

Linguaggi Relazionali

- **Algebra Relazionale**: insieme di operatori su relazioni che danno come risultato relazioni; si definiscono
 - **operatori primitivi** (ridenominazione, proiezione, unione e differenza, restrizione, prodotto)
 - **operatori derivati** (giunzioni, divisione, ...)
 - **altri operatori** (raggruppamento, order by, min, max)
- Non si usa direttamente come linguaggio di interrogazione dei DBMS ma come rappresentazione interna delle interrogazioni.
- **Calcolo Relazionale**: linguaggio dichiarativo di tipo logico dal quale è stato derivato l'SQL.

Algebra relazionale

- Data una **relazione** $R (A1: T1, \dots, An: Tn)$
 - **Tipo:** $\{(A1: T1, \dots, An: Tn)\}$
 - **Grado:** n
 - Data una ennupla $t \in R$
t.Ai valore dell'attributo Ai
- Nel modello di base:
 - relazioni come **insiemi** di ennuple
 - **non** si usa **NULL**

- **Ridenominazione (ρ)**

Data una relazione $R(X)$, con X insieme di attributi, $A \in X$ e $B \notin X$

$$\rho_{A \leftarrow B}(R)$$

relazione R dove A è ridenominato con B

$$\rho_{A \leftarrow B}(R) = \{t \mid \exists u \in R. t.B = u.A \wedge \forall C \in X - \{A\}. t.C = u.C\}$$

- Grado della nuova relazione? Tipo? Cardinalità?

R e **S** relazioni dello stesso tipo:

- **Unione (\cup)**

$$R \cup S = \{t \mid t \in R \vee t \in S\}$$

- **Differenza ($-$)**

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

- Qual è il tipo del risultato? Quante ennuple contiene il risultato?

- Se t_1 è un'ennupla non in R , allora

$$R = (R \cup \{t_1\}) - \{t_1\}$$

- **Proiezione (π)**: data $R(X)$ con $\{A_1, \dots, A_m\} \subseteq X$

$$\pi_{A_1, A_2, \dots, A_m}(R)$$

“elimina” gli attributi diversi da A_1, \dots, A_m

$$\pi_{A_1, \dots, A_m}(R) = \{ \langle t.A_1, \dots, t.A_m \rangle \mid t \in R \}$$

- Qual è il tipo del risultato? Se R contiene n ennuple quante ne contiene il risultato?

Proprietà: se L_1 e L_2 sono insiemi di attributi con $L_1 \subseteq L_2$

$$\pi_{L_1}(\pi_{L_2}(R)) = \pi_{L_1}(R)$$

- Sia data la relazione Studenti

Studenti	Nome	Cognome	<u>Matricola</u>	Anno	Provincia
	Paolo	Verdi	71523	2005	VE
	Anna	Rossi	76366	2006	PD
	Giorgio	Zeri	71347	2005	VE
	Chiara	Scuri	71346	2006	VE

- Trovare il nome, la matricola e la provincia degli studenti

- Espressione nell'algebra

π Nome, Matricola, Provincia(Studenti)

Nome	<u>Matricola</u>	Provincia
Paolo	71523	VE
Anna	76366	PD
Giorgio	71347	VE
Chiara	71346	VE

π Provincia(Studenti) ?

- **Restrizione (selezione) (σ)**

$$\sigma_{\phi}(R) = \{t \mid t \in R \wedge \phi(t)\}$$

relazione le cui ennuple sono le ennuple di R che soddisfano la **Condizione** ϕ

- Condizione ϕ è una combinazione proposizionale di (dis)uguaglianze e disequazioni tra attributi (o tra attributi e costanti)

$$\phi ::= A_i \text{ op } A_j \mid A_i \text{ op } c \mid \neg \phi \mid \phi \wedge \phi \mid \phi \vee \phi$$

dove **op** è un operatore di confronto.

- La condizione riguarda attributi di **singole ennuple**

- Qual è il tipo del risultato? Se R contiene n ennuple quante ne ha il risultato?

- **Commutativa:**

$$\sigma_{C_1}(\sigma_{C_2}(R)) = \sigma_{C_1 \wedge C_2}(R) = \sigma_{C_2}(\sigma_{C_1}(R))$$

- Trovare i dati degli studenti della provincia di Venezia:

$\sigma_{\text{Provincia} = 'VE'} (\text{Studenti})$

Nome	Cognome	<u>Matricola</u>	Anno	Provincia
Paolo	Verdi	71523	2005	VE
Giorgio	Zeri	71347	2005	VE
Chiara	Scuri	71346	2006	VE

- Trovare il nome, la matricola e l'anno di iscrizione degli studenti di Venezia:

$\pi_{\text{Nome, Matricola, Anno}} (\sigma_{\text{Provincia} = 'VE'} (\text{Studenti}))$

Nome	<u>Matricola</u>	Anno
Paolo	71523	2005
Giorgio	71347	2005
Chiara	71346	2006

- **Prodotto (\times)**

$$R \times S$$

- R e S con attributi **distinti** A_1, \dots, A_n , e B_1, \dots, B_m
 - ennuple ottenute concatenando ennuple di R e ennuple di S
 - $R \times S = \{ \langle t.A_1, \dots, t.A_n, u.B_1, \dots, u.B_m \rangle \mid t \in R \wedge u \in S \}$
- Qual è il tipo del risultato? Se R e S contengono n e m ennuple quante ne contiene il risultato?

A	B
a1	b1
a2	b2

 \times

C	D
c1	d1
c2	d2
c3	d3

 $=$

A	B	C	D
a1	b1	c1	d1
a1	b1	c2	d2
a1	b1	c3	d3
a2	b2	c1	d1
a2	b2	c2	d2
a2	b2	c3	d3

- Qual è il risultato di **Studenti** \times **Esami** ?

Studenti

Nome	Cognome	<u>Matricola</u>	Anno	Provincia
Paolo	Verdi	71523	2005	VE
Anna	Rossi	76366	2006	PD
Giorgio	Zeri	71347	2005	VE
Chiara	Scuri	71346	2006	VE

Esami

<u>Codice</u>	Materia	Candidato*	Data	Voto	Lode
B112	BD	71523	08.07.06	27	N
F31	FIS	76366	08.07.07	26	N
B247	BD	76366	28.12.06	28	S

- Trovare il nome degli studenti che hanno superato l'esame di BD con 30

$$\pi_{\text{Nome}}(\sigma_{\text{Materia}='BD' \wedge \text{Voto}=30}(\sigma_{\text{Matricola}=\text{Candidato}}(\text{Studenti} \times \text{Esami})))$$

- si introduce un operatore derivato: la giunzione!

$$\pi_{\text{Nome}}(\sigma_{\text{Materia} = 'BD' \wedge \text{Voto} = 30}(\text{Studenti} \bowtie_{\text{Matricola}=\text{Candidato}} \text{Esami}))$$

- **Giunzione**: Utile per “combinare” informazioni di relazioni correlate

$$R \bowtie_{A_i=B_j} S$$

- R e S con attributi distinti A_1, \dots, A_n , e B_1, \dots, B_m
- ovvero

$$R \bowtie_{A_i=B_j} S = \sigma_{A_i=B_j}(R \times S)$$

$$\{\langle t.A_1, \dots, t.A_n, u.B_1, \dots, u.B_m \rangle \mid t \in R \wedge u \in S \wedge t.A_i = u.B_j\}$$

- **Giunzione naturale**

$$R \bowtie S$$

- Giunzione

Utenti \bowtie Prestiti
Codice=CodUtente

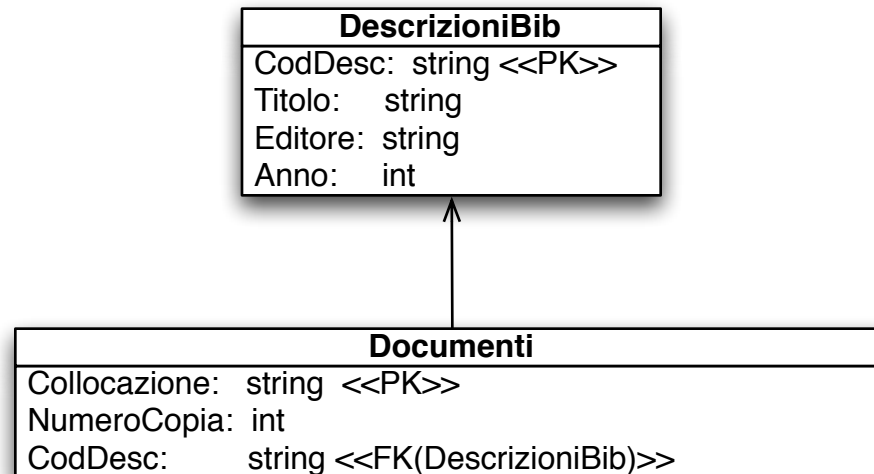
Utenti	
Codice:	string <<PK>>
NomeCognome:	string
Indirizzo:	string



Prestiti	
DataPrestito:	date
DataRestituzione:	date
CodUtente:	string <<FK(Utenti)>>
Collocazione:	string <<PK>> <<FK(Documenti)>>

- Giunzione naturale

Documenti ⋈ DescrizioniBib



- Giunzione esterna

- Intersezione

$$R \cap S$$

esprimibile come

$$R - (R - S)$$

- **Divisione**: date le relazioni $R(XY)$ e $S(Y)$ si vuole produrre una relazione $T(X)$ tale che una ennupla t è in T se e solo se **per ogni** s in S la ennupla $\langle t, s \rangle$ appare in R .

$$R \div S$$

- **Esempio**: matricola degli studenti che hanno fatto tutti gli esami che ha fatto Anna Rossi (matr. 76366).

- esami di Anna Rossi:

$$ES_AR = \pi_{Materia}(\sigma_{Candidato='76366'}(Esami))$$

- esami studenti con matricola

$$ES = \pi_{Candidato, Materia}(Esami)$$

- **Divisione**: date le relazioni $R(XY)$ e $S(Y)$ si vuole produrre una relazione $T(X)$ tale che una ennupla t è in T se e solo se **per ogni** s in S la ennupla $\langle t, s \rangle$ appare in R .

$$R \div S$$

- **Esempio**: matricola degli studenti che hanno fatto tutti gli esami che ha fatto Anna Rossi (matr. 76366).

- esami di Anna Rossi:

$$ES_AR = \pi_{Materia}(\sigma_{Candidato='76366'}(Esami))$$

- esami studenti con matricola

$$ES = \pi_{Candidato, Materia}(Esami)$$

- il risultato desiderato è quindi

$$ES \div ES_AR$$

- Usato per query che coinvolgono quantificazione universale
- Esprimibile come

$$\pi_X(R) - \pi_X((\pi_X(R) \times S) - R)$$

- Query per
 - studenti che hanno fatto un sottoinsieme degli esami di Anna Rossi
 - studenti che hanno fatto esattamente gli esami di Anna Rossi

- **Proiezione generalizzata**

$$\pi_{Exp_1 \text{ AS } A_1, Exp_2 \text{ AS } A_2, \dots, Exp_n \text{ AS } A_n}(R)$$

- Le espressioni Exp_i possono comprendere attributi, costanti, e operazioni su di essi

- **Esempio:** data una relazione $Utente(Codice, SalarioLordo, Trattenute, \dots)$

$$\pi_{Codice, SalarioLordo - Trattenute \text{ AS } Stipendio}(Utente)$$

- Le **funzioni di aggregazione** hanno come argomenti multinsiemi e ritornano come risultato un valore.
- **sum** ritorna la somma degli elementi
- **avg** ritorna la media degli elementi
- **count** ritorna il numero degli elementi
- **min** e **max** ritornano il minimo e il massimo valore degli elementi
- Se si vuole **ignorare eventuali duplicati**, si estende il nome della funzione con la stringa “**-distinct**”

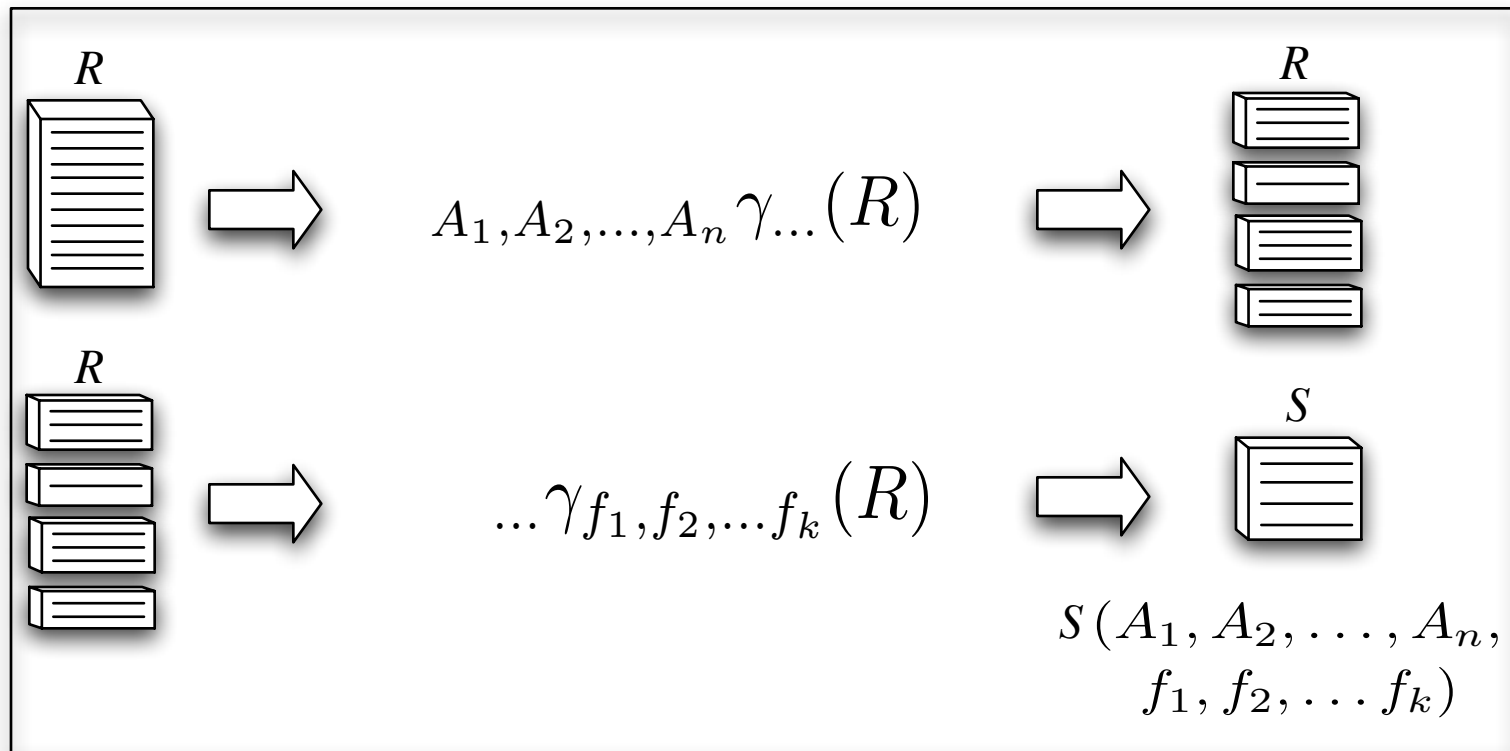
Altri Operatori (cont.)

- Raggruppamento (γ)

$$A_1, A_2, \dots, A_n \gamma f_1, f_2, \dots, f_k (R)$$

dove gli A_i sono attributi di R e le f_i sono espressioni che usano funzioni di aggregazione (min, max, count, sum, avg, ...)

$$S = A_1, A_2, \dots, A_n \gamma_{f_1, f_2, \dots, f_k} (R)$$



- Trovare per ogni candidato il numero degli esami, il voto minimo, massimo e medio

Candidato $\gamma_{\text{count}(*), \text{min}(\text{Voto}), \text{max}(\text{Voto}), \text{avg}(\text{Voto})}(\text{Esami})$

Materia	Candidato	Data	Voto	Lode
BD	71523	08.07.06	20	N
FIS	76366	08.07.07	26	N
ASD	71523	28.12.06	30	S
BD	76366	28.12.06	28	N

- raggruppamento

Materia	Candidato	Data	Voto	Lode
BD	71523	08.07.06	20	N
ASD	71523	28.12.06	30	S
FIS	76366	08.07.07	26	N
BD	76366	28.12.06	28	N

- calcolo delle funzioni

Candidato	Count(*)	min(Voto)	max(Voto)	avg(Voto)
71523	2	20	30	25
76366	2	26	28	27

- Proiezione senza l'eliminazione dei duplicati (multinsiemistica)

$$\pi_{A_1, A_2, \dots, A_n}^b(O)$$

- Eliminazione di duplicati

$$\delta(O)$$

- Ordinamento A_1, A_2, \dots, A_n attributi di O

$$\tau_{A_1, A_2, \dots, A_n}(O)$$

- Unione, Intersezione e Differenza

$$O_1 \cup^b O_2, O_1 \cap^b O_2, O_1 -^b O_2$$

- Basate su regole di equivalenza fra espressione algebriche
- Consentono di scegliere diversi ordini di join e di anticipare proiezioni e restrizioni.
- Alcuni esempi con la relazione $R(A, B, C, D)$:

$$\pi_A(\pi_{A,B}(R)) \equiv \pi_A(R)$$

$$\sigma_{C_1}(\sigma_{C_2}(R)) \equiv \sigma_{C_1 \wedge C_2}(R)$$

$$\sigma_{C_1 \wedge C_2}(R \times S) \equiv \sigma_{C_1}(R) \times \sigma_{C_2}(S)$$

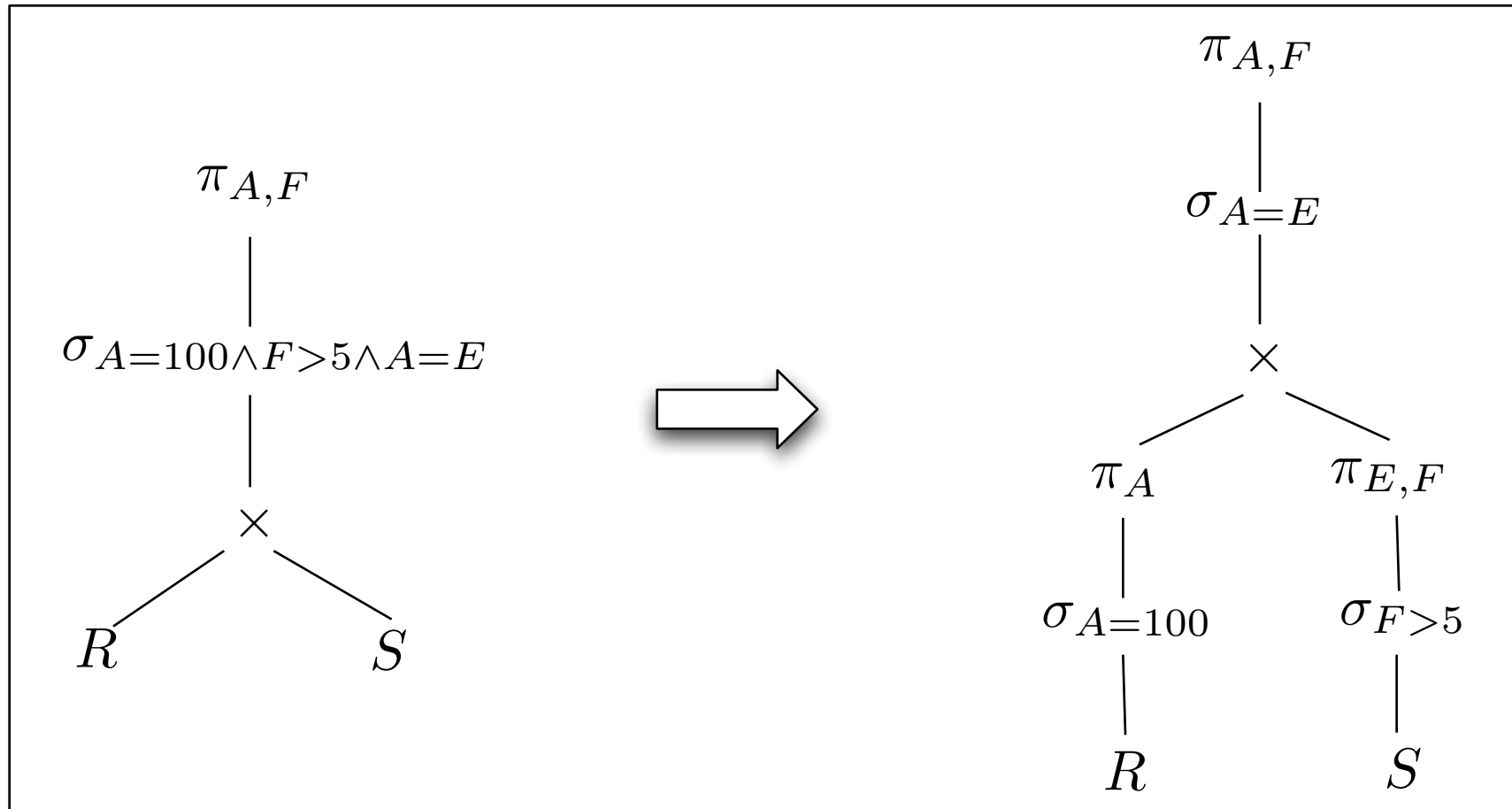
$$R \times (S \times T) \equiv (R \times S) \times T$$

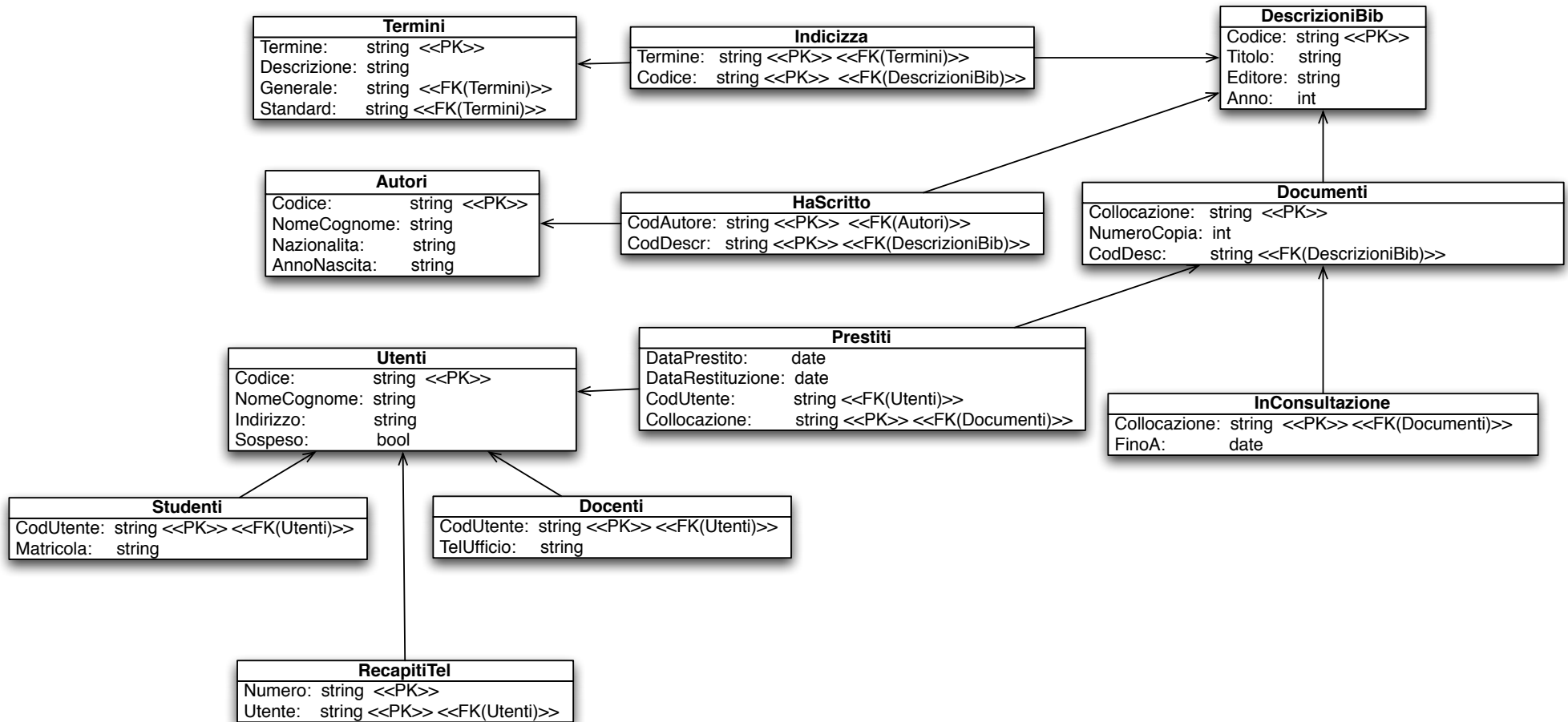
$$(R \times S) \equiv (S \times R)$$

$$\sigma_C(X \gamma_F(R)) \equiv_X \gamma_F(\sigma_C(R))$$

- Consideriamo le relazioni $R(A, B, C, D)$ e $S(E, F, G)$ e l'espressione:

$$\pi_{A,F}(\sigma_{A=100 \wedge F>5 \wedge A=E}(R \times S))$$





-
- Titolo e collocazione di tutti i documenti in prestito.
 - Nome e Cognome degli utenti che hanno documenti in prestito.

-
- Codice, Nome e Cognome di tutti gli utenti che:
 - sono studenti e hanno matricola < 7000
 - sono docenti e hanno numero di telefono tra 1200 e 1300.

 - Gli utenti (tutti gli attributi) che non hanno in prestito nessun libro.

- Codice degli utenti che hanno in prestito solo libri di fisica (si legga libro di fisica come documento la cui descrizione bibliografica è indicizzata da un termine che ha come standard “Fisica”). Si vuole una copia per ciascuna descrizione bibliografica.

-
- Codice degli utenti che hanno in prestito tutti i libri di fisica. Si vuole una copia per ciascuna descrizione bibliografica.

 - Codice degli utenti che hanno in prestito tutti e soli i libri di fisica. Si vuole una copia per ciascuna descrizione bibliografica.

-
- Nome, Cognome e Codice degli utenti che hanno in prestito più di tre libri.

 - Codice, Nome e Cognome degli autori che hanno scritto il massimo numero di libri.

Calcolo Relazionale

- L'algebra relazionale non è l'unico linguaggio formale di interrogazione per DB relazionali; un'alternativa è il **calcolo relazionale** (CR), del quale esistono due varianti:
 - **calcolo relazionale su ennuple** (CRE)
 - **calcolo relazionale su domini** (CRD)

- AR, CRE e CRD sono **espressivamente equivalenti**: ogni interrogazione esprimibile nell'uno è anche esprimibile negli altri.
- Un linguaggio relazionale espressivamente equivalente all'AR, al CRE e al CRD è detto **relazionalmente completo**
- i linguaggi dei DBMSs commerciali sono in genere non solo relazionalmente completi, ma anche di più ... in quanto includono anche altre funzionalità (e.g. aggregazione, raggruppamento, ...).

- AR è un linguaggio **procedurale**
 - un'interrogazione è una espressione che specifica, oltre a cosa va recuperato, le operazioni necessarie a recuperarlo;
- CR è un linguaggio **dichiarativo**
 - un'interrogazione è un'espressione che specifica cosa va recuperato, ma non come recuperarlo.
 - le operazioni da eseguire e la loro sequenzializzazione sono decise dal DBMS.
- Praticamente tutti i linguaggi dei DBMS relazionali commerciali sono implementazioni (più o meno fedeli ...) del CR; ad esempio SQL ~ CRE

- **termini**: denotano individui (elementi del dominio di interesse)

$$t ::= c \mid x \mid f(t_1, \dots, t_n)$$

- c costante
- x variabile
- f simbolo di funzione

- **formule**: denotano valori di verità (T o F);

$$\phi ::= p(t_1, \dots, t_n) \mid \neg \phi \mid \phi_1 \wedge \phi_2 \mid \phi_1 \rightarrow \phi_2 \mid \forall x. \phi \mid \exists x. \phi$$

- p simbolo di predicato n-ario

- Il CRE usa la logica del prim'ordine, interpretata su un dominio i cui elementi sono le ennuple della BD, per esprimere le interrogazioni
- **costanti** e le **variabili** sono di **tipo ennupla**.

- **Esempio di interrogazione:**

Nomi e cognomi degli studenti che hanno superato almeno un esame:

$$\{t.\text{Nome}, t.\text{Cognome} \mid t \in \text{Studenti} \wedge \exists e \in \text{Esami}.(t.\text{Matricola} = e.\text{Candidato})\}$$

- Un'interrogazione del CRE è un'espressione del tipo

$$\{ t_{i1}.A_1, \dots, t_{im}.A_m \mid \phi(t_1, \dots, t_n) \}$$

dove

- t_i variabili ennupla (il cui tipo, i.e. a quali relazioni appartengono, sarà indicato in ϕ);
- A_i simboli di funzione di tipo attributo ($t_i.A_i$ è una notazione alternativa per $A_i(t_i)$);
- $\phi(t_1, \dots, t_n)$ è una formula del prim'ordine in cui
 - le variabili t_1, \dots, t_n occorrono libere
 - il risultato è l'insieme delle ennuple $\langle t_{i1}.A_1, \dots, t_{im}.A_m \rangle$ tali che $\phi(t_1, \dots, t_n)$ è vera.

- Le formule atomiche possono essere

- formule di tipo

$$t \in \text{Studenti} \qquad e \in \text{Esami}$$

dichiara che t appartiene all'estensione corrente di *Studente*: quindi in ogni espressione $t.A$ nell'interrogazione, A deve essere un attributo di *Studente*;

- formule di confronto fra valori di attributi

$$t.\text{Matricola} = e.\text{Candidato}$$

- formule di confronto fra il valore di un attributo e un valore costante

$$t.\text{Provincia} = \text{'VE'}$$

● Restrizione

$\sigma_{\text{Provincia}='VE'}(\text{Studenti})$

$\{ t \mid t \in \text{Studenti} \wedge t.\text{Prov} = 'VE' \}$

● Proiezione

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

$\{ t. \text{Nome}, t. \text{Cognome} \mid t \in \text{Studenti} \}$

● Unione

$\text{Studenti} \cup \text{Docenti}$

$\{ t \mid t \in \text{Studenti} \vee t \in \text{Docenti} \}$

- Differenza

Studenti – Docenti

$$\{ t \mid t \in \text{Studenti} \wedge \neg (t \in \text{Docenti}) \}$$

- Prodotto

Studenti x Esami

$$\{ s, e \mid s \in \text{Studenti} \wedge e \in \text{Esami} \}$$

- Intersezione

Studenti \cap Docenti

$$\{ t \mid t \in \text{Studenti} \wedge t \in \text{Docenti} \}$$

-
- Esprimere nel calcolo relazionale
 - giunzione;
 - giunzione naturale.