

Basi di Dati

Algebra relazionale

Corso B

Base dati di esempio

PAZIENTI

<u>COD</u>	Cognome	Nome	Residenza	AnnoNascita
A102	Necchi	Luca	TO	1950
B372	Rossigni	Piero	NO	1940
B543	Missoni	Nadia	TO	1960
B444	Missoni	Luigi	VC	2000
S555	Rossetti	Gino	AT	2010

REPARTI

<u>COD</u>	Nome-Rep	Primario
A	Chirurgia	203
B	Pediatria	574
C	Medicina	530
L	Lab-Analisi	530
R	Radiologia	405

RICOVERI

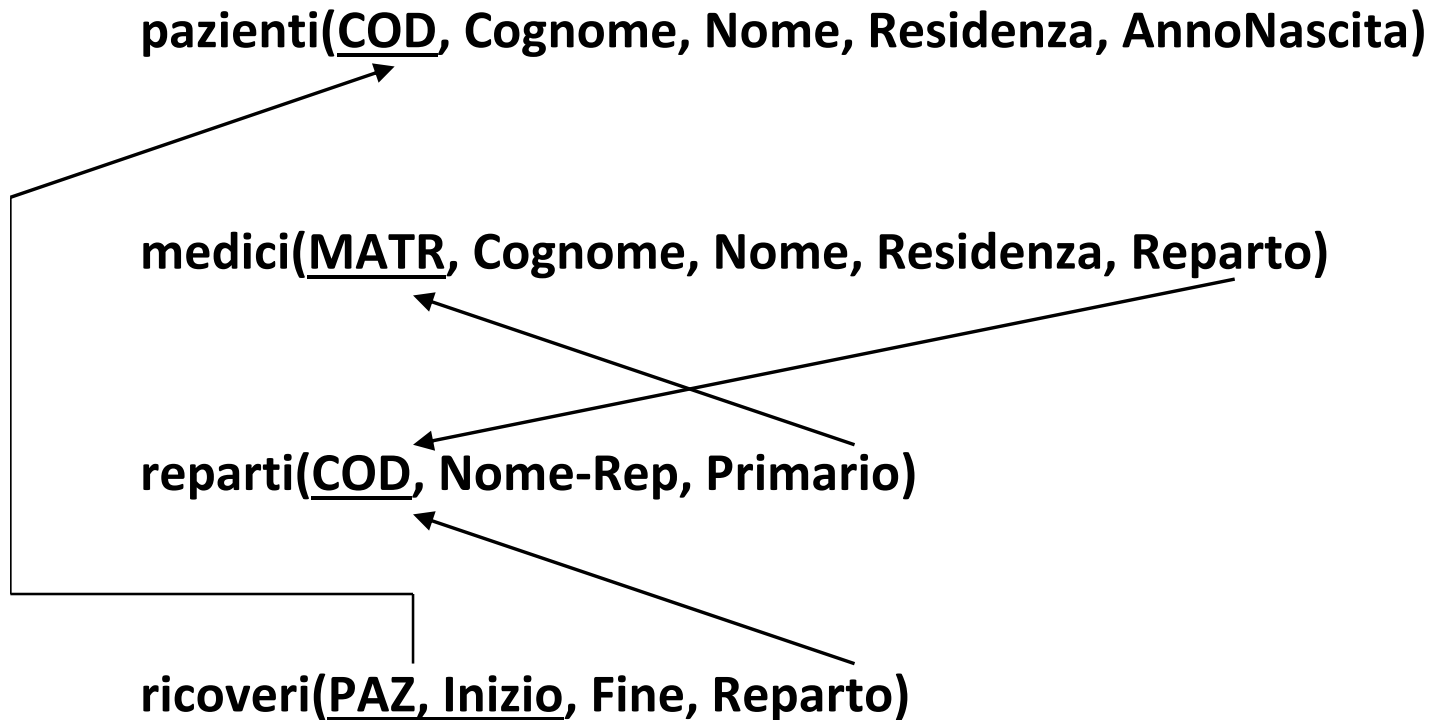
<u>PAZ</u>	Inizio	Fine	Reparto
A102	2/05/2014	9/05/2014	A
A102	2/12/2004	2/01/2005	A
S555	5/10/2014	3/12/2014	B
B444	1/12/2004	2/01/2005	B
S555	6/09/2015	1/11/2015	A

MEDICI

<u>MATR</u>	Cognome	Nome	Residenza	Reparto
203	Neri	Piero	AL	A
574	Bisi	Mario	MI	B
461	Bargio	Sergio	TO	B
530	Belli	Nicola	TO	C
405	Mizzi	Nicola	AT	R
501	Monti	Mario	VC	A

Base dati di esempio

Schema relazionale con vincoli di integrità referenziali:



DDL e DML

- Il modello relazionale descrive il modello del **DDL**, ovvero il Data Definition Language
- Ora ci occuperemo del **DML** (Data Manipulation Language) il linguaggio che si occupa di manipolare le tuple presenti nella base di dati

DML

Comandi di inserzione tuple

- `INSERT INTO pazienti(COD,Cognome,Nome,Residenza,AnnoNascita)`
`VALUES ('Z111','Rossi','Piero',' TO','1980');`

Comandi di cancellazione tuple

- `DELETE FROM pazienti WHERE condizione di selezione della tupla da cancellare;`

Modifica di tuple

- `UPDATE pazienti SET (Nome='Pietro') WHERE condizione di selezione della tupla da cancellare`

Interrogazione

- `SELECT AnnoNascita FROM pazienti WHERE condizione di selezione delle tuple da visualizzare`

DML

Il cuore del DML è il linguaggio di interrogazione

Modello dell'interrogazione di Codd

Input di un'interrogazione: **la base di dati**

Output dell'interrogazione: **una relazione**

Esempio: elencare i pazienti ricoverati nel reparto il cui primario è il Dott. Neri; visualizzare cognome, nome e nome del reparto (vedi [esempio](#))

Algebra relazionale

Codd formalizza l'interrogazione utilizzando il paradigma algebrico

Il paradigma algebrico è una costruzione procedurale dell'interrogazione (elenco dei passi per eseguire l'operazione)

La presentazione è però più astratta, utilizzando degli operatori algebrici

Operatori dell'algebra relazionale

L'algebra relazionale è costituita da operatori di base e da operatori derivati

- Operatori di base
 - Selezione
 - Proiezione
 - Prodotto cartesiano
 - Unione
 - Differenza
 - Ridenominazione
- Operatori derivati
 - Intersezione
 - Join nelle sue varie forme (equi-join, **theta-join**, natural join, ecc.)
 - Quoziente

Operatori algebrici

Gli operatori dell'algebra relazionale ricevono, come argomenti, relazioni e producono, in uscita, relazioni (dette anche relazioni virtuali)

Operatore di selezione

- Data una relazione r su uno schema A , $\sigma_p(r(A))$ è l'**operatore di selezione** dove p è un predicato e $r(A)$ è l'argomento dell'operatore
- *Per semplicità ignoreremo la possibilità che ci siano valori nulli*

Sintassi del predicato di selezione

Il predicato p è un'espressione booleana di predicati atomici

I predicati atomici sono solo di due tipi

- $A_i \theta$ (*confronto*) *costante*
- $A_i \theta$ (*confronto*) A_j

dove

- A_i e A_j sono attributi dello schema A
- Il confronto θ è un **operatore di confronto** nell'insieme $\{=, \neq, >, \geq, <, \leq\}$

Esempi

Selezione dei pazienti con residenza a Torino

$$\sigma_{Residenza='TO'}(pazienti)$$

Selezione dei pazienti residenti a Torino o a Vercelli

$$\sigma_{Residenza='TO' \vee Residenza='VC'}(pazienti)$$

Selezione dei pazienti non residenti a Torino

$$\sigma_{\neg Residenza='TO'}(pazienti)$$

Risultati di un interrogazione

Semantica dell'operatore di selezione

- La selezione produce una **relazione senza nome** che ha **l'identico schema** della relazione argomento, ma solo le tuple che **soddisfano il predicato**

Esempi

$\sigma_{Residenza='TO'} (pazienti)$

COD	Cognome	Nome	Residenza	AnnoNascita
A102	Necchi	Luca	TO	1950
B543	Missoni	Nadia	TO	1960

$\sigma_{Residenza='TO' \vee Residenza='VC'} (pazienti)$

COD	Cognome	Nome	Residenza	AnnoNascita
A102	Necchi	Luca	TO	1950
B543	Missoni	Nadia	TO	1960
B444	Missoni	Luigi	VC	2000

$\sigma_{\neg Residenza='TO'} (pazienti)$

COD	Cognome	Nome	Residenza	AnnoNascita
B372	Rossigni	Piero	NO	1940
B444	Missoni	Luigi	VC	2000
S555	Rossetti	Gino	AT	2010

Operatore di selezione

Data una relazione r su uno schema A e un predicato p , l'**operatore di selezione** $\sigma_p(r(A))$ produce una **relazione senza nome** che ha l'**identico schema** A della relazione argomento, ma solo le tuple che **soddisfano il predicato** p .

Cardinalità della selezione

La **cardinalità** della relazione virtuale prodotta dall'operazione di selezione $\sigma_p(r(A))$ è

$$0 \leq |\sigma_p(r(A))| \leq |r(A)|$$

- La cardinalità è 0 (relazione vuota) se il predicato è sempre falso
- La cardinalità è $|r(A)|$ se il predicato è sempre vero

Esempi

$\sigma_{Residenza='CN'} (pazienti)$

COD	Cognome	Nome	Residenza	AnnoNascita
-----	---------	------	-----------	-------------

$\sigma_{AnnoNascita < 2012} (pazienti)$

<u>COD</u>	Cognome	Nome	Residenza	AnnoNascita
A102	Necchi	Luca	TO	1950
B372	Rossigni	Piero	NO	1940
B543	Missoni	Nadia	TO	1960
B444	Missoni	Luigi	VC	2000
S555	Rossetti	Gino	AT	2010

Operatore di proiezione

Data una relazione $r(A)$ ed un insieme di attributi $A_{i'}, A_{j'} \dots A_k$, tutti appartenenti ad A , l'operatore di proiezione $\Pi_{A_{i'}, A_{j'} \dots A_k}(r(A))$ produce come risultato una relazione con

- schema: $\{A_{i'}, A_{j'} \dots A_k\}$
- istanza: tutte le tuple della relazione argomento, ma solo rispetto ai campi $A_{i'}, A_{j'} \dots A_k$

Esempi

Proiezione della relazione *pazienti* sull'attributo COD e sull'attributo Cognome

$\Pi_{COD, Cognome} (pazienti)$

<u>COD</u>	Cognome
A102	Necchi
B372	Rossigni
B543	Missoni
B444	Missoni
S555	Rossetti

Cardinalità della proiezione

A prima vista sembrerebbe che la cardinalità della proiezione è uguale alla cardinalità della relazione argomento, ovvero $| \Pi_{A_i, A_j, \dots, A_k}(r(A)) | = | r(A) |$

Invece, la cardinalità della proiezione $\Pi_{A_i, A_j, \dots, A_k}(r(A))$ è

$$0 \leq | \Pi_{A_i, A_j, \dots, A_k}(r(A)) | \leq | r(A) |$$

Esempio

Proiezione della relazione *pazienti* sull'attributo
Cognome

$\Pi_{\text{Cognome}}(\textit{pazienti})$

Cognome
Necchi
Rossigni
Missoni
Missoni
Rossetti

Non è più una relazione valida nel modello relazionale
teorico di Codd (non è più un insieme)

Esempio

Proiezione della relazione *pazienti* sull'attributo *Cognome*

$\Pi_{Cognome}(pazienti)$

Cognome
Necchi
Rossigni
Missoni
Rossetti

Il risultato è una istanza di relazione senza ripetizioni

Proprietà della proiezione

Se gli attributi proiettati A_i, A_j, \dots, A_k formano una **superchiave** della relazione argomento, allora

$$| \Pi_{A_i, A_j, \dots, A_k}(r(A)) | = | r(A) |$$

Composizione di operatori

L'algebra relazionale è composizionale, ovvero si possono costruire espressioni di algebra relazionale componendo insieme operatori

Esempio

Elencare codice e nome dei pazienti residenti a Torino o a Vercelli

- seleziono i pazienti residenti a Torino o Vercelli

$\sigma_{Residenza='TO' \vee Residenza='VC'} (pazienti)$

- la relazione virtuale prodotta dalla selezione può essere usata come argomento per la proiezione

$\Pi_{COD, Nome} ($

COD	Cognome	Nome	Residenza	AnnoNascita
A102	Necchi	Luca	TO	1950
B543	Missoni	Nadia	TO	1960
B444	Missoni	Luigi	VC	2000

$)$

Esempio

Elencare codice e nome dei pazienti residenti a Torino o a Vercelli

- seleziono i pazienti residenti a Torino o Vercelli

$$\sigma_{Residenza='TO' \vee Residenza='VC'} (pazienti)$$

- la relazione virtuale prodotta dalla selezione può essere usata come argomento per la proiezione

$$\Pi_{COD, Nome} (\sigma_{Residenza='TO' \vee Residenza='VC'} (pazienti))$$

Risultato della composizione

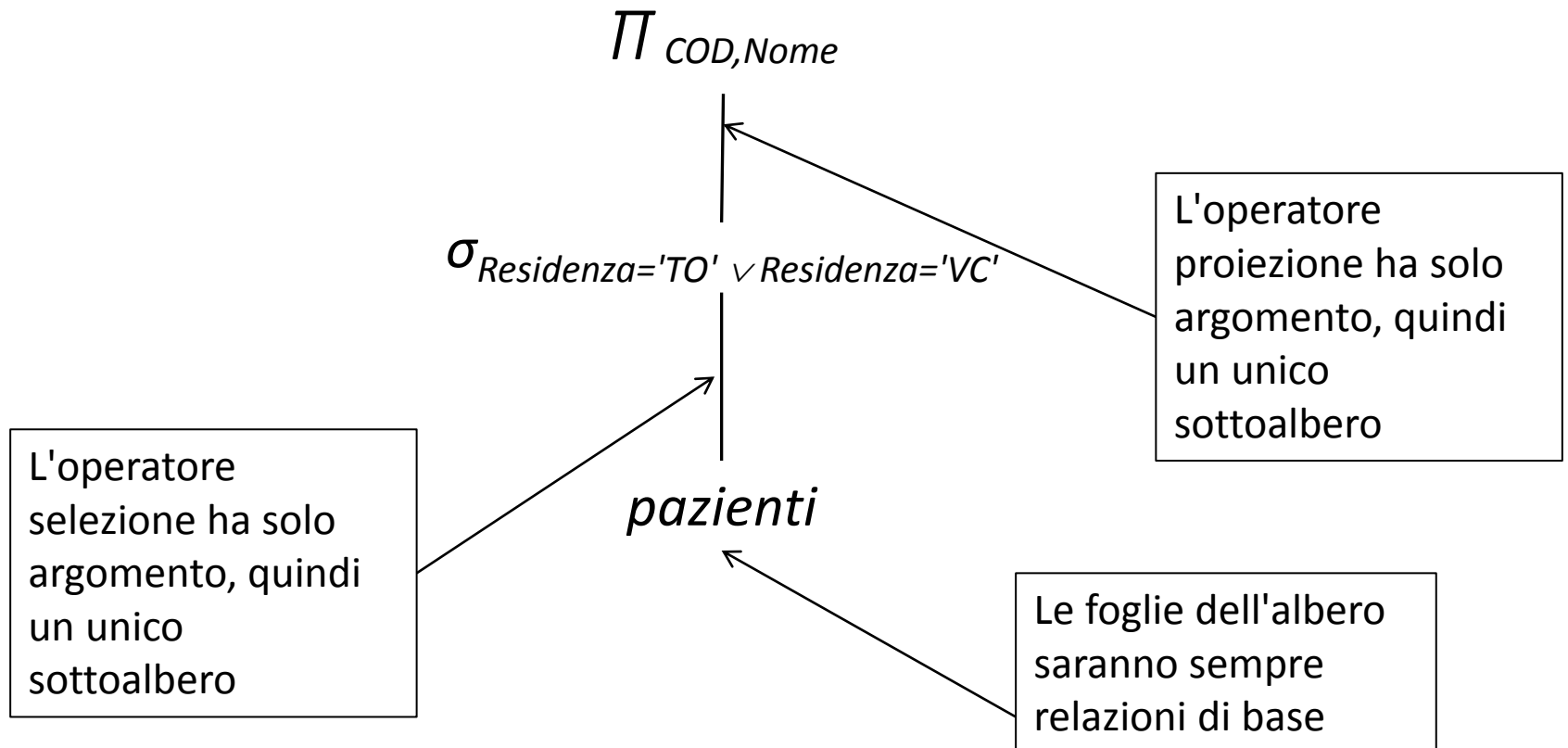
Il risultato è una relazione che ha

- schema: quello della proiezione, ovvero $\{COD, Nome\}$
- istanza: solo le tuple che soddisfano il predicato di selezione $Residenza='TO' \vee Residenza='VC'$

COD	Nome
A102	Luca
B543	Nadia
B444	Luigi

Notazione ad albero sintattico

$\Pi_{COD, Nome} (\sigma_{Residenza='TO' \vee Residenza='VC'} (pazienti))$



Correttezza sintattica

Per verificare che l'espressione algebrica sia sintatticamente corretta bisogna verificare che gli operatori algebrici siano coerenti o consistenti con gli argomenti

- $\sigma_{Residenza='TO' \vee Residenza='VC'}(pazienti)$ è corretta perché *Residenza* è un attributo della relazione *pazienti*
- $\pi_{COD, Nome}(\sigma_{Residenza='TO' \vee Residenza='VC'}(pazienti))$ è corretta perché la selezione produce tutti gli attributi dell'insieme di partenza (quindi anche *COD* e *Nome*)

Espressione non corretta

Invertendo i due operatori algebrici

$$\sigma_{Residenza='TO' \vee Residenza='VC'} (\pi_{COD, Nome} (pazienti))$$

l'espressione diventa **sintatticamente scorretta** perché la relazione virtuale prodotta dalla proiezione su pazienti non contiene l'attributo *Residenza*

Espressione corretta

Per renderla corretta, possiamo aggiungere l'attributo Residenza all'operatore di proiezione

$$\sigma_{Residenza='TO' \vee Residenza='VC'} (\pi_{COD, Nome, Residenza} (pazienti))$$

Espressione corretta

Per ottenere il risultato precedente, possiamo eseguire un'ulteriore proiezione su *COD* e *Nome*

$$\Pi_{COD, Nome} (\sigma_{Residenza='TO' \vee Residenza='VC'} (\Pi_{COD, Nome, Residenza} (pazienti)))$$

- E' possibile quindi scrivere qualsiasi combinazione di operatori, **purché sintatticamente corretta**

Operatori insiemistici

- Operatore unione \cup
- Operatore differenza $-$

Gli operatori insiemistici richiedono per definizione che gli schemi delle relazioni argomento siano identici

Operatore insiemistici

Il risultato dell'operatore insiemistico sulle relazioni argomento $r_1(A)$ ed $r_2(A)$ è una relazione che ha

- schema: lo stesso schema A delle relazioni argomento
- istanza:
 - Unione: $r_1 \cup r_2$ (unione delle tuple di r_1 ed r_2)
 - Differenza: $r_1 - r_2$ (tuple di r_1 che non sono contenute in r_2)

Cardinalità dell'unione

La cardinalità dell'unione $r_1(A) \cup r_2(A)$ è

$$0 \leq | r_1(A) \cup r_2(A) | \leq |r_1(A)| + |r_2(A)|$$

- Ovviamente la cardinalità può essere minore della somma delle cardinalità delle due relazioni se ci sono delle tuple ripetute

Esempio

Elencare cognomi e nomi di tutte le persone coinvolte nel DB "Ricoveri Ospedalieri" (quindi medici e pazienti)

$$\Pi_{\text{Cognome, Nome}}(\text{pazienti}) \cup \Pi_{\text{Cognome, Nome}}(\text{medici})$$

Cognome	Nome
Necchi	Luca
Rossigni	Piero
Missoni	Nadia
Missoni	Luigi
Rossetti	Gino

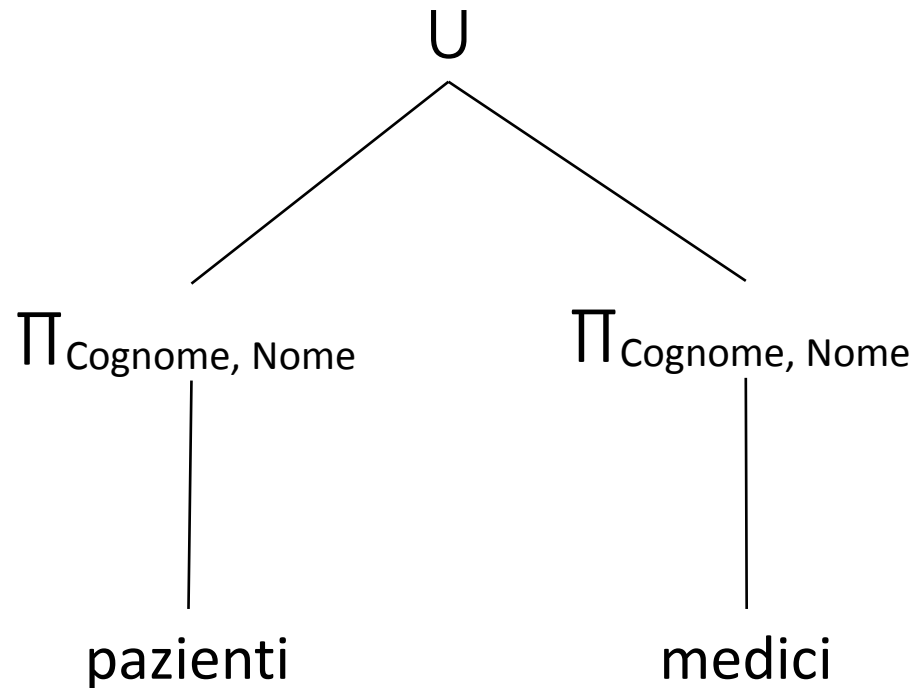
U

Cognome	Nome
Neri	Piero
Bisi	Mario
Bargio	Sergio
Belli	Nicola
Mizzi	Nicola
Monti	Mario

=

Cognome	Nome
Necchi	Luca
Rossigni	Piero
Missoni	Nadia
Missoni	Luigi
Rossetti	Gino
Neri	Piero
Bisi	Mario
Bargio	Sergio
Belli	Nicola
Mizzi	Nicola
Monti	Mario

Rappresentazione ad albero



Cardinalità della differenza

La cardinalità della differenza $r_1(A) - r_2(A)$ è

$$0 \leq |r_1(A) - r_2(A)| \leq |r_1(A)|$$

Esempio

Elencare le province di residenza dei medici in cui non risiede alcun paziente

$$\Pi_{Residenza}(medici) - \Pi_{Residenza}(pazienti)$$

Residenza	AL
MI	
TO	
AT	
VC	

—

Residenza	TO
NO	
VC	
AT	

=

Residenza	AL
MI	

Proprietà della differenza

Contrariamente all'unione, la differenza non gode della proprietà commutativa

$$\Pi_{Residenza}(pazienti) - \Pi_{Residenza}(medici)$$

Residenza	TO	NO	VC	AT
-----------	----	----	----	----

Residenza	AL	MI	TO	AT	VC
-----------	----	----	----	----	----

Residenza	NO
-----------	----

Operatore intersezione

Esattamente come gli altri due operatori, è definito su relazioni aventi lo stesso schema. Il risultato dell'operatore intersezione $r_1(A) \cap r_2(A)$ è una relazione che ha

- schema: lo stesso schema A delle relazioni argomento
- istanza: $r_1 \cap r_2$ (tuple di r_1 contenute anche in r_2)

Derivazione dell'intersezione

L'operatore di intersezione insiemistica può essere
reso con l'operatore di differenza insiemistica:

$$r_1(A) \cap r_2(A) = r_1(A) - (r_1(A) - r_2(A))$$

Cardinalità dell'intersezione

La cardinalità dell'intersezione $r_1(A) \cap r_2(A)$ è

$$0 \leq |r_1(A) \cap r_2(A)| \leq \min\{|r_1(A)|, |r_2(A)|\}$$

Problemi con gli operatori insiemistici

Con gli operatori che abbiamo visto, posso ottenere i medici che non sono primari?

MEDICI

<u>MATR</u>	Cognome	Nome	Residenza	Reparto
203	Neri	Piero	AL	A
574	Bisi	Mario	MI	B
461	Bargio	Sergio	TO	B
530	Belli	Nicola	TO	C
405	Mizzi	Nicola	AT	R
501	Monti	Mario	VC	A

REPARTI

<u>COD</u>	Nome-Rep	Primario
A	Chirurgia	203
B	Pediatria	574
C	Medicina	530
L	Lab-Analisi	530
R	Radiologia	405

L'attributo Primario della relazione REPARTI contiene la matricola (MATR) del medico che ha le funzioni di primario

Problemi con gli operatori insiemistici

Con gli operatori che abbiamo visto, posso ottenere i medici che non sono primari?

- Potrei usare la differenza insiemistica:

$$\Pi_{MATR}(medici) - \Pi_{Primario}(reparti)$$

Non posso farlo per ragioni sintattiche: lo schema del primo argomento dell'operatore differenza è **diverso** dallo schema del secondo argomento!

Operatore di ridenominazione

L'operatore di **ridenominazione** ha come argomento una relazione $r(A)$, e il suo compito è semplicemente quello di cambiar nome agli attributi (una parte o tutti) della relazione argomento

Operatore di ridenominazione

Data una relazione $r(A)$, con $A = \{A_1, \dots, A_i, \dots, A_j, \dots, A_k, \dots, A_n\}$
il risultato della ridenominazione

$$\rho_{B_i, B_j, \dots, B_k} \leftarrow_{A_i, A_j, \dots, A_k} (r)$$

è una relazione **virtuale*** $r'(A')$ con

- schema: $A' = \{A_1, \dots, B_i, \dots, B_j, \dots, B_k, \dots, A_n\}$ e
- istanza: $r' = r$ (stesse tuple)

(*) l'operatore **non modifica** la relazione della base dati

Esempio

Elencare i medici che non sono primari:

$$\Pi_{MATR}(medici) - \rho_{MATR \leftarrow \text{Primario}}(\Pi_{\text{Primario}}(reparti))$$

MATR
203
574
461
530
405
501

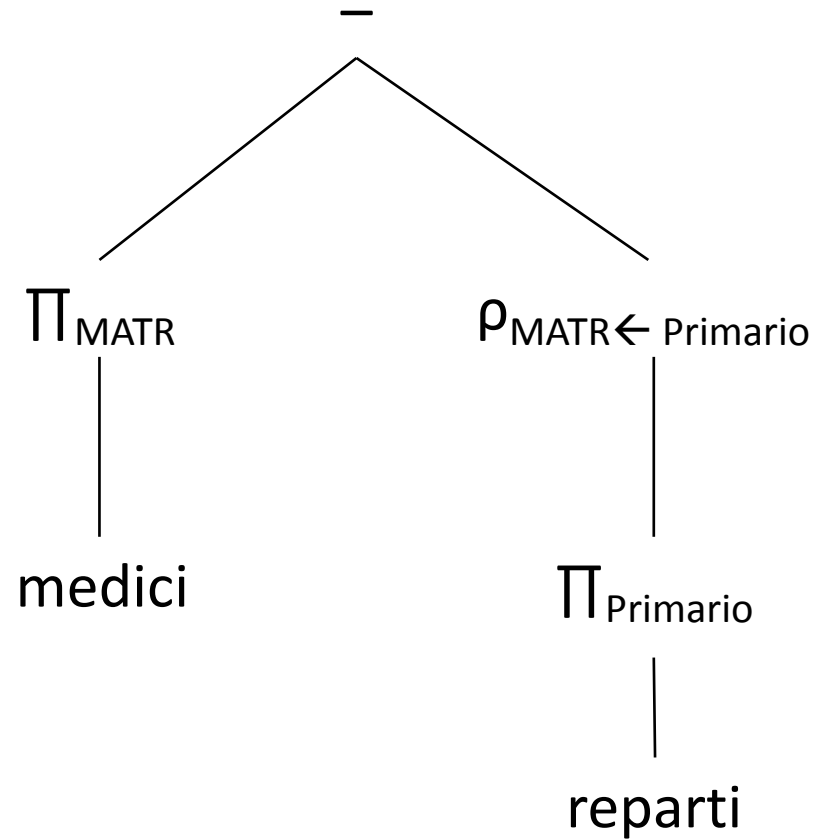
—

MATR
203
574
530
405

=

MATR
461
501

Rappresentazione ad albero



Uso corretto degli operatori insiemistici

Nell'utilizzo degli operatori insiemistici abbinati all'operatore di ridenominazione, bisogna assicurarsi sempre che gli attributi delle relazioni siano definiti su domini omogenei (coerenti)

Esempi:

- $\Pi_{MATR}(medici) - \rho_{MATR \leftarrow \text{Primario}}(\Pi_{\text{Primario}}(reparti))$
– *insiemi con domini omogenei*
- $\Pi_{Nome}(medici) - \rho_{Nome \leftarrow \text{Nome-Rep}}(\Pi_{\text{Nome-Rep}}(pazienti))$
– *insiemi con domini **non omogenei** (anche se gli attributi sono dello stesso tipo!)*

Esercizio

Elencare il codice dei pazienti che non sono mai stati ricoverati

Esercizio

Elencare le città in cui risiedono sia medici che pazienti

Ridenominazione schema

Data una relazione r , con schema $R(A)$,
 $A = \{A_1, \dots, A_i, \dots, A_j, \dots, A_{k'}, \dots, A_n\}$ il risultato della ridenominazione
dello schema

$$\rho_{R'(B_i, B_j, \dots, B_{k'})} \leftarrow R(A_i, A_j, \dots, A_{k'}) (r)$$

è una relazione **virtuale*** con schema $R'(A_1, \dots, B_i, \dots, B_j, \dots, B_{k'}, \dots, A_n)$

Esempio:

- $\rho_{UTENTI(CF, Provincia)} \leftarrow PAZIENTI(COD, Residenza) (pazienti)$
- Risultato: $UTENTI(CF, Cognome, Nome, Provincia, AnnoNascita)$

Prodotto cartesiano

Date due relazioni $r_1(A)$ ed $r_2(B)$, con $A \cap B = \emptyset$ (i due schemi non hanno attributi in comune), il prodotto cartesiano

$$r_1(A) \times r_2(B)$$

produce come risultato una relazione r' con

- *Schema: R' composto dall'unione degli schemi $A \cup B$*
- *Istanza: giustapposizione (combinazione) di tutte le tuple di $r_1(A)$ con tutte le tuple di $r_2(B)$.*

Esempio

A	B	C
a1	b1	c1
a2	b2	c2

×

D	E
d1	e1
d2	e2
d3	e3

A	B	C	D	E
a1	b1	c1	d1	e1
a1	b1	c1	d2	e2
a1	b1	c1	d3	e3
a2	b2	c2	d1	e1
a2	b2	c2	d2	e2
a2	b2	c2	d3	e3

Esempio

D	E
d1	e1
d2	e2
d3	e3

×

A	B	C
a1	b1	c1
a2	b2	c2

D	E	A	B	C
d1	e1	a1	b1	c1
d1	e1	a2	b2	c2
d2	e2	a1	b1	c1
d2	e2	a2	b2	c2
d3	e3	a1	b1	c1
d3	e3	a2	b2	c2

Proprietà del prodotto cartesiano

Il prodotto cartesiano gode della proprietà commutativa

$$r_1(A) \times r_2(B) = r_2(B) \times r_1(A)$$

Infatti, ricordiamo che nella relazione di Codd:

- nello schema di una relazione non è importante l'ordine degli attributi
- nell'istanza di una relazione non è importante l'ordine delle tuple

Relazioni identiche

A	B	C	D	E
a1	b1	c1	d1	e1
a1	b1	c1	d2	e2
a1	b1	c1	d3	e3
a2	b2	c2	d1	e1
a2	b2	c2	d2	e2
a2	b2	c2	d3	e3

D	E	A	B	C
d1	e1	a1	b1	c1
d1	e1	a2	b2	c2
d2	e2	a1	b1	c1
d2	e2	a2	b2	c2
d3	e3	a1	b1	c1
d3	e3	a2	b2	c2

Prodotto cartesiano

- Considereremo l'operatore di base prodotto cartesiano come un operatore tecnico
- Non ha un'utilità pratica diretta, ma sarà utile nella definizione di altri operatori molto importanti

Cardinalità del prodotto cartesiano

La cardinalità del prodotto cartesiano $r_1(A) \times r_2(B)$ è

$$0 \leq |r_1(A) \times r_2(B)| = |r_1(A)| \cdot |r_2(B)|$$

Infatti

- r_1 ed r_2 sono insiemi di tuple, quindi le tuple sono tutte disgiunte
- giustapponendo tuple disgiunte con altre tuple disgiunte ottengo una relazione di tuple disgiunte

Operatore di Θ -join: \bowtie_{θ}

- Serve a costruire informazioni estratte da più relazioni
- Serve a mettere in correlazione informazioni di una relazione con informazioni di un'altra relazione

Esempi

- Come posso elencare tutte le informazioni dei pazienti ricoverati?
- Come posso elencare tutte le informazioni dei primari di reparto?

Esempio di risultato

PAZ	Inizio	Fine	Reparto	COD	Cognome	Nome	Residenza	AnnoNascita
A102	2/05/2014	9/05/2014	A	A102	Necchi	Luca	TO	1950
A102	2/12/2004	2/01/2005	A	A102	Necchi	Luca	TO	1950
S555	5/10/2014	3/12/2014	B	S555	Rossetti	Gino	AT	2010
B444	1/12/2004	2/01/2005	B	B444	Missoni	Luigi	VC	2000
S555	6/09/2015	1/11/2015	A	S555	Rossetti	Gino	AT	2010

Operatore di Θ -join: \bowtie_{Θ}

Date:

- due relazioni $r_1(A)$ ed $r_2(B)$ con $A \cap B = \emptyset$ (i due schemi non hanno attributi in comune)
- una condizione (predicato) di join Θ da intendersi come una congiunzione di **confronti tra attributi** del tipo $A_i \varphi B_j$ dove $\varphi = \{<, >, =, \neq, \leq, \geq\}$

Il Θ -join (**theta-join**) è definito come

$$r_1(A) \bowtie_{\Theta} r_2(B) = \sigma_{\Theta}(r_1(A) \times r_2(B))$$

Esempio

Elencare tutte le informazioni sui pazienti ricoverati

Cosa produce il prodotto cartesiano tra ricoveri e pazienti?

Esempio

Il prodotto cartesiano produce tutte le combinazioni di tutte le tuple di *ricoveri* con tutte le tuple di *pazienti* tra cui, ad esempio, queste due:

PAZ	Inizio	Fine	Reparto	COD	Cognome	Nome	Residenza	AnnoNascita
...
A102	2/05/2014	9/05/2014	A	S555	Rossetti	Gino	AT	2010
A102	2/05/2014	9/05/2014	A	A102	Necchi	Luca	TO	1950
...

Solo un accoppiamento è però significativo!

Esempio

Impongo che il valore di *PAZ* sia uguale al valore di *COD*

PAZ	Inizio	Fine	Reparto	COD	Cognome	Nome	Residenza	AnnoNascita
...
A102	2/05/2014	9/05/2014	A	S555	Rossetti	Gino	AT	2010
A102	2/05/2014	9/05/2014	A	A102	Necchi	Luca	TO	1950
...

Di tutte le giustapposizioni, filtro solo quelle che soddisfano la condizione di join *PAZ=COD*

Esempio

Elencare tutte le informazioni sui pazienti ricoverati

$$ricoveri \bowtie_{PAZ=COD} pazienti$$

equivalente a

$$\sigma_{PAZ=COD}(ricoveri \times pazienti)$$

N.B.: gli schemi PAZIENTI e RICOVERI sono **disgiunti**

Funzionamento intuitivo del join

- Dovendo scegliere gli abbinamenti di tuple giusti secondo la condizione di join, non posso fare altro che confrontare **tutte** le coppie di tuple
- Codd si distacca dai linguaggi di programmazione e si affida all'algebra relazionale
- Il comportamento è reso dal prodotto cartesiano, che produce tutti gli abbinamenti, e dalla selezione che verifica la condizione di join

Esempio

Elencare le informazioni dei primari di ogni reparto

reparti ⋈_{Primario=MATR} *medici*

COD	Nome-Rep	Primario	MATR	Cognome	Nome	Residenza	Reparto
A	Chirurgia	203	203	Neri	Piero	AL	A
B	Pediatria	574	574	Bisi	Mario	MI	B
C	Medicina	530	530	Belli	Nicola	TO	C
L	Lab-Analisi	530	530	Belli	Nicola	TO	C
R	Radiologia	405	405	Mizzi	Nicola	AT	R

Esempio più complesso

Elenco dei reparti con le informazioni del primario solo nel caso in cui il primario afferisce al reparto che dirige

COD	Nome-Rep	Primario	MATR	Cognome	Nome	Residenza	Reparto
A	Chirurgia	203	203	Neri	Piero	AL	A
B	Pediatria	574	574	Bisi	Mario	MI	B
C	Medicina	530	530	Belli	Nicola	TO	C
L	Lab-Analisi	530	530	Belli	Nicola	TO	C
R	Radiologia	405	405	Mizzi	Nicola	AT	R

Esempio più complesso

Elenco dei reparti con le informazioni del primario solo nel caso in cui il primario afferisce al reparto che dirige

reparti ⋈_{Primario=MATR ∧ COD=Reparto} *medici*

COD	Nome-Rep	Primario	MATR	Cognome	Nome	Residenza	Reparto
A	Chirurgia	203	203	Neri	Piero	AL	A
B	Pediatria	574	574	Bisi	Mario	MI	B
C	Medicina	530	530	Belli	Nicola	TO	C
R	Radiologia	405	405	Mizzi	Nicola	AT	R

Altro esempio

Elenco dei reparti abbinati ai dati dei rispettivi primari solo nel caso in cui il primario non vi afferisca

reparti ⋈_{Primario=MATR ∧ COD ≠ Reparto} *medici*

COD	Nome-Rep	Primario	MATR	Cognome	Nome	Residenza	Reparto
L	Lab-Analisi	530	530	Belli	Nicola	TO	C

Generalità del Θ -join

Il Θ -join considera congiunzioni di **qualsiasi tipo** di confronti

Cardinalità del Θ -join

La cardinalità di $r_1(A) \bowtie_{\theta} r_2(B)$ si calcola considerando la cardinalità di $\sigma_{\theta}(r_1(A) \times r_2(B))$

$$0 \leq |\sigma_{\theta}(r_1(A) \times r_2(B))| \leq |r_1(A) \times r_2(B)|$$

Ma sappiamo che $|r_1(A) \times r_2(B)| = |r_1(A)| \cdot |r_2(B)|$, quindi

$$0 \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)| \cdot |r_2(B)|$$

N.B.: la cardinalità del Θ -join ha un range molto ampio!

Proiezione e join

reparti ⋈_{Primario=MATR} *medici*

COD	Nome-Rep	Primario	MATR	Cognome	Nome	Residenza	Reparto
A	Chirurgia	203	203	Neri	Piero	AL	A
B	Pediatria	574	574	Bisi	Mario	MI	B
C	Medicina	530	530	Belli	Nicola	TO	C
L	Lab-Analisi	530	530	Belli	Nicola	TO	C
R	Radiologia	405	405	Mizzi	Nicola	AT	R

Non è utile avere due volte l'informazione sulla matricola del medico

Proiezione e join

Elenco solo codice e nome del reparto e cognome del
primario

$\Pi_{COD, Nome-Rep, Cognome} (reparti \bowtie_{Primario=MATR} medici)$

COD	Nome-Rep	Cognome
A	Chirurgia	Neri
B	Pediatria	Bisi
C	Medicina	Belli
L	Lab-Analisi	Belli
R	Radiologia	Mizzi

Join con schemi non disgiunti

Esempio: elencare i pazienti abbinati ai medici che risiedono nella stessa città

pazienti ⋈_{Residenza=Residenza} *medici*

C'è un'ambiguità sull'attributo residenza!

Posso però ridenominare gli attributi omonimi, ad esempio:

$\rho_{CognomeP, NomeP, ResidenzaP} \leftarrow Cognome, Nome, Residenza (pazienti)$

Join con schemi non disgiunti

Posso anche ridenominare gli attributi omonimi in medici, ad esempio:

$\rho_{CognomeM, NomeM, ResidenzaM} \leftarrow Cognome, Nome, Residenza (medici)$

Quindi ora posso scrivere:

$\rho_{CognomeP, NomeP, ResidenzaP} \leftarrow Cognome, Nome, Residenza (pazienti)$

$\bowtie_{ResidenzaP=ResidenzaM}$

$\rho_{CognomeM, NomeM, ResidenzaM} \leftarrow Cognome, Nome, Residenza (medici)$

Join con schemi non disgiunti

$\rho_{\text{CognomeP, NomeP, ResidenzaP} \leftarrow \text{Cognome, Nome, Residenza}}(\text{pazienti})$

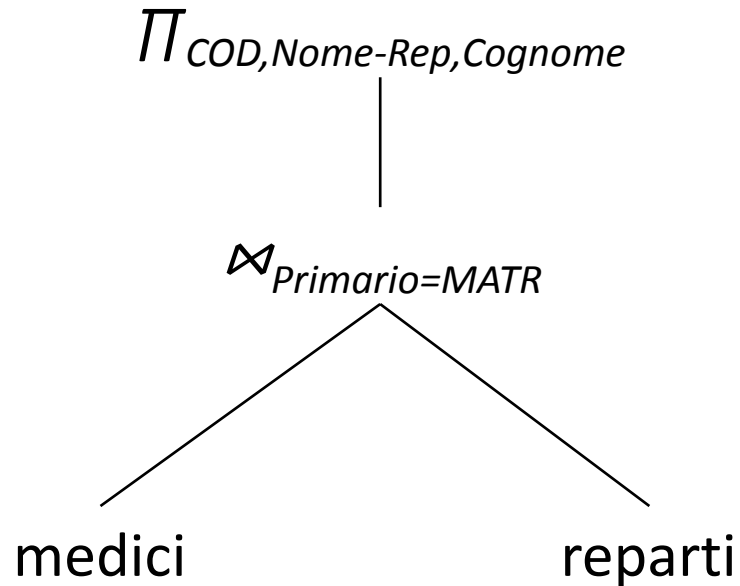
$\bowtie_{\text{ResidenzaP}=\text{ResidenzaM}}$

$\rho_{\text{CognomeM, NomeM, ResidenzaM} \leftarrow \text{Cognome, Nome, Residenza}}(\text{medici})$

COD	CognomeP	NomeP	ResidenzaP	AnnoNascita	MATR	CognomeM	NomeM	ResidenzaM	Reparto
A102	Necchi	Luca	TO	1950	461	Bargio	Sergio	TO	B
A102	Necchi	Luca	TO	1950	530	Belli	Nicola	TO	C
B543	Missoni	Nadia	TO	1960	461	Bargio	Sergio	TO	B
B543	Missoni	Nadia	TO	1960	530	Belli	Nicola	TO	C
B444	Missoni	Luigi	VC	2000	501	Monti	Mario	VC	A
S555	Rossetti	Gino	AT	2010	405	Mizzi	Nicola	AT	R

Notazione ad albero

$\Pi_{COD, Nome-Rep, Cognome} (reparti \bowtie_{Primario=MATR} medici)$



Join e selezione

Elencare i pazienti ricoverati in chirurgia

- il nome del reparto è contenuto in *reparti*
- il codice dei pazienti ricoverati è in *ricoveri*

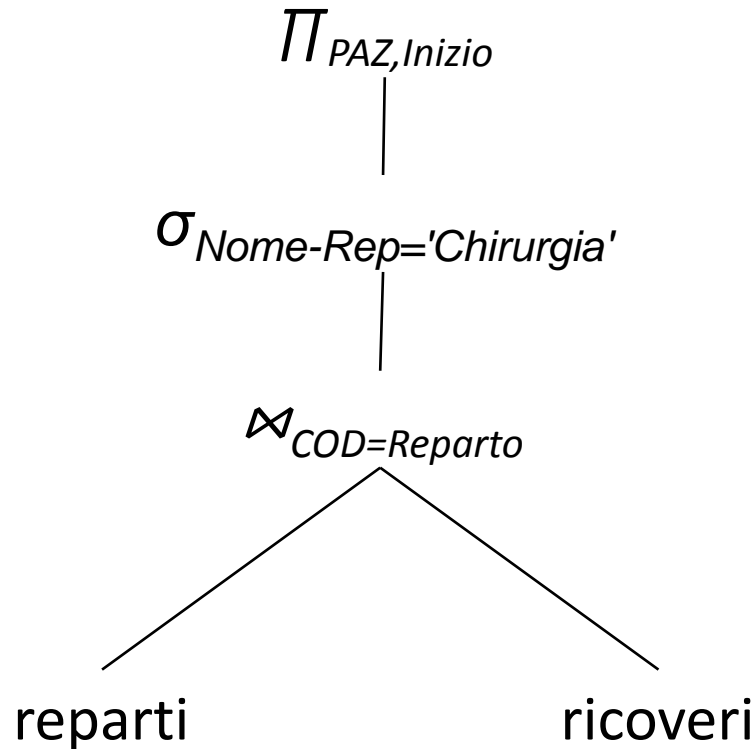
$$\sigma_{\text{Nome-Rep}='Chirurgia'}(\text{reparti} \bowtie_{\text{COD=Reparto}} \text{ricoveri})$$

Posso limitare il numero di informazioni richieste

$$\Pi_{\text{PAZ,Inizio}}(\sigma_{\text{Nome-Rep}='Chirurgia'}(\text{reparti} \bowtie_{\text{COD=Reparto}} \text{ricoveri}))$$

Join e selezione

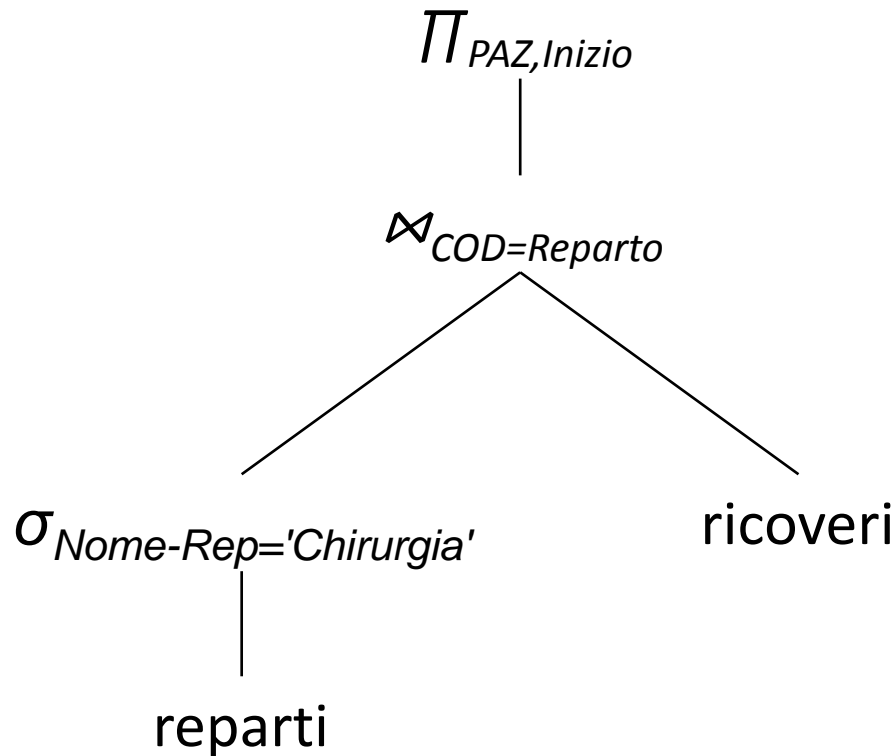
$\Pi_{PAZ, Inizio}(\sigma_{Nome-Rep='Chirurgia'}(reparti \bowtie_{COD=Reparto} ricoveri))$



Join e selezione (alternativa)

Possiamo anticipare la selezione

$$\Pi_{PAZ, Inizio}(\sigma_{Nome-Rep='Chirurgia'}(reparti) \bowtie_{COD=Reparto} ricoveri)$$



Equivalenza delle espressioni

Le due espressioni

$$\Pi_{PAZ, Inizio}(\sigma_{Nome-Rep='Chirurgia'}(reparti \bowtie_{COD=Reparto} ricoveri))$$

e

$$\Pi_{PAZ, Inizio}(\sigma_{Nome-Rep='Chirurgia'}(reparti) \bowtie_{COD=Reparto} ricoveri)$$

sono perfettamente **equivalenti**!

Esempio

Quali medici hanno avuto in cura il paziente Missoni Luigi?

- Per conoscere il nome dei pazienti ho bisogno della relazione pazienti

$$\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti)$$

Esempio

Quali medici hanno avuto in cura il paziente Missoni Luigi?

- Per sapere dov'è stato ricoverato ho bisogno della relazione ricoveri

$\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti) \bowtie_{COD=PAZ} ricoveri$

Esempio

Quali medici hanno avuto in cura il paziente Missoni Luigi?

- Per conoscere i medici afferenti al reparto ho bisogno della relazione *medici*
- La relazione *medici* ha **molti attributi omonimi** ad attributi della relazione virtuale prodotta da:

$\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti) \bowtie_{COD=PAZ} ricoveri$

Esempio

Quali medici hanno avuto in cura il paziente Missoni Luigi?

- Ridenomino gli attributi di medici:

$\rho_{CM,NM,RM,Rep} \leftarrow \text{Cognome, Nome, Residenza, Reparto}(\text{medici})$

Esempio

Quali medici hanno avuto in cura il paziente Missoni Luigi?

- Ora posso mettere in join le due espressioni

$(\sigma_{\text{Cognome}='Missoni' \wedge \text{Nome}='Luigi'}(\text{pazienti}) \bowtie_{\text{COD}=\text{PAZ}} \text{ricoveri})$
 $\bowtie_{\text{Reparto}=\text{Rep}}$

$\rho_{\text{CM},\text{NM},\text{RM},\text{Rep}} \leftarrow \text{Cognome}, \text{Nome}, \text{Residenza}, \text{Reparto}(\text{medici})$

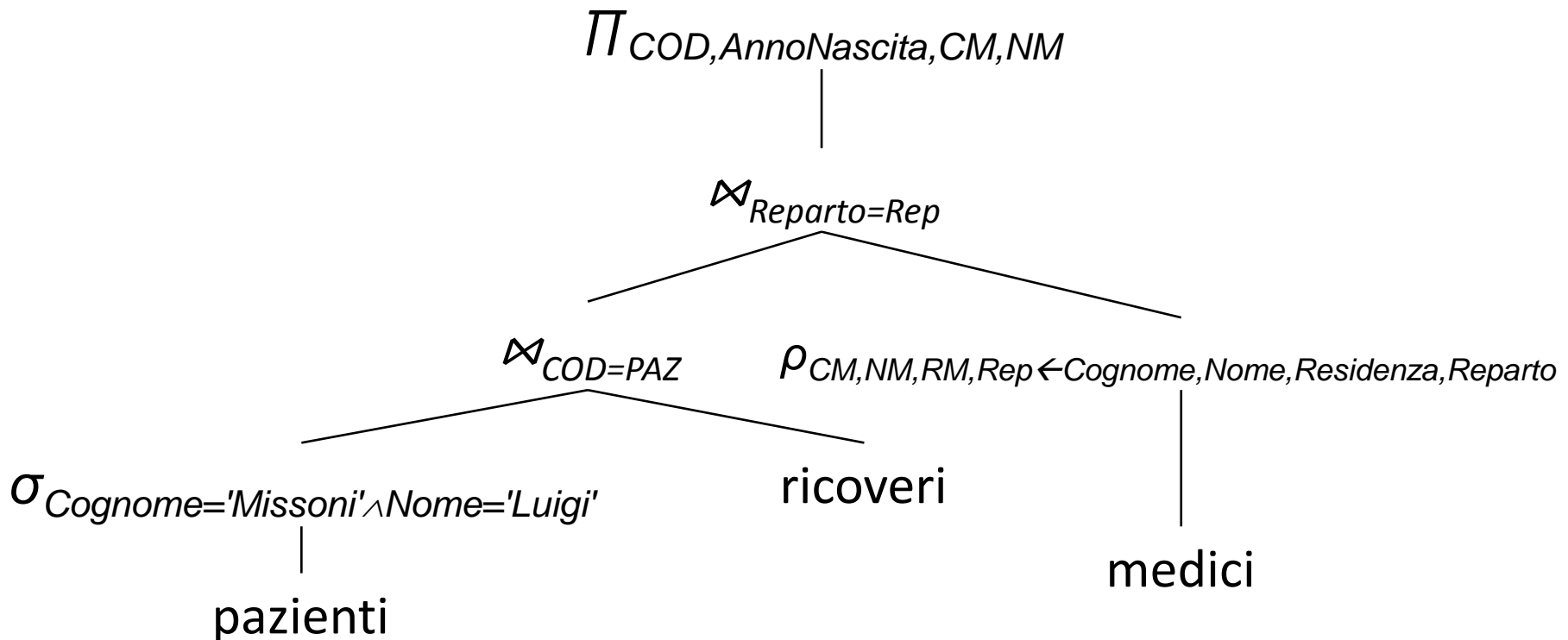
Esempio

Quali medici hanno avuto in cura il paziente Missoni Luigi?

- Infine, proietto sui dati che mi interessano, ad esempio

$$\begin{aligned} & \Pi_{COD, AnnoNascita, CM, NM} ((\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti) \\ & \quad \bowtie_{COD=PAZ} ricoveri) \bowtie_{Reparto=Rep} \\ & \quad \rho_{CM, NM, RM, Rep} \leftarrow \text{Cognome, Nome, Residenza, Reparto} (medici)) \end{aligned}$$

Esempio

$$\Pi_{COD, AnnoNascita, CM, NM}((\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti) \bowtie_{COD=PAZ} ricoveri) \bowtie_{Reparto=Rep} \rho_{CM, NM, RM, Rep \leftarrow Cognome, Nome, Residenza, Reparto}(medici))$$


Semplificazione della ridenominazione

- L'operatore di ridenominazione appesantisce la lettura delle espressioni algebriche
- Quando ho attributi con stesso nome su relazioni diverse, posso considerare il nome dell'attributo specificando la tavola a cui appartiene
- Possiamo utilizzare la **dot-notation**:

$$\begin{aligned} & \Pi_{COD, AnnoNascita, \textit{medici.Cognome}, \textit{medici.Nome}} \\ & ((\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(\textit{pazienti}) \bowtie_{COD=PAZ} \textit{ricoveri}) \\ & \quad \bowtie_{\textit{ricoveri.Reparto}=\textit{medici.Reparto}} (\textit{medici})) \end{aligned}$$

Ridenominazione e dot-notation

In pratica si sottintende un'operazione del tipo:

$\rho_{medici.Cognome, medici.Nome, medici.Residenza, medici.Reparto} \leftarrow Cognome,$
 $Nome, Residenza, Reparto$

Esercizio

Trovare cognome e nome del primario del reparto in cui è stato ricoverato il paziente Missoni Luigi

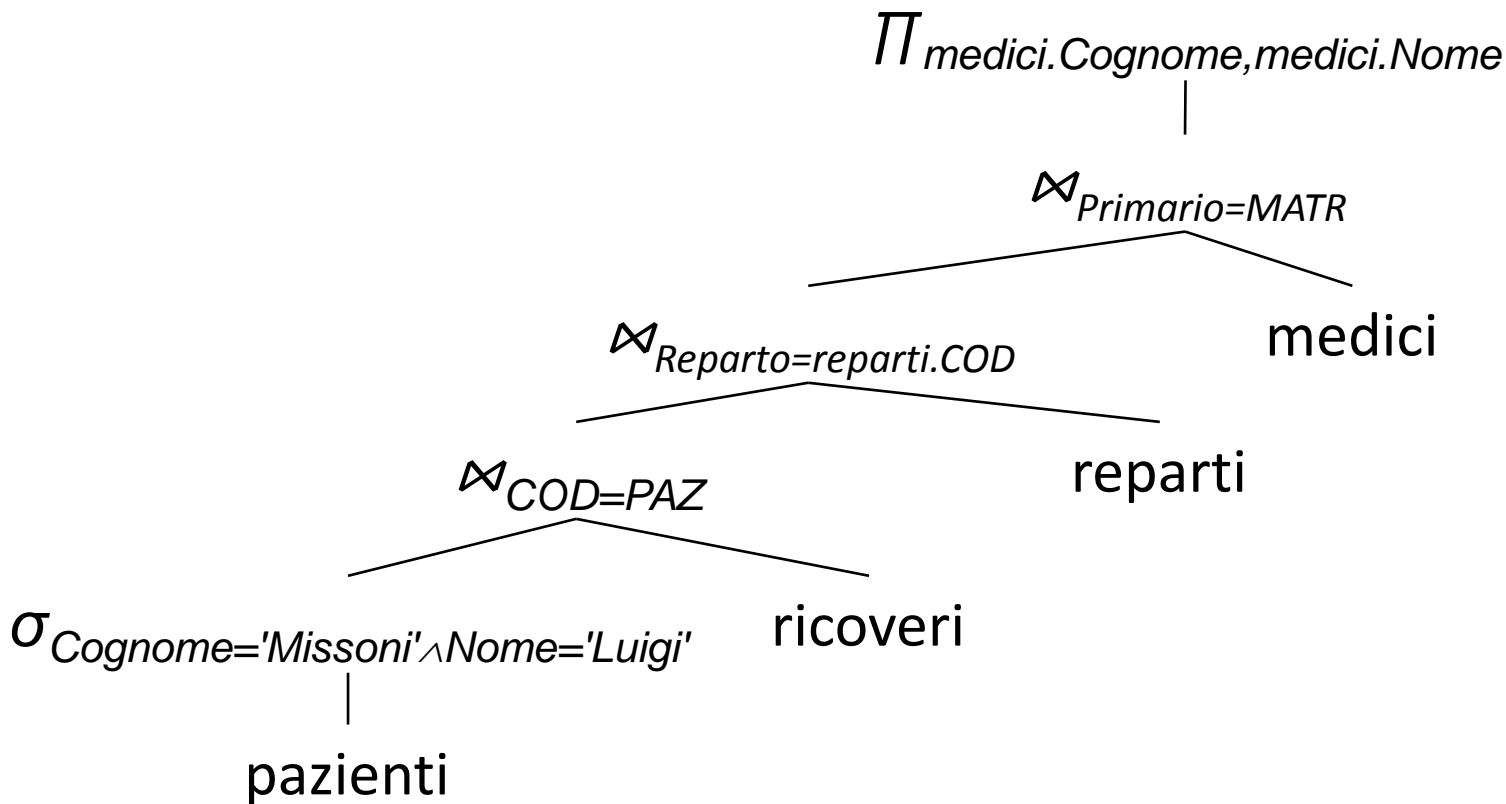
Esercizio

Trovare cognome e nome del primario del reparto in cui è stato ricoverato il paziente Missoni Luigi

$$\begin{aligned} & \Pi_{\text{medici.Cognome, medici.Nome}} \\ & (((\sigma_{\text{Cognome}='Missoni' \wedge \text{Nome}='Luigi'}(\text{pazienti}) \bowtie_{\text{COD}=\text{PAZ}} \text{ricoveri}) \\ & \bowtie_{\text{Reparto}=\text{reparti.COD}} \text{reparti}) \bowtie_{\text{Primario}=\text{MATR}} \text{medici}) \end{aligned}$$

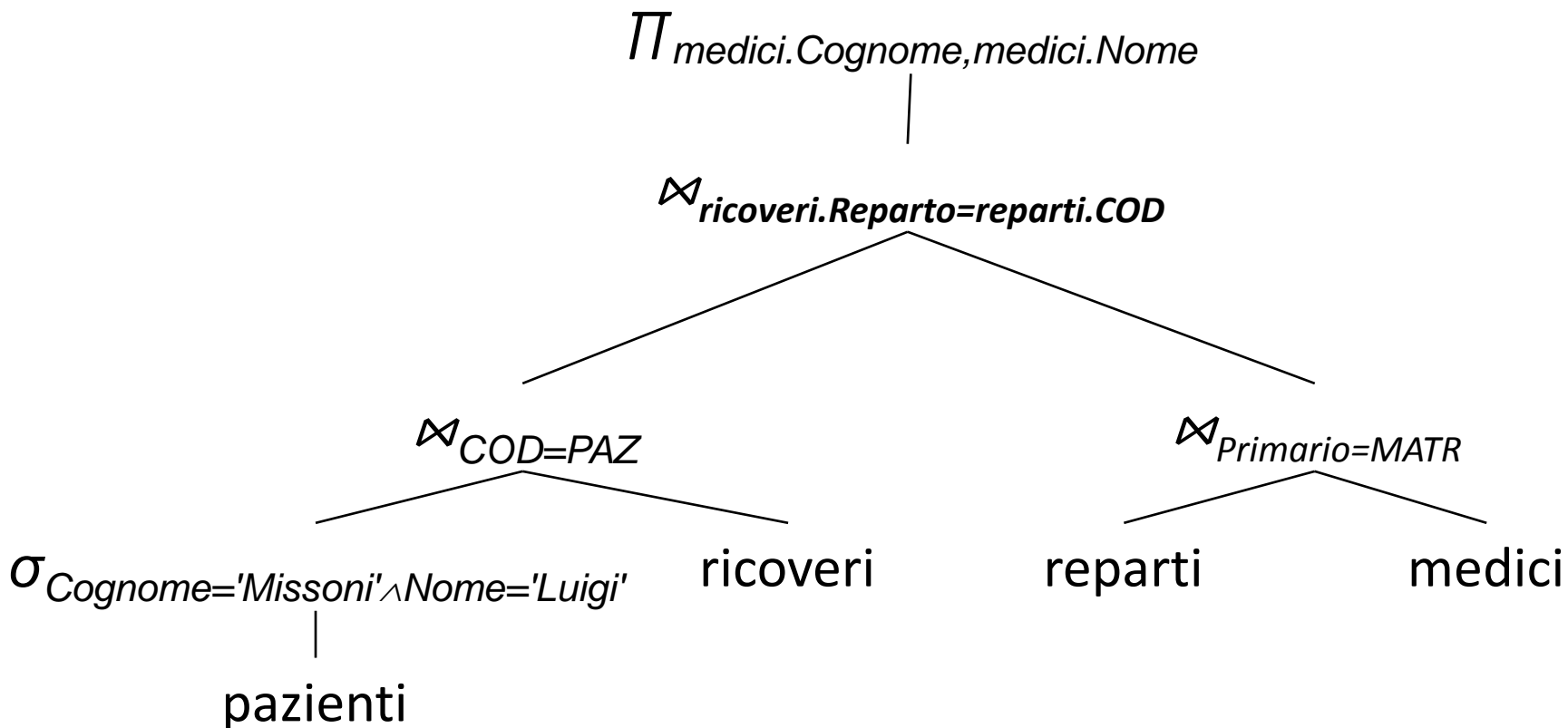
Esercizio

Trovare cognome e nome del primario del reparto in cui è stato ricoverato il paziente Missoni Luigi



Esercizio (soluzione alternativa)

Trovare cognome e nome del primario del reparto in cui è stato ricoverato il paziente Missoni Luigi



Interrogazioni con conteggi

- In algebra relazionale non è possibile effettuare interrogazioni che prevedano conteggi
 - esempio: quante volte sono stati ricoverati i pazienti?
- In SQL è possibile attraverso operatori specifici
- In algebra relazione è possibile però rispondere a domande del tipo:
 - Elencare i pazienti che hanno subito due o più ricoveri

Esempio

Elencare i pazienti che hanno subito due o più ricoveri

Intuizione: se un paziente è ricoverato più volte, sarà presente più volte nella tavola *ricoveri* con date di inizio ricovero diverse

Come confrontiamo tuple della stessa tavola?

Esempio

Elencare i pazienti che hanno subito due o più ricoveri

Posso mettere la tavola ricoveri in join con sé stessa

$$\text{ricoveri} \bowtie_{\theta} \text{ricoveri}$$

Ma devo prima rinominare tutti gli attributi!

$$\rho_{PAZ1, Inizio1, Fine1, Reparto1} \leftarrow PAZ, Inizio, Fine, Reparto \text{ricoveri} \bowtie_{\theta}$$

$$\rho_{PAZ2, Inizio2, Fine2, Reparto2} \leftarrow PAZ, Inizio, Fine, Reparto \text{ricoveri}$$

Esempio

Elencare i pazienti che hanno subito due o più ricoveri

Posso ora esplicitare il Θ

$\rho_{PAZ1, Inizio1, Fine1, Reparto1} \leftarrow PAZ, Inizio, Fine, Reparto$ (*ricoveri*)

$\bowtie_{PAZ1=PAZ2 \wedge Inizio1 \neq Inizio2}$

$\rho_{PAZ2, Inizio2, Fine2, Reparto2} \leftarrow PAZ, Inizio, Fine, Reparto$ (*ricoveri*)

Esempio

Elencare i pazienti che hanno subito due o più ricoveri

$\rho_{PAZ1, Inizio1, Fine1, Reparto1} \leftarrow PAZ, Inizio, Fine, Reparto$ (*ricoveri*)

$\bowtie_{PAZ1=PAZ2 \wedge Inizio1 \neq Inizio2}$

$\rho_{PAZ2, Inizio2, Fine2, Reparto2} \leftarrow PAZ, Inizio, Fine, Reparto$ (*ricoveri*)

PAZ1	Inizio1	Fine1	Reparto1	PAZ2	Inizio2	Fine2	Reparto2
A102	2/05/2014	9/05/2014	A	A102	2/12/2004	2/01/2005	A
A102	2/12/2004	2/01/2005	A	A102	2/05/2014	9/05/2014	A
S555	5/10/2014	3/12/2014	B	S555	6/09/2015	1/11/2015	A
S555	6/09/2015	1/11/2015	A	S555	5/10/2014	3/12/2014	B

Esempio (raffinamento)

Elencare i pazienti che hanno subito due o più ricoveri

$\rho_{PAZ1, Inizio1, Fine1, Reparto1} \leftarrow PAZ, Inizio, Fine, Reparto$ (*ricoveri*)

$\bowtie_{PAZ1=PAZ2 \wedge Inizio1 < Inizio2}$

$\rho_{PAZ2, Inizio2, Fine2, Reparto2} \leftarrow PAZ, Inizio, Fine, Reparto$ (*ricoveri*)

PAZ1	Inizio1	Fine1	Reparto1	PAZ2	Inizio2	Fine2	Reparto2
A102	2/05/2014	9/05/2014	A	A102	2/12/2004	2/01/2005	A
S555	5/10/2014	3/12/94	B	S555	6/09/2015	1/11/2015	A

Esempio

(ridenominazione dello schema)

Elencare i pazienti che hanno subito due o più ricoveri

$$\rho_{RICOVERI1} \leftarrow_{RICOVERI} (ricoveri)$$

$$\bowtie_{ricoveri1.PAZ=ricoveri2.PAZ \wedge ricoveri1.Inizio < ricoveri2.Inizio}$$

$$\rho_{RICOVERI2} \leftarrow_{RICOVERI} (ricoveri)$$

Esercizio

Elencare cognome e nome dei pazienti ricoverati più di una volta

Esercizio

Elencare nome e cognome dei pazienti residenti in città in cui risiede almeno un medico

Base di Dati "Impiegati"

IMPIEGATI

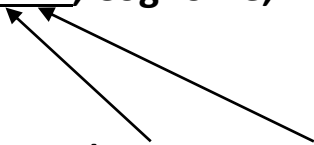
<u>MATR</u>	Cognome	Nome	Età	Stipendio
203	Neri	Piero	50	40
574	Bisi	Mario	60	60
461	Bargio	Sergio	30	61
530	Belli	Nicola	40	38
405	Mizzi	Nicola	55	60
501	Monti	Mario	25	35

ORGANIGRAMMA

Capo	Impiegato
203	405
203	501
574	203
574	530
405	461

IMPIEGATI(MATR, Cognome, Nome, Età, Stipendio)

ORGANIGRAMMA(Capo, Impiegato)



Esempio

Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

Esempio

Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

- occorre mettere in join la tavola *impiegati* con *organigramma* per conoscere lo stipendio dei capi
- per conoscere lo stipendio dei subalterni devo mettere in join un'altra volta la tavola *impiegati*

Esempio

Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

Dovendo usare due volte la tavola impiegati, devo rinominarla:

$$\rho_{CAPI} \leftarrow_{IMPIEGATI} (impiegati)$$

Esempio

Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

- occorre mettere in join la tavola *impiegati* con *organigramma* per conoscere lo stipendio dei capi

$$\rho_{CAPI \leftarrow IMPIEGATI}(impiegati) \bowtie_{CAPI.MATR = CAPO} (organigramma)$$

Esempio

Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

- per conoscere lo stipendio dei subalterni devo mettere in join un'altra volta la tavola *impiegati*

$$\begin{aligned} &(\rho_{CAPI \leftarrow IMPIEGATI}(\textit{impiegati}) \bowtie_{CAPI.MATR = CAPO}(\textit{organigramma})) \\ &\quad \bowtie_{IMPIEGATO = IMPIEGATI.MATR}(\textit{impiegati}) \end{aligned}$$

Esempio

Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

- Imponiamo ora che lo stipendio del subalterno sia superiore allo stipendio del capo

$$\begin{aligned} \sigma_{\text{capi.Stipendio} < \text{impiegati.Stipendio}} & (\rho_{\text{CAPI}} \leftarrow \text{IMPIEGATI}(\text{impiegati})) \\ \bowtie_{\text{capi.MATR} = \text{Capo}} & (\text{organigramma}) \\ \bowtie_{\text{Impiegato} = \text{impiegati.MATR}} & (\text{impiegati}) \end{aligned}$$

Esempio

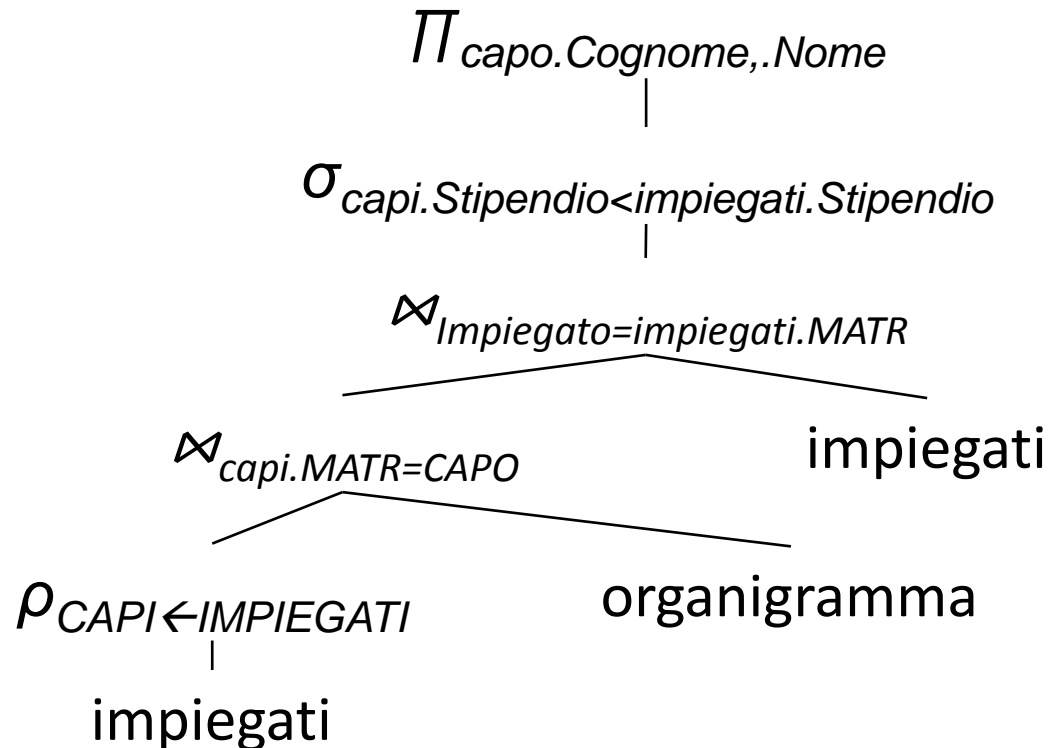
Elencare i capi che guadagnano meno di almeno uno dei suoi subalterni

- proiettiamo solo le informazioni necessarie

$$\begin{aligned} & \Pi_{capi.Cognome, capi.Nome} (\sigma_{capi.Stipendio < impiegati.Stipendio} \\ & ((\rho_{CAPI} \leftarrow IMPIEGATI(impiegati) \bowtie_{capi.MATR = Capo} (organigramma)) \\ & \quad \bowtie_{Impiegato = impiegati.MATR} (impiegati))) \end{aligned}$$

Esempio (albero sintattico)

$$\Pi_{capi.Cognome, capi.Nome}(\sigma_{capi.Stipendio < impiegati.Stipendio}((\rho_{CAPI \leftarrow IMPIEGATI}(impiegati) \bowtie_{capi.MATR = Capo}(organigramma)) \bowtie_{Impiegato = impiegati.MATR}(impiegati)))$$



Casi particolari di join

Ricordiamo la cardinalità del join

$$0 \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)| \cdot |r_2(B)|$$

In alcuni casi si può cercare di limitare l'ampiezza del range della cardinalità, ma bisogna specificare meglio il Θ -join, ovvero abbandonare la generalità del Θ

Equi-join: Θ_e -join

L'Equi-join è un caso particolare del Θ -join quando i confronti sono solo ugualianze

Date $r_1(A)$ e $r_2(B)$ la condizione (predicato) di join Θ_e da intendersi come una **congiunzione di ugualianze** del tipo $A_i = B_j$

L'equi-join (Θ_e -join) si rappresenta come

$$r_1(A) \bowtie_{\Theta_e} r_2(B)$$

Casi particolari

Caso 1

- r_1 con partecipazione completa al join

Esaminando $r_1(A) \bowtie_{\theta_e} r_2(B)$, mi rendo conto che l'istanza della base dati è tale per cui, ogni tupla di r_1 trova almeno una corrispondenza in r_2

In questo caso la cardinalità sarà

$$|r_1(A)| \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)| \cdot |r_2(B)|$$

Casi particolari

Caso 2

- r_1 in equi-join con r_2 in corrispondenza della chiave principale di r_2

Consideriamo $r_1(..., B, ...)$ ed $r_2(..., PK, ...)$ e immaginiamo una condizione di equi-join come una congiunzione di $B_i = PK_j$ dove $B_i \in B$ e $PK_j \in PK$

Casi particolari

Caso 2

- r_1 in equi-join con r_2 in corrispondenza della chiave principale di r_2

Non facciamo nessuna assunzione sull'insieme B
(potrebbe non esserci un vincolo di integrità referenziale)

Per ogni tupla di r_1 , si troverà al più **una** corrispondenza con r_2 , altrimenti sarebbe violato il **vincolo di chiave primaria!**

Casi particolari

Caso 2

- r_1 in equi-join con r_2 in corrispondenza della chiave principale di r_2

Per ogni tupla di r_1 , si troverà al più **una** corrispondenza con r_2 , altrimenti sarebbe violato il **vincolo di chiave primaria**!

Quindi:

$$0 \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)|$$

Casi particolari

Caso 3

- r_1 in equi-join con r_2 in corrispondenza della chiave principale di r_2 ed esiste un vincolo di integrità referenziale tra gli attributi della condizione

Consideriamo $r_1(\dots, B, \dots)$ ed $r_2(\dots, PK, \dots)$ e immaginiamo una condizione di equi-join come una congiunzione di $B_i = PK_j$ dove $B_i \in B$ e $PK_j \in PK$.

Inoltre vale il vincolo di integrità referenziale seguente

$$\forall t_i (t_i \in r_1) \Rightarrow \exists t_j (t_j \in r_2) \wedge (t_i[B] = t_j[PK])$$

Casi particolari

Caso 3

- r_1 in equi-join con r_2 in corrispondenza della chiave principale di r_2 ed esiste un vincolo di integrità referenziale tra gli attributi della condizione

In questo caso, ogni tupla di r_1 troverà esattamente una corrispondenza in r_2

$$|r_1(A) \bowtie_{\theta} r_2(B)| = |r_1(A)|$$

Casi particolari

Caso 3

Dimostrazione (1/3):

Vale il vincolo di integrità referenziale, quindi per ogni tupla t di r_1 con $t[B]=X$ ci sarà una tupla t in r_2 con $t[PK]=X$, quindi r_1 partecipa in modo completo al join con r_2 (**caso 1**) e avremo:

$$|r_1(A)| \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)| \cdot |r_2(B)|$$

Casi particolari

Caso 3

Dimostrazione (2/3):

Inoltre, l'equi-join è in corrispondenza con gli attributi della chiave principale di r_2 (**caso 2**), quindi vale anche:

$$0 \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)|$$

Casi particolari

Caso 3

Dimostrazione (3/3):

Quindi le due disuguaglianze

$$|r_1(\mathbf{A})| \leq |r_1(\mathbf{A}) \bowtie_{\theta} r_2(\mathbf{B})| \leq |r_1(\mathbf{A})| \cdot |r_2(\mathbf{B})|$$

e

$$0 \leq |r_1(\mathbf{A}) \bowtie_{\theta} r_2(\mathbf{B})| \leq |r_1(\mathbf{A})|$$

sono contemporaneamente soddisfatte solo se

$$|r_1(\mathbf{A}) \bowtie_{\theta} r_2(\mathbf{B})| = |r_1(\mathbf{A})|$$

Esercizio

Nell'equi-join tra *ricoveri* e *pazienti* su $PAZ=COD$, quante tuple si ottengono?

Casi particolari

Caso 1 (altra definizione, per completezza)

- con partecipazione completa al join

Esaminando $r_1(A) \bowtie_{\theta_e} r_2(B)$, mi rendo conto che l'istanza della base dati è tale per cui, ogni tupla di r_1 trova almeno una corrispondenza in r_2 **e viceversa**

In questo caso la cardinalità sarà

$$\max\{|r_1(A)|, |r_2(A)|\} \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)| \cdot |r_2(B)|$$

Natural-join

Cambio leggermente l'esempio

PAZIENTI

<u>COD</u>	Cognome	Nome	Residenza	AnnoNascita
A102	Necchi	Luca	TO	1950
B372	Rossigni	Piero	NO	1940
B543	Missoni	Nadia	TO	1960
B444	Missoni	Luigi	VC	2000
S555	Rossetti	Gino	AT	2010

REPARTI

<u>Rep</u>	Nome-Rep	MATR
A	Chirurgia	203
B	Pediatria	574
C	Medicina	530
L	Lab-Analisi	530
R	Radiologia	405

RICOVERI

COD	Inizio	Fine	Rep
A102	2/05/2014	9/05/2014	A
A102	2/12/2004	2/01/2005	A
S555	5/10/2014	3/12/2014	B
B444	1/12/2004	2/01/2005	B
S555	6/09/2015	1/11/2015	A

MEDICI

<u>MATR</u>	Cognome	Nome	Residenza	Rep
203	Neri	Piero	AL	A
574	Bisi	Mario	MI	B
461	Bargio	Sergio	TO	B
530	Belli	Nicola	TO	C
405	Mizzi	Nicola	AT	R
501	Monti	Mario	VC	A

Nome degli attributi

E' perfettamente normale utilizzare **attributi omonimi** per esprimere **concetti identici** (codice dei pazienti, matricola dei medici, codice dei reparti)

Abbiamo usato nomi diversi solo per motivi didattici (per non introdurre subito la ridenominazione)

Capita spesso di mettere in equi-join tavole su attributi omonimi (esempio: *ricoveri.COD=pazienti.COD*)

Natural-join

Date:

- due relazioni $r_1(A)$ ed $r_2(B)$ con $A=\{\mathbf{X},Y\}$ e $B=\{\mathbf{X},Z\}$ (X,Y e Z sono insiemi disgiunti di attributi), e $Y \cap Z = \emptyset$

Il **natural-join** è definito come

$$r_1(A) \bowtie r_2(B) = \prod_{X,Y,Z} (r_1 \bowtie_{\theta_e} \rho_{X' \leftarrow X}(r_2))$$

dove

$$\theta_e = (X_i = X'_i) \wedge \dots \wedge (X_k = X'_k)$$

Esempio

ricoveri ✕ pazienti

COD	Inizio	Fine	Rep	Cognome	Nome	Residenza	AnnoNascita
A102	2/05/2014	9/05/2014	A	Necchi	Luca	TO	1950
A102	2/12/2004	2/01/2005	A	Necchi	Luca	TO	1950
S555	5/10/2014	3/12/2014	B	Rossetti	Gino	AT	2010
B444	1/12/2004	2/01/2005	B	Missoni	Luigi	VC	2000
S555	6/09/2015	1/11/2015	A	Rossetti	Gino	AT	2010

Esempio

ricoveri \bowtie reparti

COD	Inizio	Fine	Rep	Nome-Rep	MATR
A102	2/05/2014	9/05/2014	A	Chirurgia	203
A102	2/12/2004	2/01/2005	A	Chirurgia	203
S555	5/10/2014	3/12/2014	B	Pediatria	574
B444	1/12/2004	2/01/2005	B	Pediatria	574
S555	6/09/2015	1/11/2015	A	Chirurgia	203

Esempio

ricoveri ∞ pazienti ∞ reparti

COD	Inizio	Fine	Rep	Cognome	Nome	Residenza	AnnoNascita	Nome-Rep	MATR
A102	2/05/2014	9/05/2014	A	Necchi	Luca	TO	1950	Chirurgia	203
A102	2/12/2004	2/01/2005	A	Necchi	Luca	TO	1950	Chirurgia	203
S555	5/10/2014	3/12/2014	B	Rossetti	Gino	AT	2010	Pediatria	574
B444	1/12/2004	2/01/2005	B	Missoni	Luigi	VC	2000	Pediatria	574
S555	6/09/2015	1/11/2015	A	Rossetti	Gino	AT	2010	Chirurgia	203

Uso corretto del natural-join

reparti ⋈ medici

A quale interrogazione sto rispondendo?

REPARTI

Rep	Nome-Rep	MATR
A	Chirurgia	203
B	Pediatria	574
C	Medicina	530
L	Lab-Analisi	530
R	Radiologia	405

MEDICI

MATR	Cognome	Nome	Residenza	Rep
203	Neri	Piero	AL	A
574	Bisi	Mario	MI	B
461	Bargio	Sergio	TO	B
530	Belli	Nicola	TO	C
405	Mizzi	Nicola	AT	R
501	Monti	Mario	VC	A

$\Pi_{Rep, Nome-Rep, MATR, Cognome, Nome, Residenza}$
(reparti ⋈_{reparti.Rep=medici.Rep ∧ reparti.MATR=medici.MATR} medici)

Uso corretto del natural-join

$\Pi_{Rep, Nome-Rep, MATR, Cognome, Nome, Residenza}$
(reparti $\bowtie_{reparti.Rep=medici.Rep \wedge reparti.MATR=medici.MATR}$ *medici)*

<u>Rep</u>	Nome-Rep	MATR	Cognome	Nome	Residenza
A	Chirurgia	203	Neri	Piero	AL
B	Pediatria	574	Bisi	Mario	MI
C	Medicina	530	Belli	Nicola	TO
L	Lab-Analisi	530	Belli	Nicola	TO
R	Radiologia	405	Mizzi	Nicola	AT

MEDICI

<u>MATR</u>	Cognome	Nome	Residenza	Rep
203	Neri	Piero	AL	A
574	Bisi	Mario	MI	B
461	Bargio	Sergio	TO	B
530	Belli	Nicola	TO	C
405	Mizzi	Nicola	AT	R
501	Monti	Mario	VC	A

Uso corretto del natural-join

reparti \bowtie medici

A quale interrogazione sto rispondendo?

<u>Rep</u>	Nome-Rep	MATR	Cognome	Nome	Residenza
A	Chirurgia	203	Neri	Piero	AL
B	Pediatria	574	Bisi	Mario	MI
C	Medicina	530	Belli	Nicola	TO
R	Radiologia	405	Mizzi	Nicola	AT

Estendere la tupla di reparto con i dati del primario corrispondente **solo nel caso** in cui il primario afferisca allo stesso reparto

Uso corretto del natural-join

Usare con cautela!

Casi limite del natural-join

Se A e B sono due schemi completamente disgiunti, cioè $A \cap B = \emptyset$, il natural-join diventa equivalente al prodotto cartesiano

$$r_1(A) \bowtie r_2(B) = r_1(A) \times r_2(B)$$

Possiamo immaginarlo come un equi-join senza alcun attributo su cui verificare l'uguaglianza (cioè, Θ_e è sempre vera)

Caso limite del natural-join

Date due r_1 ed r_2 relazioni definite sullo stesso schema A , a cosa corrisponde $r_1(A) \bowtie r_2(A)$?

Semi-join: \bowtie_{θ}

Date due relazioni $r_1(A)$ ed $r_2(B)$, definiamo il semi-join come

$$r_1(A) \bowtie_{\theta} r_2(B) = \Pi_A(r_1(A) \bowtie_{\theta} r_2(B))$$

L'operatore di semi-join funziona da selettore sulla prima tabella sfruttando la seconda tabella

Esempio

Elencare tutti i dati dei primari

$\Pi_{MATR, Cognome, Nome, Residenza, Reparto}$
 $(medici \bowtie_{MATR=Primario} reparti)$

Seleziono solo i medici che sono anche primari,
equivalente a:

$medici \bowtie_{MATR=Primario} reparti$

Esempio

Elencare tutti i dati dei pazienti ricoverati in chirurgia

$$\begin{aligned} & \Pi_{\text{pazienti.COD, Cognome, Nome, Residenza, AnnoNascita}} \\ & ((\text{pazienti} \bowtie_{\text{COD=PAZ}} \text{ricoveri}) \\ & \bowtie_{\text{Reparto=reparti.COD}} (\sigma_{\text{Nome-Rep}='Chirurgia'}(\text{reparti}))) \end{aligned}$$

equivalente a:

$$\begin{aligned} & \text{pazienti} \bowtie_{\text{COD=PAZ}} \\ & (\text{ricoveri} \bowtie_{\text{Reparto=reparti.COD}} (\sigma_{\text{Nome-Rep}='Chirurgia'}(\text{reparti}))) \end{aligned}$$

Cardinalità del semi-join

$$|r_1(A) \bowtie_{\theta} r_2(B)| = |\Pi_A(r_1(A) \bowtie_{\theta} r_2(B))|$$

La cardinalità del semi-join è uguale alla cardinalità della proiezione sullo schema A , quindi al massimo alla cardinalità della relazione costruita su A

$$0 \leq |\Pi_A(r_1(A) \bowtie_{\theta} r_2(B))| \leq |r_1(A)|$$

quindi:

$$0 \leq |r_1(A) \bowtie_{\theta} r_2(B)| \leq |r_1(A)|$$