INST327-0101 Team 1
Final Project Report
05/14/2021

# Final Project Submission

**Introduction**:

Our database is a library database that will track a collection of books. This database will include attributes of the books. It will be able to answer questions like “how many books are in a certain category”, “what books are written by the same author”, “which books are the longest", "how many books did the author publish in a certain year", etc. The database aims to provide convenience for people when they are trying to find books. People can use the author name, book name, or date published of the book, or some other information as the trail and then use this library database to find a certain book or multiple books. We created this database with the demographics of book lovers in mind, so if you were interested in a book, you can look towards this database to also find similar books in the same genre or published by the same publisher. We created this database with eight main tables that cover a variety of information about the book such as if it has restrictions in the past or the subgenre it can fall into. With all of this in mind, we hope people can gain a lot of insight and knowledge from using our database.

**Database Description**:

To be specific, in our database, there might contain thousands of books; the database will mimic a library collection in a way with a lot of varying information for each entry we add. The database will include information of the book's author, publisher, writing styles and more. It will be able to answer questions regarding what the genre is, when it was made, and how well acclaimed it is (regarding ratings). All of the data in the database should be updated regularly since there might be new books sent into the library. The database can be updated once per season or one time for half of a year. We would like our database to serve like an internet catalog for books, similar to how Worldcat runs on the Mckeldin Library’s website. There are no specific books that we are targeting for our database, but we will be using GoodReads (as stated below in more depth) to help supply it. This dataset consists of books that can interest a large demographic of readers since it includes genres like horror, romance, comedy, informative, etc. so any person interested in books can peruse the database.

**Physical Design:**

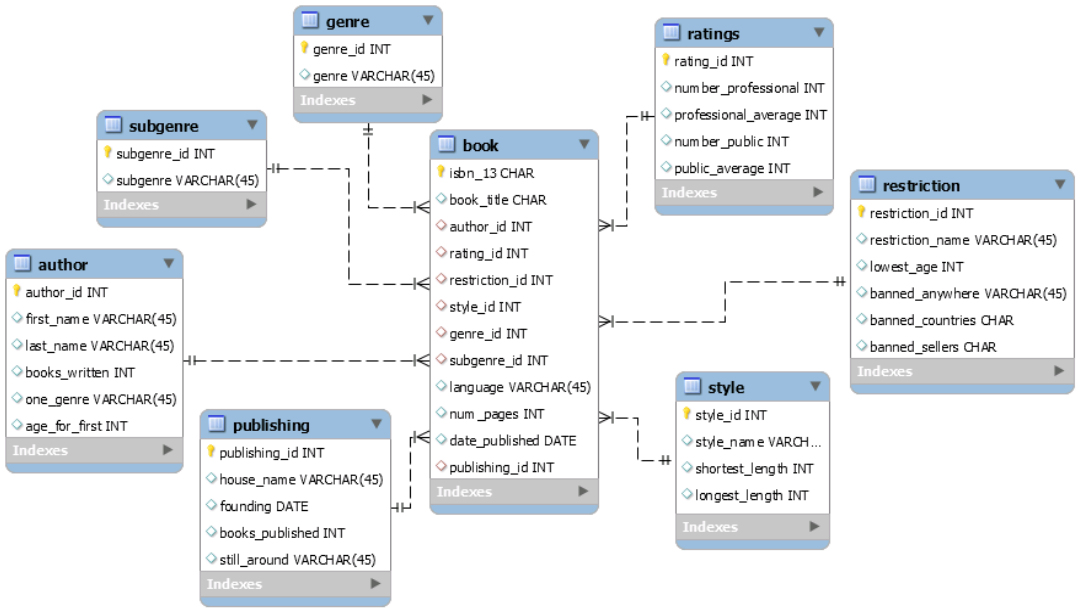


Image 1. ERD of our Database

We took a design approach that allows for readers to easily find any kinds of books that they are in search of with the ability to organize information by things like ratings, author, and genre. We started with the book table and added tables such as genre, style, ratings, and author to help users quickly find pertinent information about their desired topic. Users can utilize and organize the information using these tables for when they are looking to get information or search for reading material based on things like the sub-genre. Overall, we were aiming to make the database simplistic and very easy to use.

**Sample Data:**

The sample we took our data from was Kaggle, via a GoodReadsBook data file. We were able to find a good majority of the columns we used in our database using this, but also looked on Google to help supplement and fill any missing information and to minimize ambiguity. We believe there is reliable consistency within our dataset and the information is fairly accurate as we looked at sources that could be cross-verified. We did not look for any specific type of books, but did choose a significant amount of the data from books that are very recognizable. Below is a sample of our data within the “book” table, which includes links to other tables that provide more information on things like the author name, whether they were restricted or not, and what

publishing house they belong to.

| isbn_13 | book_title | author_id | rating_id | restriction_id | style_id | language | num_pages | date_p |
|---|---|---|---|---|---|---|---|---|
| 9780140275018 | Cod: A Biography of the Fish That Changed the... | 18 | 22 | 1 | 2 | English | 294 | 1998 |
| 9780143036388 | On the Road | 21 | 27 | 1 | 2 | English | 320 | 2005 |
| 9780345410054 | The Power of One | 15 | 18 | 1 | 3 | English | 544 | 1996 |
| 9780345448194 | Boogaloo on 2nd Avenue | 18 | 21 | 1 | 2 | English | 319 | 2006 |
| 9780374518738 | Annals of the Former World | 13 | 15 | 1 | 3 | English | 720 | 2000 |
| 9780374519322 | La Place de la Concorde Suisse | 13 | 16 | 1 | 2 | French | 160 | 1994 |
| 9780374519742 | Heirs of General Practice | 13 | 14 | 1 | 1 | English | 128 | 1986 |
| 9780380710829 | The Grass Crown | 19 | 30 | 1 | 3 | English | 1104 | 1992 |
| 9780439554893 | Harry Potter and the Chamber of Secrets | 4 | 4 | 3 | 2 | English | 352 | 2003 |
| 9780439651400 | Public Enemies | 17 | 20 | 1 | 1 | English | 150 | 2005 |
| 9780439827607 | Harry Potter Collection | 4 | 26 | 3 | 3 | English | 3342 | 2005 |
| 9780441172696 | Dune Messiah | 14 | 17 | 1 | 2 | English | 331 | 1987 |
| 9780486424545 | Howards End | 16 | 19 | 1 | 2 | English | 246 | 2002 |
| 9780553560732 | Red Mars | 3 | 3 | 1 | 3 | English | 572 | 1993 |

Image 2. Screenshot of our book table within our database

**Views/Queries:**

| | A (4) | B (3) | C (2) | D (1) | E (1) |
|---|---|---|---|---|---|
| books_and_genres | | | | | |
| titles_per_author | X | | X | | |
| restrictions | | X | | | X |
| count_Hilton_2001 | X | X | | | |
| count_french_books | | X | X | | |
| books_per_publisher | X | | | | |
| english_written_books | | X | | | |
| **Total** | **4** | **4** | **2** | **1** | **1** |

*Query Descriptions:*

books_and_genres: Uses joins and a linking table to match books to their genres

titles_per_author: Shows how many books written by each author are put in our database

restrictions: Displays all the books that are restricted in some way

count_Hilton_2001: Using a JOIN and a WHERE to find how many books Hilton published in 2001

count_french_books: uses WHERE and GROUP BY to count the number of French books in the database

books_per_publisher: shows the count of books per publisher in the database currently

english_written_books: uses a where clause to find all the books written in English

**Changes from Original Design:**

The biggest changes made were in the tables. These include additional tables, additional columns, and alternate data types in the columns. One example of this is in the publishing table. In the original project proposal, the table only had two columns - name and id. Now it has name, id, year founded, approximate books published, and whether or not it's still around today. A linking table was also added, to facilitate a many-to-many relationship between the books table and the genre table. The sales table was replaced with the restrictions table, since many of the books didn't have sales numbers. Finally, some of the datatypes were changed to allow for more characters (in the case of the isbn_13 key) or were simply not included, like "double" over "decimal" (in the case of average_rating).

**Lessons Learned**

During the semester-long project, our team worked on our database slowly but surely. We can all agree that there were many challenges and learning curves towards handling SQL and creating our own database. Within the project we first had issues with creating an ERD; although creating the project proposal assisted greatly with helping us organize our tables, transferring that information to SQL proved to be difficult for us. Following along with that idea, we also held

troubles with trying to assist our main database creator with inserting data and creating the entirety of the database practically. The other teammates can only help so much with the ERD when it needs to be created on a singular machine then shared with the group. Our modern-day solution to get around any technical issues and assistance with the creation of a database was, of course, Zoom (who would have thought!). By utilizing our extra time in labs on Friday along with weekly meetings, we were able to assist each other and help distribute the workload for each assignment efficiently. Communication is something that the group can pride themselves on as we all hold a great understanding in time commitments and responsibilities within the class and outside of it.

Overall, we learned a great amount about troubleshooting, team work, and creating a database in general through this class. Obstacles were quickly solved by the use of our individual assets and we were able to utilize our lessons learned from the course to come up with a final product we can all be proud of. Issues of downloading the creation script, making sure the database is backed up, or problems with creating views were all quickly solved by just asking questions to our team members. In places our individual selves felt like we lacked knowledge on, other group members were able to fill in and assist for.

**Potential Future Work**

Going forward in regards to this database we believe it could be interesting to continue the development of this database into a mobile application for users to quickly search information on books they would like to read either for free or to buy. However the first thing we would do is extend this database to have significantly more entries into the table to represent a more realistic database size. Making the database larger allows us to do data analysis to search for trends in ratings or a better understanding of relationships between books and their author/publisher. We could also create new tables to incorporate information on sales figures and try to predict future trends based on a specific factor such as genre. We believe that this database could be extremely useful to potential book buyers if regularly updated and could be marketed as a better and easier way to look for new books without muddling through Amazon.