

- [Extração de Dados de PDFs](#)
 - [Requisitos](#)
 - [Configuração](#)
 - [Funcionamento](#)
 - [Estrutura do Projeto](#)
 - [Dependências](#)
 - [Autor](#)

Extração de Dados de PDFs

Este projeto tem como objetivo extrair dados de arquivos PDF, inserir os dados extraídos em uma tabela MySQL e opcionalmente salvar os dados em arquivos Excel.

Requisitos

- Python 3.x
- MySQL
- Excel (opcional)

Configuração

1. **Instalar dependências:** Execute o comando abaixo para instalar as dependências do projeto:

```
pip install -r requirements.txt
```

2. **Configurar variáveis de ambiente:** Crie um ficheiro `.env` na raiz do projeto e adicione as variáveis de ambiente necessárias:

```
DB_HOST=localhost
DB_USER=root
DB_PASSWORD=123456
DB_NAME=pdf_extraction
DB_TABLE=cidadaos
```

3. **Executar o projeto:** Execute o comando abaixo para iniciar o projeto:

```
python main.py
```

4. **Logs:** Os logs do projeto são guardados no diretório **logs**.

Funcionamento

O projeto lê todos os ficheiros PDF do diretório disponibilizado pelo utilizador, extrai os dados e insere na tabela **DB_TABLE** do MySQL. Opcionalmente, os dados extraídos podem ser guardados em ficheiros excel no mesmo diretório do ficheiro pdf lido.

Estrutura do Projeto

- **main.py**: Ficheiro principal do projeto.
- **requirements.txt**: Ficheiro com as dependências do projeto.
- **.env**: Ficheiro para guardar as variáveis de ambiente.
- **.gitignore**: Ficheiro para ignorar elementos do projeto no git.
- **README.pdf**: Ficheiro com a documentação do projeto.

Dependências

- **pdfplumber**: Biblioteca para extração de dados de PDFs.
- **pandas**: Biblioteca para manipulação de dados.
- **mysql-connector-python**: Biblioteca para conexão com o MySQL.
- **python-dotenv**: Biblioteca para carregar variáveis de ambiente a partir de um ficheiro **.env**.
- **tqdm**: Biblioteca para mostrar barras de progresso.

Autor
