



UNIVERSITÀ DI CATANIA

PROGETTO

Database

Autore:
Giuseppe PITRUZZELLA

Professore:
Salvatore ALAIMO

1 Introduzione

Il progetto mira a realizzare un sistema per la gestione di un applicativo come Audible, applicazione proprietaria di Amazon, cui proposizione di valore riguarda l'ascolto di podcast ed audiolibri. Ci concentreremo, inizialmente, sulla progettazione concettuale per poi attenzionare le fasi di progettazione logica e fisica.

2 Progettazione concettuale

2.1 Specifiche sui dati

Come enunciato all'interno del capitolo precedente, il progetto ha lo scopo di creare una base di dati per un applicativo come *Audible*, un servizio di streaming di audiolibri e podcast. I protagonisti dell'applicativo, in questo caso, sono tre: (i) il *cliente*, il quale ha la possibilità di ascoltare tutti gli *audiolibri* ed i *podcast* offerti all'interno della piattaforma, (ii) l'*autore* di un audiolibro o podcast presente all'interno del servizio e (iii) il *lettore*, ovvero coloro che leggono o, più in generale, parlano all'interno di un audiolibro o podcast. Per il nostro progetto supponiamo che vi siano all'incirca 1500 clienti e 700 artisti, di cui 500 autori e 200 lettori. Rappresentiamo ogni utente che usufruisce del servizio, ossia il **cliente** mediante un username, un email, una password, la propria data di nascita ed, infine, il paese in cui vive. In particolare, ogni utente verrà identificato dalla sua email. Questi primi utenti hanno la possibilità di ascoltare, recensire e aggiungere alla propria libreria virtuale, uno o più audiolibri o podcast. D'altra parte, rappresentiamo un **autore** mediante un ID, un nome ed il paese in cui vive. Si noti che ogni autore è identificato dal solo ID e che ognuno di essi possiede almeno un genere a cui si riferiscono tipicamente le sue opere. Ogni autore può scrivere uno o più audiolibri e/o podcast. Infine, rappresentiamo il **lettore** mediante un ID, un nome ed il paese in cui vive. Ogni lettore è identificato dal solo ID. Un lettore ha la sola possibilità di leggere un audiolibro o podcast. Si noti esistano autori che sono anche lettori; è il caso, per esempio, dei creatori di podcast, i quali oltre a "creare" il proprio podcast, spesso "leggono" anche all'interno di esso. Come accennato precedentemente, le opere presenti all'interno del servizio possono essere di due tipi: *podcast* e *audiolibro*. Si noti che ogni opera appartiene ad uno specifico genere. Un **podcast** è un file audio in cui viene raccontato, in uno o più episodi, dei temi legati al tema principale che ha scaturito la nascita del podcast stesso. Ogni podcast è rappresentato da un ID, un titolo, un riassunto che spiega

il tema principale e la sua votazione media, derivata a partire dal calcolo della media dei voti nelle recensioni ricevute. Un podcast è identificato dal suo ID. Si noti che ogni podcast si compone tendenzialmente di 10 episodi, ognuno dei quali è rappresentato da un ID, un titolo, il numero dell'episodio e la sua durata. Un **audiolibro** d'altra parte, è un file audio in cui viene narrato un intero libro, cui voce narrante spesso non è la voce dello scrittore del libro. Rappresentiamo un audiolibro secondo un ID, un titolo, una breve sintesi sul libro e la sua votazione media, anche in questo caso, derivata a partire dalle recensioni degli utenti per l'audiolibro. In particolare, un audiolibro è identificato dal suo ID. Si noti che ogni audiolibro si compone tendenzialmente di 30 capitoli, ognuno dei quali è rappresentato da un ID, un titolo, il numero del capitolo e la sua durata. Si noti che, relativamente ai dati che rappresentano un audiolibro o podcast, consideriamo anche un riferimento all'autore ed al lettore per la determinata opera. Per la nostra base di dati, supponiamo che all'interno del servizio siano presenti 2 000 opere, di cui 1 250 audiolibri e 750 podcast.

2.2 Operazioni sui dati

Indice	Operazione
1	Aggiungi un cliente
2	Visualizza tutte le recensioni per un determinato audiolibro
3	Aggiungi una recensione per un opera
4	Aggiungi una nuova opera alla libreria
5	Trova tutti gli audiolibri di genere 'Poliziesco'
6	Visualizza tutti gli audiolibri scritti da uno specifico autore
7	Visualizza tutte le opere con una votazione media maggiore o uguale a 4
8	Aggiungi un episodio per un podcast

2.3 Glossario dei termini

In questo paragrafo, effettuiamo un glossario di tutti i termini relativi al progetto;

Termini	Descrizione	Sinonimi	Collegamenti
Cliente	Utente che usufruisce del servizio, ascoltando audiolibri/podcast e scrivendo recensioni su di essi.	Consumer, buyer	Autore, audiolibro, podcast
Autore	Persona che ha effettivamente scritto un'opera presente all'interno del servizio.	Creatore, artista, scrittore	Cliente, lettore, audiolibro, podcast
Lettore	Persona che effettua la lettura di uno o più audiolibri o podcast.	Oratore, artista	Autore, audiolibro, podcast
Audiolibro	Libro trasformato nella sua versione audio, possibile attraverso il lavoro di un lettore.		Podcast
Podcast	Raccolta di episodi in formato audio, incentrati su uno o più temi.		Audiolibro
Libreria	Una libreria virtuale in cui saranno presenti tutti gli audiolibri e podcast che il cliente ha volontariamente deciso di inserire.	Libreria virtuale, biblioteca, archivio	Audiolibro, podcast
Genere	Etichetta attraverso cui è possibile notare il genere di audiolibro, quindi fornire una prima descrizione dell'audiolibro. E' utile anche all'autore, il quale possiede uno o più generi in base alle opere scritte.		Audiolibro, autore
Episodio	Una parte di un podcast		Podcast
Capitolo	Una parte di un audiolibro		Audiolibro
Sintesi	Breve descrizione dell'audiolibro o podcast.	Riassunto	Audiolibro, podcast
Recensione	Opinione di un utente su un certo audiolibro		Audiolibro, cliente
Opera	Audiolibro o podcast		Audiolibro, podcast

2.4 Raggruppamento e struttura dei requisiti

- **Frase di carattere generale.** Si vuole realizzare una base di dati per un applicativo *Audible-like*, il quale permette all'utente (cliente) di poter ascoltare opere come audiolibri e podcast. All'interno della piattaforma distinguiamo tre tipi di utenti: i *clienti*, gli *autori* ed i *lettori*.
- **Frase relative ai clienti.** Rappresentiamo ogni cliente mediante un username, un email, una password, la propria data di nascita ed, infine, il paese in cui vive. In particolare, ogni utente verrà identificato dalla sua email. Questi primi utenti hanno la possibilità di ascoltare, recensire e aggiungere alla propria libreria virtuale, uno o più audiolibri o podcast.
- **Frase relative agli autori.** Rappresentiamo gli autori mediante un ID, un nome ed il paese in cui vive. Si noti che ogni autore è identificato dal solo ID e che ognuno di essi possiede almeno un genere a cui si riferiscono tipicamente le proprie opere. Ogni autore può scrivere uno o più audiolibri e/o podcast.
- **Frase relative ai lettori.** Rappresentiamo il lettore mediante un ID, un nome ed il paese in cui vive. Ogni lettore è identificato dal solo ID. Un lettore ha la sola possibilità di leggere un audiolibro o podcast. Si noti esistano autori che sono anche lettori; è il caso, per esempio, dei creatori di podcast, i quali oltre a "creare" il proprio podcast, spesso "leggono" anche all'interno di esso.
- **Frase relative agli audiolibri.** Un audiolibro è un file audio in cui viene narrato un intero libro, cui voce narrante spesso non è la voce dello scrittore del libro. Rappresentiamo un audiolibro secondo un ID, un titolo, una breve sintesi sul libro e la sua votazione media, anche in questo caso, derivata a partire dalle recensioni degli utenti per l'audiolibro. In particolare, un audiolibro è identificato dal suo ID. Si noti che ogni audiolibro si compone tendenzialmente di 30 capitoli, ognuno dei quali è rappresentato da un ID, un titolo, il numero del capitolo e la sua durata. Relativamente ai dati che rappresentano un audiolibro, verrà considerato anche un riferimento all'autore ed al lettore per la determinata opera.
- **Frase relative ai podcast.** Un podcast è un file audio in cui viene raccontato, in uno o più episodi, dei temi legati al tema principale che

ha scaturito la nascita del podcast stesso. Ogni podcast è rappresentato da un ID, un titolo, un riassunto che spiega il tema principale e la sua votazione media, derivata a partire dal calcolo della media dei voti nelle recensioni ricevute. Un podcast è identificato dal suo ID. Si noti che ogni podcast si compone tendenzialmente di 10 episodi, ognuno dei quali è rappresentato da un ID, un titolo, il numero dell'episodio e la sua durata. Relativamente ai dati che rappresentano un podcast, verrà considerato anche un riferimento all'autore ed al lettore per la determinata opera.

2.5 Generalizzazione

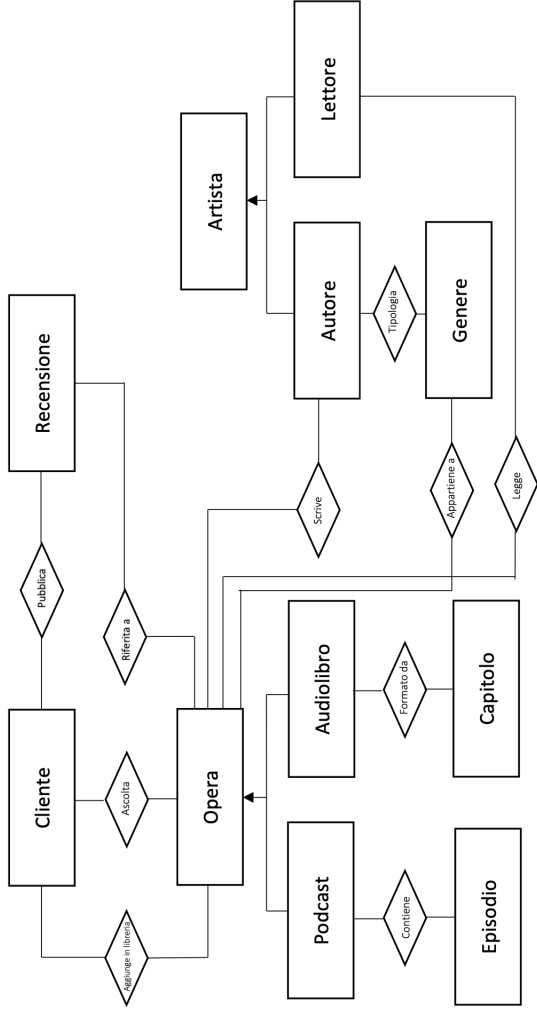
Una prima generalizzazione è possibile rispetto i creatori di contenuti per la piattaforma. Ci riferiamo agli *autori* ed i *lettori*, i quali essendo rappresentati dagli stessi dati possono essere intesi come casi specifici dell'entità "*Artista*". Questa prima generalizzazione è *totale*, poiché ogni occorrenza dell'entità genitore "*Artista*" è occorrenza di almeno una delle entità figlie, ed è inoltre *sovrapposta* poiché ogni occorrenza dell'entità genitore non è occorrenza di al più una delle entità figlie (quindi, esistono autori che sono anche lettori e viceversa). Una seconda generalizzazione avviene a partire dagli oggetti su cui si basa l'intera piattaforma, ovvero gli audiolibri ed i podcast. Quest'ultimi, infatti, sono rappresentati dalle stesse informazioni, per cui possono essere visti come entità figlie di un entità genitore "*Opera*". La generalizzazione è, in questo caso, *totale* poiché ogni occorrenza di "*Opera*" è occorrenza di almeno una delle entità figlie, ed *esclusiva* poiché ogni occorrenza di "*Opera*" è occorrenza di al più una delle entità figlie.

2.6 Strategie di progetto e schema concettuale

La strategia di progetto che abbiamo scelto per lo svolgimento del progetto è la strategia *top-down*. Attraverso la strategia *top-down*, lo schema concettuale viene prodotto mediante una serie di **raffinamenti** effettuati su uno **schema iniziale** che descrive tutte le specifiche con pochi concetti molto astratti. I raffinamenti avranno lo scopo di aumentare il dettaglio dei concetti inizialmente astratti.

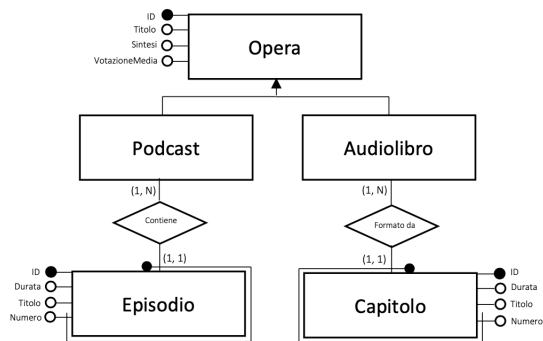
2.6.1 Schema scheletro

Procediamo inizialmente con uno schema scheletro, in cui rappresentiamo i dati da noi raccolti dal testo introduttivo.

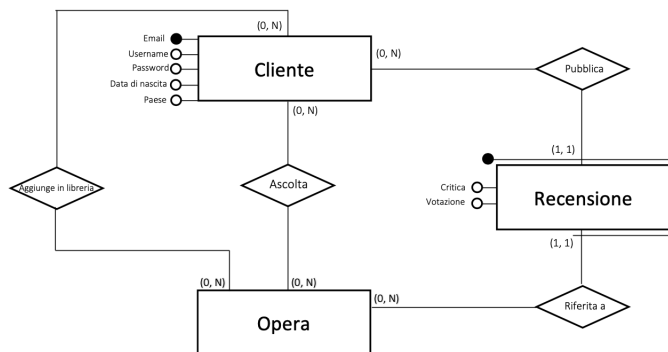


2.6.2 Raffinamenti

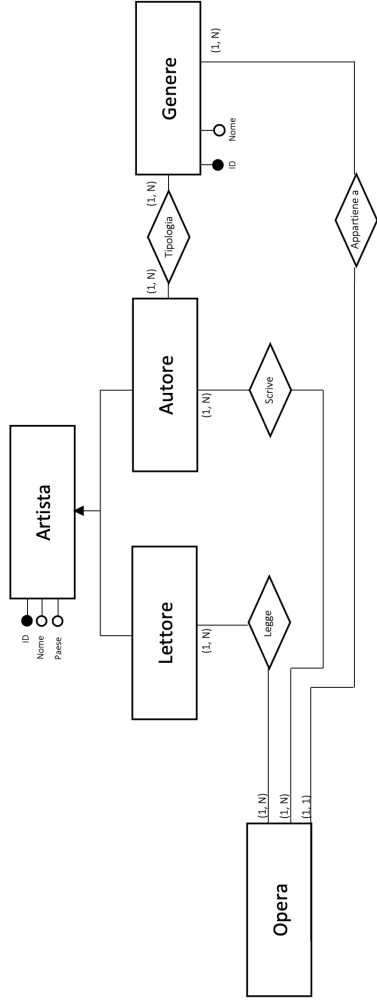
1. Un primo raffinamento è relativo alle entità (specifiche) "Podcast" ed "Audiolibro", le quali vengono raggruppate da un'entità (generale) "Opera". Queste due entità figlie, come abbiamo potuto notare, sono rappresentate dagli stessi dati. Procediamo, quindi, definendo tutti gli attributi per le entità e definire l'associazione "Contiene" che lega l'entità "Podcast" all'entità "Episodi". In particolare, poiché ogni podcast possiede da uno ad n episodi, la medesima entità partecipa all'associazione con cardinalità $(1, n)$. D'altra parte, ogni episodio si riferisce ad uno ed un solo podcast, motivo per cui partecipa con cardinalità $(1, 1)$. Eseguiamo lo stesso con l'entità "Audiolibro", per il quale creiamo un'associazione "Formato da" ed un'entità "Capitolo".



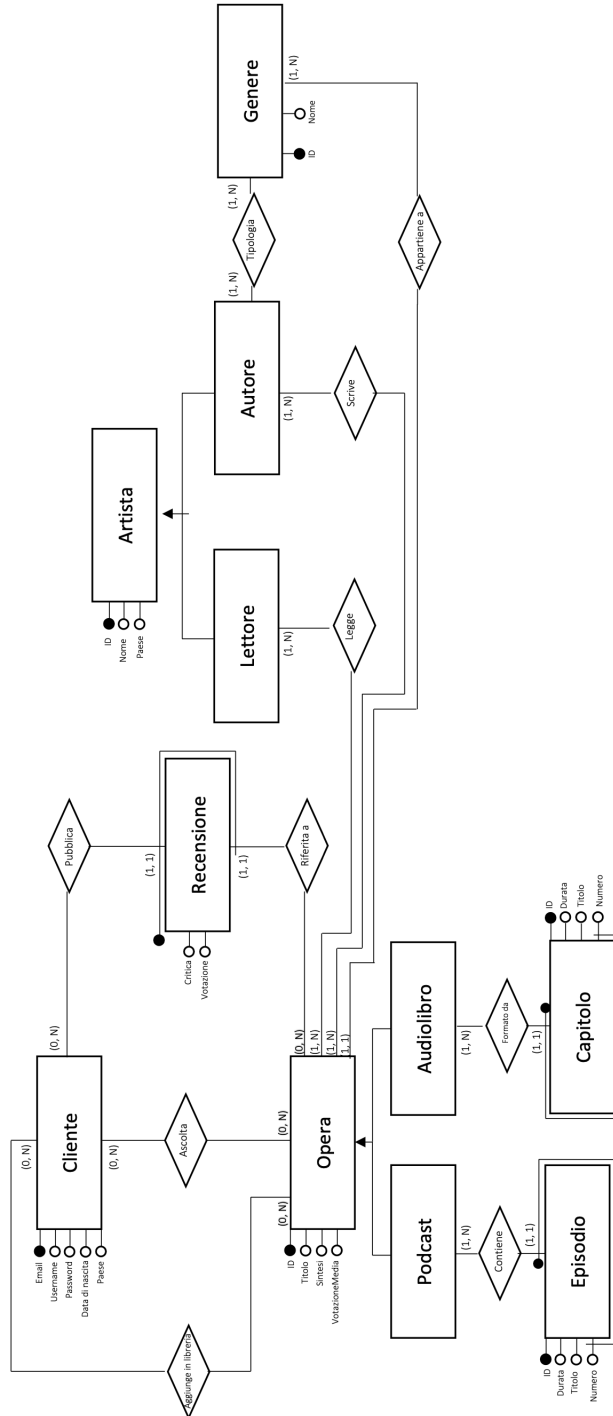
2. Un secondo raffinamento è relativo alle entità "Cliente" e "Recensione", le quali a loro volta possiedono delle associazioni in cui partecipa l'entità "Opera" (quest'ultima protagonista del raffinamento precedente). Procediamo descrivendo tutti gli attributi per le due entità e definendo le associazioni (e conseguentemente le cardinalità): "Aggiunge in libreria", "Ascolta", "Pubblica" e "Riferita a".



- Un terzo raffinamento è relativo alle entità (specifiche) "Autore" e "Lettore", raggruppate in un'entità (generale) "Artista". Queste due entità figlie sono rappresentate dagli stessi dati. Procediamo definendo gli attributi per le entità e le associazioni "Legge", "Scrive" e "Tipologia", ognuna con le rispettive cardinalità.



2.6.3 Schema finale



2.7 Vincoli non esprimibili nel diagramma e dati derivabili

2.7.1 Vincoli non esprimibili nel diagramma

Rispetto il diagramma E-R risultante, notiamo in questo caso esistere un unico vincolo legato al *voto* massimo per una recensione. Il voto, infatti, è un valore numerico compreso tra 1 e 5.

2.7.2 Dati derivabili nel diagramma

L'unico dato derivabile all'interno del diagramma è l'attributo "Votazione-Media" presente nell'entità "*Opera*" è derivabile a partire dal conteggio delle recensioni ed il successivo calcolo della media per tutte le recensioni riguardanti la specifica opera.

2.8 Dizionario dei dati

Entità	Descrizione	Attributi	Identificatore
Cliente	Utente che usufruisce del servizio, ascoltando audiolibri/podcast e scrivendo recensioni su di essi.	Username, email, password, data di nascita, paese	Email
Opera	Insieme delle opere inserite all'interno della piattaforma, quindi audiolibri o podcast.	ID, Titolo, Sintesi, Votazione-Media	ID
Podcast	Raccolta di episodi in formato audio, incentrati su uno o più temi.	ID, Titolo, Sintesi, Votazione-Media	ID
Audiolibro	Libro trasformato nella sua versione audio, possibile attraverso il lavoro di un lettore.	ID, Nome, Sintesi, Votazione-Media	ID
Episodio	Una parte di un podcast.	ID, Titolo, Numero, Durata	ID, Numero, Opera.ID
Capitolo	Una parte di un audiolibro.	ID, Titolo, Numero, Durata	ID, Numero, Opera.ID
Recensione	Critica di un cliente rispetto una specifica opera.	Critica, Valutazione	Opera.ID, Cliente.Email
Artista	Un utente autore o lettore in grado di fornire nuovi contenuti ai clienti.	ID, Nome, Paese	ID
Lettore	Utente che fornisce nuovi contenuti ai clienti effettuando la lettura di uno o più audiolibri o podcast.	ID, Nome, Paese	ID
Autore	Utente autore di uno o più contenuti all'interno della piattaforma.	ID, Nome, Paese	ID
Genere	Genere letterario di un opera o di un autore.	ID, Nome	ID

2.9 Dizionario delle relazioni

Associazione	Entità partecipanti	Descrizione	Attributi
Aggiunge in libreria	Cliente, Opera	Un cliente può aggiungere un opera all'interno della sua libreria.	
Ascolta	Cliente, Opera	Un cliente può ascoltare un opera.	
Pubblica	Cliente, Recensione	Un cliente può pubblicare una o più recensioni per un opera, ma ogni recensione si riferisce ad uno ed un solo cliente.	
Riferita a	Recensione, Opera	Ogni recensione riguarda una ed una sola opera. D'altra parte, un opera può apparire nelle recensioni da zero ad n volte.	
Contiene	Podcast, Episodio	Un podcast contiene uno o più episodi ma un episodio si riferisce ad uno ed un solo podcast.	
Formato da	Audiolibro, Capitolo	Un audiolibro contiene uno o più capitoli ma un capitolo si riferisce ad uno ed un solo audiolibro.	
Scrive	Autore, Opera	Un autore scrive una o più opere, poi caricate all'interno della piattaforma. Poiché vi possono essere più autori per la stessa opera, l'opera stessa può apparire una o più volte.	
Legge	Lettore, Opera	Un autore legge una o più opere, poi caricate all'interno della piattaforma. Poiché vi possono essere più lettori per la stessa opera, l'opera stessa può apparire una o più volte.	
Tipologia	Autore, Genere	Un autore scrive delle opere che si riferiscono ad uno o più generi letterari.	
Appartiene a	Opera, Genere	Ogni opera appartiene ad al più un genere letterario.	

3 Progettazione logica

3.1 Stime

Iniziamo la progettazione logica per la nostra base di dati effettuando tutte le opportune stime dei dati che non possono esser ricavate a partire dalla documentazione iniziale. Effettuiamo, innanzitutto, la stima delle opere ascoltate da ogni utente; consideriamo, infatti, che ogni utente ascolti in media 15 opere al mese. Inoltre, possiamo stimare che di tutte le opere ascoltate, circa il 90% di esse venga aggiunto in libreria. Infine, stimiamo il numero di recensioni con un valore uguale a tutte le opere presenti all'interno della propria libreria.

3.2 Tabella dei volumi

Concetto	Tipo	Volume
Cliente	E	1500
Opera	E	2000
Podcast	E	750
Audiolibro	E	1250
Episodio	E	7500
Capitolo	E	37500
Recensione	E	20000
Artista	E	700
Lettore	E	200
Autore	E	500
Genere	E	10
Aggiunge in libreria	R	20000
Ascolta	R	22500
Pubblica	R	20000
Riferita a	R	20000
Contiene	R	7500
Formato da	R	37500
Scrive	R	2500
Legge	R	2400
Tipologia	R	1000
Appartiene a	R	20000

3.3 Tabella delle frequenze

Indice	Tipo	Frequenza
Op.1	I	15 / Giorno
Op.2	I	20 / Giorno
Op.3	I	5 / Giorno
Op.4	I	10 / Mese
Op.5	I	10 / Giorno
Op.6	I	20 / Giorno
Op.7	I	15 / Giorno
Op.8	I	20 / Mese

3.4 Analisi delle ridondanze

Procediamo adesso con l'analisi delle ridondanze, effettuate secondo la seguente equivalenza: $1 S = 2 L$. Nello specifico verrà analizzata le sola operazione 7, l'unica a far uso di un dato ridondante, ossia *VotazioneMedia*.

In presenza di ridondanza.

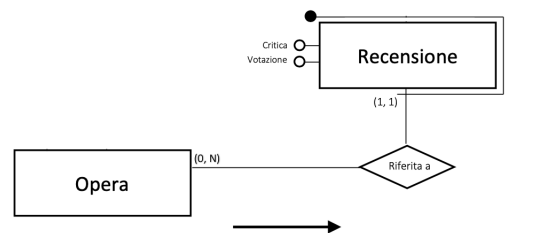
Concetto	Costrutto	Accesso	Tipo
Audiolibro	E	1	L

$1 L * 15 = 15$ accessi al giorno.

In assenza di ridondanza.

Concetto	Costrutto	Accesso	Tipo
Audiolibro	E	1	L
Riferita a	R	1	L
Recensione	E	1	L

$3 L * 15 = 45$ accessi al giorno.



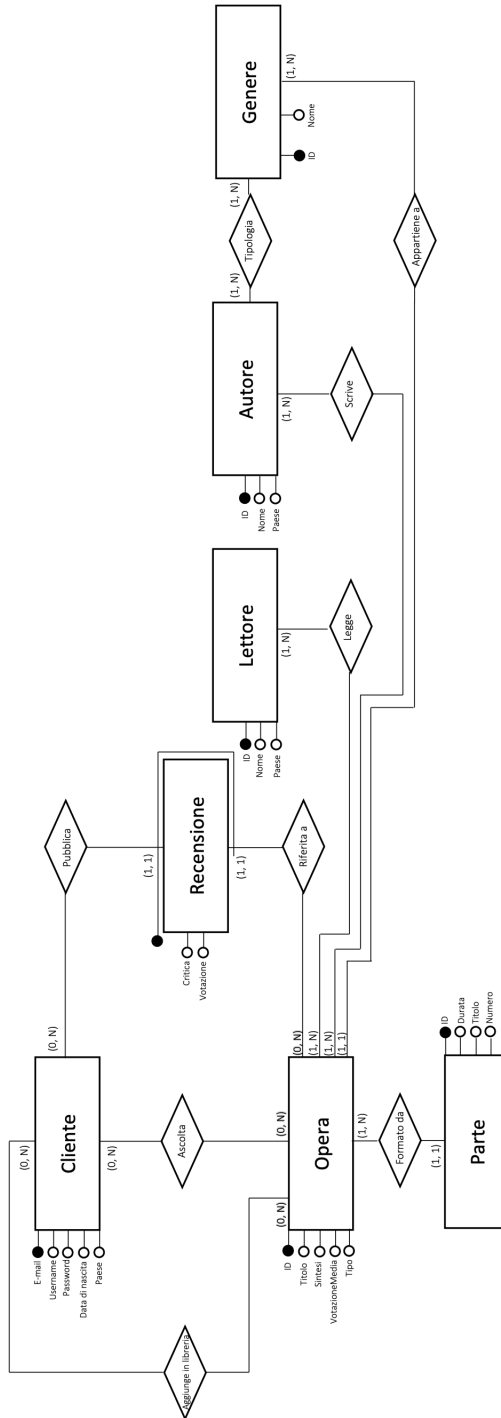
3.5 Eliminazione delle gerarchie ed accorpamento di entità

Sappiamo che i DBMS non permettono la rappresentazione di una generalizzazione ed è per questo motivo che all'interno di questo paragrafo provvederemo a rimuoverle secondo certe tecniche. La prima generalizzazione che attenzioniamo è quella relativa alle entità: *Opera*, *Audiolibro* e *Podcast*. In questo caso, decidiamo di operare secondo un *accorpamento delle entità figlie della generalizzazione dell'entità padre*, motivo per cui aggiungiamo un attributo *Tipo*, necessario a distinguere il tipo di opera all'interno della base di dati. Eliminata la prima gerarchia, però, non possiamo non notare le due associazioni ed entità relative ai due tipi di opera, ossia *Episodio* e *Capitolo*. Decidiamo a questo proposito di accorpare i due all'interno di una sola entità "*Parte*", che unisce insieme il concetto di episodio e capitolo (essendo due concetti interscambiabili) e partecipa, insieme all'entità *Opera*, all'associazione *Formato da*. La seconda ed ultima generalizzazione da attenzionare è relativa alle entità *Artista*, *Autore* e *Lettore*. Decidiamo di eliminare la generalizzazione secondo un accorpamento dell'entità genitore della generalizzazione nelle entità figlie, motivo per cui eliminiamo l'entità *Artista*. Poiché quest'ultima non partecipava a nessun associazione, ci limitiamo solamente a trasferire i suoi attributi all'interno delle entità figlie.

3.6 Scelta degli identificatori principali

Alcune entità all'interno dello schema, presentano identificatori misti formati a partire da attributi presenti all'interno di altre entità. Per semplicità, decidiamo quando possibile di rimuovere quest'ultimi lasciando un semplice identificativo *ID*.

3.7 Schema E-R ristrutturato



3.8 Traduzione di entità ed associazioni

Iniziamo adesso la traduzione di entità ed associazioni per arrivare alla formazione dello schema logico e quindi alla conclusione della progettazione logica. Procediamo, innanzitutto, con il tradurre ogni entità come una relazione con lo stesso nome avente per attributi i medesimi attributi dell'entità e per chiave il suo identificatore. Procediamo, quindi, con la traduzione delle associazioni.

- **Associazioni molti a molti**

- L'associazione *Aggiunge in libreria* viene trasformata in una relazione *OpereInLibreria*, la quale possiede come attributi gli identificatori delle entità coinvolte, ovvero: *Email* ed *ID*; in particolare, quest'ultima, verrà ridenominata con il nome *IDOpera*;
- L'associazione *Ascolta* viene trasformata in una relazione *OpereAscoltate*, la quale possiede come attributi gli identificatori delle entità coinvolte, ovvero: *Email* ed *ID*; in particolare, quest'ultima, verrà ridenominata con il nome *IDOpera*;
- L'associazione *Legge* viene trasformata in una relazione *TitoliLettore*, la quale possiede come attributi gli identificatori delle entità coinvolte, ovvero: *ID*(Lettore), e *ID*(Opera), le quali vengono ridenominate in *IDLettore* e *IDOpera*;
- L'associazione *Scrive* viene trasformata in una relazione *TitoliAutore*, la quale possiede come attributi gli identificatori delle entità coinvolte, ovvero: *ID*(Lettore), e *ID*(Opera), le quali vengono ridenominate in *IDLettore* e *IDOpera*;
- L'associazione *Tipologia* viene trasformata in una relazione *GenereAutore*, la quale possiede come attributi gli identificatori delle entità coinvolte, ossia: *ID*(Lettore), e *ID*(Genere), le quali vengono ridenominate in *IDAutore* e *IDGenere*;

- **Associazioni uno a molti**

- L'associazione *Pubblica*, nel caso ipotetico in cui fosse tradotta attraverso la creazione di una relazione, avrebbe esclusivamente una chiave "ID"(Recensione) (poiché ogni recensione fa riferimento ad uno ed un solo utente). Notiamo, quindi, che l'ipotetica relazione e la relazione *Recensione* sono in possesso dello stesso identificatore, motivo per cui scegliamo di inglobare l'associazione all'interno della relazione *Recensione*, aggiungendo l'attributo "Email".

- L’associazione *Riferita a*, segue lo stesso ragionamento precedente, motivo per cui scegliamo di inglobare l’associazione all’interno della relazione *Recensione*, secondo l’introduzione di un nuovo attributo ” *IDOpera*”.
- L’associazione *Formato da*, segue ancora una volta il ragionamento precedentemente discusso, per questo motivo scegliamo di includere l’associazione all’interno della relazione *Parte*, secondo un nuovo attributo *IDOpera*.
- L’associazione *Appartiene a*, infine, viene anch’essa inclusa nella relazione *Opera*, secondo un nuovo attributo *IDGenere*.

3.9 Schema logico

Cliente (Email, Username, Password, DataDiNascita, Paese)

OpereAscoltate (Email, IDOpera)

OpereInLibreria (Email, IDOpera)

TitoliLettore (IDLettore, IDOpera)

TitoliAutore (IDAutore, IDOpera)

GenereAutore (IDAutore, IDGenere)

PartiOpera (ID, IDOpera, Numero, Titolo, Durata)

Opera (ID, Titolo, Tipo, Sintesi, VotazioneMedia, IDGenere)

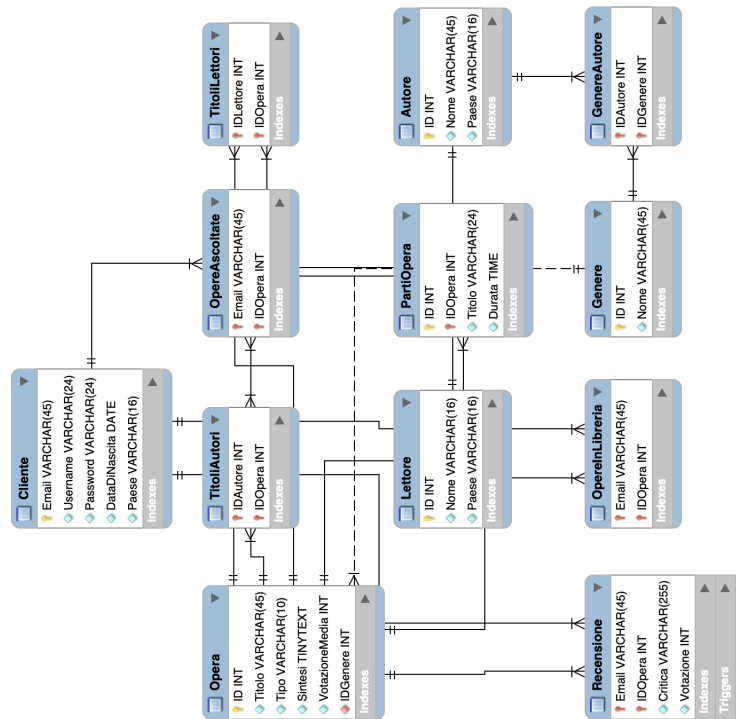
Recensione (Email, IDOpera, Critica, Votazione)

Autore (ID, Nome, Paese)

Lettore (ID, Nome, Paese)

Genere (ID, Nome)

3.10 Schema UML



4 Progettazione fisica

4.1 Traduzione in SQL

In conclusione, procediamo con la traduzione in linguaggio SQL di:

- *Tabelle*, definite secondo il precedente schema logico e presenti nel file `tables.sql`;
- *Dati*, necessari al testing e presenti nel file `data.sql`.
- *Trigger*, presenti nel file `trigger.sql`;
- *Operazioni*, definite secondo la tabella presente nella sotto-sezione 2.2 e presenti nel file `op.sql`;