



CSCE 2501

Fundamentals of Database Systems

Project - Oscars

Prof. Hossam Sharara
Dept. of Computer Science, AUC

Introduction

In this project, you will be building a database application for Academy Awards (Oscars) over their 96 iterations, that allows users to get information about the nominated and winning movies, actors, directors, etc. for the yearly academy awards.

The main site that provide such information is Wikipedia (https://en.wikipedia.org/wiki/96th_Academy_Awards). Wikipedia doesn't provide filtering or higher level aggregations for statistics. Hence, your application will be used to provide more flexibility on that front.

Your task will to build a database system backend and an application that provide the Academy Awards statistics and features described next.

Milestone I: Data Requirements

The database system you design should be based on the information provided on the nominees in each of the Academy Awards categories, along with the associated movie. The URLs for each iteration of the Academy Award list of nominations and prizes can be constructed by replacing the iteration number in this url (https://en.wikipedia.org/wiki/<iteration_number>_Academy_Awards) e.g. for the 1st iteration, the <iteration_number> should be replaced by “1st”, etc. till the 96th iteration.

Your database design should store all the information about the movies, including but not limited to: movie name, release date, actors, director, producers, production company, language and run time. For each actor / director / producer, you need to store their full name, date of birth, country of birth, and death date (if applicable). For each nomination, you need to store the nominated movie, the nomination category, the nominated staff member(s), and whether the nomination was granted (won) or not.

The system should also allow users to register on the system using their email addresses, and pick a username, gender, age, birthdate, and country. After registering, the users can add their own nominations for any movie in any category for any iteration.

Milestone I: Database Design & Implementation

Tools: MySQL

Deadline: 5th March, 2025

Description:

In this milestone, you are required to design the database ERD, create your database and schema in MySQL server for the data requirements presented on the previous slides.

Deliverables: (20 points)

- Entity-Relationship Diagram of the system
- Relational Model for your system
- A transcript for the SQL statements used to create your database and schema

Milestone II: Database Population

In this milestone, you are required to write a web crawler to crawl the pages from Wikipedia for all the Academy Awards nominations in their 96 iterations, parse the HTML you crawl and extract the relevant fields for populating the non-user tables in your schema. Note that you may need to extract some of the information from the cards on the landing page for the actors / directors / etc. on Wikipedia

At the end of this milestone, you should have all your non-user tables populated with the extracted data. For the user tables, use sample test information to populate the relevant tables.

Milestone II: Web Crawling and Data Population

Tools: Scrappy, pycharm, BeautifulSoup, selenium, python (or any other crawler library / language)

Deadline: 26th March, 2025

Description:

In this milestone, you are required to implement your crawler, extract the data from the Wikipedia website and populate your database

Deliverables: (50 points)

- Crawling script
- Populated MySQL database dump
- CSV files for the values in each table

Milestone III: Application Layer

In the final milestone, you are required to design a client application that is capable of connecting to the database hosted on remote MySQL server. The application should have the following functionalities:

- Register a user
- Add a new user nomination for a staff member for a given movie
- View existing nominations for the user
- View the top nominated movies by the system users in each category / year
- Show the total nominations and oscars for a given director, actor, and singer
- Show the top 5 birth countries for actors who won the best actor category
- Show all the nominated staff members born in a given country, along with the categories they were nominated for, number of nominations and oscars in descending order
- Dream Team - Extract the living cast members that can create the best movie ever (director who won most oscars, best leading actor, actress, supporting actor, supporting actress, producer, singer for the movie score).
- Show the top 5 production companies by the number of won Oscars
- List all non-english speaking movies that ever won an oscar, along with the year

Milestone III: Application Layer

Tools: Any language / platform to implement your application (Web-based, GUI, command-line are all acceptable)

Deadline: 16th April, 2025

Description:

In this milestone, you are required to implement an application to perform various transactions / queries on your backend database system. The database has to be hosted on an actual server (you can use any of the free DB hosting services available (e.g. www.db4free.net)) and make sure your application is communicating with this service not the localhost

Deliverables: (30 points + 5 bonus points for GUI/Web-based apps)

- Application Implementation / Source Code + Executable
- Latest Dump of your database
- Video Demo