Step1: Convert .dat File to .csv File, Which Can be Recognized by Mongodb

See java program: DatToCSV.java

Strategy:

replace "," with "&" , " \" " with" ^" , "::" with ","

(Final when we need it, we will change it back)

Step2: Open Mongodb Service

Command Line: mongo

Step3: Create Mongo Database and collection

See java program: CreateMongoDB.java

Using java to link to localhost, and create database and collection.

Step4: Import .csv to target collection

Command Line:

mongoimport -d DA -c assignment2_movies --type csv --headerline (fileLocalUrl) (E.x.) (/Users/chengaoxiang/Desktop/18_Fall_Semester/Data_Analytics/Assignment/Assignment2/ml-10M100K/Movies.csv)

Step5: Running Query to Get Result

Using Java

Problem 1:

Program : query1.java Result : query1_result_MoviesRating

Strategy:

Name[movieID] = movieName, aveRating[movieID] = Rating[movieID] / count[movieID]

For every movieID which count is not 0, output its aveRating.

Problem 2:

Program : query2.java Result : query2_result_SimilarUsers

Strategy:

{movie1, movie2,···} = { {user1,user2···} {user1, user5···}{···}···}

{user1, user2,···} = { {movie1,movie2···} {movie1, movie5···}{···}···}

For target user, return all movies he tagged, then for each of those movies, return users who have tagged it.

Problem 3:

Program : query2.java Result : query2_result_SimilarUsers

Strategy:

Extract Genres type, split with "|" , using list store all genres, when contains, add new, when exists, arr[list.indexOf(genre)]++.