

Basi di Dati, Progetto di Laboratorio

Federico Serra, matricola 898925

Alessandro Nocera, matricola 886732

Indice

1	Progettazione concettuale	1
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
2	Progettazione logica	7
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
2.3.3	Partizionamento/accorpamento di entità e associazioni	12
2.3.4	Scelta degli identificatori principali	12
2.3.5	Schema E-R ristrutturato + business rules	13

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

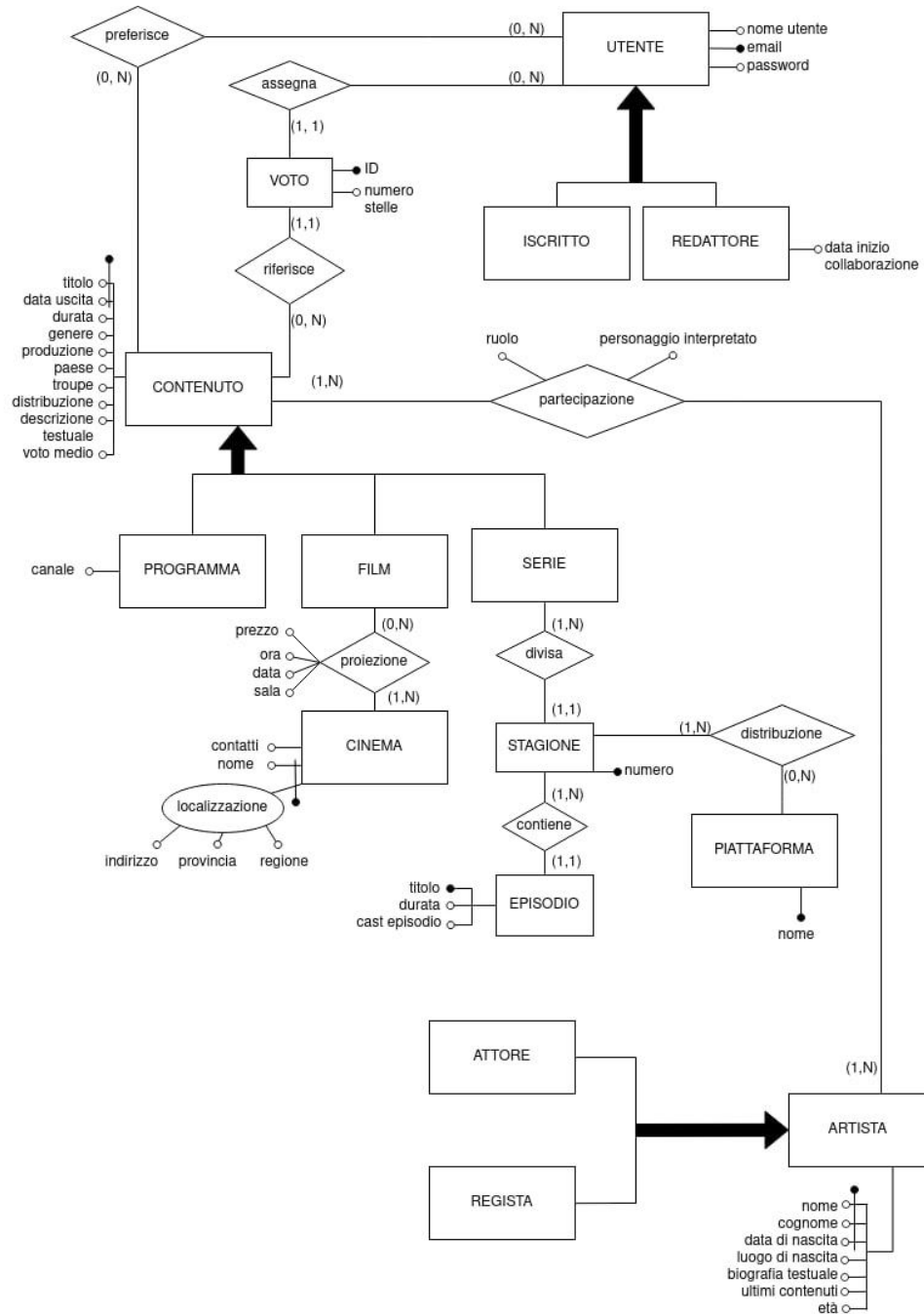
- **Frasi di carattere generale:** Si vuole realizzare una base di dati per la gestione di un sito che fornisce informazioni su film in uscita nei cinema e su serie e programmi TV in onda, liberamente ispirato a piattaforme come ComingSoon.
- **Frasi relative agli utenti:** Gli utenti possono essere 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 contenute divise 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.
- L'attributo ruolo dell'associazione partecipazione può assumere i valori regista o attore. Nel primo caso l'attributo personaggio interpretato avrà valore NULL, nel secondo invece avrà come valore il nome del personaggio interpretato dall'attore all'interno del contenuto.

2 Progettazione logica

2.1 Tavola dei volumi

<i>Concetto</i>	<i>Tipo</i>	<i>Volume</i>	<i>Motivazione</i>
Utente	Entità	50.000	Approssimazione della somma tra Iscritti e Redattori
Iscritto	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.
Redattore	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.
Preferisce	associazione	250.000	Si considera una media di 5 film preferiti per utente
Assegna	associazione	5	Media di 5 voti per utente ottenuta dividendo i voti totali per gli utenti
Voti	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
Voto contenuto	associazione	500	Si considera che il 1% di utenti vota un film.
Contenuti	Entità	40.000	Basandosi su dati di un articolo su mymovies ¹
Partecipazione	associazione	20	Si stima un media di partecipazione a 20 contenuti per artista
Programma	Entità	500	Approssimazione dati mymovies
Film	Entità	10.000	Approssimazione dati mymovies
Proiezione	associazione	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
Cinema	Entità	3.500	Numero di cinema presenti su mymovies
Serie	Entità	1.750	Approssimazione basata sul fatto che sono 600 le serie presenti su Netflix, considerando che non è presente 1 serie su 3.
Divisa	associazione	3.500	Stesso valore di Stagione.
Stagione	Entità	3.500	Media di 2 stagioni a serie.
Contiene	associazione	29.500	Stesso valore di Episodio.
Distribuzione	associazione	600	Si stima che per ogni piattaforma è presente il circa il 30% delle serie.
Episodio	Entità	29.500	Media di 8-9 episodi a stagione
Piattaforma	Entità	10	Si considerano le 10 principali piattaforme in Italia
Attore	Entità	3.000	Gli artisti rimanenti togliendo i registi
Regista	Entità	2.000	Si considera una media di 20 contenuti girati per regista
Artisti	Entità	5.000	Dati approssimati ComingSoon

¹ Articolo mymovies: www.mymovies.com

2.2 Tavola delle operazioni

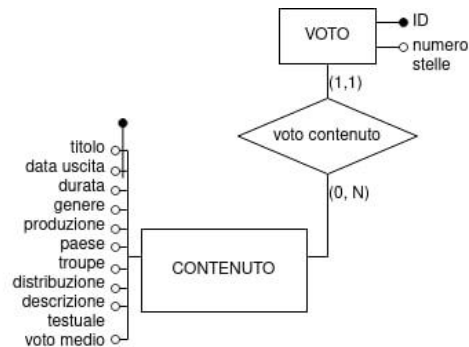
<i>Operazione</i>	<i>Tipo</i>	<i>Frequenza</i>	<i>Motivazione</i>
Registrazione utente	Interattivo	10/giorno	Si considera il numero di utenti totali attuali dopo 15 anni di attività
Cancellazione utente	Interattivo	1/giorno	Si considera che l'eliminazione avvenga meno raramente della registrazione, si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

Concetto	Costrutto	Accessi	Tipo
Contenuto	Entità	1	Lettura

Il costo è di 500000 accessi al giorno.

Accessi senza ridondanza:

Concetto	Costrutto	Accessi	Tipo
Voto	Entità	1	Lettura
Voto contenuto	associazione	1	Lettura
Contenuto	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: $4 \times 40000 = 160000 = 160Kbyte$

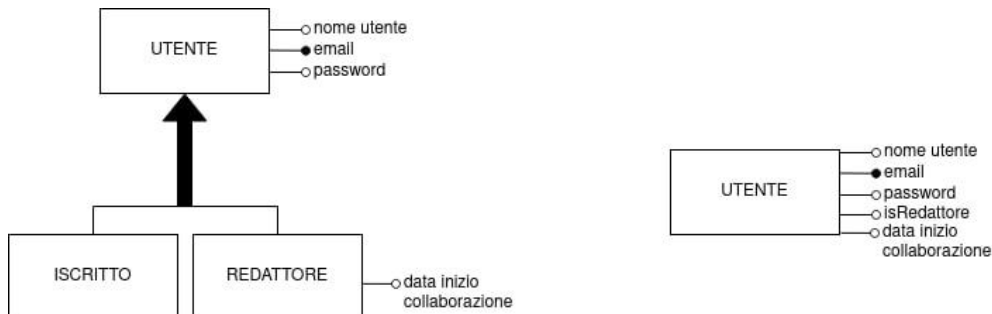
	Con ridondanza	Senza ridondanza
Numero accessi	500315	1500180
Spazio aggiuntivo	160Kbyte	0Kbyte

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.

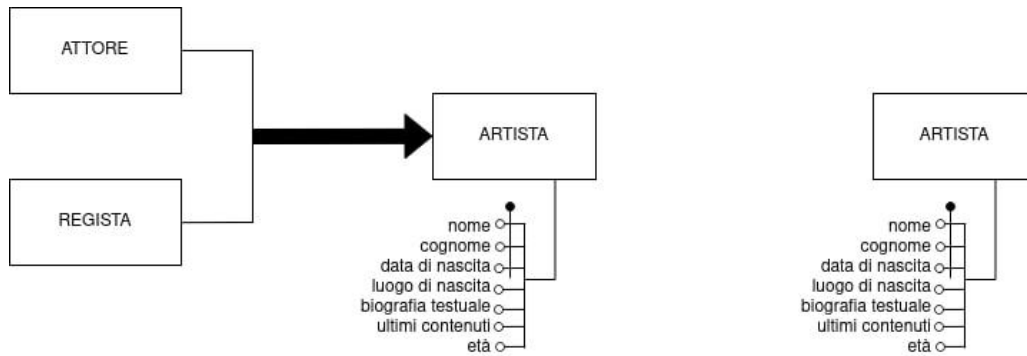
2.3.2 Eliminazione delle generalizzazioni

1. Utente - Iscritto, Redattore (*generalizzazione totale ed esclusiva*)



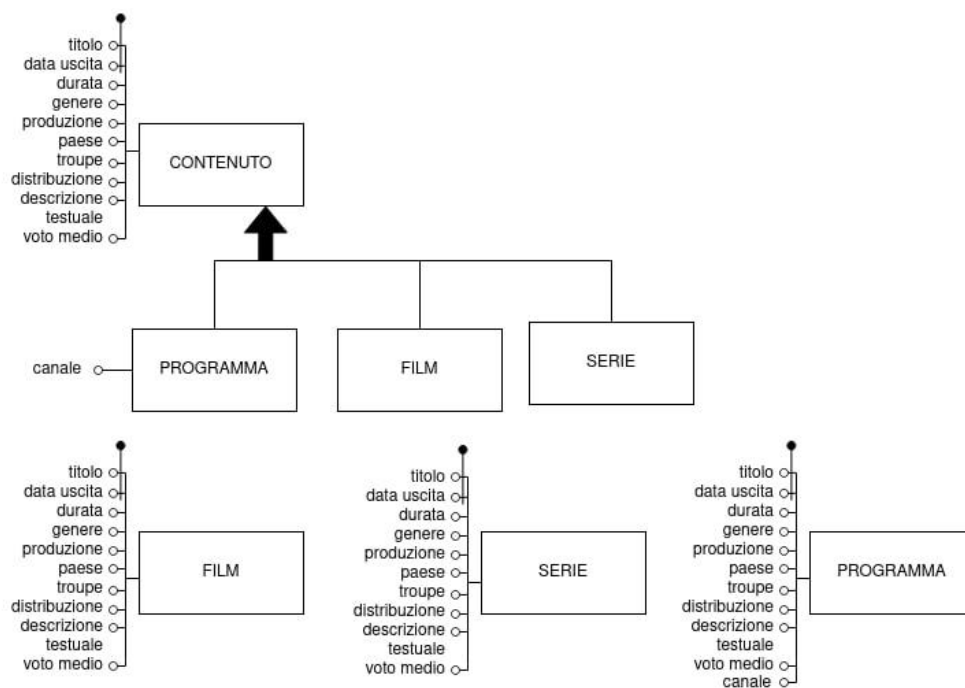
Motivazione: si è scelto di accorpare le entità figlie della generalizzazione nell'entità padre, semplificando così la base di dati. Si è aggiunto a tale scopo un flag *isRedattore* per indicare se l'utente è un iscritto o un redattore.

2. Artista - Attore, Regista (*generalizzazione totale ed esclusiva*)



Motivazione: si è scelto di accorpare le entità figlie della generalizzazione nell'entità padre, in quanto gli accessi alle tabelle sono per la maggior parte contestuali. Se osserviamo la tavola delle operazioni infatti notiamo che la visualizzazione di un artista avviene 1000 volte in un giorno, la visualizzazione di un contenuto (che richiede l'accesso ad entrambe le tabelle) invece occorre 500000 volte nello stesso tempo. Per questo accorpamento non è stata necessaria l'aggiunta di attributi, in quanto la differenza tra attore e regista viene già espressa dall'attributo ruolo della associazione partecipazione.

3. Contenuto - Programma, Film, Serie (*generalizzazione totale ed esclusiva*)



Motivazione: si è scelto di accorpare la entità padre della generalizzazione nelle entità figlie, in quanto gli accessi alla tabelle figlie sono distinti.

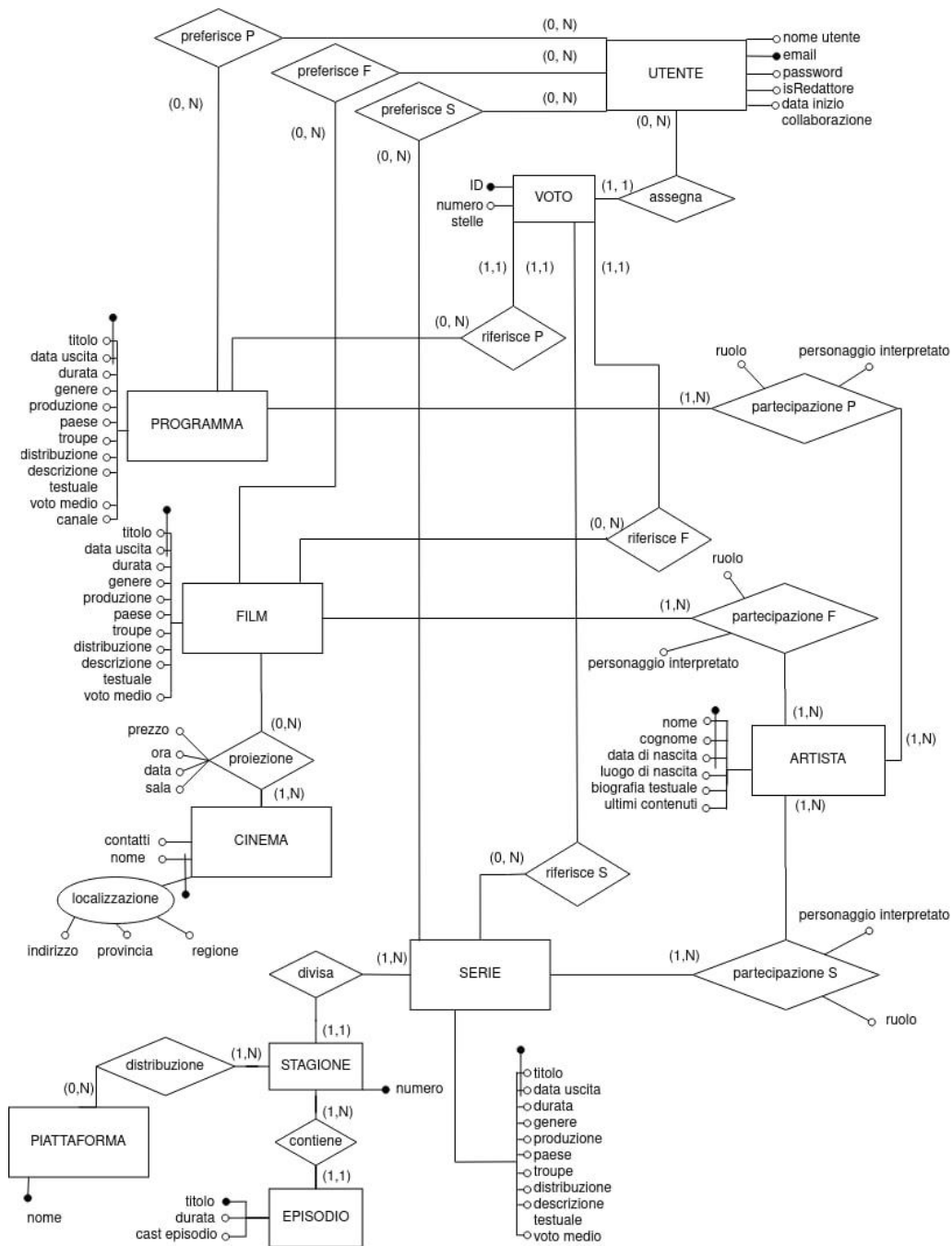
2.3.3 Partizionamento/accorpamento di entità e associazioni

Non si è valutato necessario il partizionamento e/o l'accorpamento di entità e associazioni a fronte delle operazioni previste sulla base di dati, descritte nella tavola delle operazioni.

2.3.4 Scelta degli identificatori principali

La descrizione degli identificatori principali non è stata trattata in quanto si ritiene sufficientemente auto-esplicativa a partire dallo schema E-R + business-rules.

2.3.5 Schema E-R ristrutturato + business rules



Business rules:

- Se un utente ha il flag `isRedattore = TRUE` allora l'attributo `data inizio collaborazione` IS NOT NULL.

- Se viene modificato il flag `isRedattore` a `FALSE` l'attributo `data inizio collaborazione` assume valore `NULL`.
- Se viene modificato il flag `isRedattore` a `TRUE` l'attributo `data inizio collaborazione` assume come valore la data del momento in cui viene modificato il flag (il giorno corrente).
- Il valore del voto è compreso tra 1 e 5 stelle.
- L'attributo voto medio di film indica la media dei voti assegnati dagli utenti.
- L'attributo voto medio di serie indica la media dei voti assegnati dagli utenti.
- L'attributo voto medio di programma indica la media dei voti assegnati dagli utenti.
- Per il singolo artista viene mantenuta una lista dei contenuti a cui ha partecipato.
- Gli utenti con flag `isRedattore = TRUE` aggiornano le informazioni relative a: film, serie, programmi.
- L'attributo ruolo dell'associazione partecipazione F può assumere i valori regista o attore. Nel primo caso l'attributo personaggio interpretato avrà valore `NULL`, nel secondo invece avrà come valore il nome del personaggio interpretato dall'attore all'interno del film.
- L'attributo ruolo dell'associazione partecipazione S può assumere i valori regista o attore. Nel primo caso l'attributo personaggio interpretato avrà valore `NULL`, nel secondo invece avrà come valore il nome del personaggio interpretato dall'attore all'interno della serie.
- L'attributo ruolo dell'associazione partecipazione P può assumere i valori regista o attore. Nel primo caso l'attributo personaggio interpretato avrà valore `NULL`, nel secondo invece avrà come valore il nome del personaggio interpretato dall'attore all'interno del programma.