# Terza forma normale

## Index

- Quali sono i problemi di uno schema mal progettato?
  - Considerazioni
  - Conclusioni
- Terza forma normale
  - Definizione
  - Esempi
- <u>Dipendenza parizale</u>
  - Definizione
- <u>Dipendenza transitiva</u>
  - Definizione
- Definizione alternativa di 3NF (teorema)
- Cosa vogliamo ottenere?
  - La 3NF non basta
    - Esempi
- Forma normale di Boyce-Codd
  - Esempio

# Quali sono i problemi di uno schema mal progettato?

Ritorniamo al nostro esempio di base di dati che contiene le informazioni sugli studenti e sugli esami sostenuti, e ripartiamo dalla soluzione "buona" tornata alla fine

## **Ipotesi 3**

La base di dati consiste di quattro schemi di relazione:

• Studente (Matr, CF, Cogn, Nome, Data, Com)

- Corso (C#, Tit, Doc)
- Esame (Matr, C#, Data, Voto)
- Comune (Com, Prov)

Studente	Matr	CF	Cogn		Nome	Data	Com					
	01		Rossi		Mario		Tolfa		Comu	ne Co	m	Prov
	02		Bianchi	i	Paolo		Tolfa			Tol	fa	Roma
·		•										
Es	same	Matr	C#	Data	Voto			`orso	C#	Tit	Doc	
Es	same	Matr 01	<b>C#</b>	Data	Voto		c	Corso	C#	Tit	Doc	_
Es	same						c	Corso	<b>C#</b> 10 20	Tit Fisica Chimica	Doc Pipp	00

Per progettare uno schema "buono" occorre rappresentare separatamente ogni concetto in una relazione distinta

#### Considerazioni

Poiché il numero di matricola identifica univocamente uno studente, ad ogni numero di matricola corrisponde:

- un solo codice fiscale (Matr o CF)
- un solo cognome ( $\mathrm{Matr} \to \mathrm{Cogn}$ )
- un solo nome ( $Matr \rightarrow Nome$ )
- una sola data di nascita ( $\mathrm{Matr} o \mathrm{Data}$ )
- un solo comune di nascita ( $\mathrm{Matr} \to \mathrm{Com}$ )

Quindi un'istanza di Studente per essere legale deve soddisfare la dipendenza funzionale

$$Matr \rightarrow Matr, CF, Cogn, Nome, Data, Com$$

Con considerazioni analoghe abbiamo che un'istanza di Studente per essere legale deve soddisfare la dipendenza funzionale

$$CF \rightarrow Matr, CF, Cogn, Nome, Data, Com$$

Pertanto sia Matr che CF sono chiavi di Studente

D'altra parte possiamo osservare che ci possono essere due studenti con lo stesso cognome e nomi differenti quindi possiamo avere due istanze di Studente che non

soddisfano le dipendenza funzionale  $Cogn \to Nome$ . Inoltre possiamo avere istanze di Studente che non soddisfano:  $Cogn \to Nome$ ,  $Cogn \to Data$ ,  $Cogn \to Com$  ecc.

Con ciò possiamo concludere che le uniche dipendenze funzionali non banali che devono essere soddisfatte da un'istanza legale di Studente sono del tipo

dove K contiene una chiave (Matr o CF)

#### **८** Hint

Vedremo che questa è la prima condizione che però va ulteriormente rifinita per arrivare ad una definizione precisa di terza forma normale (3NF)

Uno studente può sostenere l'esame relativo ad un corso una sola volta; pertanto per ogni esame esiste:

- una sola data (in cui è stato sostenuto)
- ullet un solo voto Quindi ogni istanza legale di  ${\it Esame}$  deve soddisfate la dipendenza funzionale

Matr, 
$$C\# \to Data$$
, Voto

D'altra parte uno studente può sostenere esami in dati differenti e riportare voti diversi nei vari esami. Pertanto esistono istanze di Esame che non soddisfano una o entrambe le dipendenze funzionali:  $Matr \to Data, Matr \to Voto$ 

Inoltre l'esame relativo ad un certo corso non può essere superato da diversi studenti in date diverse e con voti diversi. Pertanto esistono istanze di Esame che non soddisfano una o entrambe le dipendenze funzionali:  $C\# \to Data$ ,  $C\# \to Voto$  Pertanto Matr, C# è una chiave per Esame

## **ઇ Hint**

Vedremo in seguito delle procedure rigorose per identificare la/le chiavi

# Conclusioni

Per ciascun schema di relazione:

• Studente (Matr, CF, Cogn, Nome, Data, Com)

- Corso (C#, Tit, Doc)
- Esame (Matr, C#, Data, Voto)
- Comune (Com, Prov)
- $\odot$  Stiamo continuando ad assumere che  $\mathrm{Com} \to \mathrm{Prov}$ , cioè che non ci sono comuni omonimi

Le uniche dipendenze funzionali non banali che devono essere soddisfatte da ogni istanza legale di sono del tipo

dove K contiene una chiave

## Terza forma normale

Uno schema di relazione è in **3NF** se le uniche dipendenze funzionali non banali (non riflessive) che devono essere soddisfatte da ogni istanza legale sono del tipo

dove K contiene una chiave oppure X è contenuto in una chiave

## **Definizione**

Dati uno shema di relazione R e un insieme di dipendenze funzionali F su R, R è in  $\operatorname{\mathbf{3NF}}$  se

$$orall X o A \in F^+ \quad A 
otin X$$

- A appartiene ad una chiave (è **primo**, applicabile solo ai singleton)
- ullet X contiene una chiave (è una superchiave)

#### **Marning**

- è sbagliato scrivere  $\forall X \to A \in F$ , perché non sapremmo se e come valutare una dipendenza del tipo  $X \to AB$  (due o più attributi a destra)
- se sostituisco  $\forall X \to A \in F$  con  $\forall X \to Y \in F$ , non so come comportarmi se Y contiene sia attributi primi che non
- la condizione  $A \not\in X$  è importante. Infatti, per l'assioma della riflessività, se  $A \in X$  avremo sempre  $X \to A \in F^A$  e quindi in  $F^+$ , anche quando A non

è primo e X non è superchiave, e quindi se considerassimo questo tipo di dipendenze nessuno schema risulterebbe in 3NF

# Esempi

#### **≔** Esempio 1

$$R = A, B, C, D$$

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

Ho come chiavi:

•  $K_1 = A \rightarrow A \rightarrow CD$  per transitività e  $A \rightarrow BCD$  per <u>unione</u>

A 
ightarrow B 
ightarrow rispetta le condizioni per essere in 3NF

 $B \to BC \to B$  non è superchiave, controllo quindi i determinati. C e D violano la 3NF perché in entrambi i casi non fanno parte di una chiave

Lo schema non è in 3NF

#### **≔** Esempio 2

$$R=A,B,C,D$$
  $F=\{AB o CD,AC o BD,D o BC\}$ 

Ho come chiavi:

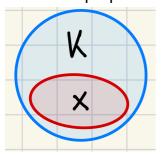
- $K_1 = AB$
- $K_2 = AC$
- $K_3 = AD$  (per il teorema dell'aumento sull'ultima dipendenza funzionale)

D nell'ultima dipendenza non è una chiave (ma un pezzo di una chiave) però il determinato è composto da B (un attributo della chiave AB) e da C (un attributo della chiave AC); dunque lo schema è in **3FN**.

Ho infatti decomposto in  $D \to B$  (B è parte di una chiave) e in  $D \to C$  (C è una parte di una chiave)

# Dipendenza parziale

Siano R uno schema di relazione e F un insieme di dipendenze funzionali su R.  $X \to A \in F^+ \mid A \not\in X$  è una **dipendenza parziale** su R se A non è primo ed X è contenuto propriamente in una chiave di R



#### **Definizione**

A dipende parzialmente da una chiave K se  $\exists X\subset R$  tale che  $K\to R\in F^+$  con  $A\not\in X$  e tale che  $X\subset K$  e A non è parte di una chiave

#### **∃** Example

Curriculum (Matr, CF, Cogn, Nome, DataN, Com, Prov, C#, Tit, Doc, Dat

Ad un numero di matricola corrisponde un solo cognome:  $Matr \to Cogn$  Quindi ad una coppia costituita da un numero di matricola e da un codice di corso corrisponde un solo cognome:  $Matr, C\# \to Cogn$ 

L'attributo Cogn dipende parzialmente dalla chiave Matr, C# (  $Matr, C\# \to Cogn$  è una conseguenza di  $Matr \to Cogn$ )

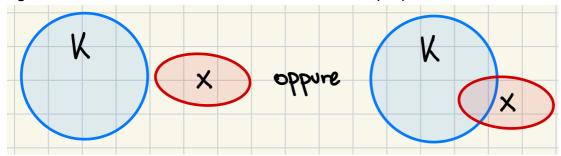
(i) Info

 $\operatorname{Matr}$  è propriamente contenuto in una chiave

# Dipendenza transitiva

Siano R uno schema di relazione e F un insieme di dipendenze funzionali su R.  $X \to A \in F^+ \mid A \not\in X$  è una **dipendenza transitiva** su R se A non è primo e per

ogni chiave K di R si ha che X non è contenuto propriamente in K e  $K-X \neq \varnothing$ 



#### **Definizione**

A dipende parzialmente da una chiave K se  $\exists X\subset R$  tale che  $K\to R\in F^+$  con  $A\not\in X$  e  $X\to A\in F^+$  e A non è parte di una chiave

#### **∃** Example

Studente (Matr, CF, Cogn, Nome, Data, Com, Prov)

Ad un numero di matricola corrisponde un solo comune di nascita:  $Matr \to Com$  Un comune si trova in una sola provincia:  $Com \to Prov$ 

Quindi ad un numero di matricola corrisponde una sola provincia:  $\mathrm{Matr} \to \mathrm{Prov}$ 

L'attributo Prov dipende transitivamente dalla chiave Matr ( $Matr \to Prov$  è una conseguenza di  $Matr \to Com$  e  $Com \to Prov$ )

## (i) Info

Per ogni chiave K di R (Matr e CF) Com non è contenuto propriamente nella chiave (Com non è sottoinsieme né di Matr né di CF) e  $K-\mathrm{Com} \neq \varnothing$  (Matr  $-\mathrm{Com} = \mathrm{Matr}$  e CF  $-\mathrm{Com} = \mathrm{CF}$ )

# Definizione alternativa di 3NF (teorema)

Dato uno schema R e un insieme funzionali F su R, R è in 3NF se e solo se in F non ci sono né dipendenze parziali né dipendenze transitive

(i) Dimostrazione

Parte ⇒

Lo schema R è in 3NF, quindi  $\forall X \to A \in F^+, A \notin X$ A appartiene ad una chiave (è primo) oppure X contiene una chiave (è superchiave)

- ullet Se A è parte di una chiave (è primo), viene a mancare la prima condizione per avere una dipendenza parziale o transitiva
- Se A non è primo (non fa parte di nessuna chiave), allora X è superchiave. Dunque si ha che  $X\supset K$  facendo mancare la seconda condizione per la dipendenza parziale (non può essere che  $X\subset K$ ); per lo stesso motivo non si può verificare che  $K-X\neq\varnothing$  quindi la dipendenza non può essere transitiva

#### Parte ←

Supponiamo per assurdo che non ci sono dipendenze parziali e transitive e lo schema non sia in 3NF. In tal caso vuol dire che ci sta almeno una dipendenza che viola la 3NF (ovvero che A non è primo e X non è una superchiave). Siccome A non è primo devo verificare il secondo punto della 3NF

- Se X non è superchiave vuol dire che  $\forall K \in R, X \not\subset K$  e  $K X \neq \emptyset$  e quindi vuol dire che ci sta una dipendenza transitiva. **CONTRADDIZIONE**
- Se X è una porzione di una chiave, ovvero se  $\exists K \in R \mid X \subset K$ , in tal caso  $X \to A$  è una dipendenza parziale su R. CONTRADDIZIONE

# Cosa vogliamo ottenere?

Abbiamo visto che uno schema in 3NF ha delle buone proprietà che lo rendono preferibile ad uno che non è in 3NF. Un obbiettivo da tener presente quando si progetta una base di dati è quello di produrre uno schema in un ogni relazione sia in 3NF. Normalmente nella fase di progettazione concettuale si usa il modello Entità-Associazione e si individuano per l'appunto i concetti che devono essere rappresentati nella base di dati

Se il lavoro di individuazione è fatto accuratamente lo schema relaziona può essere derivato con opportune regole, è in 3NF. Se tuttavia, dopo tale processo, ci ritrovassimo a produrre uno schema che non è in 3NF dovremmo procedere ad una fare di **decomposizione** di tale schema in maniera analoga a quella esaminata nell'esempio sui dati di un'Università (qui)

### La 3NF non basta

Uno schema che non è in 3NF può essere decomposto in più modi in un insieme di schemi in 3NF. Ad esempio lo schema R=ABC con l'insieme di dipendenze funzionali  $F=\{A\to B, B\to C\}$  non è in 3NF per la presenza in  $F^+$  della dipendenza transitiva  $B\to C$ , dato che la chiave è evidentemente A. R può essere decomposto in:

• 
$$R_1 = AB \operatorname{con} \{A \to B\}$$

• 
$$R_2 = BC ext{ con } \{B o C\}$$
 oppure:

• 
$$R_1 = AB \operatorname{\mathsf{con}} \left\{ A o B \right\}$$

• 
$$R_2 = AC \operatorname{con} \left\{ A o C \right\}$$

Entrambi gli schemi sono in 3NF, tuttavia la seconda soluzione non è soddisfacente. Consideriamo due istanze legali degli schemi ottenuti

L'istanza dello schema originario R che posso ricostruire da questa attraverso il join naturale è la seguente

R	A	В	С
	a1	b1	<b>c</b> 1
	a2	b1	c2

Ma non è un'istanza legale di R in quanto non soddisfa la dipendenza funzionale  $B \to C$ 

## 

Occorre preservare tutte le dipendenze in  $F^+$ 

Per **perdita di dati** si intende sia la perdita delle tuple originali sia l'aggiunta di tuple non presenti originariamente

In conclusione, quando si decompone uno schema per ottenerne uno in 3NF occorre tenere presenti altri due requisiti dello schema decomposto:

 deve preservare le dipendenze funzionali che valvono su ogni isntanza legasle dello schema originario  deve permettere di ricostruire mediante il join naturale ogni istanza legale dello schema originario (senza aggiunta di informazione estranea)

#### **ර Hint**

Se ci sono delle dipendenze che rispettano la seconda condizione del 3NF (determinato primo) devo fare attenzione ad aggiungere dei contrains nella decomposizione dello schema in quanto si potrebbe violare una dipendenza funzionale

## **Esempi**

#### **∃** Example

Consideriamo lo schema  $R = \{ \text{Matricola, Comune, Provincia} \}$  con l'insieme di dipendenze funzionali  $F = \{ \text{Matricola} \to \text{Comune, Comune} \to \text{Provincia} \}$  Lo schema non è in 3NF per la presenza in  $F^+$  della dipendenza transitiva  $\text{Comune} \to \text{Provincia}$ , dato che la chiave è evidentemente Matricola ( Provincia dipende transitivamente da Matricola)

 ${\it R}$  può essere decomposto in:

- $R_1 = (\text{Matricola}, \text{Comune}) \text{ con } \{\text{Matricola} \rightarrow \text{Comune}\}$
- $R_2 = (Matricola, Provincia) con {Comune <math>\rightarrow Provincia}$

#### oppure:

- $R_1(Matricola, Comune)$  con  $\{Matricola \rightarrow Comune\}$
- $R_2({
  m Matricola},{
  m Provincia})$  con  $\{{
  m Matricola}
  ightarrow {
  m Provincia}\}$

Entrambi gli schemi sono in 3NF, tuttavia la seconda soluzione non è soddisfacente

Consideriamo le istanze legali degli schemi ottenuti

R1	Matricola	Comune	R2	Matricola	Provincia
	501	Tivoli		501	Roma
	502	Tivoli		502	Rieti

L'istanza dello schema originario R che posso ricostruire da questa (facendo un join naturale) è la seguente

R	Matricola	Comune	Provincia
	501	Tivoli	Roma
	502	Tivoli	Rieti

Ma non è un'istanza legale di R, in quanto non soddisfa la dipendenza funzionale  $\operatorname{Comune} \to \operatorname{Provincia}$ 

#### & Hint

E' evidente che ci sia stato un errore di inserimento, ma non abbiamo potuto rivelarlo

#### **∃** Example

Consideriamo ora lo schema R=ABC con l'insieme di dipendenze funzionali  $F=\{A\to B, B\to C\}$ . Lo schema non è in 3NF per la presenza in  $F^+$  delle dipendenze funzionali parziali  $A\to B$  e  $C\to B$ , dato che la chiave è AC tale schema può essere decomposto in:

• 
$$R_1 = AB \operatorname{con} \{A \to B\}$$

• 
$$R_2 = BC \operatorname{\mathsf{con}} \left\{ C o B \right\}$$

Tale schema pur preservando tutte le dipendenze in  $F^+$  non è soddisfacente.

Consideriamo l'istanza legale di  ${\cal R}$ 

R	A	В	С
	a1	b1	<b>c</b> 1
	a2	b1	c2

In base alla decomposizione data, questa istanza di decompone in:

E dovrebbe essere possibile ricostruirla esattamente tramite join e invece se si

effettua il join si avrà come risultato il prodotto cartesiano

R	A	В	С
	a1	b1	c1
	a2	b1	c2
	a1	b1	c2
	a2	b1	c1

#### **∃** Example

Consideraimo ora lo schema

- R = (Matricola, Progetto, Capo)
- $F = \{ \text{Matricola} \rightarrow \text{Progetto}, \text{Capo} \rightarrow \text{Progetto} \}$

Il progetto ha più capi ma ogni capo ha un solo progetto, e un impiegato su un progetto dà conto ad un solo capo (ogni capo segue un gruppo).

Lo schema non è in 3NF per la presenza in F+ delle dipendenze parziali  ${
m Matricola} o {
m Progetto}$  e  ${
m Capo} o {
m Progetto}$ .

Dato che la chiave è (Matricola, Capo) tale schema può essere decomposto in:

- $R_1 = (Matricola, Progetto) con \{Matricola \rightarrow Progetto\}$
- $R_2 = ( ext{Progetto}, ext{Capo}) \text{ con } \{ ext{Capo} o ext{Progetto}\}$

Tale schema pur preservando tutte le dipendenze in  $F^+$  non è soddisfacente

Consideriamo l'istanza legale di  ${\it R}$ 

Matric	ola	Progetto	Саро
501		30	E1
502		30	E2

In base alla decomposizione data, questa istanza si decompone in:

R1	Matricola	Progetto	R2	Progetto	Саро
	501	30		30	E1
	502	30		30	E2

Ma quando la si ricostruisce tramite join si ottiene uno schema in cui sono presenti

delle tuple estranee (perdita di informazione)

R	Matricola	Progetto	Capo
	501	30	E1
	502	30	E2
	501	30	E2
	502	30	E1

# Forma normale di Boyce-Codd

La 3NF non è la più restrittiva che si può ottenere ma ne esistono altre tra cui la forma normale di **Boyce-Codd** 

#### (i) Definizione

Una relazione è in forma normale di Boyce-Codd (BCNF) quando in essa **ogni determinante è una superchiave** (ricordiamo che una chiave è anche superchiave)

Una relazione che rispetta la forma normale di Boyce-Codd è anche in terza forma normale, ma non il contrario

# **Esempio**

#### **: Example**

Consideriamo una relazione che descrive l'allocazione delle sale operatorie di un ospedale. Le sale operatorie sono prenotate, giorno per giorno, in orari previsti, per effettuare interventi su pazienti ad opera dei chirurghi dell'ospedale.

Nel corso di una giornata una sala operatoria è occupata sempre dal medesimo chirurgo che effettua più interventi, in ore diverse.

Noti i valori di Paziente e DataIntervento, sono noti anche: ora dell'intervento, chirurgo e sala operatoria utilizzata

#### Schema:

Interventi(Paziente, DataIntervento, OraIntervento, Chirurgo, Sala) In base alla presedente descrizione, nella relazione Interventi valgono le dipendenze funzionali:

- $\{Paziente, DataIntervento\} \rightarrow OraIntervento, Chirurgo, Sala$
- {Chirurgo, DataIntevento, OraIntervento}  $\rightarrow$  Paziente, Sala
- $\{Sala, DataIntervento, OraIntervento\} \rightarrow Paziente, Chirurgo$
- {Chirurgo, DataIntervento}  $\rightarrow$  Sala

#### Ho come chiavi:

- $K_1 = \{Paziente, DataIntervento\}$
- $K_2 = \{ \text{Chirurgo, DataIntervento, OraIntervento} \}$
- $K_3 = \{ Sala, DataIntervento, OraIntervento \}$

Qualunque sia la chiave primaria che scegliamo, i determinanti nelle prime 3 dipendenze funzionali sono insiemi di attributi che possono svolgere la funzione di chiave e quindi la BCNF non è sicuramente violata in questi casi.

La BCNF non è invece soddisfatta per la quarta dipendenza funzionale che ha come determinante un insieme di attributi non chiave. Ne segue che la relazione Interventi non è in BCNF

Ma Interventi è in 3NF in quanto la quarta dipendenza funzionale non viola la definizione, perché l'attributo Sala è un attributo che fa parte della chiave {Sala, DataIntervento, OraIntervento} e quindi è primo

Conseguenza: la relazione Interventi, pur essendo in terza forma normale, presenta una certa ridondanza nei dati che può creare problemi in fase di aggiornamento.

Se per qualche ragione si deve cambiare la sala operatoria utilizzata da in chirurgo in una certa data, bisogna aggiornate più righe: per esempio, per spostare Romano dalla Sala2 alla Sala3, bisogna modificare due righe della tabella

#### Interventi

Paziente	DataIntervento	OraIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2

La tabella Interventi può essere normalizzata, ottenendo i due schemi:

- OccupazioneSale(Chirurgo, DataIntervento, Sala)
- Interventi(Paziente, DataInterventi, OraIntervento, Chirurgo)

L'attributo Sala viene tolto da Interventi e compare in una nuova tabella che ha come chiave il determinante della dipendenza funzionale che non rispettava la BCNF

#### Interventi

Paziente	DataIntervento	OraIntervento	Chirurgo
Bianchi	25/10/2005	8.00	De Bakey
Rossi	25/10/2005	8.00	Romano
Negri	26/10/2005	9.30	Veronesi
Viola	25/10/2005	10.30	De Bakey
Verdi	25/10/2005	11.30	Romano

#### OccupazioneSale

Chirurgo	DataIntervento	Sala
De Bakey	25/10/2005	Sala1
Romano	25/10/2005	Sala2
Veronesi	26/10/2005	Sala1

#### **Problema**

Supponiamo di voler tenere traccia dei pazienti che devono essere sottoposti a più interventi chirurgici, in diversi reparti, per la cura di patologie più complicate. Una relazione che rappresenta questa esigenza è mostrata nella tabella sotto

#### ChirurgieMultiple

Paziente	Reparto	Chirurgo
Rossi	Cardiochirurgia	De Bakey
Rossi	Chir. Generale	Romano
Bianchi	Chir. Generale	Romano
Bianchi	Chir. Oncologica Veronesi	
Verdi	Chir. Generale	Lanzetta

Ogni n-upla della relazione ChirurgieMultiple associa un paziente al chirurgo che lo ha operato e al reparto nel quale è avvenuto l'intervento. Valgono le dipendenze funzionali:

- Chirurgo  $\rightarrow$  Reparto
- {Paziente, Reparto}  $\rightarrow$  Chirurgo

 $\{ {
m Paziente}, {
m Reparto} \}$  è chiave e la prima dipendenza viola Boyce-Codd. Proviamo quindi a procedere come prima

#### Chirurghi

Chirurgo	Reparto
De Bakey	Cardiochirurgia
Romano	Chir. Generale
Veronesi	Chir. Oncologica
Lanzetta	Chir. Generale

#### **Pazienti**

Paziente	Chirurgo
Rossi	De Bakey
Rossi	Romano
Bianchi	Romano
Bianchi	Veronesi
Verdi	Lanzetta

Questa decomposizione non conserva la seconda delle due dipendenze funzionali. Se, per esempio, si volesse registrare il fatto (errato) che il paziente Bianchi è stato operato da Lanzetta nel reparto di Cardiochirurgia, l'interimento nella tabella Paziente della coppia di valori: (Bianchi, Lanzetta) sarebbe consentito.

Solo quando si cerca di ricostruire i dati della relazione ChirurgieMultiple con un join tra Pazienti e Chirurghi la n-upla: (Bianchi, Lanzetta, Chir.Generale) poltrebbe evidenziare l'errore dei dati perché < Bianchi - Chir. Generale > dovrebbe essere associato a Romano