

## 18. Metadata en PDF

Tuesday, October 27, 2020 8:43 AM

### Módulo pyPDF2

Una biblioteca creada como un kit de herramientas PDF que es capaz de:

- Extraer información del documento
- dividir documentos página por página
- fusión de documentos página por página
- páginas de recorte
- fusionar varias páginas en una sola página
- cifrar y descifrar archivos PDF

Entre otras cosas

#### Leer documentos PDF

```
from PyPDF2 import PdfFileReader

pdfFileObj = open('Manual.pdf', 'rb')
pdf = PdfFileReader(pdfFileObj)
print("Páginas: ", pdf.getNumPages())
print("Título: ", pdf.documentInfo.title)
primera_hoja = pdf.getPage(0)
print(primera_hoja.extractText())
```

### ReportLab

Es una librería de Python que permite generar documentos PDF y gráficas.

#### Escribir archivos PDF

```
from reportlab.pdfgen.canvas import Canvas
from reportlab.lib.pagesizes import LETTER

canvas = Canvas("hello.pdf", pagesize=LETTER)
canvas.drawString(72, 72, "Hello, World")
canvas.save()
```

Sin embargo, para la creación de PDF enriquecidos y con formato de calidad se recomienda aprovechar las herramientas con las que ya cuenta el sistema operativo o, en su defecto, LaTeX.

#### Leer Metadata de archivos PDF

```
from PyPDF2 import PdfFileReader, PdfFileWriter
import os

def printMeta():
    ruta = input("Ruta de imágenes: ")
    os.chdir(ruta)
    for root, dirs, files in os.walk(".", topdown=False):
        for name in files:
            ext = name.lower().rsplit('.', 1)[-1]
            if ext in ['pdf']:
                print("[+] Metadata for file: %s " % (ruta+os.path.sep+name))
                pdfFile = PdfFileReader(open(ruta+os.path.sep+name, 'rb'))
                docInfo = pdfFile.getDocumentInfo()
                print("Tipo: ", type(docInfo))
                for metaItem in docInfo:
                    print('[+] ' + metaItem + ':' + docInfo[metaItem])
                print("\n")

printMeta()
```

Referencia:

- <https://pypi.org/project/PyPDF2/>
- <https://pypi.org/project/reportlab/>
- <https://realpython.com/creating-modifying-pdf/>