$ $\quad$ ($*$) Viene violata

Operazione di Divisione

$$M/N = \pi_{\beta}(M) - \pi_{\beta}((\pi_{\beta}(M) \times N) - M)$$

$$\boxed{M/N \triangleq \pi_{\beta}(M) - \pi_{\beta}\left[(\pi_{\beta}(M) \times N) - M\right]}$$

$$\pi_{\beta}(M) = \{\langle 1 \rangle, \langle 2 \rangle\}$$

$$\pi_{\beta}(M) \times N = \{\langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 2, 2 \rangle, \langle 2, 3 \rangle\}$$

$$\pi_{\beta}(M) \times N \supset M$$

$$\pi_{\beta}(M) \times N - M = \{\langle 2, 2 \rangle\}$$

Operazione di Divisione

$$M/N = \pi_\beta(M) - \pi_\beta((\pi_\beta(M) \times N) - M)$$

Non e' supportato come primitiva ma e' utile per esprimere query del tipo:

*Trova i marinai che hanno prenotato **tutte** le barche*
$M/N = \{ \langle x \rangle \mid \exists \langle x,y \rangle \in M \; \forall \langle y \rangle \in N \}$

L'IDEA è cogliere la natura distributiva del prodotto cartesiano. La *divisione* è l'inversa del prodotto cartesiano nel senso seguente:

M/N = Q (quoziente) **sse** esiste R tc

$$Q \times N \cup R = M$$

Operazione di Divisione

$$M/N = \pi_{sid}(M) - \pi_{sid}((\pi_{sid}(M) \times N) - M)$$

Non e' supportato come primitiva ma e' utile per esprimere query del tipo:

*Trova i marinai che hanno prenotato **tutte** le barche*

$M/N = \{ \langle sid \rangle \mid \exists \langle sid, bid \rangle \in M \ \forall \langle bid \rangle \in N \}$

sid	bid
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s3	p2

M

bid
p2

N1

bid
p2
p4

N2

M/N1 →

sid
s1
s3

M/N2 →

sid
s1

Q1

Trova il nome dei marinai (tabella S) che hanno riservato la barca numero 103.

$$\pi_{sname}(\sigma_{bid=103}(Reserves \bowtie Sailors))$$

Opt

$$\boxed{\pi_{sname}((\sigma_{bid=103}Reserves) \bowtie Sailors)}$$

In maniera più elegante:

$$\rho(Temp1, \sigma_{bid=103}Reserves)$$

$$\rho(Temp2, Temp1 \bowtie Sailors)$$

$$\pi_{sname}(Temp2)$$

Q2

Trova il nome dei marinai (tabella S) che hanno riservato la barca rossa.

$$\pi_{sname}(\sigma_{color='rossa'} \text{Boats} \bowtie \text{Reserves} \bowtie \text{Sailors})$$

Opt

$$\boxed{\pi_{sname}(\pi_{sid}((\pi_{bid}\sigma_{color='rossa'}\text{Boats})\bowtie\text{Reserves})\bowtie\text{Sailors})}$$

In maniera più elegante:

$$\rho(\text{Temp1}, \pi_{bid}\sigma_{color='rossa'}\text{Boats})$$

$$\rho(\text{Temp2}, \pi_{sid}(\text{Temp1}\bowtie\text{Reserves}))$$

$$\rho(\text{Temp3}, \text{Temp2} \bowtie \text{Sailors})$$

$$\boxed{\pi_{sname}(\text{Temp3})}$$

Q3 e Q4

*Trova il nome dei marinai (tabella S) che hanno riservato una barca rossa **o** una barca verde.*

$$\rho(TempBoats, (\sigma_{color='rossa'} Boats) \cup (\sigma_{color='verde'} Boats))$$

oppure:

$$\rho(TempBoats, (\sigma_{color='rossa' \vee color='verde'} Boats))$$

$$\pi_{sname}(TempBoats \bowtie Reserves \bowtie Sailors)$$

*Trova il nome dei marinai (tabella S) che hanno riservato una barca rossa **ed** una barca verde.*

$$\rho(TempRossa, \pi_{sid}(\sigma_{color='rossa'} Boats \bowtie Reserves))$$

$$\rho(TempVerde, \pi_{sid}(\sigma_{color='verde'} Boats \bowtie Reserves))$$

$$\pi_{sname}((TempRossa \cap TempVerde) \bowtie Sailors)$$

Perché non abbiamo utilizzato la congiunzione?

Q5 e Q6

Trova il nome dei marinai (tabella S) che hanno riservato almeno due barche

Q5 E Q6

Trova il nome dei marinai (tabella S) che hanno riservato almeno due barche

$\rho(Reservations, \pi_{\text{sid, sname,bid}}(\text{Sailors} \bowtie \text{Reserves}))$

$\rho(ReservationPairs(1 \rightarrow \text{sid1}, 2 \rightarrow \text{sname1}, 3 \rightarrow \text{bid1}, 4 \rightarrow \text{sid2}, 5 \rightarrow \text{sname2}, 6 \rightarrow \text{bid2}), \text{Reservations} \times \text{Reservations}$

$$\boxed{\pi_{\text{sname2}}\ \sigma_{(sid1=sid2) \wedge (bid1 \neq bid2)} \text{ReservationPairs}}$$

Q5 E Q6

Trova il nome dei marinai (tabella S) che hanno riservato almeno due barche

$\rho(Reservations,\ \pi_{sid,\ sname,bid}(\text{Sailors} \triangleright\triangleleft \text{Reserves}))$

$\rho(ReservationPairs(1 \rightarrow \text{sid1}, 2 \rightarrow \text{sname1}, 3 \rightarrow \text{bid1}, 4 \rightarrow \text{sid2}, 5 \rightarrow \text{sname2}, 6 \rightarrow \text{bid2}), \text{Reservations} X \text{Reservations}$

$$\boxed{\pi_{sname2}\ \sigma_{(sid1=sid2)\ \wedge (bid1 \neq bid2)} \text{ReservationPairs}}$$

Trova il sid dei marinai (tabella S) che hanno più di 20 anni e che non hanno riservato una barca rossa.

$$\pi_{sid}(\sigma_{age>20}\text{Sailors}) -$$
$$\pi_{sid}((\sigma_{color='rossa'}\text{Boats}) \triangleright\triangleleft \text{Reserves}\ )$$

Allegato A – Alcune regole di equivalenza

- **Il join naturale è commutativo ed associativo:**

$E_1 \triangleright\triangleleft E_2 \equiv E_2 \triangleright\triangleleft E_1$

$(E_1 \triangleright\triangleleft E_2) \triangleright\triangleleft E_3 = E_1 \triangleright\triangleleft (E_2 \triangleright\triangleleft E_3) = E_1 \triangleright\triangleleft E_2 \triangleright\triangleleft E_3$

- **Selezione e proiezione si possono raggruppare:**

$\sigma_{F1}(\sigma_{F2}(E)) = \sigma_{F1 \wedge F2}(E) \quad \pi_Y(\pi_{XY}(E)) = \pi_Y(E)$

- **Selezione e proiezione commutano (F si riferisce solo ad attributi in Y):**

$\pi_Y(\sigma_F(E)) = \sigma_F(\pi_Y(E))$

- **"Push-down" della selezione rispetto al join (F è nello schema di E_1):**

$\sigma_F(E_1 \triangleright\triangleleft E_2) = \sigma_F(E_1) \triangleright\triangleleft E_2$

Allegato A – Osservazione: equivalenza rispetto a identità

- **La equivalenza $\equiv$ usata nella regola:**

$$E_1 \rhd\lhd E_2 \equiv E_2 \rhd\lhd E_1$$

È diversa dalla identità = che vale invece in:

$$(E_1 \rhd\lhd E_2) \rhd\lhd E_3 = E_1 \rhd\lhd (E_2 \rhd\lhd E_3) = E_1 \rhd\lhd E_2 \rhd\lhd E_3$$

- **La equivalenza $\equiv$ non richiede la identità anche rispetto allo schema logico. Osservate che infatti** $E_1 \rhd\lhd E_2$ *ed* $E_2 \rhd\lhd E_1$ **hanno schemi logici diversi, nel caso generale. Quindi affinchè** $A \equiv B$ **basta che esistano regole di scambio di righe tali che i due schemi siano riconducibili l'uno all'altro e che le tuple così ottenute siano le stesse.**
- **L'identità algebrica espressa da** $A=B$ **richiede che i due insiemi siano identici rispetto <u>sia</u> allo schema <u>che</u> ai valori.**

DOMANDE

- Che significa che gli operatori di algebra relazionale possono essere composti?
- Qual'e' la cardinalità minima e massima di
  - $R \times \sigma_F R$
  - $\sigma_{F2} R \cup \sigma_{F1} R$
  - $(\sigma_F R1 \times R2) \cap (R1 \times \sigma_F R2)$

  in funzione di $|R|$, $|R1|$, $|R2|$?
- Dimostrare sotto quali condizioni:
  - $\sigma_F R \times A \subseteq R \times A$
  - $(A \cup B) \cap C = (A \cup \overline{B}) \setminus C$
  - $(\sigma_F R1 \times R2) \neq (R1 \times \sigma_F R2)$