<https://github.com/jaraco/PDF>

PDF

===

A Pure-Python library built as a PDF toolkit. It is capable of:

- extracting document information (title, author, ...),

- splitting documents page by page,

- merging documents page by page,

- cropping pages,

- merging multiple pages into a single page,

- encrypting and decrypting PDF files.

By being Pure-Python, it should run on any Python platform without any

dependencies on external libraries. It can also work entirely on StringIO

objects rather than file streams, allowing for PDF manipulation in memory.

It is therefore a useful tool for websites that manage or manipulate PDFs.

Examples

========

Example 1::

from PyPDF2 import PdfFileWriter, PdfFileReader

output = PdfFileWriter()

input1 = PdfFileReader(open("document1.pdf", "rb"))

# print how many pages input1 has:

print "document1.pdf has %d pages." % input1.getNumPages()

# add page 1 from input1 to output document, unchanged

output.addPage(input1.getPage(0))

# add page 2 from input1, but rotated clockwise 90 degrees

output.addPage(input1.getPage(1).rotateClockwise(90))

# add page 3 from input1, rotated the other way:

output.addPage(input1.getPage(2).rotateCounterClockwise(90))

# alt: output.addPage(input1.getPage(2).rotateClockwise(270))

# add page 4 from input1, but first add a watermark from another PDF:

page4 = input1.getPage(3)

watermark = PdfFileReader(open("watermark.pdf", "rb"))

page4.mergePage(watermark.getPage(0))

output.addPage(page4)

# add page 5 from input1, but crop it to half size:

page5 = input1.getPage(4)

page5.mediaBox.upperRight = (

page5.mediaBox.getUpperRight\_x() / 2,

page5.mediaBox.getUpperRight\_y() / 2

)

output.addPage(page5)

# encrypt your new PDF and add a password

password = "secret"

output.encrypt(password)

# finally, write "output" to document-output.pdf

outputStream = file("document-output.pdf", "wb")

output.write(outputStream)

Example 2::

from PyPDF2 import PdfFileReader, PdfFileMerger

merger = PdfFileMerger()

input1 = open("document1.pdf", "rb")

input2 = open("document2.pdf", "rb")

input3 = open("document3.pdf", "rb")

# add the first 3 pages of input1 document to output

merger.append(fileobj = input1, pages = (0,3))

# insert the first page of input2 into the output beginning after the second page

merger.merge(position = 2, fileobj = input2, pages = (0,1))

# append entire input3 document to the end of the output document

merger.append(input3)

# Write to an output PDF document

output = open("document-output.pdf", "wb")

merger.write(output)