**Removedor de HTML**

**Descrição**

O Removedor de HTML é um programa desenvolvido em Python que permite remover tags HTML de um arquivo Excel, produzindo uma versão limpa do texto sem formatação HTML.

**Requisitos**

Python 3.x

Bibliotecas Python:

pandas

BeautifulSoup

Tkinter

**Instalação**

Certifique-se de ter o Python instalado no seu computador. Se não tiver, você pode baixá-lo em python.org.

Instale as bibliotecas necessárias executando o seguinte comando no terminal:

Navegue até o diretório onde está o seu arquivo requirements.txt.

Execute o seguinte comando:

pip install -r requirements.txt

Isso instruirá o pip a ler o arquivo requirements.txt e instalar todas as dependências listadas nele automaticamente.

Não é necessário instalar a biblioteca tkinter, pois ela é uma biblioteca padrão do Python e já vem incluída na maioria das instalações.

**Utilização**

Execute o programa removedor\_de\_html.py usando o Python. Isso pode ser feito digitando o seguinte comando no terminal:

python removedor\_de\_html.py

Na janela que se abrir, você encontrará três botões:

Selecionar Arquivo Excel: Use este botão para escolher o arquivo Excel que deseja processar.

Selecionar Pasta de Destino: Use este botão para escolher a pasta onde deseja salvar o arquivo processado.

Remover HTML: Depois de selecionar o arquivo e a pasta de destino, clique neste botão para iniciar o processo de remoção de HTML.

Após clicar em "Remover HTML", o programa processará o arquivo Excel, removerá todas as tags HTML e salvará a versão limpa na pasta de destino especificada.

Uma mensagem de conclusão será exibida informando que o texto limpo foi salvo com sucesso.

**Exemplo**

Suponha que você tenha um arquivo Excel chamado dados.xlsx com conteúdo HTML em algumas células. Você deseja remover as tags HTML e salvar o texto limpo em um novo arquivo chamado texto\_limpo.xlsx. Para fazer isso:

**Execute o programa removedor\_de\_html.py.**

Clique no botão "Selecionar Arquivo Excel" e escolha o arquivo dados.xlsx.

Clique no botão "Selecionar Pasta de Destino" e escolha a pasta onde deseja salvar o arquivo processado.

Clique no botão "Remover HTML" para iniciar o processo de remoção de HTML.

Após a conclusão do processo, uma mensagem será exibida indicando que o texto limpo foi salvo com sucesso.

Notas

Este programa é útil para quem precisa extrair o texto de arquivos Excel que contenham formatação HTML, como páginas da web salvas em formato de tabela.

Certifique-se de ter permissão de escrita na pasta de destino selecionada, pois o programa salvará o arquivo processado nessa pasta.

Se o arquivo Excel for muito grande, o processo de remoção de HTML pode demorar um pouco dependendo do desempenho do seu computador.

Desenvolvedor: Gleison Amorim