

---

## Coleta de Dados com o Selenium

### *Administrativo*

Este exercício pode ser feito em **duplas, trios ou quartetos**, não sendo possível realizar individualmente. Apenas um dos integrantes deve submeter os arquivos do trabalho no moodle.

Data de Entrega: [23/04/2024](#)

Forma de Entrega: O trabalho deve ser apresentado durante o período de aula. Não haverá uma aula de apresentação para este trabalho. O grupo deve marcar com o professor um horário para realizar a apresentação. A apresentação consiste em mostrar o código e o programa em funcionamento para o professor. Poderão ser feitas perguntas sobre o funcionamento do trabalho para o grupo.

Além disso, deve ser entregue no moodle:

- Scripts Python ou Jupyter notebooks com o código que realiza as tarefas solicitadas.
- Dados extraídos (um arquivo json)
- Orientações sobre como executar os scripts (como comentário no código, arquivo README.txt ou células de texto em jupyter notebook). Incluir comentários sobre a configuração do ambiente de desenvolvimento necessário para rodar os scripts (quando for necessário).

### *Tarefa 1 – Web Scraping com Selenium*

Considerando o site <https://www.imdb.com/>.

1. Faça scraping para obter os 250 séries com as maiores avaliações do IMDB. Devem ser obtidos: Título, Ano de estreia, Número Total de Episódios, Nota do IMDB e o link para página da série.
2. Faça scraping das páginas específicas dos 250 filmes obtidos no item anterior. Obtenha dessa página a popularidade e a listagem do elenco principal (incluindo nome do ator/atriz e da personagem).
3. Salve as informações obtidas em um arquivo json. Tome cuidado para aplicar as transformações necessárias nos dados extraídos (remover números do título, conversão de números inteiros e reais para os devidos tipos).