

Intro To Database Final Project

Group 第5組 110550063 張博凱、110550116廖梓程、110550069 余駿佑




- Main idea

我們希望透過這個網站和資料庫，來達成篩選或推薦電影的目的，因此我們希望使用者在瀏覽網站時可以透過我們給定的標籤來篩選或是找尋想看的電影

- Data

由於從kaggle上載下來的資料有點凌亂，再加上有些data是有問題的(column出現NaN、重複的row、意義不明的要素等)，所以我們有寫一個python來把csv整理一下，並且把一些csv整併一下，我們也會把code丟到github

整理前：

 credits.csv	2019/9/21 上午 10:29	Microsoft Excel 逗...	185,467 KB
 keywords.csv	2019/9/21 上午 10:29	Microsoft Excel 逗...	6,086 KB
 links.csv	2019/9/21 上午 10:29	Microsoft Excel 逗...	966 KB
 movies_metadata.csv	2019/9/21 上午 10:29	Microsoft Excel 逗...	33,638 KB
 ratings_small.csv	2019/9/21 上午 10:30	Microsoft Excel 逗...	2,382 KB

整理後：





 cast.csv	2023/6/15 下午 06:26	Microsoft Excel 逗...	3,210 KB
 links.csv	2019/9/21 上午 10:29	Microsoft Excel 逗...	966 KB
 movie.csv	2023/6/15 下午 06:26	Microsoft Excel 逗...	23,530 KB
 ratings_small.csv	2019/9/21 上午 10:30	Microsoft Excel 逗...	2,382 KB

Table details:

movies.csv:

id: **Primary Key**, 每一部電影都會有一個獨特的id

來源:credit.csv

belongs_to_collection: 有續集或其他篇章的才有，記錄了該作品是否為系列作品

來源:movies_metadata.csv

budgets: 預算(單位為美金)

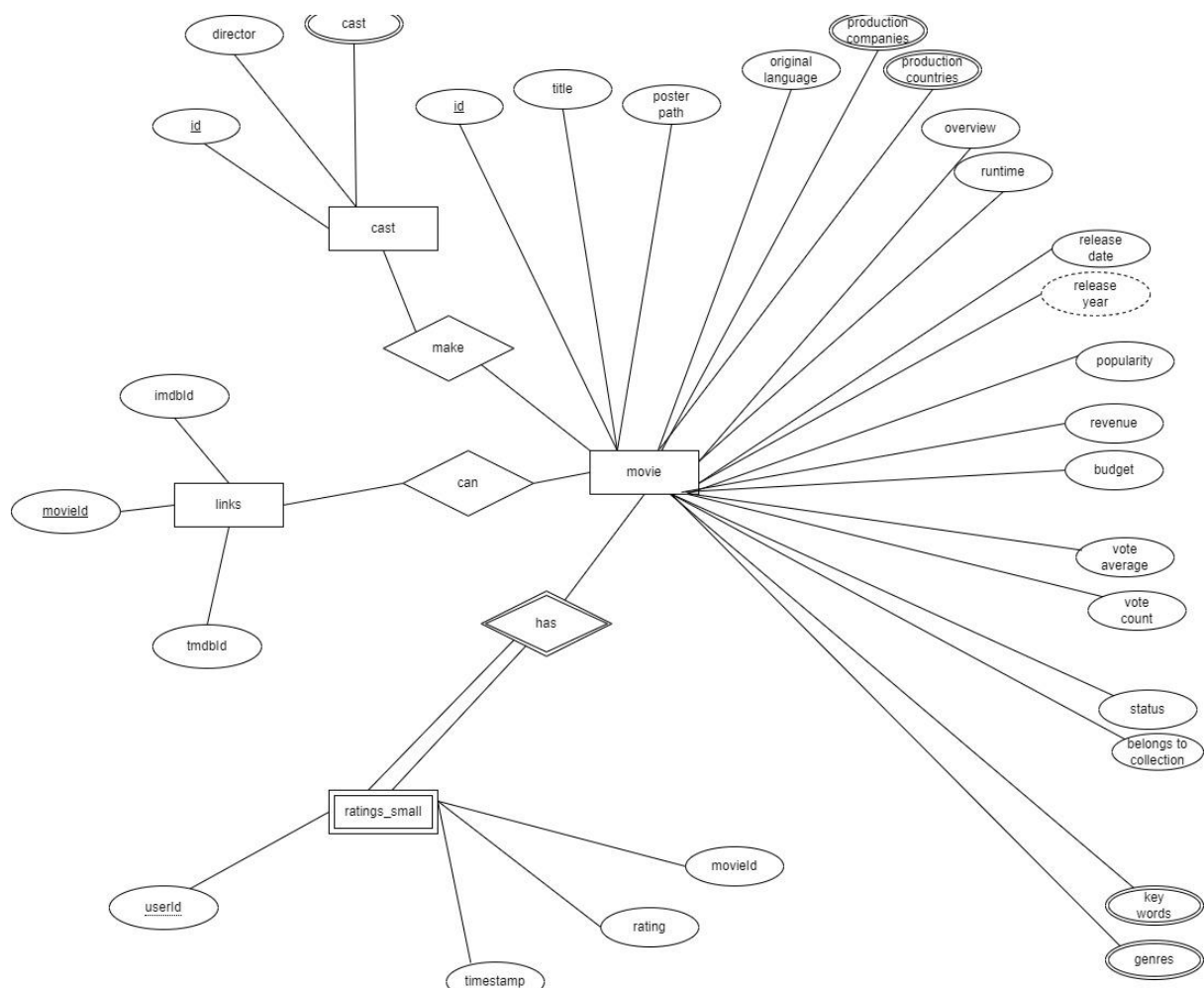
來源:movies_metadata.csv

genres: 代表電影的分類，一部電影可能有多個分類 EX:Animation,Comedy

來源:movies_metadata.csv
original_language: 原片的語言
來源:movies_metadata.csv
overview: 電影大綱
來源:movies_metadata.csv
popularity: 人氣指數(單位未知)
來源:movies_metadata.csv
poster_path: 電影海報的網址(實際上好像是檔名, 丟google可以找到海報)
來源:movies_metadata.csv
production_companies: 製作公司, 一部電影可能有多個製作公司
來源:movies_metadata.csv
production_countries: 製作國家, 一部電影可能有多個製作國家
來源:movies_metadata.csv
release_date: 電影上映日期
來源:movies_metadata.csv
revenue: 收益(單位為美金)
來源:movies_metadata.csv
runtime: 影片時長(單位為分鐘)
來源:movies_metadata.csv
status: 電影的狀態(是否上架之類的)
來源:movies_metadata.csv
title: 電影標題
來源:movies_metadata.csv
vote_average: 平均評分
來源:movies_metadata.csv
vote_count: 總投票數
來源:movies_metadata.csv
release_year: 上映年份
來源:movies_metadata.csv
keywords: 有關電影的關鍵字, 一部電影可能有多個關鍵字
來源:keywords.csv
cast.csv:
id: **Foreign Key**, 每一部電影都會有一個獨特的id
來源:credits.csv
cast: 主要演員的姓名
來源:credits.csv
director: 導演姓名
來源:credits.csv

rating_small.csv:
userId:使用者id
來源:rating_small.csv
movieId:所評分的電影id

來源:rating_small.csv
 rating:評分, 從零到五且精度為小數點一位
 來源:rating_small.csv
 timestamp:評分當下的時間戳
 來源:rating_small.csv
 link.csv:
 movieId:**Foreign Key**, 每一部電影都會有一個獨特的id
 來源:link.csv
 imdbId:在imdb網站中的id
 來源:link.csv
 tmdbId:在tmdb網站中的id
 來源:link.csv
 ER-modle:



● Database

- What database do you use

database we used: MySQL

datasets we used:

credits.csv

keywords.csv

link.csv

movies_metadata.csv

ratings_small.csv

datasets we create by eliminating some redundant information from original dataset:

cast.csv

movie.csv

- How do you maintain your database

使用者除了可以查詢電影以外，還可以為電影評價，除此之外，還可以修改或刪除既有的評價，來回饋我們的資料庫

- Describe how you connect your database to your application as detail as possible

我們的網站是用php寫的，在php中，我們可以使用mysql_query()的函式來連接資料庫。我們先在網頁上寫出我們需要的搜尋欄位，接著在這些欄位中輸入文字，透過\$_GET['欄位名稱']的函式，我們可以讓php讀到我們所輸入的文字，接著我們再透過這樣子的方法輸入所有我們想要query的資訊，接著將使用這些資訊拼接出一個合法的SQL語句，並以字串的格式儲存，最後再將這個字串丟入mysql_query()中，並將SQL查詢的結果存入到暫存器中，最後我們就可以使暫存器中的資料來show data

- Application

- Describe the interface of your application

圖(一)為電影的欄位

Movie Filter

Title

Enter title

Actor

Enter actor

Director

Enter director

Genre

Enter genre

Language

--- Select ---

Release Year

--- Select ---

Company

Enter Company

Keyword

Enter genre

Search

圖(二)為評價的新增、修改、刪除的欄位

Navbar

Home

Rating

Movie Rating

UserID

Enter title

Movie

2:Ariel

Rating

Enter genre

Rate

Delete Rating

UserID

Enter title

Movie

2:Ariel

Delete

- Describe the functions of your application

- 1.Movie_Filter**我們可以使用Title、Actor、Director、Genre、Language、Release_Year、Company、Keyword作為過濾以及篩選電影的方式，按下Search後會自動將使用者輸入的條件轉為SQL query到我們的資料庫中找尋符合的Data。
- 2.新增MovieRating** 輸入UserID、選擇要評價的電影名稱，並提供你的評價。
- 3.修改MovieRating** 輸入UserID、選擇要評價的電影名稱，如果該id及電影已經存在，則覆蓋原有的評價。
- 4.刪除MovieRating** 輸入UserID、選擇要刪除評價的電影名稱，便會刪除符合的評價資訊。

- For each function of your application describe how you make it possible in detail

Filter相關:

在查詢的部分SELECT * FROM movies INNER JOIN cast ON movies.id = cast.id WHERE 1=1, 再判斷是否有輸入, 並加上類似AND title LIKE '%\$title%', 最後得到SELECT * FROM movies INNER JOIN cast ON movies.id = cast.id WHERE 1=1 AND title LIKE '%\$title%' AND... AND...的query。之後就會列出符合條件的data。

Rating相關:

SELECT * FROM ratings WHERE userId = '\$userId' AND movieId = '\$movieId', 先找有沒有, 對於不同功能做不同的事。

更新:UPDATE ratings SET rating = '\$rating' WHERE userId = '\$userId' AND movieId = '\$movieId'

插入:INSERT INTO ratings (userId, movieId, rating, timestamp) VALUES ('\$userId', '\$movieId', '\$rating', '\$timestamp')

刪除:DELETE FROM ratings

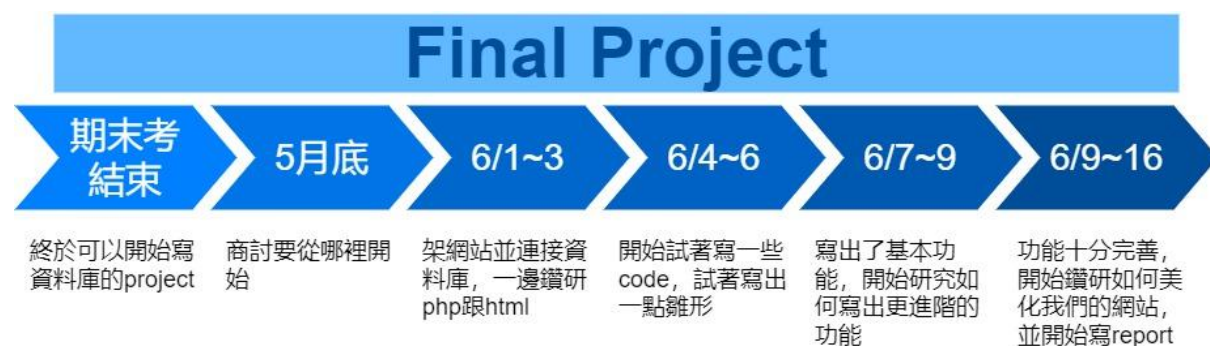
WHERE userId = '\$userId' AND movieId = '\$movieId'

之後再SELECT * FROM ratings WHERE userId = '\$userId'找出並輸出

● Others

- Draw the progress of your project

the expected progress:



the actual progress:



- What were the problems you met in the project, how did you solve them?

1. dataset的data出現問題怎麼辦

a. 使用python import pandas解決，如果還有問題，那就工人智慧

2. php語法都不會怎麼辦

a. 一言不和就google，從入門到入土

3. 做出來的成果跟想像中的成果有落差怎麼辦

a. 沒有任何事物是完美的，重要的不是結果，而是奮鬥的過程

- List the contribution of each team member in the project clearly

110550063 張博凱 33%

110550069 余駿佑 33%

110550116 廖梓程 33%

- Provide a link to your repository

<https://github.com/Duck997/2023-Spring-NYCU-Intro-To-DB>

- Provide a link to your discussion channel

都在Discord的私人群組中，如果TA想看我們可以邀你進來

