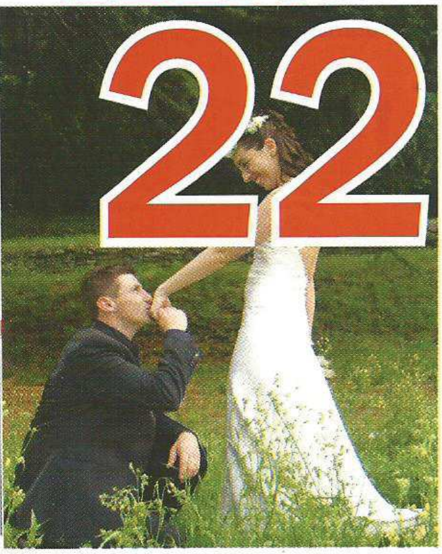
LEZIONE 22



Le operazioni relazionali: unione

In questa lezione focalizziamo l'attenzione sulle operazioni che consentono di interrogare una base di dati relazionale. Interrogare una base di dati significa ottenere le informazioni desiderate estraendo da una tabella una sottotabella, oppure combinando tra loro due o più tabelle per generare nuove relazioni. I linguaggi di programmazione utilizzati per l'interrogazione sono di tipo **non procedurale** e si basano sull'algebra relazionale e sul calcolo relazionale.

Secondo l'approccio basato sull'**algebra relazionale**, il risultato di un'interrogazione (o **query**) è una relazione. Per ottenere tale relazione si formula un'interrogazione utilizzando alcuni operatori di algebra relazionale, come ad esempio quelli di *unione*, *intersezione* e *differenza* fra relazioni viste come insiemi.

Anche secondo l'approccio basato sul **calcolo relazionale** il risultato di un'interrogazione è una relazione. Per ottenere tale relazione si formula un'interrogazione utilizzando il calcolo dei predicati del primo ordine sulle relazioni della base di dati.

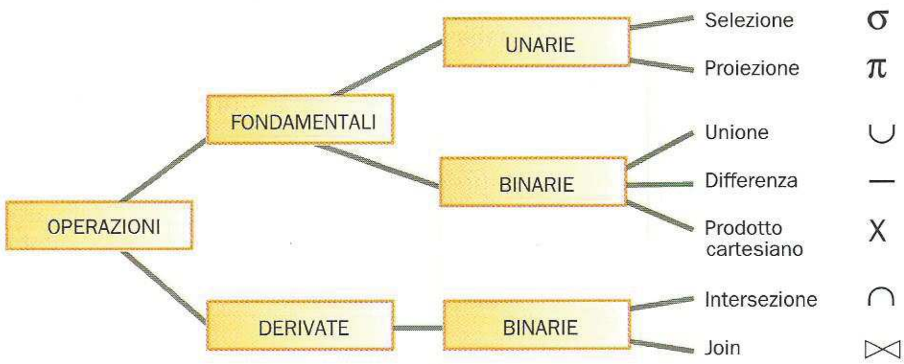
I due approcci possono considerarsi equivalenti, nel senso che la relazione risultato che deriva dall'espressione algebrica può essere espressa con un equivalente predicato del primo ordine.
Scegliamo di utilizzare l'approccio dell'algebra relazionale, data la maggiore familiarità con gli operatori algebrici rispetto al calcolo dei predicati.
Occorre definire l'insieme degli operatori che vogliamo utilizzare nelle interrogazioni. Tale insieme deve essere funzionalmente completo, cioè deve consentire di ottenere gli stessi risultati ottenibili con altri linguaggi relazionali.
Un insieme funzionalmente completo è quello formato dai seguenti cinque operatori relazionali:

1. **unione** di relazioni;
2. **differenza** di relazioni;
3. **proiezione** di relazioni;
4. **restrizione** di relazioni; (O SELEZIONE)
5. **prodotto** di relazioni.

Componendo opportunamente tali operatori, è possibile formulare qualsiasi interrogazione sulla base di dati.
Oltre a questi cinque operatori di base, comunque è opportuno introdurne altri due:

6. **intersezione** di due relazioni;
7. **giunzione naturale** di due relazioni. (JOIN

Questi ultimi operatori equivalgono a una composizione di quelli base e sono introdotti perché permettono di scrivere espressioni più semplici e sintetiche.



Unione di relazioni

Due relazioni R e S si dicono **compatibili** se:

- hanno lo stesso numero di attributi;
- ogni attributo nella stessa posizione all'interno delle due relazioni è dello stesso tipo.

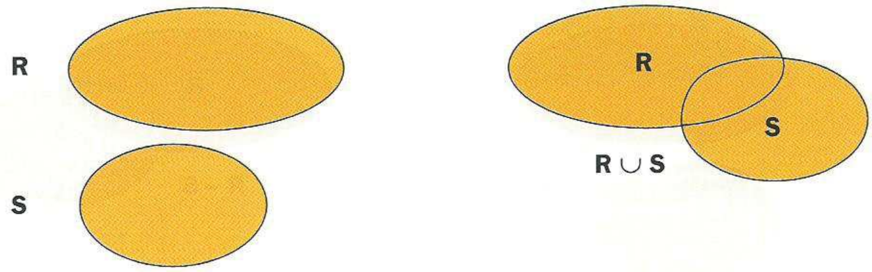
Ad esempio, le seguenti relazioni *Persona* e *Dipendente* sono compatibili:

PERSONA(Nome: STRINGA, Stipendio: INTERO, DataNascita: DATA)
DIPENDENTE(Nominativo: STRINGA, Stip: INTERO, DataNas: DATA)

Date due relazioni compatibili R e S, l'**unione** di R con S è la relazione ottenuta dall'unione insiemistica delle due relazioni:

$$R \cup S = \{ t \mid t \in R \textbf{ OR } t \in S \}$$

Utilizzando la rappresentazione insiemistica abbiamo:



Ad esempio, se R rappresenta i clienti del primo semestre di attività di un'azienda e S i clienti del secondo semestre, R ∪ S rappresenta i clienti dell'anno.

R = Clienti1Semestre

R	Nome			
	Rossi			
	Bianchi			
	Verdi			

S = Clienti2Semestre

S	Nome			
	Gialli			
	Bianchi			
	Neri			

R ∪ S = Clienti = Clienti1Semestre ∪ Clienti2Semestre

R ∪ S	Nome			
	Rossi			
	Bianchi			
	Neri			
	Verdi			
	Gialli			

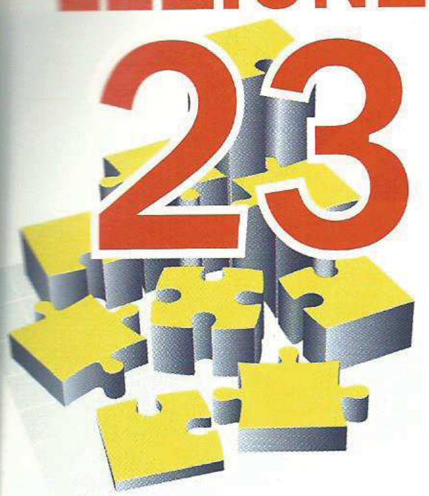
STOP

Per come è stata definita l'operazione di unione abbiamo che:

Grado(R –S) = Grado(R) = Grado(S)
Card(RS) = Card(R) + Card(S) – numero di t-uple ripetute

Ovviamente abbiamo supposto che le t-uple con nome "Bianchi" presenti in entrambe le relazioni fossero uguali (per semplicità abbiamo considerato solo *Nome* come attributo chiave primaria).

LEZIONE 23



Le operazioni relazionali: differenza, proiezione e restrizione

Differenza di relazioni

Date due relazioni compatibili R e S, la **differenza** di R con S è la relazione data dalla differenza insiemistica delle due relazioni:

$R - S = \{ t \mid t \in R \ \textbf{AND} \ t \notin S \}$

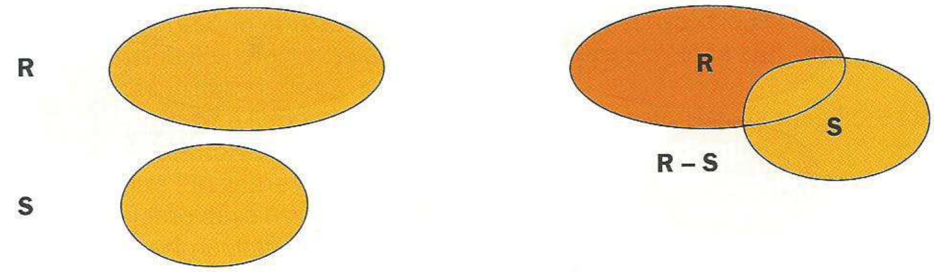
La differenza insiemistica non gode della proprietà commutativa, quindi:

$R - S \neq S - R$

Infatti:

$S - R = \{ t \mid t \in S \ \textbf{AND} \ t \notin R \}$

Utilizzando la rappresentazione insiemistica abbiamo:



Ad esempio, se R rappresenta tutti i clienti di un'azienda e S i clienti dell'anno 2009, R – S rappresenta tutti i clienti esclusi quelli relativi al 2009.

R = Clienti

R	Nome			
	Rossi			
	Bianchi			
	Verdi			

S = Clienti09

S	Nome			
	Gialli			
	Bianchi			
	Neri			

R – S = Clienti – Clienti09

R – S	Nome			
	Rossi			
	Bianchi			

Le t-uple presenti in S non vengono inserite.

STOP

Per come è stata definita l'operazione di differenza abbiamo che:

Grado(R – S) = Grado(R) – Grado(S)
Card(R – S) = Card(R) – numero di t-uple presenti anche in S

Proiezione di una relazione

Data una relazione R e un sottoinsieme $A = \{A_1, A_2, ..., A_k\}$ dei suoi attributi, si definisce **proiezione** di R su A la relazione di grado K che si ottiene considerando solo le colonne di R relative agli attributi contenuti in A ed eliminando le eventuali t-uple duplicate.

Abbiamo una relazione R, che rappresenta gli alunni di una classe terza, e una relazione S che rappresenta i testi adottati per quella stessa classe (gli attributi sono solo a titolo di esempio). Vogliamo costruire una tabella con l'elenco di tutti i testi per ogni alunno.
Scriveremo:

$\pi_{A1, A2, ..., Ak}(R) = \{ t[A_1, A_2, ..., A_k] \; t \in R \}$

L'effetto di una proiezione è quello di selezionare un certo numero di colonne dalla tabella della relazione.

PROIEZIONE

Consideriamo la seguente relazione CLIENTI relativa ai clienti di un'azienda e supponiamo di voler estrapolare solo il nome degli agenti e l'indirizzo dei clienti. $Card(S) \leq Card(R)$, infatti le t-uple presenti nella proiezione possono anche essere di numero inferiore a quelle di R, poiché le t-uple duplicate vengono scartate, come nell'esempio.

R = Clienti

R	CodCli	Nome	Agente	IndirizzoCli
	C001	Bianchi	Polis	Via Po, 23
	C002	Neri	Conte	Via Roma, 12
	C003	Verdi	Polis	Via Po, 23

S = $\pi_{Agente, IndirizzoCli}$ (R)

S	Agente	IndirizzoCli
	Polis	Via Po, 23
	Conte	Via Roma, 12

Restrizione di una relazione

Data una relazione R e un predicato P (semplice o composto) sui suoi attributi, la **restrizione** (o **selezione**) di R a P è la relazione costituita dalle t-uple di R che soddisfano P.

Scriveremo:

$\delta_P(R) = \{ t \mid t \in R \textbf{ AND } P(t) \}$

L'effetto di una restrizione è, pertanto, quello di selezionare un certo numero di righe dalla tabella della relazione. Supponiamo di avere la seguente tabella dei clienti di un'azienda e di voler selezionare le informazioni relative ai clienti della provincia di Milano. Il nostro predicato sarà: $P(t) = \{Provincia = \text{"MI"}\}$.

SELEZIONE

R = Clienti

R	CodCli	Nome	Provincia	IndirizzoCli
	C002	Neri	LE	Via Roma, 12
	C006	Bianchi	MI	Via Po, 23
	C005	Rossi	MI	Via Moro, 2

S = $\sigma_{Provincia = \text{"MI"}}$ (R)

S	CodCli	Nome	Provincia	IndirizzoCli
	C006	Bianchi	MI	Via Po, 23
	C005	Rossi	MI	Via Moro, 2

$Grado(S) = Grado(R)$

$Card(S) \leq Card(R)$, è uguale solo quando tutte le t-uple soddisfano P.

LEZIONE 24

Le operazioni relazionali: prodotto cartesiano e intersezione

Prodotto cartesiano di due relazioni

Il prodotto cartesiano tra due insiemi A e B (indicato con A x B) si definisce come l'insieme di tutte le possibili coppie che hanno come primo elemento un elemento di A e come secondo elemento un elemento di B.
Il prodotto cartesiano di due relazioni coincide con il prodotto cartesiano tra due insiemi, ed è dunque l'insieme di tutte le possibili t-uple ottenute concatenando ogni t-upla della prima relazione con ogni t-upla della seconda.
Più formalmente:

Date due relazioni R e S, rispettivamente di grado g1 e g2 e cardinalità c1 e c2, il **prodotto** di R e S è la relazione di grado g1 + g2 e cardinalità c1 x c2 le cui t-uple si ottengono concatenando ogni t-upla di R con ogni t-upla di S.

A = {1, 3, 5, 7}

B = {2, 4}

A x B = {(1, 2), (3, 2), (5, 2), (7, 2), (1, 4), (3, 4), (5, 4), (7, 4)}

A	B
1	b
2	a

C	D
3	d
4	g

A	B	C	D
1	b	3	d
2	a	4	g
2	a	3	d
2	a	4	g

Siano $r = (a_1, a_2, ..., a_{n1})$ e $s = (b_1, b_2, ..., b_{n2})$ due t-uple; la concatenazione di r e s è data da:

$$r \textbf{ conc } s = (a_1, a_2, ..., a_{n1}, b_1, b_2, ..., b_{n2})$$

Il prodotto di R per S è quindi definito come:

$$\mathbf{R \times S} = \{ t \mid t = r \textbf{ conc } s, r \in R, s \in S \}$$

Per evitare ogni ambiguità nei nomi degli attributi di *R* x *S*, occorre che i nomi degli attributi di R e S siano diversi tra loro. Se così non fosse, si può pensare di ricorrere a una preventiva operazione di ridenominazione degli attributi.
La chiave primaria della relazione prodotto è data dall'unione della chiavi primarie delle due relazioni di partenza.

Abbiamo una relazione R, che rappresenta gli alunni di una classe terza, e una relazione S che rappresenta i testi adottati per quella stessa classe (gli attributi sono solo a titolo di esempio). Vogliamo costruire una tabella con l'elenco di tutti i testi per ogni alunno.

R = Alunni

R	Matricola	Nominativo	Data di nascita
	12346	Bianchi Stefania	6/4/1988
	12347	De Pascalis Alberto	12/9/1988
	12348	Manca Roberta	10/5/1988

S = Testi

S	CodTesto	Titolo	Materia
	T1	L'italiano oggi	Italiano
	T2	Conoscere la matematica	Matematica
	T3	Informatica e Azienda	Informatica

R x S = Alunni x Testi

R x S	Matricola	Nominativo	Data di nascita	CodTesto	Titolo	Materia
	12346	Bianchi Stefania	6/4/1988	T1	L'italiano oggi	Italiano
	12346	Bianchi Stefania	6/4/1988	T2	Conoscere la matematica	Matematica
	12346	Bianchi Stefania	6/4/1988	T3	Informatica e Azienda	Informatica
	12347	De Pascalis Alberto	12/9/1988	T1	L'italiano oggi	Italiano
	12347	De Pascalis Alberto	12/9/1988	T2	Conoscere la matematica	Matematica
	12347	De Pascalis Alberto	12/9/1988	T3	Informatica e Azienda	Informatica
	12348	Manca Roberta	10/5/1988	T1	L'italiano oggi	Italiano
	12348	Manca Roberta	10/5/1988	T2	Conoscere la matematica	Matematica
	12348	Manca Roberta	10/5/1988	T3	Informatica e Azienda	Informatica

Grado(R x S) = g1 + g2 Card(R x S) = c1 x c2

Intersezione di due relazioni

Date due relazioni compatibili R e S, l'**intersezione** di R e S restituisce la relazione composta da tutte le t-uple presenti sia in R sia in S.

Scriveremo:
$R \cap S = \{ t \mid t \in R \textbf{ AND } t \in S \}$

Supponiamo di avere le informazioni relative ai clienti del 2008 e a quelli del 2009 della nostra azienda. Vogliamo ottenere una tabella con le persone che sono state nostre clienti sia nel 2008 sia nel 2009.

R = Clienti08

R	CodCli	Nome	Provincia	IndirizzoCli
	C006	Bianchi	MI	Via Po, 23
	C002	Neri	LE	Via Roma, 12
	C005	Rossi	MI	Via Moro, 2

S = Clienti09

S	CodCli	Nome	Provincia	IndirizzoCli
	C016	Verdi	CO	Via Moro, 13
	C002	Neri	LE	Via Roma, 12
	C005	Rossi	MI	Via Moro, 2

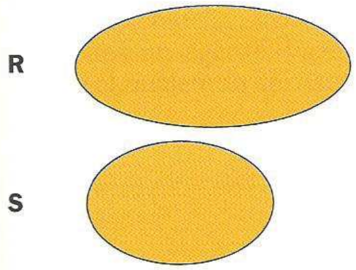
R∩S

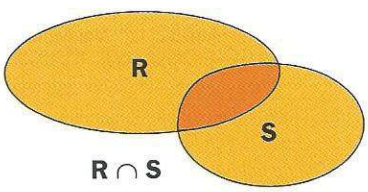
R∩S	CodCli	Nome	Provincia	IndirizzoCli
	C002	Neri	LE	Via Roma, 12
	C005	Rossi	MI	Via Moro, 2

Grado(R ∩ S) = Grado(R) = Grado(S)

Card(R ∩ S) ≤ alla cardinalità minore tra Card(R) e Card(S). In particolare si ha: Card $(R \cap S)$ = Card(R) se $R \subseteq S$
Card $(R \cap S)$ = Card(S) se $S \subseteq R$

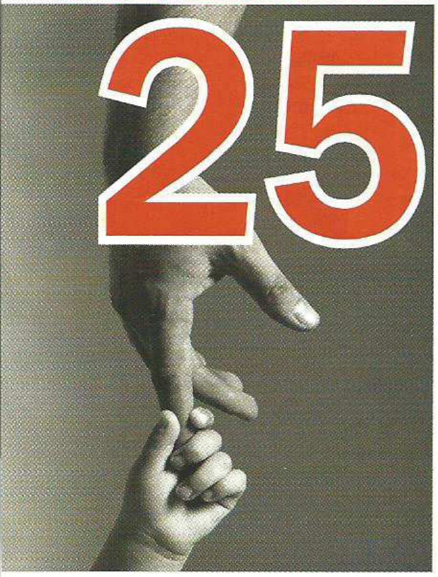
Utilizzando la rappresentazione insiemistica abbiamo:





LEZIONE

25



Le operazioni relazionali: giunzione naturale

Ancora più spesso della necessità di fare il prodotto di due relazioni, si presenta la necessità di *unirle (join)*, accoppiando solo le t-uple tra le quali esiste qualche tipo di corrispondenza. Il tipo di corrispondenza più semplice è la giunzione naturale (o join naturale) di due relazioni R e S in cui vengono accoppiate solo le t-uple di R e di S che concordano su ogni attributo comune negli schemi di R e di S. Più precisamente:

Date due relazioni R e S rispettivamente di grado g1 e g2, l'operazione di **giunzione naturale (join)** di R e S su un attributo A di R e un attributo B di S, aventi lo stesso tipo, restituisce una relazione di grado (g1 + g2 − 1) le cui t-uple si ottengono con il seguente procedimento:

1. si effettua il prodotto cartesiano di R e S;
2. sulla relazione così ottenuta si effettua una restrizione sulle t-uple aventi gli attributi A e B dello stesso valore;
3. la relazione così ottenuta ha le colonne A e B uguali, per cui si elimina una di tali colonne.

Indicheremo l'operazione con:

$$\underset{A\,=\,B}{R \bowtie S}$$

Lo scopo della giunzione naturale è quello di combinare due relazioni aventi uno o più attributi in comune, generando una nuova relazione che contiene:

- le colonne della prima e della seconda, senza duplicazioni;
- le righe della prima concatenate a quelle della seconda, secondo i valori uguali dell'attributo comune.

Il join naturale determina una relazione tramite il seguente procedimento:

1. viene effettuato un prodotto cartesiano fra le due relazioni;
2. sulla relazione così creata viene effettuata un'operazione di selezione delle t-uple in cui risulta vera la condizione posta dal predicato, e cioè che gli attributi sottoposti all'operatore di confronto siano uguali; se questa condizione non si verifica per nessun attributo, il risultato sarà una tabella vuota.
3. si usa ridenominare gli attributi che hanno lo stesso nome, in modo che non vi siano ambiguità.

Supponiamo di avere due tabelle: una relativa ai clienti della nostra azienda e l'altra relativa agli agenti. Vogliamo ottenere in un'unica tabella, per ogni cliente, anche le informazioni degli agenti da cui esso viene servito.

R = Cliente

R	CodCli	NomeCli	CodAg	IndirizzoCli
	C006	Bianchi	A0052	Via Po, 23
	C002	Neri	A0016	Via Roma, 12
	C005	Rossi	A0052	Via Moro, 2

S = Agente

S	CodAg	NomeAg	TelAgente
	A0016	Polis	346/5647523
	A0052	Rinaldi	322/7665541

Vediamo passo per passo il procedimento appena descritto nella definizione.

1. Effettuiamo il prodotto cartesiano di R e S:

► R x S

CodCli	NomeCli	CodAg	IndirizzoCli	CodAg	NomeAg	Cliente	TelAgente
C006	Bianchi	A0052	Via Po, 23	A0016	Polis	Neri	397/5647523
C006	Bianchi	A0052	Via Po, 23	A0052	Rinaldi	Bianchi	322/7665541
C002	Neri	A0016	Via Roma, 12	A0016	Polis	Neri	397/5647523
C002	Neri	A0016	Via Roma, 12	A0052	Rinaldi	Bianchi	322/7665541
C005	Rossi	A0052	Via Moro, 2	A0016	Polis	Neri	397/5647523
C005	Rossi	A0052	Via Moro, 2	A0052	Rinaldi	Bianchi	322/7665541

2. Procediamo con la restrizione del prodotto cartesiano di R e S:

► δ_{CodAg} = CodAg (R x S)

CodCli	NomeCli	CodAg	IndirizzoCli	CodAg	NomeAg	TelAgente
C006	Bianchi	A0052	Via Po, 23	A0052	Rinaldi	322/7665541
C002	Neri	A0016	Via Roma, 12	A0016	Polis	397/5647523
C005	Rossi	A0052	Via Moro, 2	A0052	Polis	322/7665541

3. Eliminando una delle colonne *CodAg* otteniamo la relazione finale della congiunzione (join) tra R e S:

► $R \underset{A=B}{\bowtie} S$

CodCli	NomeCli	IndirizzoCli	CodAg	NomeAg	TelAgente
C006	Bianchi	Via Po, 23	A0052	Rinaldi	322/7665541
C002	Neri	Via Roma, 12	A0016	Polis	397/5647523
C005	Rossi	Via Moro, 2	A0052	Rinaldi	397/5647523

Grado($R \underset{A=B}{\bowtie} S$) = (g1 + g2 − 1) poiché l'attributo comune compare una volta sola.

Card($R \underset{A=B}{\bowtie} S$) non è prevedibile a priori in quanto dipende da quante righe nelle due tabelle hanno valori uguali per gli attributi su cui si effettua il join.

Il join si applica in modo indipendente dal contenuto delle due tabelle: nel nostro esempio, la tabella *Cliente* avrebbe potuto avere un cliente al quale non corrispondesse nessun agente; viceversa la tabella *Agente* avrebbe potuto contenere un agente al quale non corrispondesse alcun cliente.