

BD 2 - Anomalie & Dipendenze Funzionali

Luca Cosmo

Università Ca' Foscari Venezia



Università
Ca' Foscari
Venezia

La Teoria della Normalizzazione

Come qualsiasi sistema di modellazione, anche il modello relazionale permette di rappresentare la stessa informazione in più modi differenti.

Problemi Fondamentali

- 1 Quando una rappresentazione è di buona qualità?
- 2 Quando due diverse rappresentazioni sono equivalenti?

La teoria della **normalizzazione** si occupa di formalizzare questi concetti e dare una risposta rigorosa a tali domande.

Qualità di Schemi Relazionali

NomeUtente	Telefono	<u>CodiceLibro</u>	Titolo	Data
Rossi Carlo	75444	XY188A	Decameron	07-07
Pastine Maurizio	66133	XY090C	Canzoniere	01-08
Paolicchi Laura	59729	XY101A	Vita Nova	05-08
Paolicchi Luca	59729	XY701B	Adelchi	14-01
Paolicchi Luca	59729	XY008C	Amleto	17-08

Problemi Principali

- 1 Ridondanza dei dati:** replicazione di informazione per gli utenti che hanno più di un libro in prestito.
- 2 Scarsa espressività:** rappresentabili solo utenti che hanno almeno un libro in prestito.

Anomalie di Aggiornamento

La ridondanza dei dati è un problema perché può portare ad anomalie in presenza di operazioni di **aggiornamento**. Più precisamente si potrebbe effettuare per errore un aggiornamento solo parziale dei dati.

Anomalia di Aggiornamento

NomeUtente	Telefono	<u>CodiceLibro</u>	Titolo	Data
Paolicchi Luca	59729	XY701B	Adelchi	14-01
Paolicchi Luca	64135	XY008C	Amleto	17-08

Anomalie di Inserimento e Cancellazione

Una cattiva progettazione può rendere impossibile rappresentare certe informazioni, conducendo ad anomalie in presenza di operazioni di **inserimento** e **cancellazione**.

Anomalia di Inserimento

NomeUtente	Telefono	<u>CodiceLibro</u>	Titolo	Data
Gianni Meucci	78425	???	???	???

Anomalia di Cancellazione

NomeUtente	Telefono	<u>CodiceLibro</u>	Titolo	Data
Rossi Carlo	75444	XY188A	Decameron	07-07

Modellazione Alternativa

<u>NomeUtente</u>	Telefono
Rossi Carlo	75444
Pastine Maurizio	66133
Paolicchi Laura	59729
Paolicchi Luca	59729

<u>CodiceLibro</u>	Titolo	Data	Nome Utente
XY188A	Decameron	07-07	Rossi Carlo
XY090C	Canzoniere	01-08	Pastine Maurizio
XY101A	Vita Nova	05-08	Paolicchi Laura
XY701B	Adelchi	14-01	Paolicchi Luca
XY008C	Amleto	17-08	Paolicchi Luca

Modellazione Errata!

<u>NomeUtente</u>	Telefono
Rossi Carlo	75444
Pastine Maurizio	66133
Paolicchi Laura	59729
Paolicchi Luca	59729

<u>CodiceLibro</u>	Titolo	Data	Telefono
XY188A	Decameron	07-07	75444
XY090C	Canzoniere	01-08	66133
XY101A	Vita Nova	05-08	59729
XY701B	Adelchi	14-01	59729
XY008C	Amleto	17-08	59729

Equivalenza di Schemi Relazionali

Chi ha preso in prestito l'Adelchi? Laura o Luca?

<u>NomeUtente</u>	Telefono
Paolicchi Laura	59729
Paolicchi Luca	59729

<u>CodiceLibro</u>	Titolo	Data	Telefono
XY101A	Vita Nova	05-08	59729
XY701B	Adelchi	14-01	59729
XY008C	Amleto	17-08	59729

Il nuovo schema ha eliminato le anomalie, ma ha condotto ad una **perdita di informazione**, quindi non è equivalente allo schema di partenza!

La Teoria della Normalizzazione

Fornisce un insieme di strumenti/algoritmi per stabilire in modo rigoroso la bontà di uno schema relazionale, e per migliorare (normalizzazione) schemi relazionali esistenti.

In queste slide formalizzeremo i seguenti concetti:

- **Dipendenze Funzionali:** stabiliscono le dipendenze tra gli attributi della relazione. Per esempio, l'attributo "Titolo" dipende da "CodiceLibro", ovvero, ogni volta che "CodiceLibro" è uguale lo è anche "Titolo".
- **Assiomi di Armstrong:** una serie di *regole* per derivare altre dipendenze funzionali.
- **Implicazione:** spesso vogliamo verificare se una certa dipendenza funzionale può essere derivata in una relazione.
- **Chiusura** di un insieme di attributi: l'insieme di tutti gli attributi derivabile da un insieme di uno o più attributi (per esempio le chiavi) secondo le regole di derivazione.

Notazione

In queste lezioni useremo la seguente notazione:

- Lettere maiuscole all'inizio dell'alfabeto, es. A, B, C : attributi
- Lettere maiuscole alla fine dell'alfabeto, es. T, X, Y : insiemi di attributi
- $R(T)$: schema di relazione costruito sull'insieme di attributi T
- r : istanza dello schema di relazione $R(T)$, cioè un insieme di ennuple
- t, u, v : ennuple di un'istanza di un dato schema di relazione
- $t[X]$: ennupla ottenuta da t considerando i soli attributi in X

Dipendenze Funzionali

Definition (Dipendenza Funzionale)

Sia $R(T)$ uno schema di relazione e siano X, Y due insiemi di attributi non vuoti tali che $X \cup Y \subseteq T$, una **dipendenza funzionale** è un qualsiasi vincolo della forma $X \rightarrow Y$.

Definition (Soddisfacibilità)

Un'istanza r di $R(T)$ **soddisfa** la dipendenza funzionale $X \rightarrow Y$ se e solo se ogni coppia di ennuple in r che coincide su X coincide anche su Y . Formalmente chiediamo $\forall t_1, t_2 \in r : t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$.

Scriveremo $R(T, F)$ per indicare uno schema di relazione $R(T)$ con le dipendenze funzionali F che tutte le sue istanze **valide** devono soddisfare.

Esempio

NomeUtente	Telefono	<u>CodiceLibro</u>	Titolo	Data
Rossi Carlo	75444	XY188A	Decameron	07-07
Pastine Maurizio	66133	XY090C	Canzoniere	01-08
Paolicchi Laura	59729	XY101A	Vita Nova	05-08
Paolicchi Luca	59729	XY701B	Adelchi	14-01
Paolicchi Luca	59729	XY008C	Amleto	17-08

Esempi di dipendenze funzionali soddisfatte o meno dall'istanza:

- ✓ NomeUtente → Telefono
- ✓ CodiceLibro → Titolo, NomeUtente, Data
- ✗ Telefono → NomeUtente

Ha senso che lo schema includa le prime due dipendenze, ma non la terza!

Dipendenze Derivate

Definition (Implicazione)

Sia $R(T, F)$ uno schema di relazione. Diciamo che F **implica logicamente** la dipendenza funzionale $X \rightarrow Y$, indicato con $F \models X \rightarrow Y$, se e solo se **ogni** istanza valida di $R(T, F)$ soddisfa anche $X \rightarrow Y$.

Example

Sia $F = \{\text{CodiceLibro} \rightarrow \text{Titolo}, \text{CodiceLibro} \rightarrow \text{NomeUtente}\}$, abbiamo:

- $F \models \text{CodiceLibro} \rightarrow \text{Titolo}, \text{NomeUtente}$
- $F \models \text{CodiceLibro} \rightarrow \text{CodiceLibro}$

Assiomi di Armstrong

Come possiamo dimostrare che $F \models X \rightarrow Y$? E' una proprietà difficile da dimostrare, perché quantifica su **tutte** le possibili istanze valide!

Assiomi di Armstrong

Regole di inferenza della forma $F \vdash X \rightarrow Y$:

$$\begin{array}{c} \text{REFL} \\ Y \subseteq X \\ \hline F \vdash X \rightarrow Y \end{array}$$

$$\begin{array}{c} \text{AUGM} \\ F \vdash X \rightarrow Y \\ \hline F \vdash XW \rightarrow YW \end{array}$$

$$\begin{array}{c} \text{TRANS} \\ F \vdash X \rightarrow Y \quad F \vdash Y \rightarrow Z \\ \hline F \vdash X \rightarrow Z \end{array}$$

Gli assiomi di Armstrong sono *corretti* e *completi* (\vdash vale \models)

Theorem

$F \vdash X \rightarrow Y$ se e solo se $F \models X \rightarrow Y$.

Derivazioni

Come possiamo dimostrare che $F \vdash X \rightarrow Y$?

Definition

Una **derivazione** di $X \rightarrow Y$ da F è una sequenza finita di dipendenze funzionali f_1, \dots, f_n tale che $f_n = X \rightarrow Y$ ed ogni f_i è un elemento di F oppure può essere ottenuta da f_1, \dots, f_{i-1} usando una regola di inferenza.

Attenzione!

E' comodo rappresentare una derivazione di $F \vdash X \rightarrow Y$ come un **albero** rovesciato la cui radice è $X \rightarrow Y$ ed ogni altro nodo è giustificato da F oppure da una regola di inferenza.

Regola Derivata: Unione

Unione

Se $X \rightarrow Y$ e $X \rightarrow Z$, allora $X \rightarrow YZ$.

Derivazione: $X \rightarrow Y, X \rightarrow XY, X \rightarrow Z, XY \rightarrow YZ, X \rightarrow YZ$.

Come albero:

$$\begin{array}{c} \text{AUGM} \frac{X \rightarrow Y, X \rightarrow Z \vdash X \rightarrow Y}{X \rightarrow Y, X \rightarrow Z \vdash X \rightarrow XY} \quad \frac{X \rightarrow Y, X \rightarrow Z \vdash X \rightarrow Z}{X \rightarrow Y, X \rightarrow Z \vdash XY \rightarrow YZ} \text{AUGM} \\ \text{TRANS} \frac{\quad}{X \rightarrow Y, X \rightarrow Z \vdash X \rightarrow YZ} \end{array}$$

Regola Derivata: Decomposizione

Decomposizione

Se $X \rightarrow YZ$, allora $X \rightarrow Y$.

Derivazione: $X \rightarrow YZ, YZ \rightarrow Y, X \rightarrow Y$.

Come albero:

$$\text{TRANS} \frac{X \rightarrow YZ \vdash X \rightarrow YZ \quad \frac{Y \subseteq YZ}{X \rightarrow YZ \vdash YZ \rightarrow Y} \text{REFL}}{X \rightarrow YZ \vdash X \rightarrow Y}$$

Regola Derivata: Indebolimento

Indebolimento

Se $X \rightarrow Y$, allora $XZ \rightarrow Y$.

Derivazione: $X \rightarrow Y, XZ \rightarrow YZ, YZ \rightarrow Y, XZ \rightarrow Y$.

Come albero:

$$\text{TRANS} \frac{\text{AUGM} \frac{X \rightarrow Y \vdash X \rightarrow Y}{X \rightarrow Y \vdash XZ \rightarrow YZ} \quad \frac{Y \subseteq YZ}{X \rightarrow Y \vdash YZ \rightarrow Y} \text{REFL}}{X \rightarrow Y \vdash XZ \rightarrow Y}$$

Problema dell'Implicazione

Le dipendenze funzionali che ci interessano veramente sono tutte quelle che possiamo derivare (e quindi implicare logicamente!)

Definition (Chiusura di F)

Dato un insieme F di dipendenze funzionali, la **chiusura** di F è definita come l'insieme $F^+ = \{X \rightarrow Y \mid F \vdash X \rightarrow Y\}$.

Definition (Problema dell'Implicazione)

Il **problema dell'implicazione** corrisponde a decidere, dati F e $X \rightarrow Y$, se $X \rightarrow Y \in F^+$ oppure no.

Si noti che calcolare F^+ applicando gli assiomi di Armstrong ha costo **esponenziale** nel numero di attributi. Calcolare F^+ è quindi un modo alitmicamente **inefficiente** per risolvere il problema dell'implicazione.

Problema dell'Implicazione

Definition (Chiusura di X)

Sia $R(T, F)$ uno schema di relazione. Dato $X \subseteq T$, la **chiusura** di X rispetto ad F è definita come l'insieme $X_F^+ = \{A \in T \mid F \vdash X \rightarrow A\}$.

Theorem

$F \vdash X \rightarrow Y$ se e solo se $Y \subseteq X_F^+$.

Pertanto per decidere se $X \rightarrow Y \in F^+$ si può controllare se $Y \subseteq X_F^+$, che vedremo essere calcolabile in tempo **polinomiale**. Questo ci consente di risolvere il problema dell'implicazione in modo efficiente.

Dimostrazione del Teorema

Theorem

$F \vdash X \rightarrow Y$ se e solo se $Y \subseteq X_F^+$.

Proof.

- (\Rightarrow) Da $F \vdash X \rightarrow Y$ abbiamo $\forall A \in Y : F \vdash X \rightarrow A$ per la regola di decomposizione. La conclusione deriva dalla definizione di X_F^+ .
- (\Leftarrow) Da $Y \subseteq X_F^+$ abbiamo $\forall A \in Y : F \vdash X \rightarrow A$ per definizione di X_F^+ . La conclusione deriva dalla regola di unione.



Calcolo di X_F^+

Per calcolare X_F^+ possiamo usare il seguente algoritmo:

function CLOSURE(X, F)

$X_{old}^+ \leftarrow \emptyset$

$X_{new}^+ \leftarrow X$

while $X_{new}^+ \neq X_{old}^+$ **do**

$X_{old}^+ \leftarrow X_{new}^+$

for all $Y \rightarrow Z \in F$ **do**

if $Y \subseteq X_{new}^+$ **then**

$X_{new}^+ \leftarrow X_{new}^+ \cup Z$

return X_{new}^+

Vediamo un esempio con $X = AB$ e
 $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow F\}$:

Calcolo di X_F^+

Per calcolare X_F^+ possiamo usare il seguente algoritmo:

```
function CLOSURE( $X, F$ )  
   $X_{old}^+ \leftarrow \emptyset$   
   $X_{new}^+ \leftarrow X$   
  while  $X_{new}^+ \neq X_{old}^+$  do  
     $X_{old}^+ \leftarrow X_{new}^+$   
    for all  $Y \rightarrow Z \in F$  do  
      if  $Y \subseteq X_{new}^+$  then  
         $X_{new}^+ \leftarrow X_{new}^+ \cup Z$   
  return  $X_{new}^+$ 
```

Vediamo un esempio con $X = AB$ e
 $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow F\}$:

$$\boxed{1} \ X_0^+ = AB$$

Calcolo di X_F^+

Per calcolare X_F^+ possiamo usare il seguente algoritmo:

```
function CLOSURE( $X, F$ )  
   $X_{old}^+ \leftarrow \emptyset$   
   $X_{new}^+ \leftarrow X$   
  while  $X_{new}^+ \neq X_{old}^+$  do  
     $X_{old}^+ \leftarrow X_{new}^+$   
    for all  $Y \rightarrow Z \in F$  do  
      if  $Y \subseteq X_{new}^+$  then  
         $X_{new}^+ \leftarrow X_{new}^+ \cup Z$   
  return  $X_{new}^+$ 
```

Vediamo un esempio con $X = AB$ e
 $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow F\}$:

1 $X_0^+ = AB$

2 $X_1^+ = AB \cup C = ABC$

Calcolo di X_F^+

Per calcolare X_F^+ possiamo usare il seguente algoritmo:

```
function CLOSURE( $X, F$ )  
   $X_{old}^+ \leftarrow \emptyset$   
   $X_{new}^+ \leftarrow X$   
  while  $X_{new}^+ \neq X_{old}^+$  do  
     $X_{old}^+ \leftarrow X_{new}^+$   
    for all  $Y \rightarrow Z \in F$  do  
      if  $Y \subseteq X_{new}^+$  then  
         $X_{new}^+ \leftarrow X_{new}^+ \cup Z$   
  return  $X_{new}^+$ 
```

Vediamo un esempio con $X = AB$ e
 $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow F\}$:

- 1 $X_0^+ = AB$
- 2 $X_1^+ = AB \cup C = ABC$
- 3 $X_2^+ = ABC \cup D = ABCD$

Calcolo di X_F^+

Per calcolare X_F^+ possiamo usare il seguente algoritmo:

```
function CLOSURE( $X, F$ )  
   $X_{old}^+ \leftarrow \emptyset$   
   $X_{new}^+ \leftarrow X$   
  while  $X_{new}^+ \neq X_{old}^+$  do  
     $X_{old}^+ \leftarrow X_{new}^+$   
    for all  $Y \rightarrow Z \in F$  do  
      if  $Y \subseteq X_{new}^+$  then  
         $X_{new}^+ \leftarrow X_{new}^+ \cup Z$   
  return  $X_{new}^+$ 
```

Vediamo un esempio con $X = AB$ e
 $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow F\}$:

- 1 $X_0^+ = AB$
- 2 $X_1^+ = AB \cup C = ABC$
- 3 $X_2^+ = ABC \cup D = ABCD$
- 4 Da qui in poi non è più possibile aggiungere altri attributi, quindi l'algoritmo termina

Complessità dell'Algoritmo

```
function CLOSURE( $X, F$ )  
   $X_{old}^+ \leftarrow \emptyset$   
   $X_{new}^+ \leftarrow X$   
  while  $X_{new}^+ \neq X_{old}^+$  do  
     $X_{old}^+ \leftarrow X_{new}^+$   
    for all  $Y \rightarrow Z \in F$  do  
      if  $Y \subseteq X_{new}^+$  then  
         $X_{new}^+ \leftarrow X_{new}^+ \cup Z$   
  return  $X_{new}^+$ 
```

Siano a il numero totale di attributi di T e d il numero di dipendenze funzionali in F :

- 1 Il while esterno viene eseguito al più a volte
- 2 Il for interno viene eseguito al più d volte
- 3 Verificare l'inclusione di insiemi ordinati di cardinalità al più a nel for interno ha costo $O(a)$

Complessità algoritmica: $O(a^2 d)$

Checkpoint

Punti Chiave

- Il problema delle anomalie ed i tre tipi principali di anomalia
- Il concetto di dipendenza funzionale e di dipendenza derivata
- La chiusura di F ed il costo esponenziale della costruzione di F^+
- Il problema dell'implicazione è risolvibile in tempo polinomiale

Materiale Didattico

Fondamenti di Basi di Dati: Capitolo 5, fino alla Sezione 5.2.3 compresa.