

A Design of Book Rating Database

Final Project Report

Pu Du
Tianyi Zhang
Pengfei Pan
Min Cui

CSC 4402 Spring 2017
Louisiana State University

April 24, 2017

Table of Contents

1. Introduction.....	3
2. E-R Diagram and Domain Analysis.....	3
2.1 E-R Diagram.....	3
2.2 Domain Analysis.....	4
3. Implementation of SQL Queries.....	5
APPDENDIX: tables.sql.....	13

1. Introduction

In this project, we design a database of Books, which contains the relationship between books and users, books and authors, books and publishers, books ratings and users. This database can help us to find out the most popular books, the locations and the authors and the publishers of the popular books, the popular books in a specific age, and others related to popular books.

In this database, book crossing dataset is collected by Cai-Nicolas Ziegler in a 4-week crawl from [bookcrossing.com](https://www.bookcrossing.com/). It contains 1,149,780 ratings of 271,379 books by 278,858 users. This dataset has too much information, so we use the modified datasets, which contains enough information but more clear.

2. E-R Diagram and Domain Analysis

In order to make our database more clearly and useful, we just put the essential information in one table. We create five tables: books, users, Authors, publishers, and book_ratings.

2.1 E-R Diagram

Each table has one primary key for identifying entities and accessing records from other tables. Also, the corresponding attributes are showed in the below E-R diagram. In order to select the result we are interested in, we have to know how all these entities are related. The attributes of these tables are presented as follows:

Books: ISBN, Book_title, Author_ID, Year_of_Publication

Users: User_ID, Location, Age

Book-ratings: User_ID, ISBN, rating

Authors: Author_ID, Name, Phone, Gender, Address

Publishers: Publisher_ID, Name, Phone, Address, Website

where the primary keys are underlined for emphasis.

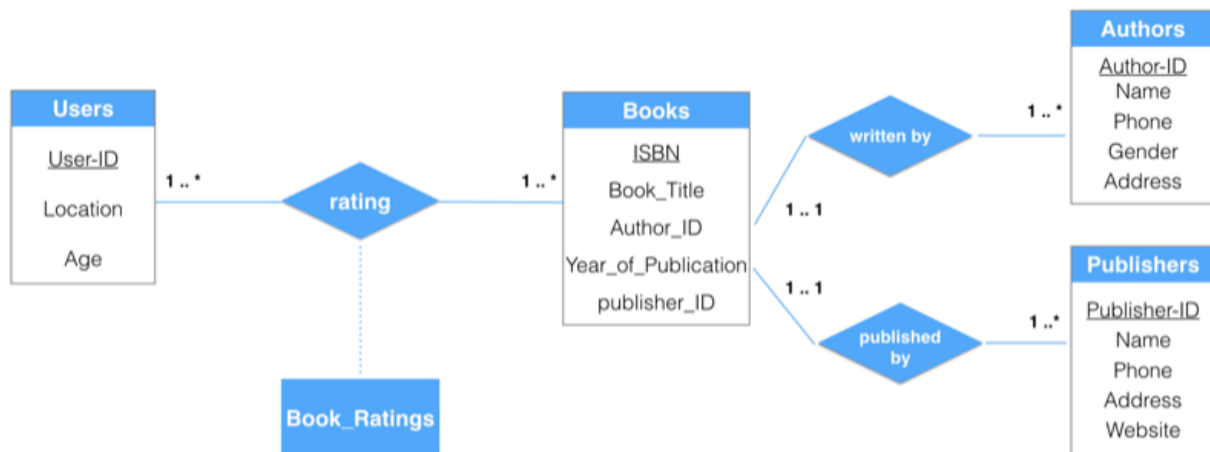


Fig. 1 E-R diagram for the database of book-rating

2.2 Domain Analysis

1. Table Books:

ISBN: the set of all text strings

Book_Title: the set of all text strings

Author_ID: the set of all integer values

Year_Of_Publication: the set of all integer values

Publisher_ID: the set of all integer values

2. Table Users:

User_ID: the set of all integer values

Location: the set of all text strings

Age: the set of all integer values

3. Table Book-rating:

ISBN: the set of all text strings

User_ID: the set of all integer values

Rating: the set of all integer values

Author_ID: the set of all integer values

4. Table Authors:

Name: the set of all text strings

Phone: the set of all integer values

Gender: the set of all text strings

Address: the set of all text strings

5. Table Publishers:

Publisher_ID: the set of all integer values

Name: the set of all text strings

Phone: the set of all integer values

Address: the set of all text strings

There are no other specific constraints about the domain. And we assume that a book only has one first author.

3. Implementation of SQL Queries

(1). "#Total users, total authors, total authors and total books". (This query implement "count" aggregation function.)

```
select count(*)
from users;
select count(*)
from authors;
select count(*)
from publishers;
select count(*)
```

from books;

```
count(*)
278858
count(*)
99198
count(*)
16550
count(*)
271065
```

(2). "#select ISBN with at least 1 full score of review"

```
select distinct ISBN
from book_ratings
where rating=10;
```

```
1853260487
0061002828
8485900057
8485900197
0440164842
8440668058
8440684096
8440699476
8478442960
9505156081
9505156103
0842382372
0786864575
0140328726
087341800X
0937274992
1570761418
1887374426
1740511786
3442441641
377040906X
0525459243
ASSORT00626
0345429230
0440506832
0486234045
0517587874
0553285408
0609800906
```

(3). "#Find the number of books with rating > 4 and users in USA." (This query

implements natural join, string operations, "count" aggregation function.)

```
select count(ISBN)
from users natural join book_ratings
where rating>4 and location like '%usa%' and rating>4;
```

```
count (ISBN)
258930
```

(4). "#Find the average rating of books with the age of user is from 30 to 50." (This query implements natural join, "avg" aggregation function.)

```
select avg(rating)
from users natural join book_ratings
where age>30 and age<50;
```

```
avg(rating)
2.6842
```

(5). "#Find the number of users in each location with book rating>8." (This query implements natural join, "count" aggregation function, group by, order by)

```
select location, count(u.user_id)
from users natural join book_ratings
where rating>8
group by location
order by location;
```

xiamen, fujian, china	1
xiaogan hubei, n/a, china	1
xix?n, asturies, spain	1
xi'an, shannxi, china	1
xxx, california, austria	1
xxxxxxx, xxxxxx, netherlands	1
yabulu, queensland, australia	1
yackandandah, victoria, australia	3
yakima, washington, usa	14
yamato-shi, kanagawa-ken, japan	1
yancey, texas, usa	7
yankalilla, south australia, australia	5
yankton, south dakota, usa	7
yardley, pennsylvania, usa	2
yardville, new jersey, usa	1
yarmouth, nova scotia,	1
yarmouth, nova scotia, canada	85
yarraville, victoria, australia	2
yate, bristol, england, united kingdom	1
yaxley, cambridgeshire, united kingdom	4
yazoo city, mississippi, usa	1
yellowknife, , canada	1
yellowknife, northwest territories, canada	5
yellowstone national park, wyoming, usa	1
yellville, arkansas, usa	1
yelm, washington, usa	2

(6). "#Find the author's name and book_title which is published after 2010." (This query implement natural join, order by)

```
select name, book_title
from authors natural join books
where year_of_publication>2010
order by name;
```

```

name      book_title
Anne Sayre    LOOK HOMEWARD ANGEL
Bohumil Hrabal Three Plays of Eugene Oneill
Charles Rosen  To Have and Have Not
Joe L. Wheeler The Royals
John Ashbery   Monkey (An Evergreen Book, E-112)
Karen Robards  MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)
Karl Evanzz    Crossing America
Larry Niven    Outline of European Architecture (Pelican S.)
Laura Scott    Das gro????e B??Äse- M??Ädchen- Lesebuch.
Lawrence Sanders FOOTBALL SUPER TEAMS : FOOTBALL SUPER TEAMS
Michael Cordy  MY TEACHER FLUNKED THE PLANET (RACK SIZE) (MY TEACHER BOOKS)
MICHAEL S. HARPER FOREST PEOPLE (Touchstone Books (Hardcover))
Nora Roberts   Heidi
RENEE ASKINS   Edgar Allen Poe Collected Poems
Robert C. Tucker In Our Time: Stories (Scribner Classic)
Ron Hansen     Owl's Amazing but True No. 2
Sapphire       Field Guide to the Birds of North America, 3rd Ed.
Wallace Stegner CLOUT
William J. Mann A Rose in Winter
William Marshall Alice's Adventures in Wonderland and Through the Looking Glass (Puffin Books)

```

(7). "#Find the author's name with book rating =10." (This query implements natural join.)

```

select a.name
from authors natural join books natural join book_ratings
where rating=10;

```

```

Jackie Robb
Florinda Donner
Douglas Borton
Carlos E Asay
Jodi Endicott
Jean Campbell
John Creasey
Henry Reed
Miguel De Unamuno
Darlene Graham
Raymond Carver
Susan Howatch
Alfred Hitchcock
Raymond Obstfeld
Lois Lowry
Cathy Hopkins
Lynn Kerstan
Arthur Conan, Sir Doyle
Peg Kehret
Lucy Gillen
Henry James
Michael Meyer
Don Pendleton

```

.....

(8) "Find the book_title, publication's name and rating of book which is published after 2010." (This query implements natural join, order by)

```
select book_title, p.name, r.rating
from books natural join book_ratings natural join publishers
where year_of_publication>2010
order by rating;
```

A Rose in Winter	St Martins Pr	0		
The Royals	New Amer Library (Mm)	0		
A Rose in Winter	St Martins Pr	0		
MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)			Berkley Publishing Group	0
Three Plays of Eugene Oneill	RoC	0		
Outline of European Architecture (Pelican S.)	Pan Macmillan	0		
CLOUT	Golden Books	0		
MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)			Berkley Publishing Group	4
MY TEACHER FLUNKED THE PLANET (RACK SIZE) (MY TEACHER BOOKS)			Apple Signature (Scholastic)	5
MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)			Berkley Publishing Group	5
MY TEACHER FLUNKED THE PLANET (RACK SIZE) (MY TEACHER BOOKS)			Apple Signature (Scholastic)	6
Crossing America	Kensington Publishing Corporation	6		
Das gro????e B??Äse- M??Ädchen- Lesebuch.	Goldmann	6		
MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)			Berkley Publishing Group	6
MY TEACHER FLUNKED THE PLANET (RACK SIZE) (MY TEACHER BOOKS)			Apple Signature (Scholastic)	7
Crossing America	Kensington Publishing Corporation	7		
Alice's Adventures in Wonderland and Through the Looking Glass	(Puffin Books)		Random House	7
Alice's Adventures in Wonderland and Through the Looking Glass	(Puffin Books)		Random House	7
The Royals	New Amer Library (Mm)	7		
Heidi	Bantam Books	8		
Heidi	Bantam Books	8		
Monkey (An Evergreen Book, E-112)	Putnam Pub Group	8		
MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)			Berkley Publishing Group	8
MY TEACHER FLUNKED THE PLANET (RACK SIZE) (MY TEACHER BOOKS)			Apple Signature (Scholastic)	8
Edgar Allen Poe Collected Poems	New Amer Library	8		
Edgar Allen Poe Collected Poems	New Amer Library	8		
Crossing America	Kensington Publishing Corporation	8		
Three Plays of Eugene Oneill	RoC	9		
Das gro????e B??Äse- M??Ädchen- Lesebuch.	Goldmann	9		
Crossing America	Kensington Publishing Corporation	9		
MY TEACHER FRIED MY BRAINS (RACK SIZE) (MY TEACHER BOOKS)			Berkley Publishing Group	9
FOREST PEOPLE (Touchstone Books (Hardcover))	Houghton Mifflin Co (J)	9		
Das gro????e B??Äse- M??Ädchen- Lesebuch.	Goldmann	9		
Three Plays of Eugene Oneill	RoC	10		
To Have and Have Not	Signet Books	10		

.....

(9). "#select the books whose rating is grater than 5 , published by author id = 2,order by rating." (This query implements natural join, descent order.)

```
Select distinct Book_Title ,Rating,Year_Of_Publication
From books natural join authors natural join book_ratings
Where Author_ID =2 and Rating>5
order by Rating desc;
```

Book_Title	Rating	Year Of Publication		
Wilfred's Hospital Ship: Wilfred Grenfell (Stories about Christian Heroes)	10	1979		
La Casa Degli Spiriti	10	1995		
The Saving Graces : A Novel	10	1999		
La Casa Degli Spiriti	9	1995		
Child of the Phoenix	9	1993		
The Saving Graces : A Novel	9	1999		
Footprints: The True Story Behind the Poem That Inspired Millions	9	1993		
The Kill Clause : A Novel	8	2003		
The Saving Graces : A Novel	8	1999		
La Casa Degli Spiriti	8	1995		
Footprints: The True Story Behind the Poem That Inspired Millions	8	1993		
There Will Be Wolves	8	1994		
Footprints: The True Story Behind the Poem That Inspired Millions	7	1993		
La Casa Degli Spiriti	7	1995		
The Saving Graces : A Novel	7	1999		

(10). "# find the books rating and the users information." (This query implements natural join.)

Select Book_Title, Rating, users.User_ID, Location, Age
 From books natural join book_ratings natural join users
 Where books.ISBN = 0001846086;

Book_Title	Rating	User_ID	Location	Age
There Will Be a Next Time	0	122874	melbourne, south australia, australia	25

(11). "# find the user 1838 rated books and rating." (This query implements natural join.)

Select User_ID, Book_Title, rating
 From book_ratings natural join books
 Where User_ID =1838 ;

User_ID	Book_Title	rating
1838	A Secret Affair	9
1838	Dangerous to Know	0
1838	Politika (Tom Clancy's Power Plays (Paperback))	9
1838	Suspicion of Deceit	0
1838	Judging Time (April Woo Suspense Novels (Paperback))	0
1838	You're Fifty--Now What?: Investing for the Second Half of Your Life	8

(12) "# find what age of user is likely to give high ratings." (This query implements natural join, group by, order by, desc)

```
Select Age,avg(Rating)
From book_ratings natural join users
Where book_ratings.User_ID = users.User_ID
Group by Age
order by avg(Rating) desc;
```

```
140      2.6000
30       2.5873
69       2.5434
38       2.5184
44       2.5104
41       2.4759
46       2.4626
45       2.4374
61       2.3940
204      2.3871
33       2.3085
51       2.2766
36       2.2743
103      2.2028
239      2.1933
104      2.1751
56       2.0732
229      2.0000
138      2.0000
67       1.8439
65       1.7936
244      1.6071
```

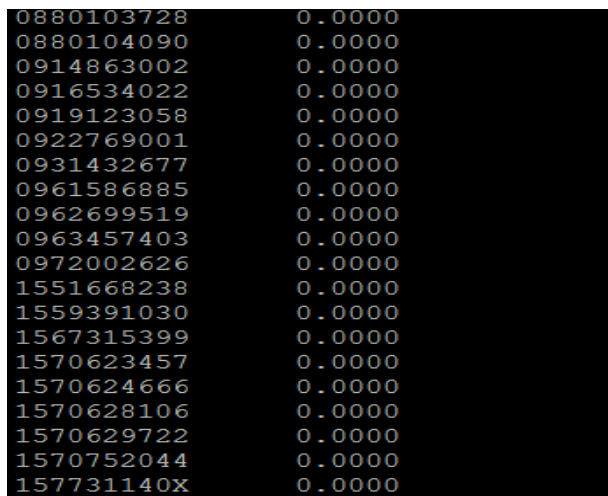
(13) "# Find books published after 2000."

```
select Book_Title
from books
where Year Of Publication > 2000;
```

```
Collected Works of Emile Zola
Murphy's Law
The New Gurus -- From Sun-Tzu And Jesus To Machiavelli And Winnie The Pooh
American Fuji
The Hearing
Madonna
The Villa
More, Now, Again: A Memoir of Addiction
Dirt Music: A Novel
Mystic Rose, The
McNally's Chance
The Community of the Ark: A Visit to the Utopian Communities of Lanza del Vasto and His Disciples of Gandhi
Cane River
99 Cows
A Darkness More Than Night
```

(14) "# Order the average score of rated book". (This query implements group by , "avg" aggregation function.)

```
select ISBN, avg(rating)
from book_ratings
group by ISBN
order by avg(rating) desc;
```



A screenshot of a terminal window displaying the results of a SQL query. The results are presented as a table with two columns: ISBN and average rating. The ISBNs are listed in descending order of their average rating, which is 0.0000 for all entries shown. The ISBNs include 0880103728, 0880104090, 0914863002, 0916534022, 0919123058, 0922769001, 0931432677, 0961586885, 0962699519, 0963457403, 0972002626, 1551668238, 1559391030, 1567315399, 1570623457, 1570624666, 1570628106, 1570629722, 1570752044, and 157731140X.

0880103728	0.0000
0880104090	0.0000
0914863002	0.0000
0916534022	0.0000
0919123058	0.0000
0922769001	0.0000
0931432677	0.0000
0961586885	0.0000
0962699519	0.0000
0963457403	0.0000
0972002626	0.0000
1551668238	0.0000
1559391030	0.0000
1567315399	0.0000
1570623457	0.0000
1570624666	0.0000
1570628106	0.0000
1570629722	0.0000
1570752044	0.0000
157731140X	0.0000

.....

APPDENDIX: tables.sql

```
Create Table books
drop table if exists books;
create table books
(ISBN varchar(13),
Book_Title varchar(255),
Author_ID int(11),
Year_Of_Publication int(10),
Publisher_ID int(11),
primary key (ISBN),
foreign key (Author_ID) references authors(Author_ID),
foreign key (Publisher_ID) references publishers(Publisher_ID)
);
```

```
Create Table users
drop table if exists users;
create table users
(User_ID int(11),
Location varchar(255),
Age int(11),
primary key (User_ID));
```

```
Create table Book_Ratings
drop table if exists book_ratings;
create table book_ratings
(User_ID int(11),
ISBN varchar(13),
Rating int(11),
primary key (User_ID, ISBN),
foreign key (User_ID) references users(User_ID),
foreign key (ISBN) references books(ISBN)
);
```

```
Create table Authors
drop table if exists authors;
create table authors
(Author_ID int(11),
Name varchar(255),
Phone int(10) Default NULL,
```

Gender varchar(1) Default NULL,
Address varchar(255) Default NULL,
primary key (Author_ID));

Create table Publishers
drop table if exists publishers;
create table publishers
(Publisher_ID int(11),
Name varchar(255),
Phone int(10) default NULL,
Address varchar(255) default NULL,
primary key (Publisher_ID));

Github repo of this project: <https://github.com/ipudu/book-rating-database>