



OSTBAYERISCHE  
TECHNISCHE HOCHSCHULE  
REGENSBURG

**Name:**

Christoph Prenissl

**Email:**

christoph.prenissl@st.oth-regensburg.de

**Student ID:**

3174997

June 9, 2021

## **IMDB Software to fetch data of Hollywood Actors and Actresses**

**Task 4 Document - Project Report**

# Contents

<b>1</b>	<b>Project Description</b>	<b>2</b>
<b>2</b>	<b>Tools, Modules and Data-Structure</b>	<b>2</b>
<b>3</b>	<b>High-level Design</b>	<b>3</b>
<b>4</b>	<b>Functionalities</b>	<b>3</b>
4.1	List of all available actors and actresses . . . . .	3
4.2	About the actor/actresses . . . . .	3
4.3	All time movie names and years . . . . .	3
4.4	Awards to actor/actresses in different years . . . . .	3
4.5	Movie genre of actor/actresses . . . . .	3
4.6	Average rating of their movies . . . . .	3
4.7	Top 5 movies, their respective years and genre . . . . .	3

# 1 Project Description

This project mainly consists of creating a Python client to fetch data of the *IMDB Top 50 Actors and Actresses* list and also gather their movie data. The client uses an API to fetch all the basic actors/actresses list (4.1), the actor/actress About section (4.2) and all their movies (4.3). For Sections 4.4 - 4.7 Web-scraping is used to get awards data and ratings of the movies. The client is presented in a window based UserInterface where the user can click to gather the wanted information.

The further details of the specifics are handled in this document.

## 2 Tools, Modules and Data-Structure

For development of the client I used *Python 3.9.4* in *Visual Studio Code* with the *Python IntelliSense* extension which helped me in code completion and understanding the structure of all the frameworks and modules. For version control I used *Git* and the helpful VS Code extension *GitLens* which helped me to keep track of my changes in the project files.

When it comes to the modules, *Requests* and *BeautifulSoup 4* were necessary to handle all the web scraping. The http context is created with the module *SSL*. For most data I used a data frame created with *Pandas*.

The Graphical UserInterface was implemented using *PyQt6*. The framework also provides Thread libraries to help with multi-threading. The UI itself was designed with *Qt Designer*. It was very convenient to have a graphical UserInterface to drag and drop widgets and have an overall understanding of all elements.

For presenting the project I also used VS Code. I wrote the reports in  $\LaTeX$  with the help of the *LaTeX Workshop* extension and *PlantUml* to present core structures and flows of the client. In the presentation of the client I used *Powerpoint*.

### 3 High-level Design

#### 4 Functionalities

- 4.1 List of all available actors and actresses
- 4.2 About the actor/actresses
- 4.3 All time movie names and years
- 4.4 Awards to actor/actresses in different years
- 4.5 Movie genre of actor/actresses
- 4.6 Average rating of their movies
- 4.7 Top 5 movies, their respective years and genre