Calcolo relazionale

Prof. Alberto Belussi

(anno accademico 2008-09)

Calcolo relazionale

E' un linguaggio di interrogazione dichiarativo: specifica le proprietà del risultato dell'interrogazione.

Esistono due versioni del calcolo relazionale:

- Il calcolo relazionale sui domini
- Il calcolo relazionale sulle tuple

Calcolo relazionale sui domini

E’ la versione più vicina al calcolo dei predicati del primo ordine.

Differenze rispetto al calcolo dei predicati:

- I simboli di predicato corrispondono alle relazioni della base di dati
- I simboli di funzione non ci sono
- Interessano solo le **formule aperte** che rappresentano le interrogazioni: il risultato è costituito dalle tuple di valori che, sostituite alle variabili libere rendono vera la formula stessa.

Calcolo relazionale sui domini

Approccio intuitivo

Le espressioni del calcolo relazionale sui domini hanno la seguente forma:

$$e = \{ A_1: x_1, \ldots, A_k: x_k \mid f \}$$

dove:

- $A_1, \ldots, A_k$ sono gli attributi del risultato
- $x_1, \ldots, x_k$ sono variabili
- f è una formula

Calcolo relazionale sui domini

La relazione risultato si ottiene applicando l'espressione ***e*** ad una istanza della base di dati e ha le seguenti caratteristiche:

- schema: { A_1, …, A_k },
- contenuto: insieme delle tuple t sullo schema {A_1, …, A_k} che rendono vera la formula ***f***.

Calcolo relazionale sui domini

TRENO(**Num***ero*, **Cat***egoria*, **Part***enza*, **Arr***ivo*, **Dest***inazione*)

FERMATA(**Treno**, **Staz***ione*, **Orario**)

Calcolo relazionale sui domini

Esempio

"Trovare l'orario di partenza di tutti i treni con destinazione Venezia, riportando il numero del treno, la categoria e l'orario di partenza"

{ Num: n, Cat: c, Part: p |
TRENO(Num: n, Cat: c, Part: p, Arr: a, Dest: d)
$\wedge$ d = "Venezia" }

Calcolo relazionale sui domini

<u>Esempio</u>

“Trovare i treni che partono dopo le 12:00 e fermano a Bologna, riportando il numero del treno, la destinazione e l’orario di partenza”

$$
\begin{aligned}
\{\ & \text{Num: } n, \text{ Dest: } d, \text{ Part: } p \mid \\
& \text{TRENO(Num: } n, \text{ Cat: } c, \text{ Part: } p, \text{ Arr: } a, \text{ Dest: } d) \\
& \wedge\ p > \text{“12:00”} \\
& \wedge\ \text{FERMATA(Treno: } n, \text{ Staz: } s, \text{ Orario: } o) \\
& \wedge\ s = \text{“Bologna”}\}
\end{aligned}
$$

Calcolo relazionale sui domini

SINTASSI

Simboli:

- costanti: $c \in D$ (dominio unico)
- variabili: $x_i \in V$ (insieme numerabile disgiunto dal dominio D)
- nomi di relazioni e attributi (dallo schema della base di dati)
- operatori di confronto: $\{\neq, =, \geq, \leq, >, <\}$
- connettivi logici: $\{\wedge, \vee, \neg\}$
- quantificatori: $\{\forall, \exists\}$

Calcolo relazionale sui domini

SINTASSI

Formule:

ATOMICHE:

- $R(A_1: x_1, \ldots, A_p: x_p)$ dove $R(A_1, \ldots, A_p)$ è lo schema di una relazione e $x_1, \ldots, x_p$ sono variabili.
- $x_1 \, \theta \, x_2$ o $x_1 \, \theta \, c$ dove x_1, x_2 sono variabili, c è una costante e θ è un operatore di confronto.

NON ATOMICHE:

- se f_1 e f_2 sono formule, allora anche $f_1 \wedge f_2$, $f_1 \vee f_2$, $\neg f_1$ e $\neg f_2$ lo sono.
- se f è una formula, allora anche $\forall x(f)$ e $\exists x(f)$ lo sono.

Calcolo relazionale sui domini

SINTASSI

Espressioni

$$\underbrace{\{A_1: x_1, \ldots, A_k: x_k}_{\text{Target List}} \mid \underset{\substack{\uparrow \\ \text{Formula}}}{f}\}$$

Calcolo relazionale sui domini

SINTASSI

Variabili libere o legate nelle formule

- nelle formule atomiche tutte le variabili sono LIBERE.
- nelle formule $\forall x(f)$ e $\exists x(f)$, *x* è una variabile LEGATA, mentre tutte le altre variabili sono LIBERE o LEGATE se LIBERE o LEGATE in *f*.
- le congiunzioni, disgiunzioni e la negazione non cambiano l'insieme delle variabili LIBERE o LEGATE.

Calcolo relazionale sui domini

SEMANTICA

Rispetto allo schema $R = \{R_1(X_1), \ldots, R_n(X_n)\}$ e da una istanza $r = \{r_1, \ldots, r_n\}$

1. $R_j(A_1\!: x_1, \ldots, A_p\!: x_p)$ è vera sui valori $(a_1, \ldots, a_p)$, se la relazione r_j in r contiene una tupla $(a_1, \ldots, a_p)$.
2. $x_1 \; \theta \; x_2$ è vera sui valori (a_1, a_2), se il confronto $a_1 \; \theta \; a_2$ è soddisfatto.

Calcolo relazionale sui domini

SEMANTICA

3. $x_1\ \theta\ c$ è vera sul valore a_1, se il confronto $a_1\ \theta\ c$ è soddisfatto.
4. $f_1 \vee f_2$ *con variabili libere* $(x_1, \ldots, x_q)$ è vera sui valori $(a_1, \ldots, a_q)$, se almeno una delle due formule f_1, f_2 è vera quando si sostituiscono i valori $(a_1, \ldots, a_q)$ *alle variabili* $(x_1, \ldots, x_q)$.
5. $f_1 \wedge f_2$ e $\neg f_1$ $(\neg f_2)$ sono vere secondo le usuali definizioni dei connettivi logici (tabelle di verità).

Calcolo relazionale sui domini

SEMANTICA

6. $\exists x(f)$ con variabili libere $(x_1, \ldots, x_q)$ è vera sui valori $(a_1, \ldots, a_q)$, se **<u>esiste almeno</u>** un valore $a \in D$ tale che f è vera quando si sostituiscono i valori $(a, a_1, \ldots, a_q)$ *alle variabili* $(x, x_1, \ldots, x_q)$.
7. $\forall x(f)$ con variabili libere $(x_1, \ldots, x_q)$ è vera sui valori $(a_1, \ldots, a_q)$, se **<u>per ogni</u>** valore $a \in D$, f è vera quando si sostituiscono i valori $(a, a_1, \ldots, a_q)$ *alle variabili* $(x, x_1, \ldots, x_q)$.

Calcolo relazionale sui domini

SEMANTICA

Interpretazione di una espressione del calcolo come interrogazione

$$e = \{A_1: x_1, \ldots, A_k: x_k \mid f(x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+n})\}$$

con $n \geq 0$, $k \geq 1$

La valutazione dell'espressione *e* produce una relazione risultato di schema $(A_1, \ldots, A_k)$ e che contiene tutte le tuple $(\underline{x}_1, \ldots, \underline{x}_k)$ tali che esiste una sostituzione $(\underline{x}_1, \ldots, \underline{x}_k, \underline{x}_{k+1}, \ldots, \underline{x}_{k+n})$ che rende vera f.

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Si consideri la seguente espressione del calcolo relazionale sui domini:

$$\{A_1: x_1, A_2: x_2 \mid R(A_1: x_1) \wedge \exists x_3 (x_2 = x_3)\}$$

Tale espressione, applicata alla base di dati contenente l'istanza *r* di *R*, produce come risultato una relazione sugli attributi (A_1, A_2) contenente le tuple (a_1, a_2) con $a_1 \in r$ e a_2 qualsiasi, vale a dire con $a_2 \in D$

Il risultato di tale espressione quindi DIPENDE da D.

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Ciò vale anche per espressioni del tipo:

$$\{A_1: x_1 \mid \neg R(A_1: x_1)\}$$

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Def.: Un’espressione di un linguaggio di interrogazione è indipendente dal dominio se il suo risultato su ciascuna istanza della base di dati non cambia al variare del dominio.

Un linguaggio di interrogazione è indipendente dal dominio se lo sono tutte le sue espressioni.

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Osservazioni

- Il calcolo relazionale sui domini non è indipendente dal dominio
- L'algebra relazionale è indipendente dal dominio.
- Dalle prime due osservazioni deriva che l'algebra relazionale non è equivalente al calcolo relazionale sui domini
- Si può dimostrare che l'algebra relazionale è equivalente al calcolo relazionale sui domini RISTRETTO alle espressioni indipendenti dal dominio (espressioni SAFE).

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Regole sintattiche per scrivere espressioni SAFE nel calcolo relazionale sui domini

Una formula *f* del calcolo relazionale sui domini è SAFE, vale a dire è indipendente dal dominio, se valgono le seguenti condizioni:

1. in *f* non compaiono quantificatori universali;

   si noti che è sempre possibile convertire un quantificatore universale in un quantificatore esistenziale poiché vale la seguente equivalenza:

$$\forall x(f) \equiv \neg \exists x(\neg f)$$

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Regole sintattiche per scrivere espressioni SAFE nel calcolo relazionale sui domini

2. ogni disgiunzione che compare in f deve avere la seguente forma:

$$f_1(x_1, \ldots, x_m) \vee f_2(x_1, \ldots, x_m)$$

dove $x_1, \ldots, x_m$ sono tutte le variabili libere di f_1 e f_2.

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Regole sintattiche per scrivere espressioni SAFE nel calcolo relazionale sui domini

3. ogni sottoformula massimale costituita dalla congiunzione di due o più formule $f_1 \wedge \ldots \wedge f_n$ deve avere tutte le variabili LIMITATE secondo la seguente definizione.
   VARIABILI LIMITATE:
   - una variabile libera x di f_i è LIMITATA se f_i non è negata e f_i non è né una formula atomica del tipo $x \ \theta \ y$ (y variabile), né del tipo $x \ \theta' \ c$ con $\theta' \neq$ '=' (quindi x in x=c è LIMITATA).
   - Se f_i è del tipo x=y e y è LIMITATA allora anche x è LIMITATA.

Calcolo relazionale sui domini

INDIPENDENZA DAL DOMINIO

Regole sintattiche per scrivere espressioni SAFE nel calcolo relazionale sui domini

4. una formula negata può comparire in f solo in congiunzione con altre formule, dove almeno una non è negata, e dove tutte le variabili sono LIMITATE. .

Calcolo relazionale sulle tuple

CARATTERISTICHE

Nel calcolo relazionale sulle tuple con espressioni di range le variabili sono associate alle tuple e non ai valori del dominio.

Il calcolo relazionale sulle tuple definisce la semantica dell'SQL "semplice" (senza operatori aggregati e senza join esterni).

Calcolo relazionale sulle tuple

SINTASSI

Le espressioni del calcolo relazionale sulle tuple hanno la seguente forma:

$$\{\, T \mid L \mid f \,\}$$

dove *T* è la target list, *L* è la range list e *f* è una formula.

Calcolo relazionale sulle tuple

SINTASSI

<u>Target List</u>

È una lista di elementi separati da virgole $e_1, \ldots, e_n$ dove ogni elemento e_i può essere:

- *Y: x.(Z)* dove *Y* e *Z* sono sequenze di attributi con la stessa cardinalità e *x* è una variabile.
- *x.(Z)* dove *Z* è una sequenza di attributi e *x* è una variabile (equivale a *Z: x.(Z)*).
- *x.** dove *x* è una variabile; in questo caso gli attributi sono quelli della relaziona associata alla variabile *x* nella *range list*.

Calcolo relazionale sulle tuple

SINTASSI

Range List

È una lista di elementi separati da virgole r_1, …, r_m dove ogni elemento r_i è così strutturato:

$$x_j\ (R)$$

dove x_j è una variabile e *R* è il nome di una relazione dello schema della base di dati.

Esiste uno e un sol elemento r_i nella *range list* per ogni variabile libera della formula *f*.

Calcolo relazionale sulle tuple

SINTASSI

Formula

Può essere una formula atomica:

- $x.A\ \theta\ c$ oppure $x_1.A_1\ \theta\ x_2.A_2$ dove x, x_1, x_2 sono variabili, A, A_1, A_2 sono attributi, c è una costante e θ è un operatore di confronto.

oppure una formula:

- se f_1 e f_2 sono formule, allora anche $f_1 \wedge f_2$, $f_1 \vee f_2$, $\neg f_1$ e $\neg f_2$ lo sono.
- se f è una formula, allora anche $\forall x(R)(f)$ e $\exists x(R)(f)$ lo sono.

Calcolo relazionale sulle tuple

SINTASSI

Si noti che la presenza della range list consente di evitare l'introduzione dei predicati per esprimere il fatto che i valori assunti dalle variabili appartengano ad un attributo di una relazione della base di dati.

Calcolo relazionale sulle tuple

SEMANTICA

Non si assegna la semantica del calcolo relazionale sulle tuple in quanto del tutto simile a quella del calcolo relazionale sui domini.

Tuttavia, si noti che le variabili assumono come valori le tuple contenute in una relazione (come dichiarato nella range list o nei quantificatori) quindi non c'è alcuna dipendenza dal dominio di riferimento.

Calcolo relazionale sulle tuple

Osservazioni

- Il calcolo relazionale sulle tuple non è equivalente all'algebra relazionale e non è equivalente al calcolo relazionale sui domini in quanto l'unione di due relazioni non è rappresentabile nel calcolo relazionale sulle tuple.
- E' invece possibile rappresentare nel calcolo relazionale sulle tuple sia l'intersezione che la differenza tra due relazioni.