CSCE 608 Database Systems
Fall 2015

Project 1 Report

Submitted By:
Deep Desai
124001412

Topic

The topic for my application is “Personal Movies Management”
The application is meant to make a movies watch-list.
A simple UI is made in form of a website - https://personal-movies-management-deepd.c9.io

Program Components

The development for this project was done on cloud9 IDE platform and git for version control. Cloud9 provides a virtual machine with Ubuntu 14.04.2 LTS installed and an IDE for online development.

The languages and softwares used for this project are as follows:

1. Relational database: MySQL
2. Web Development: HTML, PHP (server-side scripting language), CSS and Bootstrap (A free and open-source front-end framework)
3. Data manipulation: CLI commands (cat, awk, sed) and Python script

Database Schema

- Database name : c9
- Database tables : mytable (main movie database), myplan (database of watch-list, referencing mytable)

mytable	myplan
Year varchar(255)	Year varchar(255)
Length varchar(255)	Length varchar(255)
Title varchar(255)	Title varchar(255)
Subject varchar(255)	Subject varchar(255)
Actor varchar(255)	Actor varchar(255)
Actress varchar(255)	Actress varchar(255)
Director varchar(255)	Director varchar(255)
Popularity varchar(255)	Popularity varchar(255)
Awards varchar(255)	Awards varchar(255)
ID int not null auto_increment primary key	ID int primary key, FOREIGN KEY (ID) REFERENCES mytable(ID)

Data Collection and Storage

I have collected data for my application from the dataset provided on this website : "http://perso.telecom-paristech.fr/~eagan/class/as2013/inf229/labs/datasets". It has 1659 tuples.

The data obtained was semicolon-separated. The columns in the dataset were in the format:
" Year;Length;Title;Subject;Actor;Actress;Director;Popularity;Awards;Image ".

I converted this dataset in the format:
' "Year","Length","Title","Subject","Actor","Actress","Director","Popularity","Awards" '

I used the following command for conversion and to store it in a new file:

```
cat film.csv | awk '{for (i=1;i<NF;i++) $i="\""$i"\""}1' FS=";" OFS=","|awk '{$NF=""; print $0}' FS=',' OFS=',' | sed 's/,$//' > ~/newfilm-final.csv
```

I made the tables in mysql and loaded the data using a python script, seed.py:

```
import MySQLdb

db = MySQLdb.connect(host="localhost", user="deepd", passwd="", db="c9")
cur = db.cursor()
cur.execute("create table mytable(Year varchar(255), Length varchar(255), Title varchar(255), Subject varchar(255), Actor varchar(255),"\
                "Actress varchar(255), Director varchar(255) , Popularity varchar(255), Awards varchar(255), ID int not null auto_increment primary key)")
cur.execute("create table myplan(Year varchar(255), Length varchar(255), Title varchar(255),"\
          "Subject varchar(255), Actor varchar(255), Actress varchar(255), Director varchar(255) , Popularity varchar(255), Awards varchar(255),"\
        "ID int primary key, FOREIGN KEY (ID) REFERENCES mytable(ID))")
f = open('newfilm-final.csv')
for line in f.readlines():
   cur.execute("insert into mytable values(%s" % line+", NULL)")
db.commit()
```

Website Features

- Home - Lists the watch-list of movies with a remove button for each movie to remove it from the watch-list
- Search - Lists the movies according to the search, with a "Add to my plan" button for each search result to add it directly to the watch-list
  - Simple Search - Dropdown menu to search the movies database by column name (substring search)
  - Complex Search - Input complex search query
- Movies - Dropdown menu for operations on main database of movies
  - Add - Add new movie to database

- Delete - Delete button for each movie to delete the existing movie from database (only deleted if not in my plan)
- Update - Update/edit the existing movies in the database
- List - List all the movies from database
- Plan - List all the movies from database and provided a “Add to my plan” button for each entry to add it directly to my plan’s watch-list

Difficulties Faced

1. Finding dataset - It’s really hard to find a good dataset for movies, because either you find only movies list or just a list of directors/actors etc. but it’s hard to find a dataset with all the relevant information for movies. I had to search a lot to find this one.

2. PHP & HTML - It took some time to understand how PHP and HTML would work together and how one can write intermingled PHP and HTML code.

3. Update records - I was finding a lot of difficulty in figuring out how to implement update feature, I figured out after a lot of different approaches that to update a record it is better to have a unique key so that updating other columns become easy. Had to change the db schema accordingly.

4. Character Set - I was having a hard time figuring out the character set (unicode, I was using UTF-8 earlier) that would represent the characters like ó , é , ö , etc. correctly.

Things Learned

1. Learnt more about the referencing amongst tables in database systems.
2. Learnt about options like cascade on update/delete while using referencing. (I haven't used cascade in this project, because of such a requirement)
3. Learnt full-text indexing facility provided in mysql. (I haven't used it in this project, because it is not required here)
4. Learnt PHP, Bootstrap and CSS.