Project Report

Academy Awards Database

Database Systems - Further Assessment

Author: C Marteli – 19598552

Introduction

This database aims to store information on oscar nominees (Actors, Directors, Crew) the films they were nominated for (Name, Year) and whether they were winners or not. With the ability to store data on multiple Academy Awards ceremonies as well as the venue and category info.

User Guide

How to use:

#connect to mysql

:~/$ mysql -u me -p

then enter password

:~/$ <enter password>

load create tables file

:~/$ SOURCE create_tables.sql

load sample data input file

:~/$ SOURCE input_data.sql

#load query file

:~/$ SOURCE queries.sql

Design choices

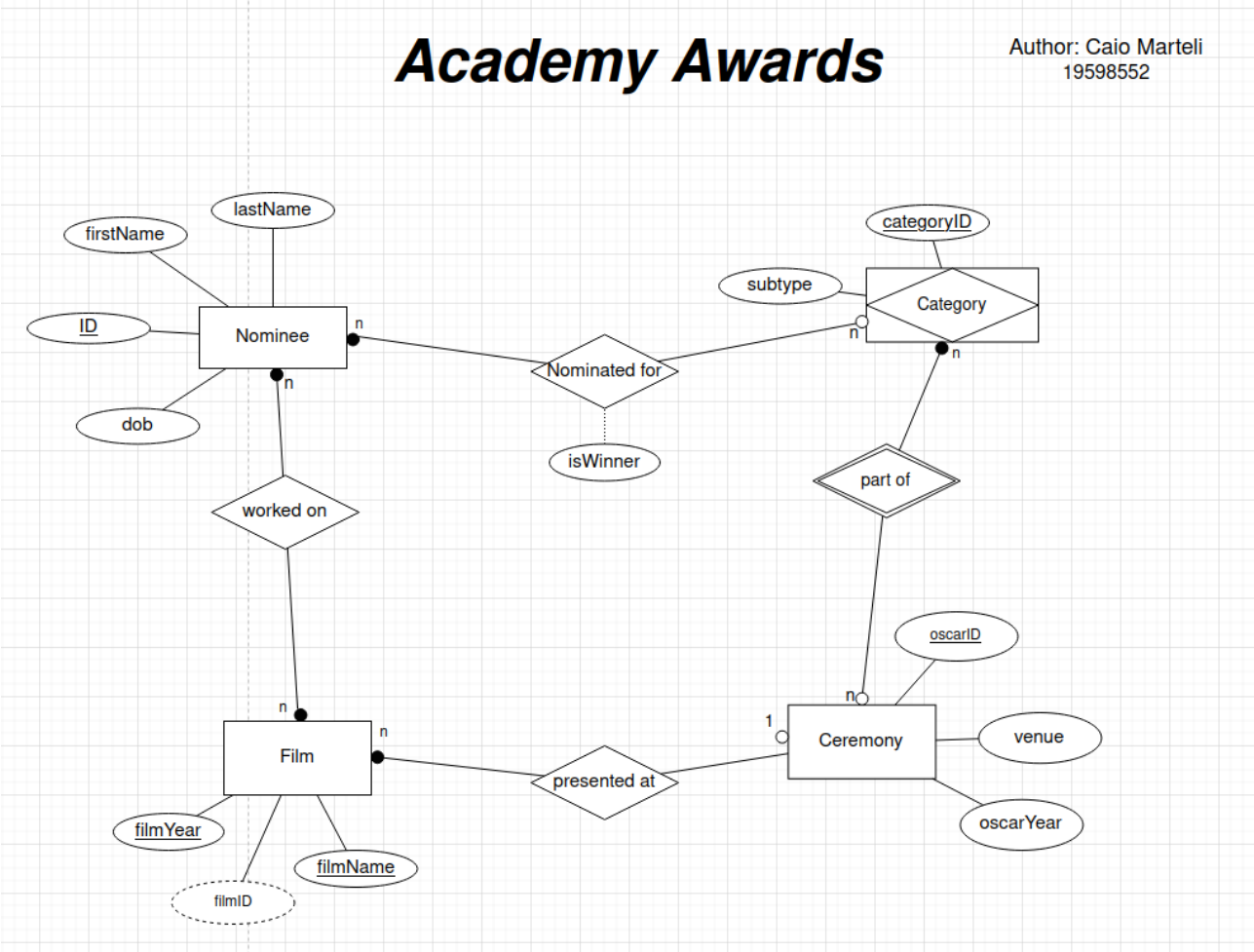
The initial database design consisted of separate entities for Actors, Directors and Staff but due to the scope of the project being too large for the time given and not enough sample data readily available, those entities were condensed into the a single “Nominee” entity. This design choice turned out to make more sense in the long run as a someone who was previously nominated for best director may in fact be nominated for best actor in a future ceremony (or any other role for that matter).

The entity “Ceremony” was an easy decision as it allows for multiple academy awards to be indexed and queried by using oscarID, simply searching by year would not suffice as the Oscars tend to cover movies from two years.

Film was another important entity as I wanted a table to easily determine how many nominations a particular film received.

Category was not added until later after the first design was finalised. Originally the relation “nominated for” was going to hold the category attribute. However, it was clear I needed category to be a foreign key to restrict data to only existing categories. This also allowed for cascade deletion of entire categories and flexible queries for example if I wanted to know who the recipients of “Acting” awards in the 91st Oscars were.

ERD:


Academy Awards
Author: Caio Marteli
19598552
lastName
firstName
ID
Nominee
dob
worked on
Nominated for
isWinner
categoryID
subtype
Category
part of
oscarID
Ceremony
venue
oscarYear
presented at
Film
filmYear
filmName
filmID
n
1

Entities:

Ceremony
The oscars or academy awards of a particular year.
Attributes:
crmyID – The number of the ceremony since the beginning of the awards (Unique)
crmyYear – The year the ceremony was held
location – The location of the Awards in a particular year

Ceremony
The multiple categories someone may be nominated for.
Attributes:
categoryID – The name of the category (unique)
subtype – The nature of the category for example “Acting” or “Technical”

Film
Attributes:
filmName – The name of the nominated feature (Composite Key)
year – Year of production of film (Composite Key)
filmID – unique identifier for a film consists of film Name+year (Derived)

Nominee
Attributes:
nomID – Unique nomination identifier
fullName – Name of Nominee(s) (can be multiple)
isShared – Indicates if it’s a shared nominations

Entity Set	Key	Other Atributes
Film	filmName, filmYear	
Nominee	ID	firstName, lastName, dob
Ceremony	oscarID	oscarYear, venue
Category	categoryID	subtype

Relationships

Nominated for
Nominee - (N : N) – Category
Nominees can be nominated to one ore more categories.
A single category may have multiple nominees.

Constraints:
An award must have a recipient even in the case of Best Picture (Director or Producer) therefore a Film must have a Nominee
A nominee must have taken part in a motion film to participate

Worked on
Nominee - (N : N) – Film
One Nominee may have worked on multiple Films
Films may have multiple nominees.

Constraints:
A film must be nominated to be at the Ceremony and therefore a nominee is required to receive the award.
Films need producers to be made as such a nominee will always exist.

Presented at
Ceremony - (1 : N) – Film
One Ceremony has many Film showcases
Films may only be showcased at one Ceremony

Constraints:
A film must be nominated to be at the Ceremony
Every Ceremony nominates films
A film must be nominated to be at the Ceremony

Part of
Ceremony - (N : N) – Category
One Ceremony has multiple categories
Category may only be showcased at multiple Ceremonies

Constraints:
A Ceremony may not have categories if it was honorary awards only for eg
Every Category has to be part of a ceremony.

Relationship Sets	Between Which Sets	Attributes of Relationship set
Nominated for	Ceremony, Film, Nominee	isWinner
Worked on	Film, Nominee	
Presented at	Film, Ceremony	
Part of	Ceremony, Category	

Relationship Sets	Cardinality Constraints	Participation/ other
Nominated for	n:n	listed above
Worked on	n:n	
Presented at	n:1	
Part of	n:n	

Relational Schema

Oscar(oscarID, oscarYear, venue*)*

*Category(*catID, subtype*)*

Films(filmName, filmYear, *oscarID FK*)

Nominees(id, firstName, lastName, dob)

Nominations(nomId, filmName, isWinner, *catID FK, id FK*)

Data types

-- TABLE **Ceremony**
oscarID SMALLINT PK
#Represents edition of Oscars since first ceremony. Numerical and less than 255. Unique.
oscarYear YEAR
#Year of ceremony. 4 digit format.
venue VARCHAR(60)
 # Venue ceremony is being held. Variable size

-- TABLE **Category**
catID VARCHAR(60) PK
#Category key to be under var size unique
subtype VARCHAR(30)
 # Sub Category award falls under (eg. Animation, Acting).

-- TABLE **Film**
filmName VARCHAR(60) **PK**
#Name of Film. composite key
filmYear YEAR **PK**
#Year of release. 4 digit format composite key

-- TABLE **Nominee**
id CHAR(8) PK
#used as main identifier generated for DB. Must be unique. Autoincrements
firstName VARCHAR(20)
#Can be null as not every nominee will have 2 names.
lastName VARCHAR(20) NOT NULL
#Must not be null needs a name to receive award.
dob DATE
#Date of birth date datatype

-- TABLE **Nominations**
id SMALLINT PK
#type of role nominated for
nomID CHAR(8) FK
filmName VARCHAR(60) NOT NULL
catID VARCHAR(60) FK

isWinner BOOLEAN NOT NULL
#true or false did they win?

SQL Queries

```
/* sampleQueries.sql: SQL file for Assignment Oscars */
/*Author: C Marteli 19598552 Created: 15/12/2021*/

-- Show all the winners from Nominations
SELECT * FROM Nominations
WHERE isWinner = TRUE;

-- Show all the Nominations ordered by film name
SELECT * FROM Nominations
ORDER BY filmName;

-- Show all the winners from Nominations and ID is less than 30
SELECT * FROM Nominations
WHERE isWinner = TRUE AND id < 30;

-- name of all nominees in "acting" subcategory only with winners at top
SELECT c.catID AS "Category", n.filmName, n.nomID, n.isWinner
FROM Category AS c
CROSS JOIN Nominations AS n
ON c.catID = n.catID
WHERE c.subtype = "Acting"
ORDER BY isWinner DESC;

-- Creates view of all winners by nominations using inner join
CREATE VIEW oscar_winners AS
SELECT CONCAT(nee.firstName," ",nee.lastName) AS "Name", ion.filmName AS "Movie",
ion.catID AS "Role", "winner" AS outcome
FROM Nominee AS nee
INNER JOIN Nominations AS ion
ON nee.id = ion.nomID
WHERE ion.isWinner = TRUE;

-- displays view
SELECT * FROM oscar_winners;

-- Creates view of all losers by nominations
CREATE VIEW oscar_losers AS
SELECT CONCAT(nee.firstName," ",nee.lastName) AS "Name", ion.filmName AS "Movie",
ion.catID AS "Role", "lost" AS outcome
FROM Nominee AS nee
INNER JOIN Nominations AS ion
ON nee.id = ion.nomID
WHERE ion.isWinner = FALSE;

-- displays view
SELECT * FROM oscar_losers;

-- now display both views as union
SELECT * FROM oscar_winners
        UNION
SELECT * FROM oscar_losers;
```

Conclusion

The main challenge would be the initial database design.

Ideally it would have been preferable to have many more entities such as Director, Technical Staff, Actors and more attributes such as gender and date of birth.

However, finding the information in a timely manner and then loading it would have been a great undertaking and time constraints would not allow it.
I was able to utilised some meaningful data from the website Kaggle (ref below) and being able to change it into SQL queries by using python scripting (script provided). Still, some of the data required to be manually altered and the process was time consuming.

In a future version of this project given more sample data I would add more info specific to the ceremonies and what the categories of each time were. As such more subcategories and locations and the already mentioned additional entities.

References

Input data adapted from:
*Kaggle (2021) The Oscar Award (Version 2.0) [.CSV File].*https://www.kaggle.com/unanimad/the-oscar-award