

Lezione 7 – Progettazione di una base di dati relazionale – Dipendenze funzionali

Prof.ssa Maria De Marsico
demarsico@di.uniroma1.it



SAPIENZA
UNIVERSITÀ DI ROMA

Definizione

- Uno **schema di relazione** R è un insieme di attributi $\{A_1, A_2, \dots, A_n\}$
- *Notazione*
 - $R = A_1 A_2 \dots A_n$
 - le prime lettere dell'alfabeto (A,B,C,...) denotano singoli attributi
 - le ultime lettere dell'alfabeto (X,Y,...) denotano insiemi di attributi
 - Se X ed Y sono insiemi di attributi XY denota $X \cup Y$

Definizione

- Dato uno schema di relazione $R = A_1 A_2 \dots A_n$,
- una **tupla** t su R è una **funzione** che associa ad ogni attributo A_i in R un valore $t[A_i]$ nel corrispondente dominio $\text{dom}(A_i)$.

	NomeStud	CognomeStud	Es. sost.	Media
t_1	Paolo	Rossi	2	26,5
t_2	Mario	Bianchi	10	28,7

$t_1[\text{NomeStud}] = \text{Paolo}$

$t_1[\text{CognomeStud}] = \text{Rossi}$

$t_1[\text{Es. sost}] = 2$

$t_1[\text{MEDIA}] = 26,5$

$t_2[\text{NomeStud}] = \text{Mario}$

$t_2[\text{CognomeStud}] = \text{Bianchi}$

$t_2[\text{Es. sost}] = 10$

$t_2[\text{MEDIA}] = 28,7$

Se X è un sottoinsieme di R e t_1 e t_2 sono due tuple su R

t_1 e t_2 **coincidono su** X ($t_1[X]=t_2[X]$)
se $\forall A \in X (t_1[A]=t_2[A])$

	NomeStud	CognomeStud	Es. sost.	Media
t_1	Paolo	Rossi	3	27
t_2	Mario	Rossi	5	27

$t_1[\text{CognomeStud Media}] = t_2[\text{CognomeStud Media}]$

$t_1[\text{CognomeStud NomeStud}] \neq t_2[\text{CognomeStud NomeStud}]$



Definizione

- Dato uno schema di relazione R ,
- una **istanza** di R è un insieme di tuple su R
- **TUTTE** le “tabelle” che abbiamo visto finora negli esempi sono istanze di qualche schema di relazione

Definizione

- Dato uno schema di relazione R ,

una **dipendenza funzionale** su R è una coppia ordinata di sottoinsiemi non vuoti X ed Y di R

- *Notazione e terminologia*
- $X \rightarrow Y$
- X **determina funzionalmente** Y oppure
- Y **dipende funzionalmente** da X
- X = parte sinistra della dipendenza o determinante
- Y = parte destra della dipendenza o dipendente

Definizione

- Dati uno schema R e una dipendenza funzionale $X \rightarrow Y$ su R ,

un'istanza r di R **soddisfa** la dipendenza funzionale $X \rightarrow Y$ se:

$$\forall t_1, t_2 \in r (t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y])$$

- **Nota:** ovviamente se $t_1[X] \neq t_2[X]$ la dipendenza è soddisfatta comunque siano i valori di $t_1[Y]$ e $t_2[Y]$

Nota:

implica



$$\forall t_1, t_2 \in r (t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y])$$

Significa che

$$\forall t_1, t_2 \in r, \text{ se } (t_1[X] = t_2[X]) \text{ allora } t_1[Y] = t_2[Y]$$



- **NOTA:** le dipendenze funzionali non fanno altro che esprimere dei **vincoli** di sui dati.
- Es. Se due tuple hanno uguale il codice fiscale (si riferiscono alla stessa persona) devono avere uguale anche la data di nascita!

- Nella relazione che rappresenta gli esami, **non abbiamo** $\text{Voto} \rightarrow \text{Lode}$ perché se $t_1[\text{Voto}] = t_2[\text{Voto}] = 27$ allora sicuramente deve essere $t_1[\text{Lode}] = t_2[\text{Lode}] = \text{'No'}$ ma ...
... se $t_1[\text{Voto}] = t_2[\text{Voto}] = 30$ e $t_1[\text{Lode}] = \text{'Si'}$ questo non determina il valore che deve avere $t_2[\text{Lode}]$ (può essere 'Si' oppure 'No' senza compromettere la correttezza del dato)
- E' possibile dire che $\text{Lode} \rightarrow \text{Voto}$?

- Nella relazione che rappresenta gli esami, **non abbiamo** $\text{Voto} \rightarrow \text{Lode}$ perché se $t_1[\text{Voto}] = t_2[\text{Voto}] = 27$ allora sicuramente deve essere $t_1[\text{Lode}] = t_2[\text{Lode}] = \text{'No'}$ ma ...
... se $t_1[\text{Voto}] = t_2[\text{Voto}] = 30$ e $t_1[\text{Lode}] = \text{'Si'}$ questo non determina il valore che deve avere $t_2[\text{Lode}]$ (può essere 'Si' oppure 'No' senza compromettere la correttezza del dato)
- E' possibile dire che $\text{Lode} \rightarrow \text{Voto}$?
Se $t_1[\text{Lode}] = t_2[\text{Lode}] = \text{'Si'}$ allora sicuramente deve essere $t_1[\text{Voto}] = t_2[\text{Voto}] = 30$ ma ...
... se $t_1[\text{Lode}] = t_2[\text{Lode}] = \text{'No'}$ e $t_1[\text{Voto}] = 27$ questo non determina il valore che deve avere $t_2[\text{Voto}]$ (può essere un qualsiasi voto tra 18 e 30). Anche in questo caso **non abbiamo** $\text{Lode} \rightarrow \text{Voto}$



- Poiché certe proprietà valgono **a prescindere dagli specifici attributi in gioco**, useremo una notazione «astratta» e supporremo che le dipendenze **siano già** state definite

R

A	B	C	D
a1	b1	c1	d1
a1	b2	c1	d2
a1	b1	c1	d3

soddisfa la dipendenza funzionale $AB \rightarrow C$

R

A	B	C	D
a1	b1	c1	d1
a1	b1	c2	d2
a1	b2	c1	d3

non soddisfa la dipendenza funzionale $AB \rightarrow C$

Dati uno schema di relazione R e un insieme F di dipendenze funzionali,

- un'istanza di R è **legale** se soddisfa **tutte** le dipendenze in F

$F = \{A \rightarrow B\}$

R

A	B	C	D
a1	b1	c1	d1
a1	b1	c1	d2
a2	b2	c1	d3

L'istanza soddisfa la dipendenza funzionale $A \rightarrow B$ (e quindi è un'istanza **legale**)

... e anche la dipendenza funzionale $A \rightarrow C$... ma...

$A \rightarrow C$ **non** è in F e non è detto che debba sempre essere soddisfatta ...

Se consideriamo una istanza diversa...

$$F = \{A \rightarrow B\}$$

R	A	B	C	D
	a1	b1	c1	d1
	a1	b1	c2	d2
	a2	b2	c1	d3

La nuova istanza soddisfa la dipendenza funzionale $A \rightarrow B$ (e quindi è anch'essa un'istanza **legale**)
ma **non soddisfa** la dipendenza funzionale $A \rightarrow C$, d'altra parte... $A \rightarrow C$ non è in F quindi ... perché dovrebbe essere comunque sempre soddisfatta?

$$F = \{A \rightarrow B, B \rightarrow C\}$$

R

A	B	C	D
a1	b1	c1	d1
a1	b1	c1	d2
a2	b2	c1	d3

ogni istanza legale (cioè ogni istanza che soddisfa sia $A \rightarrow B$ che $B \rightarrow C$ **soddisfa** sempre **anche** la dipendenza funzionale $A \rightarrow C$... possiamo considerarla allora «come se fosse in F »?



Dato uno schema di relazione R e un insieme F di dipendenze funzionali su R ci sono delle dipendenze funzionali **che non sono in F** , ma che **sono soddisfatte da ogni istanza legale di R** .

- Matricola → CodiceFiscale e CodiceFiscale → DataNascita
- devono essere sempre soddisfatte da ogni istanza legale ...
- ... ma allora sarà sempre soddisfatta anche Matricola → DataNascita
- -----
- CodiceFiscale → Nome, Cognome
- deve essere soddisfatta da ogni istanza legale
- ... ma allora saranno sempre soddisfatte anche
- CodiceFiscale → Nome e
- CodiceFiscale → Cognome

Definizione

- Dato uno schema di relazione R e un insieme F di dipendenze funzionali su R
la **chiusura di F** è l'insieme delle dipendenze funzionali che sono soddisfatte da ogni istanza legale di R

Notazione

- F^+

- Se F è un insieme di dipendenze funzionali su R ed r è un'istanza di R che soddisfa **tutte** le dipendenze in F , diciamo che r è **un'istanza legale** di R .
- La *chiusura* di F , denotata con F^+ , è l'insieme di dipendenze funzionali che sono soddisfatte **da ogni** istanza legale di R
- Banalmente si ha che $F \subseteq F^+$.

- *Definizione*
- Dati uno schema di relazione R e un insieme F di dipendenze funzionali
- un sottoinsieme K di uno schema di relazione R è una **chiave** di R se:
 1. $K \rightarrow R \in F^+$
 2. non esiste un sottoinsieme proprio K' di K tale che $K' \rightarrow R \in F^+$



Consideriamo lo schema:

Studente=Matr Cognome Nome Data

Il numero di matricola viene assegnato allo studente per identificarlo



non ci possono essere due studenti con lo stesso numero di matricola



un'istanza di Studente per rappresentare correttamente la realtà non può contenere due tuple con lo stesso numero di matricola



Matr → Matr Cognome Nome Data
deve essere soddisfatta da ogni istanza legale



Matr è una chiave per Studente



- Dati uno schema di relazione R e un insieme F di dipendenze funzionali, possono esistere **più chiavi** di R
- In SQL una di esse verrà scelta come **chiave primaria** (non può assumere valore nullo)
- ESEMPIO
Studente=Matr **CF** Cognome Nome Data

Dipendenze funzionali banali



SAPIENZA
UNIVERSITÀ DI ROMA
DIPARTIMENTO DI INFORMATICA

- Dati uno schema di relazione R e due sottoinsiemi non vuoti X, Y di R tali che $Y \subseteq X$ si ha:

ogni istanza r di R soddisfa la dipendenza funzionale $X \rightarrow Y$

R	A	B	C	D
	a1	b1	c1	d1
	a1	b2	c1	d2
	a1	b1	c1	d3

$X=ABC$

$Y=AB$

$X \rightarrow Y$ è soddisfatta



Pertanto, se $Y \subseteq X$
 $X \rightarrow Y \in F^+$

Una tale dipendenza funzionale è detta **banale**

Dipendenze funzionali (proprietà)



- Dati uno schema di relazione R e un insieme di dipendenze funzionali F , si ha:

$$X \rightarrow Y \in F^+ \Leftrightarrow \forall A \in Y (X \rightarrow A \in F^+)$$

$X \rightarrow Y$ deve essere soddisfatta da **ogni** istanza legale di R

- Se $t_1[X] = t_2[X]$ allora deve essere $t_1[Y] = t_2[Y]$
- Ovviamente se $A \in Y$ e $t_1[A] \neq t_2[A]$, non può essere $t_1[Y] = t_2[Y]$
- Ovviamente se $\forall A \in Y t_1[A] = t_2[A]$, avremo $t_1[Y] = t_2[Y]$

R	A	B	C	D
	a1	b1	c1	d1
	a2	b2	c1	d2
	a1	b1	c1	d3

$$A \rightarrow BC \in F^+$$

$\Downarrow \Uparrow$

$$A \rightarrow B \in F^+$$

$$A \rightarrow C \in F^+$$

- Vedremo più in là, parlando della **terza forma normale (3NF)** che l'insieme F^+ è molto importante
- Ma come si fa a calcolare F^+ ?