

## E40 – Metadata de archivos pdf y docx

**Objetivo del ejercicio:** Usar Python para extraer y mostrar metadata básica como autor, título, fecha de creación y modificación.

### Requisitos

Instala estas librerías antes de ejecutar el script:

```
pip install python-docx PyPDF2
```

### Instrucciones

1. Crea una carpeta llamada `archivos_prueba` y coloca dentro algunos archivos `.pdf` y `.docx`.
2. Guarda el siguiente script como `extraer_metadata.py` en la misma carpeta donde está `archivos_prueba`.
3. Ejecuta el script:

```
python extraer_metadata.py
```

4. Observa en consola la metadata extraída de cada archivo.
5. Toma capturas de pantalla mostrando la ejecución y los resultados.

### Código completo (`extraer_metadata.py`)

```
import os
from docx import Document
from PyPDF2 import PdfReader

def metadata_docx(path):
    doc = Document(path)
    props = doc.core_properties
    return {
        "tipo": "DOCX",
        "autor": props.author,
        "título": props.title,
        "creado": props.created,
        "modificado": props.modified
    }

def metadata_pdf(path):
    pdf = PdfReader(path)
    info = pdf.metadata
    return {
        "tipo": "PDF",
        "autor": info.author,
        "título": info.title,
        "creado": info.creation_date,
        "modificado": info.modification_date
    }

# Carpeta a analizar
CARPETA = "archivos_prueba"

print(f"📁 Analizando carpeta: {CARPETA}\n")
for archivo in os.listdir(CARPETA):
    ruta = os.path.join(CARPETA, archivo)
    if archivo.lower().endswith(".docx"):
        meta = metadata_docx(ruta)
```

```
elif archivo.lower().endswith(".pdf"):\n    meta = metadata_pdf(ruta)\nelse:\n    continue\nprint(f"📄 Archivo: {archivo}")\nfor clave, valor in meta.items():\n    print(f"    {clave}: {valor}")\nprint("-" * 40)
```



## Entregables

- Capturas de pantalla mostrando:
  - La ejecución del script en consola.
  - Los resultados de metadata por archivo.