

```

import os

import PyPDF2

import re

from openpyxl import Workbook

# Define o caminho para a pasta com os arquivos PDF
pdf_folder_path = 'C:\\Users\\Pasta\\Areadetrabalho\\Nome_pasta'

# Cria uma lista para armazenar os caminhos dos arquivos PDF
pdf_files = []

# Percorre todos os arquivos da pasta e adiciona a lista apenas os arquivos com extensão .pdf
for filename in os.listdir(pdf_folder_path):
    if filename.endswith('.pdf'):
        pdf_files.append(os.path.join(pdf_folder_path, filename))

# Cria uma nova planilha do Excel
wb = Workbook()
ws = wb.active

# Adiciona cabeçalhos de coluna
headers = [
    'Número do Processo',
    'Nome',
    'CPF/CNPJ/Número Processo',
    'Cidade',
    'Estado',
    'E-mail',
]

ws.append(headers)

```

```

# Define as expressões regulares para buscar os dados
regex_num_processo = r'Número do Processo:\s*(.*)'
regex_nome = r'Nome:\s*(.*)'
regex_cpf_cnpj_site = r'CPF/CNPJ/Número site:\s*(.*)'
regex_cidade = r'Cidade:\s*(.*)'
regex_estado = r'Estado:\s*(.*)'
regex_email = r'e-mail:\s*(.*)'

# Percorre todos os arquivos PDF da lista
for pdf_file in pdf_files:
    # Abre o arquivo PDF em modo leitura binária
    with open(pdf_file, 'rb') as pdf:
        # Cria um objeto PDF Reader
        pdf_reader = PyPDF2.PdfReader(pdf)

        # Obtém o número de páginas do arquivo PDF
        num_pages = len(pdf_reader.pages)

        # Percorre todas as páginas do arquivo PDF
        for page_num in range(num_pages):
            # Obtém a página atual
            page = pdf_reader.pages[page_num]

            # Extrai o texto da página
            page_text = page.extract_text()

            # Busca os dados na página usando as expressões regulares
            num_processo = re.findall(regex_num_processo, page_text)
            nome = re.findall(regex_nome, page_text)
            cpf_cnpj_site = re.findall(regex_cpf_cnpj_site, page_text)

```

```
cidade = re.findall(regex_cidade, page_text)
estado = re.findall(regex_estado, page_text)
email = re.findall(regex_email, page_text)
```

```
# Adiciona os dados coletados como linhas na planilha do Excel
```

```
row = [
    num_processo[0] if num_processo else "",
    nome[0] if nome else "",
    cpf_cnpj_site[0] if cpf_cnpj_site else "",
    cidade[0] if cidade else "",
    estado[0] if estado else "",
    email[0] if email else "",
]
```

```
# Adiciona a linha de dados apenas se ela não estiver vazia
```

```
if any(row):
    ws.append(row)
```

```
# Salva a planilha do Excel no diretório correto
```

```
excel_file_path =
'C:\\Users\\Pasta\\Areadetrabalho\\Nome_do_Arquivo_Extractor\\dados_site.xlsx'
wb.save(excel_file_path)
```