

## Biblioteki do obsługi plików PDF w Pythonie

### Wprowadzenie

W Pythonie istnieje wiele bibliotek umożliwiających pracę z plikami PDF. Poniżej opisano najpopularniejsze biblioteki, ich funkcjonalności oraz przykłady użycia.

### 1. PyPDF2

PyPDF2 to popularna biblioteka służąca do manipulacji plikami PDF, takimi jak odczyt, scalanie, dzielenie czy wyodrębnianie tekstu.

**\*\*Instalacja:\*\***

```
pip install PyPDF2
```

**\*\*Przykład użycia:\*\***

```
from PyPDF2 import PdfReader, PdfWriter

# Odczyt pliku PDF
reader = PdfReader("plik.pdf")
for page in reader.pages:
    print(page.extract_text())

# Tworzenie nowego pliku PDF z wybranymi stronami
writer = PdfWriter()
writer.add_page(reader.pages[0]) # Dodaj pierwszą stronę
with open("nowy_plik.pdf", "wb") as f:
    writer.write(f)
```

### 2. ReportLab

ReportLab to biblioteka służąca do generowania plików PDF od podstaw, umożliwiająca dodawanie tekstu, obrazów i wykresów.

**\*\*Instalacja:\*\***

```
pip install reportlab
```

**\*\*Przykład użycia:\*\***

```
from reportlab.pdfgen import canvas

# Tworzenie nowego pliku PDF
c = canvas.Canvas("example.pdf")
c.drawString(100, 750, "Hello, PDF!") # Dodanie tekstu na stronie
c.save()
```

### 3. pdfplumber

pdfplumber to biblioteka pozwalająca na zaawansowane wyodrębnianie tekstu i tabel z plików PDF.

**\*\*Instalacja:\*\***

```
pip install pdfplumber
```

**\*\*Przykład użycia:\*\***

```
import pdfplumber

with pdfplumber.open("plik.pdf") as pdf:
    for page in pdf.pages:
        print(page.extract_text())
```

### 4. FPDF

FPDF to lekka biblioteka do tworzenia plików PDF, idealna do prostych zadań, takich jak dodawanie tekstu, obrazów czy tabel.

**\*\*Instalacja:\*\***

```
pip install fpdf
```

**\*\*Przykład użycia:\*\***

```
from fpdf import FPDF

pdf = FPDF()
pdf.add_page()
pdf.set_font("Arial", size=12)
pdf.cell(200, 10, txt="Hello, PDF!", ln=True, align="C")
pdf.output("example.pdf")
```

### 5. PyMuPDF (fitz)

PyMuPDF to zaawansowana biblioteka do analizy i manipulacji plików PDF, umożliwiająca wyodrębnianie tekstu, obrazów i metadanych.

**\*\*Instalacja:\*\***

```
pip install pymupdf
```

**\*\*Przykład użycia:\*\***

```
import fitz  # PyMuPDF

# Otwieranie pliku PDF
```

```
pdf = fitz.open("plik.pdf")
for page in pdf:
    print(page.get_text())
```

## Podsumowanie

Biblioteki do obsługi plików PDF w Pythonie różnią się funkcjonalnością. PyPDF2 i PyMuPDF są świetne do analizy i manipulacji istniejących plików PDF, podczas gdy ReportLab i FPDF służą do generowania nowych dokumentów PDF.