

Actor Filmography Tracker

Aim

Write a web scraping script that takes input as the name of any film actor and gives the output of the filmography of that actor in descending order.

Usage

Just run the code and enter the name of the desired actor. Just make sure that the name's spelling is perfectly correct and the first letter of First, Middle and Last name are capital and others are small since this code is case sensitive. After this if the data for that actor is available it will print all the movies by that actor otherwise will throw an exception.

Prerequisite

For running the given code directly the jupyter and python environment should be set up in the user's system. Along with this we need to first install the required python packages(Request package and BeautifulSoup package) and their dependencies in the current environment. And needless to say the system should be connected to internet and there shouldn't be any restriction on jupyter notebook by operating system or anti-virus from accessing the internet.

Code Structure

This project is developed using python and the code is shared in the form of a Jupyter Notebook. For data the wikipedia URL is used and the format for the URL is

https://en.wikipedia.org/wiki/actor_name_filmography#Film. Here actor_name is replaced by the name entered by the user.

After making this link a call is made using **request package** to get the details from page in the form of HTML code. If status 200 is achieved then we proceed to parse this data using **BeautifulSoup** package else an exception is raised.

Now according to the structure of web page relevant table of content is retrieved and by traversing through this table we get the desired details like movie name and release date which is stored in a set and then sorted in descending order to print the latest movies first.

Scope of improvement

Since the format of response from the wikipedia page is not response for all the requests, so for some searches exceptions may be triggered. For example the format for the page of some XYZ actor maybe a bit different than the one considered here so in those cases the code will throw an exception. For better results more reliable source could be used if available.