**IMDB Software of Hollywood Actresses and Actors**

Submission deadline: 13.01.2022

# **1. Description the project**

In this project was created a user-friendly software that stores and extracts information about the top 50 popular Hollywood actor and actresses from the webpage http://www.imdb.com/list/ls053501318/.

# **2. Modules, data structures, tools used in the project.**

Modules used in the project:

* requests –used to query a website.
* Beautiful soup – used to parse the data returned from the website.
* pandas – used for data manipulation and analysis.
* mysql and mysql.connector – used to make a connection to a MySQL server and work with MySQL database.
* json – used to find json-files in html script and to extract data from them.
* selenium and selenium.webdriver – used to open buttons "show more" in filmography from pages of actresses/actor.
* time – used to provide some delay when it is needed
* flask– used to launch webserver, show html-based interactive webpages.
* flaskext.mysql – used to make a connection to a MySQL server and work with MySQL database in flask framework.

Data Structures:

Web scraped data is collected in lists, then lists are used as columns of pandas dataframes.

For regular storing data about actors and actresses is used MySQL – an open-source relational database management system. Database contains four relational tables for storing actresses and actors (and information about them), films, awards, and genres. In structure of tables are used primary keys and foreign keys to link tables together.

From database data is read using SQL queries and all necessary computations such as find average rating of all films of actor are made by database system.

Tools:

As IDE for programming in Python PyCharm for Windows is used because it provides code analysis, a graphical debugger and convenient file management system, as there are used different types for this project.

User-friendly software is web-based using html scripts inside of the python code, so html or css files are not used.

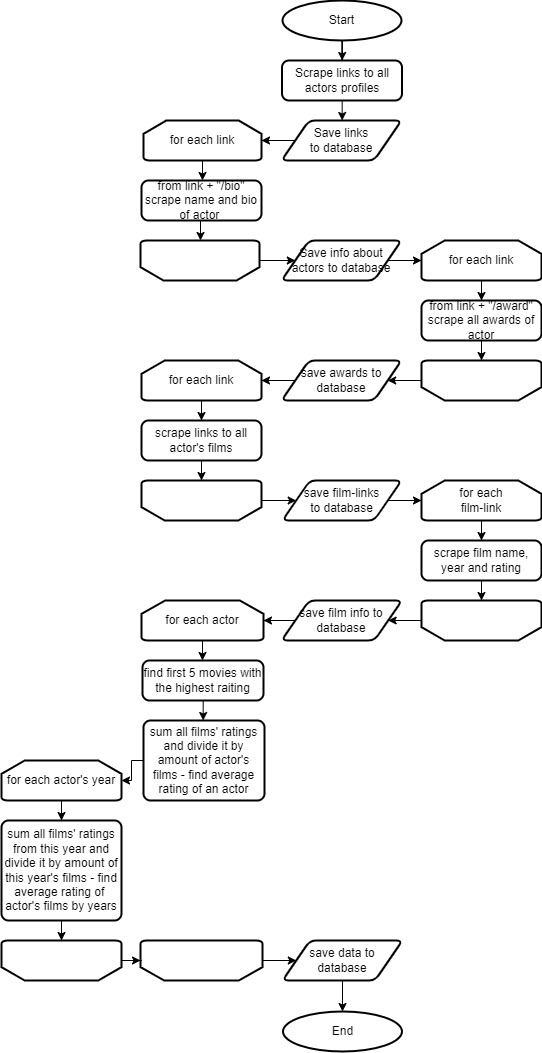
# **3. Modules of project**

Project has two .py files:

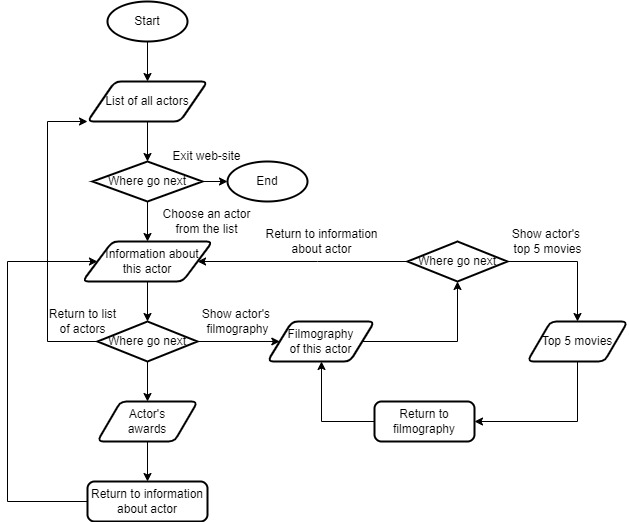
* webscraping.py – it is the script for webscraping and saving data from imdb website into the database.
* app.py – it is the web application script that should be executed from terminal by command “flask --app app run”

# **4. Flowchart of all problems addressed in the project**

Here is flowchart illustrating sequence of web-scraping steps, in which needed data was found and saved to database:



This is a flowchart of how users will be working with application.



# **5. Screenshots of the Graphical User Interface of the project**

This is the main page with all actresses/actors as links.



After clicking on some actor, in this example Matt Damon, personal profile will be opened with actor’s biography, top 3 genres, top 5 movies and links to filmography, awards and all actors page.

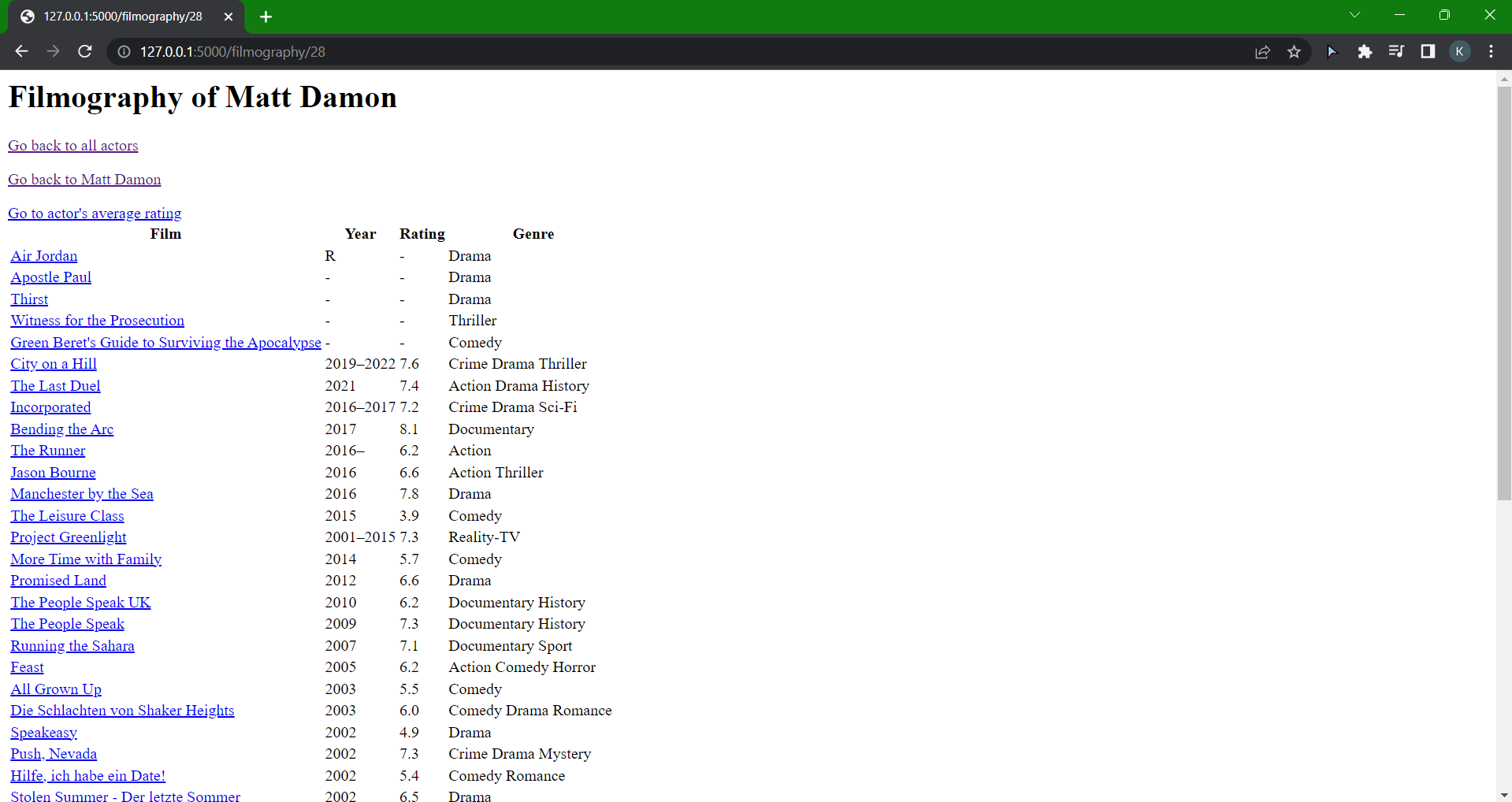
Зображення, що містить текст

Автоматично згенерований опис

Зображення, що містить текст

Автоматично згенерований опис

After clicking on filmography link the page with all actor’s films will appear. It also contains links to average ratings, back to actor and to all actors. If user clicks on the film, it will redirect to film’s imdb page.



After clicking on rating page, it will show average rating of all actor`s films and average rating of each year respectively.



If user goes back to actor`s personal page and clicks on awards, this page will appear.

Зображення, що містить текст

Автоматично згенерований опис