

1. **System development process**

Based on the description of class objects, an ER diagram for online book review is drawn. ER diagram needs to capture all the relationships and attributes which have been described in the description of class objects. The database schema will be mapped according to the ED diagram. Following the real-world rule, functional dependencies get established and are minimised into a set of minimal functional dependencies for this database schema. From the set of minimal function dependencies, the database schema is checked under the condition of 3NF or BCNF. If any relational schema does not satisfy those conditions, it will be decomposed further into 3NF or BCNF.

2. **The difference between the initial design and the current database schema**

The previous database schema had 6 tables which were Authors, Publishers, Books, Genres, Users and Reviews. It was only able to keep the record of an author writing a book but not multiple books. Genres relation kept the record of book's type with the primary key was genre name and book id. In the new database, there are 10 tables which are Authors, Publishers, Books, Write, Genres, Type, Users and Reviews. Write is the relationship between Books and Authors which allows the database to keep multiple records about an author writing multiple books and a book being written by multiple authors. Type which is a relationship between Books and Genres, keeps the information of a book with many genres more consistent. The new database schema also has two new classes for professional and amateur users.

3. **Relational Database schema**

- PUBLISHERS(Publisher_ID, Publisher_name)
- BOOKS(Book_ID, Title, Year_published, Publisher_ID*)
- AUTHORS(Author_ID, Author_name)
- WRITE(Book_ID*, Author_ID*)
- GENRES(Genres_name, Description)
- TYPE(Book_ID *, Genres_name*)
- USERS(User_ID)
- REVIEWS(Book_ID *, User_ID*, rate)
- PROFESSIONAL(User_ID*, Name, DOB, Sex, Email)
- AMATEUR(User_ID*, Name, DOB, Sex, Email)

All the above relations are in BCNF.

4. **SQL DDL for database tables**

- CREATE TABLE Authors (Author_ID INTEGER (3) PRIMARY KEY NOT NULL, Author_name VARCHAR (30));
- CREATE TABLE Write(Book_ID CHAR(5) NOT NULL , Author_ID INTEGER (3) NOT NULL, PRIMARY KEY (Author_ID, Book_ID), FOREIGN KEY(Author_ID) REFERENCES Authors (Author_ID) ON UPDATE CASCADE ON DELETE CASCADE, FOREIGN KEY(Book_ID) REFERENCES Books (Book_ID) ON UPDATE CASCADE ON DELETE CASCADE);

- CREATE TABLE Books (Book_ID CHAR (5) PRIMARY KEY NOT NULL, Title TEXT NOT NULL, Publisher_ID INTEGER (3) NOT NULL, year_published YEAR, FOREIGN KEY (Publisher_ID) REFERENCES Publishers (Publisher_ID) ON UPDATE CASCADE ON DELETE CASCADE);
- CREATE TABLE Publishers (Publisher_ID INTEGER (3) PRIMARY KEY NOT NULL, Publisher_name VARCHAR (30));
- CREATE TABLE Genres (Genres_name VARCHAR (20) PRIMARY KEY NOT NULL, Description VARCHAR (64));
- CREATE TABLE Type (Genres_name VARCHAR (20) NOT NULL, Book_ID CHAR (5) NOT NULL, PRIMARY KEY (Genres_name, Book_ID), FOREIGN KEY (Genres_name) REFERENCES Genres (Genres_name) ON UPDATE CASCADE ON DELETE CASCADE, FOREIGN KEY (Book_ID) REFERENCES Books (Book_ID) ON UPDATE CASCADE ON DELETE CASCADE);
- CREATE TABLE Users (User_ID INTEGER (5) PRIMARY KEY NOT NULL);
- CREATE TABLE Reviews (User_ID INTEGER (5) NOT NULL, Book_ID CHAR (5) NOT NULL, Rate INTEGER (1), PRIMARY KEY (User_ID, Book_ID), FOREIGN KEY (User_ID) REFERENCES Users (User_ID) ON UPDATE CASCADE ON DELETE CASCADE, FOREIGN KEY (Book_ID) REFERENCES Books (Book_ID) ON UPDATE CASCADE ON DELETE CASCADE);
- CREATE TABLE Professional (User_ID INTEGER (5) PRIMARY KEY NOT NULL, Name VARCHAR (30), Sex CHAR (1), DOB DATE, Email TEXT (40) UNIQUE, FOREIGN KEY (User_ID) REFERENCES Users (User_ID) ON UPDATE CASCADE ON DELETE CASCADE);
- CREATE TABLE Amateur (User_ID INTEGER (5) PRIMARY KEY NOT NULL, Name VARCHAR (30), Sex CHAR (1), DOB DATE, Email TEXT (40) UNIQUE, FOREIGN KEY (User_ID) REFERENCES Users (User_ID) ON UPDATE CASCADE ON DELETE CASCADE);

5. Data analysis task

- Information of authors who have written the highest numbers of books (Figure 11)
SELECT author_name, Author_ID, count(Book_ID) total
FROM Write NATURAL JOIN Authors GROUP BY author_ID
HAVING total = (select max(total)
from (SELECT count(Book_ID) total FROM write group by Author_ID));
- The books of authors who have written the highest number of books (Figure 12)
SELECT Book_ID, Title, Publisher_ID, year_published
FROM Write
NATURAL JOIN Books
WHERE author_ID in (SELECT author_ID FROM Write
NATURAL JOIN Authors
GROUP BY author_ID
HAVING count(Book_ID) = (select max(total)

```

from (SELECT count( Book_ID) total
      FROM write group by
Author_ID));

```

- authors who have never co-authored with anyone (Figure 13)

```

SELECT distinct author_ID FROM Write
      WHERE author_ID not in (SELECT w1.author_ID
                              FROM write w1 JOIN write w2
                              WHERE w1.book_ID =
w2.book_ID
                              AND w1.author_ID <>
w2.author_ID);

```

- The highest rated books were written by a single author. Where the author has never co-authored with anyone (Figure 14)

```

SELECT distinct author_ID, book_ID, avg(rate), title, Publisher_ID, year_published
FROM Write
NATURAL JOIN Reviews NATURAL JOIN Books
WHERE author_ID not in (SELECT w1.author_ID
                        FROM write w1 JOIN write w2
                        WHERE w1.book_ID = w2.book_ID AND w1.author_ID <>
w2.author_ID)
GROUP BY author_Id, book_ID
HAVING avg(rate) = ( select max(average)
                    from (SELECT avg(rate) average
                          FROM Write Natural JOIN Reviews
                          WHERE author_ID not in (SELECT w1.author_ID FROM write
w1 JOIN write w2
                          WHERE w1.book_ID =
w2.book_ID
                          AND w1.author_ID <>
w2.author_ID)
                    GROUP BY author_Id, book_ID));

```

- Books with highest average rate where the genre is not children and romance; and description has no word 'emotional' (Figure 15)

```

SELECT book_ID, Title, avg(rate)
FROM reviews
NATURAL JOIN Type NATURAL JOIN genres NATURAL JOIN Books
WHERE lower(genres_name) <> 'romance' AND lower(genres_name) NOT LIKE 'children'
AND (lower(description) NOT LIKE '% emotional %' AND lower(description) NOT LIKE '
emotional %'
AND lower(description) NOT LIKE '% emotional' AND lower(description) <> 'emotional')
GROUP BY book_ID
HAVING avg(rate) = (select max(average)
                    from (SELECT avg(rate) average
                          FROM reviews NATURAL JOIN Type NATURAL JOIN genres
                          WHERE lower(genres_name) <> 'romance'

```

```

AND lower(genres_name) NOT LIKE '%a%'
AND (lower(description) NOT LIKE '% sad %' AND
lower(description) NOT LIKE 'sad %'
AND lower(description) NOT LIKE '% sad' AND
lower(description) <> 'sad')
GROUP BY book_ID));

```

- Books have the highest average rate by professional (Figure 16)

```

SELECT B.Book_ID , title, avg(rate)
FROM Reviews R
JOIN Professional P JOIN Books B
WHERE R.User_ID = p.User_ID AND B.Book_ID = R.Book_ID
group by B.book_ID
having avg(rate) = ( select max (average)
                    from (SELECT avg(rate) average
                          FROM Reviews R JOIN Professional P JOIN Books B
                          WHERE R.User_ID = p.User_ID AND B.Book_ID = R.Book_ID
                          group by B.book_ID));

```

- Books have not been rated (Figure 17)

```

select book_ID from books
except
select book_ID from reviews;

```

- The most popular genre by number of reviews (Figure 18)

```

select genres_name, count(rate)
from type natural join reviews
group by genres_name
having count(rate) = (select max(count)
                    from (select count(rate) count
                          from type natural join reviews
                          group by genres_name));

```

- Author and publisher names of the most popular books by number of reviews (Figure 19)

```

select w.book_ID, title, author_name,publisher_name
from write w
join books b join authors a join publishers p
where w.book_Id = b.book_ID and w.author_ID = a.author_ID and p.publisher_ID =
b.publisher_ID
and w.book_ID in (select b.book_Id
                  from reviews r join books b
                  where r.book_ID = b.book_ID
                  group by b.book_ID
                  having count(rate) = (select max(count)
                                      from (select book_Id, count(rate) count from
reviews group by book_ID)));

```

- Title of the most popular books by number of reviews and its the average rating (Figure 20)

```

select book_ID, title, avg(rate)
from reviews
natural join books
where book_ID in (select book_Id
                  from reviews
                  group by book_ID
                  having count(rate) = (select max(count) from (select book_Id, count(rate)
                                                                    count
                                                                    from reviews
                                                                    group by book_ID)))
Group by book_ID;

```

6. Discussion

6.1. Limitation

This database schema does not record the history of a user rating on a book many times.

6.2 Further enhancement

Reviews should also be a weak entity that have one to many relationships to Books and Users, where user_id, book_id and rate together form a primary. That would allow users to rate on books many times.

Appendix:

	Author_ID	Author_name
1	101	Mark Morford
2	102	Richard Wright
3	103	Gina Kolata
4	104	Amy Tan
5	105	Scott Turow
6	106	David Cordingly
7	107	David Richards
8	108	Sheila Heti
9	109	Loren Estleman
10	110	Robert Hendrickson
11	111	Julia Oliver
12	112	John Grisham
13	113	Toni Morrison

	Publisher_ID	Publisher_name
1	201	Oxford University Press
2	202	Harper Flamingo Canada
3	203	Farrar Straus Giroux
4	204	Pubnam Pub Group
5	205	Audioworks
6	206	Random House
7	207	Emblem Editions
8	208	House of Anansi Press
9	209	Brilliance Audio- Trade
10	210	Kensington Publishing Crop
11	211	River City Pub
12	212	Dell
13	213	Plume

	Genres_name	Book_ID
1	Fiction	01951
2	Romance	00250
3	Politics	03065
4	Novel	04208
5	Narrative	03982
6	Horror	06714
7	Thriller	04208
8	Children	07070
9	History	08840
10	Poetry	15681
11	Fantasy	15737
12	Comedy	18889
13	Literature	04443

	Book ID	Author ID
1	01951	101
2	00250	102
3	03065	103
4	03982	104
5	06714	105
6	04208	106
7	07070	107
8	08840	108
9	15681	109
10	15737	110
11	18889	111
12	04443	112
13	04564	113

	User_ID
1	54014
2	54015
3	54018
4	54331
5	54245
6	54210
7	54002
8	54019
9	54022
10	54027
11	54011
12	54024
13	54025

Figure 1: Authors
Figure 4: Users

Figure 2: Publisher
Figure 5: Users

Figure 3: Reviews

	Book_ID	Title	Publisher_ID	year_published
1	01951	Classical Mysthology	201	2002
2	00250	Clara Callan	202	2001
3	03065	Don Quixote	203	1999
4	03982	The Kitchen Gods Wife	204	1991
5	06714	Pleading Guilty	205	1993
6	04208	Under the Black Flag: The Romance and The Reality Of Life	206	1996
7	07070	Night Below Station Street	204	1988
8	08840	The Middle Stories	208	2004
9	15681	The witchfinder	209	1998
10	15737	A Social History Of Rats And Man	210	1999
11	18889	Goodbye To The Buttermilk Sky	211	1994
12	04443	The Testament	212	1999
13	04564	Beloved	213	1994

	User_ID	Book_ID	Rate
1	54014	01951	3
2	54014	03065	5
3	54014	07070	4
4	54014	03355	3
5	54014	01222	5
6	54015	01951	4
7	54018	07070	4
8	54019	07070	3
9	54022	03065	4
10	54022	07070	2
11	54022	08840	5
12	54022	15681	5
13	54022	01478	3

	Genres_name	Book_ID
1	Fiction	01951
2	Romance	00250
3	Politics	03065
4	Novel	04208
5	Narrative	03982
6	Horror	06714
7	Thriller	04208
8	Children	07070
9	History	08840
10	Poetry	15681
11	Fantasy	15737
12	Comedy	18889
13	Literature	04443

Figure 5: Books

Figure 6: Reviews

Figure 7: Reviews

	User_ID	Name	Sex	DOB	Email
1	54019	Laura Johnson	F	1988-10-29	ljo123@ymail.com
2	54022	Smith Lachlan	M	1969-04-27	sunday@icloud.net.com
3	54027	Mike Shawn	M	1988-09-07	hallowbooks@ymail.com
4	54011	Peter Roberto	M	2000-08-30	son_d@outlook.com
5	54024	Robert Sten	M	2001-01-19	s.ro123@outlook.com
6	54025	Lola Henry	F	1977-01-27	cookie_net@gmail.com
7	54026	Ringtons Andrew	F	1966-06-14	network123@gmail.com
8	54039	Linda Mathew	F	1997-08-17	m.daas@ymail.com
9	54028	Sony Swenden	M	2003-05-25	heydata@icloud.net.com
10	54021	Lola Henry	F	1950-04-07	dophin@yahoo.com
11	54111	Lawrence Steward	F	1980-09-17	forestharvest@gmail.com
12	54212	Farid Dominic	M	1996-04-17	f.do45@outlook.com

	User_ID	Name	Sex	DOB	Email
1	54014	Linda Anderson	F	1999-11-08	linda.A@gmail.com
2	54015	Mark Martin	M	1987-08-07	m.martin@yahoo.com
3	54018	John Rawd	M	1999-09-17	services@outlook.com
4	54331	Stacey Lawrence	F	2000-09-19	s.la_anf@gmail.com
5	54245	Alice Devon	F	1993-07-29	adevon@gmail.com
6	54210	Harry Song	M	1981-06-28	laywerservices@yahoo.com
7	54002	Willow Hang	M	1992-01-17	folwersgradern@gmail.com
8	54213	Eron Tan	M	1974-12-28	chemisttan@gmail.com
9	54321	Audrey Fing	F	1978-09-18	bibibap@yahoo.com

Figure 8: Amateur

Figure 9: Professional

	Genres_name	Description
1	Fiction	literature in the form of prose, especially novels, that describes imaginary events and people
2	Romance	a feeling of excitement and mystery associated with love
3	Politics	the activities associated with the governance of a country or area, especially the debate between pa
4	Novel	a fictitious prose narrative of book length, typically representing character and action with some d
5	Narrative	a spoken or written account of connected events; a story
6	Horror	an intense feeling of fear, shock, or disgust
7	Thriller	a very exciting contest or experience
8	Children	written about children
9	History	the study of past events, particularly in human affairs
10	Poetry	iterary work in which the expression of feelings and ideas
11	Fantasy	A fantasy is something you imagine
12	Comedy	a play characterized by its humorous or satirical tone
13	Literature	written works, especially those considered of superior or lasting artistic merit
14	Adventure	an unusual and exciting or daring experience
15	Crime	an action or omission which constitutes an offence and is punishable by law
16	Detective	investigate and solve crimes
17	Spirituality	the quality of being concerned with the human spirit or soul as opposed to material or physical thin
18	Drama	an exciting, emotional, or unexpected event or circumstance

Figure 10: Genres

	author_name	Author_ID	 total
1	Amy Tan	104	7

Figure 11: Information of authors who have written the highest numbers of books

	Book ID	Title	Publisher ID	year published
1	01478	Catcher In The Rye	209	1999
2	03065	Don Quixote	203	1999
3	03982	The Kitchen Gods Wife	204	1991
4	04443	The Testament	212	1999
5	04545	They Both Die At The End	201	2002
6	08840	The Middle Stories	208	2004
7	09874	The Lion	204	2012

Figure 12: The books of authors who have written the highest number of books.

	author ID
1	106
2	110

Figure 13: authors who have never co-authored with anyone

	author ID	book ID	avg(rate)	title	Publisher I	year publi
1	106	24511	3.375	Emotional Female	213	2000

Figure 14: The highest rated books were written by a single author. Where the author has never co-authored with anyone

	book ID	Title	avg(rate)
1	03065	Don Quixote	4.75

Figure 15: Books with highest average rate where the genre is not children and romance; and description has no word 'emotional'

	Book ID	title	avg(rate)
1	01222	Mirror Man	5
2	03065	Don Quixote	5
3	24511	Emotional Female	5

Figure 16: Books have the highest average rate by professional

	book ID
1	00250
2	03544
3	03982
4	04443
5	04564
6	15737
7	25487

Figure 17: Books have not been rated

	genres name	count(rate)
1	Children	14

Figure 17: The most popular genre by number of reviews

	book ID	title	author name	publisher name
1	24511	Emotional Female	David Cordingly	Plume

Figure 19: Author and publisher names of the most popular books by number of reviews

book ID	title	avg(rate)
24511	Emotional Female	3.375

Figure 20: Title of the most popular books by number of reviews and its the average rating