**Project 5**

**File and Database**

**Due date: 12/12/2014**

1. **Project description:**

In this project, you will implement an application to read data from a set of files, create a database to store the given information, and update a database for a movie recommender system service. We have three files: movies.dat, user.dat, ratings.dat and our job is to create a set of tables in the database that store information from these files and execute some queries for a researcher who is doing research for movie recommender system domain.

* 1. Formats of data files

The format of the file *movies.dat*:

MovieID: maximum 6 characters

Title: maximum 100 characters

Genre: maximum 50 characters

Year: integer

MPAA rating: maximum 20 characters

Directors: maximum 50 characters

Company: maximum 50 characters

The format of the file *users.dat*:

UserID: maximum 8 characters

Password: minimum 6 characters and maximum 10 characters

Age: integer

Gender: one character, either ‘F’ or ‘M’

The format of the file *ratings.dat:*

USERID: maximum 8 characters

MOVIEID: maximum 6 characters

RATING: an integer from 1 to 5.

All these fields will be separated by a comma in every data file.

* 1. Structure of tables: we are going to create three tables that will help us to store data from these three files accordingly. One table corresponds to movie.dat file, we will name this table as MOVIE. One table corresponds to users.dat file and we will name this table as USERS. The last table corresponds to ratings.dat file and we will name this table as RATINGS.

Table MOVIE:

MovieID: char(6) - Primary key

Title: char(100)

Genre: char (50)

MovieYear: integer

MPAARating: char(20)

Directors: char(50)

Company: char(50)

Table USER

UserID: char(8) - Primary key

Password: char(10)

Age: integer

Gender: char(1)

Table RATINGS

USERID: char (8) Primary key

MOVIEID: char (6) Primary key

RATING: integer

* 1. Search requirements:

You will be reading data from these three files (small sample given in D2L to get us started). Your job is to write Java code to create the three tables to store those data and create the following reports from those tables. Your program will allow 2 search options:

Search by title:

The user will enter the movie title and the program will search the movie table for the movie with matching title. If found, it will display the information of the movie: title, year, genre, director, and mpaa rating. If not found, it will display a message indicating that the movie is not found.

Search by star rating:

The user will enter the start rating (say 3), the program will query the whole database (probably need to use join query here) to find movies with at least that star rating (such as at least 3 star ratings). If found, it will display information about these movies: title, year, genre, director and mpaa rating. If not found, it will display a message indicating that the movies with at least such ratings are not found.

II. **Submission**

\*.java files.

*Your own* data files. I’ve given three small data files to get you started. You should populate your data file with at least 20 records for each file. The records should represent well all three age group (i.e including users who is younger than 20 years old, as well as users from 20 to 40 years old and users who is older than 40 years old.

Compress them all together in a \*.zip or \*.rar file and drop it to the dropbox Project 5 by the due date (midnight)