# Project1 Report

Xin Jin 426009396

## (a) Project Description

In this project, I collect information about top 1000 movies from IMDB which are ranked by rating. I collect basic attributes of the movie, title, rating, classification, category, publish time and length. I also collect the director, writer, and two stars from each movie, and the detail information about director, writer, star. In this project, you can select, update, insert and delete the tuples in the data system. In the database system, I create the schema called project 1. In this schema, there are 7 tables in the following:

1. movietop1000, it includes the basics attributes about the top 1000 movies.
2. Starsinmovie, it describes the stars who belong to the specific movie. The attributes are title, star1, star1, which describes the relationship about star and movie. For every movie, I collect the most important two stars. And the titles in this table are the same with the titles in movietop1000.
3. Writerinmovie, the same with the starsinmovie, but I only collect the one writer for each movie, so the attributes are only title and writer. The titles in the table are the same with the titles in movietop1000.
4. Directorinmovie, the same with the writerinmovie except the attribute name.

5.6.7 star/director/writer, in these three tables, they describe the name, born date and born place of them.

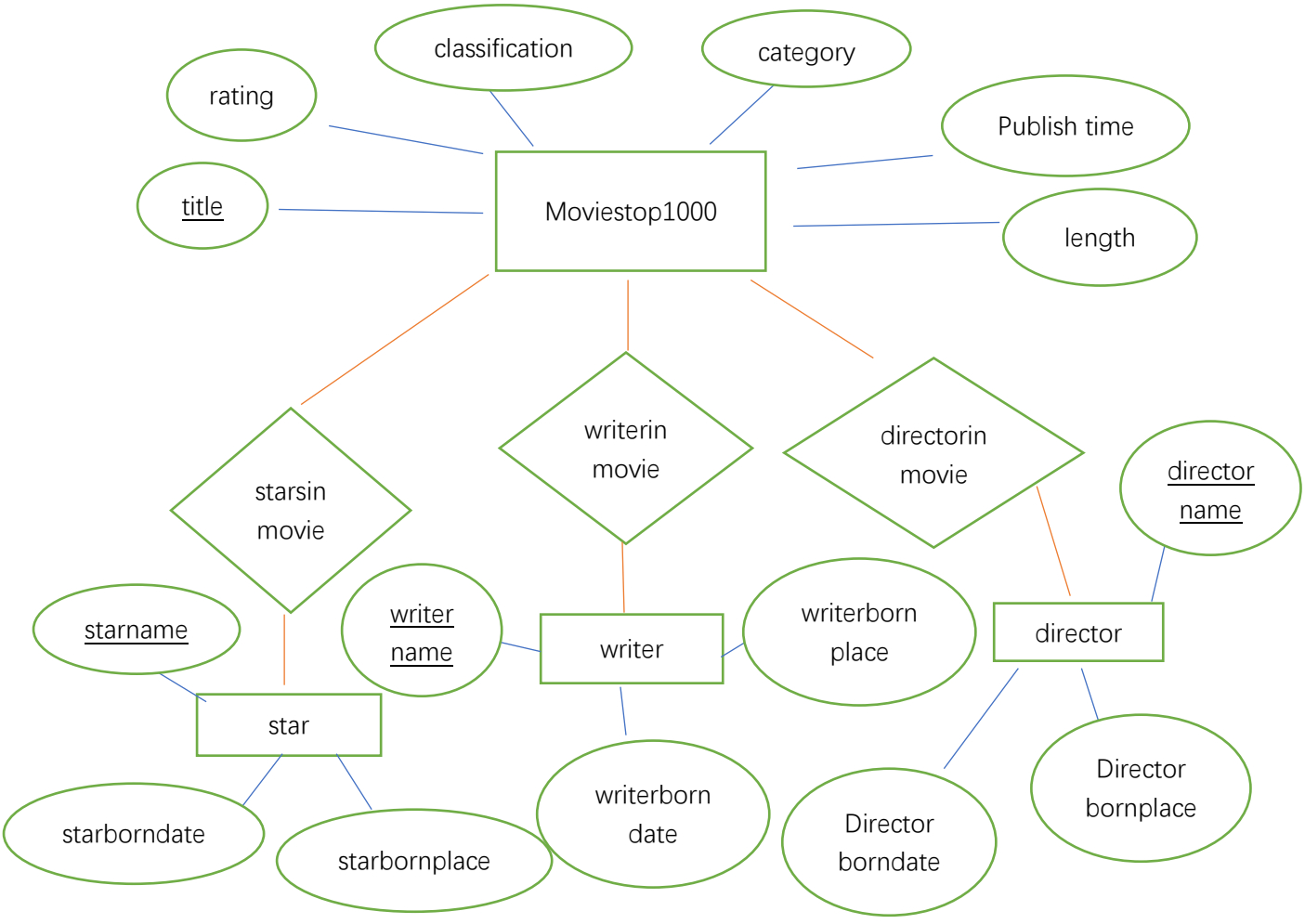
In the application, I provide four applications, the first one is the most important one in my application. It is query the object you want, which is the SELECT command in SQL. For selecting, I do separate that in 7 parts which will make users to select efficiently. 1.Use basic attributes to select movie, in this part, you can use select movie by rating or publish time or any combination of the basic attributes. For example, you want to find the rating is higher than 8 and publish time is after 2016, input 8 and 2016 in the blanks of score from and publish time from and you click the submit. You will get the results. 2. 3. 4.Use stars/director/writer to select movie, for example, you want to find movies in which Leonardo DiCaprio acts. Input the name and you will find the movies which have Leonardo DiCaprio. 5.6.7 Use movie to find the star/director/writer, just put the movie's title and you will find the name you want, after that you can get further detail about the star/director/writer, the detail includes name, born date and born place. Aside from selection, we can update the database in 7 ways, it is the same with select because of the structure of the database. 1.Update the basic attribute of the movie. For updating, we should input the information about the old one correctly, which means we should input the six attributes correctly and then do the update. 2.3.4 Update the relationship between movie and director/star/writer, in these parts, we do input the old movie title and director/star/writer and the new one and the tuple will be updated to be the new one. 5.6.7 Update the detail about director/star/writer, input the old information and new information to do the update.

Different from select and update, the process of delete and insert look easier. In insert, you should put in all the attributes you want to insert one time, includes the basic 6 attributes and the star, director, writer and their details. You can just insert some of them, and you can leave others blank. For example, you want to insert a movie, but you don't know its rating, so you can leave the rating blank. However, you must input the title of the movie, because in this structure, the movie title is the core and If you insert information without the title, it will make no sense. For the delete, it is much easier, just provide the title and it will remove all the information related to this title, also the detail of director, star, and writer. Also, it will make no sense if you just want to delete information about the writer/director/star. Because they have the relation with the movie, if one of them is deleted, this database will not be complete since it is a database about movie.

## (b) Data collection

In this project, the data are all from IMDB. For collection, I use a software called Octopus collection. In that software, I first open the web page and then make a loop about flip page. And then I make the loop of clicking every movie to get a new page about the detail of the movie. In the new page, I use regular transformation to get the specific data I need. After the collection process I get the data in Excel and insert the data into the tables.

## (c) The E-R diagram



## (d) Table normalization

The tables are in Boyce-Codd normal form. I did not need to change them.

## (e) User interface

I first make a homepage, and then there are four buttons for select, update, insert and delete. Click the button and then the page will jump to the specific page, there are 7 ways in select and update. Only one way in insert and delete. They specific detail are in the (a) project description. The ways how user use the way is to input the information and then submit, and then they will get the results, the details are also in the (a) project description. Every page has a button to back to the home page which will make the user to use it efficiently.

## (f) Project code

https://github.com/auznjx/project1-for-cse-608

## (g)Discussion

During this experience, I learnt quite a lot. First, I learnt how to collect data from the website, I believe this will do a lot help for me in the future. Second, I am more skillful in using mysql, when I just started doing this project, I need to see the lecture and book to write the command. After some days, I found I could write that without book, and I could do that efficiently. Third, I learnt how to use PHP to writer web page. This is the most important part in my project. Before doing this project, I did not hear about this programming language. However, after doing this project, I have a basic knowledge about this language. I know how to do some expression and how to use it to manipulate the database. I know how to let user to use the interface to manipulate the data. I now know more about webpage and how it works. I think I may need to learn more about PHP because I think this language is widely-used and it can help me to do more application in the future.

Aside from the things I learnt, I also met some difficulties. There was a very big problem which made me stuck in a long time. In my database, there are some titles include some French and Russian letters, when I want to output this kind of movie, I found the letter was outputted as messy code. I did not why this happened that time, so I searched this question in google and started to learn the basic knowledge about Unicode. However, I put the write command and I did not get the right output as I expected. At last, I found that command in PHP is different from HTML, in PHP it should be utf8, and in html, it should be utf-8. I solved the problem at last. It was a valuable experience for me and I learnt more about the Unicode and how it works.