Riassunto Basi di Dati - 9 crediti

Schiabel Alberto

22 novembre 2017

Indice

1 Definizioni

1.1 Modello Relazionale

1.1.1 Definizione chiave e superchiave

Un sottoinsieme K di attributi è superchiave per uno schema di relazione r se, per ogni coppia di tuple distinte, i valori assunti dalle tuple in corrispondenza non sono tutti uguali.
Una chiave è una superchiave minimale, ovvero una superchiave la quale, tolto un qualunque attributo, non è più superchiave. In altre parole, non esiste un'altra superchiave K^1 di r che sia contenuta in K come sottoinsieme proprio.
Ogni chiave è superchiave, ma in generale non vale il viceversa.

1.1.2 Definizione Vincolo di integrità

Si tratta di una proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione. Ogni vincolo può essere visto come un predicato che associa ad ogni istanza il valore *vero* o $falso$. Se e solo se il predicato assume il valore *vero*, allora si dice che l'istanza soddisfa il vincolo. Lo scopo di questi vincoli è di vietare una situazione indesiderata all'interno della base di dati.

1.1.3 Definizione Vincolo intrarelazionale

Vincolo il cui soddisfacimento è definito rispetto a singole relazioni della base di dati. Talvolta, il vincolo riguarda le tuple separatamente le une dalle altre.

1.1.4 Definizione Vincolo interrelazionale

Vincolo che coinvolge più relazioni. Permette di vietare una situazione indesiderata ad esempio richiedendo che un numero di matricola compaia nella relazione ESAMI solo se compare anche nella relazione STUDENTI.

1.1.5 Definizione Vincolo di integrità referenziale

Un vincolo di integrità referenziale tra un insieme di attributi X di due relazioni R_1 e R_2 è soddisfatto se i valori su X di ciascuna tupla dell'istanza di R_1 compaiono come valori della chiave primaria dell'istanza di R_2.

1.2 Algebra e Calcolo Relazionale

1.2.1 Definizione di Selezione

Operatore monadico il cui risultato ha gli stessi attributi dell'operando, e contiene solo le ennuple dell'operando che soddisfano la condizione specificata. La condizione deve essere un'espressione booleana.
Si dice che la selezione effettua una decomposizione orizzontale.
Simbolo: $\sigma_F(R)$

1.2.2 Definizione di Proiezione

Operatore monadico il cui risultato ha solo la parte specificata Y degli attributi X dell'operando, ma contiene tutte le ennuple dell'operando considerate solo sui valori di Y.
Si dice che la proiezione effettua una decomposizione verticale.
Simbolo: $\Pi_{Y_1,\ldots,Y^N}(R)$

1.2.3 Cardinalità Proiezione

Definedo N la cardinalità della proiezione:

1. se non esistono duplicati nelle ennuple dlel'operando r_1 allora $N = |r_1|$
2. altrimenti, a causa dell'eliminazione dei duplicati si ottiene $0 \leq N < |r_1|$

1.2.4 Definizione di Join

Operatore che permette di correlare dati contenuti in relazioni diverse, confrontando i valori contenuti in esse. Esiste in più varianti.

1.2.5 Definizione di Natural Join

Operatore binario che correla dati in relazioni diverse, sulla base di valori uguali, in attibuti con lo stesso nome.
Simbolo: $\bowtie$
Proprietà:

1. Commutatività: $r1 \bowtie r2 = r2 \bowtie r1$
2. Associatività $r1 \bowtie (r2 \bowtie r3) = (r1 \bowtie r2) \bowtie r3$
3. Se gli insiemi X_1 e X_2 di attributi di due tuple sono uguali, allora i Natural Join coincide con un'intersezione.
4. Se gli insiemi X_1 e X_2 di attributi di due tuple sono disgiunti, allora i Natural Join coincide con il prodotto cartesiano.

1.2.6 Definizone di Theta Join

Operatore definito come il prodotto cartesiano seguito da una selezione, nel modo seguente (dove F è una formula proposizionale utilizzabile in una selezione, e dove le relazioni r_1 e r_2 non hanno attributi in comune):

$$r_1 \bowtie_F r_2 = \sigma_F(r_1 \bowtie r_2)$$

1.2.7 Definizione di Equi Join

L'Equi Join non è altro che un Theta Join in cui la condizione di selezione F sia una congiunzione di uguaglianza, con un attributo della prima relazione $r1$ e uno della seconda $r2$.

1.3 Cardinalità Join

Definedo N la cardinalità del join:

1. se il join di r_1 e r_2 è completo allora $0 \leq N \leq max(|r_1|, |r_2|)$
2. se il join coinvolge una chiave di r_2, allora $0 \leq N \leq |r_1|$
3. se il join coinvolge una chiave di r_2 ed $\exists$ un vincolo di integrità referenziale tra un attributo di r_1 e la chiave di r_2, allora $N = |r_1|$

1.4 Progettazione Concettuale

1.4.1 Definizione strategia di progetto top-down

Nella strategia top-down, lo schema concettuale viene prodotto mediante raffinamenti successivi a partire da uno schema iniziale che, pur descrivendo tutte le specifiche, resta astratto. Tale schema viene a via a via raffinato aumentando il livello di dettagli, ma mantiene le medesime informazioni. Tutti gli aspetti presenti nello schema finale sono presenti a ogni livello di raffinamento.
PRO: il progettista può inizialmente descrivere tutte le specifiche dei dati trascurandone i dettagli
CONTRO: è necessario possedere sin dall'inizio una visione globale di tutte le componenti del sistema

1.4.2 Definizione strategia di progetto bottom-up

Nella strategia bottom-up si suddividono le specifiche in modo da sviluppare diversi schemi elementari ma dettagliati, che successivamente vengono integrati tra di loro. Tale strategia favorisce lo sviluppo in team.

1.4.3 Definizione strategia di progetto inside-out

Si tratta di un caso particolare della bottom-up. Inizialmente si individuano solo alcuni concetti importanti e poi si procede, a partire da questi, a "macchi d'olio"; si rappresentano cioè prima i concetti in relazione con quelli iniziali, per poi muoversi verso quelli più lontani, in una sorta di "navigazione" tra le specifiche.

2 Algebra Relazionale

2.1 Principio di complementarietà

Data un'interrogazione q da realizzare in algebra relazionale, spesso la si può scomporre in sottointerrogazioni che possono corrispondere a diversi idiomi, frequentemente richiesti in sede d'esame. Molti di questi si basano sul **Principio di Complementarietà**:

$$\sigma_p(R) \equiv R - \sigma_{\neg p}(R)$$

La selezione fatta su una relazione R con un predicato p, è uguale alla relazione R stessa meno la selezione su R con predicato p negato.

Studente	**Corso**	**Data**	**Voto**
Luca	Basi di Dati	23/11/2017	30
Anna	Logica	22/11/2017	22
Marco	Programmazione	20/09/2017	18
Anna	Programmazione	21/10/2017	30

Per esempio nella relazione ESAMI, l' insieme degli studenti che ha preso 30 è dato dall' insieme di tutti gli studenti che hanno dato almeno un esame (quelli nella tabella **ESAMI**) meno tutti gli studenti che non hanno preso 30. Gli idiomi di interrogazione principali sono:

1. Minimo e Massimo (assoluti o relativi)
2. Cardinalità
3. Per ogni
4. Inclusione

2.2 Idiomi frequenti di interrogazione

2.2.1 Minimo e Massimo assoluto

Dato lo schema relazionale $R(A, B)$, trovare il minimo/massimo in R. Si supponga di voler determinare il minimo B:

$$\Pi_B(R) - \Pi_B(R \bowtie_{B>B^1} (\rho A^1, B^1 \leftarrow A, B(R)))$$

Nella seconda parte vengono trovati tutti quei valori che non sono il minimo. Per fare ciò si deve joinare la relazione R con un'altra istanza di se stessa, con gli attributi ridenominati. La condizione del theta join indica che ogni attributo B deve essere maggiore degli stessi attributi ridenominati. In tal modo vengono mantenute tutte le tuple tranne quella in cui l'attributo B assume il valore minore.

Per il Principio di Complementarietà, sottraendo dall'insieme iniziale l'insieme delle tuple dove B non è il minimo, si ottiene proprio il valore minimo cercato.

Esempio con la relazione **ESAMI** e l'attributo Voto:

$$S1 := \rho_{Studente^1,Corso^1,Data^1,Voto^1 \leftarrow Studente,Corso,Data,Voto}(\mathbf{ESAMI})$$
$$\Pi_{Voto}(\mathbf{ESAMI}) - \Pi_{Voto}(\mathbf{ESAMI} \bowtie_{Voto>Voto^1} (S1))$$

2.2.2 Minimo e Massimo relativo

Dato lo schema relazionale $R(A, B)$, trovare per ogni A il minimo/massimo in R. Si supponga di voler determinare il massimo B in A:

$$\Pi_{A,B}(R) - \Pi_{A,B}(R \bowtie_{A=A^1 \wedge B<B^1} (\rho A^1, B^1 \leftarrow A, B(R)))$$

È molto simile al massimo assoluto. Il theta join in questo caso seleziona tutti i valori minimi di B per ogni attributo A.
Esempio con la relazione **ESAMI** in cui A sia Studente e B sia Voto:

$$S1 := \rho_{Studente^1,Voto^1 \leftarrow Studente,Voto}(\mathbf{ESAMI})$$
$$\Pi_{Nome,Voto}(\mathbf{ESAMI}) - \Pi_{Nome,Voto}(\mathbf{ESAMI} \bowtie_{Studente=Studente^1 \wedge Voto<Voto^1} (S1))$$

2.2.3 Cardinalità

Dato lo schema relazionale $R(A, B)$, trovare gli A che sono associati ad almeno 2 B:

$$\Pi_A(R \bowtie_{A=A^1 \wedge B\neq B^1} (\rho A^1, B^1 \leftarrow A, B(R)))$$

Viene ancora fatto un theta join tra la relazione R e se stessa con gli attributi ridenominati. Il predicato del join consente di mantenere tutte quelle tuple in cui l'attributo A è uguale e B è diverso. Queste tuple sono proprio tutte le tuple di B associate almeno 2 volte ad ogni elemento di A.

Dato lo schema relazionale $R(A, B)$, trovare gli A che sono associati ad almeno 3 B:

$$S1 := \rho_{A^1,B^1 \leftarrow A,B}(R)$$
$$S2 := \rho_{A^2,B^2 \leftarrow A,B}(R)$$
$$S3 := R \times S1 \times S2 \Pi_A(\sigma_{A=A^1 \wedge A=A^2 \wedge B\neq B^1 \wedge B\neq B^2 \wedge B^1\neq B^2}(S3))$$

Viene fatto il prodotto cartesiano della relazione e dei suoi due duplicati con gli attributi rinominati. Dalla relazione che otteniamo cosi facendo, vengono selezionate le tuple che soddisfano il predicato di selezione, ovvero tutte le tuple in cui A è associato ad almeno 3 B. I generale per trovare gli A che sono associati ad almeno n B bisogna fare:

$(n-1)$	prodotti cartesiani
$(n-1)$	condizioni della forma $A = A^i$
$\frac{n(n-1)}{2}$	condizioni della forma $B^i \neq B^j$

2.2.4 Per Ogni

Dato lo schema relazionale $R(A, B, C)$, trovare gli A per i quali tutti i C sono positivi.

$$\Pi_A(R) - \Pi_A(\sigma_{C \leq 0}(R))$$

A tutti gli A si vogliono togliere quegli A per cui C è negativo, ottenendo così gli A per i quali tutti i C sono positivi.

Dato lo schema relazionale $R(A, B, C)$, trovare gli A per i quali tutti i C sono uguali.

$$\Pi_A(R) - \Pi_A(R \bowtie_{A=A^1 \wedge C \neq C^1} (\rho_{A^1,B^1,C^1 \Leftarrow A,B,C}(R)))$$

Il theta join permette di trovare tutti gli elementi di A che non hanno tutti i C uguali. Sottraendo ciò che troviamo dal join con l'insieme di tutti gli elementi, otteniamo gli A per i quali tutti i C sono uguali.

2.2.5 Inclusione

Dati gli schemi relazionali $R(A, B)$ e $S(B)$, trovare gli A per i quali l'insieme dei B associati include tutti gli elementi di S.

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

Esempio:

ESAME(Studente, CCorso, Voto) **Corso**(CCorso, Docente)

Determinare gli studenti che hanno passato tutti gli esami:

$$S1 := \Pi_{Studente,CCorso}(\mathbf{STUDENTE})$$
$$S2 := \Pi_{CCorso}(\mathbf{CORSO})$$
$$\Pi_{Studente}(S1) - \Pi_{Studente}((\Pi_{Studente}(S1) \times S2) - S1)$$

2.3 Esercizio Venditori, Prodotto, Listino

2.3.1 Quesito

È dato uno schema di basi di dati costituito dalle relazioni:

VENDITORE(<u>vid</u>, vnome, indirizzo)
PRODOTTO(<u>pid</u>, pnome, colore, peso)
LISTINO(<u>vid</u>, <u>pid</u>, prezzo)

Esistono dei vincoli di integrità referenziale tra vid di VENDITORE e vid di LISTINO, e tra pid di PRODOTTO e pid di LISTINO.

Formulare in algebra relazionale le seguenti interrogazioni:

1. Trovare i nomi dei venditori che forniscono prodotti rossi o prodotti verdi

2. Trovare i nomi dei venditori che hanno a listino almeno due prodotti rossi

3. Trovare l'id dei venditori che hanno a listino solo prodotti verdi

4. Trovare l'id dei prodotti a listino più pesanti

2.3.2 Soluzioni

1. $\Pi_{vnome}(VENDITORE \bowtie LISTINO \bowtie \sigma_{colore="rosso" \vee color="verde"}(PRODOTTO))$

2. S1 := $\sigma_{color="rosso"}(PRODOTTO)$
   S2 := $\Pi_{pid,vid}(LISTINO \bowtie S1)$
   S3 := $\rho_{pid1,vid1 \leftarrow pid,vid}(S2)$
   $\Pi_{vnome}(\sigma_{pid \neq pid1 \wedge vid=vid1}(S2 \bowtie S3) \bowtie PRODOTTO$

3. S1 := $LISTINO \bowtie \sigma_{colore \neq "verde"}(PRODOTTO))$
   $\Pi_{pid}(LISTINO - S1)$

4. S1 := $\rho_{pid1,pnome1,colore1,peso1 \leftarrow pid,nome,colore,peso}(PRODOTTO)$
   S2 := $\sigma_{peso > peso1}(PRODOTTO \bowtie S1)$ $\Pi_{pid}(S2)$