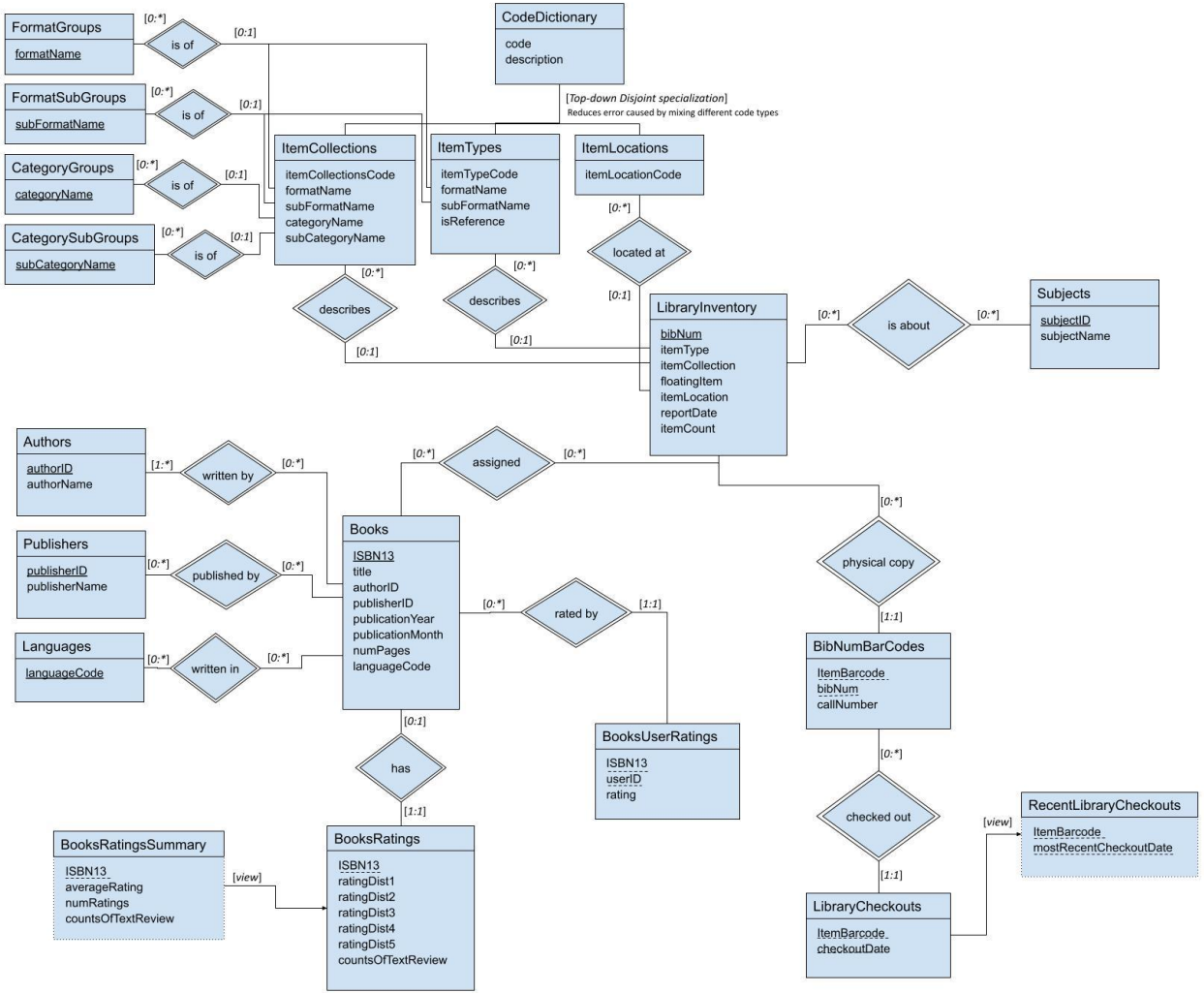
20-Books

ER Schema by sqguo

Client Application by sqguo

Data-mining by sqguo



ER Diagram

Translates to 21 tables and 3 Views.

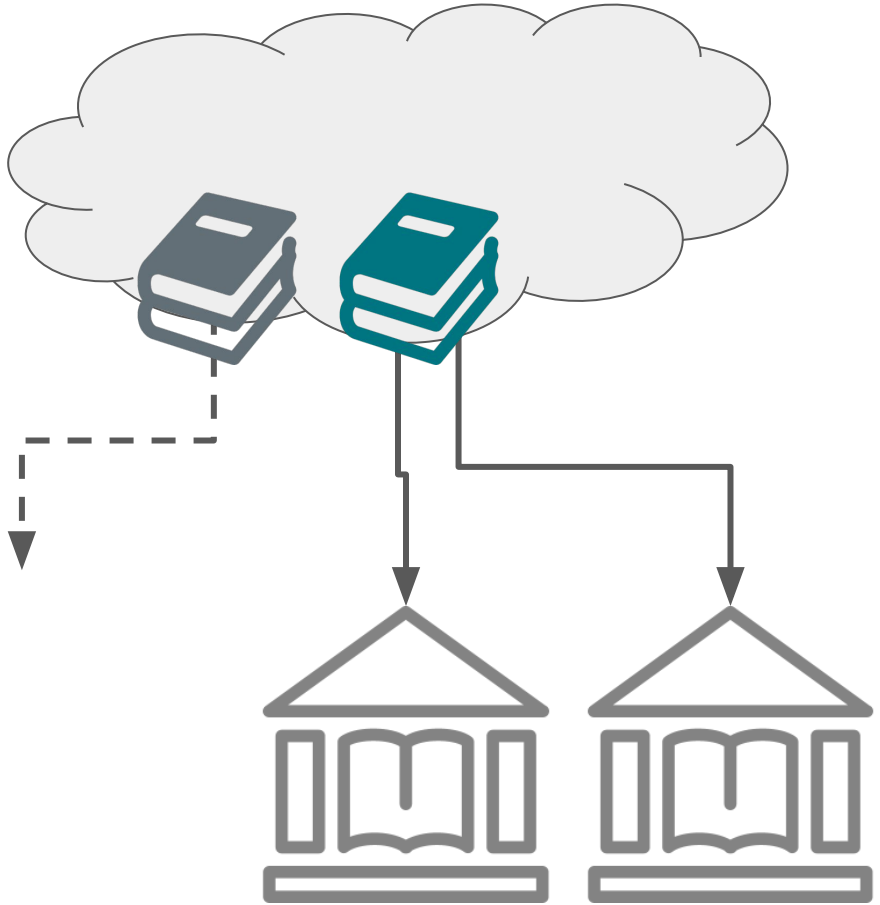
2,173,401 unique ISBNs books.

30,974,900 Records in the LibraryCheckouts (2013 to 2017)

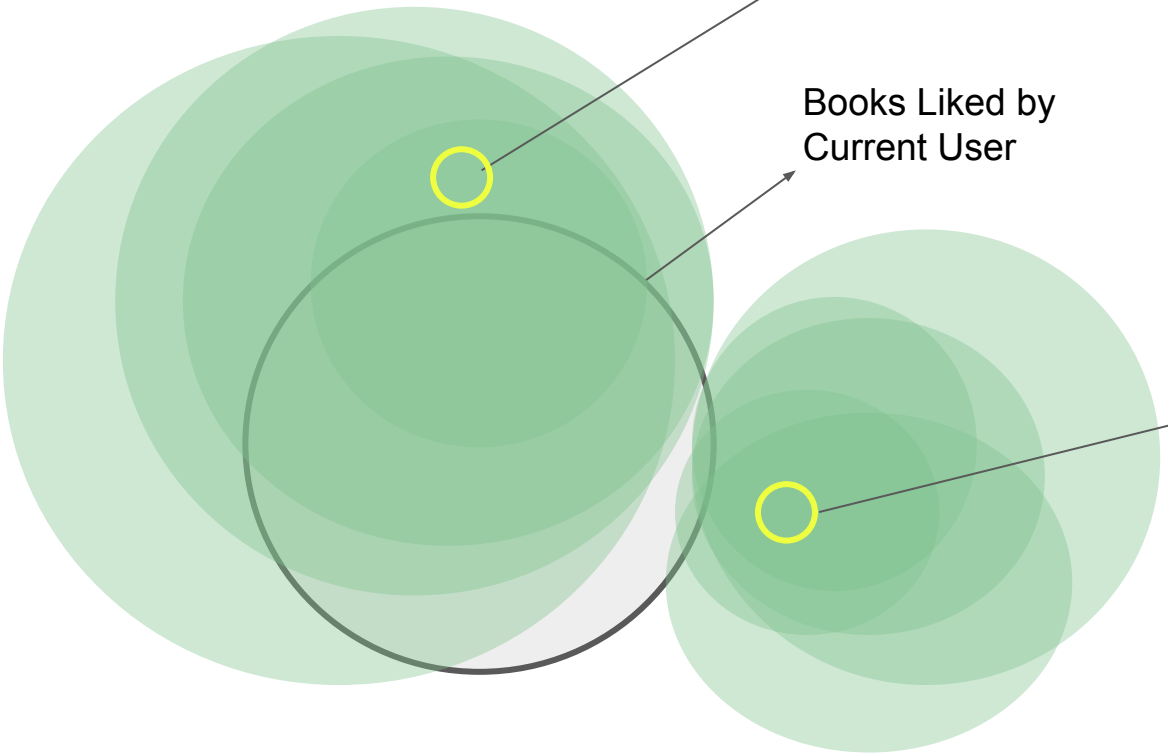
Client Application

Helps Users find the book they need before going to the library

`lookup a book`	`f`
`borrow a book`	`b`
`publish a book`	`p`
`review a book`	`r`
`checkout a book`	`c`
`recommend a book`	`e`
`exit`	`q`



Recommendation Engine




```
WITH mylikedISBNs AS (SELECT ISBN13 FROM BooksUserRatings WHERE
userID = {} AND rating >= 4), mydislikedISBNs AS (SELECT ISBN13 FROM
BooksUserRatings WHERE userID = {} AND rating <= 3), mylikedISBNsDups
AS (SELECT ISBN13 FROM Books WHERE LOWER(title) IN (SELECT LOWER
(title) FROM mylikedISBNs LEFT JOIN Books USING(ISBN13))),
mydislikedISBNsDups AS (SELECT ISBN13 FROM Books WHERE LOWER(title)
IN (SELECT LOWER(title) FROM mydislikedISBNs LEFT JOIN Books USING
(ISBN13))), otherUsersWhoLikedmy AS (SELECT userID, COUNT(*) AS
userSimilarity FROM mylikedISBNsDups LEFT JOIN BooksUserRatings USING
(ISBN13) WHERE userID != {} AND rating >=4 GROUP BY userID),
topOtherUsersWhoLikedmy AS (SELECT*FROM otherUsersWhoLikedmy ORDER BY
userSimilarity DESC LIMIT 20), similarISBNs AS (SELECT ISBN13, (SUM
(userSimilarity)) AS overlapscore FROM topOtherUsersWhoLikedmy LEFT
JOIN BooksUserRatings USING(userID) WHERE rating >=4 and ISBN13 NOT
IN (SELECT*FROM mylikedISBNsDups) and ISBN13 NOT IN (SELECT*FROM
mydislikedISBNsDups) GROUP BY ISBN13), ISBNsToRecommand AS (SELECT
MIN(ISBN13) AS defISBN13, LOWER(title), MAX(overlapscore) as aggScore
FROM similarISBNs LEFT JOIN Books USING(ISBN13) GROUP BY LOWER
(title)), ISBNsUniqueReomand AS (SELECT defISBN13 AS ISBN13, aggScore
FROM ISBNsToRecommand ORDER BY aggScore DESC LIMIT 20) SELECT*FROM
ISBNsUniqueReomand LEFT JOIN Books USING (ISBN13) LEFT JOIN Authors
USING(authorID) LEFT JOIN Publishers USING(publisherID) LEFT JOIN
BooksRatingsSummary USING(ISBN13);
```

A will always be recommended over **B** if there is a choice.

Data Mining

Library Checkouts (2015 - 2017)	Goodreads Ratings			
Popularity	1	2	3	4
Extremely Unpopular	135	898	8221	4001
Very Unpopular	111	892	11776	5883
Unpopular	42	390	7205	3733
Normal	50	362	9046	5270
Popular	21	247	7071	4673
Very Popular	10	115	5971	4586
Extremely Popular	3	17	3898	4337

Data Mining (Validation)

Library Checkouts (2013 - 2014)	Goodreads Ratings			
Popularity	1	2	3	4
Extremely Unpopular	123	886	8131	4177
Very Unpopular	117	872	11193	6026
Unpopular	58	401	6439	3476
Normal	55	385	8186	4621
Popular	22	255	6900	4133
Very Popular	17	179	6490	4415
Extremely Popular	4	50	4503	4138