

DESAFIO WEBSCRAPING

Nosso objetivo não é avaliar o quanto você sabe sobre cada tópico, mas sim sua capacidade de resolver problemas e como você se comporta diante de desafios, mesmo sem ter todo o conhecimento necessário. Lembre-se que todos nós aprendemos com o tempo, e o maior acervo público de conhecimento que já existiu está ao seu alcance.

A tarefa proposta é desenvolver um código em Python para coletar informações de um site e extrair os dados para um arquivo CSV.

Sua tarefa é extrair o TÍTULO, DATA e RESUMO das 10 primeiras postagens do site <https://gizmodo.uol.com.br/> e armazená-los em um dicionário de chaves e valores. Existem diversas formas de coletar dados de sistemas, sejam diretas (com acesso ao banco de dados ou dataset) ou indiretas (com acesso apenas ao front-end), onde uma solução de usuário é necessária para a coleta de dados. Uma das formas mais comuns de realizar isso é através do Web Scraping.

Ferramentas Úteis:

- Para criar uma POC e testar seu código rapidamente, recomendamos a instalação do Anaconda e o uso do Jupyter Notebook para testes de código. O Anaconda é uma distribuição Python pronta para manipulação de dados. Acesse <https://www.anaconda.com/download/> para mais informações.
- O Selenium Webdriver é uma opção útil para realizar o Web Scraping. Após instalar o Anaconda, o Selenium pode ser instalado com o comando PIP (ex. pip install selenium). Consulte a documentação do Selenium Webdriver para Python ([Selenium Get Started](#)) para auxiliá-lo.

Pesquisas como "web scraping com python" e "web scraping com selenium" serão bastante úteis nesse processo. O recurso de inspecionar elementos do Google Chrome também pode ser de grande ajuda.

Lembre-se de que será necessário um arquivo chamado chromedriver.exe.

Salve o dicionário como um arquivo CSV utilizando o módulo pandas.

- O módulo Pandas do Python, já instalado com o Anaconda, permite que você organize os dados e os salve em um arquivo. Assim, você poderá criar um arquivo CSV a partir de um dicionário de chaves e valores com os dados capturados.

Crie uma conta no GitHub.com (caso ainda não tenha uma) e um novo repositório. Recomendamos a instalação do GitHub Desktop para facilitar o processo. Com ele, você pode realizar todas as ações necessárias através de uma interface gráfica, sem a necessidade de comandos de código.

Links úteis:

<https://blog.rocketseat.com.br/iniciando-com-git-github/>

<http://www.devfuria.com.br/git/tutorial-iniciando-git/>

Realize um commit e um push de todo o seu trabalho para o repositório do GitHub e envie-o para o recrutador.

NOTAS:

Não se preocupe em terminar tudo; o mais importante é deixar registrado todo o processo de desenvolvimento, pois tudo será avaliado.

Em caso de dúvidas, não hesite em perguntar ao recrutador.