

code_jpeg_pdfs

August 14, 2024

1. Transformar mapas “jpeg” em “pdf”

```
[ ]: !pip install Pillow
from google.colab import drive
drive.mount('/content/drive') #rodar 1 vez e autorizar conexão

[ ]: !ls "/content/drive/My Drive/caminho/" #conferir se a pasta de destino está
↪correta

[ ]: #transformando arquivos jpeg em pdf
from PIL import Image
import os

folder_path = "/content/drive/My Drive/caminho/JPEG/"

jpeg_files = [os.path.join(folder_path, f) for f in os.listdir(folder_path) if
↪f.endswith(".jpeg") or f.endswith(".jpg")]

output_folder = "/content/drive/My Drive/caminho/PDFs/"

os.makedirs(output_folder, exist_ok=True)

for jpeg_file in jpeg_files:
    image = Image.open(jpeg_file)
    pdf_path = os.path.join(output_folder, os.path.splitext(os.path.
↪basename(jpeg_file))[0] + '.pdf')
    image.convert('RGB').save(pdf_path, 'PDF', resolution=100.0)

print("Conversão concluída. PDFs salvos em:", output_folder)
```

2. Juntar PDFs.

Para isso, lembrar de mudar a pasta em cada código, antes de executar as células (executar apenas daqui pra baixo)

```
[ ]: #rodar apenas uma vez
!pip install PyPDF2
from google.colab import drive
drive.mount('/content/drive')
```

```
[ ]: #rodar apenas uma vez
import os
from PyPDF2 import PdfMerger

[ ]: mapas_folder = '/content/drive/My Drive/PDF/' #lembrar de colocar o endereço da
↳pasta de entrada
dossies_folder = '/content/drive/My Drive/PDFSJUNTADOS/' #lembrar de colocar o
↳endereço da pasta de saída

[ ]: #lista de arquivos pdf na pasta de entrada
mapas_files = {f.lower(): f for f in os.listdir(mapas_folder) if f.endswith(".
↳pdf")}
dossies_files = {f.lower(): f for f in os.listdir(dossies_folder) if f.
↳endswith(".pdf")}

#junta os arquivos pdf correspondentes ou mover os arquivos de entrada
for dossies_file_lower, dossies_file in dossies_files.items():
    dossies_path = os.path.join(dossies_folder, dossies_file)
    merger = PdfMerger()

#garantindo que há correspondência entre as pastas
    if dossies_file_lower in mapas_files:
        mapas_file = mapas_files[dossies_file_lower]
        mapas_path = os.path.join(mapas_folder, mapas_file)
        merger.append(dossies_path) #adiciona o pdf de entrada
        merger.append(mapas_path) #adiciona o pdf de saída
    else:
        merger.append(dossies_path) #apenas adiciona pdfs com correspondência
↳de nome entre as pastas

    output_path = os.path.join(dossies_folder, dossies_file)
    merger.write(output_path)
    merger.close()

print("Processamento concluído. PDFs salvos na pasta.")
```

3. Limpar pdfs (rodar esse código desfaz todo o processo, use-o em caso de algum erro e depois rode os códigos do item #2 novamente).

```
[ ]: import os
from PyPDF2 import PdfReader, PdfWriter

#pasta a ser formatada
pdf_folder = '/content/drive/My Drive/PDFSJUNTADOS/'

#lista de arquivos PDF na pasta
pdf_files = [f for f in os.listdir(pdf_folder) if f.endswith(".pdf")]
```

```

#remover a última página de PDFs com exatamente X páginas
for pdf_file in pdf_files:
    pdf_path = os.path.join(pdf_folder, pdf_file)
    reader = PdfReader(pdf_path)

    if len(reader.pages) == 4:
        writer = PdfWriter()
        for i in range(1):
            writer.add_page(reader.pages[i])

        #salvar o PDF modificado
        with open(pdf_path, "wb") as output_pdf:
            writer.write(output_pdf)

print("Processamento concluído. Última página removida dos PDFs com exatamente_
↪X páginas.")

```