

Mesclar arquivos PDF

Perguntei 12 anos, 5 meses atrás Modificado 28 dias atrás Visto 268 mil vezes



É possível, usando Python, mesclar arquivos PDF separados?

255

Supondo que sim, preciso estender isso um pouco mais. Espero percorrer as pastas em um diretório e repetir este procedimento.



E posso estar abusando da sorte, mas é possível excluir uma página que está contida em cada um dos PDFs (minha geração de relatório sempre cria uma página extra em branco).

Pitão pdf arquivo-io pypdf

Compartilhar Seguir

editado em 12 de outubro de 2021
às 1:37



Woody1193

6.524 4 39 75

perguntado em 9 de agosto de
2010 às 22h23



Btibert3

38k 44 127 167

15 respostas

Classificado por:

Pontuação mais alta (padrão)



Você pode usar a classe [PyPdf2](#) s [PdfMerger](#) .

419

Concatenação de arquivos



Você pode simplesmente [concatenar](#) arquivos usando o [append](#) método.



```
from PyPDF2 import PdfMerger

pdfs = ['file1.pdf', 'file2.pdf', 'file3.pdf', 'file4.pdf']

merger = PdfMerger()

for pdf in pdfs:
    merger.append(pdf)

merger.write("result.pdf")
merger.close()
```

Você pode passar identificadores de arquivo em vez de caminhos de arquivo, se desejar.

Mesclagem de arquivos

Se você deseja um controle mais refinado da mesclagem, existe um [merge](#) método do

Junte-se ao Stack Overflow para encontrar a melhor resposta para sua pergunta técnica, ajude outras pessoas a responderem às suas.

Sign up



```
for pdf in ['file1.pdf', 'file2.pdf', 'file3.pdf']:
    with fitz.open(pdf) as mfile:
        result.insertPDF(mfile)

result.save("result.pdf")
```

Com muitas opções, detalhadas no [wiki](#) de projetos .

Compartilhar Seguir

editado em 22 de maio de 2022 às 20:31

respondido em 21 de junho de 2016 às 13h12



Martin Thoma

118k 152 590 910



Paul Rooney

20,3k 9 41 61

- 5 O PyMuPDF foi muito mais rápido do que o PyPDF2 para o que eu tinha que fazer (Concatenar 280 PDFs de página única). Obrigado! – [Skusku](#) 7 de janeiro de 2022 às 11h34
 - 2 PyPDF2 agora também é mantido novamente :-). Acabei de vincular à sua resposta: pypdf2.readthedocs.io/en/latest/user/merging-pdfs.html – [Martin Thoma](#) 9 de abril de 2022 às 13h58
 - 1 pyPDF2 está funcionando bem, mas os links internos não estão funcionando, ou seja, não movendo para a seção especificada ao clicar. Qualquer ideia? Eu testei mesmo com um único arquivo, ainda o mesmo problema – [Rami](#) 11 de maio de 2022 às 17h20 ✎
- Eu nunca tentei isso. Este é apenas um processo básico de fusão, pode ser necessário entrar e corrigir os links internos depois. Fico feliz em resolver isso, se você puder fornecer mais detalhes. Você pode até considerar fazer uma nova pergunta? – [Paul Rooney](#) 11 de maio de 2022 às 22h47
- 1 @Skusku O mesmo para mim, tive que combinar páginas de diferentes documentos PDF em um "one-pager". Com PyPDF2 (usando `merge_page()` e transformações), isso levou 12 minutos (!), com PyMuPDF apenas 2 segundos usando a abordagem [aqui](#) . – [Splines](#) 16 de junho de 2022 às 13h22

Use [Py.pdf](#) ou seu sucessor [PyPDF2](#) :



153



Uma biblioteca Pure-Python construída como um kit de ferramentas PDF. É capaz de:

- dividir documentos página por página,
- mesclando documentos página por página,

(e muito mais)

Aqui está um exemplo de programa que funciona com ambas as versões.

```
#!/usr/bin/env python
import sys
try:
    from PyPDF2 import PdfFileReader, PdfFileWriter
except ImportError:
    from pyPdf import PdfFileReader, PdfFileWriter

def pdf_cat(input_files, output_stream):
```

Junte-se ao Stack Overflow para encontrar a melhor resposta para sua pergunta técnica, ajude outras pessoas a responderem às suas.

Sign up



```
with open("result.pdf", "wb") as fout:
    merger.write(fout)
```

How would I make the same code above today

```
from glob import glob
from PyPDF2 import PdfMerger

def pdf_merge():
    '''Merges all the pdf files in current directory'''
    merger = PdfMerger()
    allpdfs = [a for a in glob("*.pdf")]
    [merger.append(pdf) for pdf in allpdfs]
    with open("Merged_pdfs.pdf", "wb") as new_file:
        merger.write(new_file)

if __name__ == "__main__":
    pdf_merge()
```

Share Follow

edited Jan 4 at 10:24

answered Nov 17, 2017 at 17:40



PythonProgrammi

21.6k 3 39 34

1 I used this successfully – [Merlin](#) Apr 8, 2021 at 21:42

pyPDF2 is promising, but internal links are not working. Any idea? I tested even with single file, still same issue – [Rami](#) May 11, 2022 at 17:20 ✎

2 As of December 2022, I found that PdfFileMerger was deprecated when I tried running the code in the second part. You need to replace PdfFileMerger with PdfMerger and it'll work just fine – [R41nMak3R](#) Dec 27, 2022 at 12:33



15



The [pdfcrow library](#) can do this quite easily, assuming you don't need to preserve bookmarks and annotations, and your PDFs aren't encrypted. [cat.py](#) is an example concatenation script, and [subset.py](#) is an example page subsetting script.

The relevant part of the concatenation script -- assumes `inputs` is a list of input filenames, and `outfn` is an output file name:

```
from pdfcrow import PdfReader, PdfWriter

writer = PdfWriter()
for inpf in inputs:
    writer.addpages(PdfReader(inpf).pages)
writer.write(outfn)
```

Junte-se ao Stack Overflow para encontrar a melhor resposta para sua pergunta técnica, ajude outras pessoas a responderem às suas.

Sign up

