Progettazione logica

La progettazione logica consiste nella traduzione dello schema ER in uno specifico modello logico, in questo caso quello relazionale.

Costrutti principali

Domini di base

Consistono nei domini da cui si scelgono i valori delle proprietà delle istanze di informazione da rappresentare (es: caratteri, stringhe, interi, decimali a virgola mobile, ...).

Costrutto relazione

Il costrutto relazione può essere rappresentato come una tabella formata da una lista di colonne, dove:

- I dati sono scritti nelle righe;
- Ogni riga descrive le caratteristiche di una istanza dell'informazione da rappresentare;
- I valori contenuti nelle colonne descrivono sempre la stessa proprietà delle istanze di informazione da rappresentare.

Esempio:

MILANO	20100	1300000
VERONA	37100	350000
BRESCIA	25100	250000

Definizione formale come insieme di ennuple (lists): dati n insiemi di valori (domini) $D_1, \ldots, D_n$ con $n > 0$ e indicato con $D_1 \times \cdots \times D_n$ il loro prodotto cartesiano:

$$D_1 \times \cdots \times D_n = \{(v_1, \ldots, v_n) \mid v_1 \in D_1 \wedge \cdots \wedge v_n \in D_n\}$$

una relazione ρ di grado n è un qualsiasi sottoinsieme di $D_1 \times \cdots \times D_n$:

$$\rho \subseteq D_1 \times \cdots \times D_n$$

dove:

- $(v_1, \ldots, v_n)$ è una ennupla della relazione;
- $|\rho|$ è la cardinalità della relazione (numero di ennuple).

I domini $D_1, \ldots, D_n$ possono essere a cardinalità infinita, mentre le relazioni sono sempre a cardinalità finita; non esiste alcun ordinamento delle tuple; non è possibile duplicare una ennupla; i valori all'interno delle ennuple sono ordinati.

Esempio

Relazione delle città:

$$\rho = \{(MILANO, 20100, 1.300.000), (VERONA, 37100, 350.000), (BRESCIA, 25100, 250.000)\}$$
$$\rho \subseteq D_1 \times D_2 \times D_3$$

- $D_1 =$ Stringhe di caratteri;
- $D_2 =$ Numeri interi;
- $D_3 =$ Numeri interi.

Se t è una ennupla $(v_1, \ldots, v_n)$ il valore posto in i-esima posizione si indica con la notazione:

$$t[i]$$

Poichè questa modalità di accesso ai valori non è efficace per l'uso pratico delle relazioni, si preferisce quindi assegnare un nome alle colonne, portando all'introduzione della definizione di relazione come insieme di tuple.

Definizione formale come insueme di tuple (mappings): sia X un insieme di nomi e sia Δ l'insieme di tutti i domini di base ammessi dal modello. Allora si può definire una funzione

$$DOM : X \rightarrow \Delta$$

che associa ad ogni nome A di X un dominio $DOM(A)$ di Δ. I nomi di X si dicono attributi.
Una tupla su X è una funzione:

$$t : X \rightarrow \bigcup_{A \in X} DOM(A)$$

dove

$$t[A] = v \in DOM(A)$$

Una relazione su X è un insieme di tuple su X, dove X è l'insieme di attributi della relazione.

Esempio

Relazione delle città:

- X = {Nome, CAP, Abitanti};
- DOM(Nome) = Stringhe di caratteri;
- DOM(CAP) = Numeri interi;
- DOM(Abitanti) = Numeri interi.

$$\rho_X = \{t_1, t_2, t_3\}$$

t_1 [Nome] = MILANO	t_2 [Nome] = VERONA	$t_3[\ldots] = \ldots$
t_1 [CAP] = 20100	t_2 [CAP] = 37100	
t_1 [Abitanti] = 1.300.000	t_2 [Abitanti] = 350.000	

Una relazione è un insieme di tuple e quindi non può contenere tuple duplicate; i domini per gli attributi possono essere solo domini di base (non sono ammessi altri domini, né il prodotto cartesiano di domini); in generale una base di dati relazionale è costituita da più relazioni.

Progettazione

Nel caso in cui in una relazione, ovvero una tabella, si uniscono concetti disomogenei e con un'esistenza autonoma (come raggruppare tutti i dati nella stessa tabella), si rischia di creare ridondanza, ovvero una ripetizione inutile di più tuple.
La ridondanza può, a sua volta, creare le seguenti anomelie:

- Anomalia di **aggiornamento** -> per aggiornare il valore di un attributo si è obbligati a modificare tale valore su più tuple della base di dati.
- Anomalia di **inserimento** -> per inserire una nuova istanza di un concetto è necessario inserire valori al momento sconosciuti (sostituibili con valori nulli) per gli attributi non disponibili.
- Anomalia di **cancellazione** -> per cancellare un'istanza di un concetto è necessario cancellare valori ancora validi oppure inserire valori nulli per gli attributi da cancellare.

Il metodo principale per eliminare le ridondanze consiste nel dividere la tabella in tabelle più piccole.

Esempio:

PROPRIETÀ

CodiceFiscale	Cognome	Nome	DataNas	CodiceCatasto	Via	NumeroCivico	Subalterno	Tipo

La tabella unica PROPRIETÀ può essere divisa come segue:

PROPRIETARIO

CodiceFiscale	Cognome	Nome	DataNas

UNITÀ_IMMOBILIARE

CodiceCatasto	Via	NumeroCivico	Subalterno	Tipo

La suddivisione dell'esempio sopra, però, non riesce a rappresentare la stessa informazione della tabella unica, sebbene ne elimini la ridondanza. Infatti non conserva l'informazione che descrive l'associazione tra proprietario e unità immobiliare.
Per mantenere anche questa informazione è necessario aggiungere una terza tabella che, per sua natura, conterrà delle ridondanze (seppur limitate).

Esempio

PROPRIETARIO

CodiceFiscale	Cognome	Nome	DataNas

UNITÀ_IMMOBILIARE

CodiceCatasto	Via	NumeroCivico	Subalterno	Tipo

PROPRIETÀ

CodiceFiscale	CodiceCatasto

La nuova tabella replicherà parte degli attributi, scegliendo tra questi quelli che hanno la proprietà di identificare il concetto verso il quale si vuole generare il legame.

Modello valued - based In base a questo, si può definire il modello relazionale come valued - based, ovvero:

- È totalmente indipendente dalla rappresentazione fisica (tutta l'informazione è nei valori) e non ci sono meccanismi per gestire riferimenti o puntatori tra istanze di informazione;
- I legami logici tra tuple diverse si realizzano attraverso la replicazione di alcuni attributi (che hanno la proprietà di identificare il concetto e quindi di rappresentarlo): il legame tra due tuple si intende stabilito quando esse presentano gli stessi valori negli attributi replicati;
- È facile trasferire i dati da una base di dati all'altra;
- È rappresentato solo ciò che è rilevante per l'applicazione.

Terminologia

Schema di una relazione Uno schema di una relazione è costituito dal nome della relazione e da un insieme di nomi per i suoi attributi:

$$R(A_1, \ldots, A_n) \quad \text{oppure} \quad R(A_1 : D_1, \ldots, A_n : D_n)$$

dove $DOM(A_i) = D_i$

Schema di una base di dati Uno schema di una base di dati è un insieme di schemi di relazione:

$$S = \{R_1(A_{1,1}, \ldots, A_{1,n_1}), \ldots, R_m(A_{m,1}, \ldots, A_{m,n_m})\}$$

dove $R_1 \neq \cdots \neq R_m$

Istanza di una relazione Un'istanza di una relazione di schema $R(A_1, \ldots, A_n)$ con $X = \{A_1, \ldots, A_n\}$: è un insieme r di tuple su X.

Istanza di una base di dati Un'istanza di una base di dati di schema $S = \{R_1(A_{1,1}, \ldots, A_{1,n_1}), \ldots, R_m(A_{m,1}, \ldots, A_{m,n_m})\}$:
è un insieme di istanze di relazioni $db = \{r_1, \ldots, r_m\}$ dove ogni r_i è un'istanza della relazione di schema $R_i(A_{i,1}, \ldots, A_{i,n_i})$.