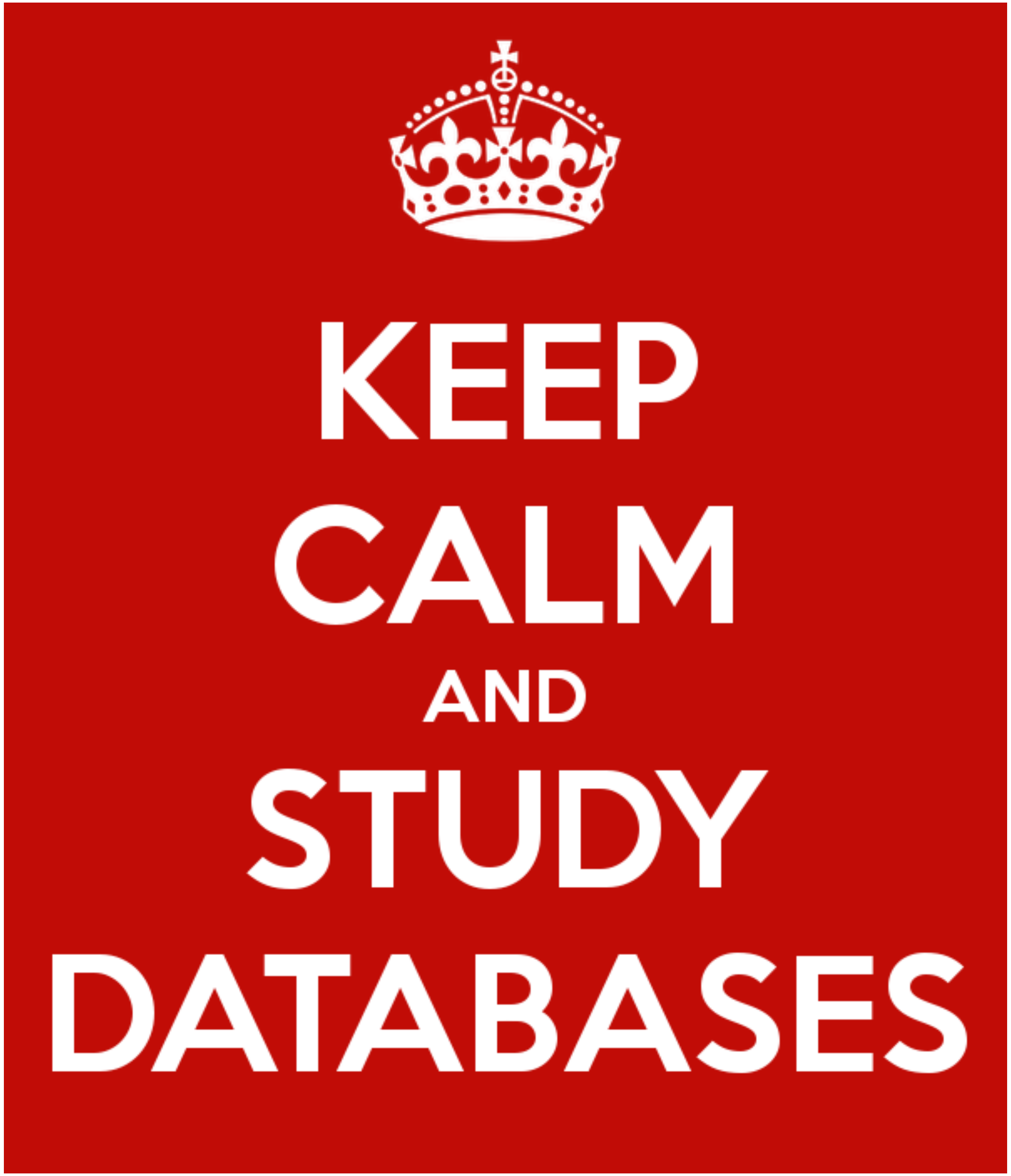
Basi di Dati
Algebra relazionale

Contenuti - Roadmap

Lab (progettazione)	Corso di Teoria	Lab (SQL)
• Metodologie e modello Entità Associazioni • Progettazione concettuale e logica	• Modello Relazionale • **Algebra relazionale** • Ottimizzazione logica • Calcolo relazionale • La normalizzazione • Metodi di accesso e indici • Gestione della concorrenza • Gestione del ripristino	• Linguaggio SQL


KEEP
CALM
AND
STUDY
DATABASES

Base di dati «Ricoveri»

- Schema relazionale con vincoli di integrità referenziale

PAZIENTI(COD, Cognome, Nome, Residenza, AnnoNascita)

MEDICI(MATR, Cognome, Nome, Residenza, Reparto)

REPARTI(COD, Nome-Rep, Primario)

RICOVERI(PAZ, Inizio, Fine, Reparto)

Base di dati «Ricoveri»

pazienti

COD	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

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

ricoveri

PAZ	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

MATR	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

Algebra relazionale

Codd formalizza le interrogazioni utilizzando un'algebra.

L'algebra relazionale:

- non «manipola» numeri ma relazioni,
- fornisce un fondamento teorico ai linguaggi di interrogazione delle basi di dati,
- le sue proprietà vengono sfruttate internamente dai DBMS per ottimizzare le query,
- è una costruzione *procedurale* dell'interrogazione, cioè un elenco dei passi da eseguire per rispondere all'interrogazione. Ogni passo è definito astrattamente attraverso gli operatori algebrici.

Operatori dell'algebra relazionale

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

- Operatori di base
  - Selezione σ_p
  - Proiezione π_A
  - Prodotto cartesiano $\times$
  - Unione $\cup$
  - Differenza $-$
  - Ridenominazione $\rho_{B \leftarrow A}$
- Operatori derivati
  - Intersezione $\cap$
  - Join (theta, equi, natural) $\bowtie$
  - Quoziente $\div$

Operatori algebrici

Ogni operatore dell’algebra relazionale riceve, come argomento, una relazione e produce, in uscita, una relazione (detta anche relazione virtuale)

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 per ora ignoreremo la possibilità che ci siano valori nulli*

Operatore di selezione

Data una relazione *r(A)* su uno schema *A*, l'operatore di selezione

$$\sigma_p(r(A))$$

produce come risultato una relazione con

- schema: A
- istanza: le tuple della relazione *r* che soddisfano il predicato *p*.

Sintassi del predicato di selezione

Il predicato p è un'espressione booleana formata componendo predicati atomici di due tipi

- $A_i \ \varphi$ *costante*
- $A_i \ \varphi \ A_j$

dove

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

- Esempi:
  - $\sigma_{Nome='Luca'}(pazienti)$
  - $\sigma_{Nome=Cognome}(pazienti)$
  - $\sigma_{Nome='Luca' \vee Nome='Chiara'}(pazienti)$

Esempi

Selezione dei pazienti con residenza a Torino (‘TO’)

Esempi

Selezione dei pazienti con residenza a Torino (‘TO’)

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

Esempi

Selezione dei pazienti con residenza a Torino ('TO')

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

Selezione dei pazienti residenti a Torino o a Vercelli ('VC')

Esempi

Selezione dei pazienti con residenza a Torino (‘TO’)

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

Selezione dei pazienti residenti a Torino o a Vercelli (‘VC’)

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

Esempi

Selezione dei pazienti con residenza a Torino ('TO')

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

Selezione dei pazienti residenti a Torino o a Vercelli ('VC')

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

Selezione dei pazienti non residenti a Torino

Esempi

Selezione dei pazienti con residenza a Torino (‘TO’)

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

Selezione dei pazienti residenti a Torino o a Vercelli (‘VC’)

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

Selezione dei pazienti non residenti a Torino

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

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

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 falso per ogni tupla
- La cardinalità è $|r(A)|$ se il predicato è vero per ogni tupla

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)* e un insieme di attributi $A_i, A_j, \ldots A_k$, tutti appartenenti ad *A*, l'operatore di proiezione

$$\pi_{A_i,\, A_j,\, \ldots\, A_k}(r(A))$$

produce come risultato una relazione con

- schema: $\{A_i, A_j, \ldots A_k\}$
- istanza: tutte le tuple della relazione argomento, ma solo rispetto agli attributi $A_i, A_j, \ldots A_k$

Esempi

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

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

COD	**Cognome**
A102	Necchi
B372	Rossigni
B543	Missoni
B444	Missoni
S555	Rossetti

Cardinalità della proiezione

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

Invece, la cardinalità della proiezione $\pi_{A_i, A_j, \ldots A_k}(r(A))$ *è*

$$0 \leq |\pi_{A_i, A_j, \ldots A_k}(r(A))| \leq |r(A)|$$

Esempio

Proiezione della relazione *pazienti* sull'attributo Cognome

$$\pi_{Cognome}(pazienti)$$

Cognome
Necchi
Rossigni
Missoni
Missoni
Rossetti

È una relazione valida?

Esempio

Proiezione della relazione *pazienti* sull'attributo Cognome

$$\pi_{Cognome}(pazienti)$$

Cognome
Necchi
Rossigni
Missoni
Missoni
Rossetti

Non è più una relazione valida nel modello relazionale teorico di Codd (infatti 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, … A_k formano una **superchiave** della relazione argomento, allora

$$|\pi_{A_i, A_j, \ldots A_k}(r(A))| = |r(A)|$$

Composizione di operatori

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

Esempio

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

- Ottenuti 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

- Ottenuti 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 di

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

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

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

$$\pi_{COD,Nome}$$

$$\sigma_{Residenza='TO' \vee Residenza='VC'}$$

$$pazienti$$

L'operatore proiezione ha solo un argomento, quindi un unico sottoalbero

L'operatore di selezione ha solo un argomento, quindi un unico sottoalbero

Le foglie dell'albero saranno sempre relazioni di base

Correttezza sintattica

Per verificare che l'espressione algebrica sia sintatticamente corretta bisogna verificare che gli operatori algebrici siano coerenti 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 gli attributi *COD* e *Nome* sono compresi nello schema restituito dalla selezione

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 aggiungere *un'ulteriore* proiezione su *COD* e *Nome*

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

- È possibile 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

Operatori insiemistici

Il risultato dell'operatore insiemistico sulle relazioni argomento $r_1(A)$ e $r_2(A)$ 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 e r_2)
  - Differenza: $r_1 - r_2$ (insieme delle tuple contenute in r_1 che non sono contenute in r_2)

Cardinalità dell'unione

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

$$max\{|r_1(A)|,|r_2(A)|\} \leq |r_1(A) \cup r_2(A)| \leq |r_1(A)|+|r_2(A)|$$

- Ovviamente, se ci sono delle tuple ripetute, la cardinalità è strettamente minore della somma delle cardinalità delle due relazioni

Esempio

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

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

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

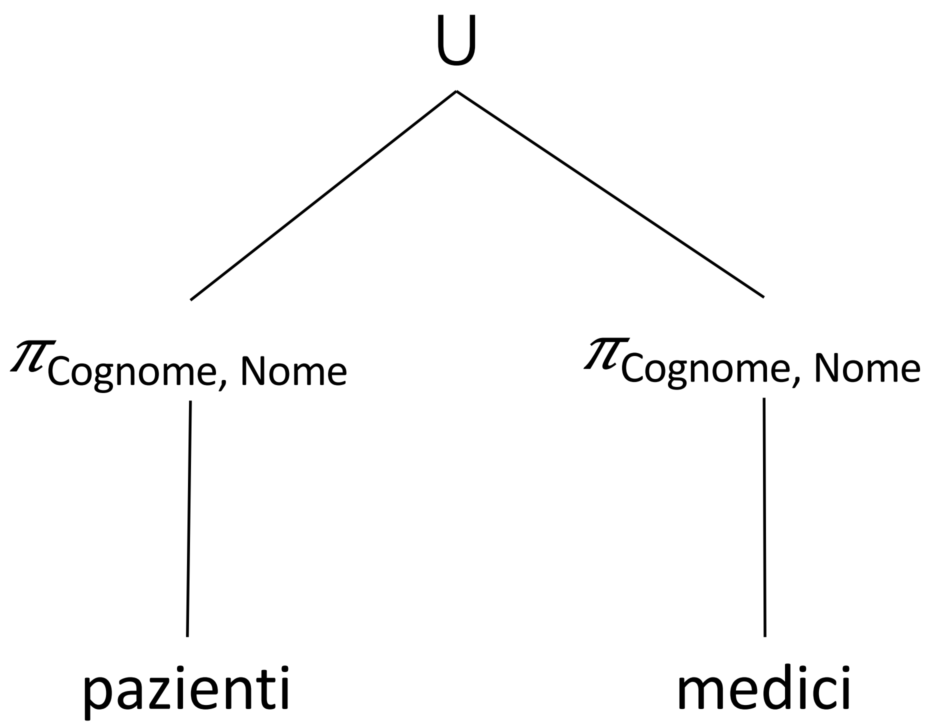
∪

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: $r_1 \cap r_2$ (insieme delle tuple di r_1 contenute anche in r_2)

Derivazione dell'intersezione

L'operatore di intersezione insiemistica può essere derivato dall'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

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

Problemi con gli operatori insiemistici

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

Medici

MATR	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

COD	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:

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

- Ma l'interrogazione è sintatticamente scorretta: lo schema del primo argomento dell'operatore differenza **è diverso** dallo schema del secondo argomento (gli attributi hanno nome diverso)!

Operatore di ridenominazione

L'operatore di **ridenominazione** ha come argomento una relazione *r(A).*

Il suo compito è semplicemente quello di cambiare nome agli attributi (una parte o tutti) della relazione argomento

Operatore di ridenominazione

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

$$\rho_{B_i,B_j,\ldots,B_k \leftarrow A_i,A_j,\ldots A_k}(r)$$

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

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

(*) cioè l'operatore **non modifica** la relazione originale nella base di dati, ma produce una nuova relazione con uno schema diverso

Esempio

Elencare i medici che non sono primari:

$$\pi_{MATR}(medici) - \rho_{\mathrm{MATR} \leftarrow \mathrm{Primario}}(\pi_{Primario}(reparti))$$

MATR
203
574
461
530
405
501

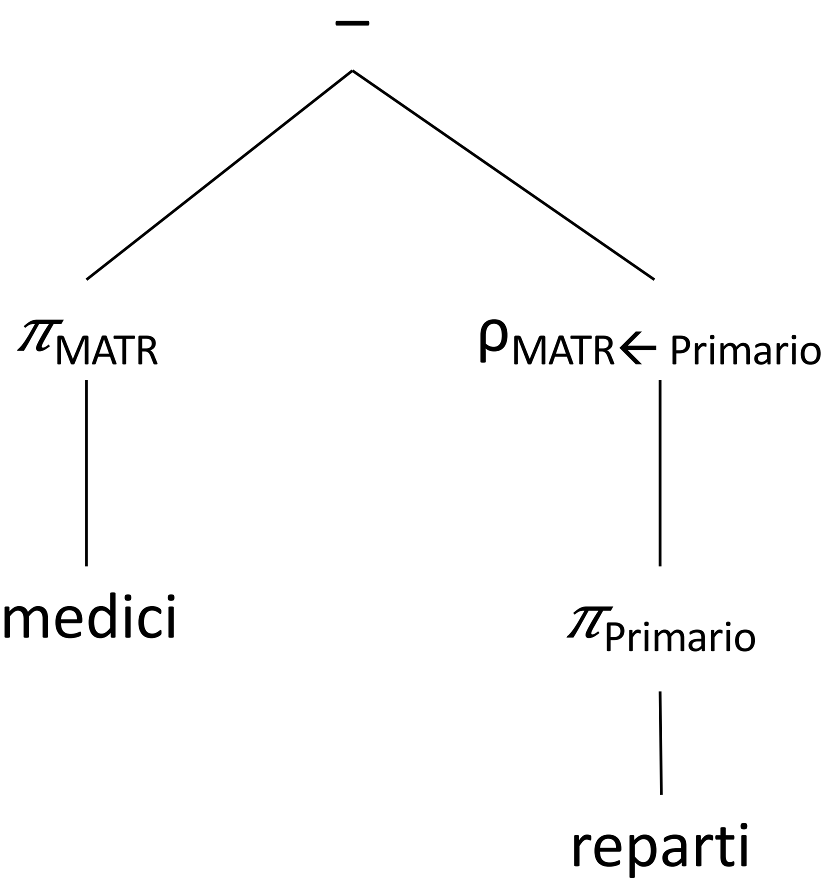
−

MATR
203
574
530
405

=

MATR
461
501

Rappresentazione ad albero



Esercizio

Elencare i codici dei pazienti che non sono mai stati ricoverati

Esercizio

Elencare i codici dei pazienti che non sono mai stati ricoverati

Suggerimento: potreste usare la differenza

Esercizio

Elencare i codici dei pazienti che non sono mai stati ricoverati

$$\pi_{COD}(pazienti) - \rho_{COD \leftarrow PAZ}(\pi_{PAZ}(ricoveri))$$

Esercizio

Elencare le città in cui risiedono sia medici che pazienti

Esercizio

Elencare le città in cui risiedono sia medici che pazienti

Suggerimento: potreste usare l'intersezione

Esercizio

Elencare le città in cui risiedono sia medici che pazienti

$$\pi_{Residenza}(medici) \cap \pi_{Residenza}(pazienti)$$

Ridenominazione dell'intero schema

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

$$\rho_{R'(B_i,B_j,\ldots,B_k) \leftarrow R(A_i,A_j,\ldots A_k)}(r)$$

è una relazione **virtuale**(*) con schema $R'(A_1,\ldots,B_i,\ldots,B_j,\ldots,B_k,\ldots,A_n)$

Esempio:

- $\rho_{UTENTI(CF,Provincia) \leftarrow PAZIENTI(COD,Residenza)}(pazienti)$
- Lo schema del risultato è *UTENTI(CF,Cognome,Nome,Provincia,AnnoNascita)*
- (gli attributi non ridenominati rimangono)

Prodotto cartesiano

Date due relazioni $r_1(A)$ e $r_2(B)$ 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

$\times$

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

$\times$

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»

- Cioè 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 e r_2 sono insiemi di tuple, quindi le tuple sono tutte distinte
- giustapponendo tuple distinte con altre tuple distinte ottengo una relazione di tuple distinte

Operatore di ϴ-join: $\bowtie_{\theta}$

- Serve a costruire informazioni estratte da più relazioni

- Mette 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: $\bowtie_{\Theta}$

Date:

- due relazioni $r_1(A)$ e $r_2(B)$ con $A \cap B = \emptyset$ (i due schemi non hanno attributi in comune)
- una condizione (predicato) Θ di join (tipicamente Θ è una formula proposizionale con **confronti tra attributi del tipo A_i φ B_j** *o* **confronti tra attributi e valori A_i φ costante** dove φ *= {<,>,=,≠,≤,≥}*)

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**

Esempio

Elencare le informazioni dei primari di ogni reparto

$$reparti \bowtie_{Primario=MATR} medici$$

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 afferisca 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 afferisca al reparto che dirige

$$\textit{reparti} \bowtie_{Primario=MATR \wedge COD=Reparto} \textit{medici}$$

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

$$\textit{reparti} \bowtie_{Primario=MATR \wedge COD \neq Reparto} \textit{medici}$$

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

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 intervallo molto ampio!

Proiezione e join

$$\textit{reparti} \bowtie_{Primario=MATR} \textit{medici}$$

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 nel risultato due volte l'informazione sulla matricola del medico, quindi spesso dopo un join si applica una proiezione per mantenere solo gli attributi di interesse.

Proiezione e join

Elenco solo codice e nome del reparto e cognome del primario

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

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

$$pazienti \bowtie_{Residenza=Residenza} medici$$

Problema: c'è un'ambiguità sull'attributo residenza!

Posso però ridenominare gli attributi omonimi, ad es.:

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

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

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

$$\bowtie_{ResidenzaP=ResidenzaM}$$

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

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

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*

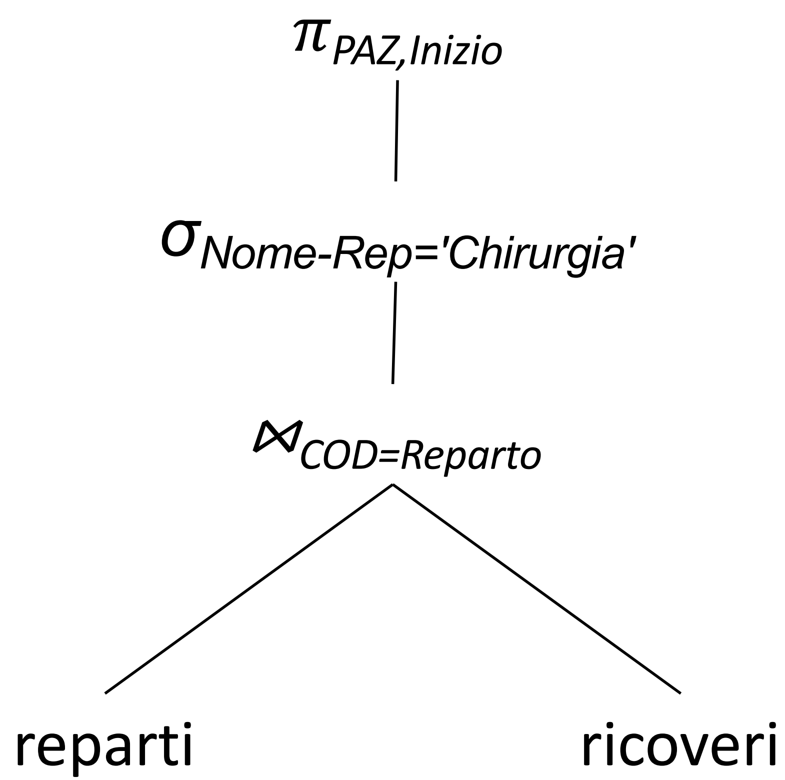
$$\sigma_{Nome\text{-}Rep='Chirurgia'}(reparti \ltimes_{COD=Reparto}\ ricoveri)$$

Posso limitare il numero di informazioni richieste

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

Join e selezione

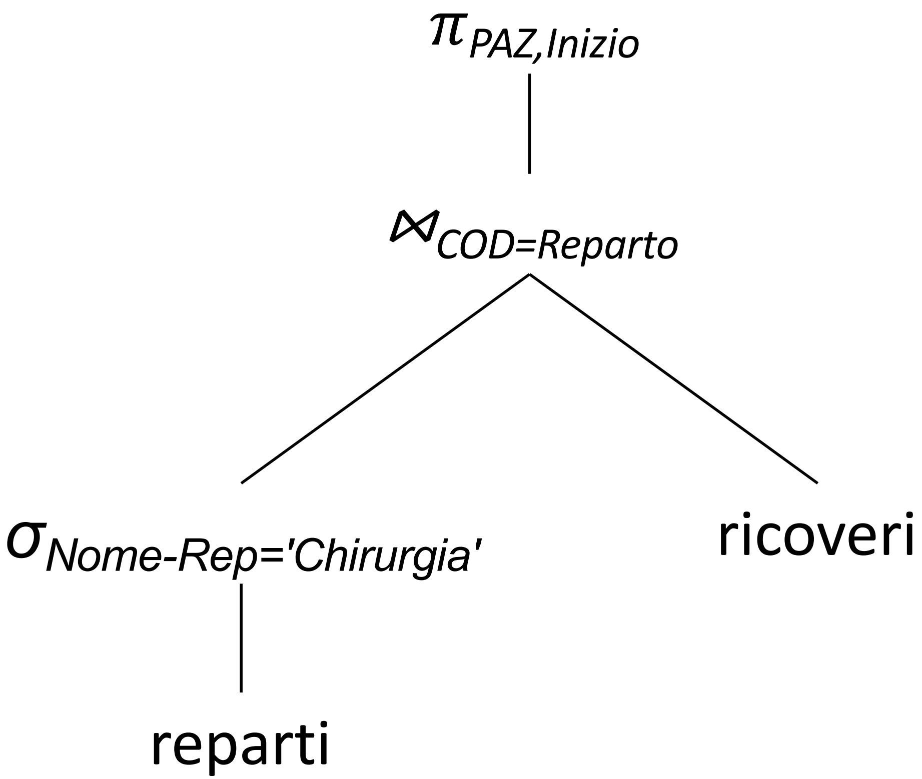
$$\pi_{PAZ,Inizio}(\sigma_{Nome\text{-}Rep='Chirurgia'}(reparti \bowtie_{COD=Reparto} ricoveri))$$



Join e selezione (alternativa)

Possiamo anticipare la selezione

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



Equivalenza delle espressioni

Le due espressioni

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

e

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

sono perfettamente **equivalenti**!

Esempio

Quali medici hanno avuto in cura il paziente Luigi Missoni?

- Per conoscere i dati del paziente ho bisogno della relazione pazienti…

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

Esempio

Quali medici hanno avuto in cura il paziente Luigi Missoni?

- … per sapere dove 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 Luigi Missoni?

- … per conoscere i medici afferenti al reparto ho bisogno della relazione *medici*…
- La relazione *medici* ha **molti attributi con lo stesso nome** di 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 Luigi Missoni?

- … ridenomino gli attributi di medici…

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

Esempio

Quali medici hanno avuto in cura il paziente Luigi Missoni?

- … ora posso mettere in join le due espressioni…

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

Esempio

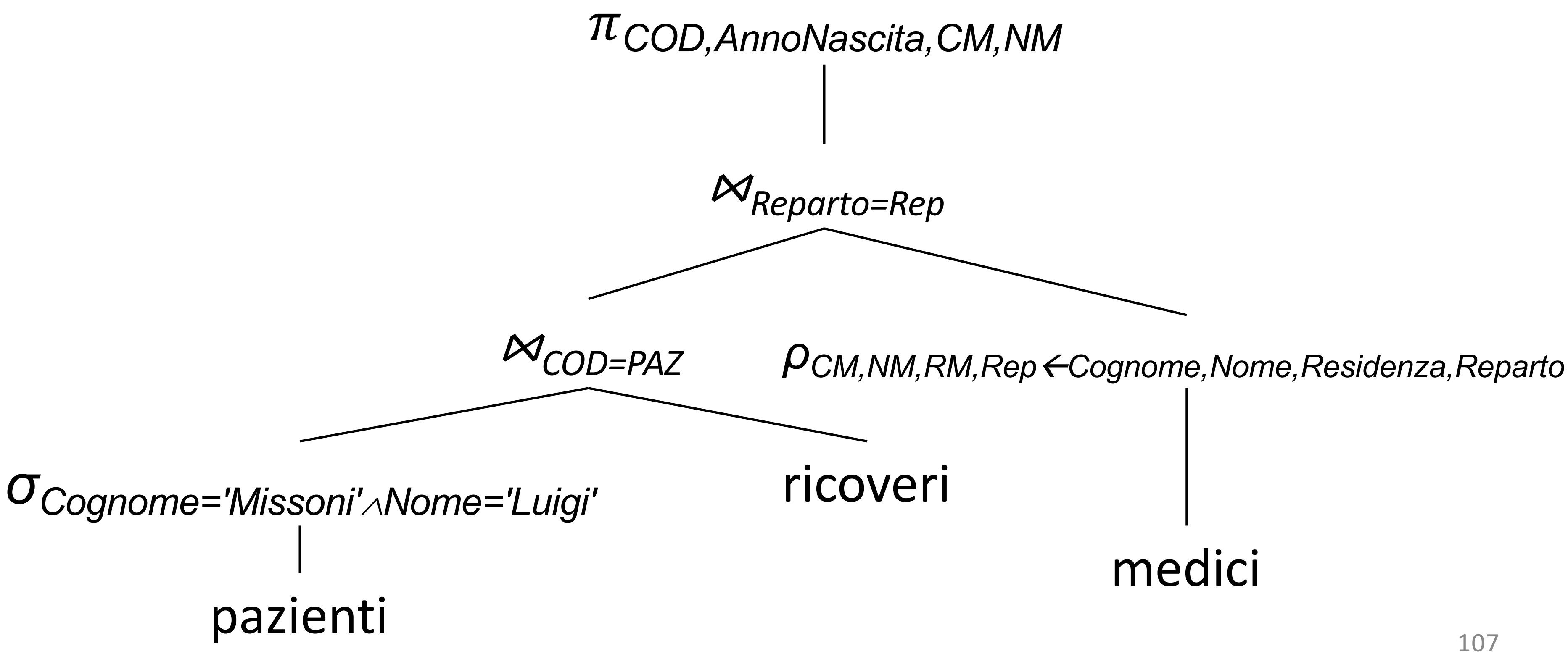
Quali medici hanno avuto in cura il paziente Luigi Missoni?

- … infine, proietto sui dati che mi interessano, ad esempio:

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

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 lo stesso nome in relazioni diverse, posso disambiguare il nome dell'attributo specificando la relazione a cui appartiene utilizzando la **dot-notation**:

$$\pi_{COD,AnnoNascita,\mathbf{MEDICI.Cognome},\mathbf{MEDICI.Nome}}$$
$$((\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti) \bowtie_{COD=PAZ} ricoveri)$$
$$\bowtie_{\mathbf{RICOVERI.Reparto}=\mathbf{MEDICI.Reparto}} (medici))$$

Ridenominazione e dot-notation

In pratica si sottintende un'operazione del tipo:

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

Esercizio

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

Esercizio

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

$\pi_{\textbf{MEDICI.Cognome,MEDICI.Nome}}$ $(((\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(pazienti) \bowtie_{PAZIENTI.COD=PAZ} ricoveri) \bowtie_{\textbf{Reparto=REPARTI.COD}} reparti) \bowtie_{Primario=MATR} medici)$

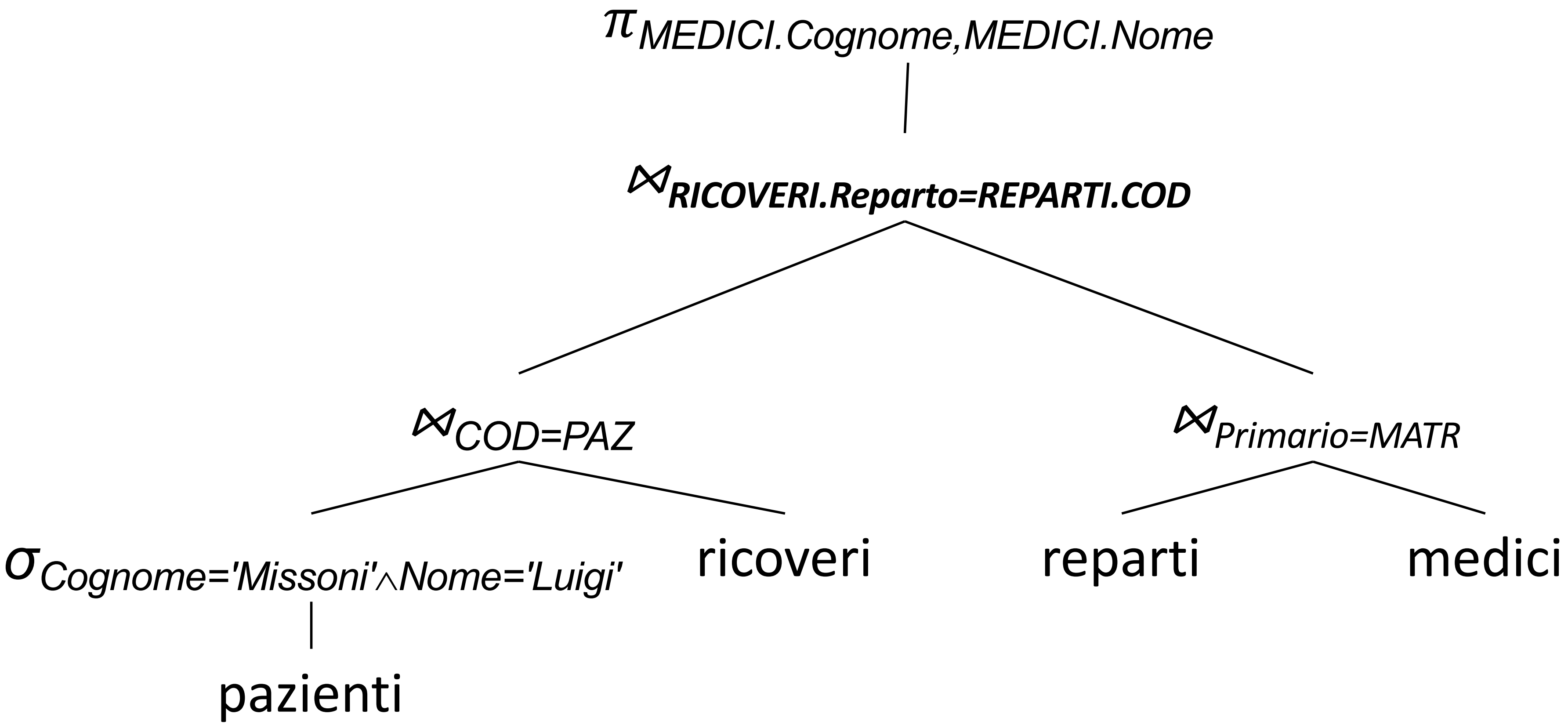
Esercizio

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

$$\pi_{MEDICI.Cognome, MEDICI.Nome}\left(\left(\left(\sigma_{Cognome='Missoni' \wedge Nome='Luigi'}(\text{pazienti}) \bowtie_{COD=PAZ} \text{ricoveri}\right) \bowtie_{Reparto=REPARTI.COD} \text{reparti}\right) \bowtie_{Primario=MATR} \text{medici}\right)$$

Esercizio (soluzione alternativa)

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



Interrogazioni con aggregazioni

- Non vedremo operatori in algebra relazionale che effettuano aggregazioni (ad es. conteggi: quante volte sono stati ricoverati i pazienti?)
- In SQL sono possibili attraverso operatori specifici

- Vediamo che si può comunque rispondere in algebra relazionale a domande del tipo:
  - Elencare i pazienti che hanno avuto due o più ricoveri

Esempio

Elencare i pazienti che hanno avuto due o più ricoveri

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

Come confrontiamo tuple della stessa relazione?

Esempio

Elencare i pazienti che hanno avuto due o più ricoveri

Posso mettere la relazione ricoveri in join con se stessa

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

Ma devo prima ridenominare tutti gli attributi!

$$\rho_{PAZ1,Inizio1,Fine1,Reparto1 \leftarrow PAZ,Inizio,Fine,Reparto}(ricoveri) \bowtie_{\theta}$$
$$\rho_{PAZ2,Inizio2,Fine2,Reparto2 \leftarrow PAZ,Inizio,Fine,Reparto}(ricoveri)$$

Esempio

Elencare i pazienti che hanno avuto due o più ricoveri

Posso ora esplicitare il Θ

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

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

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

Esempio

Elencare i pazienti che hanno avuto due o più ricoveri

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

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

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

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 avuto due o più ricoveri

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

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

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

PAZ1	Inizio1	Fine1	Reparto1	PAZ2	Inizio2	Fine2	Reparto2
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

Se vogliamo ricavare solo i codici dei pazienti senza duplicati, possiamo fare inoltre una proiezione su PAZ1 (o PAZ2)

Esempio (ridenominazione dello schema)

Elencare i pazienti che hanno avuto due o più ricoveri

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

$$\bowtie_{\text{RICOVERI1.PAZ=RICOVERI2.PAZ} \wedge \text{RICOVERI1.Inizio<RICOVERI2.Inizio}}$$

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

Esercizio

Elencare cognome e nome dei pazienti ricoverati più di una volta

Esercizio

Elencare cognome e nome dei pazienti ricoverati più di una volta

…

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

$$\bowtie_{\text{RICOVERI1.PAZ=RICOVERI2.PAZ} \wedge \text{RICOVERI1.Inizio<RICOVERI2.Inizio}}$$

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

…

Esercizio

Elencare cognome e nome dei pazienti ricoverati più di una volta

$$\pi_{\mathbf{Cognome,Nome}}$$

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

$$\bowtie_{\mathrm{RICOVERI1.PAZ=RICOVERI2.PAZ} \wedge \mathrm{RICOVERI1.Inizio<RICOVERI2.Inizio}}$$

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

$$\bowtie_{\mathbf{RICOVERI1.PAZ=PAZIENTI.COD}} \mathbf{pazienti}))$$

Esercizio

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

Esercizio

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

- Suggerimento: posso sfruttare un join tra pazienti e medici

Esercizio

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

$$\pi_{\text{PAZIENTI.Nome,PAZIENTI.Cognome}}$$

$$(\text{pazienti} \bowtie_{\text{PAZIENTI.Residenza=MEDICI.Residenza}} \text{medici})$$

Base di Dati "Impiegati"

impiegati

MATR	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 loro subalterni…

Esempio

Elencare i capi che guadagnano meno di almeno uno dei loro subalterni…

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

Esempio

Elencare i capi che guadagnano meno di almeno uno dei loro subalterni…

1) Dovendo usare due volte la relazione impiegati, devo ridenominarla:

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

Esempio

Elencare i capi che guadagnano meno di almeno uno dei loro subalterni…

2) occorre mettere in join la relazione *impiegati* con *organigramma* per conoscere lo stipendio dei capi

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

Esempio

Elencare i capi che guadagnano meno di almeno uno dei loro subalterni…

3) per conoscere lo stipendio dei subalterni devo mettere in join un’altra volta la relazione *impiegati*

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

$$\bowtie_{\text{IMPIEGATO = IMPIEGATI.MATR}} impiegati$$

Esempio

Elencare i capi che guadagnano meno di almeno uno dei loro subalterni…

4) Imponiamo ora che lo stipendio del capo sia inferiore dello stipendio del subalterno

$\sigma_{CAPI.Stipendio<IMPIEGATI.Stipendio}((\rho_{CAPI \leftarrow IMPIEGATI}(impiegati)$

$\bowtie_{\text{CAPI.MATR = Capo}}$ *organigramma)*

$\bowtie_{\text{Impiegato = IMPIEGATI.MATR}}$ *impiegati)*

Esempio

Elencare i capi che guadagnano meno di almeno uno dei loro subalterni…

5) proiettiamo solo le informazioni necessarie

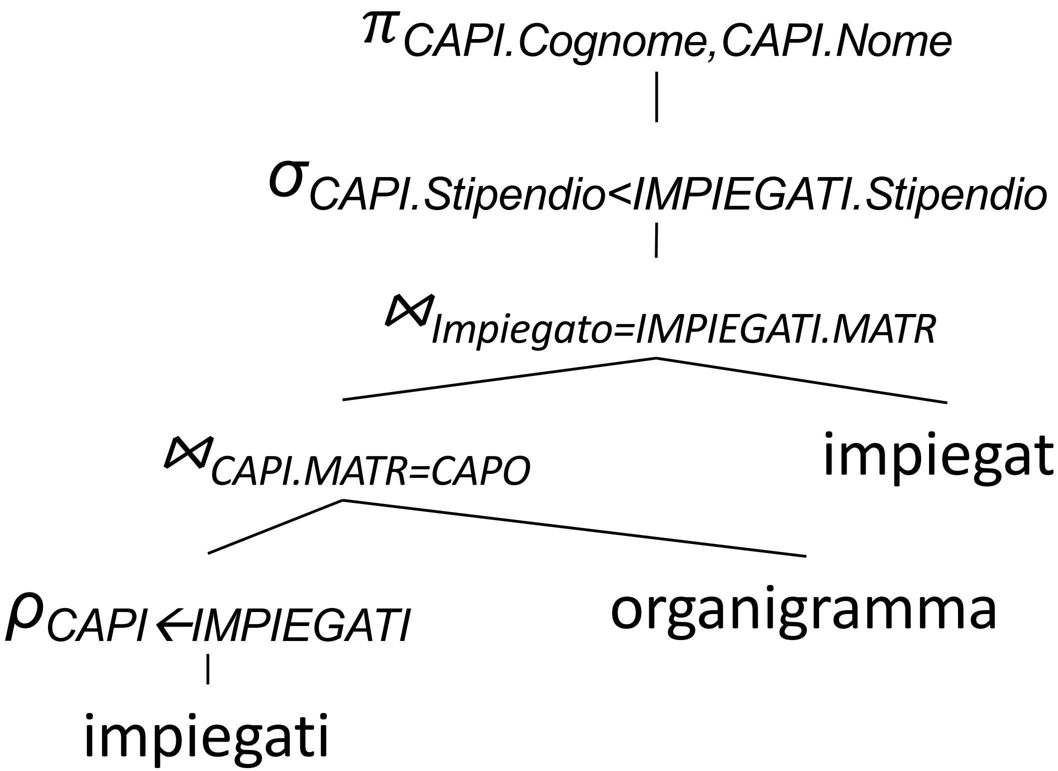
$\pi_{CAPI.Cognome,CAPI.Nome}(\sigma_{CAPI.Stipendio<IMPIEGATI.Stipendio}$

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

$\bowtie_{\text{Impiegato = IMPIEGATI.MATR}} impiegati)$

Esempio (albero sintattico)

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



Casi particolari di join

Ricordiamo la **cardinalità** del Ө-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 essere **più precisi** sull'intervallo della cardinalità **restringendo il predicato Ө a casi particolari**

Equi-join: Θ_e-join

L'Equi-join è un caso particolare del Θ-join in cui i confronti sono solo uguaglianze.

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

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

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

$$\Theta_e := A_{i1}=B_{j1} \wedge A_{i2}=B_{j2} \dots \wedge A_{in}=B_{jn}$$

Equi-join: Θ_e-join

Per semplicità consideriamo un solo predicato atomico

$$r_1(A) \ \bowtie_{Ai=Bj} \ r_2(B)$$

A seconda del numero di tuple in r_2 che troviamo per ogni tupla di r_1, possiamo individuare due casi particolari:

Caso 1: per ogni tupla di r_1 abbiamo ***al più una*** tupla in r_2

Caso 2: per ogni tupla di r_1 abbiamo ***esattamente una*** tupla in r_2

Casi particolari

$$r_1(A) \ltimes_{Ai=Bj} r_2(B)$$

Caso 1 («al più una»)

- Assumiamo che B_j sia la chiave primaria di $r_2(B)$
- Non facciamo nessuna assunzione su A (per es. potrebbero non esserci vincoli di integrità referenziale)

Per ogni tupla di r_1, si troverà **al più una corrispondenza con r_2**: se ci fossero più corrispondenze sarebbe violato il **vincolo di chiave primaria**!

Quindi:

$$0 \leq |r_1(A) \ltimes_{\Theta_e} r_2(B)| \leq \boldsymbol{|r_1(A)|}$$

Casi particolari

$$r_1(A) \bowtie_{Ai=Bj} r_2(B)$$

Caso 2 («esattamente una»)

- Assumiamo che B_j sia la chiave primaria di $r_2(B)$
- Inoltre assumiamo che esista un vincolo di integrità referenziale tra A e B («$r(A_i)$ referenzia $r(B_j)$»). Formalmente $\forall t_i \in r_1 \; (\exists t_j \in r_2 \; (t_i[A_i]=t_j[B_j]))$

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

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

Esercizio

Quante tuple genera l'equi-join
ricoveri $\bowtie_{PAZ=COD}$ *pazienti*?

Esercizio

Quante tuple genera l'equi-join
ricoveri $\bowtie_{PAZ=COD}$ *pazienti*?

In *ricoveri* $\bowtie_{PAZ=COD}$ *pazienti* la condizione di join PAZ=COD coinvolge la chiave primaria di pazienti COD ed esiste un vincolo di integrità referenziale tra ricoveri(PAZ) e pazienti(COD).

Quindi per ogni tupla di ricoveri esiste una e una sola tupla di pazienti che rispetta la condizione.

Quindi

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

Natural join

Cambiamo leggermente il database RICOVERI dando nomi uguali agli attributi che rappresentano lo stesso concetto

pazienti

COD	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

Rep	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

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

Natural join

Date:

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

il **natural join** è definito come

$$r_1(A) \bowtie r_2(B) := \pi_{R1.X,Y,Z}(\rho_{R1.X \leftarrow X}(r_1) \bowtie_{\theta_e} \rho_{R2.X \leftarrow X}(r_2))$$

dove

$$\theta_e := (R1.X_1 = R2.X_1) \wedge \ldots \wedge (R1.X_k = R2.X_k)$$

Esempio

ricoveri $\bowtie$ *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 $\bowtie$ *pazienti* $\bowtie$ *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*

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

Vengono messi in uguaglianza **tutti** gli attributi omonimi.
Quindi a quale interrogazione sto rispondendo?

Uso corretto del natural join

$$reparti \bowtie medici$$

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

Vengono messi in uguaglianza **tutti** gli attributi omonimi. Quindi a quale interrogazione sto rispondendo?

$$\pi_{Rep,Nome\text{-}Rep,MATR,Cognome,Nome,Residenza}$$
$$(reparti \bowtie_{REPARTI.Rep=MEDICI.Rep \wedge REPARTI.MATR=MEDICI.MATR} medici)$$

Ricavo i medici che sono primari del reparto cui afferiscono

Uso corretto del natural join

Usare con cautela!

Casi limite del natural join

Se A e B sono due schemi 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)

Casi limite del natural join

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

Casi limite del natural join

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

Possiamo immaginarlo come un equi-join su tutti gli attributi, quindi prendiamo le tuple uguali su tutti gli attributi in r_1 e r2, quindi $r_1(A) \bowtie r_2(A) = r_1(A) \cap r_2(A)$

Semi-join: $\ltimes_{\theta}$

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

$$r_1(A) \ltimes_{\theta} r_2(B) := \pi_A(r_1(A) \bowtie_{\theta} r_2(B))$$

L'operatore di semi-join funziona come un filtro sulla prima relazione sfruttando la seconda relazione

Esempio

Elencare tutti i dati dei primari

$$\textit{medici} \ltimes_{MATR=Primario} \textit{reparti}$$

equivalente a:

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

seleziono solo i medici che sono anche primari grazie al join

Esempio

Elencare tutti i dati dei pazienti ricoverati in chirurgia

$$pazienti \ltimes_{COD=PAZ} (ricoveri \ltimes_{Reparto=REPARTI.COD} (\sigma_{Nome\text{-}Rep='Chirurgia'}(reparti)))$$

Cardinalità del semi-join

$$|r_1(A) \ltimes_\theta r_2(B)| = |\ \pi_A(\ r_1(A) \bowtie_\theta r_2(B)\ )|$$

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

$$0 \leq |r_1(A) \ltimes_\theta r_2(B)| \leq |r_1(A)|$$