



1° lezione database

Due Date	@October 4, 2024
Materia	Database
REVISIONATA	<input checked="" type="checkbox"/>
Status	Done

Il modello relazionale dei dati

Un modello dei dati è un insieme di *meccanismi di astrazione* per definire una base di dati, con associato un insieme predefinito di *operatori* e di *vincoli di integrità*.

Relazioni

si hanno 3 accezioni di relazioni:

1. **matematica:** come nella teoria degli insiemi
2. **secondo il modello relazionale dei dati**
3. **relazione di una classe di fatti (Entità associazioni)**

▼ Relazioni matematiche

Supponendo di avere due domini $D_1 = a, b$ $D_2 = x, y, z$ possiamo definire il prodotto cartesiano che è dato da $D_1 * D_2$ sono tutte le possibili coppie che si possono ottenere a partire dai due insiemi D_1 e D_2 .

Una relazione è un sotto insieme del prodotto cartesiano in simboli: $r \subseteq D_1 * D_2$

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

a	x
a	z
b	y

Generalizzando il concetto di prodotto cartesiano con D_n con n insiemi anche non distinti.

▼ Prodotto cartesiano

Il prodotto cartesiano è l'insieme di tutte le n-uple (d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$

▼ Relazione matematica su D_1, \dots, D_n

è un sottoinsieme del prodotto cartesiano di $D_1 * \dots * D_n$

con D_n sono domini della relazione

Una relazione è un insieme di n-uple ordinate

$$(d_1, \dots, d_n) \text{ tali che } d_1 \in D_1, \dots, d_n \in D_n$$

La relazione è un insieme dove:

- non c'è ordinamento fra le n-uple
- le n-uple sono distinte
- ciascuna n-upla è ordinata: l' i-esimo valore proviene dall'i-esimo dominio

$$Partite \subseteq \text{stringa} \times \text{stringa} \times \text{intero} \times \text{intero}$$

Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

Una chiave è un insieme di attributi che identifica in maniera univoca uno e un solo record all'interno della tabella. Formalmente si può definire che un insieme X di attributi di uno schema di relazione R è una super chiave dello schema se ogni istanza valida dello schema non contiene due n-uple distinte t_1 e t_2 con $t_1[X] = t_2[X]$

Ciascuno dei domini ha ruoli diversi, distinguibili attraverso la posizione. La struttura è posizionale

Distinguiamo due tipi di struttura

▼ Struttura posizionale

Nella struttura posizionale, i valori degli attributi sono distinti in base alla loro **posizione** nella tupla.

$$Partite \subseteq \text{stringa} \times \text{stringa} \times \text{intero} \times \text{intero}$$

Possiamo dire che ogni tupla rappresenta una partita, e la posizione dei vari attributi è chiara:

- Primo attributo: Squadra 1 (stringa)
- Secondo attributo: Squadra 2 (stringa)
- Terzo attributo: Gol squadra 1 (intero)
- Quarto attributo: Gol squadra 2 (intero)

Il significato di ogni valore nella tupla dipende dalla sua posizione

▼ Struttura non posizionale

Nella struttura non posizionale, a ciascun valore viene associato un **nome unico** (detto "attributo") che ne definisce il ruolo, indipendentemente dalla posizione nella tupla. In questo caso, ogni valore ha un significato basato sull'attributo a cui è associato, e non dalla sua posizione.

Se trasformassimo la tua tabella in una struttura **non posizionale**, potremmo assegnare un nome a ciascun attributo, ad esempio:

- `squadra1`
- `squadra2`
- `gol_squadra1`
- `gol_squadra2`

Il ruolo di ciascun valore viene determinato dal suo nome (attributo), non dalla sua posizione.

Schema di relazione

è una struttura che si prepara a contenere dei dati è formata da dal nome di relazione (R) e da un tipo di relazione T. R: {T}. Non contiene dati ma definisce la struttura secondo una certa semantica.

Definizione Tipo Relazione(T)

Si basano sui tipi primitivi come interi, reali, stringhe.

Indichiamo con T_n i tipi primitivi e A_n gli attributi. $(A_1 : T_1, A_2 : T_2, \dots, A_n : T_n)$ è un tipo n-upla di garando n
 $\{T\}$ è un tipo relazione cioè un tipo insieme di n-uple

Da cosa è costituito uno schema relazionale?

Uno schema relazionale è costituito da:

- un insieme di *schemi di relazione* $R_i : \{T_i\}, i=1, 2, \dots, k$
- un insieme di *vincoli di integrità* relativi a tali schemi.

Lo schema relazionale costituisce l'aspetto *intenzionale*.

▼ L'aspetto intenzionale

è la parte in cui si vanno a definire le strutture delle relazioni.

▼ L'aspetto estenzionale

Consiste nel valorizzare i singoli attributi inserendo delle n-uple

$(A_1 := v_1, A_2 := v_2, \dots, A_n := v_n)$ di tipo $(A_1 : T_1, A_2 : T_2, \dots, A_n : T_n)$ è un insieme di coppie (A_i, v_i) con

Una relazione R di tipo $\{(A_1 : T_1, A_2 : T_2, \dots, A_n : T_n)\}$ è un insieme finito di n-uple di tipo $(A_1 : T_1, A_2 : T_2$

La carnalità di una relazione è il numero di n-uple

Esempio di relazione

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

- E' una relazione di tipo
 $\{(Nome : Stringa, Matricola : intero, Indirizzo : Stringa, Telefono : intero)\}$
- Ciascuna riga è una n-upla della relazione.

Esempio di schema relazionale

```
{
  Studenti : { (Nome : stringa, Matricola : intero(6),
               Indirizzo : stringa, Telefono : intero(6)) }

  Esami : { (Corso : stringa, Matricola : intero(6), Voto : {18,19,...,30}) }

  Corsi : { (Corso : stringa, Professore : stringa) }
}
```

- Studenti (Nome, Matricola, Indirizzo, Telefono);
- Esami (Corso, Matricola, Voto);
- Corsi (Corso, Professore);

Dominio di un attributo

Il dominio di un attributo è l'insieme di valori che l'attributo può assumere in simboli. L'attributo si indica con A_i .

Esempio: nella tabella Esami(Corso, Matricola, Voto)

il dominio è $\text{dom}(\text{Voto}) = \{18, 19, \dots, 30\}$

Vincoli di integrità

Il vincolo di integrità ha come obiettivo quello di garantire l'integrità delle informazioni e migliorare la qualità di queste contenute all'interno di una base di dati. Un vincolo è un predicato, ovvero una funzione che assume valori vero o falso e deve essere soddisfatta da ogni n-upla nella base dei dati. Si ha un'istanza valida quando questa soddisfa tutti i vincoli di integrità stabiliti in precedenza definite su di essi.

Esempi di vincoli di integrità

- Il voto deve essere compreso tra 18 e 30
- La lode può apparire solo se voto = 30
- Ogni studente deve avere un numero di matricola
- Il numero di matricola di uno studente deve essere univoco
- Esami dati devono fare riferimento solo a corsi offerti (esempio di vincoli intrarelazionali)

Si hanno tre vincoli di integrità fondamentali

1. Quali attributi non possono assumere il valore NULL (vincoli intrarelazionali)
2. Quali attributi sono chiave (vincoli intrarelazionali)
3. Quali attributi sono chiave esterne (vincoli intrarelazionali)

Esempio di valori nulli

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	NULL	999999
Enzo Gialli	456789	Via Catania 3	NULL

Si hanno due casi:

1. Il dato che è nullo, non è disponibile
2. Il dato che è nullo non è stato valorizzato

Le chiavi

Una chiave è un insieme di attributi che identifica in maniera univoca uno e un solo record all'interno di una tabella. Formalmente, un insieme X di attributi di uno schema di relazione R è una super chiave se, per ogni istanza valida dello schema, non esistono due n-uple distinte t_1 e t_2 tali che $t_1[X] = t_2[X]$. In altre parole, la super chiave garantisce l'unicità dei record rispetto agli attributi che la compongono.

Consideriamo la seguente tabella: Dipendente(nome, ruolo, matricola, matricola_capo). Supponiamo che l'attributo matricola sia una chiave. Una super chiave è un qualsiasi insieme di attributi che identifica in modo univoco un record nella tabella. Ad esempio, l'insieme di tutti gli attributi {nome, ruolo, matricola, matricola_capo} {nome, ruolo, matricola, matricola_capo} è una super chiave, perché garantisce che ogni record è univoco. Tuttavia, in questa tabella, l'attributo matricola da solo è sufficiente per identificare univocamente ogni dipendente.

Il concetto di chiave si riferisce a una super chiave minimale, cioè una super chiave che non contiene attributi superflui. Poiché matricola da sola è sufficiente per identificare un dipendente, essa è una chiave. Le altre super

chiavi che includono più attributi (ad esempio, {matricola,nome}{matricola,nome} o {matricola,matricolacapo}{matricola,matricolacapo}) non sono minimali, perché includono informazioni ridondanti.

In sintesi, una chiave è una super chiave minimale. Nel nostro esempio, l'attributo matricola è la chiave della tabella, poiché è l'insieme minimale di attributi che identifica univocamente ogni record.

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

In questo esempio, la coppia a nome matricola è una super chiave, ma non una chiave infatti, in questo caso la chiave è la matricola. Se vuoi definire comunicare un attributo o un insieme di attributi, quando comunque preso un'istanza valida, non si riesce mai a trovare una coppia di record che assumano lo stesso valore sugli attributi che si sono scelti come chiave. Nel caso in cui però ci fossero due persone con lo stesso nome, il nome non potrebbe essere un altro punto chiave.

L'importanza delle chiavi

L'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della basi dei dati. le chiavi permettono di correlare i dati in relazione diverse, il modello nazionale basato sui valori. in presenza dei valori nulli, i valori della chiave non permettono di identificare le n-uple e di realizzare facilmente i riferimenti ad altre relazioni.

Cos'è la chiave primaria?

La chiave primaria è una delle chiavi scelta per un dato schema. Non ammette un nulli e si indica sottolineandola. Se si ha una chiave primaria, composta gli attributi scelti non devono essere nulli.

Nome	<u>Matricola</u>	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

Nell'esempio di sopra la matricola è una chiave primaria.

Chiavi esterne

Una live esterna all'interno della tabella è un attributo o una coppia di attributi che risulta essere chiave primaria in una qualche altra tabella all'interno dello schema relazionale. F

Un insieme di attributi {A1, A2, ..., An} di uno schema di relazione R è una chiave esterna che fa riferimento a una chiave primaria {B1, B2, ..., Bn} di un altro schema S se, in ogni istanza valida della base di dati, per ogni n-upla rdell'istanza di R esiste una n-upla s (riferita da r) dell'istanza di S tale che $r.A_i = s.B_i$ per $i = 1, \dots, n$.

Questo significa che per ogni record di R, deve esistere un record corrispondente in S in cui gli attributi specificati come chiavi esterne in R coincidono con gli attributi della chiave primaria in S. La chiave esterna stabilisce quindi un collegamento referenziale tra due tabelle, garantendo l'integrità dei dati.

Studenti			
Nome	<u>Matricola</u>	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

Esami		
Corso	<u>Matricola</u>	Voto
Programmazione 1	345678	27
Architettura	123456	30
Matematica discreta	234567	19
Algoritmi	345678	28

Esami

Corso	Matricola	Voto
Programmazione 1	345678	27
Architettura	123456	30
Matematica discreta	234567	19
Basi di Dati	345678	28

Corsi

Corso	Professore
Architettura	Barbanera
Programmaizione 1	Cincotti
Matematica discreta	Milici
Algoritmi	Cantone

Algebra relazionale

È formato da un set di operatori che sono definiti sulle relazioni e producono come risultato un'altra relazione. Gli operatori possono essere combinati per formare espressioni più complesse.

Quali sono gli operatori dell'algebra relazionale?

Gli operatori primitivi sono:

- Ridenominazione
- Unione
- Differenza
- Proiezione
- Selezione
- Prodotto

I simboli che si utilizzano per la tradizionale sono:

- R, S denotano relazioni
- A, B sono attributi
- X, Y, Z sono insiemi di attributo

▼ Operatore di ridenominazione

Questo operatore consente di cambiare un nome ad un attributo che sta in una certa tabella.

È un operatore lunario.

Esempio di ridenominazione

STUDENTE		
Corso	Matricola	Voto
Programmazione	123456	27
EINN	23456	28

$\delta_{mstricola \rightarrow \text{Codice Studente}}(\text{Studente})$

▼ Unione, Differenza, Intersezione (\cup \cap $-$)

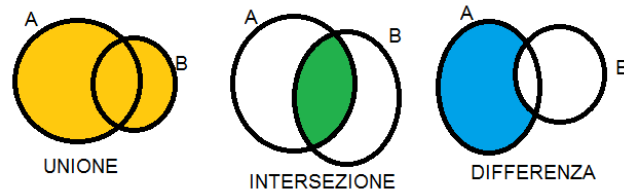
Questi operatori sono definiti su relazioni con lo stesso schema (stessi attributi) perché il risultato deve essere un set omogeneo di n-uple. Sono tutti operatori binari, cioè con due tabelle in ingresso.

Siano R ed S relazione dello stesso tipo allora:

$$R \cup S = \{t | t \in R \vee t \in S\}$$

$$R \cap S = \{t | t \in R \wedge t \in S\}$$

$$R - S = \{t | t \in R \wedge t \notin S\}$$



▼ Esempio di unione

Graduates

Number	Surname	Age
7274	Robinson	37
7432	O'Malley	39
9824	Darkes	38

Managers

Number	Surname	Age
9297	O'Malley	56
7432	O'Malley	39
9824	Darkes	38

Graduates \cup Managers

Number	Surname	Age
7274	Robinson	37
7432	O'Malley	39
9824	Darkes	38
9297	O'Malley	56

▼ Esempio di intersezione

Graduates

Number	Surname	Age
7274	Robinson	37
7432	O'Malley	39
9824	Darkes	38

Managers

Number	Surname	Age
9297	O'Malley	56
7432	O'Malley	39
9824	Darkes	38

Graduates \cap Managers

Number	Surname	Age
7432	O'Malley	39
9824	Darkes	38

L' intersezione è un operatore derivato, si può ottenere sfruttando gli operatori di base che sono l'unione e la differenza.

▼ Esempio di Differenza

Graduates

Number	Surname	Age
7274	Robinson	37
7432	O'Malley	39
9824	Darkes	38

Graduates - Managers

Number	Surname	Age
7274	Robinson	37

Managers

Number	Surname	Age
9297	O'Malley	56
7432	O'Malley	39
9824	Darkes	38

La differenza non è commutativa.

Se nelle tabelle non si hanno gli stessi attributi allora l'unione non si può eseguire, pertanto si ricorre a ridenominare gli attributi nel seguente modo:

Paternity

Father	Child
Adam	Cain
Adam	Abel
Abraham	Isaac
Abraham	Ishmael

Maternity

Mother	Child
Eve	Cain
Eve	Seth
Sarah	Isaac
Hagar	Ishmael

$$\delta_{Father \rightarrow Parent}(Paternity) \cup \delta_{Mother \rightarrow Parent}(Maternity)$$

Parent	Child
Adam	Cain
Adam	Abel
Abraham	Isaac
Abraham	Ishmael
Eve	Cain
Eve	Seth
Sarah	Isaac
Hagar	Ishmael

Proiezione

La proiezione si indica con il simbolo π . Eseguire una proiezione significa selezionare solo alcuni attributi di una relazione, mantenendo però tutti i record (n-upla) che appartengono a quella relazione. In questo modo, si ottiene un'altra relazione che ha lo stesso numero di record della relazione originale, ma contiene solo gli attributi specificati nella proiezione, pertanto si deduce che si ottiene un'altra relazione che ha la stessa cardinalità della relazione di partenza ma l'ordine cambia in base a quanti attributi si specificano.

Ad esempio, se applichiamo una proiezione su una relazione prendendo solo gli attributi . Indicheremo l'operazione come π_{C_2, C_3} . Questo significa che nella relazione risultante, verranno mantenuti tutti i record, ma ogni record conterrà solo i valori degli attributi C_2 e C_3 .

Definizione formale della proiezione

$$\pi_{A_1, A_2, \dots, A_n}(R) = \{t[A_1, A_2, \dots, A_n] \mid t \in R\}$$

Esempio di proiezione

Employees

Surname	FirstName	Department	Head
Smith	Mary	Sales	De Rossi
Black	Lucy	Sales	De Rossi
Verdi	Mary	Personnel	Fox
Smith	Mark	Personnel	Fox

$\pi_{\text{Surname, FirstName}}(\text{Employees})$

Surname	FirstName
Smith	Mary
Black	Lucy
Verdi	Mary
Smith	Mark

Employees

Surname	FirstName	Department	Head
Smith	Mary	Sales	De Rossi
Black	Lucy	Sales	De Rossi
Verdi	Mary	Personnel	Fox
Smith	Mark	Personnel	Fox

$\pi_{\text{Department, Head}}(\text{Employees})$

Department	Head
Sales	De Rossi
Personnel	Fox

Vantaggio importante nella proiezione è il fatto che ha un effetto anche su quelli che sono i tempi di ricerca.

La selezione

La selezione si indica con il simbolo σ . Quando si esegue una selezione, si prendono tutti gli attributi della relazione, mantenendo quindi lo stesso numero di campi (attributi), ma si selezionano solo i record che soddisfano una determinata condizione o predicato. Questo significa che, sebbene il numero di attributi resti invariato, la **cardinalità** della relazione (cioè il numero di record) può cambiare, in quanto vengono mantenuti solo i record che rispettano la condizione specificata.

Ad esempio, l'operazione $\sigma_{\text{attributo1}=000}$ applicata su una tabella seleziona solo i record in cui il valore dell'attributo1 è uguale a 000. Di conseguenza, la relazione risultante avrà gli stessi campi della relazione originale, ma solo i record che soddisfano il predicato verranno mantenuti.

In generale, dopo l'operazione di selezione, si ottiene una relazione con lo stesso numero di campi, ma con una cardinalità **ridotta**, ovvero $K \leq M \leq M$, dove K è il numero di record nella relazione selezionata e M è il numero di record nella relazione originale.

Definizione formale della selezione

Sia R una relazione allora

$$\sigma_{\lambda}(R) = \{t \mid t \in R \wedge \lambda(t) = \text{TRUE}\}$$

dove λ è una formula proposizionale costruita a partire dagli atomi $A \theta B$, utilizzando i connettivi proposizionali. A e B sono attributi di R o costanti, e $\theta \in \{=, <, >, \neq, \leq, \geq\}$.

Employees

Surname	FirstName	Age	Salary
Smith	Mary	25	2000
Black	Lucy	40	3000
Verdi	Nico	36	4500
Smith	Mark	40	3900

Selezione

$$\sigma_{\text{Age} < 30 \vee \text{Salary} > 4000}(\text{Employees})$$

Surname	FirstName	Age	Salary
Smith	Mary	25	2000
Verdi	Nico	36	4500

Citizens

Surname	FirstName	PlaceOfBirth	Residence
Smith	Mary	Rome	Milan
Black	Lucy	Rome	Rome
Verdi	Nico	Florence	Florence
Smith	Mark	Naples	Florence

$$\sigma_{\text{PlaceOfBirth} = \text{Residence}}(\text{Citizens})$$

Surname	FirstName	PlaceOfBirth	Residence
Black	Lucy	Rome	Rome
Verdi	Nico	Florence	Florence

ATTENZIONE per stabilire se si tratta di una proiezione o di una selezione vedere se è presente la condizione(selezione), se la query richiede un attributo di un'entità si ha una selezione quindi si deduce che vuole sapere solo la colonna

Prodotto cartesiano

Supponiamo di avere due relazioni R ed S con attributi tutti diversi.

In simboli

$R(A_1 : T_1, \dots, A_n : T_n)$ e $S(A_{n+1} : T_{n+1}, \dots, A_{n+m} : T_{n+m})$ due relazioni con $\{A_1, \dots, A_n\} \cap \{A_{n+1}, \dots, A_{n+m}\} = \emptyset$

allora si pone $R \times S = \{tu \mid t \in R \wedge u \in S\}$

Esempio prodotto cartesiano

Employees

Employee	Project
Smith	A
Black	A
Black	B

Projects

Code	Name
A	Venus
B	Mars

Employees X Projects

Employee	Project	Code	Name
Smith	A	A	Venus
Black	A	A	Venus
Black	B	A	Venus
Smith	A	B	Mars
Black	A	B	Mars
Black	B	B	Mars