

# Basi di Dati, Progetto di Laboratorio

Federico Serra, matricola 898925

Alessandro Nocera, matricola 886732

## Indice

<b>1</b>	<b>Progettazione concettuale</b>	<b>1</b>
1.1	Requisiti iniziali . . . . .	1
1.2	Glossario dei termini . . . . .	2
1.3	Requisiti rivisti e strutturati in gruppi di frasi omogenee . . . . .	3
1.4	Schema E-R + business rules . . . . .	5
1.4.1	Scherma E-R . . . . .	5
1.4.2	Business rules . . . . .	6
<b>2</b>	<b>Progettazione logica</b>	<b>7</b>
2.1	Tavola dei volumi . . . . .	7
2.2	Tavola delle operazioni . . . . .	8
2.3	Ristrutturazione dello schema E-R . . . . .	9
2.3.1	Analisi delle ridondanze . . . . .	9
2.3.2	Eliminazione delle generalizzazioni . . . . .	10

## 1 Progettazione concettuale

### 1.1 Requisiti iniziali

Si vuole realizzare una base di dati per la gestione di una piattaforma che fornisce informazioni su film in uscita nei cinema e su serie e programmi TV in onda, liberamente ispirato a piattaforme come ComingSoon.

Un numero limitato di utenti della redazione si occupa di aggiornare le informazioni sui contenuti disponibili inserendo i dati di film e serie, comprese le date di uscita e programmazione in sala e in TV. Degli utenti si memorizzano nome utente, password e indirizzo email. Per i redattori si tiene traccia anche della data di inizio collaborazione.

I contenuti hanno una serie di caratteristiche ~~come evidenziato in Figura 1~~, quali: data di uscita, genere, anno, regia, paese, durata, distribuzione, sceneggiatura, fotografia, musiche e produzione. In particolare, a ogni film e serie TV è associata una lista di attori, con l'indicazione del personaggio interpretato. Notare che alcune informazioni tipiche dei film, ad esempio la regia, per le serie TV sono specifiche dei singoli episodi. Attori e registi hanno una propria scheda ~~(esempio in Figura 2)~~ dove vengono visualizzati foto, nome,

cognome, dati anagrafici e biografia. Qui può essere anche mostrata una lista dei contenuti più recenti a cui la persona ha partecipato.

I film vengono proiettati in svariati cinema in tutto il paese. Per ogni cinema, si tiene traccia del suo nome, contatti, e localizzazione (regione, provincia e indirizzo). La proiezione avviene a degli orari specifici in diverse date, in una delle sale disponibili al cinema, e comporta il pagamento di un biglietto il cui prezzo è indicato sul sito.

Programmi e serie TV vanno invece in onda su diverse piattaforme (es. Netflix, Rai Gulp, ...), che possono essere canali TV tradizionali o piattaforme di streaming video. Le serie TV sono organizzate per genere, come i film, es. commedia, drammatico, etc. In più, i serial sono composti da episodi, racchiusi in una o più stagioni, dove ogni episodio ha uno specifico titolo, durata, regista, e sceneggiatore. Anche per i programmi sono memorizzate alcune informazioni, come l'anno di messa in onda, genere, paese, durata e una descrizione testuale.

I visitatori del sito, previa registrazione tramite email e password, possono lasciare un voto (da 1 a 5 stelline) su qualunque contenuto. La media dei voti ricevuti da film e programmi viene mostrata nelle rispettive schede. Inoltre, gli utenti possono cliccare su un link "inserisci tra i preferiti" presente nelle schede dei contenuti, in modo da poter visualizzare in seguito i contenuti salvati. Nel momento della cancellazione di un utente dal sito, i suoi preferiti possono essere rimossi dal database, mentre i voti espressi dal visitatore rimangono in memoria.

## **1.2 Glossario dei termini**

<i>Termine</i>	<i>Descrizione</i>	<i>Sinonimi</i>	<i>Collegamento</i>
<b>Piattaforme</b>	Servizi informatici che rendono disponibili agli utenti le stagioni di una serie	Piattaforme di streaming video	Serie
<b>Contenuti</b>	Insieme di produzioni audiovisive (film, serie e programmi)		Film, serie, programmi
<b>Utente</b>	Persona registrata sul sito	Visitatori	Voto
<b>Redattore</b>	Utente che aggiorna informazioni riguardanti i contenuti presenti sul sito	Utenti della redazione	Voto
<b>Serie</b>	Serie TV disponibili sul sito	Serie TV, serial	Stagione, episodio, piattaforma
<b>Stagione</b>	Raggruppamento in episodi di una serie		Episodio, serie
<b>Episodio</b>	Singola unità di una stagione		Stagione, attore, regista
<b>Film</b>	Film presenti sul sito		Cinema, attore, regista
<b>Attore</b>	Colui che recita in un contenuto interpretando un ruolo	Persona	Contenuto
<b>Regista</b>	Colui che si occupa della regia dei contenuti	Regia, persona	Contenuto
<b>Cinema</b>	Luogo in cui è proiettato un film		Film
<b>Voto</b>	Voto numerico (da 1 a 5 stelle) espresso da un utente del sito		Contenuto, utente

### 1.3 Requisiti rivisti e strutturati in gruppi di frasi omogenee

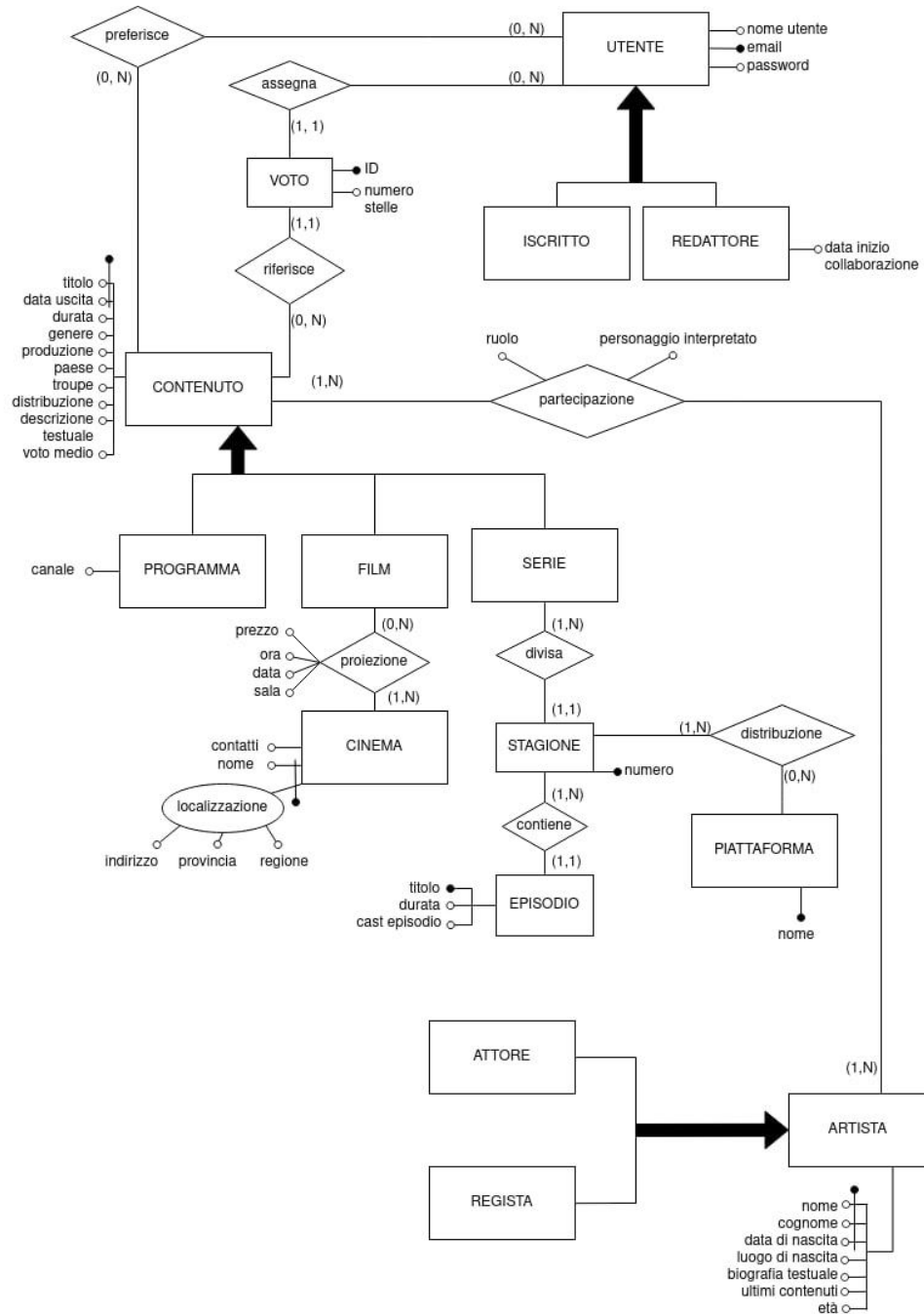
- **Frasi di carattere generale:** Si vuole realizzare una base di dati per la gestione di un sito che fornisce informazioni su film in uscita nei cinema e su serie e programmi TV in onda, liberamente ispirato a piattaforme come ComingSoon.
- **Frasi relative agli utenti:** Gli utenti possono essere di due tipologie: iscritti o redattori. Entrambi si registrano tramite email e password. Possono votare e aggiungere ai preferiti qualunque contenuto. Nel momento della cancellazione di un utente dal sito, i suoi preferiti possono essere rimossi dal database, mentre i voti espressi rimangono in memoria. Per gli utenti si memorizzano nome utente, password e indirizzo email.
- **Frasi relative ai redattori:** I redattori sono utenti che si occupano di aggiornare le informazioni sui contenuti disponibili inserendo i dati dei contenuti, comprese le date di uscita e programmazione nei cinema e nei canali televisivi. Per i redattori si registra anche della data di inizio collaborazione.
- **Frasi relative ai contenuti:** I contenuti possono essere di tre tipi: serie, film e programmi. Per i contenuti rappresentiamo: titolo, data di uscita, genere, paese, durata, distribuzione, produzione e i membri della troupe (attori con relativo personaggio interpretato, regista, sceneggiatori, i direttori della fotografia, i responsabili delle musiche).

Per i programmi memorizziamo anche il canale dove è disponibile il contenuto.  
Per i film memorizziamo invece anche il cinema dove saranno disponibili.

- **Frasi relative alle serie:** Le serie sono contenuti divisi in stagioni. Le stagioni vanno in onda su siti di streaming online e sono divise in episodi. Per ogni episodio si memorizza titolo, durata, regista, sceneggiatore e la lista di attori che vi partecipano.
- **Frasi relative ai cinema:** Per i cinema rappresentiamo nome, contatti, e localizzazione (regione, provincia e indirizzo). La proiezione dei film è caratterizzata da: una sala del cinema, una data e un'ora, il prezzo del biglietto.
- **Frasi relative agli artisti:** Gli artisti possono essere attori o registi. Per gli artisti è registrata una scheda personale contenente: foto, nome, cognome, età, data e luogo di nascita, biografia e gli ultimi contenuti a cui ha partecipato.

## 1.4 Schema E-R + business rules

### 1.4.1 Scherma E-R



#### **1.4.2 Business rules**

- Quando un utente viene cancellato i voti espressi rimangono, dati personali e preferenze vengono rimossi.
- Il valore del voto è compreso tra 1 e 5 stelle.
- L'attributo voto medio di un contenuto indica la media dei voti assegnati dagli utenti.
- Per il singolo artista viene mantenuta una lista dei contenuti a cui ha partecipato.
- I redattori aggiornano le informazioni relative ai contenuti. (? e degli artisti)

## 2 Progettazione logica

### 2.1 Tavola dei volumi

<i>Concetto</i>	<i>Tipo</i>	<i>Volume</i>	<i>Motivazione</i>
<b>Utente</b>	Entità	50.000	Approssimazione della somma tra Iscritti e Redattori
<b>Iscritto</b>	Entità	50.000	Ci si basa sulla piattaforma italiana mymovies che contava nel 2018 una media di 500.000 utenti unici giornalieri, si suppone che il 10% sia registrato.
<b>Redattore</b>	Entità	10	Mymovies conta 5 content manager, più altri collaboratori. Si sceglie 10 come numero di profili abilitati alla modifica calcolandone uno per ogni content manager più altri 5 divisi tra i collaboratori.
<b>Preferisce</b>	Relazione	250.000	Si considera una media di 5 film preferiti per utente
<b>Assegna</b>	Relazione	5	Media di 5 voti per utente ottenuta dividendo i voti totali per gli utenti
<b>Voti</b>	Entità	250.000	Da mymovies si evince che ci sono circa 200.00 recensioni totali, si considera un voto per recensione più un voto per utente che non lascia il commento
<b>Voto contenuto</b>	Relazione	500	Si considera che il 1% di utenti vota un film.
<b>Contenuti</b>	Entità	40.000	Basandosi su dati di un articolo su mymovies <sup>1</sup>
<b>Partecipazione</b>	Relazione	20	Si stima una media di partecipazione a 20 contenuti per artista
<b>Programma</b>	Entità	500	Approssimazione dati mymovies
<b>Film</b>	Entità	10.000	Approssimazione dati mymovies
<b>Proiezione</b>	Relazione	420.000	Si stima che in ogni cinema ci siano 4 spettacoli al giorno, moltiplicati per i 30 giorni (un film rimane in programmazione un mese), moltiplicati per i cinema in Italia
<b>Cinema</b>	Entità	3.500	Numero di cinema presenti su mymovies
<b>Serie</b>	Entità	1.750	Approssimazione basata sul fatto che sono 600 le serie presenti su Netflix, considerando che non è presente 1 serie su 3.
<b>Divisa</b>	Relazione	3.500	Stesso valore di Stagione.
<b>Stagione</b>	Entità	3.500	Media di 2 stagioni a serie.
<b>Contiene</b>	Relazione	29.500	Stesso valore di Episodio.
<b>Distribuzione</b>	Relazione	600	Si stima che per ogni piattaforma è presente il circa il 30% delle serie.
<b>Episodio</b>	Entità	29.500	Media di 8-9 episodi a stagione
<b>Piattaforma</b>	Entità	10	Si considerano le 10 principali piattaforme in Italia
<b>Attore</b>	Entità	3.000	Gli artisti rimanenti togliendo i registi
<b>Regista</b>	Entità	2.000	Si considera una media di 20 contenuti girati per regista
<b>Artisti</b>	Entità	5.000	Dati approssimati ComingSoon

<sup>1</sup> Articolo mymovies: [www.mymovies.com](http://www.mymovies.com)

## 2.2 Tavola delle operazioni

<i>Operazione</i>	<i>Tipo</i>	<i>Frequenza</i>	<i>Motivazione</i>
<b>Registrazione utente</b>	Interattivo	10/giorno	Si considera il numero di utenti totali attuali dopo 15 anni di attività
<b>Cancellazione utente</b>	Interattivo	1/giorno	Si considera che l'eliminazione avvenga meno raramente della registrazione, si ottiene così una media di +9 utenti/giorno coerente con i valori riportati nella tavola dei volumi
<b>Modifica dati utente</b>	Interattivo	5/mese	Operazione fatta raramente
<b>Assegnazione di un voto a contenuto</b>	Interattivo	45/giorno	Si considera il numero di voti dopo 15 anni di attività
<b>Aggiornamento voto medio di un contenuto</b>	Batch	45/giorno	Uguale al numero di assegnazioni di voto
<b>Aggiunta di un contenuto ai preferiti</b>	Interattivo	45/giorno	Numero totale dei preferiti suddivisi in 15 anni di attività
<b>Eliminazione di un contenuto tra i preferiti</b>	Interattivo	1/giorno	Molto inferiore al numero di aggiunta dei preferiti
<b>Aggiunta di un contenuto</b>	Interattivo	3/giorno	Si considerano film e serie presenti sulla piattaforma dopo 15 anni di attività
<b>Modifica contenuto</b>	Interattivo	1/mese	Si suppone di dover modificare un contenuto molto raramente
<b>Aggiornamento cinema di proiezione del film</b>	Interattivo	90/mese	1/mese per i film attualmente al cinema, contando una media di 1000 film prodotti all'anno
<b>Aggiunta stagione alla serie</b>	Interattivo	1/mese	Si considera la media uscita di una nuova serie di 1 al mese
<b>Aggiunta artista</b>	Interattivo	1/mese	Si considera che l'aggiunta della scheda di un artista avvenga raramente, più raramente della modifica
<b>Modifica/aggiornamento artista</b>	Interattivo	3/giorno	Legato all'aggiunta di un contenuto
<b>Visualizzazione preferiti</b>	Interattivo	800/giorno	Si considera che un utente guarda i propri preferiti raramente, una volta ogni due mesi
<b>Visualizzazione contenuto</b>	Interattivo	500.000/giorno	Ci si basa sui dati di mymovies, è l'operazione in assoluto più frequente
<b>Visualizzazione artista</b>	Interattivo	1.000/giorno	Minore della visualizzazione del contenuto, si stima $\frac{1}{5}$ delle visualizzazioni del contenuto

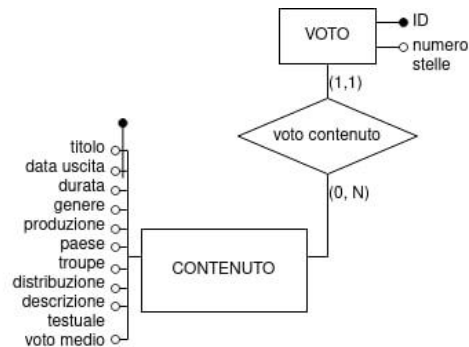


## 2.3 Ristrutturazione dello schema E-R

### 2.3.1 Analisi delle ridondanze

1. Attributo “età” di “artista” (attributo derivabile): l’attributo “età” è derivabile considerando l’anno di nascita e la data odierna. Il mantenimento di questo attributo comporta l’aggiornamento costante di un dato (“età”) secondo la data di nascita dell’artista. Per il precedente motivo si è scelto di eliminare la ridondanza, diminuendo gli aggiornamenti dei dati relativi all’artista.
2. Attributo “Voto medio” di “contenuto” (attributo derivabile da entità e conteggio): l’attributo “voto medio” è derivabile dal conteggio delle occorrenze dell’entità “voto” facente riferimento ad un dato “contenuto”. In questa somma, se si tiene anche conto del valore dei punteggi (il numero di stelle), si può facilmente derivare che  $\frac{\text{totalepunteggi}}{\text{cardinalitàvoto}} = \text{votomedio}$ .

Si è scelto di analizzare la seconda ridondanza in quanto ritenuta più significativa.



#### Operazione 1 - assegnazione di un voto a un contenuto -

*Accessi con ridondanza:*

Concetto	Costrutto	Accessi	Tipo
<b>Voto</b>	Entità	1	Scrittura
<b>Voto</b>	Entità	1	Lettura
<b>Voto contenuto</b>	Relazione	1	Scrittura
<b>Contenuto</b>	Entità	1	Scrittura

Il costo è di  $(45 \times 2) + 45 + (45 \times 2) + (45 \times 2) = 315$  accessi al giorno, contando gli accessi in scrittura come doppi.

*Accessi senza ridondanza:*

Concetto	Costrutto	Accessi	Tipo
<b>Voto</b>	Entità	1	Scrittura
<b>Voto contenuto</b>	Entità	1	Scrittura

Il costo è di  $(45 \times 2) + (45 \times 2) = 180$  accessi al giorno, contando gli accessi in scrittura come doppi.

## Operazione 2 - visualizzazione contenuto (include il voto medio) -

*Accessi con ridondanza:*

<i>Concetto</i>	<i>Costrutto</i>	<i>Accessi</i>	<i>Tipo</i>
<b>Contenuto</b>	Entità	1	Lettura

Il costo è di 500000 accessi al giorno.

*Accessi senza ridondanza:*

<i>Concetto</i>	<i>Costrutto</i>	<i>Accessi</i>	<i>Tipo</i>
<b>Voto</b>	Entità	1	Lettura
<b>Voto contenuto</b>	Relazione	1	Lettura
<b>Contenuto</b>	Entità	1	Lettura

Il costo è di  $500000 + 500000 + 500000 = 1500000$  accessi al giorno.

## Operazione 1 + operazione 2 - costi totali -

*Con ridondanze (operazione 1 + operazione 2):*

Costo totale in numero di accessi  $315 + 500000 = 500315$ .

*Senza ridondanze (operazione 1 + operazione 2):*

Costo totale in numero di accessi  $180 + 1500000 = 1500180$ .

*Costi aggiuntivi in termini di spazio:*

Ipotesi: si utilizzano 4byte per memorizzare il valore del voto medio.

Spazio totale necessario:  $4 \times 40000 = 160000 = 160Kbyte$

	<i>Con ridondanza</i>	<i>Senza ridondanza</i>
<i>Numero accessi</i>	500315	1500180
<i>Spazio aggiuntivo</i>	160 Kbyte	0 Kbyte

## Decisione:

Considerata la differenza di circa 1000000 di accessi e lo spreco di memoria non ingente, si è scelto di lasciare la ridondanza, essendo anche l'operazione 2 la più frequente sulla base di dati.

### 2.3.2 Eliminazione delle generalizzazioni