

# Lezione 2 – Il modello relazionale

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



SAPIENZA  
UNIVERSITÀ DI ROMA



- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati
- Disponibile in DBMS reali nel 1981 (non è facile implementare l'indipendenza con efficienza e affidabilità!)

- Basato sulla nozione matematica di relazione
- Le relazioni si traducono in maniera naturale in tabelle (infatti useremo sempre il termine relazione invece di tabella)
- Dati e relazioni (riferimenti) tra dati di insiemi (tabelle) diversi sono rappresentati come valori
- **ATTENZIONE** ai diversi significati con cui può essere utilizzato il termine relazione



- **relazione matematica**: come nella teoria degli insiemi
- **relazione** secondo il modello relazionale dei dati
- **relazione** (dall'inglese **relationship**) che rappresenta una classe di fatti, nel modello concettuale Entità-Relazioni (Entity-Relationship); tradotto anche con **associazione** perchè rappresenta appunto un tipo di collegamento (relazione) concettuale tra entità diverse (ad esempio Esame mette in relazione uno Studente ed un Insegnamento) ... ma questo sarà trattato nel secondo modulo del corso



- **Dominio:** un insieme possibilmente infinito di valori;
- esempi:
  - l'insieme dei numeri interi è un dominio
  - l'insieme dei numeri decimali è un dominio
  - l'insieme delle stringhe di caratteri di lunghezza = 20 è un dominio
  - $\{0,1\}$  è un dominio

Siano  $D1, D2, \dots, Dk$  domini, non necessariamente distinti. Il prodotto Cartesiano di tali domini, denotato da

$$D1 \times D2 \times \dots \times Dk$$

è l'insieme

$$\{(v1, v2, \dots, vk) \mid v1 \in D1, v2 \in D2, \dots, vk \in Dk\}$$

  
lista ordinata di valori

  
tale che

  
appartiene



- Una relazione matematica è un qualsiasi sottoinsieme del prodotto Cartesiano di uno o più domini
- Una relazione che è sottoinsieme del prodotto Cartesiano di  $k$  domini si dice di grado  $k$
- Gli elementi di una relazione sono detti tuple (oppure n-uple oppure ennuple). Il numero di tuple di una relazione è la sua cardinalità
- Ogni tupla di una relazione di grado  $k$  ha  $k$  componenti ordinate (l' $i$ -esimo valore viene dall' $i$ -esimo dominio) ma non c'è ordinamento tra le tuple
- Le tuple di una relazione sono tutte distinte (almeno per un valore) ... dai vincoli sugli insiemi!

- Supponiamo  $k = 2$
- $D1 = \{\text{bianco}, \text{nero}\}$ ,  $D2 = \{0, 1, 2\}$

$D1 \times D2 = \{(\text{bianco}, 0), (\text{bianco}, 1), (\text{bianco}, 2), (\text{nero}, 0), (\text{nero}, 1), (\text{nero}, 2)\}$

$\{(\text{bianco}, 0), (\text{nero}, 0), (\text{nero}, 2)\}$  è una relazione di grado 2, cardinalità 3 e con tuple  $(\text{bianco}, 0), (\text{nero}, 0), (\text{nero}, 2)$

$\{(\text{nero}, 0), (\text{nero}, 2)\}$  è una relazione di grado 2, cardinalità 2 e con tuple  $(\text{nero}, 0), (\text{nero}, 2)$



- $D1=\{a,b\}$
- $D2=\{x,y,z\}$

- prodotto cartesiano  $D1 \times D2 \longrightarrow$

a	x
a	y
a	z
b	x
b	y
b	z


- una relazione  $r \subseteq D_1 \times D_2 \longrightarrow$

a	x
a	z
b	y



- Quasi sempre si usano domini predefiniti, comuni ai linguaggi di programmazione
- La seguente è una relazione con due 4-ple su String, String, Integer, Real

Paolo	Rossi	2	26,5
Mario	Bianchi	10	28,7

- Se  $r$  è una relazione di grado  $k$
- se  $t$  è una tupla di  $r$
- se  $i$  è un numero intero nell'insieme  $\{1, \dots, k\}$   
 $t[i]$  (oppure  $t.i$ ) indica la  $i$ -sima componente di  $t$
- esempio:
  - se  $r = \{(0,a), (0,c), (1,b)\}$
  - $t = (0,a)$  è una tupla di  $r$  
  - $t[2] = a$
  - $t[1] = 0$
  - $t[1,2] = (0, a)$

0	a
0	c
1	b

- Problema:
- Come interpretare i dati nella tabella?

Paolo	Rossi	2	26,5
Mario	Bianchi	10	28,7

- Soluzione:
- Assegnare nomi alla tabella e alle colonne

Studente	Nome	Cognome	Es. sost.	Media
	Paolo	Rossi	2	26,5
	Mario	Bianchi	10	28,7

**Da dati a informazioni!**



- Un **attributo** è definito da un nome  $A$  e dal **dominio** dell'attributo  $A$  che indichiamo con  $dom(A)$ .
- Sia  $R$  un insieme di attributi. Un'ennupla (tuple, in inglese) *su*  $R$  è una funzione definita su  $R$  che associa ad ogni attributo  $A$  in  $R$  un elemento di  $dom(A)$ .
- Se  $t$  è un'ennupla su  $R$  ed  $A$  è un attributo in  $R$ , allora con  $t(A)$  indicheremo il valore assunto dalla funzione  $t$  in corrispondenza dell'attributo  $A$ .

nome di relazione



attributi

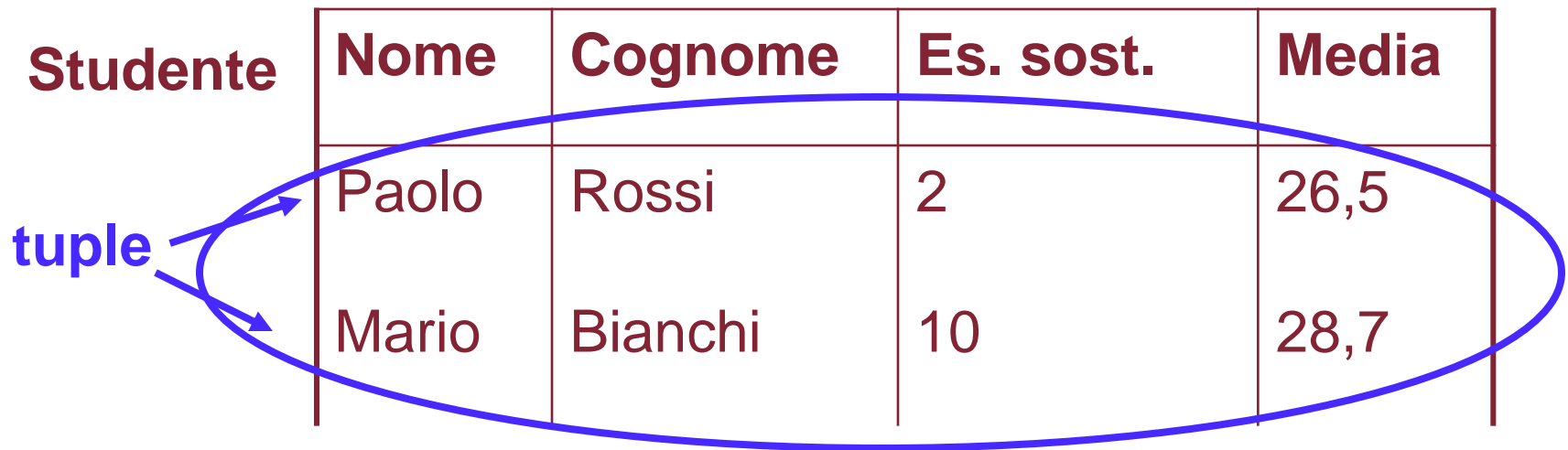
schema di relazione

Studente	Nome	Cognome	Es. sost.	Media

**Studente**

Nome	Cognome	Es. sost.	Media
Paolo	Rossi	2	26,5
Mario	Bianchi	10	28,7

**tuple**



**istanza di relazione**



- Una relazione può essere implementata come una tabella in cui ogni riga è una tupla della relazione **differente da ogni altra** e ogni colonna corrisponde ad una componente (valori omogenei, cioè provenienti dallo stesso dominio)
- Le colonne corrispondono ai domini  $D_1, D_2, \dots, D_k$  e hanno associati dei nomi univoci all'interno della tabella, detti nomi degli attributi (descrivono il ruolo)
- La coppia (nome di attributo, dominio) è chiamata attributo. L'insieme di attributi di una relazione è detto schema
- Se una relazione è denominata  **$R$**  e i suoi attributi hanno nomi  $A_1, A_2, \dots, A_k$ , lo schema è spesso indicato da
$$R(A_1, A_2, \dots, A_k)$$

- **Schema di relazione:** un nome di relazione  $R$  con un insieme di nomi di attributi (differenti ....)

$$R(A1, A2, \dots, Ak)$$

- Ricordiamo che lo **schema** di una relazione è sostanzialmente invariante nel tempo, e ne descrive la **struttura** (aspetto intensionale):
- **Istanza** di una relazione con schema  $R(X)$ : insieme  $r$  di tuple su  $X$
- Ricordiamo che l'**istanza** contiene i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale)

- **Schema di base di dati:** un insieme di schemi di relazione con nomi differenti
- **Schema di base di dati relazionale:** insieme  $\{R_1, R_2, \dots, R_n\}$  di schemi di relazione
- **Base di dati relazionale** con schema  $\{R_1, R_2, \dots, R_n\}$ : insieme  $\{r_1, r_2, \dots, r_n\}$  dove  $r_i$  è una istanza di relazione con schema  $R_i$

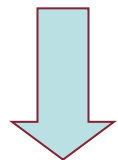
- Schema Info\_Città(Città,Regione,Popolazione)
- Istanza di relazione Info\_Città

Città	Regione	Popolazione
Roma	Lazio	3000000
Milano	Lombardia	1500000
Genova	Liguria	800000
Pisa	Toscana	150000



- Nell'ultima definizione di modello relazionale, le componenti di una relazione sono indicate dai nomi degli attributi, anziché dalla posizione
- $t[A_i]$  indica il valore dell'attributo con nome  $A_i$  della tupla  $t$
- Se  $t$  è la seconda tupla nell'esempio precedente, allora  $t[\text{Regione}] = \text{Lombardia}$
- Se  $Y$  è un sottoinsieme di attributi dello schema  $X$  di una relazione ( $Y \subseteq X$ ) allora  $t[Y]$  è il sottoinsieme dei valori nella tupla  $t$  che corrispondono ad attributi contenuti in  $Y$  (detto anche **restrizione** di  $t$ )

- Oggetto = tupla (implementata come record)
- Campi = Informazioni di interesse → **SCHEMA** della relazione



- Oggetto = “Membro dello Staff”
- Informazioni di interesse = Codice, Cognome, Nome, Ruolo, Anno di assunzione

CODICE	COGNOME	NOME	RUOLO	ASSUNZIONE
COD1	Rossi	Mario	Analista	1995

- Tabella = Insieme di tuple di tipo omogeneo → una particolare **ISTANZA** della relazione



- Tabella STAFF = Insieme di tuple di tipo “Membro dello Staff”

CODICE	COGNOME	NOME	RUOLO	ASSUNZIONE
COD1	Rossi	Mario	Analista	1995
COD2	Bianchi	Pietro	Analista	1990
COD3	Neri	Paolo	Amministratore	1985



- Nel modello relazionale i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di **valori** dei domini che compaiono nelle ennuple



**studenti**

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Data di nascita</b>
6554	Rossi	Mario	05/12/1978
8765	Neri	Paolo	03/11/1976
9283	Verdi	Luisa	12/11/1979
3456	Rossi	Maria	01/02/1978

**esami**

<b>Studente</b>	<b>Voto</b>	<b>Corso</b>
3456	30	04
3456	24	02
9283	28	01
6554	26	01

**corsi**

<b>Codice</b>	<b>Titolo</b>	<b>Docente</b>
01	Analisi	Mario
02	Chimica	Bruni
04	Chimica	Verdi

- I valori NULLI rappresentano mancanza di informazione o il fatto che l'informazione non è applicabile
- es. numero di telefono:
  - la persona non ha il telefono
  - non so se la persona ha il telefono
  - la persona ha il telefono ma non ne conosco il numero
- Non possiamo non inserire il campo (la tupla deve aderire allo schema completo)! Al massimo possiamo stabilire un default
- Attenzione. Alcuni campi NON DOVREBBERO MAI assumere valori nulli
  - es. Matricola di Studente (non posso più identificarlo)
  - ... ma anche Voto in Esame (manca una informazione essenziale)

- Cattiva abitudine: usare valori del dominio “inutilizzati”
  - Potrebbero non esistere
  - Potrebbero essere utilizzati in seguito
  - Potrebbero falsare i calcoli (non conosco lo stipendio di un impiegato, ma uno 0 pesa in un calcolo del valore medio come un qualsiasi altro numero)
- Valore speciale: **NULL**
- NULL: valore polimorfo = non appartiene a nessun dominio ma può sostituire valori in qualsiasi dominio
- Due valori NULL, anche se sullo stesso dominio, sono considerati diversi
- Attenzione! NULL non è 0 (numero intero)

# Troppi valori nulli



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

studenti	Matricola	Cognome	Nome	Data di nascita
	6554	Rossi	Mario	05/12/1978
	9283	Verdi	Luisa	12/11/1979
	NULL	Rossi	Maria	01/02/1978

esami	Studente	Voto	Corso
	NULL	30	NULL
	NULL	24	02
	9283	28	01

corsi	Codice	Titolo	Docente
	01	Analisi	Mario
	02	NULL	NULL
	04	Chimica	Verdi

# Una base di dati “scorretta”



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

## IMPIEGATO

CODICE      COGNOME      NOME      RUOLO      ASSUNZIONE      DIP

COD1	Rossi	Mario	Analista	1795	01
COD2	Bianchi	Pietro	Analista	1990	05
COD2	Neri	Paolo	Amministratore	1985	01

## DIPARTIMENTO

NUMERO      NOME

01	Progettazione
02	Amministrazione

Cosa c'è che non va ???  
Sintatticamente è corretto ...

# Una base di dati "scorretta"



Esami	Studente	Voto	Lode	Corso
	276545	32		01
	276545	30	si	02
	787643	27	si	03
	739430	15		04

Studenti	Matricola	Cognome	Nome
	276545	Rossi	Mario
	787643	Neri	Piero
	787643	Bianchi	Luca



- Vincolo di integrità: proprietà che deve essere soddisfatta da ogni istanza della base di dati (legata quindi allo schema)
- I vincoli descrivono proprietà specifiche del campo di applicazione, e quindi delle informazioni ad esso relative modellate attraverso la base di dati
- Una istanza di base di dati è corretta se soddisfa **tutti** i vincoli di integrità associati al suo schema

# Come evitare le “scorrettezze”



## IMPIEGATO

**CODICE      COGNOME      NOME      RUOLO      ASSUNZIONE      DIP**

COD1	Rossi	Mario	Analista	1795	01
COD2	Bianchi	Pietro	Analista	1990	05
COD2	Neri	Paolo	Amministratore	1985	01

## DIPARTIMENTO

**NUMERO      NOME**

01	Progettazione
02	Amministrazione

(ASSUNZIONE > 1980)

COD2 UNIQUE

DIP REFERENCES

DIPARTIMENTO.NUMERO



# Come evitare le “scorrettezze”



Esami	Studente	Voto	Lode	Corso	
	276545	32		01	(Voto $\geq 18$ ) AND (Voto $\leq 30$ )
	276545	30	si	02	
	787643	27	si	03	(Voto = 30) OR NOT (Lode = „si“)
	739430	24		04	vincoli intra-relazionali

Studente references Studenti.Matricola      vincoli inter-relazionali

Studenti	Matricola	Cognome	Nome	
	276545	Rossi	Mario	
	787643	Neri	Piero	
	787643	Bianchi	Luca	Matricola unique



- Vincoli di dominio
  - $\text{ASSUNZIONE} > 1980$
  - $(\text{Voto} \geq 18) \text{ AND } (\text{Voto} \leq 30)$
- Vincoli di tupla
  - $(\text{Voto} = 30) \text{ OR NOT } (\text{Lode} = \text{„si"})$
- Vincoli tra tuple della stessa relazione
  - COD2 UNIQUE
  - Matricola unique
- Vincoli tra valori in tuple di relazioni diverse
  - $\text{DIP REFERENCES DIPARTIMENTO.NUMERO}$
  - $\text{Studente references Studenti.Matricola}$



- Vincoli intrarelazionali: definiti sui valori di singoli attributi (di dominio) o tra valori di attributi di una stessa tupla o tra tuple della stessa relazione
- Vincoli interrelazionali: definiti tra più relazioni



- Vincolo di chiave primaria (primary key): unica e mai nulla
- Vincoli di dominio (es. ASSUNZIONE > 1980)
- Vincoli di unicità (unique)
- Vincoli di esistenza del valore per un certo attributo (not null)
- Espressioni sul valore di attributi della stessa tupla (es.  $\text{data\_arrivo} < \text{data\_partenza}$ ) se arrivo e partenza ad esempio da un albergo



- Occorre identificare univocamente le tuple di una istanza di relazione
- Una chiave di una relazione (non è detto che sia unica) è un attributo o insieme di attributi che identifica univocamente una tupla

- Un insieme  $X$  di attributi di una relazione  $R$  è una chiave di  $R$  se soddisfa le seguenti condizioni:
  - 1) Per ogni istanza di  $R$ , non esistono due tuple distinte  $t_1$  e  $t_2$  che hanno gli stessi valori per tutti gli attributi in  $X$ , tali cioè che  $t_1[X] = t_2[X]$
  - 2) Nessun sottoinsieme proprio di  $X$  soddisfa la condizione 1)

- Istanza di relazione Staff

CODICE	COGNOME	NOME	RUOLO	ASSUNZIONE
COD1	Rossi	Mario	Analista	1995
COD2	Bianchi	Pietro	Analista	1990
COD3	Neri	Paolo	Amministratore	1985

Chiave? ... NON basta che in UNA CERTA ISTANZA il valore non venga ripetuto. Dobbiamo essere sicuri che non lo sarà MAI

- Istanza di relazione Info\_Città

Città	Regione	Popolazione
Roma	Lazio	3000000
Milano	Lombardia	1500000
Genova	Liguria	800000
Pisa	Toscana	150000

Chiave? Dipende ... dalla «grandezza» di città



- Una relazione potrebbe avere più chiavi alternative
- Si sceglie quella più usata o quella composta da un numero minore di attributi = chiave **primaria**
- La chiave primaria non ammette valori nulli
- Esiste sempre almeno una chiave ... perché ? ... NON POSSIAMO AVERE DUE TUPLE UGUALI!
- Sono le chiavi che consentono di mettere in relazione dati in tabelle diverse

- Istanza di relazione Staff

CF	CODICE	COGNOME	NOME	RUOLO	ASSUNZIONE
RSI...	COD1	Rossi	Mario	Analista	1995
BA...	COD2	Bianchi	Pietro	Analista	1990
NRI...	COD3	Neri	Paolo	Amministratore	1985

- Chiavi?

- Secondo la definizione, è possibile che (Cognome, Ruolo) sia una chiave ?



- Vincolo di integrità referenziale (foreign key): porzioni di informazione in relazioni diverse sono correlate attraverso valori di chiave
- Se una relazione fa riferimento al valore di un attributo o di un insieme di attributi che dovrebbe comparire in una seconda relazione, dobbiamo assicurarci che ciò avvenga realmente
- Un vincolo di integrità referenziale fra gli attributi  $X$  di una relazione  $R_1$  e un'altra relazione  $R_2$  impone ai valori su  $X$  in  $R_1$  di comparire come valori della chiave primaria di  $R_2$

# Infrazioni



SAPIENZA

INFORMATICA

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

## Vigili

<u>Matricola</u>	Cognome	Nome
3987	Rossi	Luca
3295	Neri	Piero
9345	Neri	Mario
7543	Mori	Gino

# Infrazioni



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

Auto	<u>Prov</u>	<u>Numero</u>	Cognome	Nome
	MI	39548K	Rossi	Mario
	TO	E39548	Rossi	Mario
	PR	839548	Neri	Luca



- Vincoli di integrità referenziale fra:
  - l'attributo Vigile della relazione INFRAZIONI e l'attributo Matricola (chiave) della relazione VIGILI
  - gli attributi Prov e Numero di INFRAZIONI e gli attributi Prov e Numero (chiave) della relazione AUTO

## Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

## Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	E39548	Rossi	Mario
TO	F34268	Rossi	Mario
PR	839548	Neri	Luca

ATTENZIONE AI  
VINCOLI SU PIU'  
ATTRIBUTI



- Giocano un ruolo fondamentale nel concetto di “modello basato su valori”
- In presenza di valori nulli i vincoli possono essere resi meno restrittivi
- E' possibile definire „azioni“ compensative a seguito di violazioni
- Attenzione ai vincoli su più attributi



## Impiegati

Il vincolo di integrità referenziale non viene violato dal valore NULL

### Matricola **Cognome** **Progetto**

<b>34321</b>	<b>Rossi</b>	<b>IDEA</b>
<b>53524</b>	<b>Neri</b>	<b>XYZ</b>
<b>64521</b>	<b>Verdi</b>	<b>NULL</b>
<b>73032</b>	<b>Bianchi</b>	<b>IDEA</b>

## Progetti

<u><b>Codice</b></u>	<b>Inizio</b>	<b>Durata</b>	<b>Costo</b>
<b>IDEA</b>	<b>01/2000</b>	<b>36</b>	<b>200</b>
<b>XYZ</b>	<b>07/2001</b>	<b>24</b>	<b>120</b>
<b>BOH</b>	<b>09/2001</b>	<b>24</b>	<b>150</b>



- Non tutte le proprietà di interesse possono essere rappresentate tramite vincoli espliciti nel modello logico ...
- ... ma questo sarà trattato nel secondo modulo
- Come formalizziamo i vincoli?
- Vincoli intrarelazionali  $\supseteq$  relazioni tra valori di attributi della stessa tupla  $\rightarrow$  **dipendenze funzionali definite sullo schema**



- Una dipendenza funzionale stabilisce un particolare legame semantico tra due insiemi non-vuoti di attributi  $X$  e  $Y$  appartenenti ad uno schema  $R$
- Tale vincolo si scrive  $X \rightarrow Y$  e si legge  $X$  *determina*  $Y$

- Supponiamo di avere uno schema di relazione  
VOLI (CodiceVolo, Giorno, Pilota, Ora)
- con i vincoli ... dettati dal buon senso ...
  - Un volo con un certo codice parte sempre alla stessa ora
  - Esiste un solo volo con un dato pilota, in un dato giorno ad una data ora
  - C'è un solo pilota di un dato volo in un dato giorno.
- I vincoli corrispondono alle dipendenze funzionali
  - $\text{CodiceVolo} \rightarrow \text{Ora}$
  - $\{\text{Giorno}, \text{Pilota}, \text{Ora}\} \rightarrow \text{CodiceVolo}$
  - $\{\text{CodiceVolo}, \text{Giorno}\} \rightarrow \text{Pilota}$



- Diremo che una relazione  $r$  con schema  $R$  **soddisfa** la dipendenza funzionale  $X \rightarrow Y$  se
  - (i) la dipendenza funzionale  $X \rightarrow Y$  è *applicabile* ad  $R$ , nel senso che sia  $X$  sia il  $Y$  sono sottoinsiemi di  $R$ ;
  - (ii) le ennuple in  $r$  che concordano su  $X$  concordano anche su  $Y$ , cioè per ogni coppia di ennuple  $t1$  e  $t2$  in  $r$ 
$$t1[X] = t2[X] \Rightarrow t1[Y] = t2[Y].$$

**Nota importante:** l'implicazione significa che se le tuple sono uguali su  $X$ , allora devono essere uguali anche su  $Y$

- (1)  $test := \text{VERO}$ .
- (2) Con il criterio lessicografico ordinare le ennuple in  $r$  usando  $X$  come base dell'ordinamento. Sia  $L$  la lista risultante.
- (3) Per ogni coppia  $t_1$  e  $t_2$  di elementi consecutivi della lista  $L$ ,  
  
se  $t_1[X] = t_2[X]$  e  $t_1[Y] \neq t_2[Y]$ , allora  $test := \text{FALSO}$  ed Uscire.

- Caso particolare: se  $X \cup Y = R$  possiamo invece usare un passo 3 alternativo

per ogni coppia  $t_1$  e  $t_2$  di elementi consecutivi della lista  $L$ , se  $t_1[X] = t_2[X]$ , allora  $test := \text{FALSO}$  ed Uscire.

**Nota:** una relazione è un **insieme** di ennuple