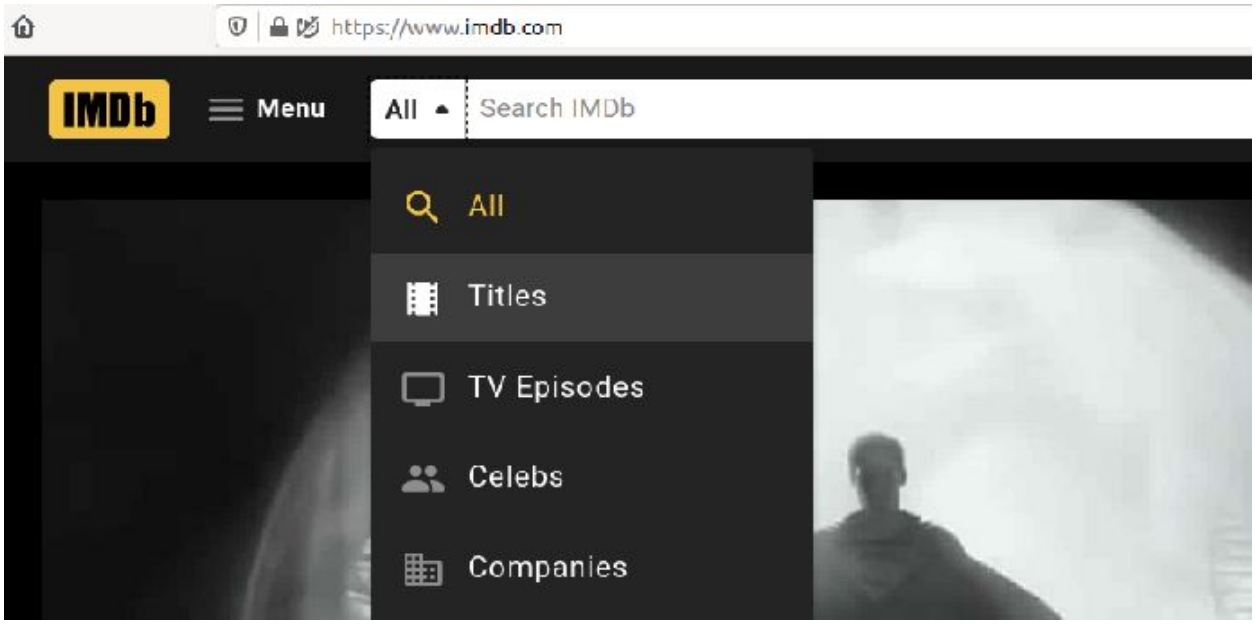
Experis Academy Advanced Java Course Preliminary Assignment

This take-home assignment is part of the selection process for the advanced server side Java course at Experis Academy.

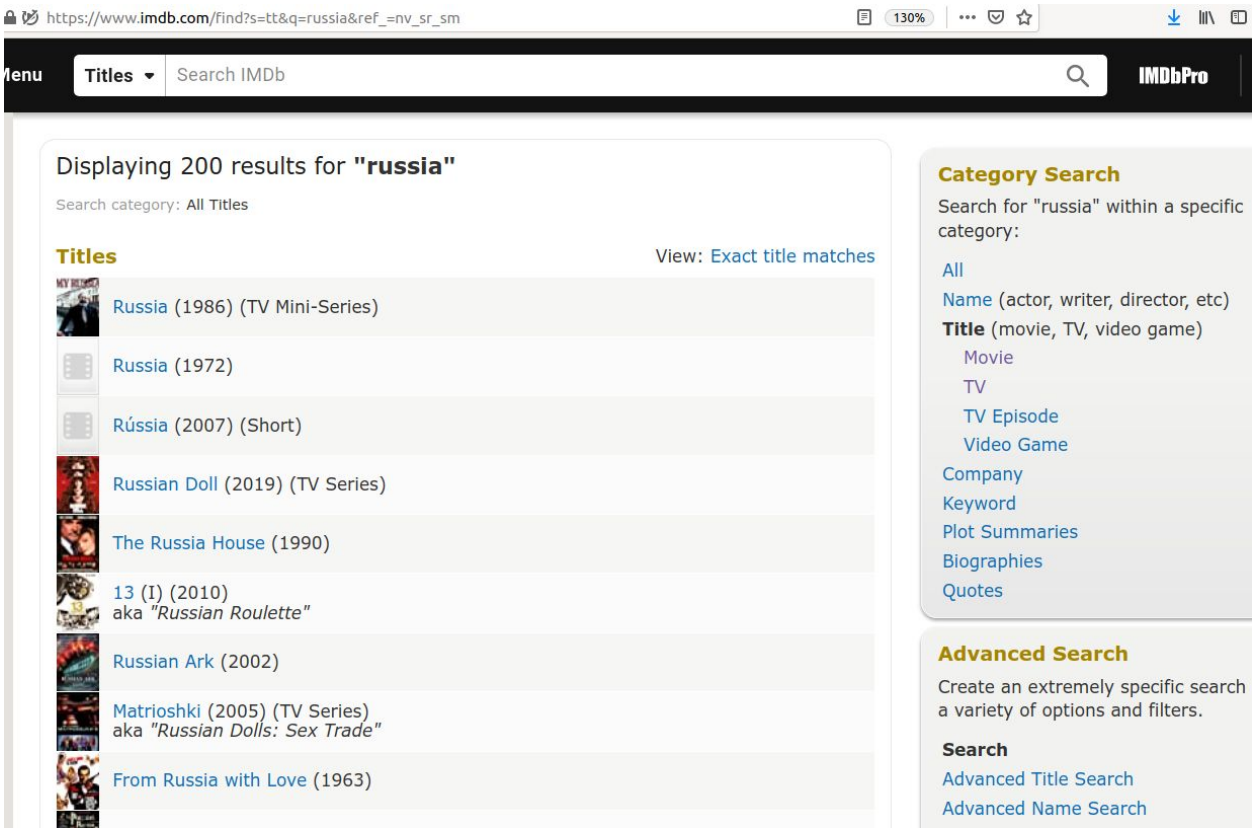
You are expected to develop and deliver the specified "*product*" in 2 days.

Introduction

A user can search for movies on the following page: https://www.imdb.com/ by choosing "**Titles**" from the drop down and entering a string:



Here is an example looking for all movies with word: **russia**



The search can be further narrowed to display only movies with titles containing the search string by selecting **"Movie"** in the category search box on the right.

Assignment

You are expected to develop a command line or a UI tool that can execute queries for movie titles using the web site https://www.imdb.com
The tool should use web scraping to accomplish this.
You are free to choose whatever implementation technology/programming language you prefer. We usually recommend:

1. Python
2. NodeJS/javascript.

But you can select top do it with Java, C# or anything else if you wish.

The tool developed will accept a search term and then produce a file with information about the movies with titles containing the search term.
The resulting file should contain one record for each movie, records shall be separated by a new line. Each record shall contain the following fields separated by vertical bar (|) :

1. Movie title
2. Genre : list of genre
3. MPAA rating. (R, PG-13, etc)

4. Movie duration
5. Director or list of directors if more than one (comma separated)
6. Star or list of stars (comma separated)

Example records:

- The Russia House|Drama, Romance, Thriller|R|2h 3min|Fred Schepisi|Sean Connery, Michelle Pfeiffer, Roy Scheider
- Avengers: Endgame|Action, Adventure, Drama|PG-13|3h 1min| Anthony Russo, Joe Russo|Robert Downey Jr., Chris Evans, Mark Ruffalo

Notes:

- Some movie titles are in the development phase. These should not be included in the resulting file. See The Lord Before the Rings

- Some Movies lack some information. Maybe not rated yet, or duration is unknown ,etc. If a field is missing you should output that as an empty field. See: Skull Commandos
- IMDb search facility is a fuzzy search:
    - Searching for *Elm* will also return:
        - *S**elm**a*
        - *Hot**el M**umbai*
        - *Ste**el M**agnolias*

      The second & third results are obviously not desired and should be filtered out an and not included in the output file.

Methodology

The simplest approach is to write a tool that can do *web scraping* using the imdb.com web site. Such a tool will simulate what a browser can do but will automate the process.

If you choose to use python to develop the tool, you can find many packages that can help you download HTML content from sites (urllib) and parse this content to extract information (Beautiful Soup).

If you choose JavaScript/NodeJS, then you can also find many equivalent packages: *Axios*, *Superagent*, and *Request packages to download HTML content and Cheerio for parsing and extracting data from it.*

Note: these are recommendations, feel free to substitute with other technologies/packages/languages as you see fit.

Deliverables

On completion of this assignment you will submit the following in a zip archive:

- README.txt file describing how to install and operate the too developed
- All Sources for the tool.
- A sample output file from a run of the tool searching for: "star trek"

Duration

You have 3 days to complete the task. All submitted tasks will be reviewed. We encourage you to submit your solution even if it's not perfect (Who is!?). Just make sure to describe the missing features or the unsolved bugs if any.

May the code be with you!