Basi di Dati - IV (seconda parte)

Corso di Laurea in Informatica

Anno Accademico 2013/2014

Alessandra Raffaetà

raffaeta@dsi.unive.it

Linguaggi Relazionali

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, ...)
  - altri operatori (raggruppamento, order by, min, max)
  - Non si usa come 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

Notazione

- Data una **relazione R (A1: T1, ..., An: Tn)**
  - **Tipo**: {(A1: T1, ..., An: Tn)}
  - **Grado**: n
  - Data una ennupla t ∈ R

    t.Ai valore dell'attributo Ai

- Nel modello di base:
  - relazioni come **insiemi** di ennuple
  - non si usa **NULL**

Algebra Relazionale: Ridenominazione

- **Ridenominazione (δ)**

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

$$\delta_{A \to B}(R)$$

  relazione R dove A è ridenominato con B

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

- Grado della nuova relazione? Tipo? Cardinalità?

Algebra Relazionale: Unione e Differenza

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

- **Unione (∪)**

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

- **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

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

Algebra Relazionale: Proiezione

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

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

"elimina" gli attributi diversi da A1, ..., Am

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

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

- **Proprietà**: se L1 e L2 sono insiemi di attributi con L1 ⊆ L2

$$\pi_{L1}(\pi_{L2}(R)) = \pi_{L1}(R)$$

Proiezione: Esempi

- Sia data la relazione Studenti

Studenti

Nome	Cognome	<u>Matricola</u>	Anno	Prov
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

Proiezione: Esempi

- Espressione nell'algebra

$$\pi_{\text{Nome, Matricola, Provincia}}(\text{Studenti})$$

Nome	Matricola	Provincia
Paolo	71523	VE
Anna	76366	PD
Giorgio	71347	VE
Chiara	71346	VE

$$\pi_{\text{Provincia}}(\text{Studenti}) \; ?$$

Algebra Relazionale: Restrizione

- **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)

$$\phi \;::=\; A_i\, op\, A_j \;\mid\; A_i\, op\, c \;\mid\; \neg\phi \;\mid\; \phi \wedge \phi \;\mid\; \phi \vee \phi$$

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

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

Algebra Relazionale: Restrizione (cont.)

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

- Commutativa:

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

Restrizione: Esempi

- Trovare i dati degli studenti della provincia di Venezia:

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

Nome	Cognome	Matricola	Anno	Prov
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:

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

Nome	Matricola	Anno
Paolo	71523	2005
Giorgio	71347	2005
Chiara	71346	2006

Algebra Relazionale: Prodotto

- **Prodotto (×)**

$$R \times S$$

  - R e S con attributi distinti A1, ..., An, e B1, ..., Bm
  - ennuple ottenute concatenando ennuple di R e ennuple di S
  - $R \times S = \{\langle t.A_1, \ldots, t.A_n, u.B_1, \ldots, u.B_m \rangle \mid t \in R \ \wedge \ u \in S\}$

- Qual è il tipo del risultato? Se R e S contengono n e m ennuple quante ne contiene il risultato?

Prodotto: Esempio

A	B
a1	b1
a2	b2

×

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

Algebra Relazionale: Esempi

- Qual è il risultato di Studenti × Esami ?

Studenti

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

Esami

Codic	Materia	Candidato*	Data	Voto	Lode
B112	BD	71523	08.07.06	27	N
F31	FIS	76366	08.07.07	26	N
B247	BD	71523	28.12.06	30	S

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

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

- si introduce un operatore derivato: la giunzione!

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

Operatori derivati: Giunzione (o Join)

- **Giunzione**: Utile per "combinare" informazioni di relazioni correlate

$$R \underset{A_i = B_j}{\bowtie} S$$

  - R e S con attributi distinti A1, ..., An, e B1, ..., Bm
  - ovvero

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

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

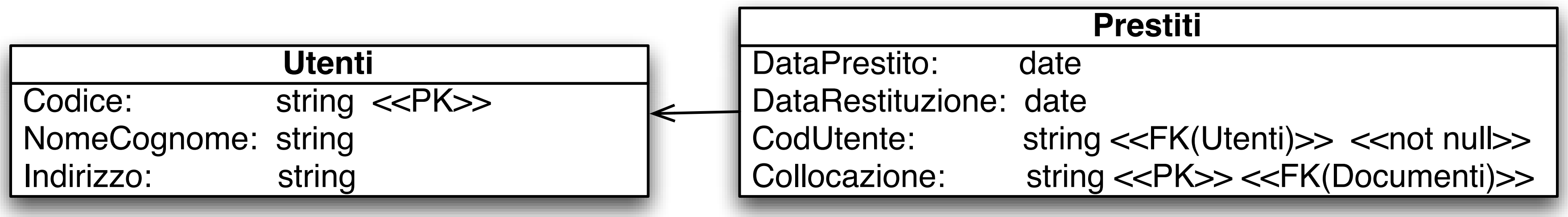
- **Giunzione naturale**

$$R \bowtie S$$

Esempio

- Giunzione

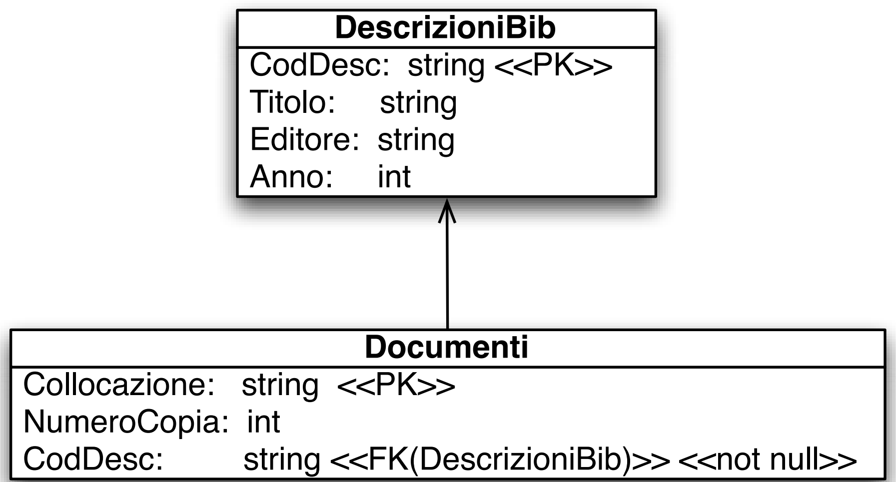
$$\text{Utenti} \underset{\text{Codice}=\text{CodUtente}}{\bowtie} \text{Prestiti}$$



Esempio

- Giunzione naturale

$$\text{Documenti} \bowtie \text{DescrizioniBib}$$



Operatori Derivati: intersezione

- **Giunzione esterna**

- **Intersezione**

$$R \cap S$$

esprimibile come

$$R - (R - S)$$

Altri Operatori

- **Proiezione generalizzata**

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

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

- **Esempio**: data una relazione Utente(Codice,SalarioLordo,Trattenute, ...)

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

Altri Operatori

- Proiezione senza l'eliminazione dei duplicati (multinsiemistica)

$$\pi^{b}_{A_1, A_2, \ldots, A_n}(R)$$

- Ordinamento (liste di ennuple)

$$\tau_{A_1, A_2, \ldots, A_n}(R)$$

Altri Operatori (cont.)

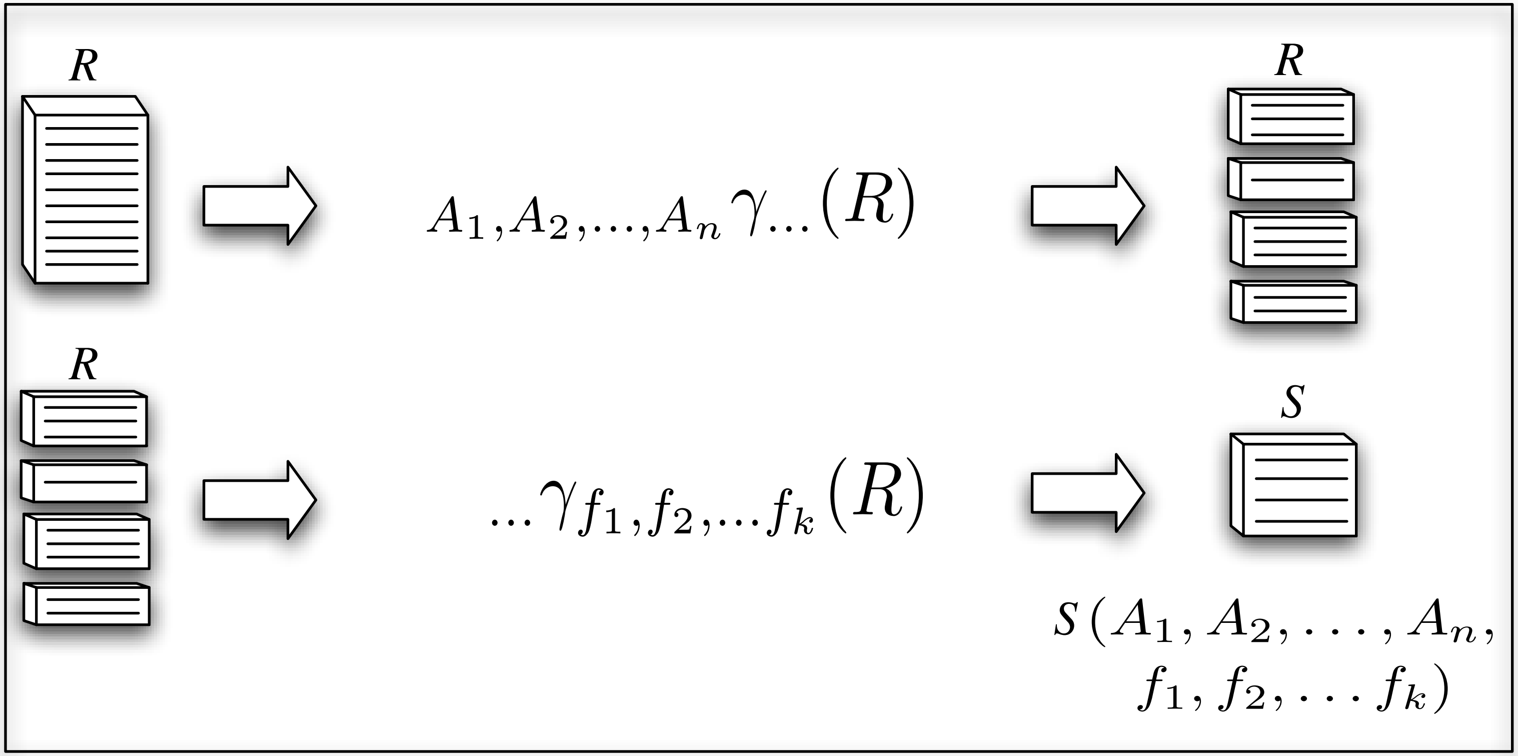
- Raggruppamento (γ)

$$_{A_1, A_2, \ldots, A_n}\gamma_{f_1, f_2, \ldots f_k}(R)$$

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

Significato del raggruppamento (γ)

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



Esecuzione del raggruppamento

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

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

Materi	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

Esecuzione del raggruppamento (cont.)

- raggruppamento

Materi	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

Trasformazioni Algebriche

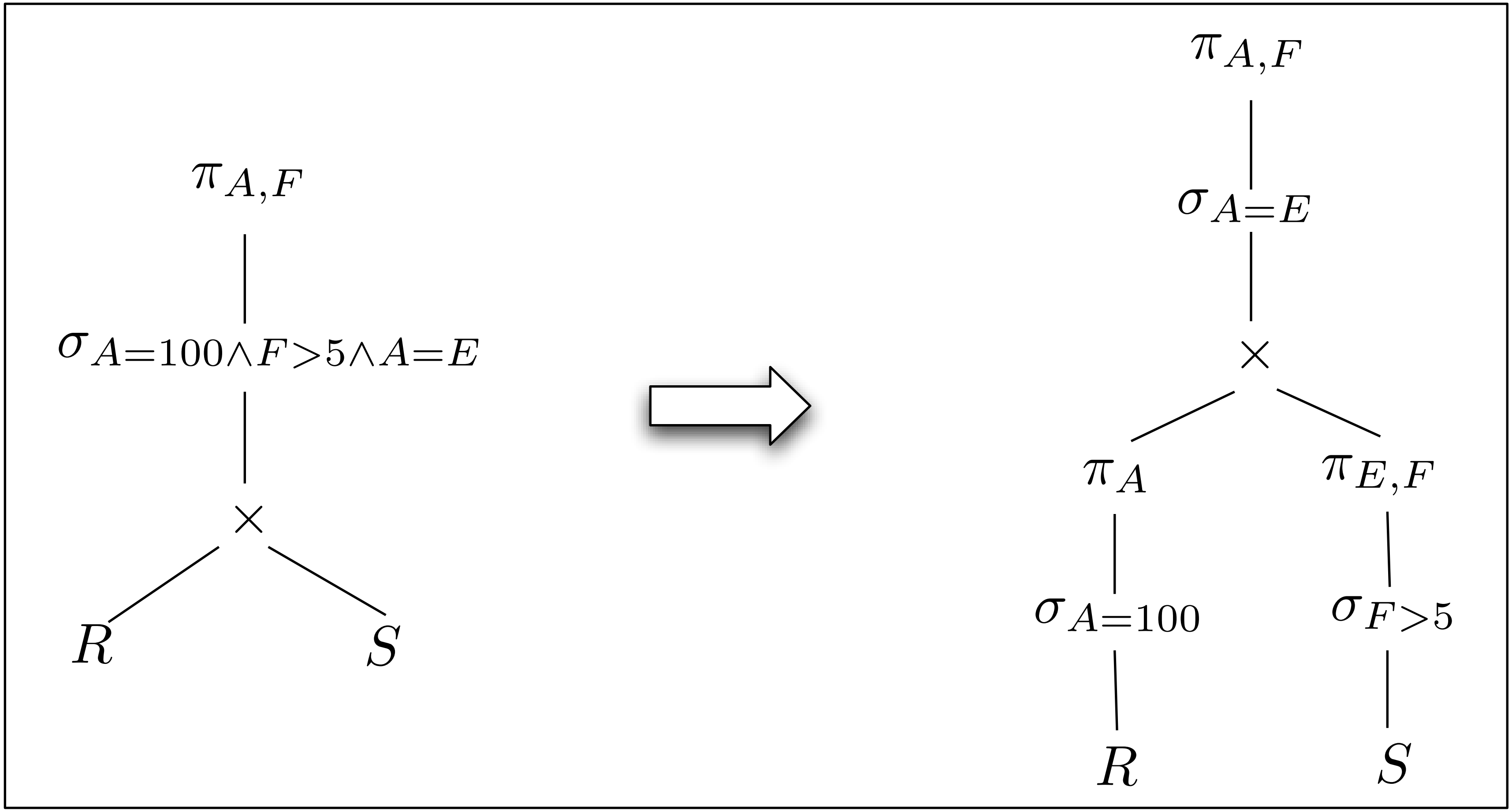
- Basate su regole di equivalenza fra espressione algebriche
- Consentono di scegliere diversi ordini di join e di anticipare proiezioni e restrizioni.
- Alcuni esempi con la relazione R(A, B, C, D):

$$
\begin{aligned}
&\pi_A(\pi_{A,B}(R)) \equiv \pi_A(R)\\
&\sigma_{C_1}(\sigma_{C_2}(R)) \equiv \sigma_{C_1 \wedge C_2}(R)\\
&\sigma_{C_1 \wedge C_2}(R \times S) \equiv \sigma_{C_1}(R) \times \sigma_{C_2}(S)\\
&R \times (S \times T) \equiv (R \times S) \times T\\
&(R \times S) \equiv (S \times R)\\
&\sigma_C({}_X\gamma_F(R)) \equiv_X \gamma_F(\sigma_C(R))
\end{aligned}
$$

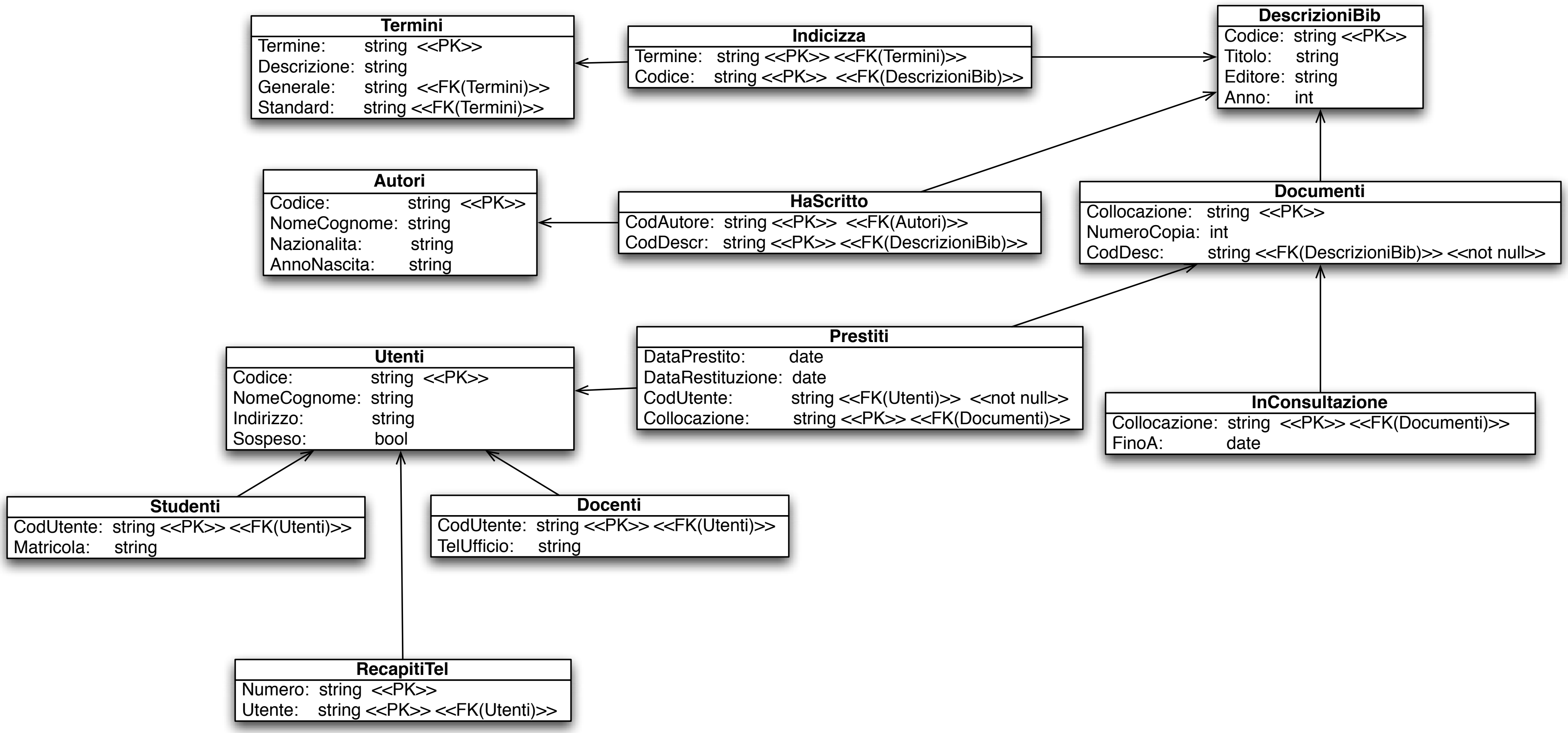
Alberi Logici e Trasformazioni Algebriche

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

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



Esempio: Biblioteca (semplificato)



Query sulla biblioteca

- 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 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")

Query sulla biblioteca (cont.)

- Codice degli utenti che hanno in prestito tutti i libri (almeno un documento per ogni descrizione bibliografica).
- Codice degli utenti che hanno in prestito tutti i libri di fisica.
- Codice degli utenti che hanno in prestito tutti e soli i libri di fisica.
- 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

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)

Completezza Relazionale

- 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, ...).

Differenze fra AR e CR

- 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

Logica del prim'ordine

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

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

  - c costante
  - x variabile
  - f simbolo di funzione

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

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

  - p simbolo di predicato n-ario

Calcolo Relazionale su ennuple

- 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:

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

Calcolo Relazionale su ennuple

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

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

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 $A_i(t_i)$);
- $\phi(t_1,\ldots t_n)$ è una formula del prim'ordine in cui
  - le variabili $t_1,\ldots t_n$ occorrono libere
  - il risultato è l'insieme delle ennuple $< t_{i1}.A_1, \ldots, t_{im}.A_m >$ tali che $\phi(t_1,\ldots t_n)$ è vera.

Calcolo Relazionale su ennuple

- Le formule atomiche possono essere
  - formule di tipo

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

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

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

Esprimibilità dell'AR in CRE

- Selezione

  $\sigma_{Prov=\text{'VE'}}(\text{Studente})$

  $\{ t \mid t \in \text{Studente} \wedge t.\text{Provincia}=\text{'VE'} \}$

- Proiezione

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

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

- Unione

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

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

Esprimibilità dell'AR in CRE (cont.)

- Differenza

  Studenti − Docenti

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

- Prodotto cartesiano

  Studenti x Esami

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

- Intersezione

  Studenti ∩ Docenti

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

Esercizi

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