Progettazione di una base di dati relazionale - Dipendenze funzionali

---

Index

Schema di relazione

Uno **schema di relazione** R è un insieme di attributi $\{A_1, A_2, \ldots, A_n\}$
Notazione:

- $\text{R} = A_1, A_2, \ldots A_n$
- le prime lettere dell'alfabeto $(\text{A,B,C,} \ldots)$ denotano i singoli attributi
- le ultime lettere dell'alfabeto $(\text{X, Y,} \ldots)$ denotano insiemi di attributi
- Se X e Y sono insiemi di attributi XY denota $X \cup Y$

Tupla

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

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

t_1[NomeStud]=Paolo
t_1[CognomeStud]=Rossi
t_1[Es. sost]=2
t_1[MEDIA]=26,5

t_2[NomeStud]=Mario
t_2[CognomeStud]=Bianchi
t_2[Es. sost]=10
t_2[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[CognomeStud Media]=t_2[CognomeStud Media]

t_1[CognomeStud Nomestud]≠t_2[CognomeStud Nomestud]

Istanza di relazione

Dato uno schema di relazione R una **istanza** di R è un insieme di tuple su R

> ⓘ **Info**
>
> Tutte le "tabelle" che abbiamo visto finora negli esempi sono istanze di qualche schema di relazione

---

Dipendenze funzionali

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:

- $\mathrm{X} \rightarrow \mathrm{Y}$ → si legge X determina funzionalmente Y oppure Y dipende funzionalmente da X
- X → parte sinistra della dipendenza o determinante
- Y → parte destra della dipendenza o dipendente

Dati uno schema R e una dipendenza funzionale $\mathrm{X} \rightarrow \mathrm{Y}$ su R un'istanza r di R **soddisfa** la dipendenza funzionale $\mathrm{X} \rightarrow \mathrm{Y}$ se:

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

> ⓘ **Le dipendenze funzionali non fanno altro che esprimere dei vincoli sui dati**

Nota

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

E' possibile dire che $\mathrm{Lode} \rightarrow \mathrm{Voto}$?
Se $t_1[\mathrm{Lode}] = t_2[\mathrm{Lode}] = \text{'Si'}$ allora sicuramente deve essere $t_1[\mathrm{Voto}] = t_2[\mathrm{Voto}] = 30$.
Ma se $t_1[\mathrm{Lode}] = t_2[\mathrm{Lode}] = \text{'No'}$ e $t_1[\mathrm{Voto}] = 27$ questo non determina il valore che deve avere $t_2[\mathrm{Voto}]$ (può essere un qualsiasi voto tra 18 e 30)

Esempio

R

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

Soddisfa da dipendenza funzionale $\mathrm{AB} \rightarrow \mathrm{C}$

R

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

Non soddisfa la dipendenza funzionale $\mathrm{AB} \rightarrow \mathrm{C}$

Istanza legale

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

Osservazione

F={A →B}

R

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

L'istanza soddisfa la dipendenza funzionale $\mathrm{A} \rightarrow \mathrm{B}$ (e quindi è un'istanza legale) e anche la dipendenza funzionale $\mathrm{A} \rightarrow \mathrm{C}$ ma $\mathrm{A} \rightarrow \mathrm{C}$ non è in F e non è detto che debba sempre essere soddisfatta

F={A →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 $\mathrm{A} \rightarrow \mathrm{B}$ (e quindi è anch'essa un'istanza legale) ma non soddisfa la dipendenza funzionale $\mathrm{A} \rightarrow \mathrm{C}$, d'altra parte $\mathrm{A} \rightarrow \mathrm{C}$ 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 $\mathrm{A} \rightarrow \mathrm{B}$ che $\mathrm{B} \rightarrow \mathrm{C}$ soddisfa sempre anche la dipendenza funzionale $\mathrm{A} \rightarrow \mathrm{C}$). Possiamo considerarla allora "come se fosse in F"?

Dunque 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**

Esempio

$$\text{Matricola} \rightarrow \text{CodiceFiscale} \rightarrow \text{DataNascita}$$

devono essere sempre soddisfatte da ogni istanza legale ma allora sarà sempre soddisfatta anche $\text{Matricola} \rightarrow \text{DataNascita}$

$$\text{CodiceFiscale} \rightarrow \text{Nome, Cognome}$$

deve essere soddisfatta da ogni istanza legale ma allora saranno sempre soddisfatte anche:

- $\text{CodiceFiscale} \rightarrow \text{Nome}$
- $\text{CodiceFiscale} \rightarrow \text{Cognome}$

Chiusura di un insieme di dipendenze funzionali

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^+

F ed 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 $\mathrm{F} \subseteq \mathrm{F}^+$

Due insiemi di dipendenze funzionali che hanno la stessa chiusura avranno le stesse istanze legali

$$F \subseteq F^+ = G^+ \supseteq G$$

Chiave

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** se $\mathrm{K} \to \mathrm{R} \in \mathrm{F}^+$ e non esiste un sottoinsieme proprio K' di K tale che $\mathrm{K}' \to \mathrm{R} \in \mathrm{F}^+$

Esempio

Consideriamo lo schema

$$\mathrm{Studente}{=}\mathrm{Matr},\ \mathrm{Cognome},\ \mathrm{Nome},\ \mathrm{Data}$$

Il numero di matricola viene assegnato allo studente per identificarlo

$$\Downarrow$$

Quindi non i possono essere due studenti con lo stesso numero di matricola

$$\Downarrow$$

Quindi un'istanza di $\mathrm{Studente}$ per rappresentare correttamente la realtà non può contenere due tuple con lo stesso numero di matricola

$$\Downarrow$$

Quindi $\mathrm{Matr} \to \mathrm{Matr},\ \mathrm{Cognome},\ \mathrm{Nome},\ \mathrm{Data}$ deve essere soddisfatta da ogni istanza legale

$$\Downarrow$$

Quindi Matr è una chiave per $\mathrm{Studente}$

Chiave primaria

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: $\mathrm{Studente} = \mathrm{Matr}, \mathbf{CF}, \mathrm{Cognome}, \mathrm{Nome}, \mathrm{Data}$
Se prendiamo $\mathbf{CF}$ come chiave primaria Matr deve essere UNIQUE

Dipendenze funzionali banali

Dati uno schema di relazione R e due sottoinsiemi non vuoti X, Y di R tali che $\mathrm{Y} \subseteq \mathrm{X}$ si ha che ogni istanza r di R soddisfa la dipendenza funzionale $\mathrm{X} \to \mathrm{Y}$

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

X=ABC
Y=AB

X→Y è soddisfatta

Pertanto se $\mathrm{Y} \subseteq \mathrm{X}$ allora $\mathrm{X} \to \mathrm{Y} \in \mathrm{F}^+$
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:

$$\mathrm{X} \to \mathrm{Y} \in \mathrm{F}^+ \Leftrightarrow \forall \mathrm{A} \in \mathrm{Y} \, (\mathrm{X} \to \mathrm{A} \in \mathrm{F}^+)$$

$\mathrm{X} \to \mathrm{Y}$ deve essere soddisfatta sa **ogni** istanza legale di R

- se $t_1[\mathrm{X}] = t_2[\mathrm{X}]$ allora deve essere $t_1[\mathrm{Y}] = t_2[\mathrm{Y}]$
- ovviamente se $\mathrm{A} \in \mathrm{Y}$ e $t_1[\mathrm{A}] \neq t_2[\mathrm{A}]$, non può essere $t_1[\mathrm{Y}] = t_2[\mathrm{Y}]$
- ovviamente se $\forall \mathrm{A} \in \mathrm{Y} \, t_1[\mathrm{A}] = t_2[\mathrm{A}]$, avremo $t_1[\mathrm{Y}] = t_2[\mathrm{Y}]$

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

$A \to BC \in F^+$

$\Downarrow\Uparrow$

$A \to B \in F^+$
$A \to C \in F^+$