

Basi di dati

Giuseppe De Giacomo

***Dipartimento di Informatica e Sistemistica "Antonio Ruberti"
Università di Roma "La Sapienza"***

Anno Accademico 2006/07
Canale M-Z

<http://www.dis.uniroma1.it/~degiacomo/didattica/basidati/>

4. La progettazione concettuale

4.1 introduzione alla progettazione di basi di dati

- 1. introduzione alla progettazione di basi di dati**
2. modello ER
3. metodologia per la progettazione concettuale
4. esempio

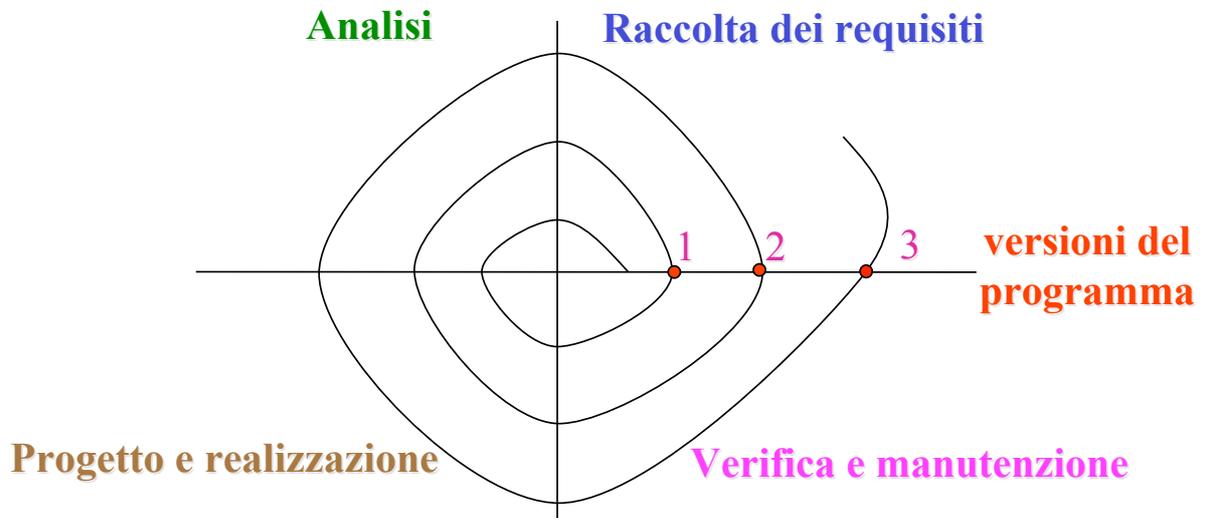
Inquadramento del progetto di basi di dati

- Il progetto di basi di dati ha lo scopo di produrre lo schema logico e lo schema fisico della basi di dati, secondo le indicazioni raccolte durante la fase di raccolta dei requisiti
- **Due scenari tipici:**
 1. il progetto di basi di dati viene svolto **nell'ambito della più ampia attività di progetto** del sistema informatico
 2. il progetto di basi di dati viene svolto come un **processo a sé stante**

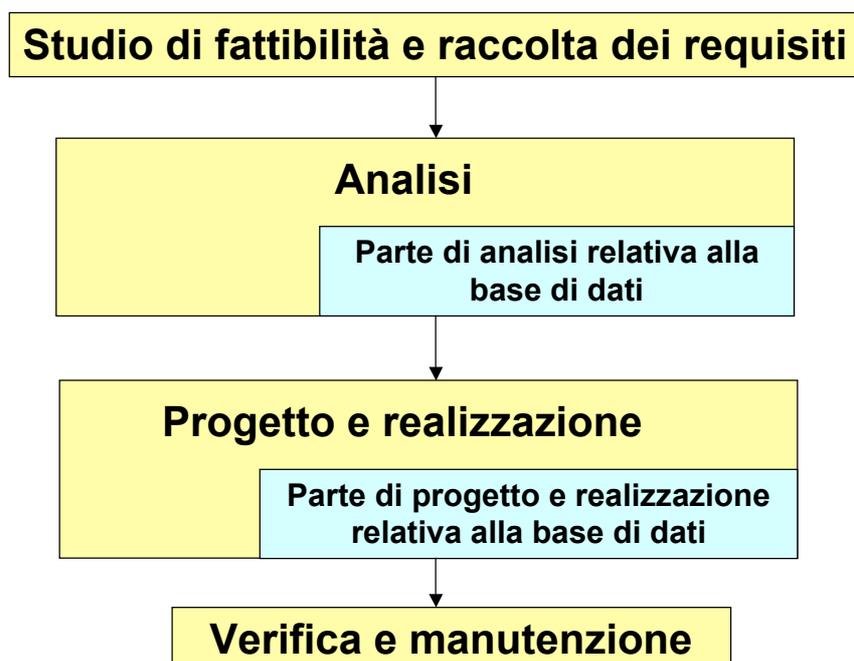
Ciclo di vita di un'applicazione software

1. Studio di fattibilità e raccolta dei requisiti
 - valutare costi e benefici
 - pianificare le attività e le risorse del progetto
 - raccogliere i **requisiti**
2. Analisi
 - si occupa del **cosa**
 - descrivere il dominio dell'applicazione e specificare le funzioni delle varie componenti: lo **schema concettuale**
3. Progetto e realizzazione
 - si occupa del **come**
 - definire l'architettura del programma
 - scrivere il **codice del programma** e produrre la **documentazione**
4. Verifica
 - Il programma svolge correttamente, completamente, efficientemente il compito per cui è stato sviluppato?
5. Manutenzione
 - Correzione e aggiornamento del programma

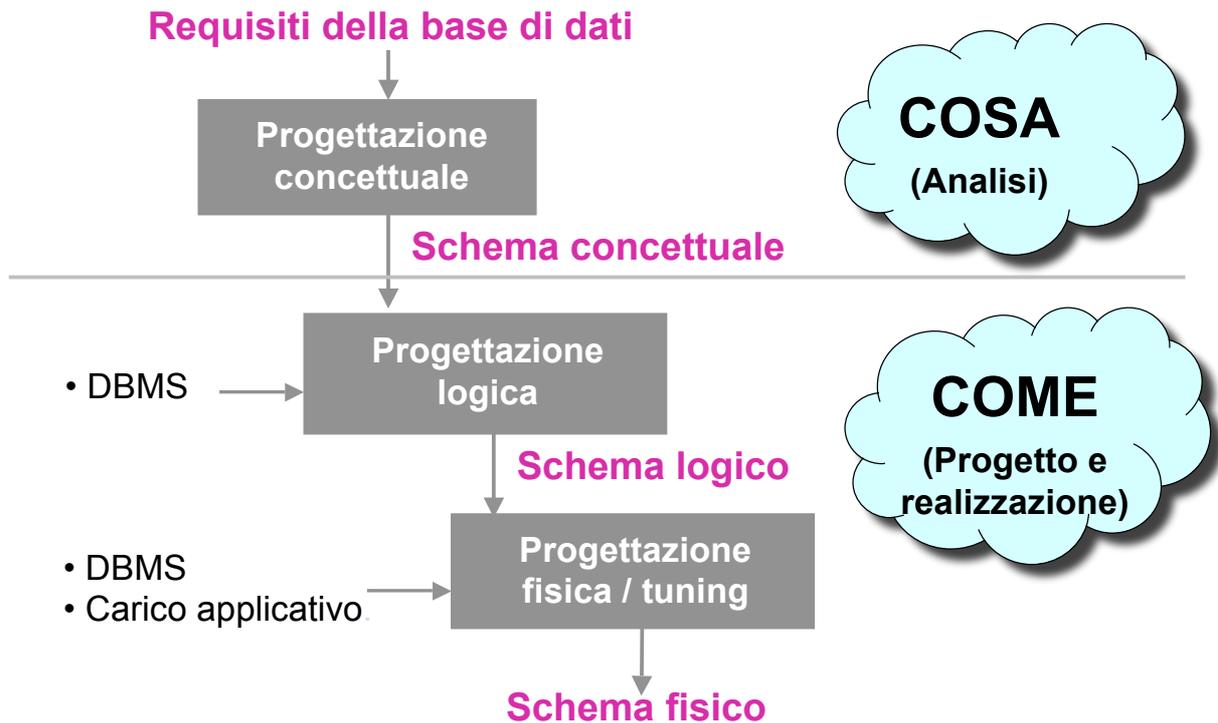
Ciclo di vita (modello a spirale)



Scenario 1: progetto complessivo del sistema informatico



Scenario 2: Progetto della basi di dati



Confronto tra i due scenari

- In entrambi gli scenari, il processo di progettazione di basi di dati si compone delle seguenti fasi:
 1. Progettazione concettuale
 2. Progettazione logica
 3. Progettazione fisica / tuning
- Le **differenze tra i due scenari** sono:
 - Nel **primo scenario** la **progettazione concettuale** è **parte della fase complessiva di analisi** e si deve coordinare con gli altri aspetti, mentre nel **secondo** è una **fase autonoma**
 - Nel **primo scenario** le fasi di **progettazione logica** e **progettazione fisica / tuning** sono **parti della fase complessiva di progetto e realizzazione**, e si devono coordinare con gli altri aspetti, mentre nel **secondo** sono **fasi autonome**

Progettazione concettuale

Rappresentare i dati della realtà d'interesse in termini di un modello (descrizione) formale, ad alto livello, indipendente dal DBMS

- **Input:**
 - Descrizione della realtà d'interesse: *(per sapere quali dati rappresentare)*
- **Output:**
 - **Schema concettuale** *(rappresentazione dei dati in un modello di dati concettuale, eventualmente con vincoli esterni)*
- **Qualità:**
 - correttezza e completezza della rappresentazione

Progettazione logica

- Rappresentare i dati della realtà d'interesse in termini dei costrutti logici di una classe di DBMS (quelli basati su un determinato modello logico)
- **Input:**
 - Schema concettuale *(prodotto dalla progettazione concettuale)*
 - DBMS scelto *(per sapere quale modello logico di dati adottare)*
 - Previsioni del carico applicativo *(per ottimizzare la base di dati rispetto ad esso)*
- **Output:**
 - **Schema logico** *(rappresentazione dei dati in un modello di dati logico, eventualmente con vincoli complessi)*
- **Qualità:**
 - Correttezza e completezza della rappresentazione
 - Efficienza della manipolazione dei dati

Progettazione fisica / tuning

Rappresentare i dati della realtà di interesse attraverso le strutture dati di uno specifico DBMS

- **Input:**

- Schema logico (*prodotto della progettazione logica*)
- DBMS scelto (*per sapere quali strutture dati fisiche utilizzare*)
- Previsioni del carico applicativo (*per ottimizzare la base di dati rispetto ad esso*)

- **Output:**

- **Schema fisico** (*Tipicamente nei DBMS attuali consiste in uno schema logico + alcune scelte relative all'effettiva implementazione su DBMS di detto schema*)

- **Qualità:**

- Efficienza rispetto al carico applicativo

Compromesso tra correttezza/completezza ed efficienza

- Si noti che passando dallo schema concettuale allo schema logico e poi allo schema fisico, stiamo **degradando la correttezza/completezza** della rappresentazione strutturale, degrado eventualmente compensato da opportuni vincoli (*qualità semantiche*)
- D'altra parte in questi passaggi poniamo sempre **maggiore attenzione all'efficienza** (*qualità computazionali*)

Correttezza e completezza del sistema

- Si noti che la correttezza (adeguatezza e completezza rispetto alle specifiche) del sistema **basi di dati + applicazioni** è **irrinunciabile**
- Quindi, se la nostra rappresentazione dei dati **non è completa** (come è tipicamente il caso per lo schema logico e lo schema fisico) **oppure non garantisce la correttezza in tutti i casi** (come può accadere per lo schema fisico), allora saranno le **applicazioni** a garantire di manipolare i dati in modo da preservare correttezza e completezza

Modelli di dati

- I prodotti della varie fasi sono schemi basati su specifici **modelli di dati**:
 - Schema concettuale → **modello concettuale**
 - Schema logico → **modello logico (relazionale)**
 - Schema fisico → **modello logico + parametri fisici**

Modelli dei dati adottati

- Schema concettuale:
 - Modello Entità-Relazione (ER)

- Schema logico:
 - Modello relazionale + vincoli d'integrità
 - DDL: SQL standard

- Schema fisico:
 - Modello relazionale + strutture fisiche
 - DDL: SQL supportato dal DBMS scelto

4. La progettazione concettuale

4.2 modello ER

1. introduzione alla progettazione di basi di dati
- 2. modello ER**
3. metodologia per la progettazione concettuale
4. esempio

Il modello Entità-Relazione (ER) è un modello di dati

Un **modello (concettuale) di dati** è caratterizzato dai costrutti ammessi. Ogni costrutto ha un impatto sia a livello intensionale (**schema concettuale**), sia a livello estensionale (**istanze** dello schema). Per definire precisamente un modello di dati, occorre specificare per ogni costrutto

- il suo significato intuitivo, che determina il modo con cui si utilizza (**pragmatica**)
- la forma che assume a livello intensionale (**sintassi**)
- l'impatto sul livello estensionale (**semantica**)

Nel modello ER la struttura dello schema concettuale è descritta in forma grafica (diagramma dello schema)

Differenza tra schemi e istanze nei modelli di dati

- Lo **schema** descrive la struttura, cioè l'aspetto **intensionale** (ad esempio, nel modello relazionale, l'intestazione delle tabelle)
- Ad ogni schema corrispondono più **istanze** (aspetto **estensionale**), anche se, istante per istante, solo una è quella significativa (ad esempio, nel modello relazionale l'istanza corrente è formata dalle tuple che popolano le tabelle)

Costrutti fondamentali del modello Entità-Relazione

- Entità
- Attributi di entità
- Relazioni
- Ruoli
- Attributi di relazione
- IS-A e Generalizzazioni
- Vincoli di identificazione
- Vincoli di cardinalità
- Altri vincoli

ER e UML

I costrutti del modello ER sono analoghi a quelli usati in altri modelli concettuali, ad esempio i diagrammi delle classi di UML

ER	UML
<ul style="list-style-type: none">• Entità• Relazioni• Attributi• Cardinalità• IS-A e Generalizzazioni• Identificatori• -----	<ul style="list-style-type: none">• Classi• Associazioni/aggregazioni• Attributi• Molteplicità• IS-A e Generalizzazioni• -----• Operazioni

Il modello ER è specializzato alla rappresentazione dei dati

La nozione di entità

Una entità è una **classe di oggetti** (fatti, persone, cose) che sono di interesse per l'applicazione, che hanno esistenza autonoma, e che hanno proprietà comuni

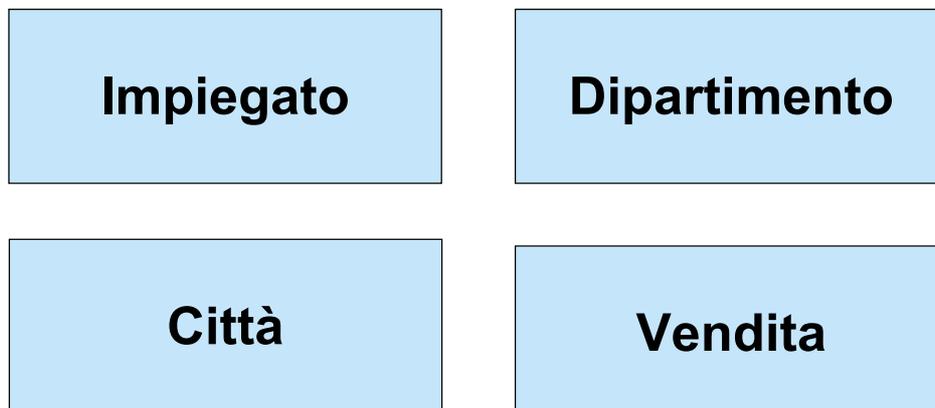
Esempi:

- ▶ **impiegato**
- ▶ **vendita**
- ▶ **dipartimento**
- ▶ **ordine**
- ▶ **città**
- ▶ **studente**

Sintassi: rappresentazione grafica di entità

Ogni **entità** ha nome che la identifica in modo univoco nello schema, ed è rappresentata da un rettangolo nel diagramma che descrive lo schema stesso

Esempi:



Semantica delle entità

- A livello **estensionale** un'entità è costituita da un insieme di oggetti, che sono chiamati le sue **istanze**
- Ciò significa che, se in uno schema **S** è definita una entità **E**, in ogni istanza **I** dello schema **S**, alla entità **E** è associato un insieme di oggetti (che viene denotato *istanze(I,E)* o semplicemente *istanze(E)*)
 $\{e_1, e_2, e_3, \dots\}$
che viene detto anche l'estensione di **E** nella istanza **I** di **S**
- Una istanza di entità non è un valore che identifica un oggetto, ma è l'oggetto stesso

Entità e istanze d'entità

- Una entità (nello schema concettuale **S**) denota un insieme (o classe) di oggetti "simili"
- Una istanza dell'entità (in una istanza di **S**) denota un oggetto dell'insieme rappresentato dall'entità

Nota: nello schema concettuale rappresentiamo le entità, non le singole istanze ("astrazione")

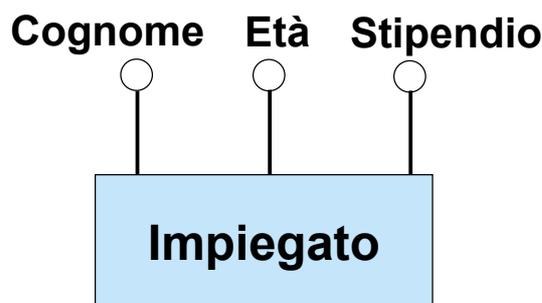
La nozione di attributo di entità

- Un attributo di entità è una proprietà locale di un'entità, di interesse ai fini dell'applicazione
- Un attributo associa ad ogni istanza di entità un valore appartenente ad un insieme detto **dominio** dell'attributo (tipicamente, interi, caratteri, stringhe, ecc.)
- Si definisce un attributo per l'entità E quando si vuole rappresentare una proprietà **locale** delle istanze dell'entità E. Una proprietà di un oggetto si dice locale quando in ogni istanza dello schema il valore di tale proprietà dipende **solamente** dall'oggetto stesso, e **non ha alcun rapporto** con altri elementi dell'istanza dello schema

Sintassi: rappresentazione grafica di attributo

Ogni attributo di entità ha nome che lo identifica in modo univoco nell'ambito della entità, ed è rappresentato da un cerchio collegato alla entità a cui appartiene.

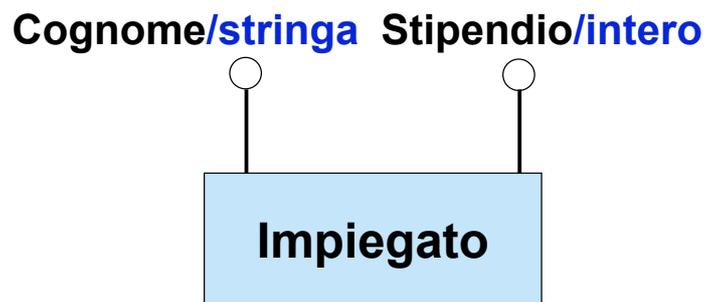
Esempi:



Sintassi: rappresentazione grafica di dominio

Il dominio associato ad un attributo viene generalmente tralasciato nella rappresentazione grafica (ma si indica nel dizionario dei dati – si veda in seguito).

Se si indica nel diagramma, la notazione è:



Semantica degli attributi di entità

Se I è una istanza di uno schema S dove A è un attributo dell'entità E su un dominio D , $istanze(I,A)$, o semplicemente $istanze(A)$, è un insieme di coppie (x,y) tali che x è in $istanze(E)$, y è in D , ed esiste una ed una sola coppia in $istanze(A)$ per ogni x in $istanze(E)$. In altre parole, a livello **estensionale** l'attributo A è una funzione totale

$$A : istanze(E) \rightarrow D$$

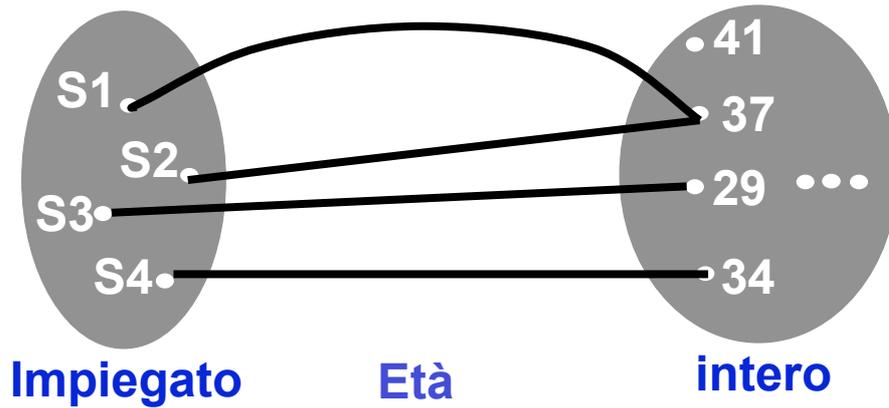
cioè, in ogni istanza I è una istanza di S , l'attributo A è una funzione che associa ad ogni istanza di E un valore appartenente a D .

Esercizio 1: attributi di entità

Dato lo schema:



La seguente è una sua istanza ?

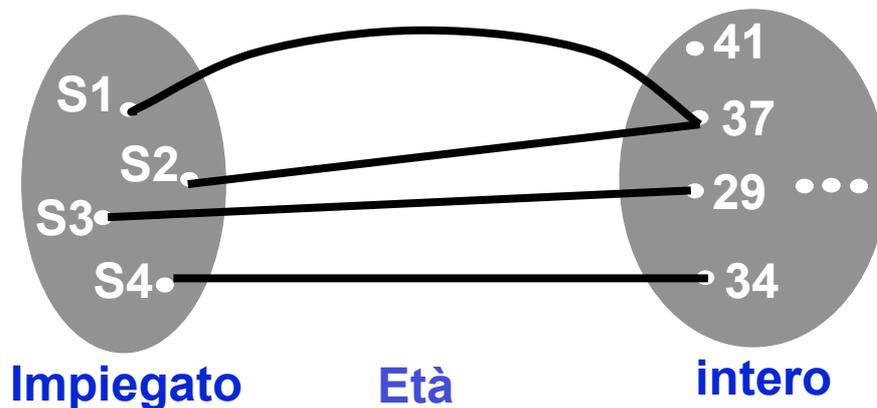


Esercizio 1: soluzione

Dato lo schema:



La seguente è una sua istanza **Sì!** ?

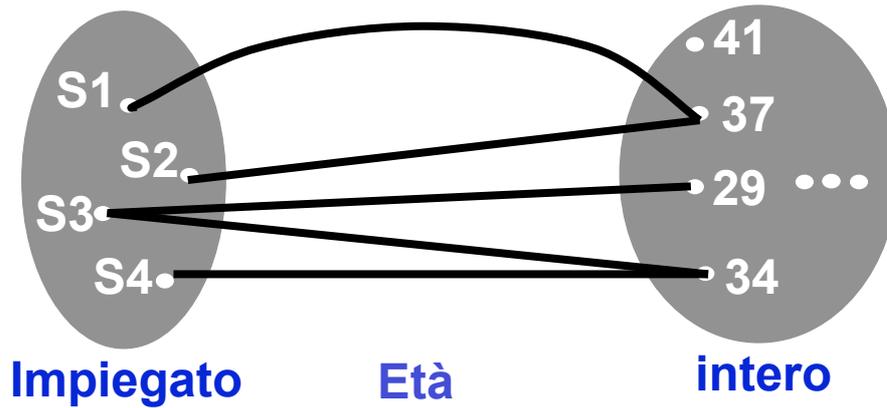


Esercizio 2: attributi di entità

Dato lo schema:



La seguente è una sua istanza ?



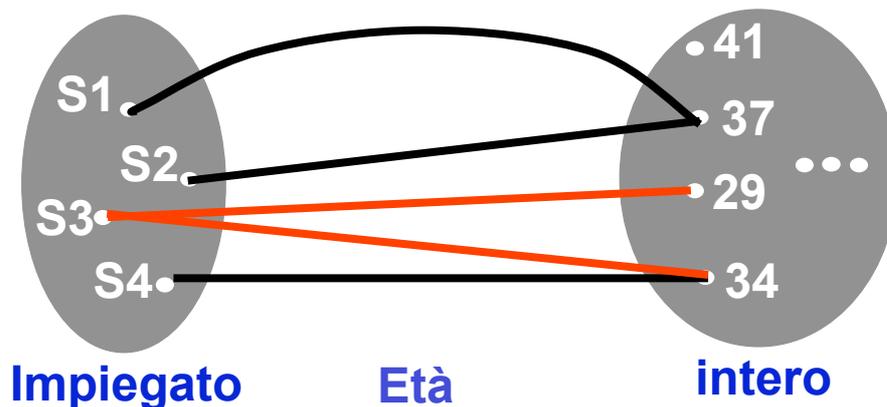
Esercizio 2: soluzione

Dato lo schema:



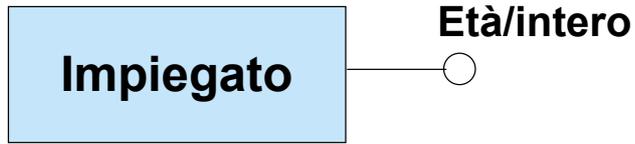
La seguente è una sua istanza ?

No, l'attributo Età associa due valori a S3!
Ricordarsi che un attributo è una funzione.

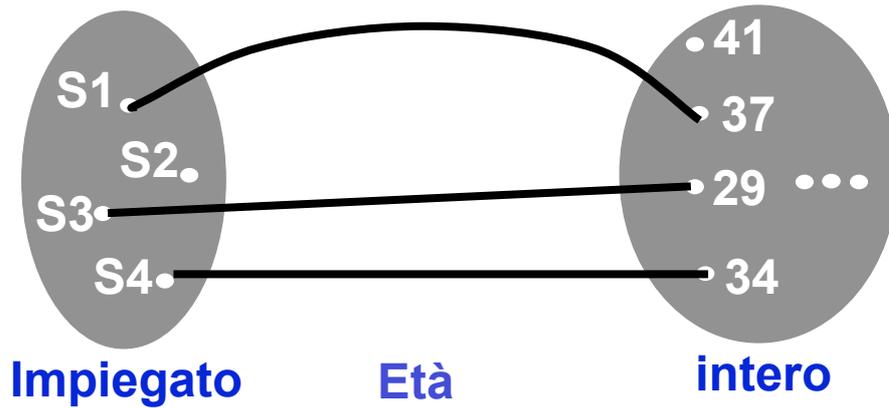


Esercizio 3: attributi di entità

Dato lo schema:



La seguente è una sua istanza ?



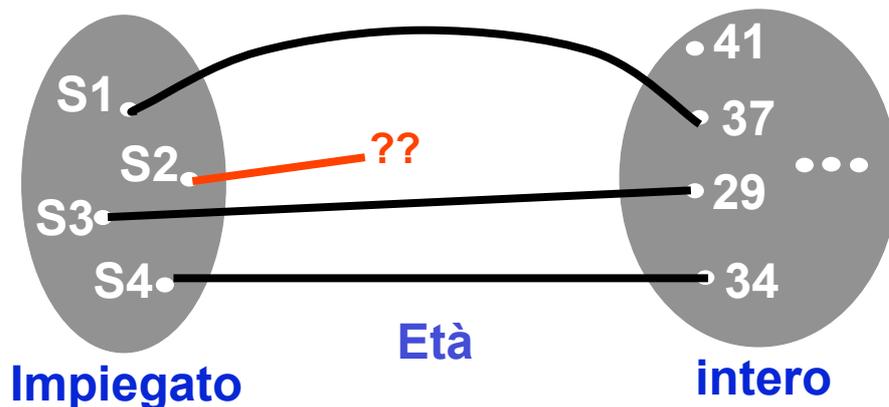
Esercizio 3: soluzione

Dato lo schema:



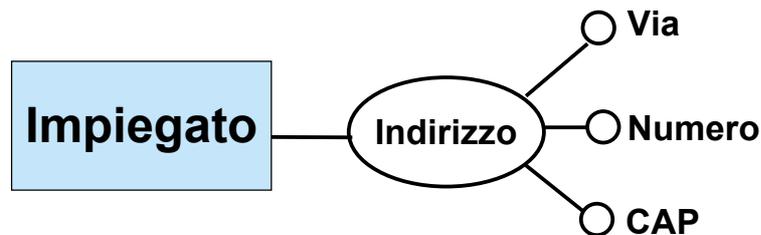
La seguente è una sua istanza ?

No, manca il valore di età per S2! Ricordarsi che un attributo è una funzione totale



Attributi composti

- Un attributo può anche essere definito su un dominio complesso. Di particolare interesse è il caso di dominio di tipo “record”. Un attributo il cui dominio è il tipo record si dice **composto**.
- Notazione:



“Indirizzo” è un attributo composto, e il suo dominio è un tipo record con campi “Via”, “Numero” e “CAP”.

La nozione di relazione

Una **relazione** (o associazione) si definisce su due o più entità, e rappresenta un **legame** fra tali entità. Il numero di entità coinvolte in una relazione determina il suo **grado**

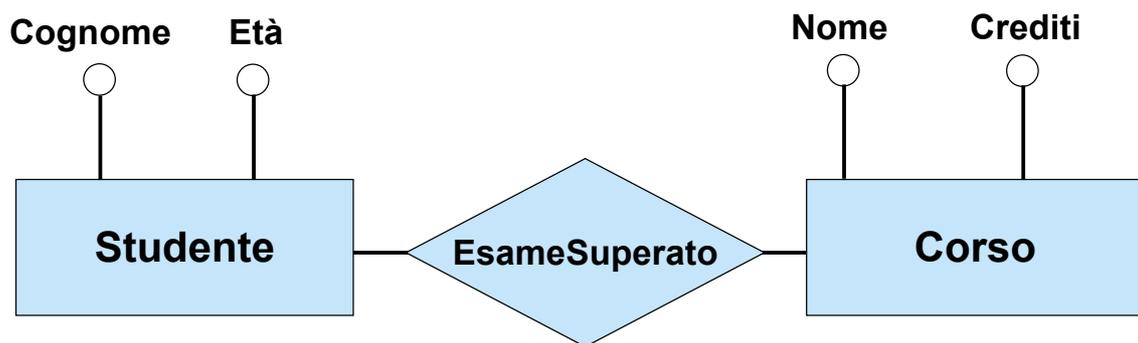
Esempi

- ▶ **Residenza** (tra Persona e Città)
- ▶ **Afferenza** (tra Impiegato e Dipartimento)

Sintassi: rappresentazione grafica di relazione

Ogni **relazione** ha un nome che la identifica in modo univoco nello schema, ed è rappresentata nel diagramma che descrive lo schema da un rombo collegato alle entità sulle quali è definita la relazione

Esempio di relazione di grado 2 (relazione binaria):



Semantica delle relazioni binarie (prima versione)

- A livello **estensionale** una relazione **R** tra le entità **E** ed **F** è costituita da un insieme di coppie **(x,y)**, tali che **x** è una istanza di **E**, ed **y** è una istanza di **F**. Ogni coppia è detta **istanza** della relazione **R**
- Ciò significa che, se in uno schema **S** è definita una relazione **R** sulle entità **E** ed **F**, in ogni istanza **I** dello schema **S**, alla relazione **R** è associato un insieme di coppie (denotato da *istanze(I,R)* o semplicemente *istanze(R)*)

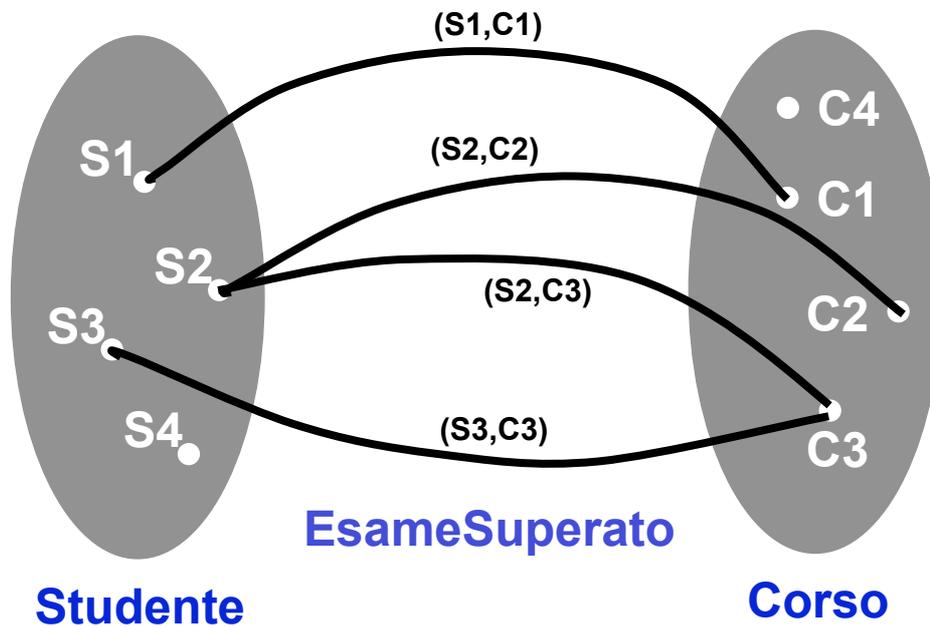
$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots\}$$

che viene detto anche l'estensione di **R** nella istanza **I** di **S**

- In altre parole, una relazione nel modello ER è, dal punto di vista della semantica, una relazione matematica. In ogni istanza dello schema si ha:

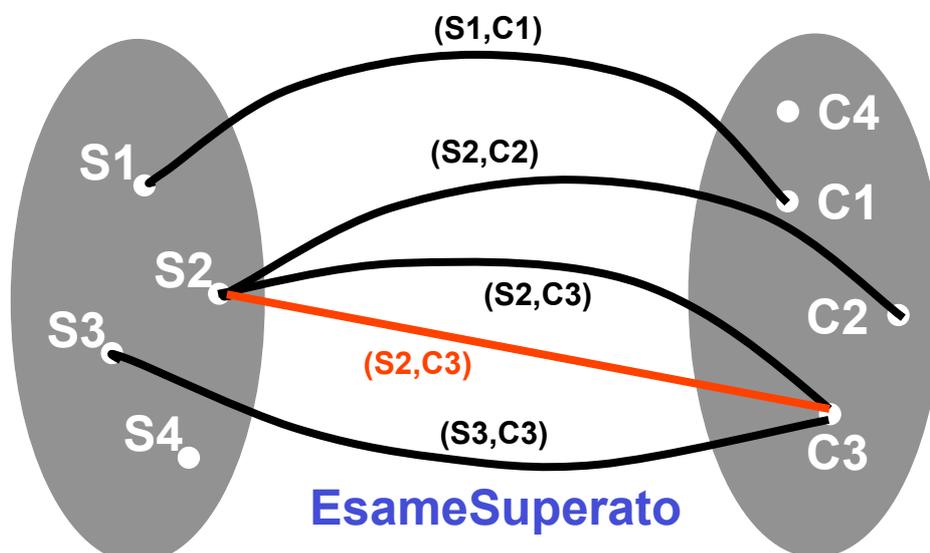
$$istanze(I,R) \subseteq istanze(I,E) \times istanze(I,F)$$

Relazione a livello estensionale



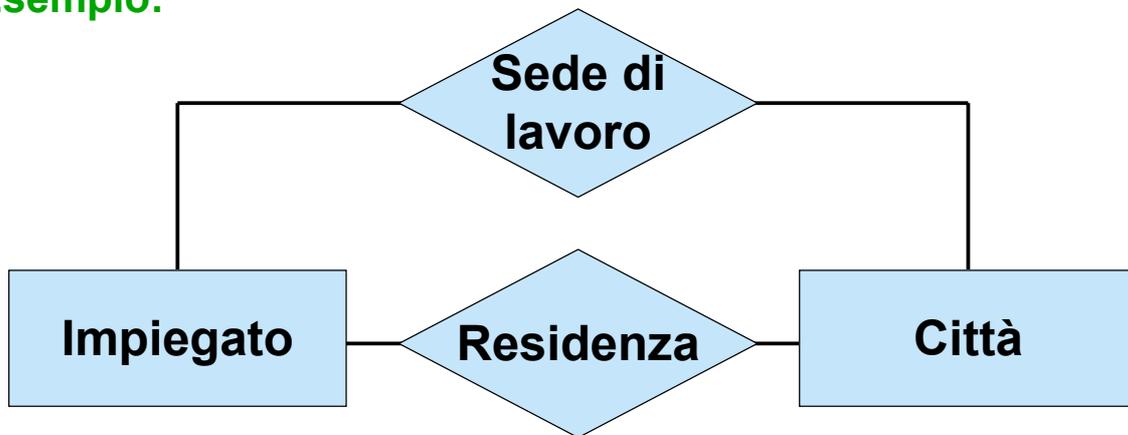
Osservazione importante

Dalla semantica delle relazioni segue immediatamente che **non** possono esistere due istanze della stessa relazione che coinvolgono le stesse istanze di entità.



Due entità possono essere coinvolte in più relazioni

Esempio:

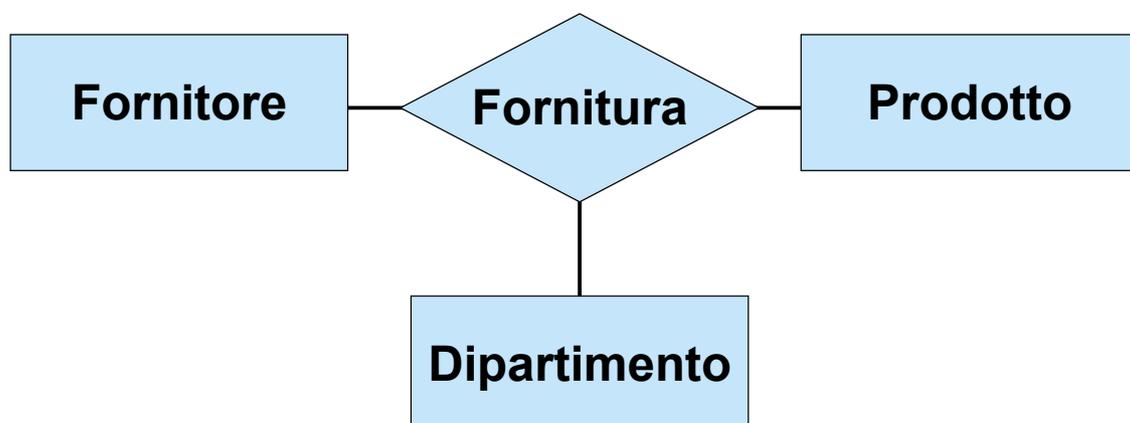


“Sede di Lavoro” e “Residenza” rappresentano due legami diversi tra le stesse entità “Impiegato” e “Città”

Relazioni n-arie (grado maggiore di 2)

Una relazione di grado maggiore di 2 si dice n-aria.

Esempio:

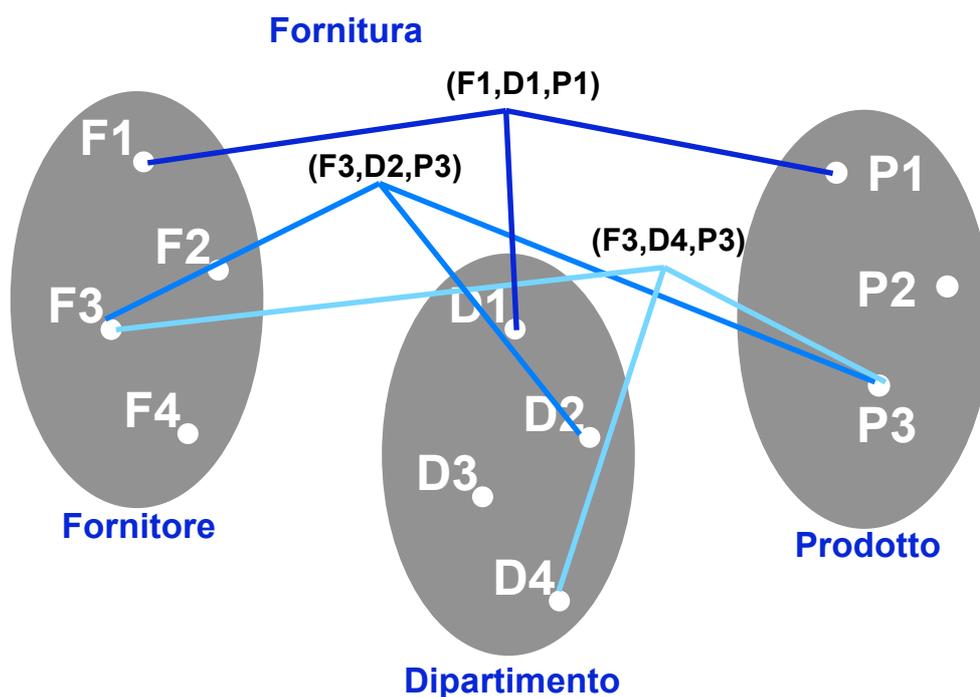


Semantica delle relazioni n-arie (prima versione)

- A livello **estensionale** una relazione **R** tra le entità **E₁, E₂, ..., E_n** è costituita da un insieme di n-ple (o tuple) **(x₁, x₂, ..., x_n)**, tali che **x₁** è una istanza di **E₁**, **x₂** è una istanza di **E₂**, ..., **x_n** è una istanza di **E_n**. Ogni n-ple è detta **istanza** della relazione **R**
- Quindi, in ogni istanza **I** dello schema si ha:

$$istanze(I, R) \subseteq istanze(I, E_1) \times \dots \times istanze(I, E_n)$$

Relazione n-aria a livello estensionale



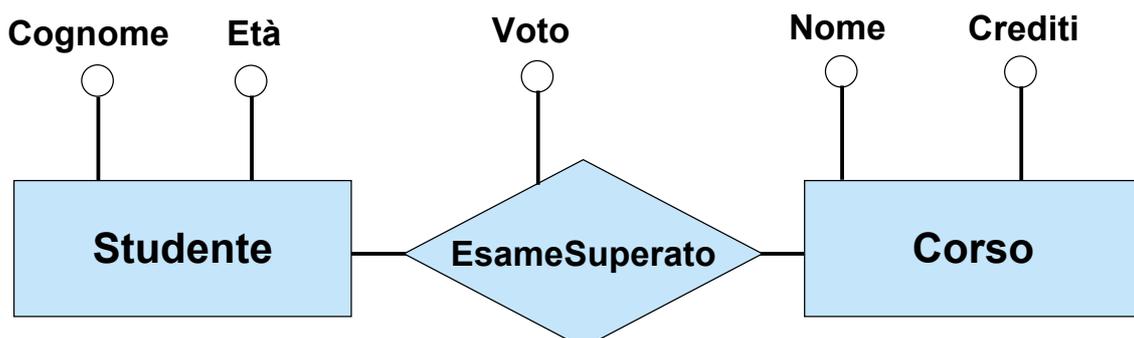
La nozione di attributo di relazione

- Un attributo di relazione è una proprietà locale di una relazione, di interesse ai fini dell'applicazione
- Un attributo della relazione R tra le entità E_1, E_2, \dots, E_n modella una proprietà **non** di E_1 , **non** di E_2, \dots , **non** di E_n , ma del **legame** tra E_1, E_2, \dots, E_n rappresentato da R
- Un attributo associa ad ogni istanza di relazione un valore appartenente ad un insieme detto **dominio** dell'attributo

Sintassi: rappresentazione grafica di attributo di relazione

Ogni attributo di relazione ha un nome che lo identifica in modo univoco nell'ambito della relazione, ed è rappresentato da un cerchio collegato alla relazione a cui appartiene.

Esempio



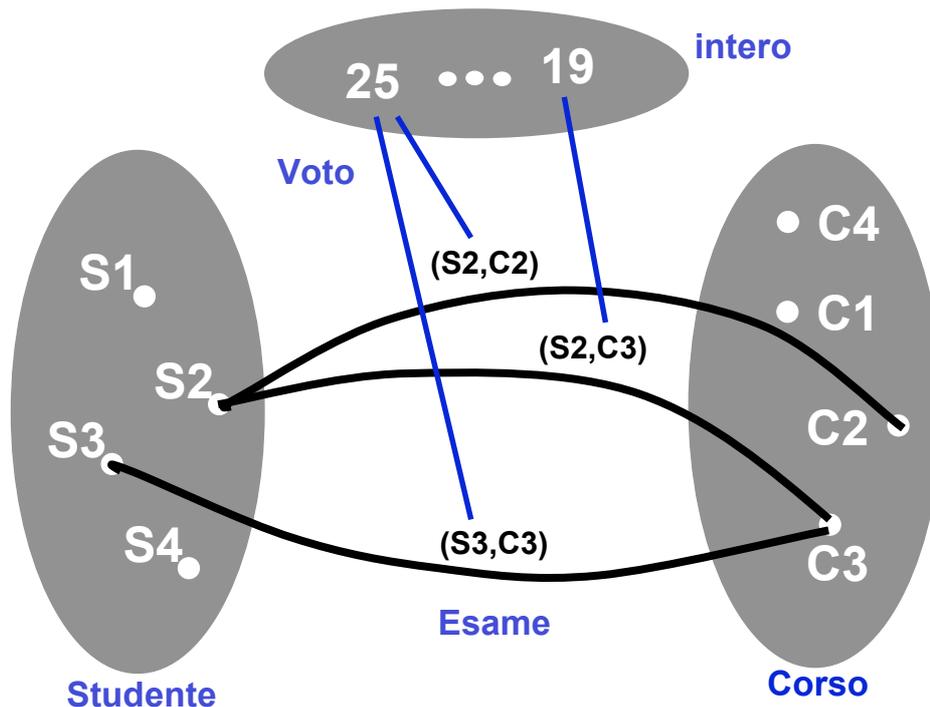
Semantica degli attributi di relazione

Se I è una istanza di uno schema S dove A è un attributo della relazione R su un dominio D , $istanze(I,A)$ è un insieme di coppie (x,y) tali che x è in $istanze(I,R)$, y è in D , ed esiste una ed una sola coppia in $istanze(I,A)$ per ogni x in $istanze(I,R)$. In altre parole, a livello **estensionale** l'attributo A è una funzione totale

$$A : istanze(R) \rightarrow D$$

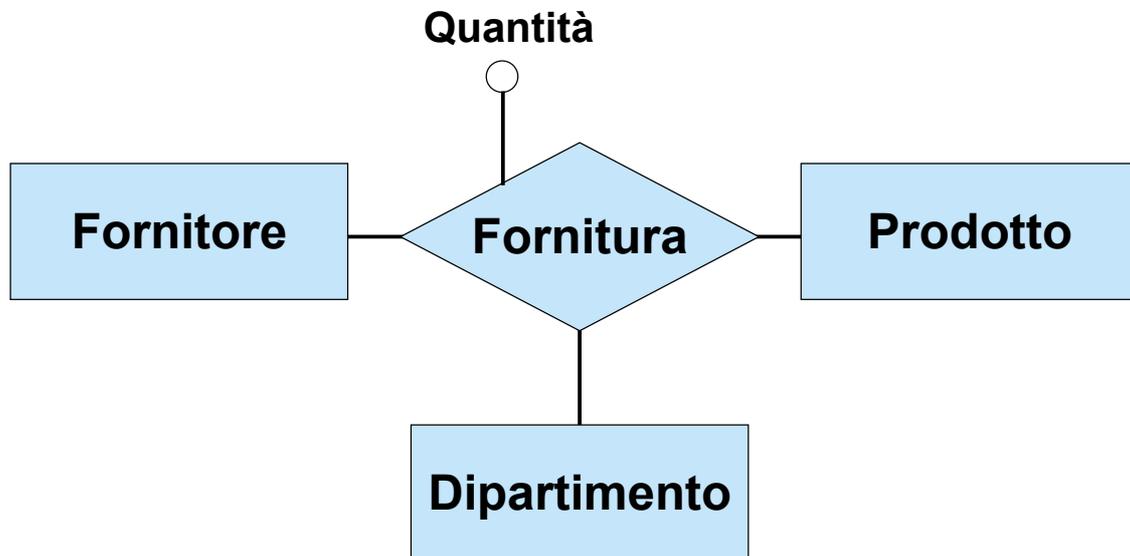
cioè, in ogni istanza I è una istanza di S , l'attributo A è una funzione che associa ad **ogni** istanza di R un valore appartenente a D .

Attributi di relazione binaria: livello estensionale

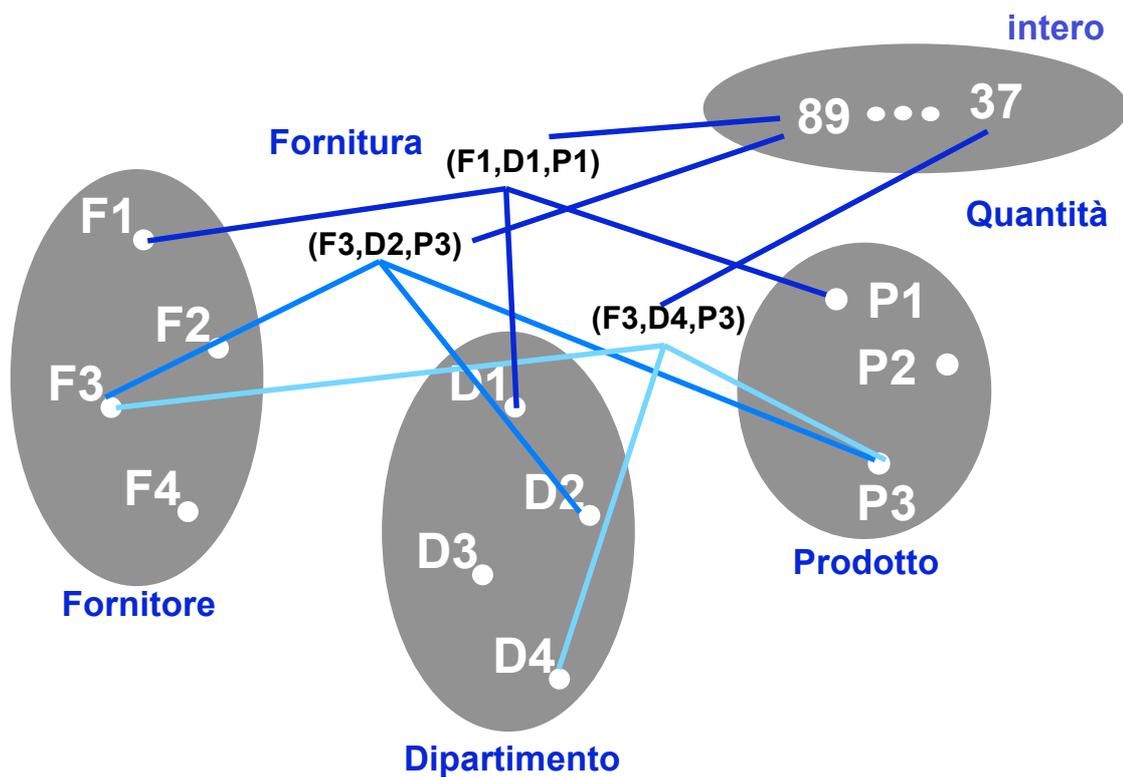


Relazioni n-arie con attributi

Esempio di relazione n-aria con attributo



Attributi di relazione n-aria: livello estensionale

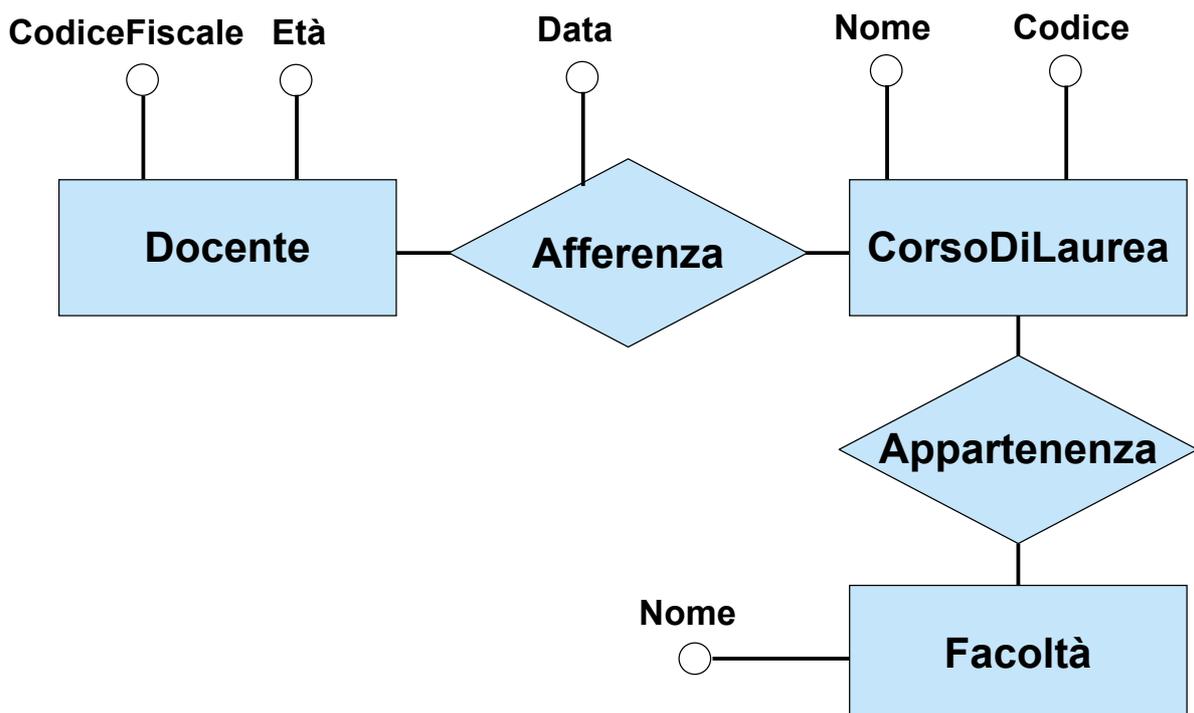


Esercizio 4: un semplice schema concettuale

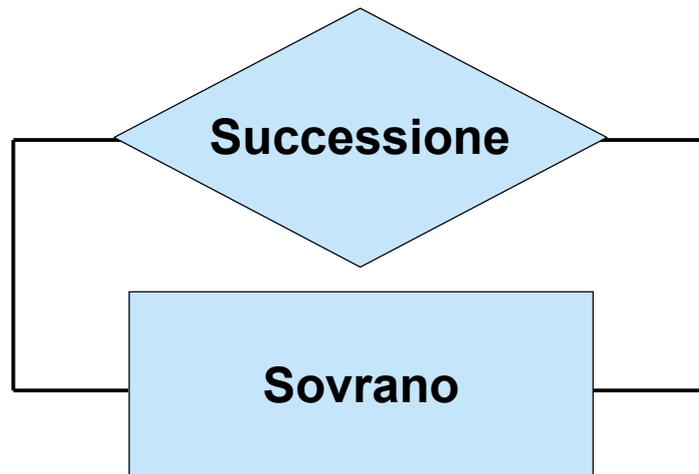
Descrivere lo schema concettuale della seguente realtà:

I docenti hanno un codice fiscale ed una età. I docenti operano nei corsi di laurea (si dice che afferiscono ai corsi di laurea). Interessa la data di afferenza dei docenti ai corsi di laurea. I corsi di laurea hanno un codice ed un nome, ed appartengono alle facoltà. Ogni facoltà ha un nome.

Esercizio 4: soluzione



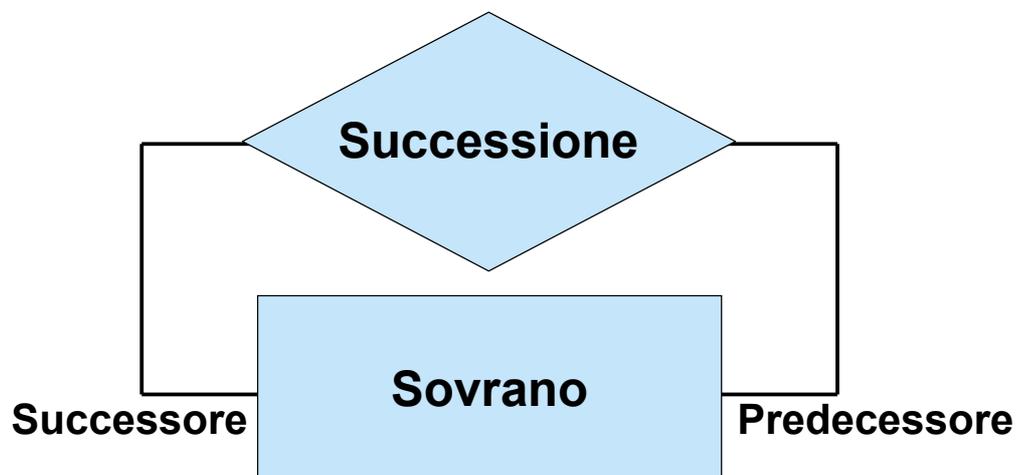
Una relazione può coinvolgere due o più volte la stessa entità



Problema: in una istanza di questo schema, data una coppia che è istanza di "Successione", non si può individuare chi è il sovrano predecessore e chi il sovrano successore.

Relazioni sulla stessa entità

Nelle relazioni dove una stessa entità è coinvolta più volte è necessario aggiungere la specifica dei "ruoli"



La nozione di ruolo

- È possibile aggiungere ad una relazione l'informazione che specifica il **ruolo** che una entità gioca nella relazione
- Se per una relazione **R** che coinvolge l'entità **E** (e altre entità) si stabilisce che il ruolo giocato da **E** in **R** è **U**, questo significa che in ogni istanza dello schema, per ogni tupla istanza della relazione **R** che coinvolge l'istanza **e** di **E**, il ruolo "giocato" da **e** nella tupla è **U**

Sintassi: rappresentazione grafica di ruolo

- La specifica del ruolo è necessaria nel caso di relazioni che coinvolgono la stessa entità più volte, mentre non sarebbe necessaria negli altri casi. Tuttavia, per uniformità, **noi richiediamo che, per ogni relazione R dello schema, per ogni entità E coinvolta in R (e, se E è coinvolta in R più di una volta, per ogni volta in cui è coinvolta), si specifichi il ruolo giocato da E in R**
- Graficamente, per ogni collegamento tra una relazione ed una entità coinvolta in essa, noi **specificheremo il nome del ruolo vicino al collegamento stesso**
- Se l'esplicita specifica del nome del ruolo manca per un collegamento tra R ed E, allora **si assume che il nome del ruolo coincida con il nome della entità E**
- Ovviamente, ogni collegamento tra R ed una entità coinvolta in R deve essere identificata univocamente dal nome di ruolo

Semantica dei ruoli e delle relazioni (versione finale)

- L'impatto dei ruoli sul livello **estensionale** è tale da modificare la semantica delle relazioni
- A livello **estensionale** una relazione **R** tra le entità E_1, E_2, \dots, E_n (non necessariamente tutte distinte) con rispettivi ruoli U_1, U_2, \dots, U_n (tutti distinti) è costituita da un insieme di **n-ple (o tuple) etichettate**

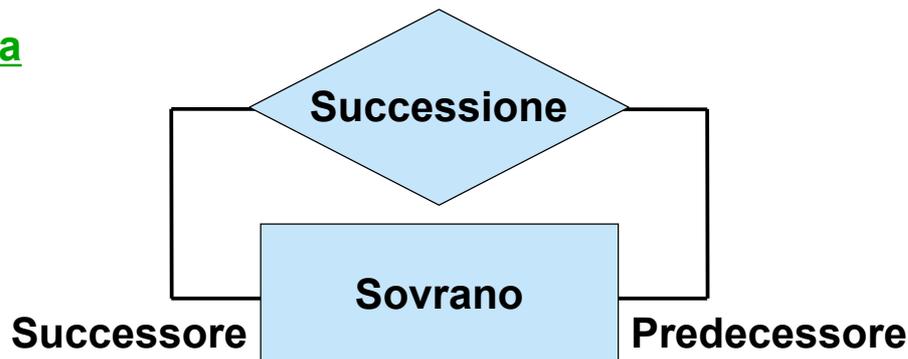
$$(U_1:x_1, U_2:x_2, \dots, U_n:x_n)$$

tali che x_1 è una istanza di E_1 , x_2 è una istanza di E_2, \dots, x_n è una istanza di E_n . Ogni n-ple etichettata è detta **istanza** della relazione **R**

Nelle slide, quando ciò non genera ambiguità, ometteremo, per ragioni di spazio, le etichette nelle tuple etichettate

Semantica dei ruoli: esempio

Schema



Istanza dello schema

$istanze(\text{Sovrano}) = \{ \text{romolo}, \text{numa}, \text{tullo}, \text{anco} \}$

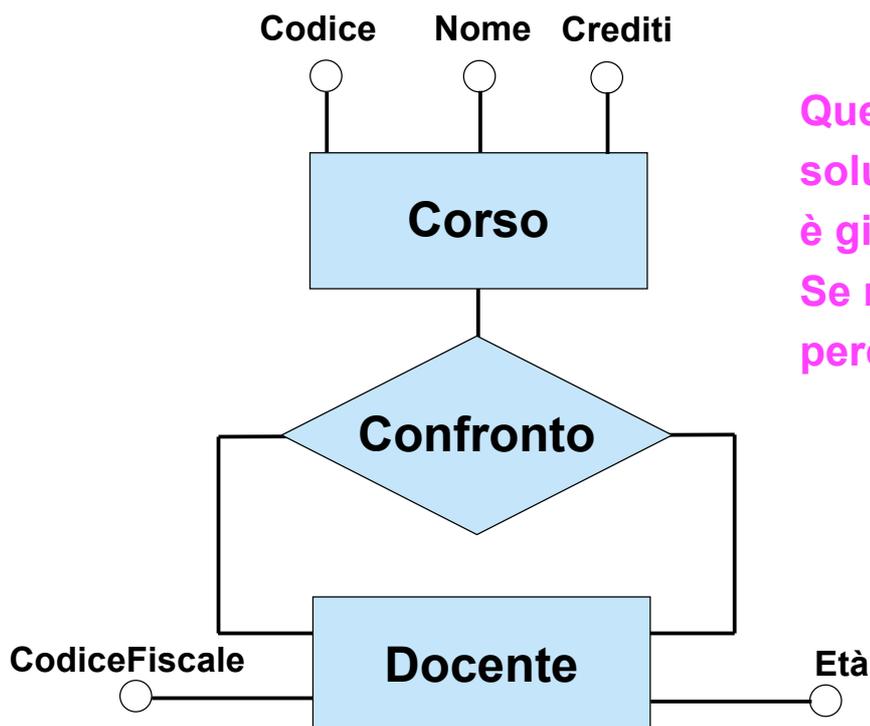
$istanze(\text{Successione}) = \{ (\text{Predecessore:romolo}, \text{Successore:numa}),$
 $(\text{Predecessore:numa}, \text{Successore:tullo}),$
 $(\text{Predecessore:tullo}, \text{Successore:anco})$
 $\}$

Esercizio 5: schema concettuale

Descrivere lo schema concettuale della seguente realtà:

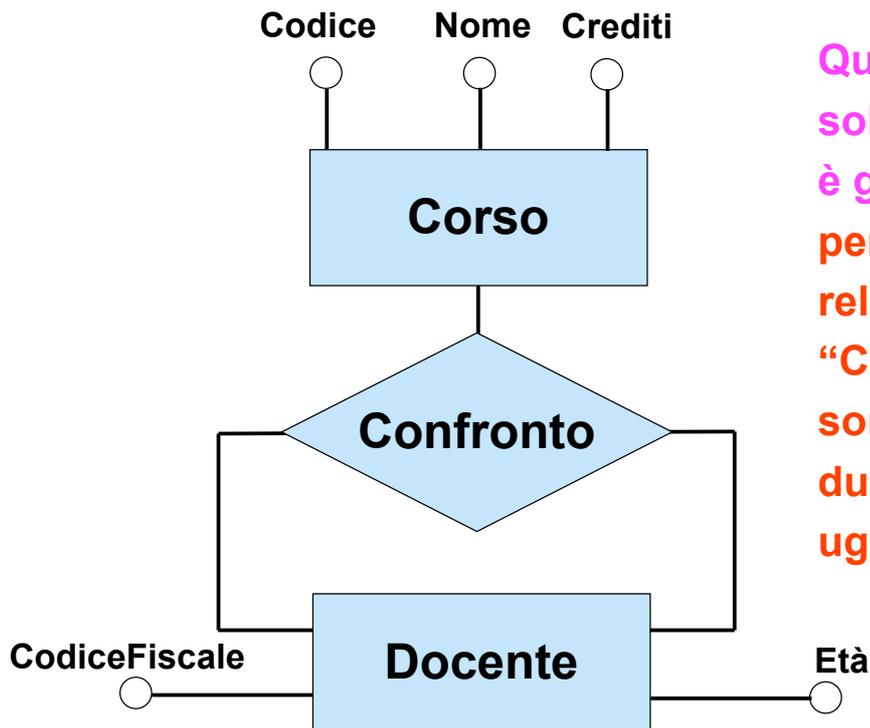
I docenti hanno un codice fiscale ed una età. I corsi hanno un codice identificativo, un nome, ed il numero di crediti. Sfruttando i moduli di valutazione dei corsi e dei docenti da parte degli studenti, si vuole rappresentare nella base di dati l'informazione se un docente è migliore di un altro nell'insegnare un corso.

Esercizio 5: soluzione 1



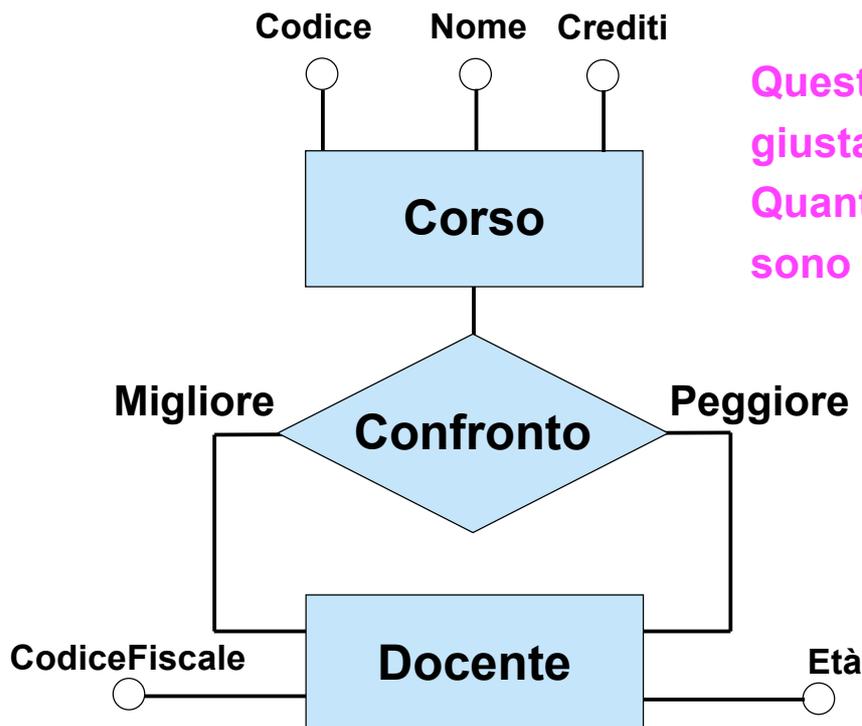
Questa soluzione è giusta? Se no, perché ?

Esercizio 5: soluzione 1



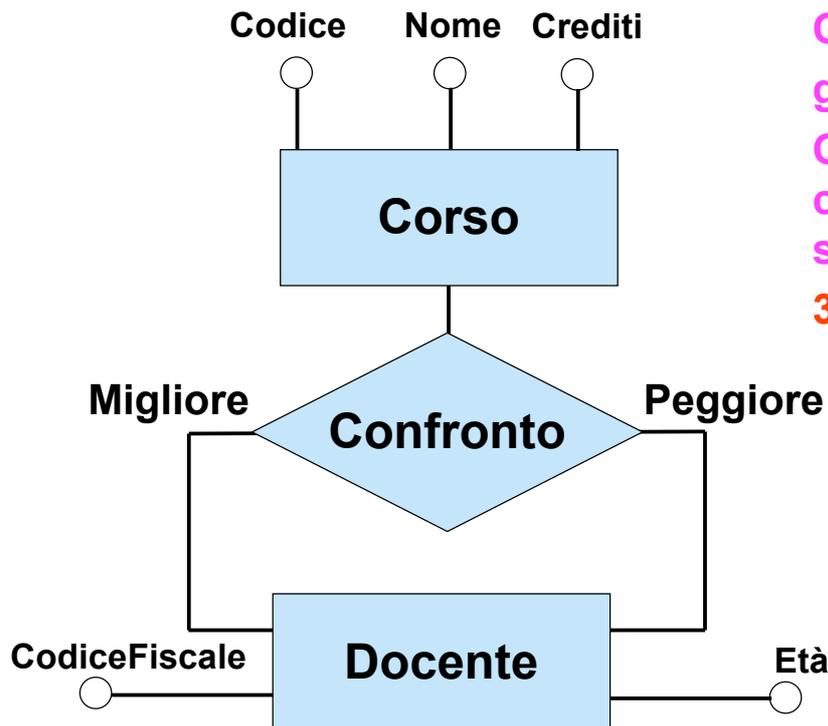
Questa soluzione è giusta? **NO**, perché nella relazione "Confronto" sono definiti due ruoli uguali

Esercizio 5: soluzione 2



Questa soluzione è giusta?
Quanti e quali ruoli ci sono nello schema ?

Esercizio 5: soluzione 2



Questa soluzione è giusta? **SI'**

Quanti e quali ruoli ci sono nello schema ?

3: Corso, Migliore, Peggior

Esercizio 6: livello estensionale

Descrivere l'istanza dello schema concettuale precedente (soluzione 2) corrispondente ai seguenti dati (si ignorino gli attributi):

Docenti: Paolo, Mario, Grazia

Corsi: Basi di Dati, Logica, Algebra

Graduatorie:

Basi di dati

1. Mario
2. Grazia
3. Paolo

Logica

1. Grazia
2. Mario
3. Paolo

Algebra

1. Grazia
2. Paolo

Esercizio 6: soluzione

Istanza I:

istanze(Docente): { Paolo, Mario, Grazia }

istanze(Corso): { BasiDiDati, Logica, Algebra }

istanze(Confronto):

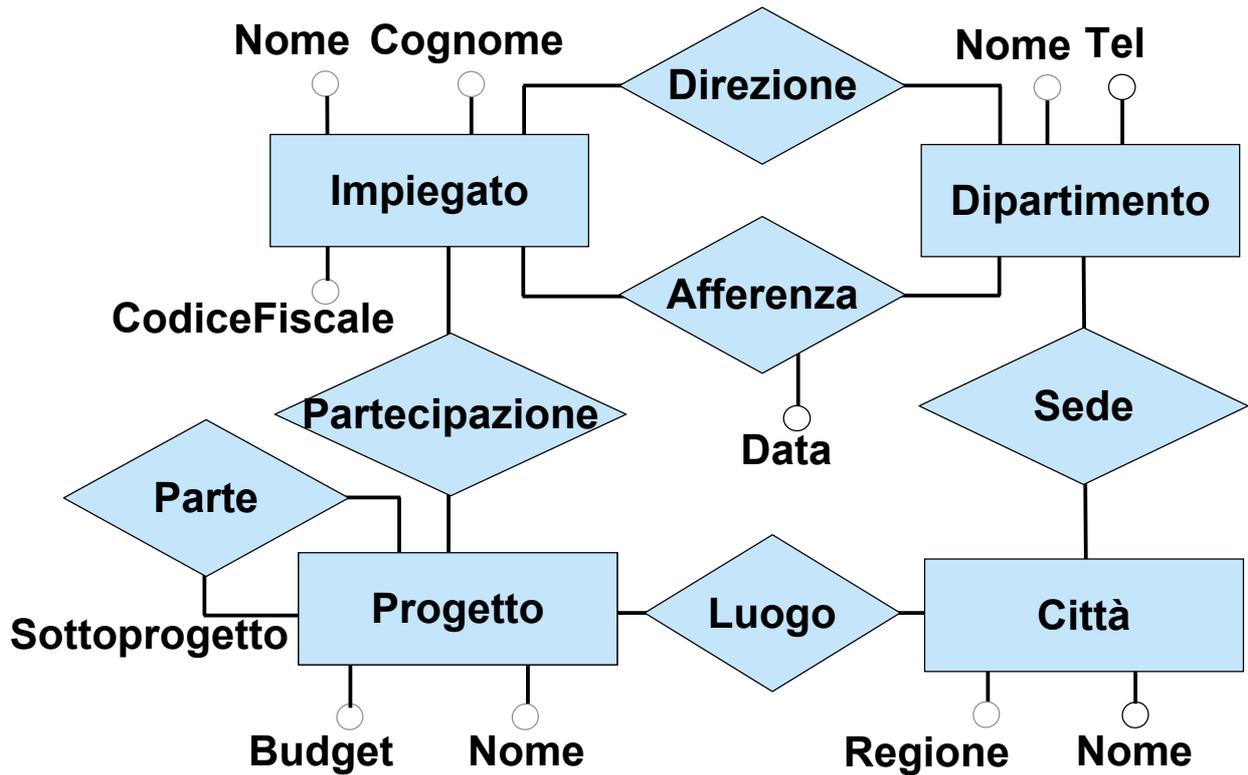
{ (Corso:BasiDiDati, Migliore:Mario, Peggioro: Grazia),
(Corso:BasiDiDati, Migliore:Mario, Peggioro: Paolo),
(Corso:BasiDiDati, Migliore:Grazia, Peggioro: Paolo),
(Corso:Logica, Migliore:Grazia, Peggioro: Mario),
(Corso:Logica, Migliore:Grazia, Peggioro: Paolo),
(Corso:Logica, Migliore:Mario, Peggioro: Paolo),
(Corso:Algebra, Migliore:Grazia, Peggioro: Paolo) }

Esercizio 7: schema concettuale

Descrivere lo schema concettuale della seguente realtà:

Degli impiegati interessa il codice fiscale, il nome, il cognome, i dipartimenti ai quali afferiscono (con la data di afferenza), ed i progetti ai quali partecipano. Dei progetti interessa il nome, il budget, e la città in cui hanno luogo le corrispondenti attività. Alcuni progetti sono parti di altri progetti, e sono detti loro sottoprogetti. Dei dipartimenti interessa il nome, il numero di telefono, gli impiegati che li dirigono, e la città dove è localizzata la sede. Delle città interessa il nome e la regione.

Esercizio 7: soluzione

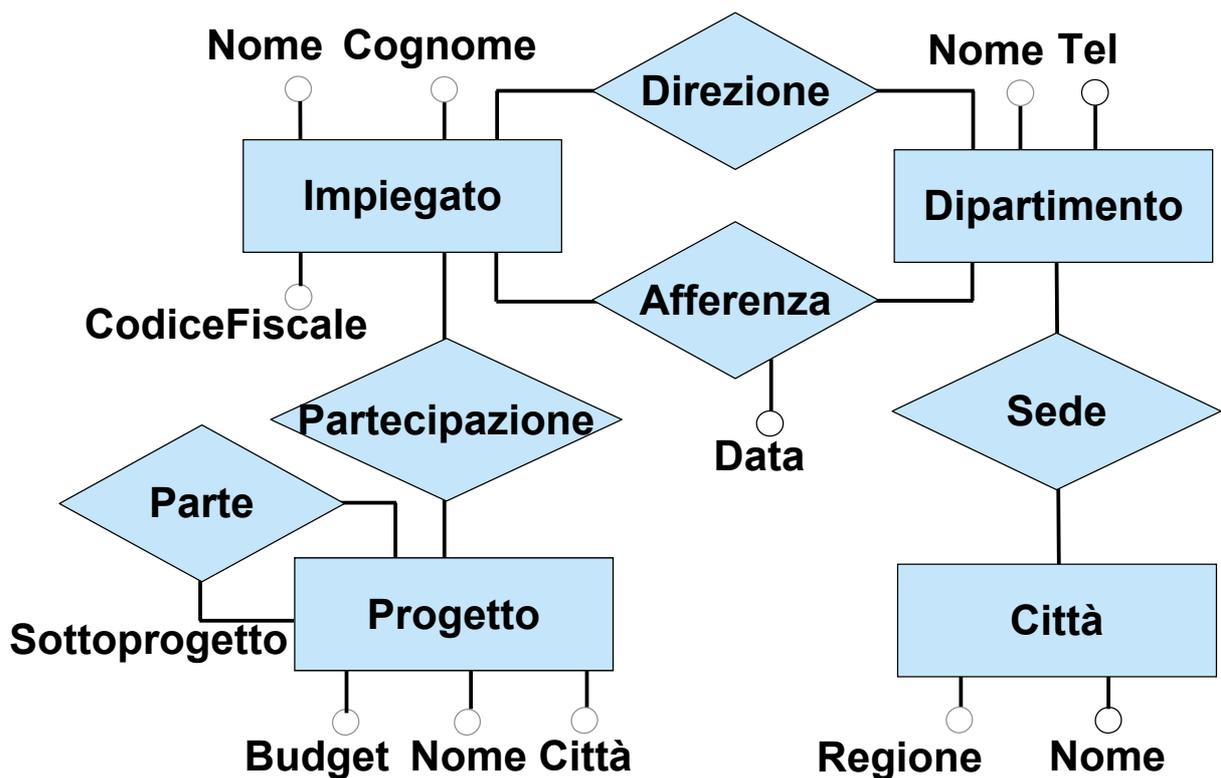


Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 67

Esercizio 8: qual è l'errore in questo schema?



Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 68

Esercizio 8: soluzione

L'errore sta nell'aver rappresentato il luogo di svolgimento di un progetto come un attributo. In realtà il luogo di svolgimento è una città, e questo viene perso nello schema.

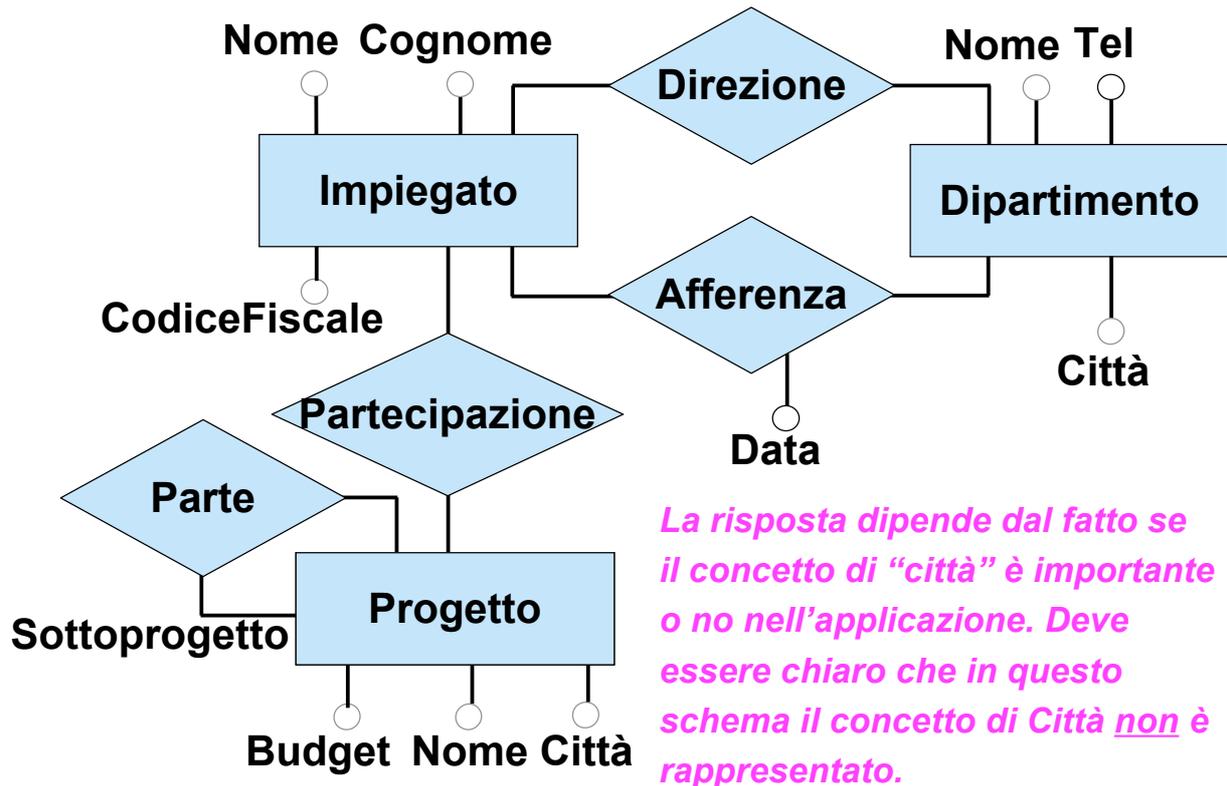
Infatti, il nome dell'attributo (Città) **non ha nulla a che vedere con la semantica** dello schema (che invece è caratterizzata dalle istanze dello schema): ciò che viene perso è che, in ogni istanza dello schema, il luogo di svolgimento di un progetto è una istanza dell'entità Città.

Detto in altro modo, nello schema il luogo di progetto è rappresentato come una proprietà locale di Progetto, perdendo così il legame con l'entità Città.

Esercizio 9: è corretto questo schema?



Esercizio 9: soluzione



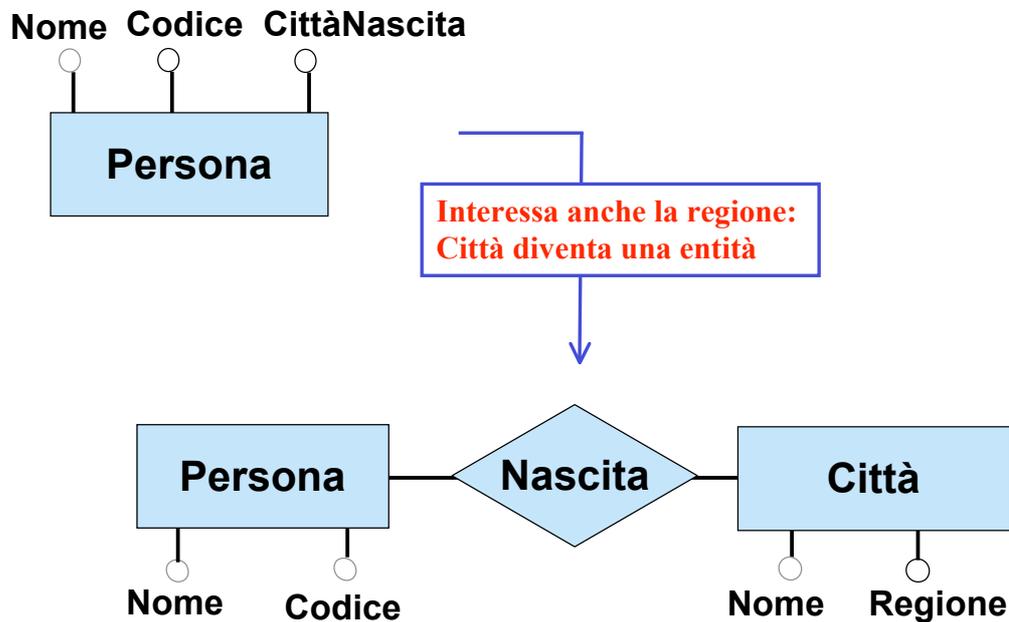
Scelta tra entità e attributo

Un concetto verrà modellato come

- una **entità**
 - se le sue istanze sono concettualmente significative indipendentemente da altre istanze
 - se ha o potrà avere delle proprietà indipendenti dagli altri concetti
 - se il concetto è importante nell'applicazione
- un **attributo** di una entità o relazione
 - se le sue istanze non sono concettualmente significative
 - se non ha senso considerare una sua istanza indipendentemente da altre istanze
 - se serve solo a rappresentare una proprietà locale di un altro concetto

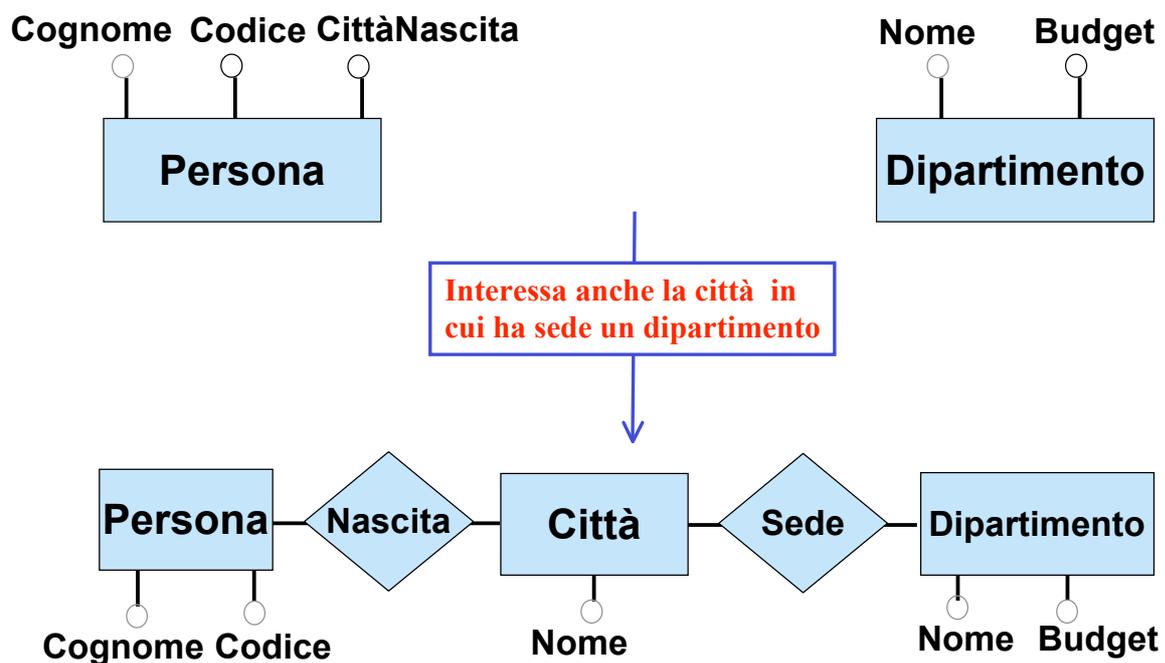
Scelta tra entità e attributo

Le scelte possono cambiare durante l'analisi. Esempio:



Scelta tra entità e attributo

Le scelte possono cambiare durante l'analisi. Esempio:



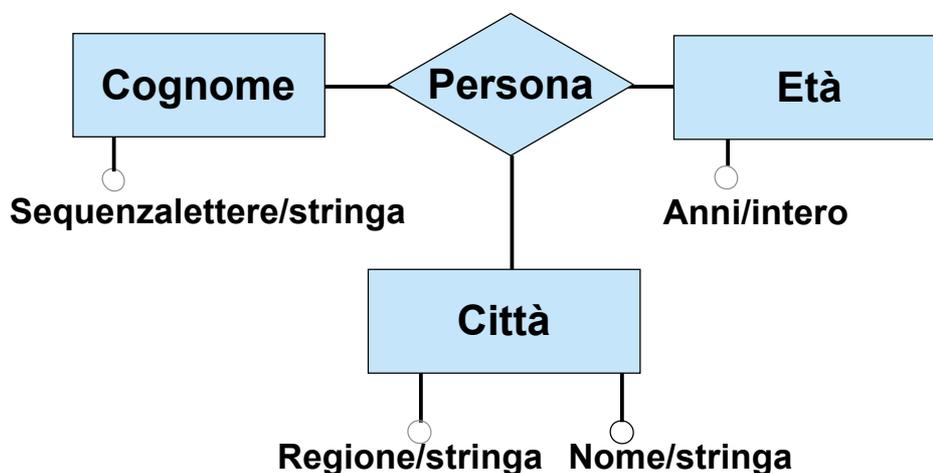
Scelta tra entità e relazione

Un concetto verrà modellato come

- una **entità**
 - se le sue istanze sono concettualmente significative indipendentemente da altre istanze
 - se ha o potrà avere delle proprietà indipendenti dagli altri concetti
 - se ha o potrà avere relazioni con altri concetti
- una **relazione**
 - se le sue istanze non sono concettualmente significative indipendentemente da altre istanze, cioè se le sue istanze rappresentano n-ple di altre istanze
 - se non ha senso pensare alla partecipazione delle sue istanze ad altre relazioni

Esercizio 10: scelta tra entità e relazione

In un contesto in cui occorre rappresentare le persone e le loro proprietà, questo schema è chiaramente sbagliato. Perché?



Esercizio 10: soluzione

- Nello schema, la classe delle persone non è modellato, nel senso che, a livello estensionale non esistono oggetti con significato autonomo che sono istanze di persone
- A livello estensionale ogni persona sarà quindi rappresentata da una tupla, che, per definizione, esiste solo perchè esistono le sue componenti. E' come dire che una persona esiste solo perchè ha un cognome, una età, ed una città di nascita
- Modellando Cognome come entità, lo schema asserisce implicitamente che i cognomi sono oggetti la cui esistenza autonoma è di interesse per la nostra applicazione, e questo è chiaramente assurdo. Stessa considerazione vale per Età
- Infine, per la semantica delle relazioni, lo schema asserisce implicitamente che non esistono due persone con lo stesso cognome, la stessa età e la stessa città di nascita, e anche questo è assurdo.

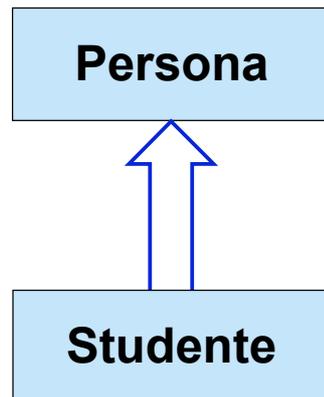
La nozione di relazione ISA tra entità

- Fino ad ora non abbiamo detto nulla sul fatto se due entità possano o no avere istanze in comune
- E' facile verificare che, in molti contesti, può accadere che tra due classi rappresentate da due entità nello schema concettuale sussista la **relazione ISA** (o **relazione di sottoinsieme**), e cioè che ogni istanza di una sia anche istanza dell'altra.
- La relazione ISA nel modello ER si può definire tra due entità, che si dicono "entità padre" ed "entità figlia" (o sottoentità, cioè quella che rappresenta un sottoinsieme della entità padre)

Sintassi: rappresentazione grafica della relazione ISA tra entità

La relazione ISA si rappresenta nel diagramma dello schema concettuale mediante una freccia dalla sottoentità alla entità padre

Esempio

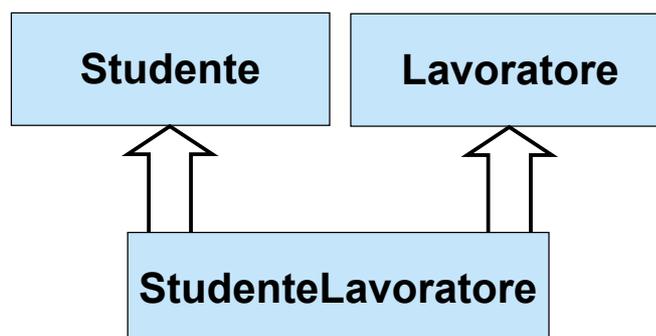


*Si dice che **Studente** è in relazione ISA con **Persona**, o, in alternativa, che **Studente ISA Persona***

Sintassi: regole sulla relazione ISA tra entità

ATTENZIONE: vige la regola che una entità può avere al **massimo una** entità padre. In altre parole, il modello ER **non** ammette **ereditarietà multipla**

Esempio



*Questo schema è **scorretto***

Semantica della relazione ISA tra entità

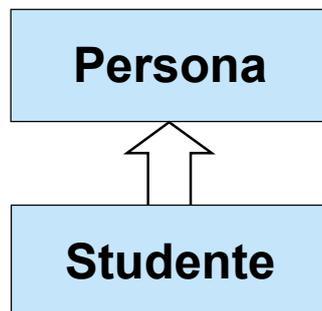
A livello estensionale la relazione ISA impone semplicemente che in ogni istanza **I** di uno schema in cui **E₁** ISA **E₂**:

$$\text{istanze}(\mathbf{I}, \mathbf{E}_1) \subseteq \text{istanze}(\mathbf{I}, \mathbf{E}_2)$$

Si noti che dalla definizione segue che la relazione ISA è **riflessiva e transitiva**.

Esempio

Schema



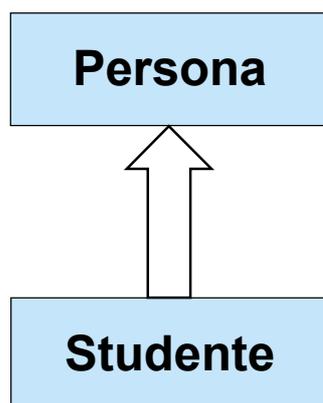
Istanza I:

$$\text{istanze}(\text{Persona}) = \{ a, b, c, d \}$$

$$\text{istanze}(\text{Studente}) = \{ b, d \}$$

Relazione ISA a livello estensionale

Schema



Istanza I1:

$$\text{istanze}(\text{Persona}) = \{ a, b, c, d \}$$

$$\text{istanze}(\text{Studente}) = \{ b, d \}$$

Istanza I2:

$$\text{istanze}(\text{Persona}) = \{ a, b, c, d \}$$

$$\text{istanze}(\text{Studente}) = \{ a, b, c, d \}$$

Questa non è un'istanza:

$$\text{istanze}(\text{Persona}) = \{ a \}$$

$$\text{istanze}(\text{Studente}) = \{ a, b \}$$

Istanza I3:

$$\text{istanze}(\text{Persona}) = \{ a, b, c, d \}$$

$$\text{istanze}(\text{Studente}) = \{ \}$$

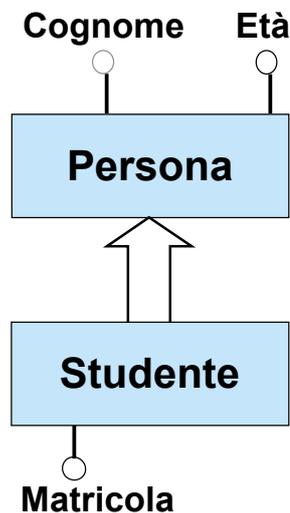
Ereditarietà su entità nel modello ER

Principio di ereditarietà: ogni proprietà dell'entità padre è anche una proprietà della sottoentità, e non si riporta esplicitamente nel diagramma. L'entità figlia può avere ovviamente ulteriori proprietà

Dal fatto che

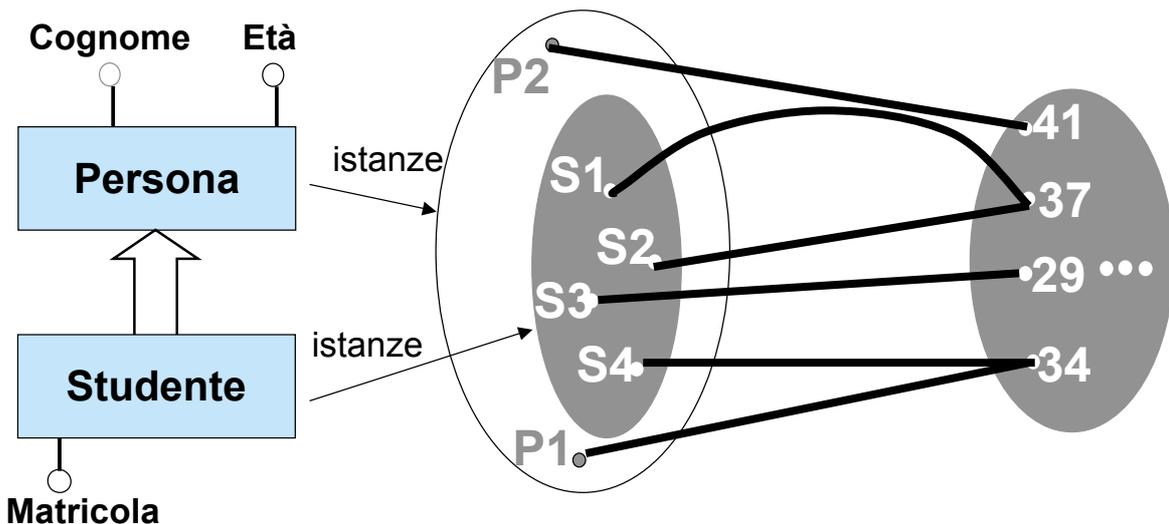
1. Ogni istanza di Persona ha una Età
 2. Ogni istanza di Studente è istanza di Persona
- segue logicamente che**
3. Ogni istanza di Studente ha una Età

Ragionamento sillogistico (cfr. opera di Aristotele più di due millenni fa)



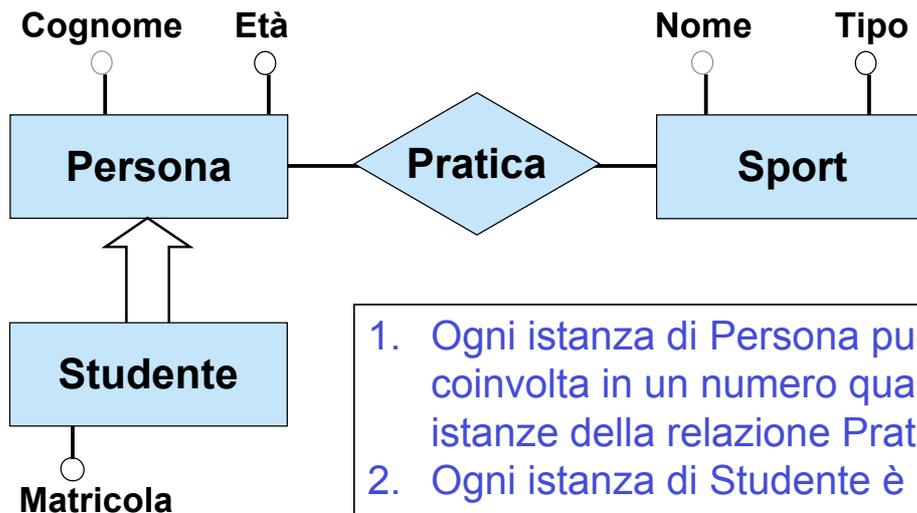
Cognome, Età ereditati da Persona, Matricola ulteriore proprietà

Ereditarietà: livello estensionale



L'entità Studente eredita l'attributo Età dalla entità padre Persona. Ciò significa che ogni studente ha una età: infatti, se esistesse una istanza di Studente che non ha un valore per l'attributo Età, quella sarebbe una istanza di Persona senza un valore per Età, contraddicendo il significato dell'attributo Età (funzione totale) per Persona

Ereditarietà delle relazioni

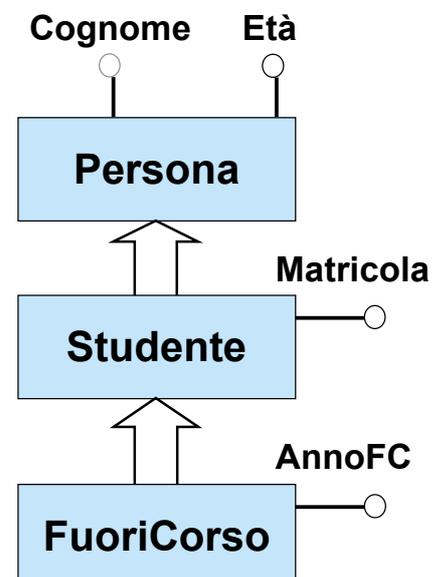


1. Ogni istanza di Persona può essere coinvolta in un numero qualunque di istanze della relazione Pratica
 2. Ogni istanza di Studente è una istanza di Persona
- quindi**
3. Ogni istanza di Studente può essere coinvolta in un numero qualunque di istanze della relazione Pratica

Ereditarietà della relazione ISA: transitività

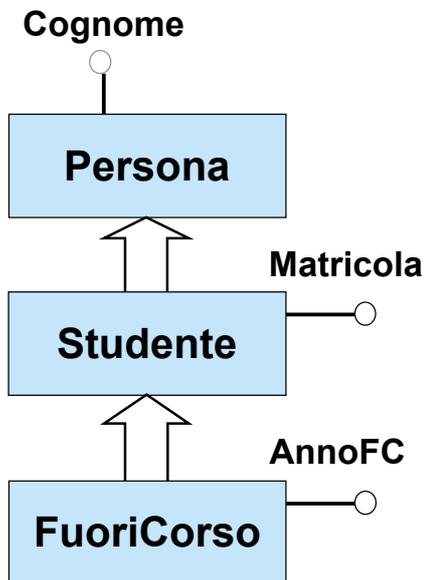
Principio di ereditarietà: anche la relazione ISA si eredita. Questo conferma che la relazione ISA è transitiva (oltre a essere riflessiva)

- Dal fatto che**
1. Ogni istanza di Studente è una istanza di Persona
 2. Ogni istanza di FuoriCorso è una istanza di Studente
- segue logicamente che**
3. Ogni istanza di FuoriCorso è una istanza di Persona



Esercizio 11: I1 è una istanza dello schema?

Schema



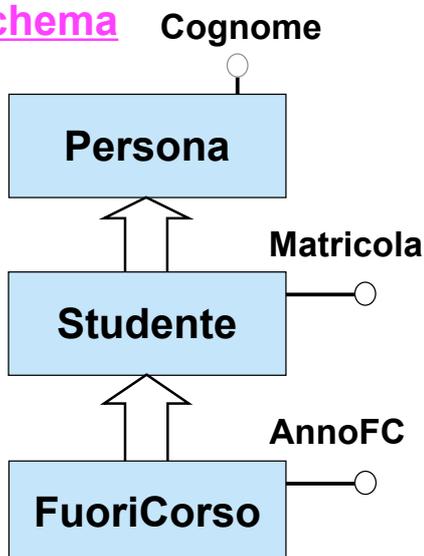
I1:

istanze(Persona) = { a,b,d }
istanze(Studente) = { b,d }
istanze(FuoriCorso) = { b,d }
istanze(Cognome) =
{ (a,'a'), (b,'b'), (d,'d') }
istanze(Matricola) =
{ (b,'12'), (d,'34'), (a,'89') }
istanze(AnnoFC) = { (b,2) }

Esercizio 11: soluzione

La risposta è NO, perchè: - (a,'89') in Matricola
- manca AnnoFC per d

Schema



I1:

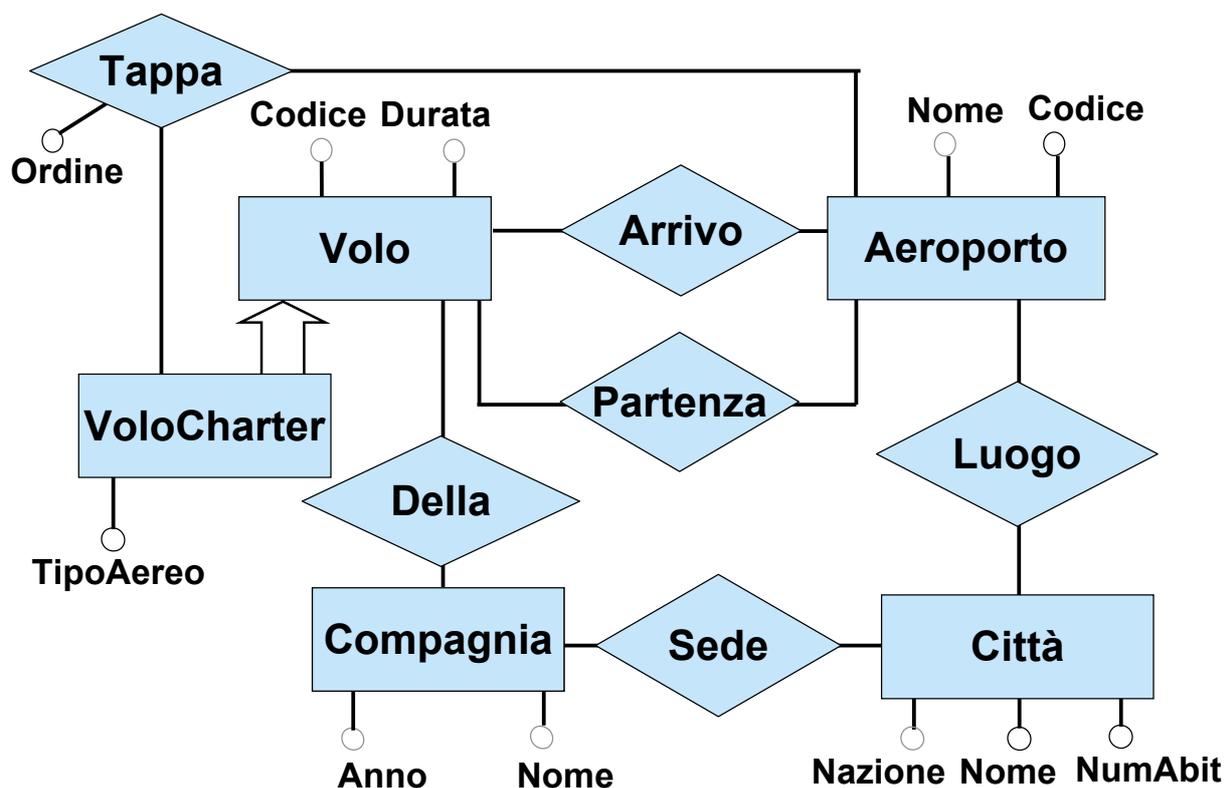
istanze(Persona) = { a,b,d }
istanze(Studente) = { b,d }
istanze(FuoriCorso) = { b,d }
istanze(Cognome) =
{ (a,'a'), (b,'b'), (d,'d') }
istanze(Matricola) =
{ (b,'12'), (d,'34'), (a,'89') }
istanze(AnnoFC) = { (b,2) }

Esercizio 12: schema concettuale

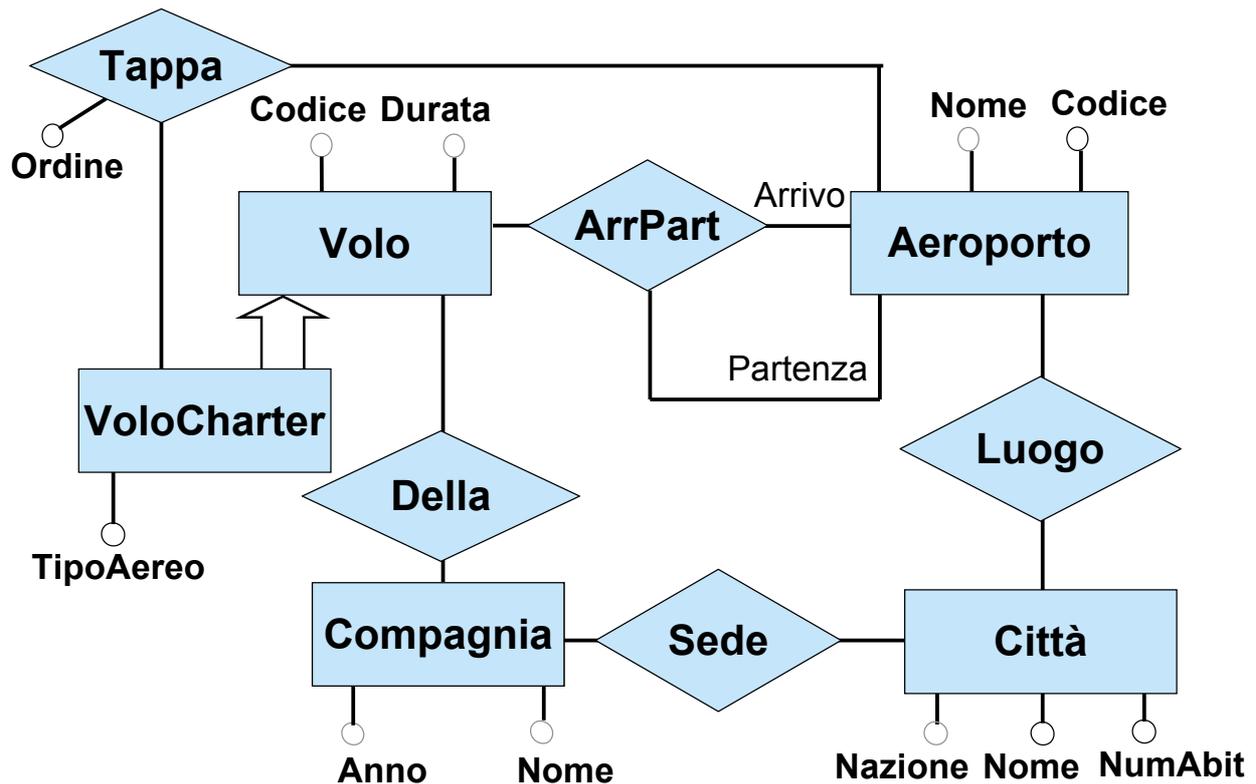
Descrivere lo schema concettuale corrispondente ad un'applicazione riguardante voli aerei, per la quale valgono le seguenti specifiche.

Dei voli interessa: codice, durata in minuti, compagnia aerea, aeroporto di partenza e aeroporto di arrivo. Degli aeroporti interessa: codice, nome, città (con nome e numero di abitanti) e nazione. Delle compagnie aeree interessa il nome, l'anno di fondazione, e la città in cui ha sede la direzione. I voli charter sono particolari voli, che possono prevedere tappe intermedie in aeroporti. Delle tappe intermedie di un volo charter interessa l'ordine con cui esse si susseguono (ad esempio, il volo 124, che parte da "Milano Linate" e arriva a "Palermo Punta Raisi", prevede prima l'aeroporto di Bologna e poi quello di Napoli come tappe intermedie). Infine, dei voli charter interessa anche il tipo di aereomobile utilizzato per il volo.

Esercizio 12: soluzione



Esercizio 12: altra soluzione



Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 91

La nozione di generalizzazione tra entità

Finora, abbiamo considerato la relazione ISA che stabilisce che l'entità padre è più generale della sottoentità. Talvolta, però, l'entità padre può generalizzare diverse sottoentità rispetto ad un unico criterio. In questo caso si parla di **generalizzazione**.

Nella generalizzazione, le sottoentità hanno insiemi di istanze disgiunti a coppie (anche se in alcune varianti del modello ER, si può specificare se due sottoentità della stessa entità padre sono disgiunte o no).

Una generalizzazione può essere di due tipi:

- **Completa:** l'unione delle istanze delle sottoentità è uguale all'insieme delle istanze dell'entità padre
- **Non completa**

Giuseppe De Giacomo

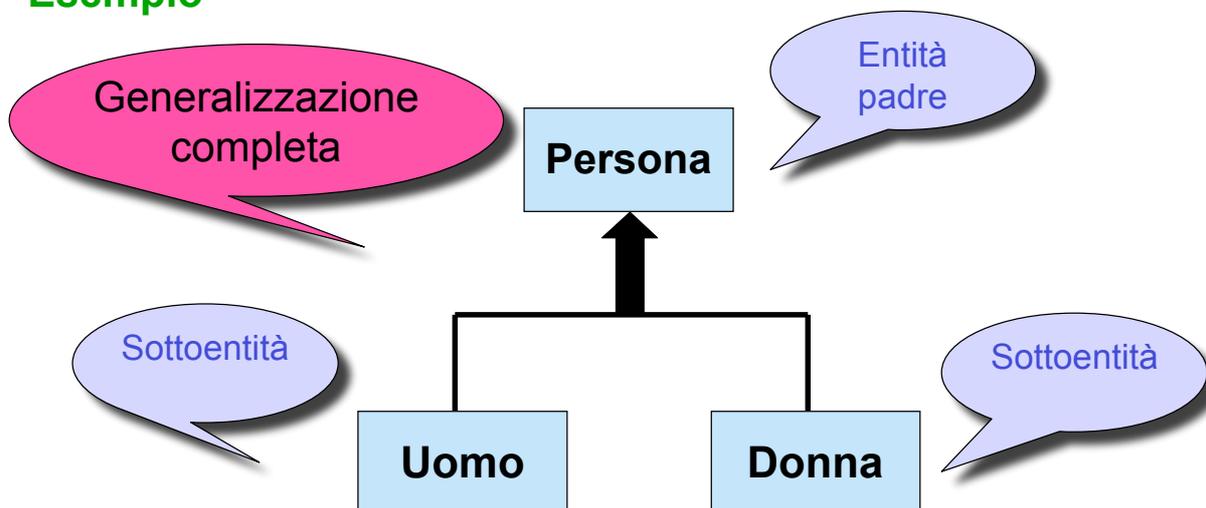
Basi di Dati

Progettazione concettuale - 92

Sintassi: rappresentazione della generalizzazione

La generalizzazione si indica collegando mediante un arco le sottoentità, e collegando con una freccia tale arco alla entità padre. La freccia è annerita se la generalizzazione è completa.

Esempio



Giuseppe De Giacomo

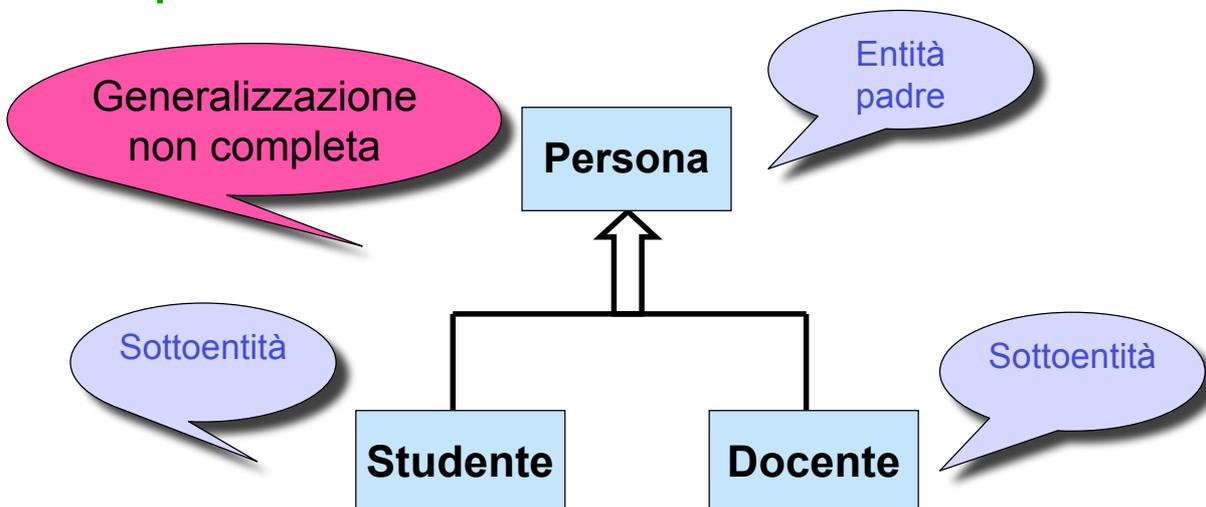
Basi di Dati

Progettazione concettuale - 93

Sintassi: rappresentazione della generalizzazione

La freccia è non è annerita se la generalizzazione non è completa.

Esempio



Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 94

Semantica della generalizzazione

Se in uno schema **S** è definita una generalizzazione tra una entità padre **F** e le sottoentità **E₁, E₂, ..., E_n**, in ogni istanza **I** dello schema **S** si ha:

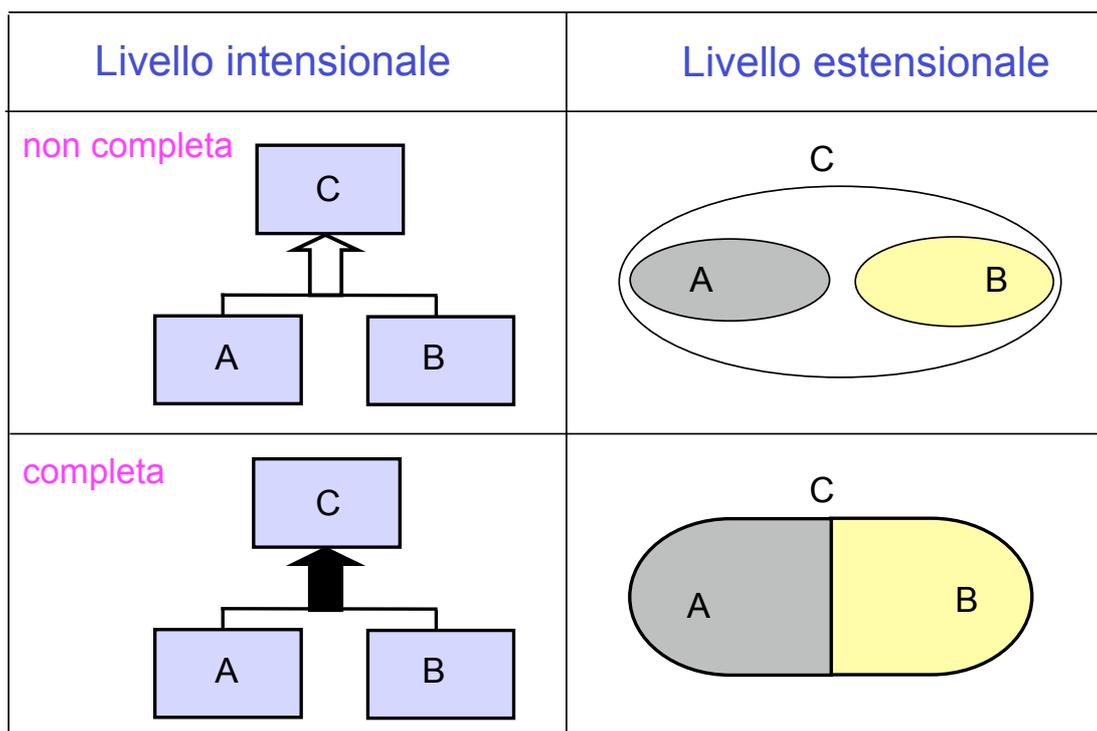
- $istanze(I, E_1) \subseteq istanze(I, F)$
- $istanze(I, E_2) \subseteq istanze(I, F)$
-
- $istanze(I, E_n) \subseteq istanze(I, F)$
- $istanze(I, E_i) \cap istanze(I, E_k) = \emptyset$, per ogni $1 \leq i, k \leq n, i \neq k$

In più, se la generalizzazione è completa, si ha che:

- $istanze(I, E_1) \cup \dots \cup istanze(I, E_n) = istanze(I, F)$

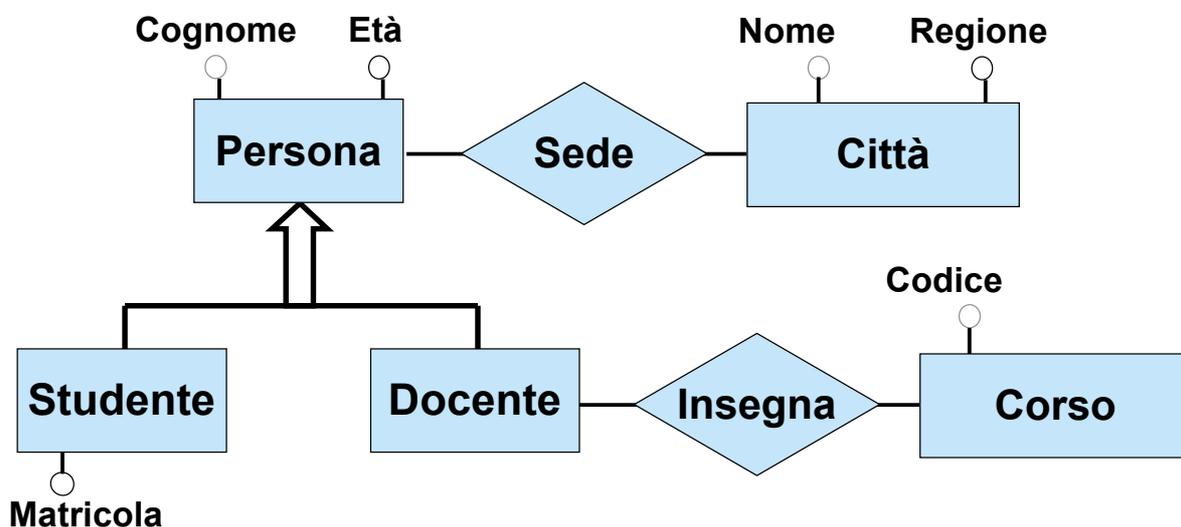
Si noti che ciò implica che in una generalizzazione ogni **E_i** è in relazione ISA con **F**, e che in una generalizzazione completa le sottoentità formano **partizione** dell'entità padre.

Generalizzazioni a livello estensionale



Generalizzazioni ed ereditarietà

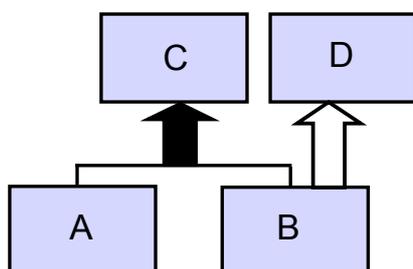
Il principio di ereditarietà vale anche per le generalizzazioni: ogni proprietà dell'entità padre è anche una proprietà della sottoentità, e non si riporta esplicitamente nel diagramma. L'entità figlia può avere ovviamente ulteriori proprietà.



Regole sulla generalizzazione

Attenzione: continua a valere la regola che una entità può avere al massimo una entità padre (tenendo conto sia della relazione ISA sia delle generalizzazioni)

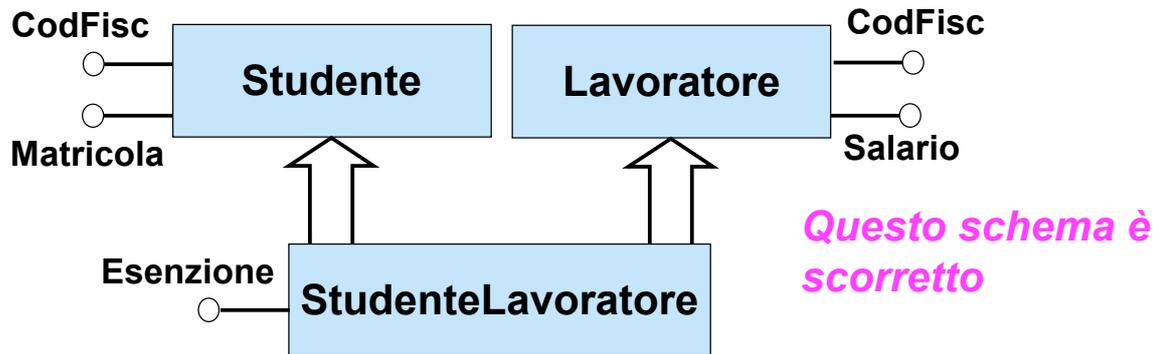
Esempio



Questo schema è scorretto

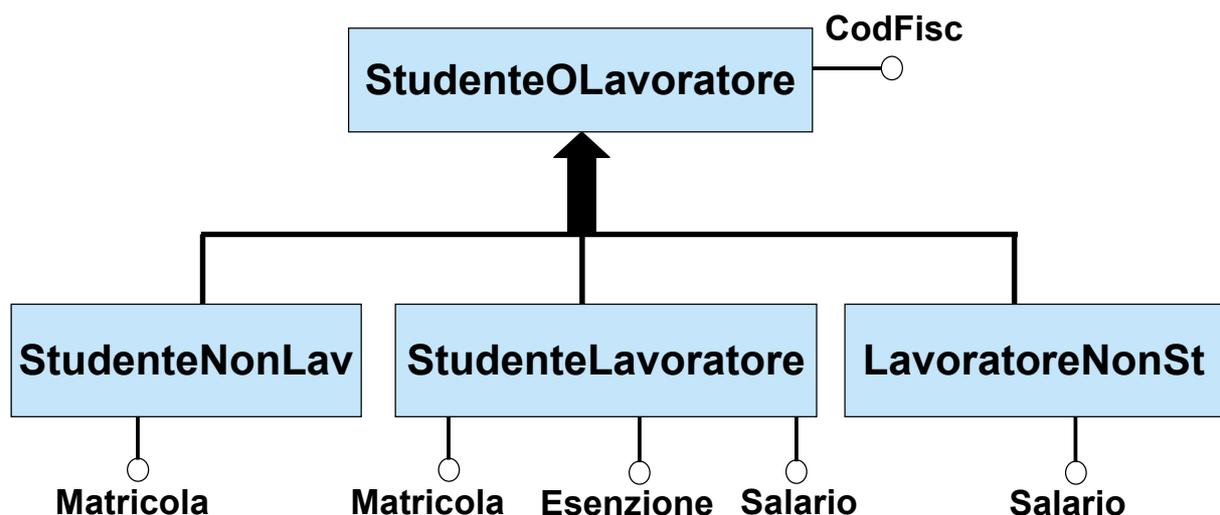
Dato uno schema S, si può costruire un grafo che ha le entità di S come nodi ed ha un arco da una entità E ad una entità F se F è padre di E secondo la relazione ISA o secondo la generalizzazione. Diremo che una entità G è **antenata** di E in uno schema se esiste un cammino da E ad G nel grafo associato

Esercizio 13: relazione ISA e generalizzazioni



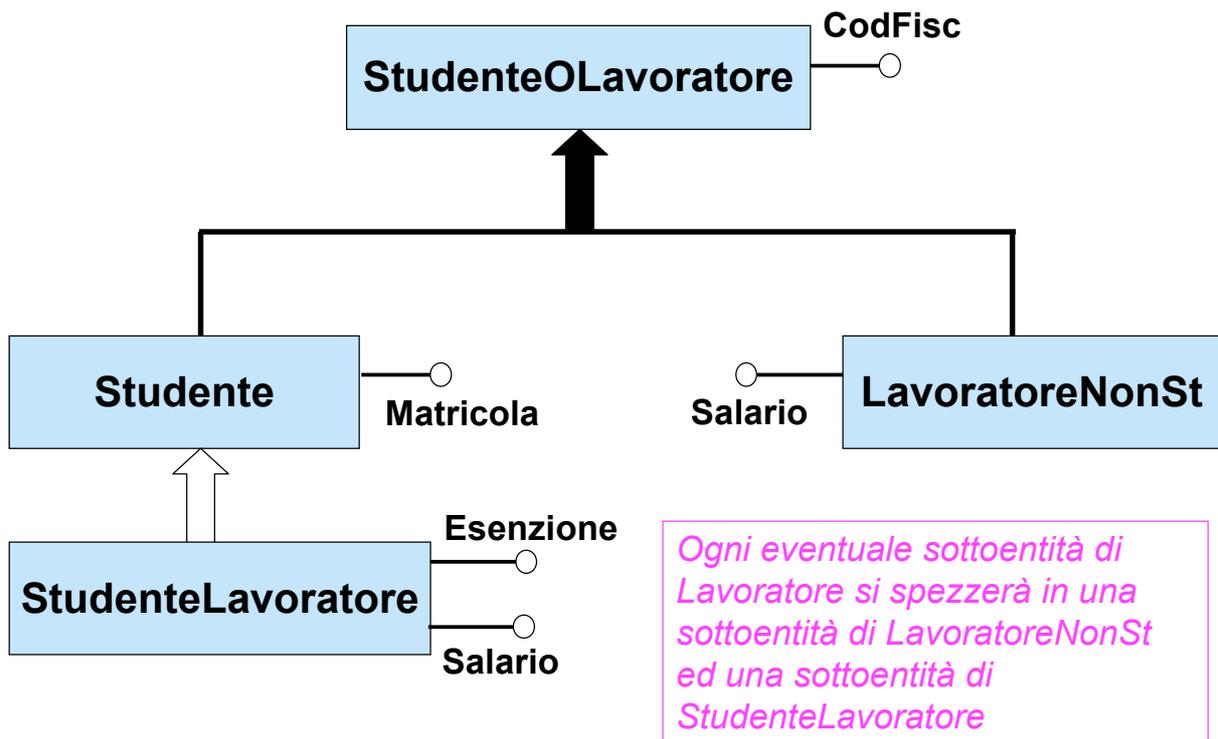
Si può ristrutturare lo schema in modo che lo schema risultante sia corretto, e colga sostanzialmente la stessa semantica?

Esercizio 13: soluzione 1

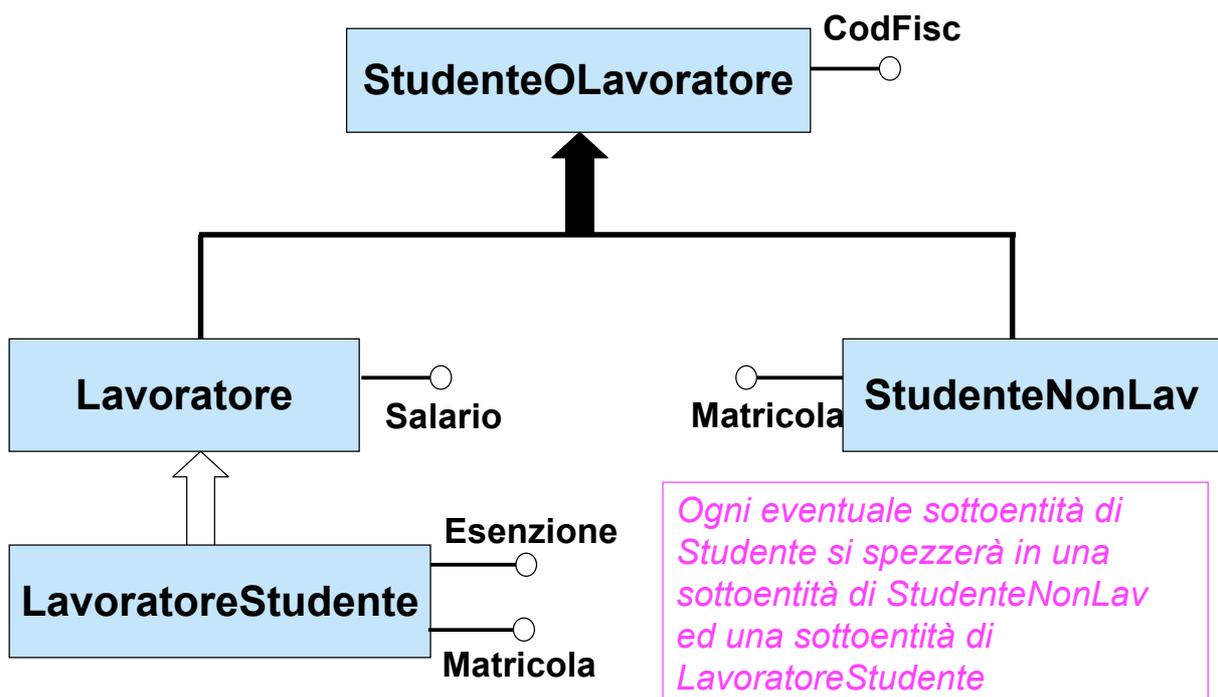


Ogni eventuale sottoentità di Studente si spezzerà in una sottoentità di StudenteNonLav ed una sottoentità di StudenteLavoratore. Analogamente, ogni sottoentità di Lavoratore si spezzerà in una sottoentità di LavoratoreNonSt ed una sottoentità di StudenteLavoratore

Esercizio 13: soluzione 2

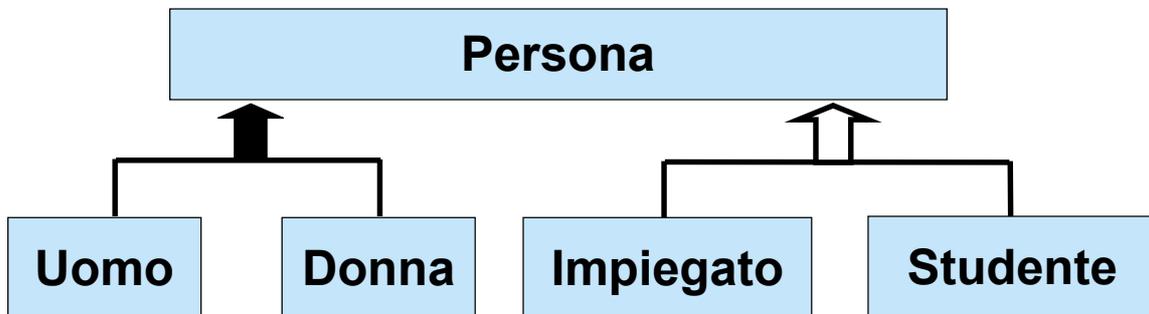


Esercizio 13: soluzione 3



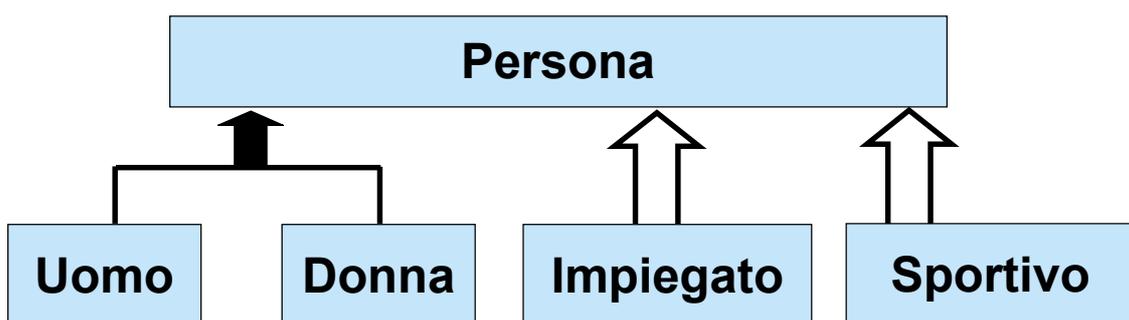
Diverse generalizzazioni della stessa classe

La stessa entità può essere padre in diverse generalizzazioni



Concettualmente, non c'è alcuna correlazione tra due generalizzazioni diverse, perchè rispondono a due criteri diversi di classificare le istanze della entità padre

Differenza tra due isa e una generalizzazione



Le due sottoclassi derivano da uno stesso criterio di classificazione delle istanze della superclasse

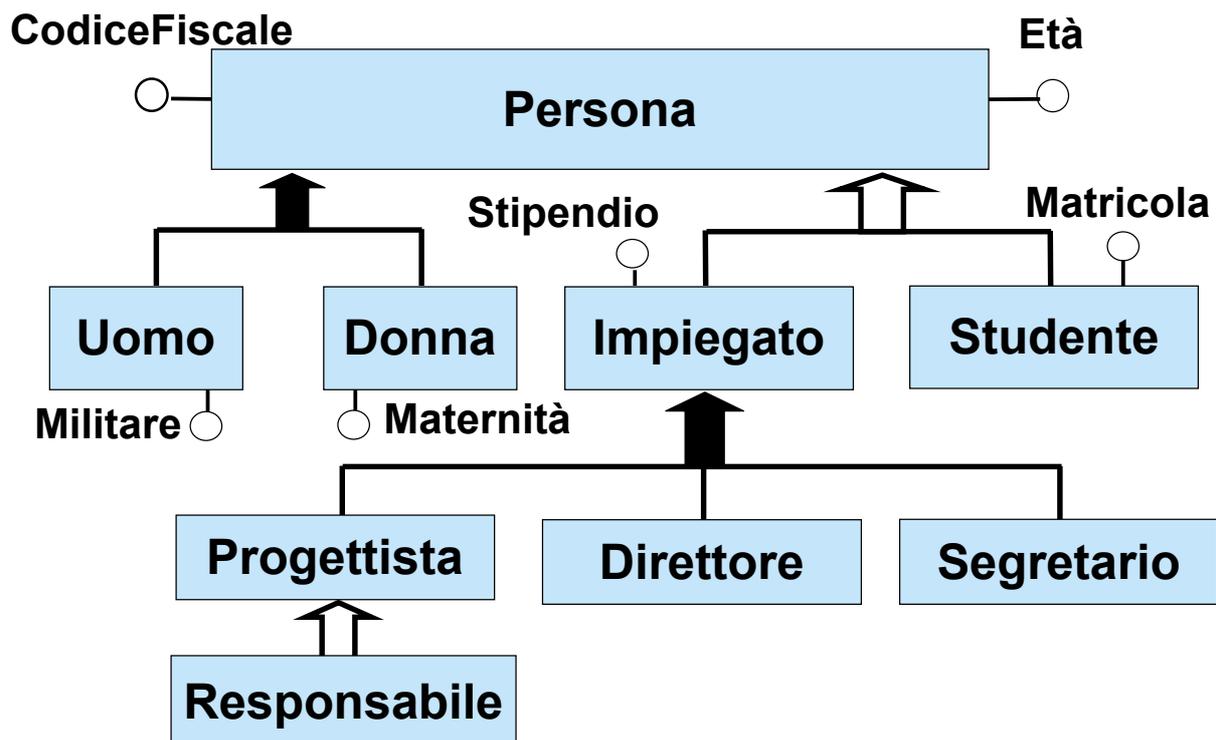
Le due sottoentità sono indipendenti, nel senso che il loro significato non deriva dallo stesso criterio di classificazione delle istanze della entità padre

Esercizio 13: generalizzazione

Descrivere lo schema concettuale corrispondente alle seguenti specifiche.

Le persone hanno un codice fiscale ed una età. Gli uomini anche la posizione militare, e le donne anche il numero di maternità. Gli impiegati hanno lo stipendio e possono essere o segretari, o direttori o progettisti. Alcuni progettisti sono responsabili di progetto. Gli studenti (che non possono essere impiegati) hanno un numero di matricola. Esistono persone che non sono né impiegati né studenti.

Esercizio 13: soluzione

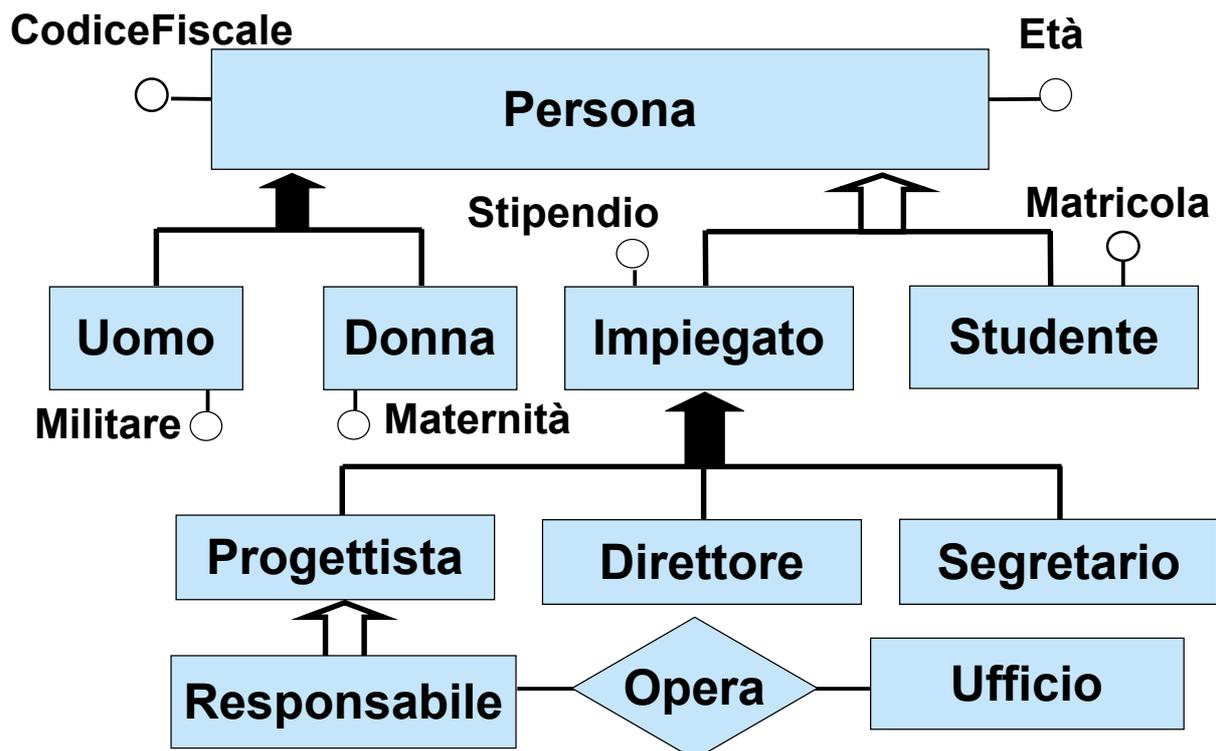


Entità disgiunte

Da quanto detto sinora, non è ancora chiaro quando, date due entità di uno schema, esse sono disgiunte (cioè non hanno istanze comuni in tutte le istanze dello schema). A questo proposito, valgono queste osservazioni:

- In ogni schema si assume che le entità che non hanno padre nello schema siano in realtà sottoentità di una generalizzazione completa la cui entità padre è una entità speciale detta **TOP** (ne segue che tali entità sono disgiunte a coppie, e che in ogni istanza dello schema, **TOP** rappresenta tutti gli oggetti che sono istanze di qualche entità)
- E' facile verificare che due entità E ed F sono **disgiunte** in uno schema S se esistono due entità G ed H (non necessariamente diverse da E ed F) nello schema S tali che:
 - G ed H sono sottoentità della stessa entità padre in una generalizzazione
 - G è antenata di E, ed H è antenata di F

Esercizio 14: elencare tutte le coppie di entità disgiunte in questo schema



Esercizio 14: soluzione

Coppie di entità disgiunte:

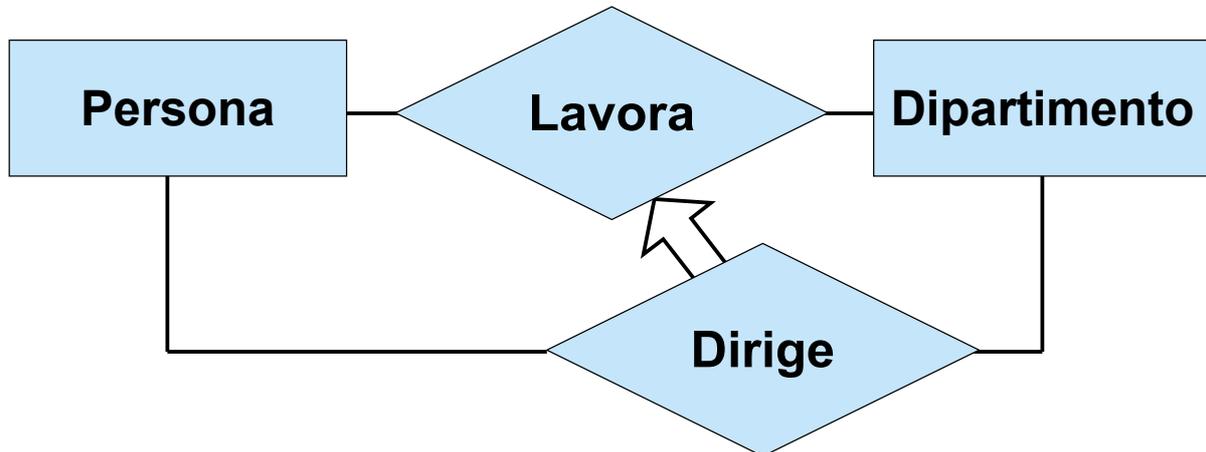
- Ufficio e ognuna delle altre entità
- Uomo e Donna
- Impiegato e Studente
- Progettista e Studente
- Segretario e Studente
- Direttore e Studente
- Responsabile e Studente
- Progettista e Direttore
- Progettista e Segretario
- Direttore e Segretario
- Responsabile e Direttore
- Responsabile e Segretario

Relazione ISA e generalizzazioni tra relazioni

- La relazione ISA può essere definita anche tra relazioni, ma con la seguente **regola**: si può definire la relazione ISA tra la relazione R (relazione figlia) e la relazione Q (relazione padre) in uno schema S solo se R e Q
 - hanno lo stesso grado
 - hanno gli stessi ruoli
 - per ogni ruolo U, l'entità corrispondente ad U in R è una entità figlia dell'entità corrispondente ad U in Q
- La semantica non cambia rispetto al caso della relazione ISA tra entità: se in uno schema **S**, è definita la relazione ISA tra **R** e **Q** (**R** ISA **Q**, dove **R** e **Q** sono due relazioni con lo stesso grado e gli stessi ruoli), allora in ogni istanza **I** dello schema **S**, $istanze(I, R) \subseteq istanze(I, Q)$
- Analogamente, si possono definire generalizzazioni tra relazioni

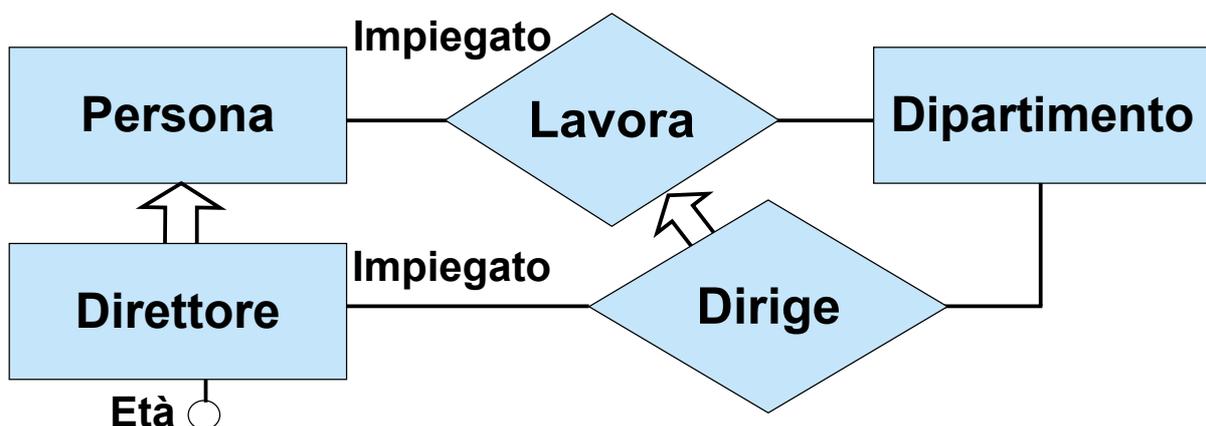
Esempio di relazione ISA tra relazioni

Il direttore di un dipartimento deve lavorare in quel dipartimento

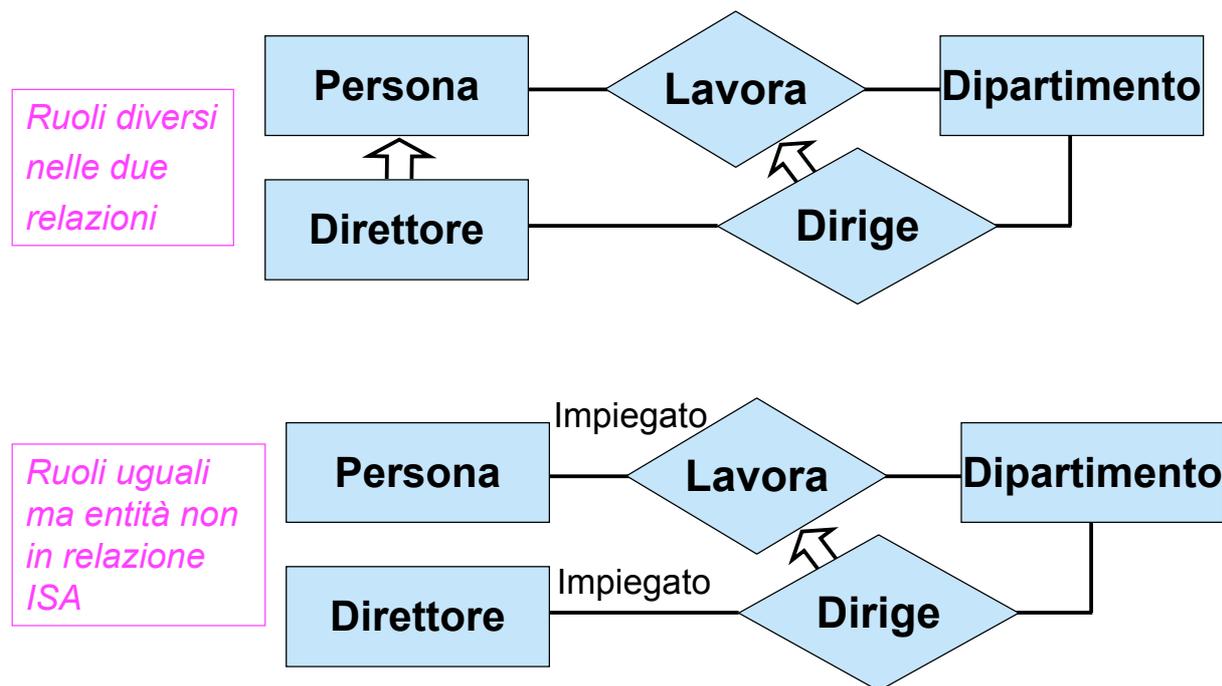


Esempio di relazione ISA tra relazioni

Il direttore di un dipartimento deve lavorare in quel dipartimento, ed ha proprietà aggiuntive (ad esempio l'età) rispetto agli impiegati



Esempi di schemi non corretti



Relazioni disgiunte

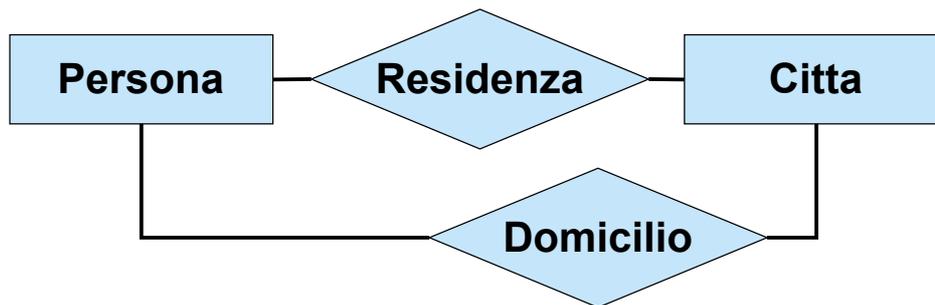
Per le relazioni valgono osservazioni analoghe a quelle che abbiamo fatto sulle entità.

- In ogni schema si assume che le relazioni dello stesso grado k e con gli stessi ruoli U_1, U_2, \dots, U_k che non hanno padre nello schema siano in realtà sottoentità in una ISA la cui relazione padre è una relazione speciale detta **TOP** _{U_1, U_2, \dots, U_k} (ne segue che, al contrario del caso delle entità, tali relazioni non sono disgiunte a coppie).
- Due relazioni P ed R che hanno lo stesso grado e gli stessi ruoli sono **disgiunte** in uno schema S se esistono due relazioni Q e T (non necessariamente diverse da P ed R) nello schema S tali che:
 - Q e T sono relazioni figlie della stessa relazione padre in una generalizzazione
 - Q è antenata di P , e T è antenata di R

Relazioni disgiunte

Da quanto detto segue che, al contrario del caso delle entità, due relazioni (dello stesso grado e con gli stessi ruoli) che non hanno padre nello schema **non sono disgiunte**.

Esempio: Residenza e Domicilio non sono disgiunte



Vincoli di integrità nel modello ER

Ricordiamo che un **vincolo di integrità** è una regola che si esprime sullo schema (livello intensionale), ma che specifica una condizione che deve valere **per ogni istanza** dello schema (livello estensionale).

Classificazione dei vincoli di integrità nel modello ER:

- Vincoli di cardinalità sulle relazioni
- Vincoli di cardinalità sugli attributi
- Vincoli di identificazione d'entità
- Altri vincoli (esterni)

Vincoli di cardinalità sulle relazioni

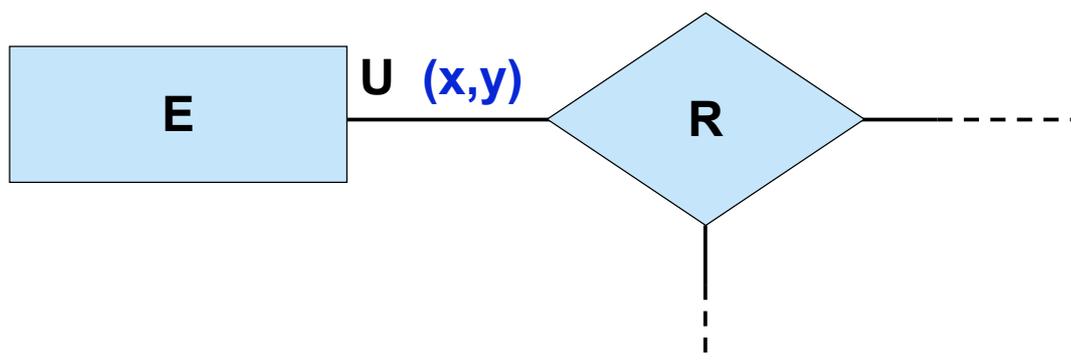
Un **vincolo di cardinalità** si associa ad un ruolo U (corrispondente ad una entità E) in una relazione R, ed impone un limite minimo ed un limite massimo di istanze della relazione a cui può partecipare ogni istanza dell'entità E nel ruolo U.

Serve a caratterizzare meglio il significato di una relazione.

Sintassi dei vincoli di cardinalità

Un vincolo di cardinalità si esprime mediante **una coppia (x,y) associata al ruolo U della relazione R**, dove:

- **x** è la cardinalità minima, cioè un **intero ≥ 0**
- **y** è la cardinalità massima, che è
 - "n", oppure
 - un **intero positivo $\geq x$**



Semantica dei vincoli di cardinalità

Se in uno schema **S** è definito un vincolo di cardinalità (x,y) associato ad un ruolo **U** (corrispondente ad una entità **E**) in una relazione **R**, allora in ogni istanza **I** dello schema **S**, per ogni **e** in $istanze(I,E)$, il numero di istanze di **R** che hanno **e** come componente nel ruolo **U** è:

- maggiore o uguale a x
- minore o uguale a y (se y è diverso da “n”)

Attenzione: il vincolo di cardinalità si definisce su una relazione, ma in realtà stabilisce una condizione che deve valere per tutte le istanze della entità a cui è associato il vincolo stesso (tramite il ruolo)

Esempio di vincolo di cardinalità

- Ad ogni impiegato sono assegnati da 1 a 5 incarichi
- Ogni incarico è assegnato ad al più 50 impiegati



Istanza:

$istanze(Impiegato) = \{ a,b,c \}$

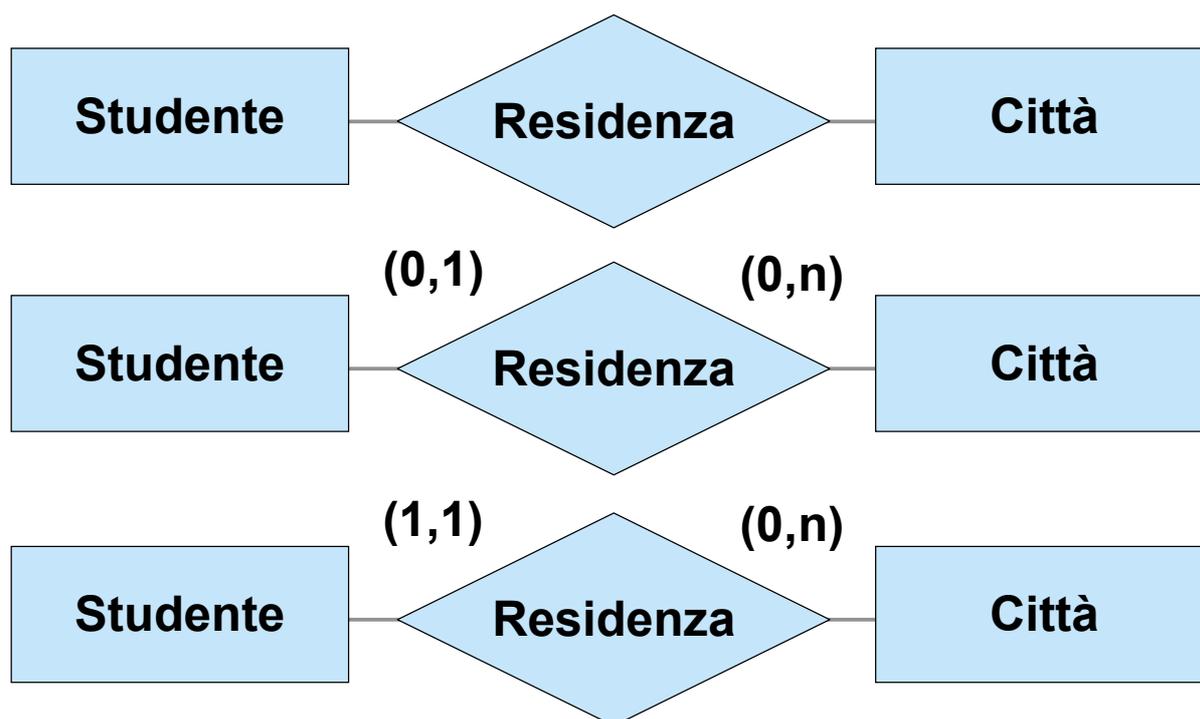
$istanze(Incarico) = \{ x,y,v,w,z \}$

$istanze(Assegnazione) = \{ (a,w), (b,v), (b,w), (c,y), (c,w), (c,z) \}$

Cardinalità maggiormente utilizzate

- Le cardinalità di interesse sono tipicamente tre: 0, 1, n.
- Per la **cardinalità minima**:
 - 0 significa “partecipazione opzionale”
 - 1 significa “partecipazione obbligatoria”
- Per la **cardinalità massima**:
 - 1 significa: “l’entità partecipa al più una volta sola alla relazione”
 - n significa: “l’entità partecipa un numero qualsiasi di volte alla relazione”
- La **mancaza del vincolo di cardinalità** è equivalente alla coppia (0,n)

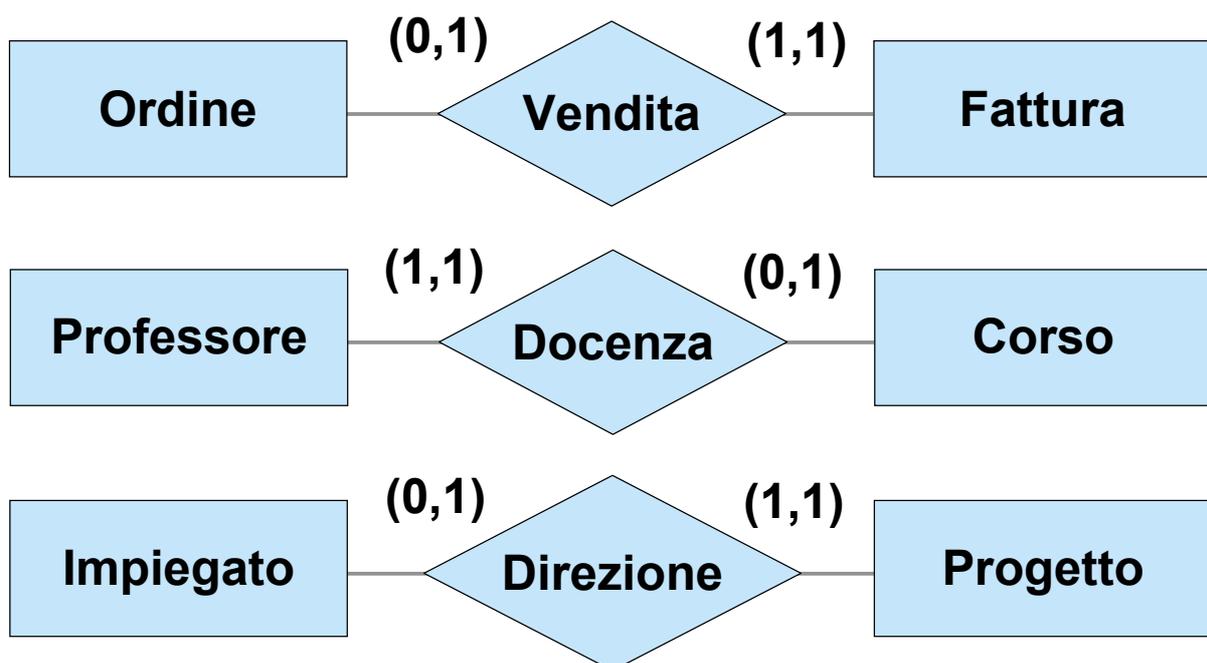
Esempi di vincolo di cardinalità



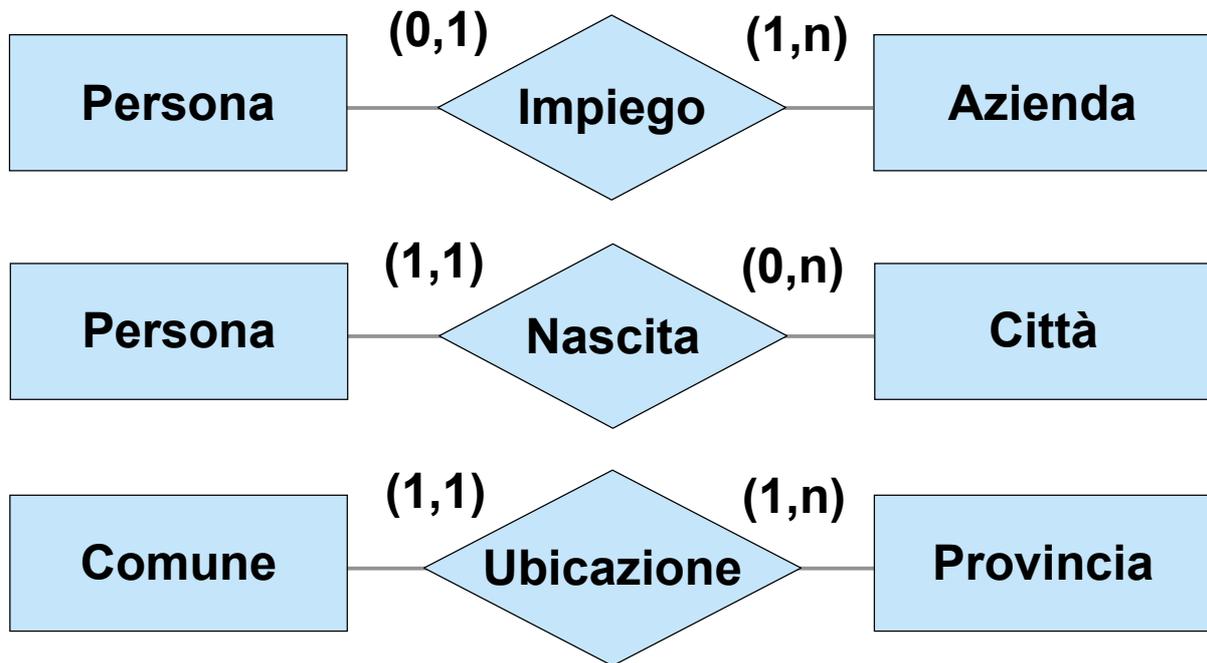
Classificazione di relazioni binarie

- Discutiamo la cardinalità massime sulle relazioni binarie
- Con riferimento alle cardinalità **massime**, le relazioni binarie possono essere di tipo:
 - uno a uno
 - uno a molti
 - molti a molti

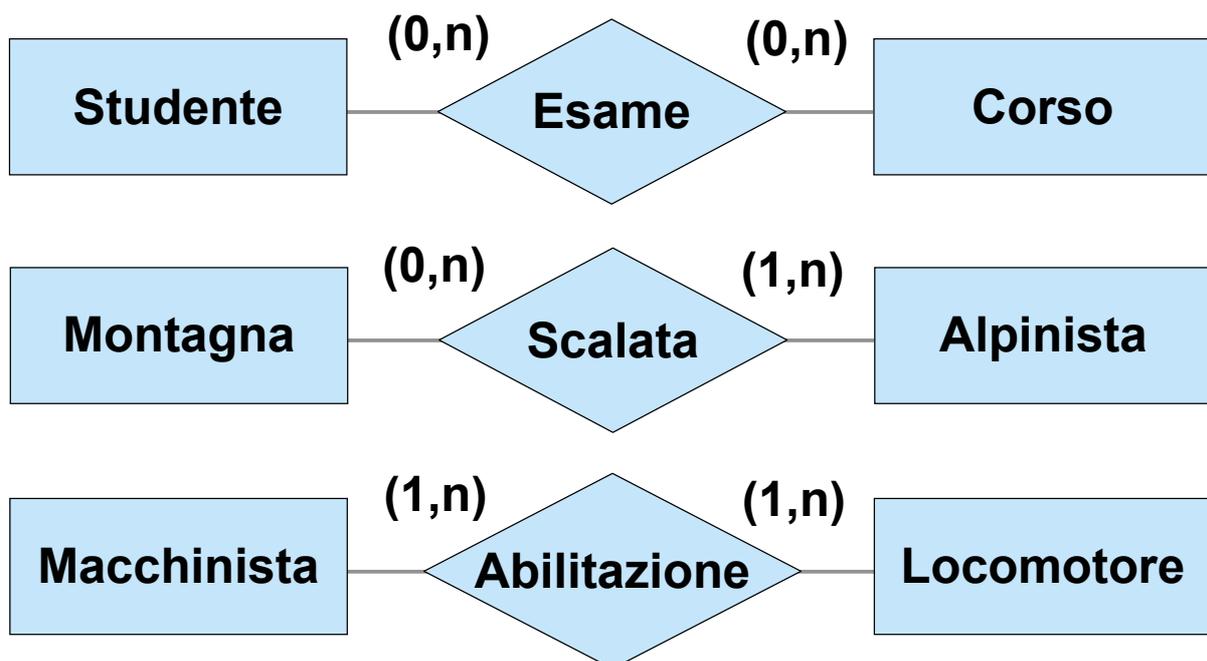
Relazioni binarie “uno a uno”



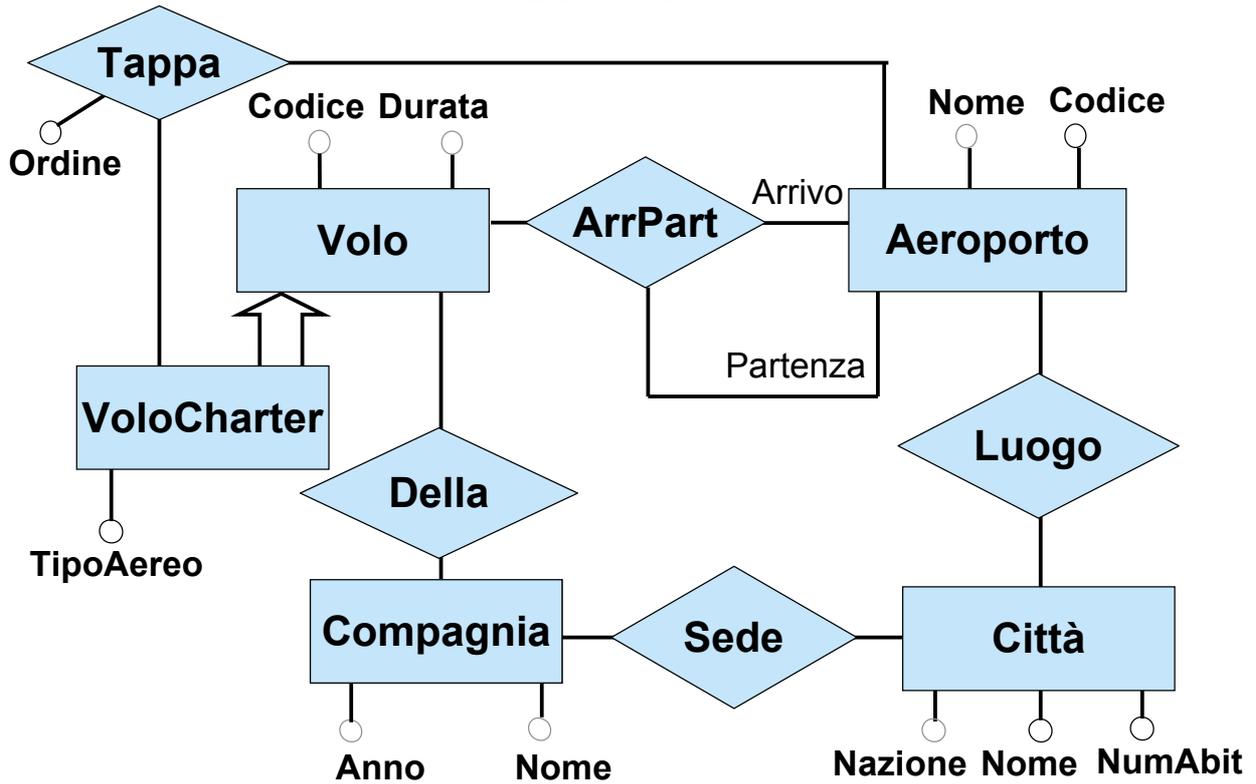
Relazioni binarie “uno a molti”



Relazioni binarie “molti a molti”



Esercizio 15: aggiungere le cardinalità

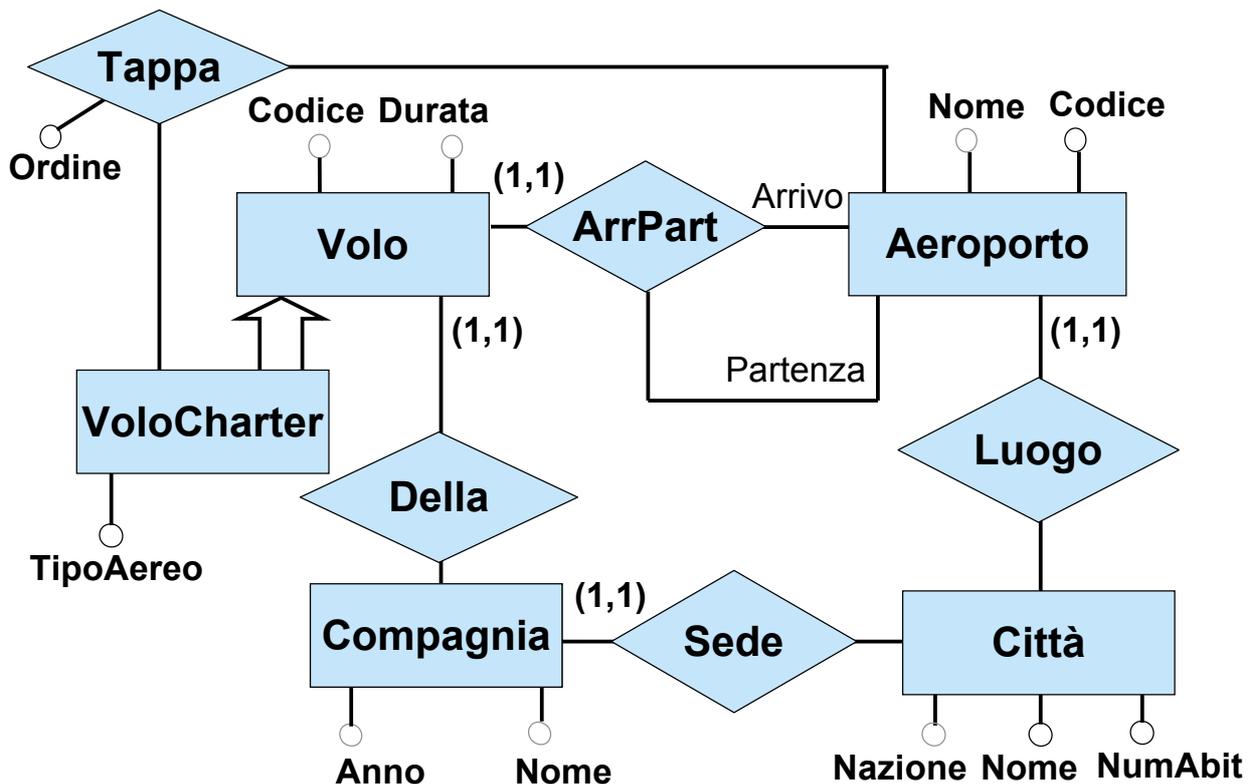


Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 127

Esercizio 15: soluzione

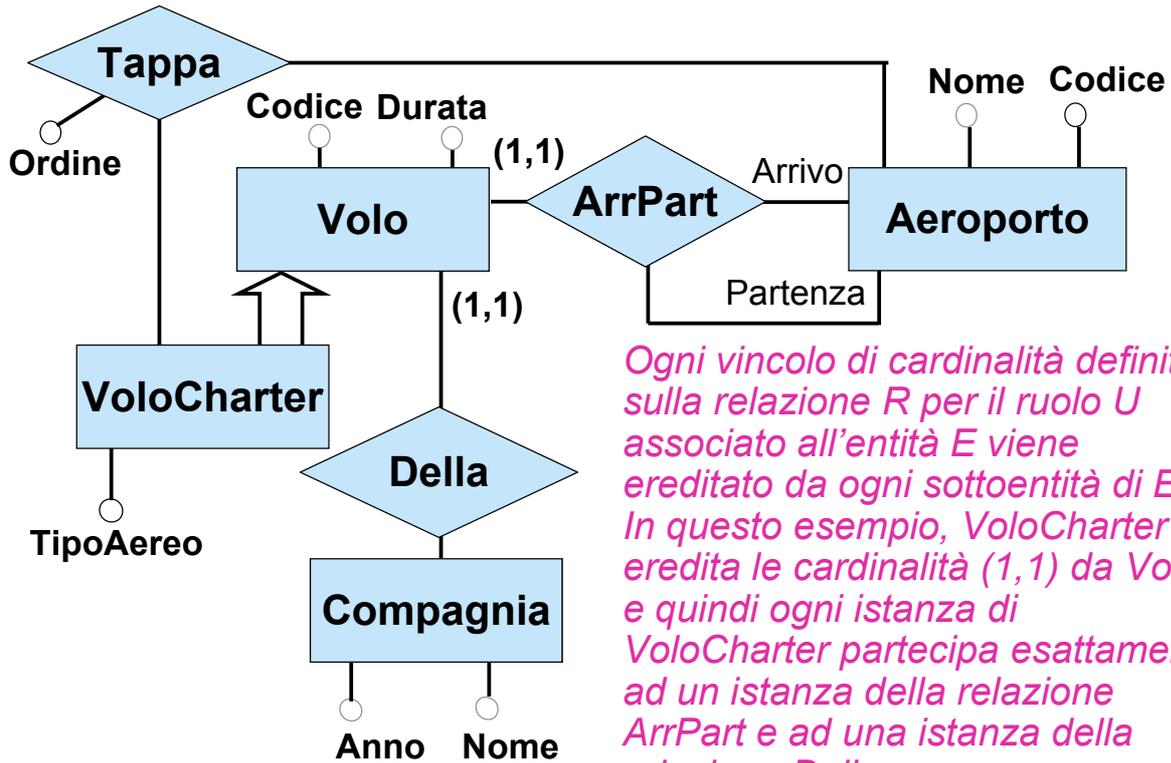


Giuseppe De Giacomo

Basi di Dati

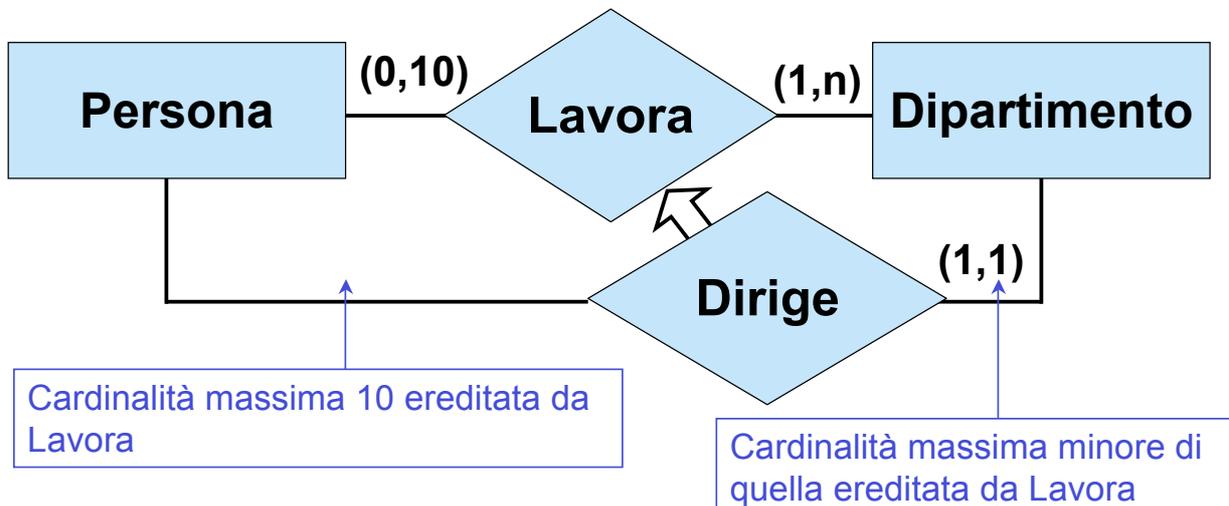
Progettazione concettuale - 128

Vincoli di cardinalità ed ereditarietà su entità



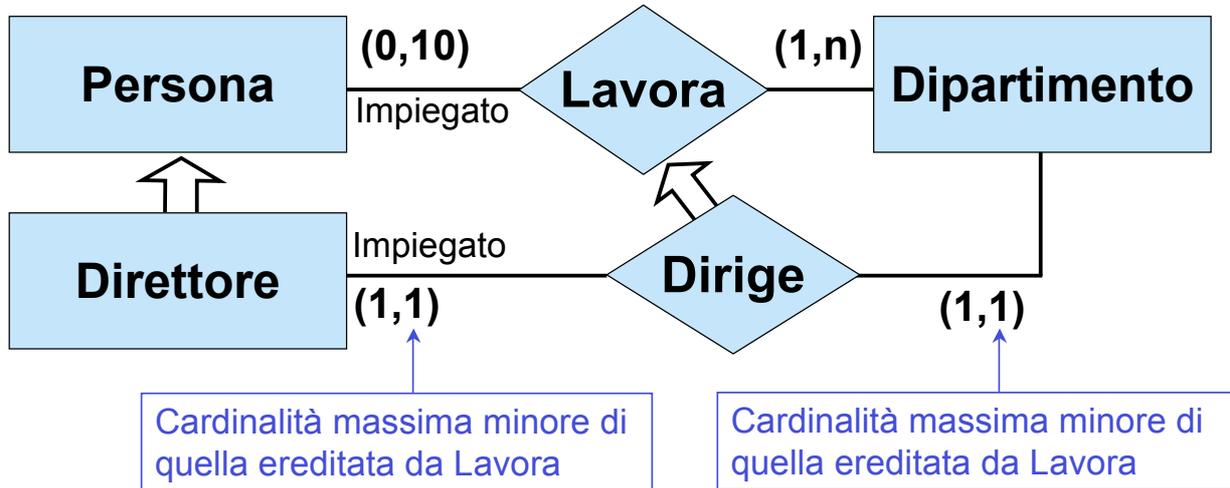
Vincoli di cardinalità ed ereditarietà su relazioni

Se Q ISA R , ogni vincolo di cardinalità **massima** definito sulla relazione R (per il ruolo U) viene ereditato dalla relazione Q (per lo stesso ruolo U), che può però avere cardinalità massima più stringente di quella ereditata, ovvero **cardinalità massima minore**.



Vincoli di cardinalità ed ereditarietà su relazioni

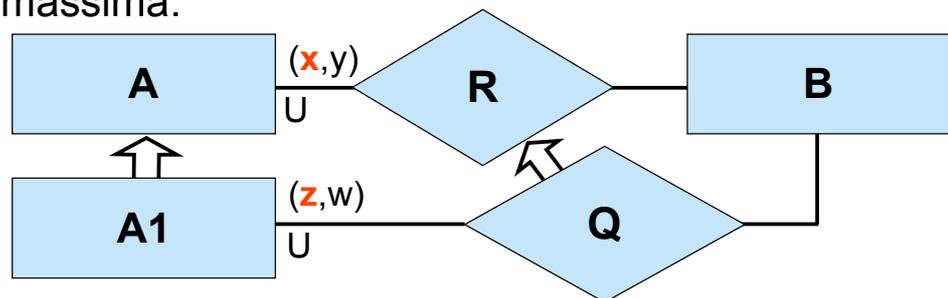
La regola sulla ereditarietà della cardinalità massima vale anche se l'entità corrispondente al ruolo U nella relazione figlia è diversa dalla (cioè è una sottoentità della) entità corrispondente al ruolo U nella relazione padre



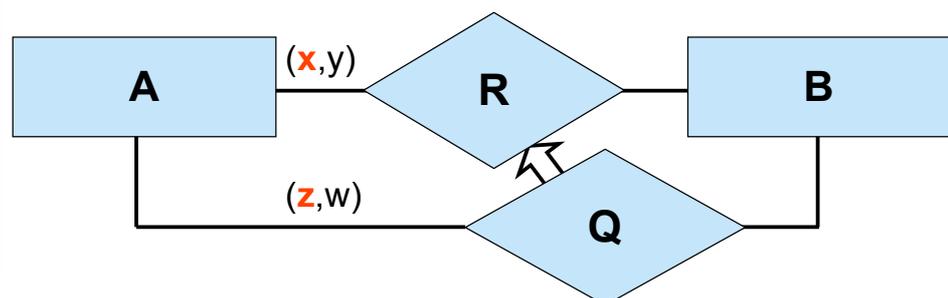
Vincoli di cardinalità ed ereditarietà su relazioni

Se Q ISA R , il rapporto tra i vincoli di cardinalità **minima** su Q ed R sono più complessi rispetto al caso dei vincoli di cardinalità massima.

x e z sono totalmente scorrelati



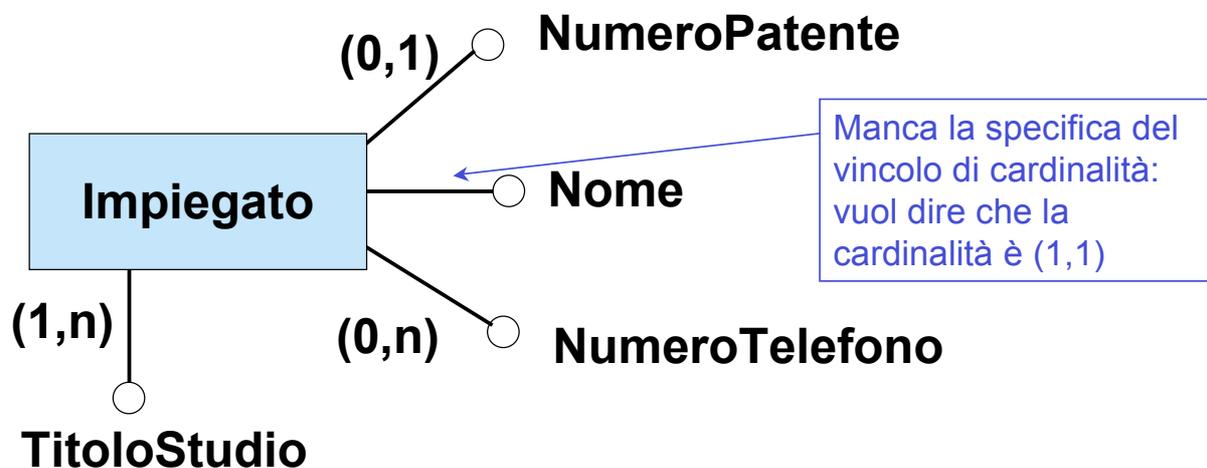
Se x è minore di z, si può sostituire x con z ed ottenere uno schema equivalente più accurato



Vincoli di cardinalità sugli attributi

- E' possibile definire vincoli di cardinalità anche sugli attributi, con due scopi:
 - indicare **opzionalità**
 - indicare **attributi multivalore**
- La semantica di un attributo con cardinalità diverso da (1,1) è diversa dal caso di cardinalità (1,1):
 - Quando la cardinalità massima di un attributo non è 1, l'attributo si dice multivalore, e non rappresenta più una funzione ma una **relazione**; le cardinalità stabiliscono il numero minimo e massimo di valori che l'attributo associa ad ogni istanza dell'entità
 - Quando la cardinalità minima di un attributo è 0, la relazione (o la funzione, se la cardinalità massima è uguale a 1) rappresentata dall'attributo **non è più totale**
- Se la specifica del vincolo manca, si intende che la cardinalità dell'attributo sia (1,1), e la semantica è quella usuale

Rappresentazione grafica dei vincoli di cardinalità sugli attributi



Osservazione sulla semantica degli attributi

Se in uno schema **S** un attributo **A** è definito in n entità E_1, \dots, E_n , rispettivamente con domini D_1, \dots, D_n , e con cardinalità $(x_1, y_1), \dots, (x_n, y_n)$, in ogni istanza dello **I** dello schema **S**, **A** è una relazione (o una funzione se la cardinalità massima è sempre 1) del tipo

$$\text{istanze}(\mathbf{I}, \mathbf{A}) \subseteq (\text{istanze}(\mathbf{I}, E_1) \cup \dots \cup \text{istanze}(\mathbf{I}, E_n)) \times (D_1 \cup \dots \cup D_n)$$

e tale che, per ogni i vale la seguente condizione:

- se $a \in \text{istanze}(\mathbf{I}, E_i)$, allora il numero di coppie (a, b) in $\text{istanze}(\mathbf{I}, \mathbf{A})$ è soggetto ai vincoli di cardinalità (x_i, y_i) , e per ogni $(a, b) \in \text{istanze}(\mathbf{I}, \mathbf{A})$, si ha che $b \in D_i$

Da ciò segue che occorre porre attenzione ai nomi assegnati agli attributi: quando si assegna lo stesso nome a due attributi di entità diverse, si deve ricordare che questi, a livello estensionale, rappresenteranno un'unica relazione.

Questa osservazione vale anche per uno stesso attributo definito su relazioni diverse.

Vincoli di identificazione di entità

- Un identificatore di una entità è un insieme di proprietà (attributi o relazioni) che permettono di identificare univocamente le istanze di un'entità. In altre parole non esistono due istanze di una data entità che assumono lo stesso valore per tutte le proprietà che formano l'identificatore.
- Un **vincolo di identificazione** per una entità E definisce un identificatore per E
- Su ogni entità si può definire un numero qualunque di vincoli di identificazione (tipicamente almeno uno)

Sintassi dei vincoli di identificazione di entità

- Un identificatore di una entità E può essere
 - **Interno**, ossia formato solo da attributi di E
 - **Esterno**, ossia formato da attributi di E e da ruoli di relazioni che coinvolgono E , oppure solo da ruoli di relazioni che coinvolgono E con la regola che tutti gli attributi e tutti i ruoli di relazione che concorrono ad un identificatore di entità devono avere cardinalità $(1,1)$
- Notazione per gli identificatori interni:
 - Se l'identificatore è formato da un solo attributo, si annerisce il corrispondente pallino.
 - Se l'identificatore è formato da più attributi, si uniscono gli attributi con una linea che termina con pallino annerito.
- Notazione per gli identificatori esterni:
 - Se l'identificatore è formato da attributi e relazioni (o meglio ruoli), si indica unendo gli attributi ed i ruoli con una linea che termina con pallino annerito.

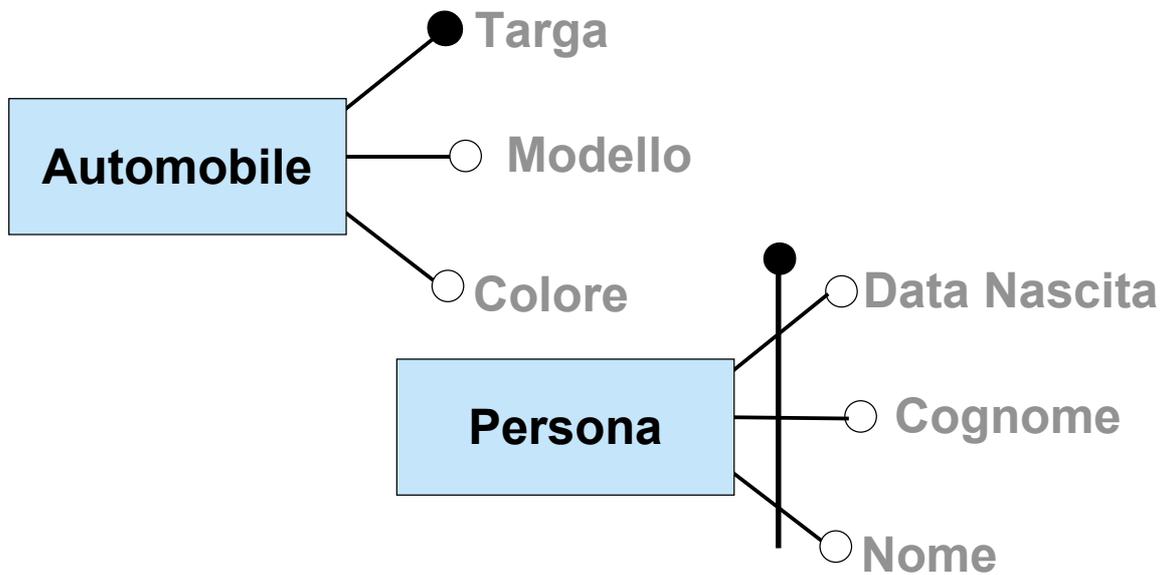
Semantica dei vincoli di identificazione di entità

Se in uno schema S è definito un vincolo di identificazione che specifica un identificatore per l'entità E formato

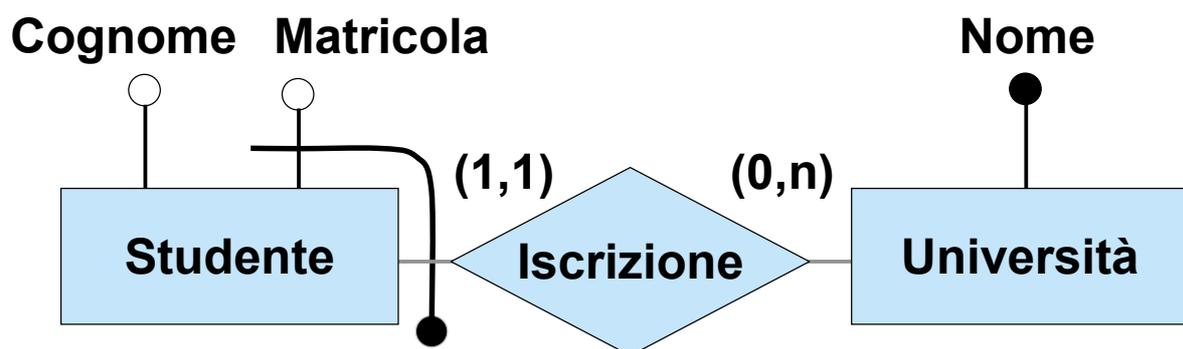
- dagli attributi A_1, A_2, \dots, A_k
- dalle relazioni R_1 (ruolo U_1), R_2 (ruolo U_2), ..., R_h (ruolo U_h)

allora in ogni istanza I dello schema S , prese due diverse istanze qualunque e_1 ed e_2 in $istanze(I, E)$, esse differiscono per il valore di almeno un A_i o per la partecipazione con ruolo U_i in almeno una R_i

Esempi di identificatori interni

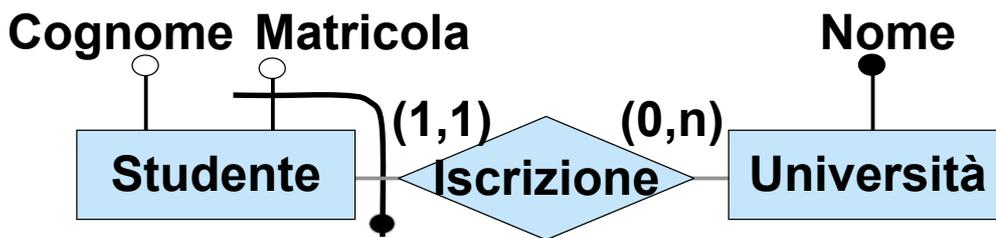


Esempio di identificatore esterno



Nota: Un'entità che ha un identificatore esterno viene detta **entità debole**

Identificatore esterno: livello estensionale



Istanza I:

$istanze(\text{Studente}) = \{ a, b, c \}$

$istanze(\text{Università}) = \{ v, z \}$

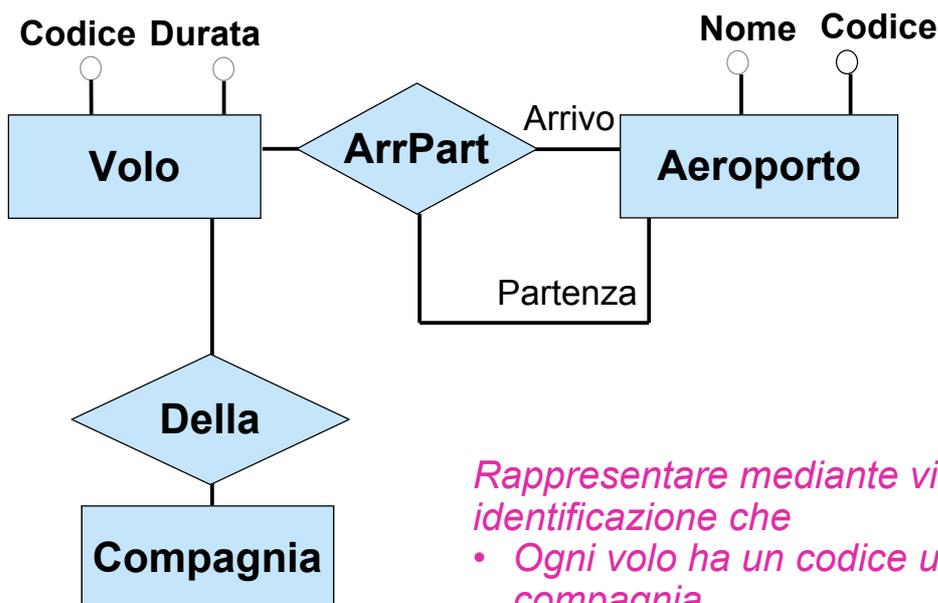
$istanze(\text{Matricola}) = \{ (a,12), (b,24), (c,12) \}$

$istanze(\text{Cognome}) = \{ (a, \text{'Rossi'}), (b, \text{'Finzi'}), (c, \text{'Rossi'}) \}$

$istanze(\text{Nome}) = \{ (v, \text{'Viterbo'}), (z, \text{'Camerino'}) \}$

$istanze(\text{Iscrizione}) = \{ (a,v), (b,v), (c,z) \}$

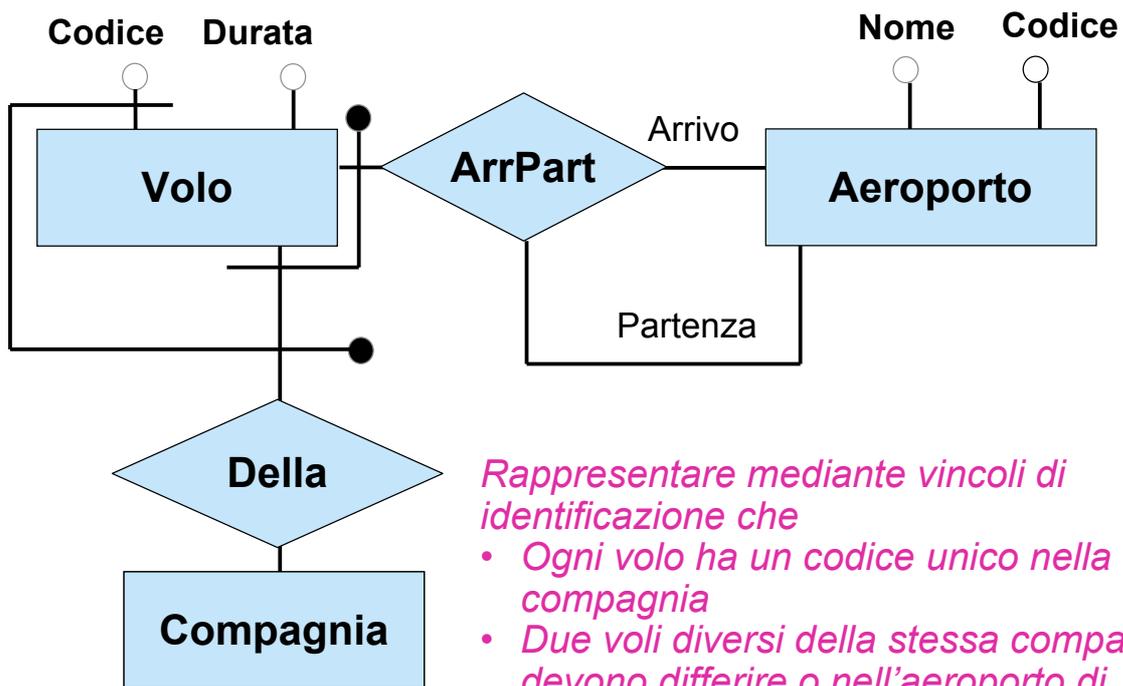
Esercizio 16: vincoli di identificazione



Rappresentare mediante vincoli di identificazione che

- Ogni volo ha un codice unico nella compagnia
- Due voli diversi della stessa compagnia devono differire o nell'aeroporto di arrivo o nell'aeroporto di partenza

Esercizio 16: soluzione

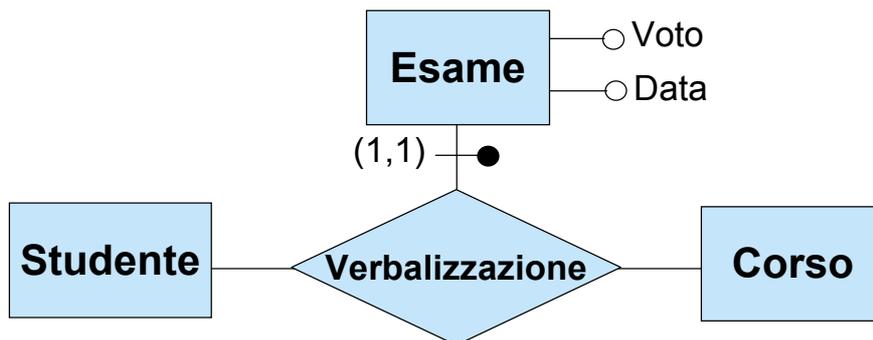


Rappresentare mediante vincoli di identificazione che

- Ogni volo ha un codice unico nella compagnia
- Due voli diversi della stessa compagnia devono differire o nell'aeroporto di arrivo o nell'aeroporto di partenza

Esempi di identificazione esterna

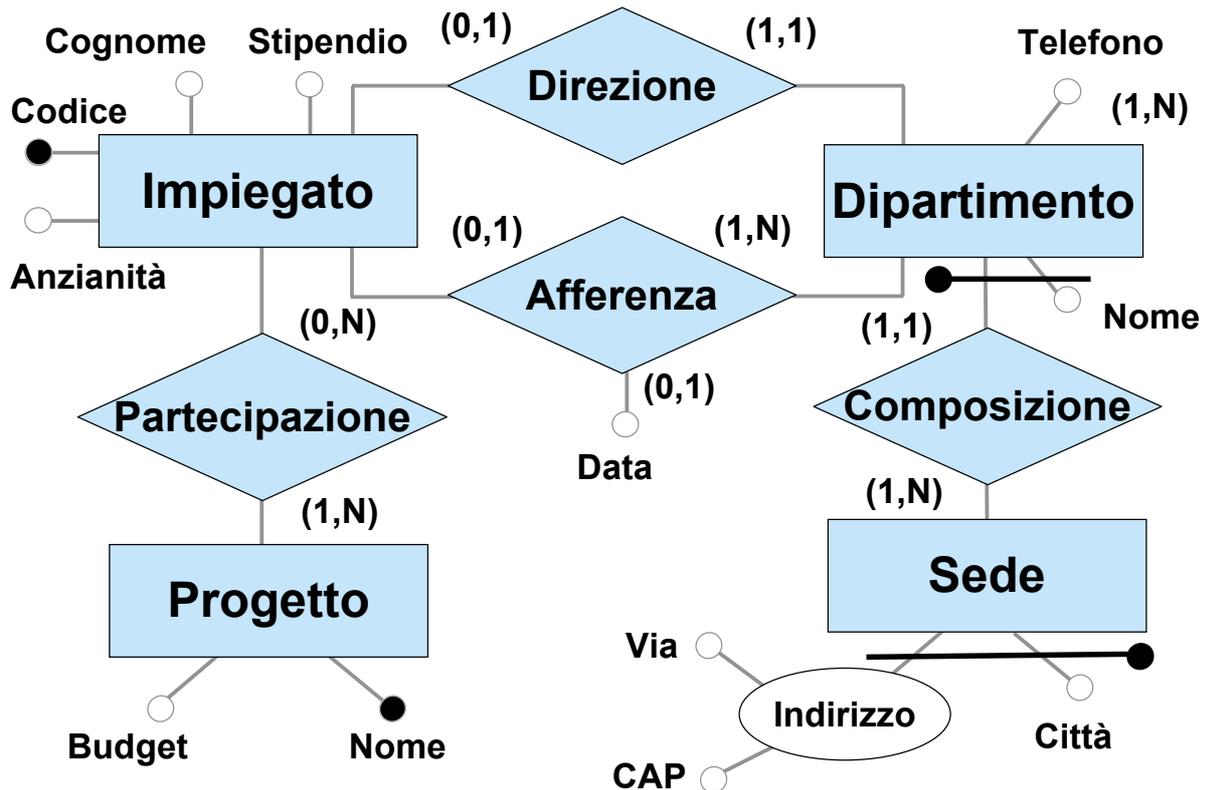
Un identificatore esterno può anche non comprendere attributi



Una entità che partecipa con ruolo U e con cardinalità (1,1) ad una relazione R nella quale almeno uno degli altri ruoli hanno cardinalità massima 1, è implicitamente identificata esternamente da R tramite U. Ad esempio,



Esempio di vincoli di cardinalità e di identificaz.



Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 145

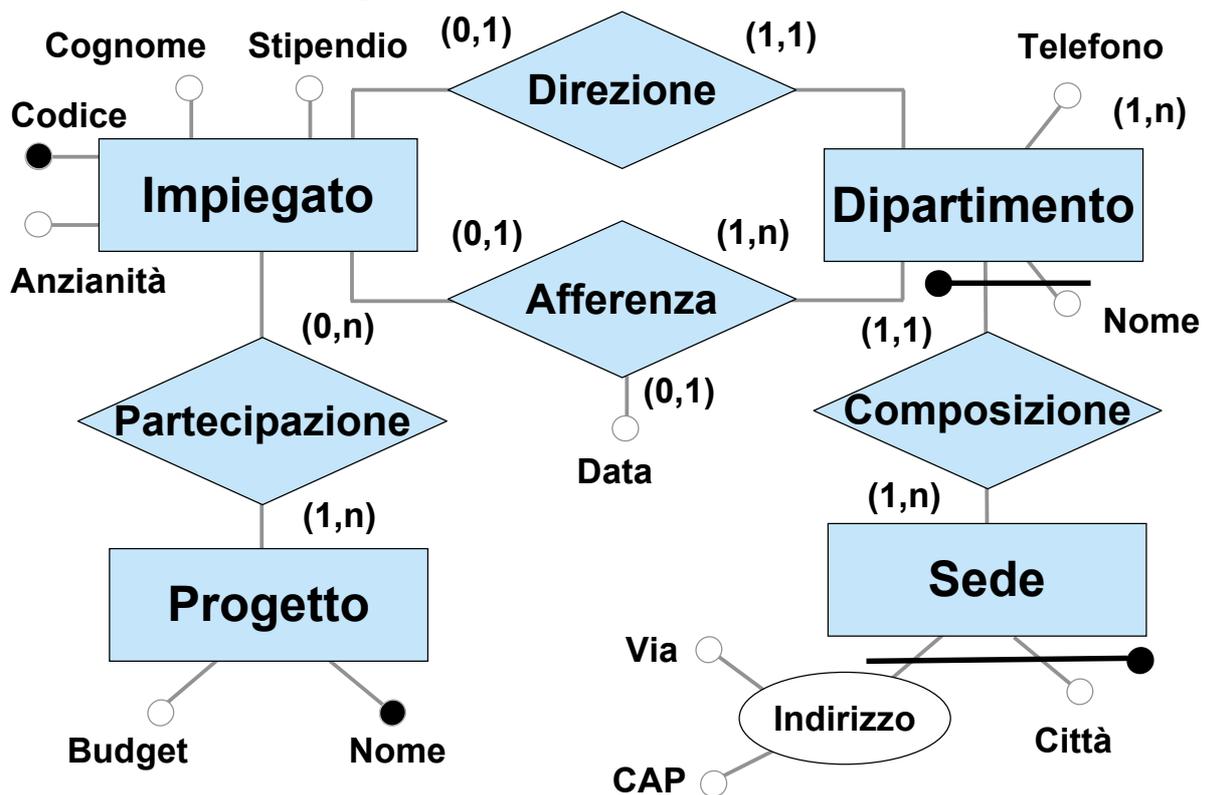
Vincoli non esprimibili nel diagramma ER

- Gli schemi ER permettono di cogliere la maggior parte delle interrelazioni tra i dati del dominio d'interesse
- Tuttavia alcune interrelazioni non possono essere colte direttamente da uno schema ER
- Tali interrelazioni vanno in ogni caso tenute presenti attraverso delle asserzioni aggiuntive dette **vincoli esterni al diagramma**, o semplicemente **vincoli esterni**

Vincoli non esprimibili nel diagramma ER

- Come rappresentiamo tali vincoli?
- Attraverso formalismi opportuni (es, in logica matematica)
- Attraverso delle asserzioni in linguaggio naturale (che devono essere il più possibile precise e non ambigue)

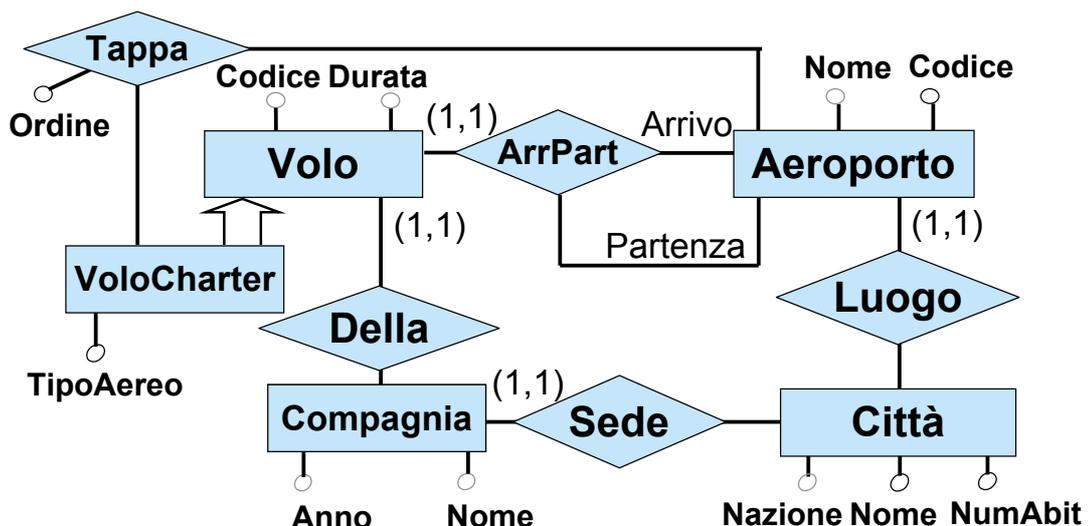
Esempio di schema concettuale ...



... con vincoli esterni

Vincoli di integrità esterni
(1) Il direttore di un dipartimento deve afferire a tale dipartimento da almeno 5 anni
(2) Un impiegato non deve avere uno stipendio maggiore del direttore del dipartimento al quale afferisce
(3) Un dipartimento con sede a Roma deve essere diretto da un impiegato con più di dieci anni di anzianità
(4) Un impiegato non può partecipare ad un numero di progetti maggiore di due volte il numero di dipartimenti ai quali afferisce

Vincolo esterno sullo schema dell'esercizio 12



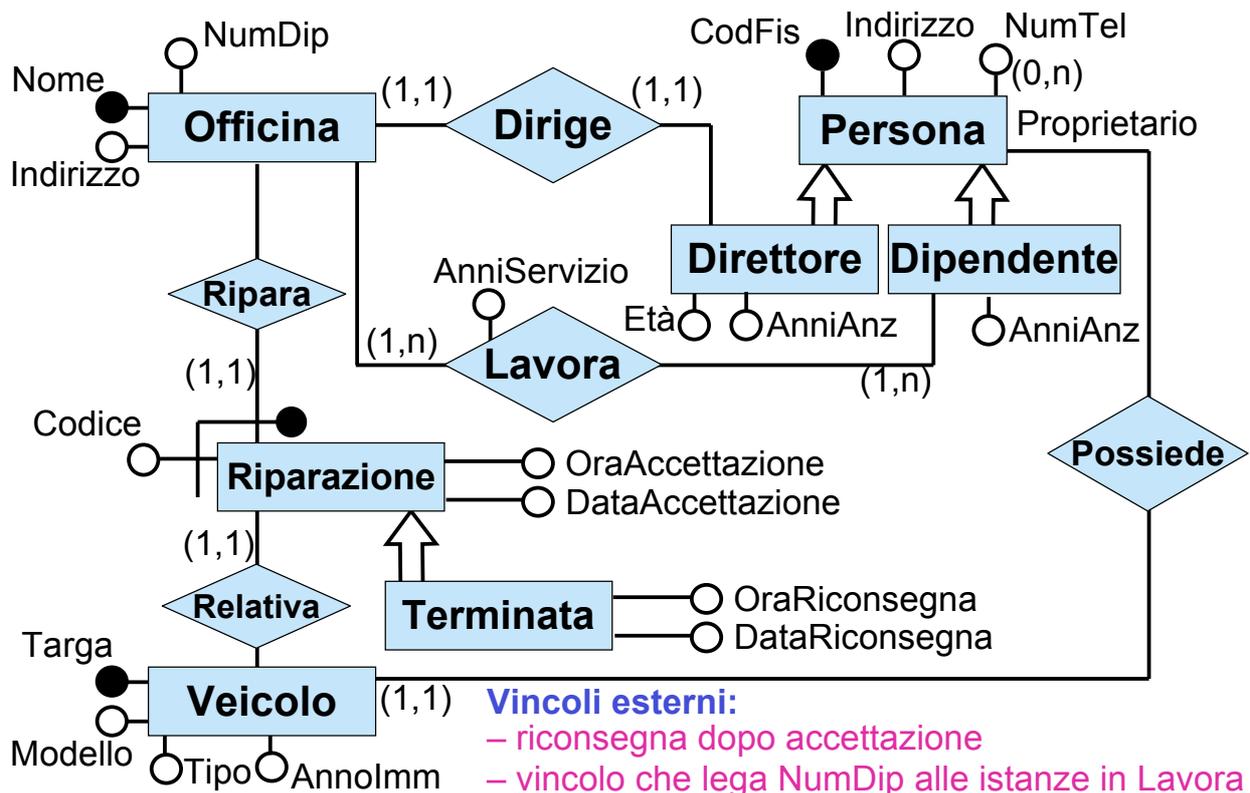
Vincolo esterno: Per ogni v in VoloCharter, se $(v, a_1), \dots, (v, a_n)$ sono tutte le coppie in Tappa alle quali partecipa v , e se o_1, \dots, o_n sono i valori assegnati a tali coppie dall'attributo Ordine, allora per ogni $i=1, \dots, n$, esiste un o_j tale che $o_j = i$.

Esercizio 17: progettazione concettuale

Descrivere lo schema concettuale corrispondente ad un'applicazione riguardante un insieme di officine, facendo riferimento alla seguenti specifiche.

Delle officine interessano: nome, indirizzo, numero di dipendenti (maggiore di 0), dipendenti (con l'informazione su quanti anni di servizio), e direttore. Si noti che ogni officina ha uno ed un solo direttore, ed un direttore dirige una ed una sola officina. Dei dipendenti e dei direttori interessano: codice fiscale, indirizzo, numeri di telefono, e anni di anzianità. Dei direttori interessa anche l'età. Si noti che un direttore non è necessariamente un dipendente di officina (ma può esserlo). Ogni riparazione è effettuata da una ed una sola officina, e riguarda uno ed un solo veicolo. Di ogni riparazione interessano: codice (univoco nell'ambito dell'officina), ora e data di accettazione del veicolo, e, nel caso di riparazione terminata, ora e data di riconsegna del veicolo. Dei veicoli interessano: modello, tipo, targa, anno di immatricolazione, e proprietario. Ogni veicolo ha uno ed un solo proprietario. Dei proprietari di veicoli interessano: codice fiscale, indirizzo, e numeri di telefono.

Esercizio 17: soluzione

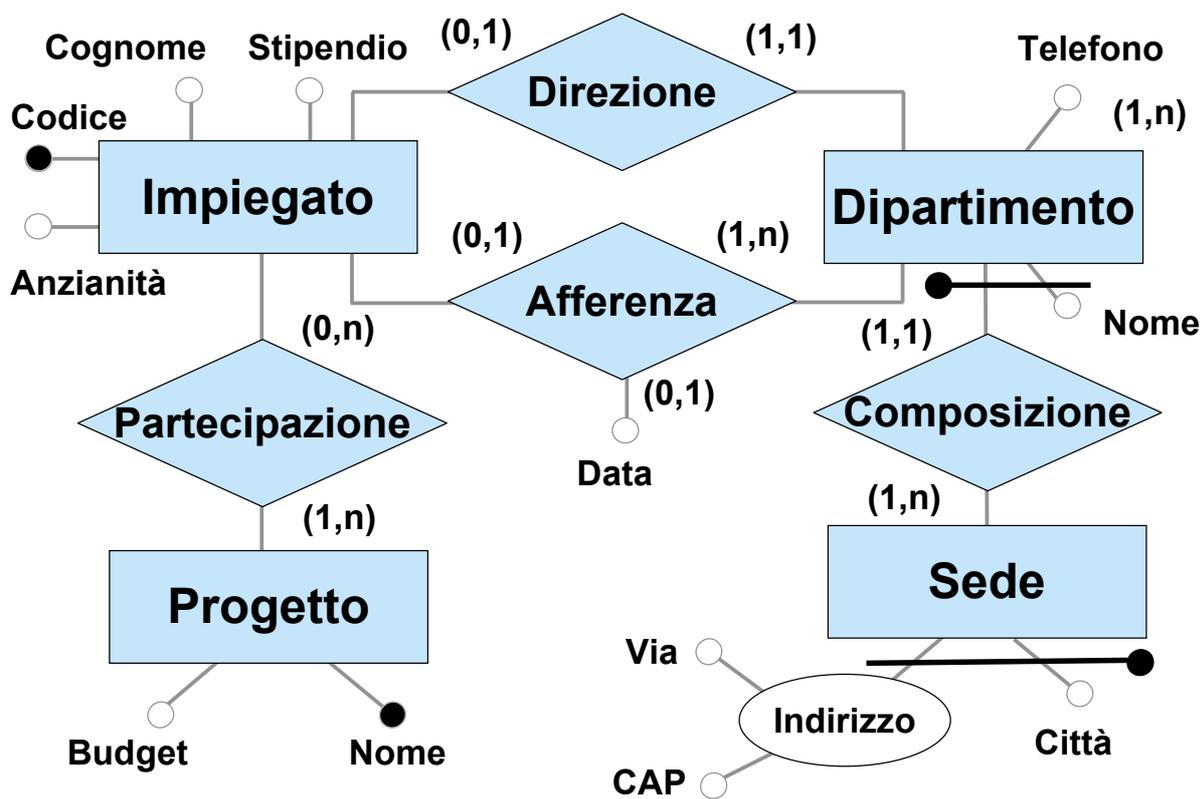


Documentazione associata agli schemi ER

Oltre al diagramma ER, lo schema concettuale è descritto dal cosiddetto **dizionario dei dati**

- Il dizionario dei dati è costituito dalle tabelle di
 - Entità
 - Relazioni
 - Attributi
 - Vincoli esterni

Esempio di schema concettuale



Dizionario dei dati: entità

Entità	Descrizione	Attributi	Identificatori
Impiegato	Dipendente dell'azienda	Codice Cognome Stipendio Anzianità	{ Codice }
Progetto	Progetti aziendali	Nome Budget	{ Nome }
Dipartimento	Struttura aziendale	Nome Telefono	{ Nome, Sede }
Sede	Sede dell'azienda	Città Indirizzo (Via, CAP)	{ Città, Indirizzo }

Dizionario dei dati: relazioni

Relazione	Descrizione	Componenti	Attributi
Direzione	Direzione di un dipartimento	Impiegato, Dipartimento	
Afferenza	Afferenza ad un dipartimento	Impiegato, Dipartimento	Data
Partecipazione	Partecipazione ad un progetto	Impiegato, Progetto	
Composizione	Composizione dell'Azienda	Dipartimento, Sede	

Dizionario dei dati: attributi

Attributo	Entità/Relazione	Dominio	Descrizione
Codice	Impiegato	Intero	Codice identificativo di impiegati
Cognome	Impiegato	Stringa	Cognome di impiegato
Stipendio	Impiegato	Reale	Stipendio di impiegato
Nome	Progetto	Stringa	Nome del progetto
...

Dizionario dei dati: vincoli esterni

Vincoli di integrità esterni
(1) Il direttore di un dipartimento deve afferire a tale dipartimento
(2) Un impiegato non deve avere uno stipendio maggiore del direttore del dipartimento al quale afferisce
(3) Un dipartimento con sede a Roma deve essere diretto da un impiegato con più di dieci anni di anzianità
(4) Un impiegato che non afferisce ad alcun dipartimento non deve partecipare ad alcun progetto

4. La progettazione concettuale

4.2 Metodologia per la progettazione concettuale

1. introduzione alla progettazione di basi di dati
2. modello Entità-Relazione
3. **metodologia per la progettazione concettuale**
4. esempio

Attività nella progettazione concettuale

La progettazione concettuale è preceduta dalla fase di raccolta (o acquisizione) dei requisiti

A partire dai requisiti raccolti, la progettazione concettuale comprende le seguenti **attività**:

1. strutturazione e organizzazione dei requisiti raccolti
2. costruzione del glossario (opzionale)
3. costruzione dello schema concettuale (diagramma e dizionario dei dati)
4. controllo di qualità

Queste attività sono tipicamente interconnesse

Commenti sulla raccolta dei requisiti

La raccolta dei requisiti è un'attività difficile, per cui sono state studiate e proposte metodologie specifiche. Noi faremo solo alcune considerazioni di carattere generale.

Scopo della raccolta dei requisiti: produrre documentazione che descriva in modo “corretto” e “completo” i requisiti della base di dati.

Possibili fonti di requisiti:

- **utenti**, attraverso:
 - interviste
 - documentazione apposita
- **documentazione esistente:**
 - normative (leggi, regolamenti di settore)
 - regolamenti interni, procedure aziendali
 - realizzazioni preesistenti
- **modulistica**

Raccolta attraverso interviste

Osservazioni:

- utenti diversi possono fornire informazioni diverse
- utenti a livello più alto hanno spesso una visione più ampia ma meno dettagliata
- le interviste portano spesso ad una acquisizione dei requisiti “**per raffinamenti successivi**”

Alcune regole per l'interazione con gli utenti:

- effettuare spesso verifiche di comprensione e coerenza
- verificare anche per mezzo di esempi (generalisti e relativi a casi limite)
- richiedere definizioni e classificazioni
- far evidenziare gli aspetti essenziali rispetto a quelli marginali

Documentazione descrittiva dei requisiti

Alcune regole generali:

- scegliere il corretto livello di astrazione
- standardizzare la struttura delle frasi
- suddividere le frasi articolate
- separare le frasi sui dati da quelle sulle funzioni

Fase 1: Strutturazione e organizzazione dei requisiti

I requisiti raccolti devono venire strutturati e organizzati per gli scopi dell'analisi.

Alcune regole generali per l'organizzazione di termini e concetti:

- costruire un glossario dei termini
- individuare omonimi e sinonimi e unificare i termini
- rendere esplicito il riferimento fra termini
- riorganizzare le frasi per concetti

Esempio di raccolta dei requisiti: base di dati bibliografica

Si procede per **raffinamenti successivi**.

Primo raffinamento:

Si vogliono organizzare i dati di interesse per automatizzare la gestione dei riferimenti bibliografici.

Raccolta requisiti: base di dati bibliografica (2)

Secondo raffinamento:

Si vogliono organizzare i dati di interesse per automatizzare la gestione dei riferimenti bibliografici, **con tutte le informazioni da riportarsi in una bibliografia. Per ogni pubblicazione deve esistere un codice identificante costituito da sette caratteri, indicanti le iniziali degli autori, l'anno di pubblicazione e un carattere aggiuntivo per la discriminazione delle collisioni.**

Raccolta requisiti: base di dati bibliografica (3)

Terzo raffinamento:

Si vogliono organizzare i dati di interesse per automatizzare la gestione dei riferimenti bibliografici, con tutte le informazioni da riportarsi in una bibliografia. **Le pubblicazioni sono di due tipi, monografie (per le quali interessano editore, data e luogo di pubblicazione) e articoli su rivista (con nome della rivista, volume, numero, pagine e anno di pubblicazione); per entrambi i tipi si debbono ovviamente riportare i nomi degli autori.** Per ogni pubblicazione deve esistere un codice identificante costituito da sette caratteri, indicanti le iniziali degli autori, l'anno di pubblicazione e un carattere aggiuntivo per la discriminazione delle collisioni.

Esempio di organizzazione dei requisiti: società di formazione

Si vuole realizzare una base di dati per una società che eroga corsi, di cui vogliamo rappresentare i dati dei partecipanti ai corsi e dei docenti. Per gli studenti (circa 5000), identificati da un codice, si vuole memorizzare il codice fiscale, il cognome, l'età, il sesso, il luogo di nascita, il nome dei loro attuali datori di lavoro, i posti dove hanno lavorato in precedenza insieme al periodo, l'indirizzo e il numero di telefono, i corsi che hanno frequentato (i corsi sono in tutto circa 200) e il giudizio finale.

Rappresentiamo anche i seminari che stanno attualmente frequentando e, per ogni giorno, i luoghi e le ore dove sono tenute le lezioni. I corsi hanno un codice, un titolo e possono avere varie edizioni con date di inizio e fine e numero di partecipanti. Se gli studenti sono liberi professionisti, vogliamo conoscere l'area di interesse e, se lo possiedono, il titolo. Per quelli che lavorano alle dipendenze di altri, vogliamo conoscere invece il loro livello e la posizione ricoperta.

Per gli insegnanti (circa 300), rappresentiamo il cognome, l'età, il posto dove sono nati, il nome del corso che insegnano, quelli che hanno insegnato nel passato e quelli che possono insegnare. Rappresentiamo anche tutti i loro recapiti telefonici. I docenti possono essere dipendenti interni della società o collaboratori esterni.

Strutturazione dei requisiti in gruppi di frasi omogenee (1)

Frase di carattere generale:

Si vuole realizzare una base di dati per una società che eroga corsi, di cui vogliamo rappresentare i dati dei partecipanti ai corsi e dei docenti.

Frase relative ai partecipanti:

Per i partecipanti (circa 5000), identificati da un codice, rappresentiamo il codice fiscale, il cognome, l'età, il sesso, la città di nascita, i nomi dei loro attuali datori di lavoro e di quelli precedenti (insieme alle date di inizio e fine rapporto), le edizioni dei corsi che stanno attualmente frequentando e quelli che hanno frequentato nel passato, con la relativa votazione finale in decimi.

Strutturazione dei requisiti in gruppi di frasi omogenee (2)

Frase relative ai datori di lavoro:

Relativamente ai datori di lavoro presenti e passati dei partecipanti, rappresentiamo il nome, l'indirizzo e il numero di telefono.

Frase relative ai corsi:

Per i corsi (circa 200), rappresentiamo il titolo e il codice, le varie edizioni con date di inizio e fine e, per ogni edizione, rappresentiamo il numero di partecipanti e il giorno della settimana, le aule e le ore dove sono tenute le lezioni.

Strutturazione dei requisiti in gruppi di frasi omogenee (3)

Frasi relative a tipi specifici di partecipanti:

Per i partecipanti che sono liberi professionisti, rappresentiamo l'area di interesse e, se lo possiedono, il titolo professionale. Per i partecipanti che sono dipendenti, rappresentiamo invece il loro livello e la posizione ricoperta.

Frasi relative ai docenti:

Per i docenti (circa 300), rappresentiamo il cognome, l'età, la città di nascita, tutti i numeri di telefono, il titolo del corso che insegnano, di quelli che hanno insegnato nel passato e di quelli che possono insegnare. I docenti possono essere dipendenti interni della società di formazione o collaboratori esterni.

Fase 2: Costruzione del glossario (opzionale)

Termine	Descrizione	Sinonimi	Collegamenti
Partecipante	Persona che partecipa ai corsi	Studente	Corso, Società
Docente	Docente dei corsi. Può essere esterno	Insegnante	Corso
Corso	Corso organizzato dalla società. Può avere più edizioni.	Seminario	Docente
Società	Ente presso cui i partecipanti lavorano o hanno lavorato	Posti	Partecipante

Fase 3: Costruzione dello schema concettuale

- Si parte dalla specifica dei requisiti raccolti, strutturati e organizzati, e dall'eventuale glossario
- Si fa uso del **modello Entità-Relazione** e si costruisce il diagramma e il dizionario dei dati
- Si utilizzano opportune **strategie** per sviluppare lo schema concettuale
- Nello sviluppare il progetto si tiene conto dei **fattori di qualità** che uno schema concettuale deve possedere

Rappresentazione dei concetti della specifica

Quale costrutto ER va utilizzato per rappresentare un concetto presente nelle specifiche?

Bisogna basarsi sulle significato dei costrutti del modello ER!

- se il concetto ha **proprietà significative** e descrive oggetti con **esistenza autonoma**:
 - **entità**
- se il concetto è un **valore semplice** e non ha proprietà:
 - **attributo**
- se il concetto **correla due o più** concetti:
 - **relazione**
- se il concetto è **caso particolare** di un altro:
 - **is-a** o **generalizzazione**

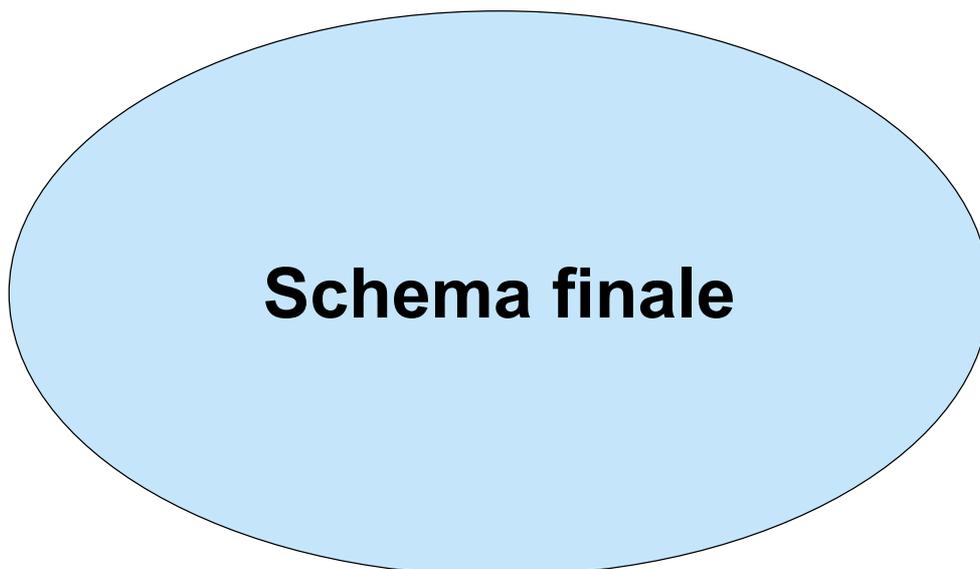
Strategie di progetto

Si possono adottare diverse **strategie di progetto**:

- top-down
- bottom-up
- inside-out
- mista

Ciascuna strategia prevede opportune **primitive di raffinamento** che specificano in che modo sostituire o integrare una parte dello schema con una versione più raffinata della stessa.

Strategia top-down

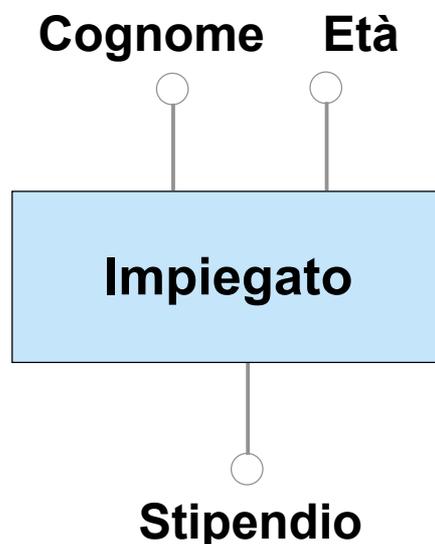


Primitive di raffinamento top-down

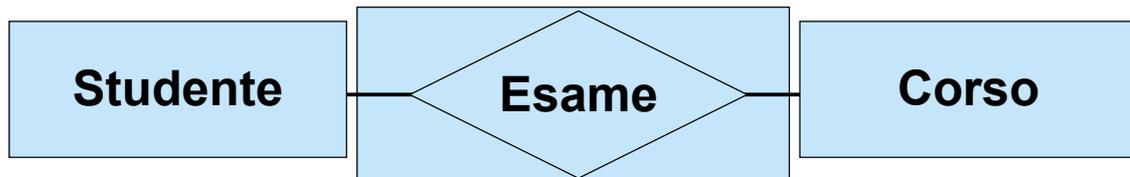
Si parte da uno **schema scheletro** (cioè scarno e generale) e si procede per raffinamenti successivi:

- introduzione di attributi di un'entità o una relazione
- introduzione degli identificatori di un'entità
- introduzione delle cardinalità di una relazione
- sostituzione di un'entità con più entità connesse da una relazione
- sostituzione di un'entità con una generalizzazione o con più entità connesse da legami is-a
- sostituzione di una relazione con più relazioni

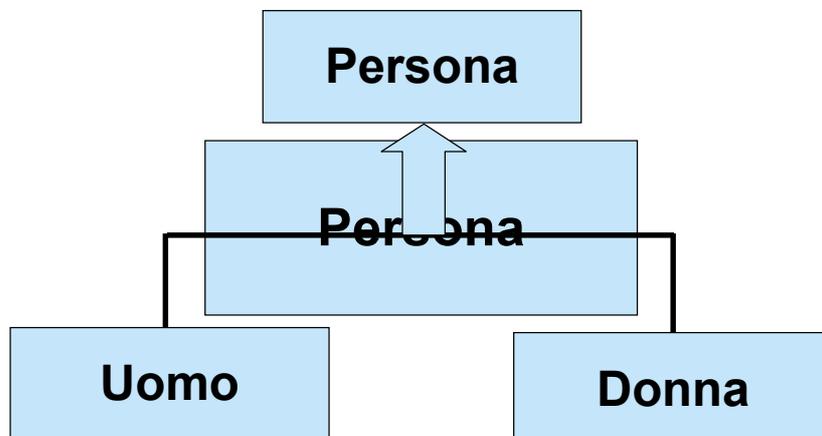
Primitive di raffinamento top-down: introduzione di attributi



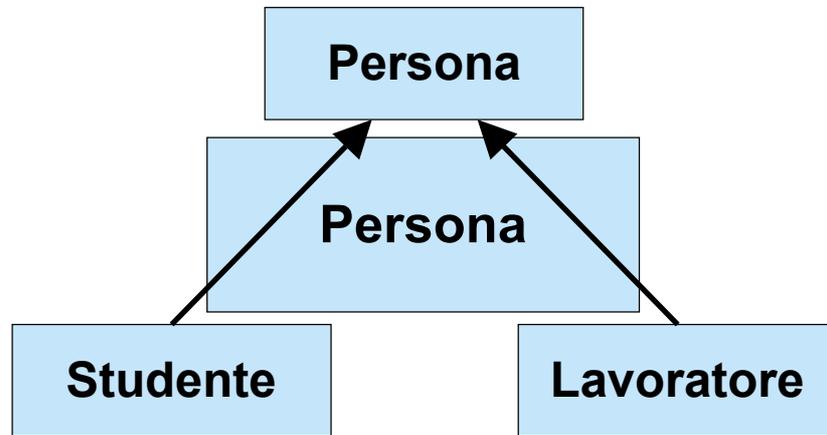
Primitive di raffinamento top-down: introduzione di una relazione



Primitive di raffinamento top-down: introduzione di una generalizzazione



Primitive di raffinamento top-down: introduzione di un is-a



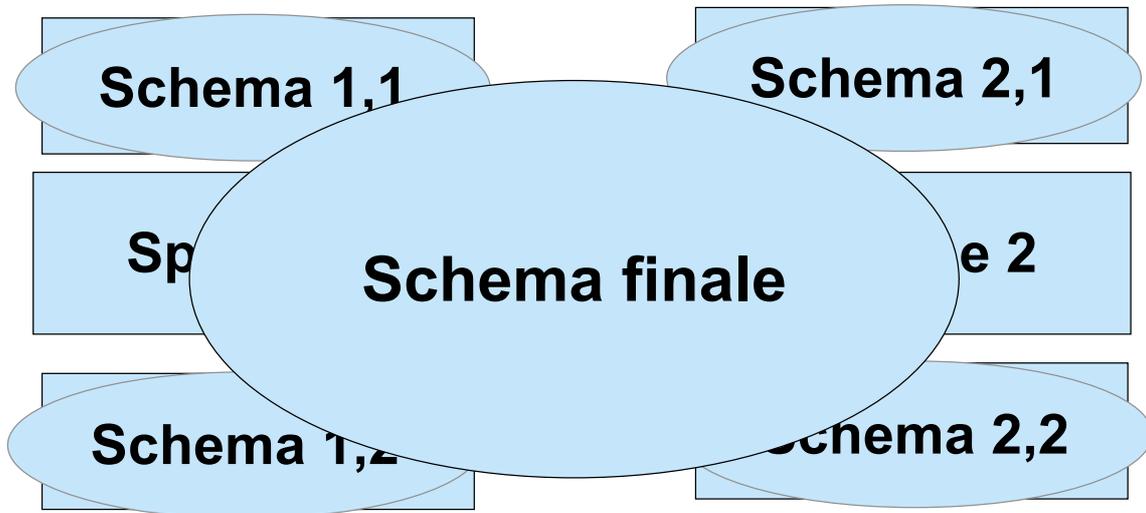
Aspetti metodologici nella costruzione top-down dello schema ER

Un metodo comunemente usato per la costruzione dello schema scheletro e per i successivi raffinamenti prevede i seguenti passi:

- Individua le entità di interesse
- Individua gli attributi delle entità
- Individua le relazioni tra le entità
- Individua gli attributi delle relazioni
- Determina le cardinalità di relazioni (e attributi)
- Determina gli identificatori delle entità
- Individua le generalizzazioni e le relazioni is-a, partendo o dall'entità più generale e scendendo nella gerarchia, oppure dalle entità più specifiche e risalendo nella gerarchia
- **Controllo di qualità**

Correggi,
modifica,
raffina,
estendi

Strategia bottom-up



Primitive di raffinamento bottom-up

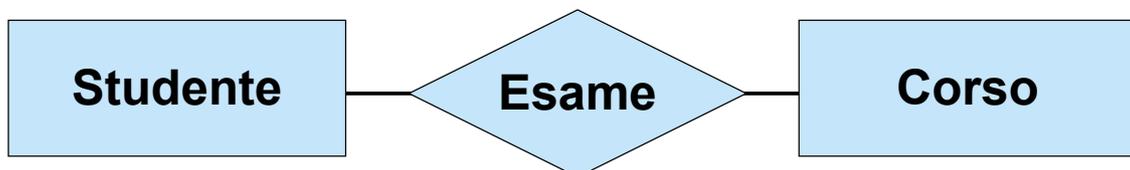
Si parte da elementi dello schema esistenti e si compongono ed integrano:

- introduzione di un'entità
- collegamento di entità con una relazione
- organizzazione di entità in una generalizzazione o attraverso legami is-a

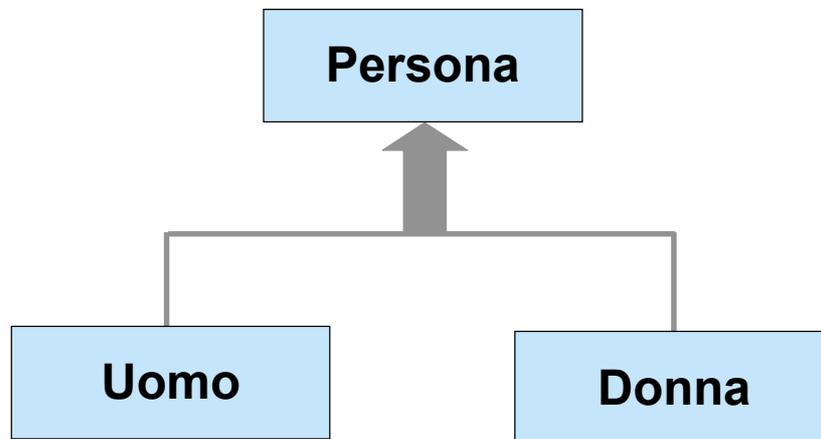
Primitive di raffinamento bottom-up: introduzione di un'entità



Primitive di raffinamento bottom-up: collegamento di entità con una relazione

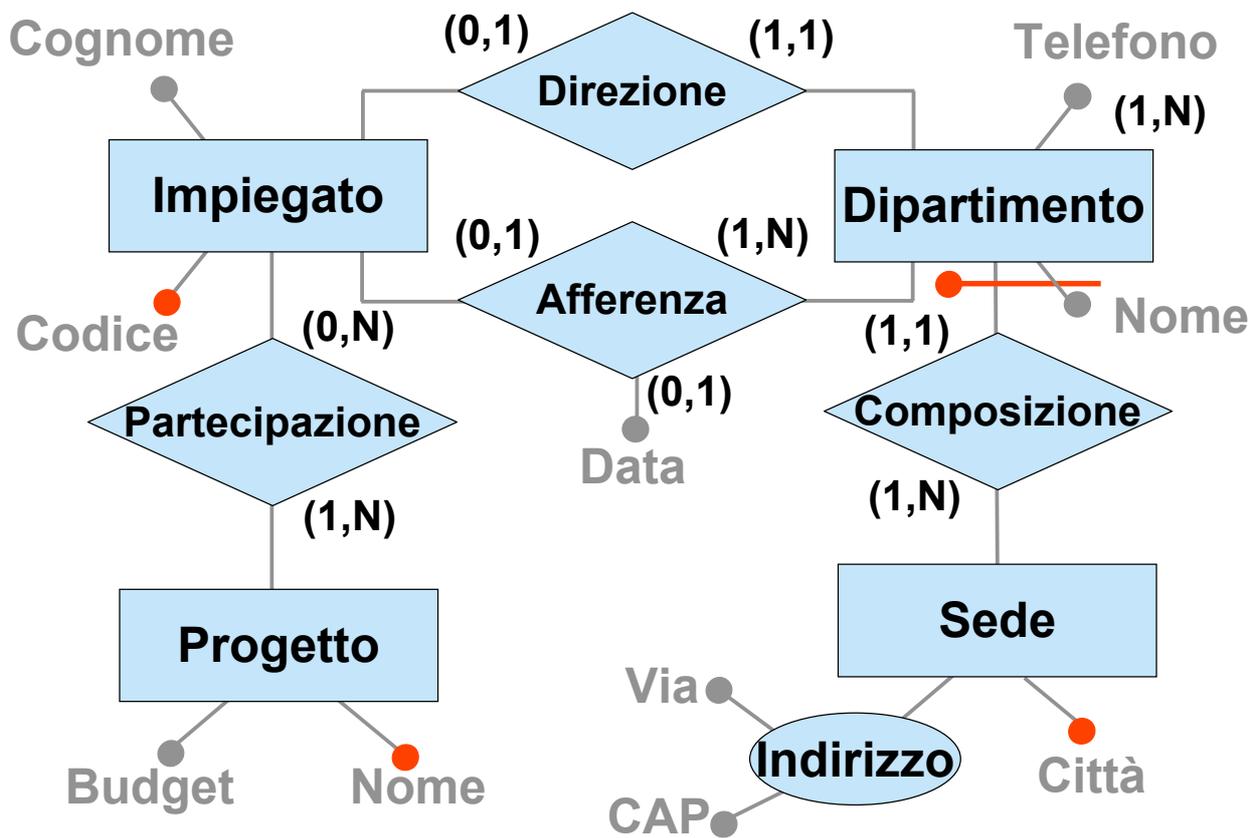


Primitive di raffinamento bottom-up: organizzazione di entità in una generalizzazione



Strategia inside-out

- Si parte da uno o più concetti particolarmente significativi della specifica.
- Si sviluppa lo schema a partire da questi concetti, introducendo nuovi concetti e relazioni che li collegano a quelli esistenti.



In pratica

Si procede di solito con una strategia ibrida (detta anche **mista**):

- si realizza uno **schema scheletro**
 - si individuano i concetti più importanti, ad esempio perché più citati o perché indicati esplicitamente come cruciali
 - si organizzano tali concetti in un semplice schema concettuale
 - ci si concentra sugli aspetti essenziali: molti attributi, cardinalità, is-a e gerarchie possono essere rimandate ad un approfondimento successivo
- si effettuano i **controlli di qualità**
- sulla base dello schema scheletro si può decomporre, **raffinare**, espandere

Trasformazioni

Indipendentemente dalla strategia adottata, è spesso necessario durante la progettazione concettuale effettuare trasformazioni che **preservino l'equivalenza**. A livello intuitivo, una trasformazione dallo schema S allo schema T preserva l'equivalenza se esiste una funzione che associa ad ogni istanza di S una istanza di T, ed una funzione che associa ad ogni istanza di T una di S.

Le più comuni sono:

- Trasformazione di un attributo semplice di un'entità in una relazione (e del corrispondente dominio in entità)
- Trasformazione di un attributo composto di un'entità in un'altra entità connessa alla prima da una relazione
- Trasformazione di una relazione in una entità
- Trasformazione di un attributo di una relazione R in una relazione (e del corrispondente dominio in entità). Si noti che questa trasformazione implica anche la trasformazione di R in entità

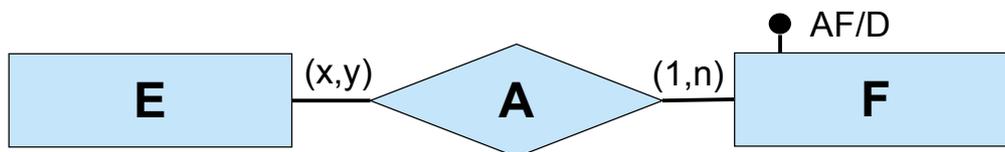
Esercizio 18: descrivere precisamente tali trasformazioni

Esercizio 18: soluzione per il caso a – sottocaso 1

Trasformazione di un attributo semplice di un'entità in una relazione (e del corrispondente dominio in entità).



Nel caso in cui E è l'unica entità con attributo A, si trasforma in:



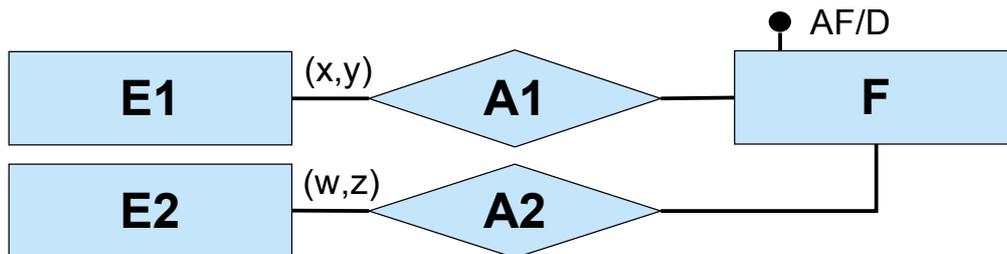
La cardinalità $(1,n)$ sulla partecipazione di F alla relazione A denota che interessano solo le istanze di F che rappresentano valori dell'attributo A effettivamente assunti dalle istanze di E nello schema originario.

Esercizio 18: soluzione per il caso a – sottocaso 2

Trasformazione di un attributo semplice di un'entità in una relazione (e del corrispondente dominio in entità).



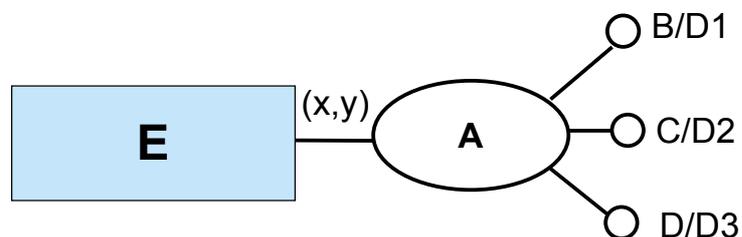
Nel caso in cui vi sono più entità con attributo A e con stesso dominio D associato ad A, si trasforma in:



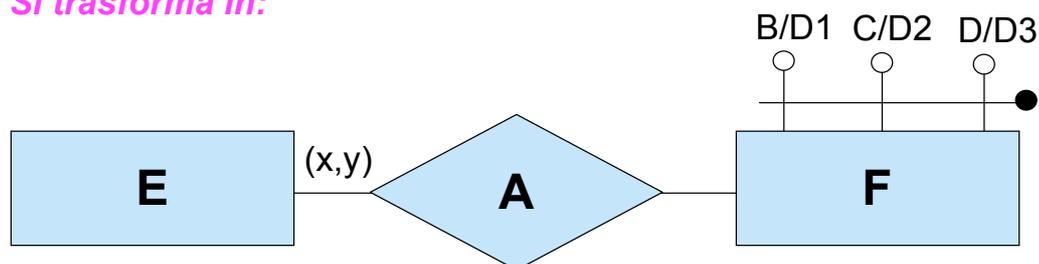
È necessario aggiungere un **vincolo esterno** di cardinalità (1,n) sulla partecipazione di F all'unione delle relazioni A1 e A2.

Esercizio 18: soluzione per il caso b

Trasformazione di un attributo composto di un'entità in un'altra entità connessa alla prima da una relazione

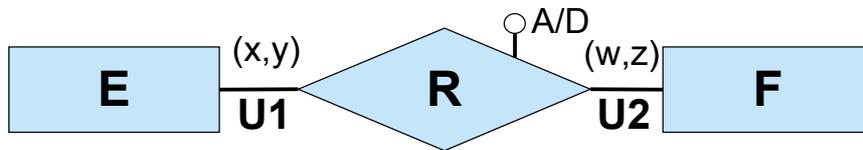


Si trasforma in:

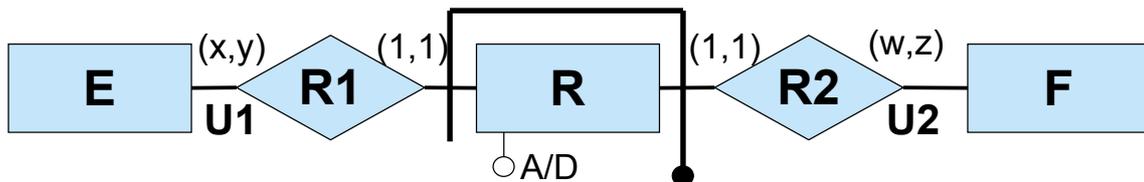


Esercizio 18: soluzione per il caso c

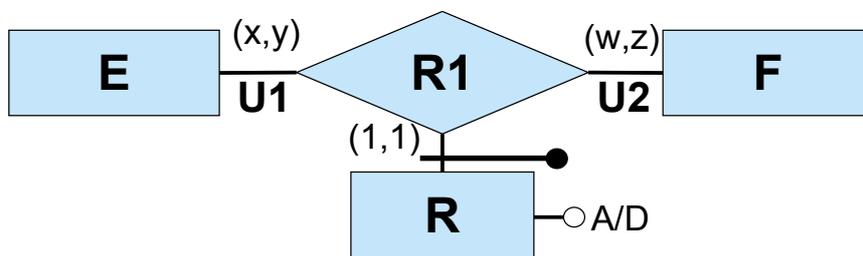
Trasformazione di una relazione in una entità



Si trasforma in

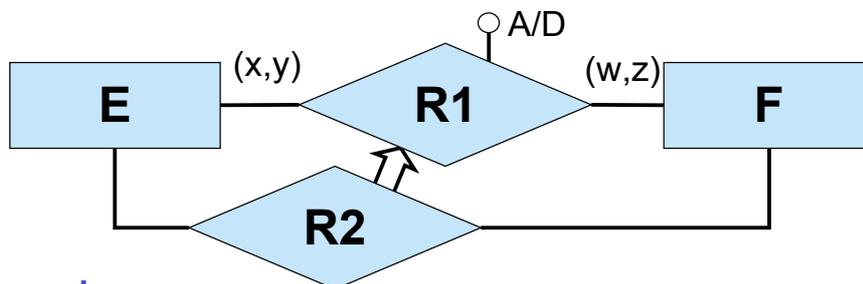


Oppure in

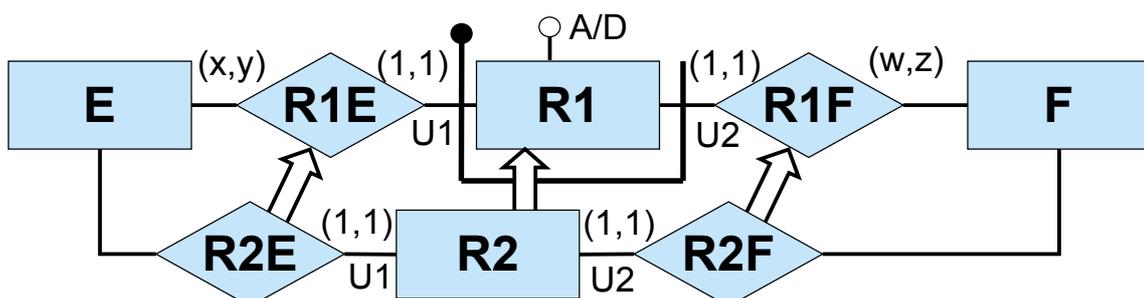


Esercizio 18: altra soluzione per il caso c

Nella trasformazione di una relazione R in una entità, anche le relazioni in ISA con R devono essere trasformate in entità.

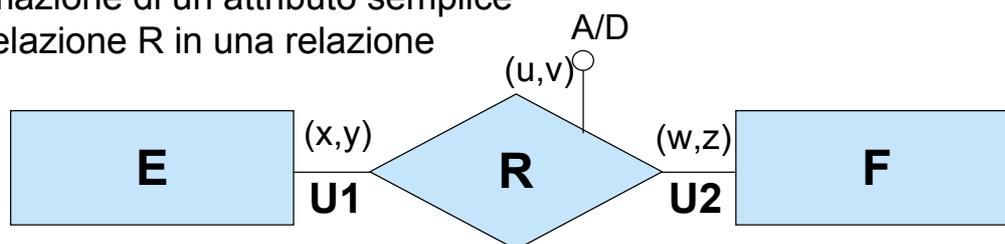


Si trasforma in:

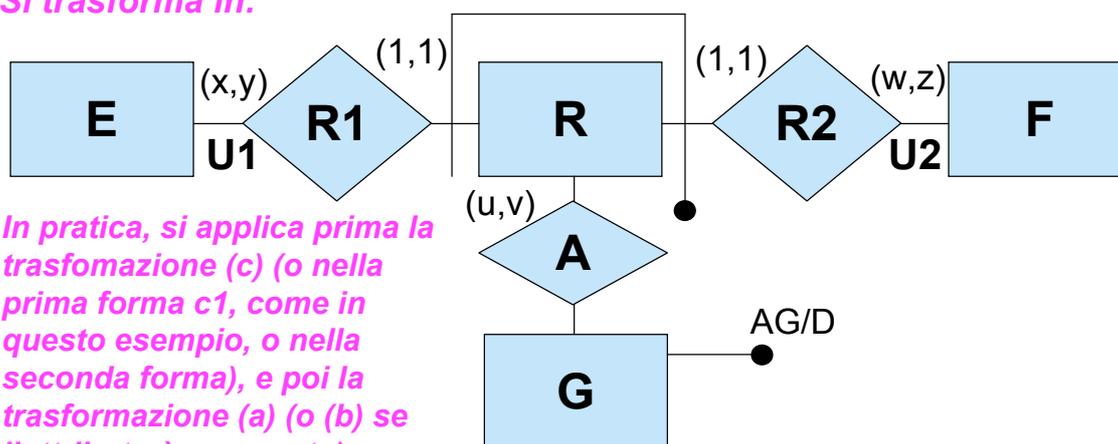


Esercizio 18: soluzione per il caso d

Trasformazione di un attributo semplice di una relazione R in una relazione



Si trasforma in:



In pratica, si applica prima la trasformazione (c) (o nella prima forma c1, come in questo esempio, o nella seconda forma), e poi la trasformazione (a) (o (b) se l'attributo è composto)

Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 197

Fase 4: Controllo di qualità dello schema concettuale

Cosa guida nelle scelte di progetto da fare per arrivare allo schema ER finale?

Fattori di qualità:

- **Correttezza: adeguatezza e completezza**

- La realtà d'interesse deve essere colta in modo adeguato e completo

- **Minimalità**

- Evitare quanto possibile di rappresentare più volte la stessa proprietà, e comunque documentare eventuale ridondanze (*nota che le ridondanze, se non riconosciute, possono compromettere la correttezza dello schema*)

- **Leggibilità**

- Progettare uno schema che sia di facile e diretta interpretazione (*es. evitare vincoli esterni, se è possibile farlo senza perdere la correttezza e senza pregiudicare la semplicità di interpretazione del diagramma*)

Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 198

La correttezza

Adeguatezza

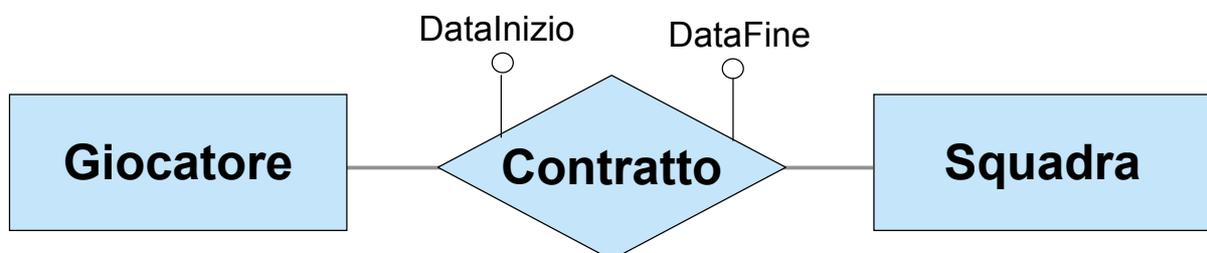
- **Rispetto al modello:** rispettare le regole del modello dei dati
- **Rispetto alle specifiche:** descrivere in modo fedele la realtà, secondo quanto descritto nelle specifiche

Completezza

- la realtà d'interesse deve essere colta in modo completo

Esercizio 19: correttezza

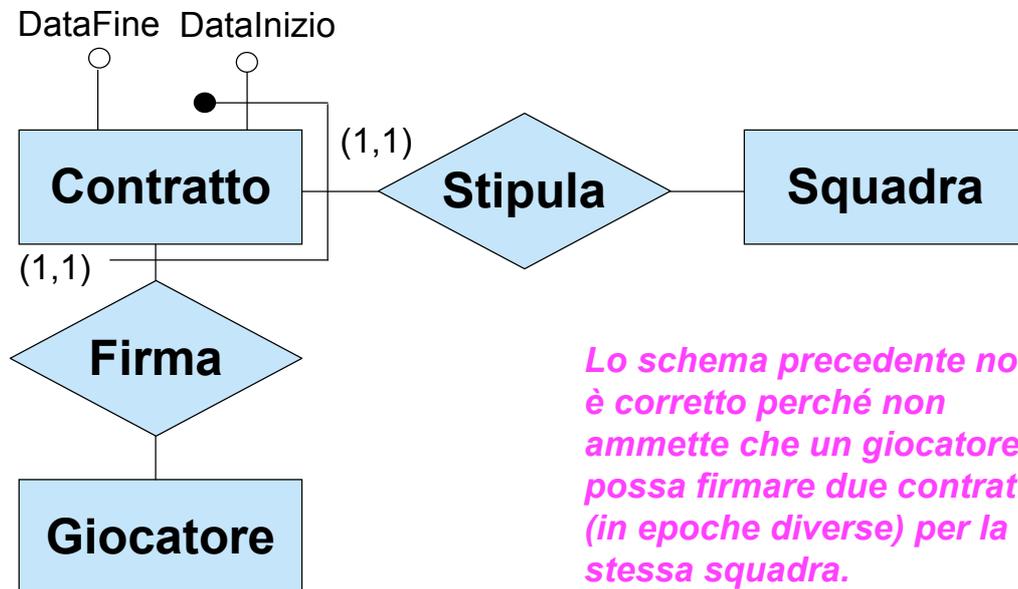
Si vogliono rappresentare le squadre in cui un giocatore milita attualmente, ed ha militato nel passato, con data inizio e data fine del contratto



E' corretto lo schema?

Esercizio 19: soluzione

Si vogliono rappresentare le squadre in cui un giocatore milita attualmente, ed ha militato nel passato, con data inizio e data fine del contratto.



Lo schema precedente non è corretto perché non ammette che un giocatore possa firmare due contratti (in epoche diverse) per la stessa squadra.

La minimalità

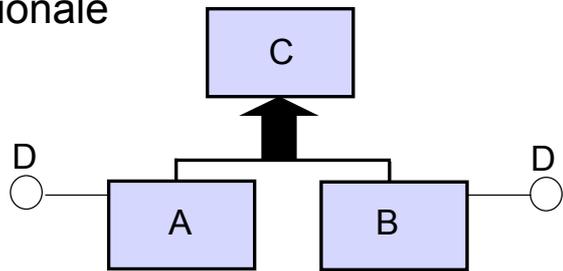
Evitare quanto possibile di rappresentare più volte la stessa proprietà: lasciare le ridondanze nello schema solo quando è rilevante dal punto di vista concettuale

- Ridondanze intensionali (in genere da evitare): si può “semplificare” lo schema preservando l’equivalenza?
Esempio importante: evitare cicli nella relazione ISA
- Ridondanze estensionali: nelle istanze dello schema la stessa proprietà estensionale è rappresentata più volte, implicitamente o esplicitamente

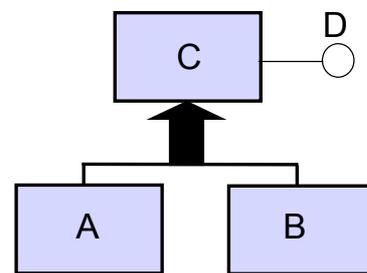
Se lo schema contiene ridondanze estensionale, occorre documentarle: se una ridondanza è rilevante a livello concettuale, occorre comunque esplicitarla, in genere mediante opportuni **vincoli di integrità**

La minimalità: ridondanza intensionale

Esempio di ridondanza intensionale



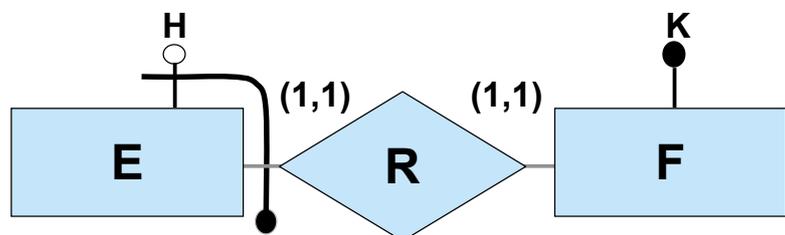
Risoluzione della ridondanza



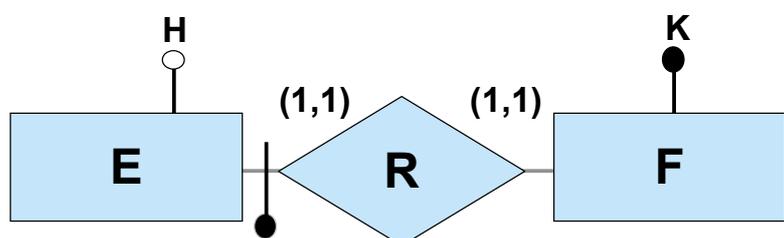
La minimalità: ridondanza intensionale

Un esempio di ridondanza intensionale è rappresentato da un identificatore di una entità che è costituito da un soprainsieme delle proprietà che formano un altro identificatore della stessa entità.

Esempio di ridondanza intensionale



Risoluzione della ridondanza

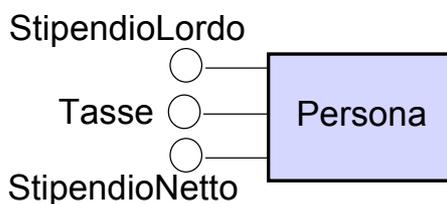


La minimalità: ridondanze estensionali

Due tipi importanti di ridondanze estensionali:

- il valore di un attributo si calcola sulla base di altre proprietà
- una relazione si ottiene da altre relazioni (per esempio da cammini di relazioni)

Esempio del primo tipo



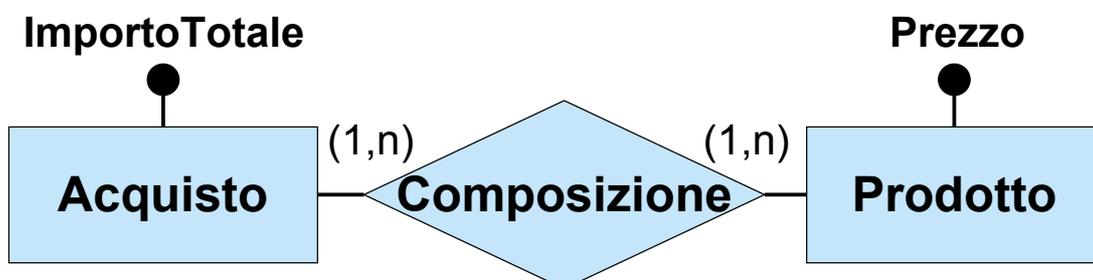
Vincolo esterno:

$$\text{StipendioNetto} = \text{StipendioLordo} - \text{Tasse}$$

La minimalità: ridondanze estensionali

Ulteriore esempio del primo tipo

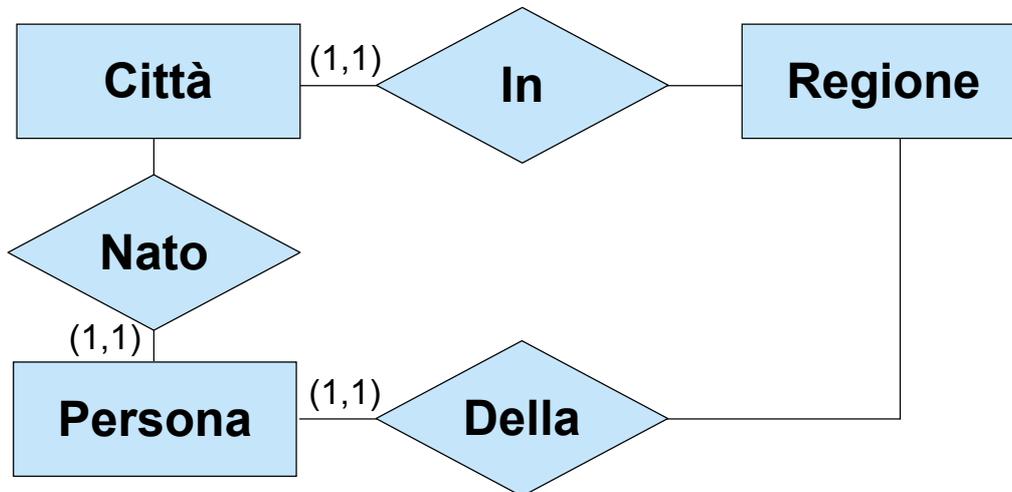
attributo calcolato sulla base degli attributi di un'altra entità



Vincolo esterno: per ogni istanza a di Acquisto, il valore di ImportoTotale è dato dalla somma dei valori di Prezzo per tutte le istanze di Prodotto connesse ad a tramite Composizione .

La minimalità: ridondanze estensionali

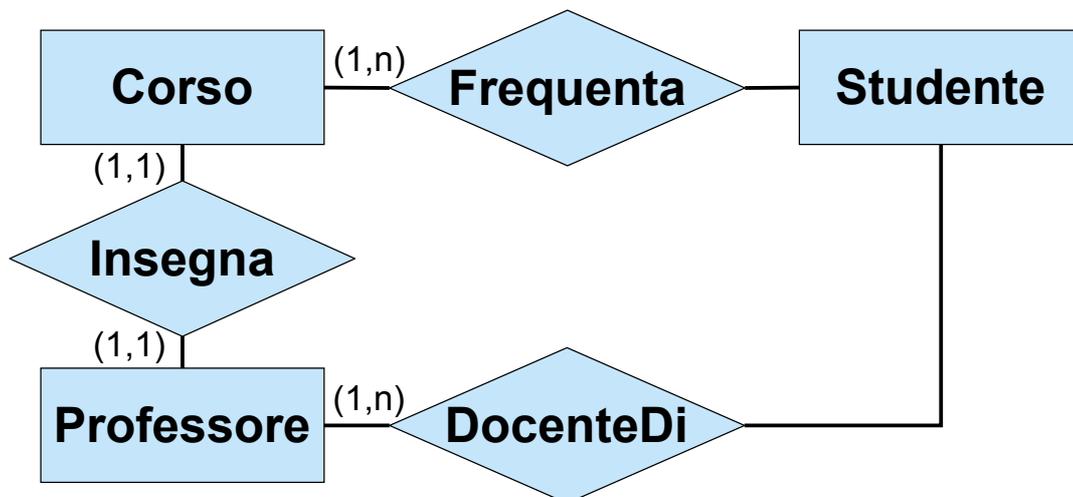
Esempio del secondo tipo



Vincolo esterno: per ogni istanza p di Persona, seguendo la relazione Nato e poi la relazione In, si ottiene la stessa istanza di Regione che si ottiene seguendo la relazione Della da p

La minimalità: ridondanze estensionali

Ulteriore esempio del secondo tipo



Vincolo esterno: per ogni istanza p di Professore, seguendo la relazione Insegna e poi la relazione Frequenta, si ottiene un'istanza di Studente che si ottiene anche seguendo la relazione DocenteDi da p .

La leggibilità

Oculata scelta dei concetti, dei loro nomi e delle strutture di rappresentazione scelte

Leggibilità del diagramma

- Produrre il diagramma secondo criteri “estetici naturali”, ad esempio cercando di produrre diagrammi che corrispondono a grafi planari, o comunque in cui il numero di incroci sia minimizzato

Leggibilità dei vincoli di integrità

- se possibile, scegliere lo schema concettuale che rappresenta un proprietà mediante vincoli espliciti e **non** mediante vincoli esterni

Produrre una buona documentazione (dizionario dei dati)

Controllo di qualità sullo schema ER

- È stata fatta una scelta oculata su come modellare i vari concetti?
 - se con attributi o con entità
 - se con entità o con relazioni
- Sono stati colti tutti gli aspetti importanti delle specifiche?
- Le generalizzazioni e le relazioni is-a sono corrette?
 - Non formano cicli?
 - Gli attributi sono associati alle entità giuste?
 - Le entità in una stessa generalizzazione sono disgiunte?
- Si possono applicare ulteriori generalizzazioni?

4. La progettazione concettuale

4.4 Esempio

1. introduzione alla progettazione di basi di dati
2. modello Entità-Relazione
3. metodologia per la progettazione concettuale
4. **esempio**

Base di dati corso di istruzione

Vogliamo memorizzare dati relativi ai partecipanti ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le città in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le lezioni che i partecipanti hanno frequentato, con i (o il), docenti che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Base di dati corso di istruzione

Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Base di dati corso di istruzione

Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER (1)

partecipante

Base di dati corso di istruzione

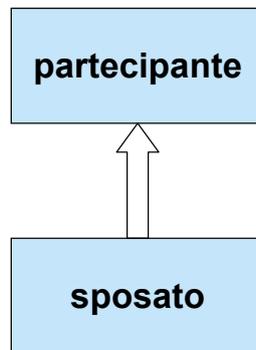
Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER (2)



Base di dati corso di istruzione

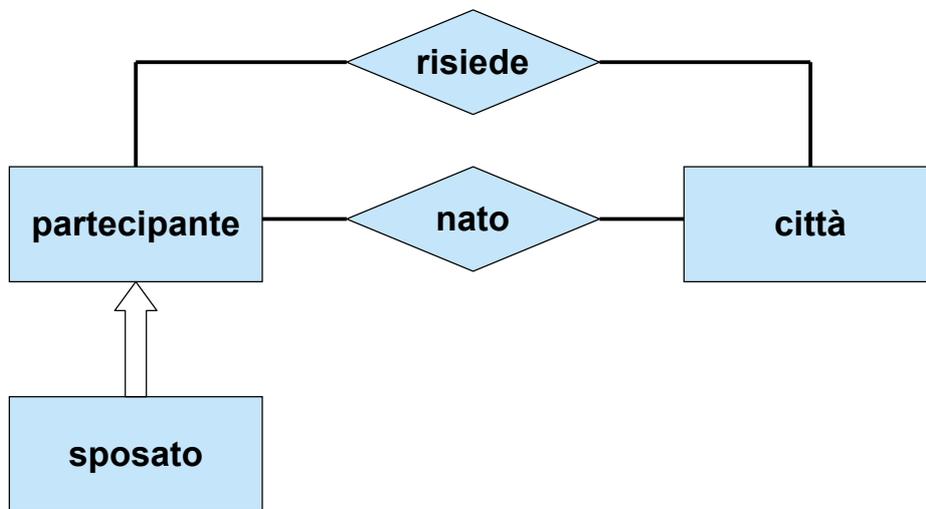
Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER (3)



Base di dati corso di istruzione

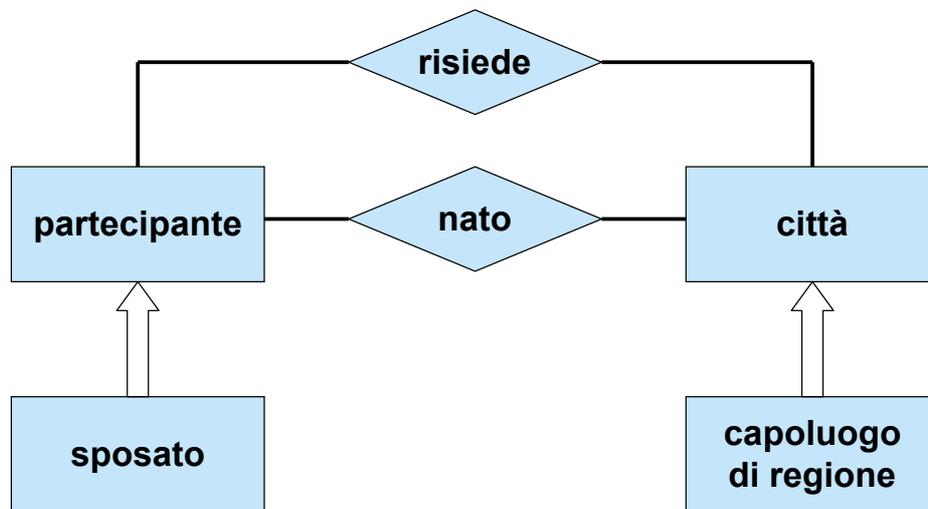
Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER (4)



Base di dati corso di istruzione

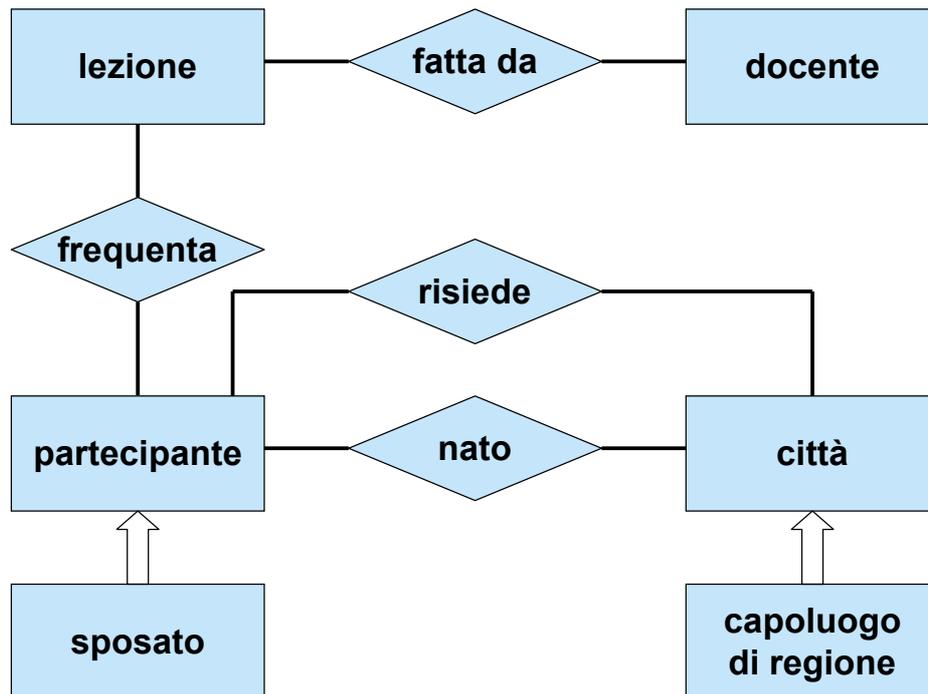
Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER (5)



Base di dati corso di istruzione

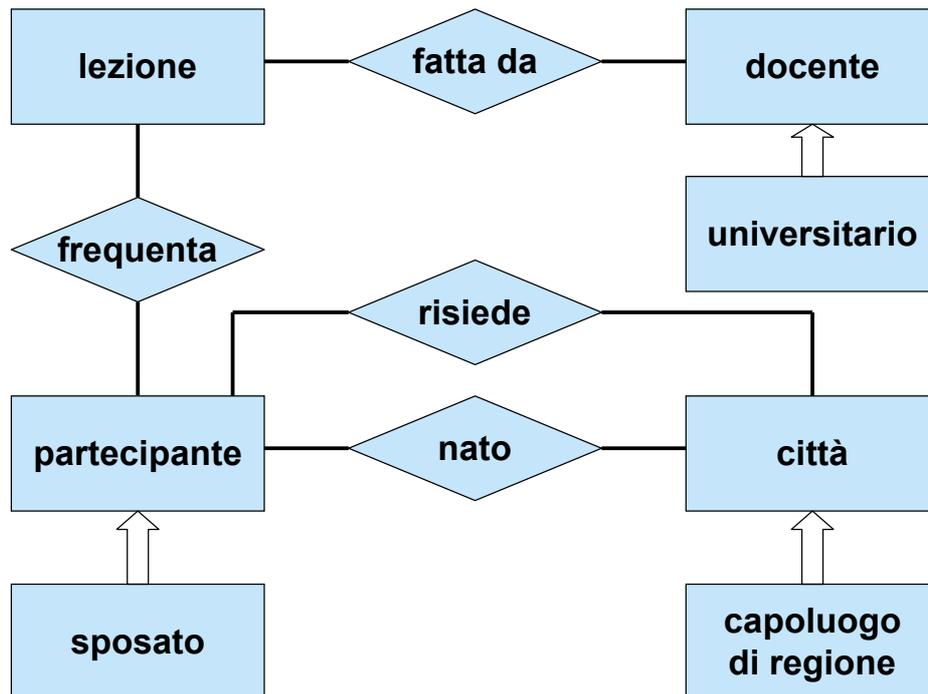
Vogliamo memorizzare dati relativi ai **partecipanti** ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le **città** in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le **lezioni** che i partecipanti hanno frequentato, con i (o il), **docenti** che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER (6)



Base di dati corso di istruzione

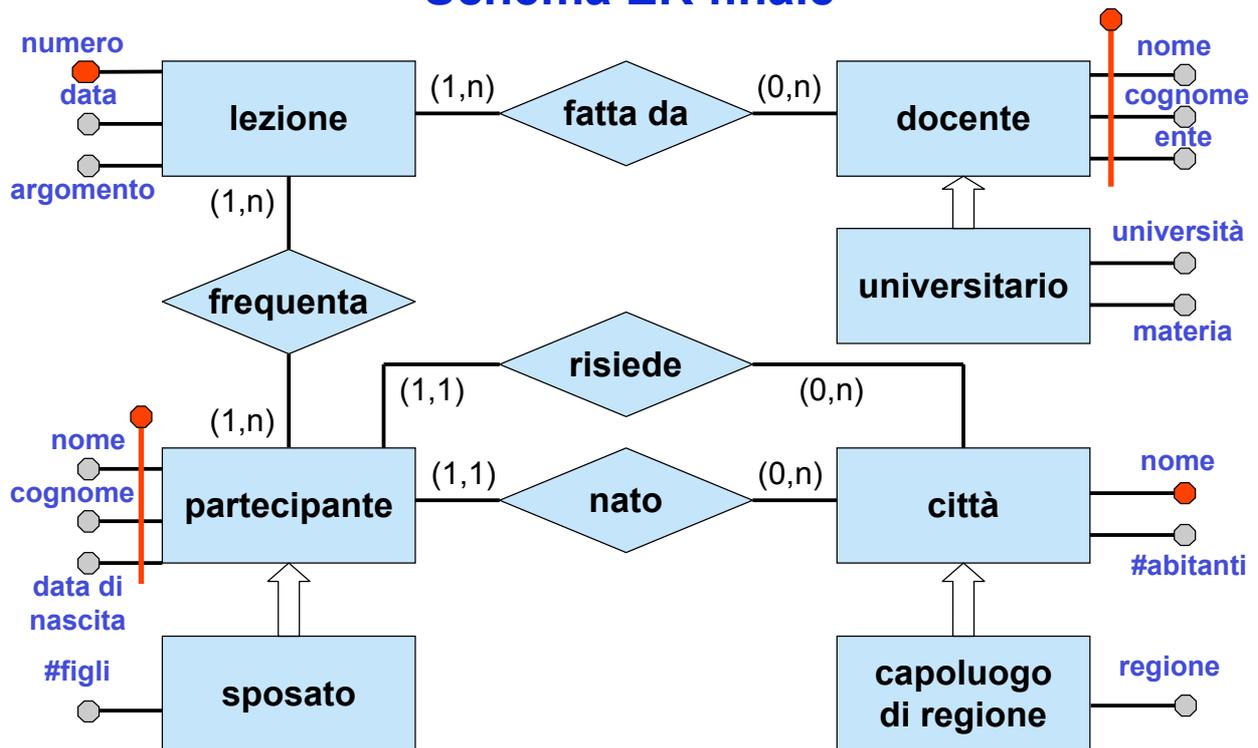
Vogliamo memorizzare dati relativi ai partecipanti ad un corso. Per ciascuno, vogliamo ricordare nome, cognome, data di nascita, se è sposato, e, nel caso lo sia, il numero di figli.

Vogliamo, poi ricordare le città in cui risiedono e le città in cui sono nati, insieme al numero di abitanti. Per le città capoluogo di regione, vogliamo ricordare la regione.

Vogliamo poi sapere le lezioni che i partecipanti hanno frequentato, con i (o il), docenti che le hanno svolte (nome, cognome, e tipo di enti di provenienza), il corrispondente argomento ed il giorno in cui si sono svolte. Ad ogni lezione va associato un numero progressivo.

Relativamente ai docenti provenienti dall'Università, si vuole ricordare la Università da cui provengono e la materia che hanno li in affidamento.

Schema ER finale



Giuseppe De Giacomo

Basi di Dati

Progettazione concettuale - 227

Vincoli non esprimibili in ER

- Ogni istanza dell'entità città, o è la città in cui è nato un partecipante oppure è la città in cui risiede un partecipante.

Giuseppe De Giacomo

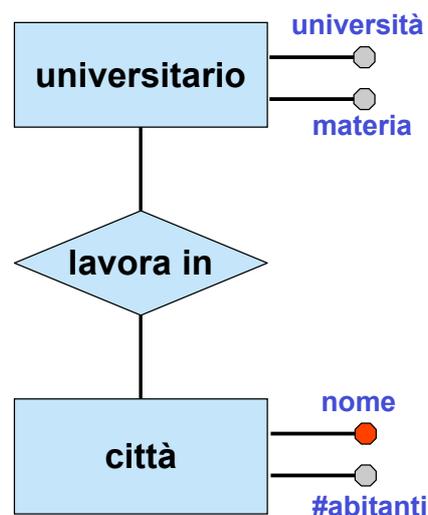
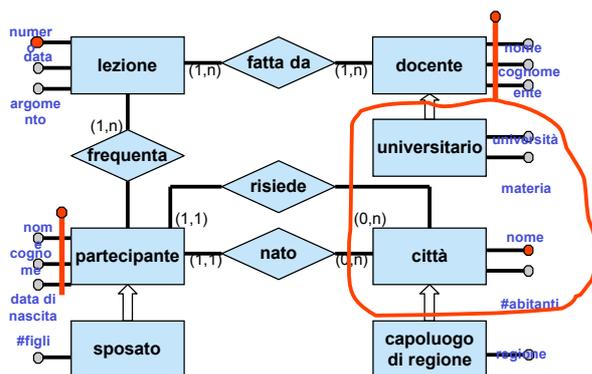
Basi di Dati

Progettazione concettuale - 228

Esercizio 20: modifica dello schema 1

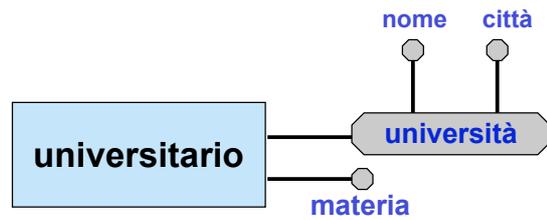
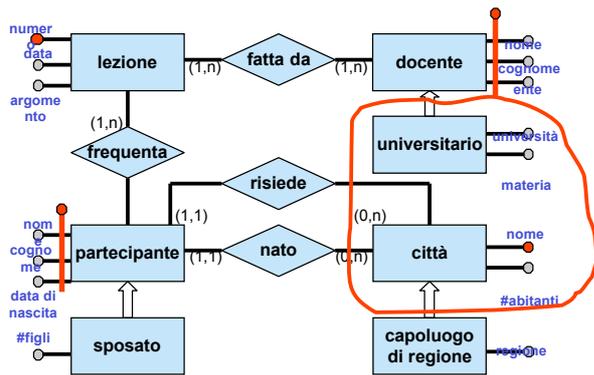
Per i docenti universitari vogliamo anche sapere la città in cui è ubicata l'Università in cui insegnano.

Esercizio 20: soluzione 1 [sbagliata!]

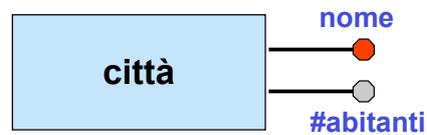


Per i docenti universitari vogliamo anche sapere la città in cui è ubicata l'Università in cui insegnano.

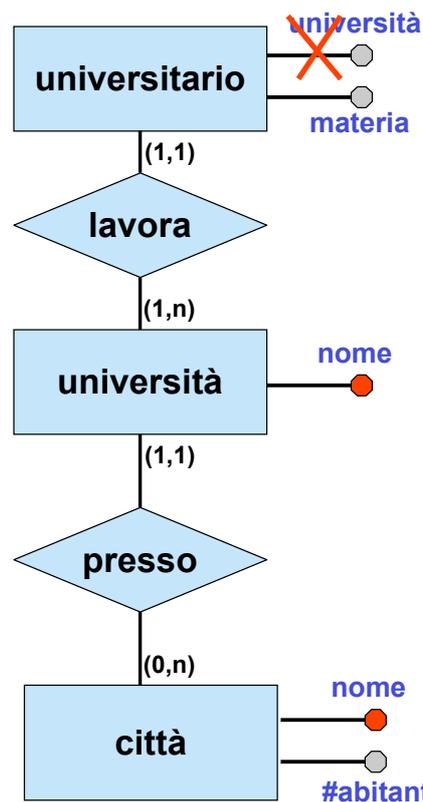
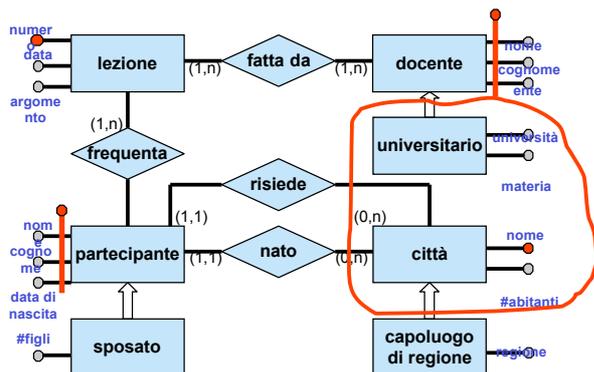
Esercizio 20: soluzione 2 [sbagliata!]



Per i docenti universitari vogliamo anche sapere la città in cui è ubicata l'Università in cui insegnano.



Esercizio 20: soluzione 3



Per i docenti universitari vogliamo anche sapere la città in cui è ubicata l'Università in cui insegnano.

Esercizio 20: confronto tra le tre soluzioni

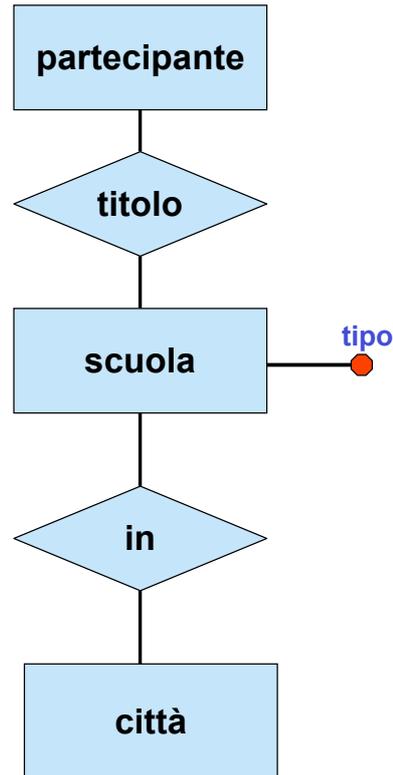
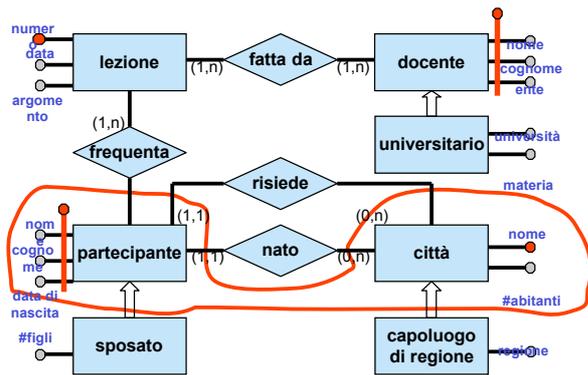
- La (2) non coglie il legame tra docente universitario e città (l'attributo città e l'entità città non sono correlate in alcun modo nel modello). **Si noti che questo è un errore!**
- La (1) e la (3) colgono il legame tra docente universitario e città.
- La (1) riferisce a docente universitario una proprietà (l'università) locale e non esplicita la dipendenza tra questa e la città.

→ *La soluzione corretta è la (3).*

Esercizio 21: modifica dello schema 2

Si vuole anche sapere in quale tipo di scuola i partecipanti hanno ottenuto l'ultimo titolo di studio, e in quale città la scuola ha sede.

Esercizio 21: soluzione 1 [sbagliata!]



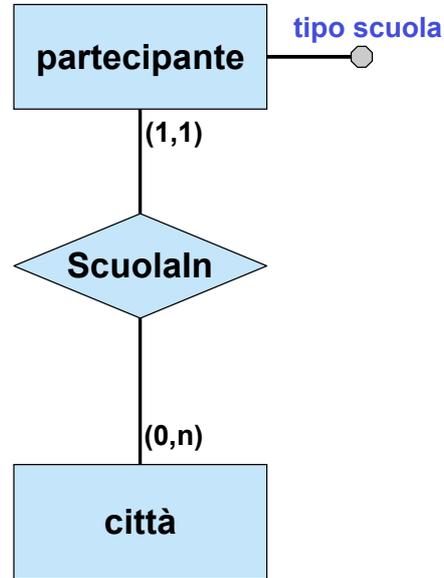
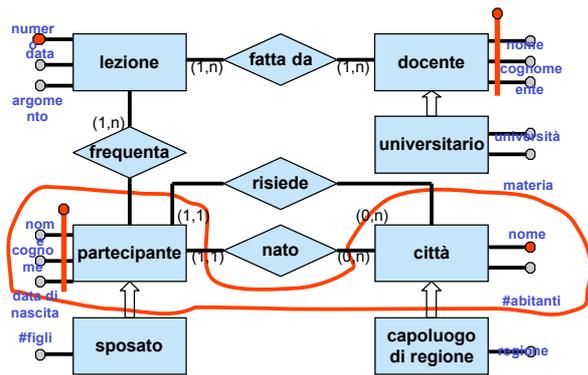
Si vuole anche sapere in quale tipo di scuola i partecipanti hanno ottenuto l'ultimo titolo di studio, e in quale città la scuola ha sede.

Esercizio 21: perché la soluzione è sbagliata?

Perché noi vogliamo sapere il tipo di scuola frequentata dal partecipante e la città in cui la scuola frequentata è situata.

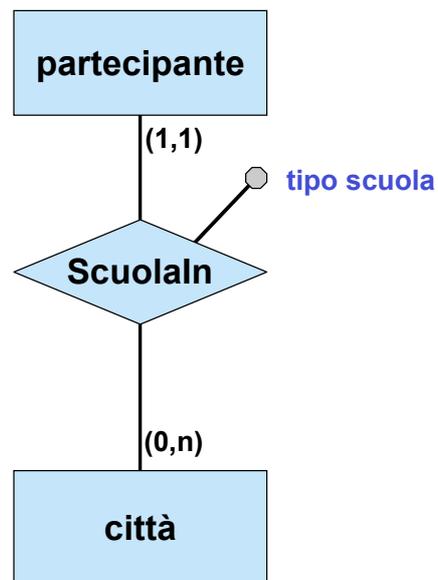
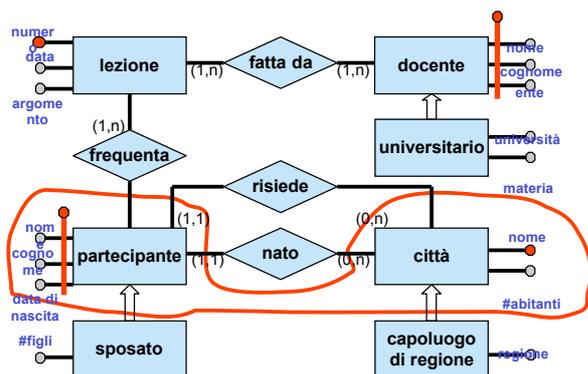
Nello schema invece viene rappresentato il tipo di scuola e le città in cui tale tipo di scuola è presente!

Esercizio 21: soluzione 2



Si vuole anche sapere in quale tipo di scuola i partecipanti hanno ottenuto l'ultimo titolo di studio, e in quale città la scuola ha sede.

Esercizio 21: soluzione 3



Si vuole anche sapere in quale tipo di scuola i partecipanti hanno ottenuto l'ultimo titolo di studio, e in quale città la scuola ha sede.

Esercizio 21: confronto tra le tre soluzioni

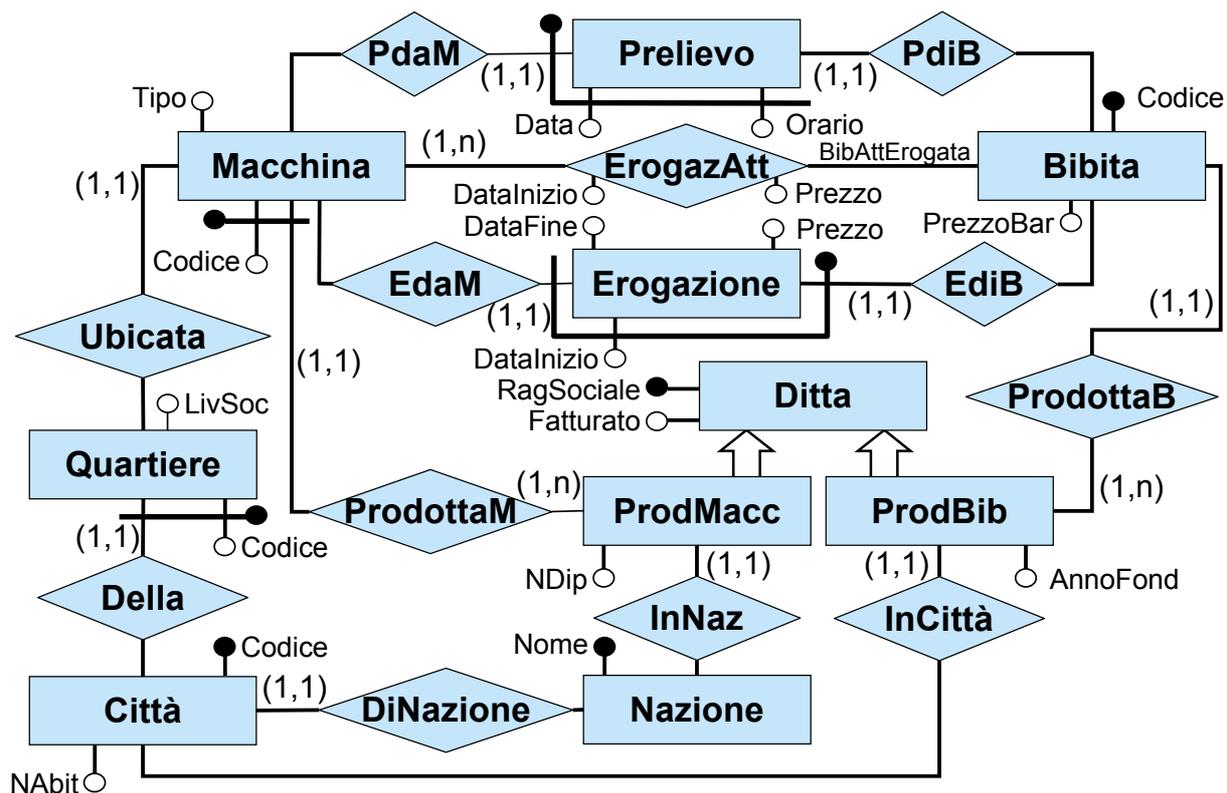
1. La soluzione (1) non è corretta.
2. La soluzione (2) non coglie che TipoScuola e ScuolIn si riferiscono alla stessa scuola.
3. La soluzione (3) coglie questo legame ponendo TipoScuola come attributo della relazione ScuolIn.

→ *La soluzione migliore è la (3).*

Esercizio 22: progettazione concettuale

Si richiede di progettare lo schema concettuale Entità-Relazione di un'applicazione relativa alle macchine che erogano bibite. Di ogni macchina erogatrice interessa il codice (unico fra quelle prodotte dalla stessa ditta produttrice), la ditta produttrice (ogni macchina erogatrice è prodotta da una ed una sola ditta), il tipo, e le bibite che la macchina eroga attualmente (ogni macchina erogatrice eroga attualmente almeno una bibita), con il relativo prezzo praticato, e la data in cui si è iniziato a praticare tale prezzo. Di ogni macchina interessa anche i prezzi eventualmente praticati precedentemente per le bibite attualmente erogate (con il relativo periodo specificato con data di inizio e data di fine), e le bibite che la macchina ha erogato e che non eroga più, sempre con i relativi prezzi nei vari periodi. Tutti i prezzi sono espressi in Euro. Ad esempio, la macchina erogatrice numero 5 della ditta "Pluto" eroga attualmente il Chinotto al prezzo di 35 Euro dal 1 marzo 2002, mentre lo erogava al prezzo di 30 Euro dal 1 gennaio 2001 al 28 febbraio 2002, e al prezzo di 25 Euro dal 1 gennaio del 1998 al 31 dicembre 2000. Inoltre, la stessa macchina erogatrice erogava il tea alla pesca, che adesso non eroga più, dal 1 gennaio 1995 al 31 dicembre del 1999 al prezzo di 25 Euro, e dal 1 aprile 2000 al 31 dicembre 2000 al prezzo di 28 Euro. Per ogni macchina erogatrice e per ogni bibita, interessa poi avere informazioni su ogni prelievamento effettuato dai clienti, con data e orario (espresso in ora, minuti e secondi) in cui è stato effettuato. Ogni prelievamento riguarda una ed una sola bibita, e si assume che le macchine erogatrici consentano al massimo un prelievamento al secondo. Di ogni macchina erogatrice interessa anche in quale quartiere di quale città è ubicata. Di ogni quartiere interessa il codice (unico all'interno della città), ed il livello sociale (numero intero positivo). Ogni città è identificata da un codice, e di ogni città interessa la nazione (della nazione interessa solo il nome) ed il numero di abitanti. Di ogni bibita interessa il codice (identificativo), il prezzo standard praticato al bar, e la ditta produttrice (ogni bibita è prodotta da una ed una sola ditta). Di ogni ditta produttrice di macchine erogatrici interessa la ragione sociale (identificativo), il fatturato, il numero di dipendenti, e la nazione in cui è situata la sede ufficiale (della nazione interessa solo il nome). Di ogni ditta produttrice di bibite interessa la ragione sociale (identificativo), il fatturato, l'anno di fondazione, e la città in cui è situata la sede ufficiale.

Esercizio 22: soluzione – schema ER



Esercizio 22: soluzione – vincoli esterni (1)

- Le date di inizio e di fine erogazione sono coerenti tra di loro:
per ogni $e \in \text{istanze}(\text{Erogazione})$: $\text{DataInizio}(e) \leq \text{DataFine}(e)$
 - I periodi di erogazione di una stessa bibita da parte di una stessa macchina sono disgiunti a coppie:
per ogni $e_1, e_2 \in \text{istanze}(\text{Erogazione})$, $m \in \text{istanze}(\text{Macchina})$,
 $b \in \text{istanze}(\text{Bibita})$
se $(e_1, m), (e_2, m) \in \text{istanze}(\text{EdaM})$,
 $(e_1, b), (e_2, b) \in \text{istanze}(\text{EdiB})$, e
 $\text{DataInizio}(e_1) \leq \text{DataInizio}(e_2)$
allora $\text{DataFine}(e_1) \leq \text{DataInizio}(e_2)$
- per ogni $e \in \text{istanze}(\text{Erogazione})$, $(m, b) \in \text{istanze}(\text{ErogazAtt})$,
se $(e, m) \in \text{istanze}(\text{EdaM})$ e $(e, b) \in \text{istanze}(\text{EdiB})$
allora $\text{DataFine}(e) \leq \text{DataInizio}((m, b))$

Esercizio 22: soluzione – vincoli esterni (2)

- I prelievi sono coerenti con le erogazioni:
per ogni $p \in \text{istanze}(\text{Prelievo})$, $m \in \text{istanze}(\text{Macchina})$,
 $b \in \text{istanze}(\text{Bibita})$
se $(p,m) \in \text{istanze}(\text{PdaM})$ e $(p,b) \in \text{istanze}(\text{PdiB})$
allora esiste $e \in \text{istanze}(\text{Erogazione})$
con $(e,m) \in \text{istanze}(\text{EdaM})$, $(e,b) \in \text{istanze}(\text{EdiB})$, e
 $\text{DataInizio}(e) \leq \text{Data}(p) \leq \text{DataFine}(e)$
oppure $(m,b) \in \text{istanze}(\text{ErogazAtt})$ e
 $\text{DataInizio}((m,b)) \leq \text{Data}(p)$
- Per le ditte che producono sia macchine che bibite, le informazioni su città e nazione della sede ufficiale sono coerenti:
per ogni $d \in \text{istanze}(\text{ProdMacc}) \cap \text{istanze}(\text{ProdBib})$
se $(d,c) \in \text{istanze}(\text{InCittà})$ e $(c,n) \in \text{istanze}(\text{DiNazione})$
allora $(d,n) \in \text{istanze}(\text{InNaz})$