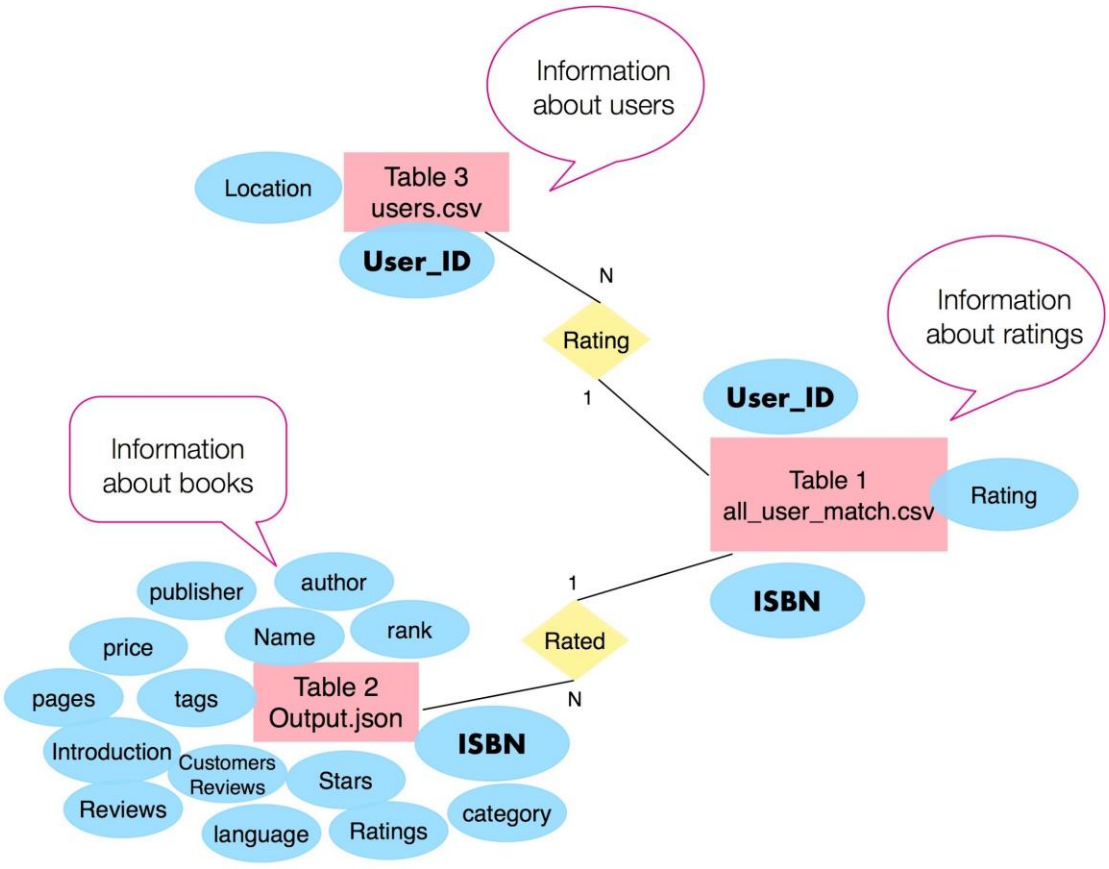
Data Structure and Entity Relationship

Youtian Guo, Zhiyi Guo, Sili Wang, Xintong Zhou, Jiachen Zou



We have one json file (*Output.json*) and two csv files (*all_user_match.csv* and *users.csv*). The first one was obtained from Amazon by web scraping(*Web_Scraping.ipynb*) and contains 14 attributes about each book. The latter two files were from Book Crossing Community. *users.csv* contains users' personal information and *all_user_match.csv* contains how these users rate books. We cleaned the data and input them into MongoDB as three collections. The diagram above was made based on collections in our database, which is a little bit different from the original files.

The relational data schema is shown in the following three tables:

Table 1: all_user_match	
Attribute	**Type of Values**

User ID (from 1 to 6 digits)	Varchar
ISBN (10-digit)	Varchar
Rating (0-5)	Float

Table 2: Book Information	
Attribute	**Type of Values**
ISBN (10-digit)	Varchar
Auther	Varchar
Publisher	Varchar
Rank (34-18,781,568)	Float
Name	Varchar
Price (0.01-902)	Float
Tags	Varchar
Pages	Float
Introduction	Varchar
Reviews	Varchar
Customer Reviews (1890000+)	Float
Language (17)	Varchar
Stars distribution (0-5)	Float
Rating (0-5)	Float
Category (36)	Varchar

Table 3: User Information	
Attribute	**Type of Values**
User ID (from 1 to 6 digits)	Varchar
Location	Varchar