Algebra relazionale

L'algebra relazionale è utilizzata per l'interrogazione di una base di dati. Specifica il procedimanto per ottenere il risultato.
L'algebra relazionale consiste in un insieme di operazioni chiuse sulle relazioni e si compone di operatori di tipo:

- Unario -> $op(r_1) \rightarrow r_2$;
- Binario -> $op(r_1, r_2) \rightarrow r_3$;

Dal punto di vista funzionale, gli operatori possono essere divisi in tre gruppi:

- Operatori insiemistici;
- Operatori specifici;
- Operatori di giunzione (o Join), che uniscono più tabelle imponendo condizioni.

Dal punto di vista della derivabilità, possono essere classificati in:

- Operatori di base;
- Operatori derivati (ottenuti combinando operatori di base).

Operatori insiemistici

Gli operatori insiemistici applicano alle relazioni le operazioni dell'algebra legate agli insiemi. Questo è possibile poichè le relazioni sono insiemi di tuple omogenee.

Definizione: Due tuple si dicono omogenee se hanno gli stessi attributi.

Gli operatori insieminìstici si possono applicare, quindi, solo a relazioni con lo stesso schema.

Date r_1 e r_2 relazioni di schema $R_1(x)$ e $R_2(X)$, è possibile applicare loro i seguenti operatori insiemistici:

Unione

Operatore binario di base.

$$r_1 \cup r_2 = r_3$$

- Contenuto di r_3 -> $\{t|(t \in r_1)\ OR\ (t \in r_2)\}$;
- Schema di r_3 -> insieme X, non cambia.

Cardianlità: $\begin{cases} |r_1 \in r_2| \leq |r_1| + |r_2| \\ |r_1 \in r_2| \geq MAX(|r_1|, |r_2|) \end{cases}$

Differenza

Operatore unario di base.

$$r_1 - r_2 = r_3$$

- Contenuto di r_3 -> $\{t|(t \in r_1)\ AND\ (t \notin r_2)\}$;
- Schema di r_3 -> insieme X, non cambia.

Cardianlità: $\begin{cases} |r_1 - r_2| \leq |r_1| \\ |r_1 - r_2| \geq 0 \end{cases}$

Intersezione

Operatore binario derivato.

$$r_1 \cap r_2 = r_1 - (r_1 - r_2)$$

Cardianlità: $\begin{cases} |r_1 \cap r_2| \leq min(|r_1|, |r_2|) \\ |r_1 \cap r_2| \geq 0 \end{cases}$

Operatori specifici

Data r relazione di schema $R(X)$ con $X = \{A_1, ..., A_n\}$, gli operatori specifici applicabili sono:

Ridenominazione

Operatore unario di base. Consente la modifica dello schema di una relazione (ovvero permette di modificare il valore dei suoi attributi).
Dato un insieme di attributi $y = \{B_1, ..., B_n\}$, con $|y| = |b|$, allora:

$$\rho_{A_1,...,A_n \to B_1,...,B_n} = \{t | \exists t' \in r \text{ tale che } \forall i \in \{\,...,n\} \text{ allora } t'[A_i] = t[B_i]\}$$

- Contenuto di r_3 -> risultato ;
- Schema di r_3 -> y.

Esempio

Popolazione:

Nome	**Cap**	**Abitanti**
Verona	37100	35000
Vicenza	50100	15000

Città:

Comune	**Cap**	**Popolazione**
Milano	20100	2500000

Non posso fare l'unione, devo prima ridenominare:

$$\rho_{\text{Abitanti} \to \text{Popolazione}}(\text{Popolazione}) \cup (\text{Città})$$

Selezione

Operatore unario di base. Consente di estrarre da una relazione solo le tuple che soddisfano una certa condizione F (taglio in orizzontale).

$$\sigma_F(r) = \begin{cases} \text{schema} & X \\ \text{istanza} & \{t \,|\, \exists t' \in r : \; F(t)(\text{la tupla t rende vera F})\} \end{cases}$$

La condizione F è una formula proposizionale che si ottiene combinando attraverso i connettivi logici $\wedge, \vee, \neg$ formule *atomiche* del tipo:

- $A\theta B$;
- $A\theta c$.

Dove:

- $\theta \in \{=, \neq, >, <, \geq, \leq\}$;
- $A, B \in X$;
- $c \in DOM(A)$ o è compatibile con $DOM(A)$.

Una formula atomica del tipo $A\theta B$ è vera sulla tupla $t \in r$ se vale:

$$t[A] \text{ oppure } t[B]$$

Una formula atomica del tipo $A\theta c$ è vera sulla tupla $t \in r$ se vale:

$$t[A] \text{ oppure } c$$

Cardianlità: $0 \geq |\sigma(r)| \geq |r|$[1]

[1] F è tanto più selettiva quanto più la cardinalità di avvicina a 0.

Proiezione	Operatore unario di base. Consente di eliminare alcuni attributi delle tuple di una relazione. Sia $y = \{A_1, ..., A_n\}$ un sottoinsieme degli attributi di X, allora: $$\Pi_Y(r) = \begin{cases} \text{schema} & Y \\ \text{istanza} & \{t \mid \exists t' \in r : \ t = t'[Y]\} \end{cases}$$ dove $t'[Y]$ è una tupla E su Y tale che $\forall A_i \in Y_i$ vale $E[A_i] = t'[A_i]$. Cardianlità: $\begin{cases} 1 \leq \|Pi_Y(r)\| \leq \|r\| & \text{in generale} \\ \|Pi_Y(r)\| = \|r\| & \text{se } y \text{ è una superchiave per } r \end{cases}$

Operatori di giunzione

Gli operatori di giunzione consentono di unire in un'unica relazione tuple contenute in due relazioni distinte costruendo coppie di tuple che soddisfano una condizione di joint e generando per ogni coppia una tupla nel risultato.

$\bar{r}_1$	A	B	C
	a_1	b_2	c_1
	a_2	b_2	c_2
	a_3	b_1	c_2
	a_4	b_4	c_3

$\bar{r}_2$	C	D
	c_1	d_0
	c_2	d_9

$r_1 \bowtie r_2$	A	B	C	D
	a_1	b_1	c_1	d_0
	a_2	b_2	c_2	d_9
	a_3	b_1	c_2	d_9

Gli operatori di giunzione si dividono in:

Join naturale	Operatore binario di base. La condizione di join è implicita e dipende dallo schema di relazioni coinvolte. Due tuple costituiscono una coppia generata dal join se presentano gli stessi valori (uguaglianza) negli attributi comuni alle due relazioni. Siano r_1 e r_2 due relazioni di schema $R_1(X_1)$ e $R_2(X_2)$, allora: $$r_1 \bowtie r_2 = \begin{cases} \text{schema:} X_1 \cup X_2 \\ \text{istanza: } \{t \mid \exists t_1 \in r_1 \ \wedge \exists t_2 \in r_2 : \quad t_1 = t[X_1] \ \wedge t_2 = t[X_2]\} \end{cases}$$ Nel caso in cui non vi siano attributi in comune, il join naturale è detto prodotto cartesiano e produce tutte le coppie possibili. Cardianlità: - $0 \leq \|r_1 \bowtie r_2\| \leq \|r_1\| \cdot \|r_2\|$ in generale; - $\|0 \leq \|r_1 \bowtie r_2\| \leq \|r_1\|$ se $x_1 \cap x_2$ è superchiave per r_2; - $\|r_1 \bowtie r_2\| = \|r_1\|$ se $x_1 \cap x_2$ è soggetto a un vincolo di integrità referenziale che vincola $x_1 \cap x_2$ su r_1 rispetto a r_2 e $x_1 \cap x_2$ è superchiave per r_2. Proprietà: Date r_1 e r_2 due relazioni di schema $R_1(X_1)$ e $R_2(X_2)$, allora: - Il join naturale $r_1 \bowtie r_2$ si dice completo se $$\forall t_1 \in r_1 \text{ tali che } \exists t \in r_1 \bowtie r_2 \text{ vale che } t[X_1] = t_1$$ $$\wedge$$ $$\forall t_2 \in r_2 \text{ tali che } \exists t \in r_1 \bowtie r_2 \text{ vale che } t[X_2] = t_2$$ In questo caso la cardinalità diventa: $MAX(\|r_1\|, \|r_2\|) \leq \|r_1 \bowtie r_2\| \leq \|r_1\| \cdot \|r_2\|$;

	- Il join naturale è commutativo; - Il join naturale è associativo; - Se le due relazioni hanno lo stesso schema ($X_1 = X_2$) allora $$r_1 \bowtie r_2 = r_1 \cap r_2$$ - Se le due relazioni hanno schemi disgiunti ($X_1 \cap X_2 = \emptyset$) allora $$r_1 \bowtie r_2 = r_1 \text{ x } r_2 \text{ (prodotto cartesiano)}$$
Θ-join	Operatore binario derivato. Produce come risultato il join tra r_1 e r_2 dove la condizione di join è esplicitata come parametro. Per poterlo usare è necessario che le due relazioni abbiano schemi discìgiunti ($X_1 \cap X_2 = \emptyset$). $$r_1 \bowtie_F r_2$$ Un θ-join si dice EQUI-join se la condizione F è una congiunzione di uguaglianza tra attributi di r_1 e r_2.

Equivalenza tra gli operatori di join

Il join naturale tra due relazioni r_1 di schema X_1 e r_2 di schema X_2 dove $X_1 \cap X_2 = \{c_1, \ldots, c_m\}$ equivale alla seguente espressione contenente un Θ-Join :

$$r_1 \bowtie r_2 \equiv \Pi_{X_1 \cap X_2}(r_1 \bowtie_{c'_1=c_1 \wedge \cdots \wedge c'_m=c_m} (\rho_{c_1,c_2,\ldots,c_m \to c'_1,c'_2,\ldots,c'_m}(r_2)))$$