

Merging multiple PDFs into a single PDF using a Python script

This is one off-post, irrelevant to my blog's main focus.

Merging multiple PDFs into a single document is one activity which most of us have to do. Almost on a daily basis or on a weekly or monthly basis. There are of course many websites which offer this as a service. The ones which allow you to merge PDFs for free often have some limits. Either based on number of files or the time between every merging operation.

In this article, I am presenting two different methods for merging many PDF files into a single document. Using the Python tool kit – PyPDF2.

Before we go further, I emphasize that there is no “one-method-fits-all” approach. And I do not claim that these methods are the best. These are two methods that have worked fine for me so far. So, I thought that I would share it in this platform.

Prerequisites before you try either of these methods:

Make sure that

1. You have installed latest version of Python (that's obvious, duh!)
2. You have installed the PyPDF2 tool kit
3. Saved the PDF files that you want to merge in Python's working directory. Of course, you can change the directory using Python code. For simplicity of code, I am placing the PDF files on the working directory for these two methods that I am going to present here..

Method 1:

This method is directly taken from Chapter 13 of the book “Automate the Boring Stuff with Python” by Al Sweigart.

When is method 1 suitable?

1. When you have lesser number of files
2. When the group of files to be merged do not have a common filename pattern

How this method works?

In the following sequence.

1. Import the PyPDF2 tool kit which has the tools that we need for playing with PDFs
2. Open each and every file by entering the file name
3. Read each and every file which was opened in Step 2 using PdfFileReader

4. Create a blank PDF file using PdfFileWriter where you can store the merged output
5. Loop through every page in every file which was read in Step 3 using for loop and copy all the information
6. Give a name for the output file and then paste all the copied information in Step 5
7. Close all the files

If you find the above sequence difficult to understand, have a look at the code below. Python is very reader-friendly. So I hope you would get the idea.

```
1  import PyPDF2
2
3  # Open the files that have to be merged one by one
4  pdf1File = open('FirstInputFile.pdf', 'rb')
5  pdf2File = open('SecondInputFile.pdf', 'rb')
6
7  # Read the files that you have opened
8
9  pdf1Reader = PyPDF2.PdfFileReader(pdf1File)
10 pdf2Reader = PyPDF2.PdfFileReader(pdf2File)
11
12 # Create a new PdfFileWriter object which represents
13 pdfWriter = PyPDF2.PdfFileWriter()
14
15 # Loop through all the pagenumbers for the first doc
16 for pageNum in range(pdf1Reader.numPages):
17     pageObj = pdf1Reader.getPage(pageNum)
18     pdfWriter.addPage(pageObj)
19
20 # Loop through all the pagenumbers for the second doc
21 for pageNum in range(pdf2Reader.numPages):
22     pageObj = pdf2Reader.getPage(pageNum)
23     pdfWriter.addPage(pageObj)
24
25 # Now that you have copied all the pages in both the
26 pdfOutputFile = open('MergedFiles.pdf', 'wb')
27 pdfWriter.write(pdfOutputFile)
28
29 # Close all the files - Created as well as opened
30 pdfOutputFile.close()
31 pdf1File.close()
32 pdf2File.close()
```

Method 2:

This method is more elegant and has just 5 lines of code. It's my favorite and it uses the PdfFileMerger module.

When is method 2 suitable?

1. When you have a lot of PDF files (I mean a looooot – Like for example, hundreds of PDF files or even more)
2. If all the PDF files that you want to merge follow a naming convention for their file names.

How this method works?

In the following sequence.

1. Import PdfFileMerger and PdfFileReader tools
2. Loop through all the files that have to be merged and append them
3. Write the appended files into an output document and specify a name for it.

That's it. It's simple but powerful.

So let's look into the code now. Before we go there, I will show how my input files are named. And remember that these files are placed in Python's working directory.