Fondamenti di Teoria delle Basi di Dati

Riccardo Torlone

Parte 3: Algebra relazionale

Linguaggi per basi di dati

- operazioni sullo schema
  - DDL: data definition language
- operazioni sui dati
  - DML: data manipulation language
    - interrogazione ("query")
    - aggiornamento

Linguaggi di interrogazione

- Dichiarativi
  - specificano le proprietà del risultato ("che cosa")
- Procedurali
  - specificano le modalità di generazione del risultato ("come")

Linguaggi di interrogazione relazionali

- Algebra relazionale: procedurale (teorico/pratico)
- Calcolo relazionale:
  dichiarativo (teorico)
- SQL (Structured Query Language):
  parzialmente dichiarativo (pratico)
- QBE (Query by Example):
  dichiarativo (pratico)

Nozioni di base

- Sia U_A un insieme infinito contabile di attributi e assumiamo che tutti gli attributi in U_A abbiano lo stesso dominio D
- Sia U_r l'insieme di tutte le possibili relazioni su attributi in U_A
- Dato uno schema di basi di dati $\mathbf{R}$, sia infine $I(\mathbf{R})$ l'insieme di tutte le sue istanze
- Una query Q è una funzione ricorsiva parziale

$$\mathrm{Q} : I(\mathbf{R}) \rightarrow U_r$$

- Data **r** in I(**R**), Q(**r**) indica il risultato di Q applicato ad **r**

Espressioni

- Le query si esprimono mediante espressioni la cui sintassi è definita da un linguaggio
- Una espressione E rappresenta una query Q se la funzione definita da E è Q
- Due espressioni E_1 ed E_2 sono equivalenti ($E_1 \equiv E_2$) su uno schema **R** se $E_1(\mathbf{r})=E_2(\mathbf{r})$ per ogni istanza **r** di I(**R**).
- Si può definire anche il contenimento di query
- Un linguaggio L_1 è espressivo almeno quanto un altro linguaggio L_2 se per ogni espressione E_2 di L_2 c'è una espressione E_1 di L_1 tale che $E_1 \equiv E_2$

Algebra relazionale

- Insieme di operatori
  - su relazioni
  - che producono relazioni
  - e possono essere composti

Operatori dell'algebra relazionale

- unione, intersezione, differenza
- ridenominazione
- selezione
- proiezione
- join (e varianti)

Operatori insiemistici

- le relazioni sono insiemi
- i risultati debbono essere relazioni
- è possibile applicare unione, intersezione, differenza solo a relazioni definite sugli stessi attributi

Unione

Laureati

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45

Quadri

Matricola	Nome	Età
9297	Neri	33
7432	Neri	54
9824	Verdi	45

Laureati $\cup$ Quadri

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45
9297	Neri	33

Intersezione

Laureati

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45

Quadri

Matricola	Nome	Età
9297	Neri	33
7432	Neri	54
9824	Verdi	45

Laureati ∩ Quadri

Matricola	Nome	Età
7432	Neri	54
9824	Verdi	45

Differenza

Laureati

Matricola	Nome	Età
7274	Rossi	42
7432	Neri	54
9824	Verdi	45

Quadri

Matricola	Nome	Età
9297	Neri	33
7432	Neri	54
9824	Verdi	45

Laureati – Quadri

Matricola	Nome	Età
7274	Rossi	42

Ridenominazione

- operatore monadico (con un argomento)
- "modifica lo schema" lasciando inalterata l'istanza dell'operando

Paternità

Padre	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

$\rho_{\text{Genitore} \leftarrow \text{Padre}}$ **(Paternità)**

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

$\rho_{\text{Genitore} \leftarrow \text{Padre}}$ **(Paternità)**

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco

$\rho_{\text{Genitore} \leftarrow \text{Madre}}$ **(Maternità)**

Genitore	Figlio
Eva	Abele
Eva	Set
Sara	Isacco

$\rho_{\text{Genitore} \leftarrow \text{Padre}}$ **(Paternità)**

$\cup$

$\rho_{\text{Genitore} \leftarrow \text{Madre}}$ **(Maternità)**

Genitore	Figlio
Adamo	Abele
Adamo	Caino
Abramo	Isacco
Eva	Abele
Eva	Set
Sara	Isacco

Definizione della ridenominazione

- Sia *r(X)* una relazione ed *f* una funzione iniettiva su *X*

$$\rho_f(r) = \{t \mid \exists t' \in r \ \ such \ \ that \\ t'[A] = t[f(A)], \forall A \in X\}$$

Richiamo sulle funzioni

$$f : S \rightarrow T$$

- S dominio, T codominio,
- f è iniettiva se per ogni s_1 ed s_2 in S tali che $s_1 \neq s_2$ risulta $f(s_1) \neq f(s_2)$
- f è suriettiva se per ogni $t \in T$ esiste $s \in S$ tale che $f(s) = t$
- f e una biezione se è iniettiva e suriettiva
- Se $f \circ g$ è l'identità f (g) è l'inversa sinistra (destra) di g (f)
- Teorema:
  - Una funzione è iniettiva sse ha un'inversa sinistra
  - Una funzione è suriettiva sse ha un'inversa destra

Selezione

- operatore monadico
- produce un risultato che
  - ha lo stesso schema dell'operando
  - contiene un sottoinsieme delle tuple dell'operando,
  - quelle che soddisfano una condizione

Impiegati

Matricola	Cognome	Filiale	Stipendio
7309	Rossi	Roma	55
5998	Neri	Milano	64
9553	Milano	Milano	44
5698	Neri	Napoli	64

- impiegati che
  - guadagnano più di 50
  - guadagnano più di 50 e lavorano a Milano
  - hanno lo stesso nome della filiale presso cui lavorano

Selezione, sintassi e semantica

- sintassi

$$\sigma_{Condizione}\ (Operando)$$

  - *Condizione*: espressione booleana
- semantica
  - il risultato contiene le tuple dell'operando che soddisfano la condizione

Semantica più rigorosa

- Formule proposizionali sono:
  - Atomi su X del tipo $A_1 \Theta A_2$ oppure $A \Theta c$ dove
    - A_1 e A_2 sono in X
    - a è una costante
    - Θ è un simbolo di confronto
  - Se F_1 ed F_2 sono formule proposizionali su X allora
    - $\neg F_1$, $F_1 \wedge F_2$, $F_1 \vee F_2$ sono formule su X
- Semantica di formule proposizionali:
  - F è vera su una tupla t definita su X se:
    - $F=A_1 \Theta A_2$ ($A \Theta c$) e $t[A_1] \Theta t[A_2]$ ($t[A] \Theta c$) è vero
    - $F = \neg F_1$, $F = F_1 \wedge F_2$, $F = F_1 \vee F_2$ e valgono le regole dei connettivi logici su F_1 ed F_2

Definizione dell'operatore di selezione

- Sia *r(X)* una relazione ed *F* una f.p. su *X*

$$\sigma_F(r) = \{t \ on \ X \mid t \in r \ and \ F(t) = vero\}$$

Proiezione

- operatore monadico
- produce un risultato che
  - ha parte degli attributi dell'operando
  - contiene tuple cui contribuiscono tutte le tuple dell'operando

Impiegati

Matricola	Cognome	Filiale	Stipendio
7309	Neri	Napoli	55
5998	Neri	Milano	64
9553	Rossi	Roma	44
5698	Rossi	Roma	64

- per tutti gli impiegati:
  - matricola e cognome
  - cognome e filiale

Proiezione, sintassi e semantica

- sintassi

$$\pi_{ListaAttributi}\ (Operando)$$

- semantica
  - il risultato contiene le tuple ottenute da tutte le tuple dell'operando ristrette agli attributi nella lista

Semantica più rigorosa

- Sia *r(X)* una relazione ed $Y \subseteq X$

$$\pi_Y(r) = \{t\ on\ Y \mid \exists t' \in r\ \ s.t.\ t'[Y] = t\}$$

- matricola e cognome di tutti gli impiegati

Matricola	Cognome
7309	Neri
5998	Neri
9553	Rossi
5698	Rossi

$$\pi_{\text{Matricola, Cognome}}\ (\text{Impiegati})$$

Join naturale

- operatore binario (generalizzabile)
- produce un risultato
  - sull'unione degli attributi degli operandi
  - con tuple costruite ciascuna a partire da una tupla di ognuno degli operandi

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
A	Mori
B	Bruni

Impiegato	Reparto	Capo
Rossi	A	Mori
Neri	B	Bruni
Bianchi	B	Bruni

- ogni tupla contribuisce al risultato:
  - join completo

Un join non completo

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
B	Mori
C	Bruni

Impiegato	Reparto	Capo
Neri	B	Mori
Bianchi	B	Mori

Un join vuoto

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
D	Mori
C	Bruni

Impiegato	Reparto	Capo

Un join completo, con n x m tuple

Impiegato	Reparto
Rossi	B
Neri	B

Reparto	Capo
B	Mori
B	Bruni

Impiegato	Reparto	Capo
Rossi	B	Mori
Rossi	B	Bruni
Neri	B	Mori
Neri	B	Bruni

Definizione dell’operatore di join

- Siano r(X) ed r(Y) due relazioni

$$r_1 \bowtie r_2 = \{t \; on \; XY \mid \exists t_1 \in r_1 \; and \; t_2 \in r_2$$
$$s.t. \; t[X] = t_1[X] \; and \; t[Y] = t_2[Y]\}$$

Join e proiezioni

Impiegato	Reparto
Rossi	A
Neri	B
Bianchi	B

Reparto	Capo
B	Mori
C	Bruni

Impiegato	Reparto	Capo
Neri	B	Mori
Bianchi	B	Mori

Impiegato	Reparto
Neri	B
Bianchi	B

Reparto	Capo
B	Mori

Proiezioni e join

Impiegato	Reparto	Capo
Neri	B	Mori
Bianchi	B	Bruni
Verdi	A	Bini

Impiegato	Reparto
Neri	B
Bianchi	B
Verdi	A

Reparto	Capo
B	Mori
B	Bruni
A	Bini

Impiegato	Reparto	Capo
Neri	B	Mori
Neri	B	Bruni
Bianchi	B	Mori
Bianchi	B	Bruni
Verdi	A	Bini

Formalizziamo

Consideriamo n relazioni $r_1(X_1), \ldots, r_n(X_n)$. Valgono le seguenti proprietà:

- $\sigma_F(\bowtie_{j=1}^{n} r_j) = \bowtie_{j=1}^{n} (\sigma_F(r_j))$
- $\pi_{X_i}(\bowtie_{j=1}^{n} r_j) \subseteq r_i$, for $i = 1, \ldots, n$
- $\pi_{X_i}(\bowtie_{j=1}^{n} r_j) = r_i$, for $i = 1, \ldots, n$ iff $r_1 \ldots, r_n$ hanno un join completo
- $\bowtie_{j=1}^{n} (\pi_{X_j}(r)) \supseteq r$

Si dice che r ha una decomposizione lossless rispetto a $X_1, \ldots, X_n$ se $\bowtie_{j=1}^{n} (\pi_{X_j}(r)) = r$

Una dimostrazione

Lemma *Siano* $r_1(X_1), \ldots, r_n(X_n)$ n *relazioni, risulta* $\pi_{X_i}(\bowtie_{j=1}^{n} r_j) \subseteq r_i$ *per* $i = 1, \ldots, n$.

Proof L'enunciato si dimostra facendo vedere che per un certo $1 \leq i \leq n$, se $t \in \pi_{X_i}(\bowtie_{j=1}^{n} r_j)$, allora $t \in r_i$.

Se $t \in \pi_{X_i}(\bowtie_{j=1}^{n} r_j)$ allora per definizione di proiezione esiste $t' \in \bowtie_{j=1}^{n} r_j$ tale che $t'[X_i] = t$. Per definizione di join esistono $t_1 \in r_1(X_1), \ldots,$ $t_n \in r_n(X_n)$ tali che $t_1[X_1] = t'[X_1], \ldots, t_n[X_n] = t'[X_n]$. Quindi esiste $t_i[X_i] = t'[X_i] = t \in r_i(X_i)$. C.v.d.

Chiavi e dipendenze funzionali

Definition *Data una relazione* $r(X)$, $K \subseteq X$ *è una* **superchiave** *per* r *se non esistono due tuple* t_1 *e* t_2 *in* r *tali che* $t_1[K] = t_2[K]$ *ed è* **chiave** *se nessun sottoinsieme proprio di* K *è una superchiave.*

Definition *Let* $r(X)$ *be a relation a let* Y *and* Z *be two subsets of* X*, then* r *satisfies the functional dependency (FD)* $Y \rightarrow Z$ *if* $\forall t_1, t_2 \in r$ *if* $t_1[Y] = t_2[Y]$ *then* $t_1[Z] = t_2[Z]$*.*

Lossless decomposition & FDs

Definition Let $r(U)$ be a relation and X, X_1 and X_2 be three subsets of U with $X_1 X_2 = U$ and $X_1 \cap X_2 = X$. Then r has a lossless decomposition w.r.t. X_1 and X_2 if $r = \pi_{X_1}(r) \bowtie \pi_{X_2}(r)$

Theorem *Let $r(U)$ be a relation and X, X_1 and X_2 be three subsets of U with $X_1 X_2 = U$ and $X_1 \cap X_2 = X$. Then r has a lossless decomposition w.r.t. X_1 and X_2 if it satisfies at least one of the FDs $X \to X_1$, $X \to X_2$.*

Proof

$r \subseteq \pi_{X_1}(r) \bowtie \pi_{X_2}(r)$: trivial

Proof (cont.)

$$r \supseteq \pi_{X_1}(r) \bowtie \pi_{X_2}(r)$$

Assume that r satisfies $X \rightarrow X_2$ and let $t \in \pi_{X_1}(r) \bowtie \pi_{X_2}(r)$.

By definition of join there are $t_1 \in \pi_{X_1}(r)$ and $t_2 \in \pi_{X_2}(r)$ such that $t_1[X_1] = t[X_1]$, $t_2[X_2] = t[X_2]$ and $t_1[X] = t[X] = t_2[X]$.

By definition of projection there two tuples $t'_1, t'_2 \in r$ such that $t'_1[X_1] = t_1[X_1]$, $t'_2[X_2] = t_2[X_2]$ and, since $t_1[X] = t_2[X]$, $t'_1[X] = t'_2[X]$.

Since $X \rightarrow X_2$ and $t'_1[X] = t'_2[X]$, it follows that $t'_1[X_2] = t'_2[X_2]$ and so $t'_1[X_1] = t_1[X_1] = t[X_1]$ and $t'_1[X_2] = t_2[X_2] = t[X_2]$. Then we have that $t = t'_1 \in r$. c.v.d.

Esercizi

Siano date n relazioni $r_1(X_1), \ldots, r_n(X_n)$. Dimostrare che valgono le seguenti proprietà:

- $\bowtie_{j=1}^{n} (\pi_{X_j}(r)) \supseteq r$
- $\bowtie_{j=1}^{n} (\pi_{X_j}(r)) \not\subseteq r$
- $\sigma_F(\bowtie_{j=1}^{n} r_j) = \bowtie_{j=1}^{n} (\sigma_F(r_j))$