

DESAFIO TÉCNICO

Seu desafio é automatizar o processo de extração de dados do site de notícias. Link para o site de notícias: www.nytimes.com

Você deve ter 3 variáveis configuradas (você pode salvá-las no arquivo de configuração, mas fique à vontade para fazer da maneira que achar melhor de acordo com as melhores práticas):

- Frase de pesquisa
- Categoria ou seção de notícias
- Número de meses durante os quais você precisa receber notícias

Exemplo de como isso deve funcionar: 0 ou 1 - apenas o mês atual, 2 - mês atual e anterior, 3 - mês atual e dois meses anteriores, e assim por diante

O projeto de função deve ser desenvolvido com base na arquitetura REFramework.

As principais etapas:

1. Abra o site seguindo o link
2. Insira uma frase no campo de pesquisa
3. Na página de resultados, aplique os seguintes filtros:

Ao selecionar uma categoria ou seção de notícias, sua automação deverá ter a opção de escolher entre nenhuma ou qualquer número de categorias/seções. Isso deve ser especificado através do arquivo de configuração ou/ou da forma que você achar melhor de acordo com as melhores práticas.

o escolha as notícias mais recentes (ou seja, as mais recentes).

4. Obtenha os valores: título, data e descrição.

5. Armazene em um arquivo Excel:

- título
- data
- descrição (se disponível)
- nome do arquivo da imagem
- contagem de frases de pesquisa no título e na descrição
- Verdadeiro ou Falso, dependendo se o título ou a descrição contém alguma quantia em dinheiro

Formatos possíveis: \$ 11,1 | US\$ 111.111,11 | 11 dólares | 11 dólares

6. Baixe a imagem da notícia e especifique o nome do arquivo no arquivo Excel

7. Siga as etapas 4 a 6 para todas as notícias que estiverem dentro do período exigido.

É mandatória a utilização de WSL2 (Ubuntu) e Docker para build da imagem e execução da automação.

Use a forma que achar melhor para disponibilizar o arquivo final, seja por volumes ou outro método 😊

Orientação: subir no git, e exportar o pacote.