

Documentação do Script "PDF Search"

Descrição:

O "PDF Search" é uma ferramenta que permite a busca em arquivos PDF por meio de palavras-chave. Ele utiliza técnicas de processamento de imagens e texto para buscar as palavras-chave nos arquivos PDF disponíveis no diretório especificado pelo usuário.

Bibliotecas Utilizadas:

tkinter: Utilizada para criar a interface gráfica.

filedialog, messagebox, scrolledtext: Módulos do tkinter utilizados para interação com o sistema de arquivos, exibição de mensagens e criação de uma área de texto rolável.

concurrent.futures.ThreadPoolExecutor: Utilizado para realizar operações em paralelo.

PIL (Python Imaging Library): Utilizada para processamento de imagens.

PyPDF2: Utilizada para extrair texto de arquivos PDF.

fitz: Utilizada para manipular arquivos PDF.

os, io: Utilizados para operações com o sistema de arquivos.

Instalação de Dependências Externas:

Para instalar as bibliotecas necessárias, execute o seguinte comando:

```
pip install pillow pytesseract PyPDF2 PyMuPDF
```

Variáveis Globais:

tesseract_path: Armazena o caminho para o executável do Tesseract OCR.

root_directory: Armazena o diretório raiz onde os arquivos PDF serão pesquisados.

config_file: Nome do arquivo de configuração (config.ini).

Principais Funções:

carregar_configuracoes(): Carrega as configurações previamente salvas do arquivo config.ini.

salvar_configuracoes(): Salva as configurações no arquivo config.ini.

selecionar_tesseract_path(): Abre uma janela para o usuário selecionar o executável do Tesseract OCR.

selecionar_root_directory(): Abre uma janela para o usuário selecionar o diretório raiz para pesquisa de PDFs.

confirmar_configuracoes(): Confirma as configurações definidas pelo usuário e esconde a seção de configuração.

buscar_pdfs(): Inicia a busca por PDFs com base na palavra-chave fornecida pelo usuário.

executar_busca(): Executa a busca de PDFs com base na palavra-chave e exibe os resultados na área de texto.

limitar_caracteres(evento): Limita o número de caracteres na entrada de pesquisa.

Principais Componentes da Interface:

Seção de Configuração: Permite ao usuário definir o caminho para o executável do Tesseract e o diretório raiz para pesquisa de PDFs.

Seção do PDF Search: Permite ao usuário realizar buscas por palavra-chave em arquivos PDF.

Entrada de Pesquisa: O usuário insere a palavra-chave desejada.

Botão "Buscar": Inicia a busca por PDFs com base na palavra-chave inserida.

Área de Texto Rolável: Exibe os resultados da busca, incluindo os nomes dos arquivos PDF encontrados.