Databases Final Project

Tyler Lee ([tlee93@jhu.edu](mailto:tlee93@jhu.edu)) in 600.315

Jeffrey Sham ([jsham2@jhu.edu](mailto:jsham2@jhu.edu)) in 600.415

**Jeffrey + Tyler’s MOVIE AWARDS Database**

**2) Description**: Our database contains information about the Academy Awards and Golden Globe Awards winners and nominees between the years 2000 and 2010. Using PHP and HTML, we created a web interface in which you can search our database, advanced search our database, and edit our database. We used mysql on dbase and our website can be found at ugrad.cs.jhu.edu/~tlee93.

**3) Design Update:** Our database design has changed slightly. We do not have a table for production companies because this information was not readily available to us in the csv files that we downloaded from the internet. We decided that information pertaining to a film, such as genre, language, country, would be inputted as null fields in our database by default to allow for the user to input this information later. The idea is the director (and only the director) can log into the database and input this information for his or her own film.

**4) Loading the database**: We downloaded csv files containing the information from aggdata.com. Those csv files are called academy\_awards.csv and golden\_globes.csv. We wrote scripts in java to parse the csv into SQL queries. These files are ParseExcel.java for the academy awards data, and ParseExcel2.java for the golden globes data. The data was formatted slightly differently in the csv files, so there are modifications in the java files. We decided to limit the amount of input to only go back to 2000. We considered the acting and directing awards to merit a nomination for the person, while all other nominations went directly to the film in our database. We had to do a few tricks to account for commas and apostrophes contained in peoples’ names and in the titles of films, such as Mo’Nique and Pan’s Labyrinth. The SQL code that was generated from these scripts can be found in sql\_code\_for\_academy\_awards\_new, and sql\_code\_for\_golden\_globes\_new, respectively.

This code also generated a file called personOut.txt (as well as personOut1.txt, movieOut.txt, and movieOut1.txt) which contains names and email addresses that we generated by appending first\_name to an ID number we assigned. Using this code, we found information on individual birthdates and birth places on imdb.com. The SQL code that updated the database to reflect this new information can be found in Update\_Oscar\_People\_and\_Places.

The file SQL\_code.sql contains the relational table specification.

You’ll notice that each nomination has its own ID, which means that a film that has 8 nominations will be found inserted into the film table 8 times with 8 different IDs in the NominatedFor field. Likewise, people with multiple nominations will appear in the table Person multiple times with different IDs in the NominatedFor field. Another thing to note is that some films will show different years because the golden globes list them in a different year than the academy awards do. The golden globes list the year of the award ceremony, while the academy awards list the year the film was released, which is typically the previous year but not always.

**6) User’s Guide:**

Mysql on dbase information:

Username: cs41515\_jsham2

Database: cs41515\_jsham2\_db

Password: PTIASBIT

Go to ugrad.cs.jhu.edu/~tlee93 to find our website.

In the top left corner you’ll find fields to enter a username and password. Usernames are unique and they are automatically generated email addresses. We made one for you: username is [yarowsky@jhu.edu](mailto:yarowsky@jhu.edu) and password is also [yarowsky@jhu.edu](mailto:yarowsky@jhu.edu). Using this login information, you will be able to edit and vote, which you would not be able to do if you were not logged in. You would only be able to edit your own profile, which contains the date of birth field. You will then be sent an email to notify you. This was implemented using triggers, which Shyam gave us permission for. You may also vote by selecting a year. You are only permitted to vote once, and after you’ve voted you will receive an email notification. We have a column that keeps track if you have voted yet. The idea here is that in a given year, voting members can submit their vote, and they may only vote once.

If you try logging in as a director, you are also able to edit your own films! Try logging in with username: [clint222@jhu.edu](mailto:clint222@jhu.edu) and password [clint222@jhu.edu](mailto:clint222@jhu.edu). This is Clint Eastwood. You can edit information about his award-nominated films. You must submit all fields. Hypothetically if Clint Eastwood had this email, he would get a notification, but these emails are triggered to us. If the film was nominated for multiple awards, it will have multiple tuples. These each have a different NominatedFor ID, which is hidden from this view.

You do not have to log in to do basic search or advanced search. The basic search lets you select a table from the dropdown menu and then input your search. For example, if you type Tom Hanks using the Person dropdown, you will see three rows because he received three Oscar or golden globe nominations between 2000 and 2010 for Castaway and Charlie Wilson’s War. Or try typing in Slumdog Millionaire using the Film dropdown. You will see many nominations.

The advanced search lets you limit your search by entering multiple fields. Click the advanced search button to take you to that page. There are specific instructions at the top of the page that are critical if you want to query in 2 or more fields contained in different tables. Try an example: if you simply search for Black Swan in the InFilm field, you will get three results. If you go back and limit it further by searching for F in the Gender field, you will get two results: Natalie Portman and Mila Kunis, because director Darren Aronofsky is male.

**7) Specialization**: We did major work in creating a forms-based interface with sophisticated report generation—this is the advanced search function. We also did significant work in triggers. We receive email notifications when someone votes or edits information in the database. You ([yarowsky@jhu.edu](mailto:yarowsky@jhu.edu)) will receive email notification if you edit your profile or vote. We also did significant work in security. The Film table is only editable by directors, and further, these directors may only edit their own films. Our database is also secure in that voters may only submit one vote.

**8) Limitations:** Our database voting mechanism only allows one vote per person, but ideally we would like to allow one vote per person per category. If we wanted to get really particular, a person would only be allowed to vote in certain categories and for awards to which they are members. For example, as an actor, Colin Firth should only be allowed to vote for acting awards, and further as a member of AMPAS and not the HFPA, he would only be allowed to vote for the Oscars, not the Golden Globes.

Another limitation is that we have data duplication because a film or a person can have multiple nominations, so we gave them multiple tuples with one unique NominatedFor ID per nomination. Ideally, we would have just one tuple for a certain film or person. Perhaps we could allow the Award table to take in Film.ID/Person.ID as a foreign key, joining the two tables by the film rather than by the award. But sometimes we also specify a person, not just a film, for the nomination. The Person.ID would have to be left null in cases where the award is for a film like Best Animated Feature. A disadvantage to this method would be that it would not be immediately obvious how many awards Toy Story 3 was nominated for in the Film table; you would have to look to the Award table too. There are advantages and disadvantages to both methods.

**9) Outputs**: You can find interesting sample outputs on the homepage at ugrad.cs.jhu.edu/~tlee93. Here you will find five sample queries and the resulting tables. One such query required us to calculate age by using date of birth and the year in which an award was given.

**10) Full Relational Specification:** This can be found in the file SQL\_code.sql. Note that Person and Film do not have primary keys because a single person or film can have multiple tuples for multiple nominations, therefore the ID would not be unique. Though NominatedFor is a unique column, it is not particularly helpful for identifying the Person or Film because the ID is not intuitive.

**11) SQL Code**: You will find the SQL queries located on the homepage in the file queriesForTheWebsite.txt. You will also find our SQL code for building the database in our tar file.

**12) Files included in TAR**:

Subdirectory: cs41515\_jsham2\_tlee93

* academy\_awards.csv
* advanced\_search.php
* basic\_search.php
* edit\_dob.php
* edit\_film.php
* edit\_film\_attr.php
* edit\_info.php
* edit\_profile.html
* golden\_globes.csv
* index.php
* loggedIn.php
* login.php
* movieOut.txt
* movieOut1.txt
* ParseExcel.java
* ParseExcel2.java
* personOut.txt
* personOut1.txt
* queriesForTheWebsite.txt: containing SQL queries
* README2: txt file containing write-up composing phase II
* set\_up\_adv\_search.html
* SQL\_code.sql: contains full relational table specifications
* sql\_code\_for\_academy\_awards\_new.txt: containing SQL queries to build database
* sql\_code\_for\_golden\_globes\_new.txt: containing SQL queries to build database
* submit\_vote.php
* Update\_Oscar\_People\_and\_Places.txt: containing SQL queries to build databases
* vote.php

Git repository containing the same files: https://github.com/jeffreysham/DatabasesFinalProject.git