**BuscarFechasPDF**

**Importaciones**

python

import re

from PyPDF2 import PdfReader

* import re: Importa el módulo de expresiones regulares, que se utiliza para buscar patrones en cadenas de texto.
* from PyPDF2 import PdfReader: Importa la clase PdfReader del módulo PyPDF2, que se utiliza para leer archivos PDF.

**Función para Extraer Fechas**

python

def extraer\_fechas(pdf\_path):

# Abrir el archivo PDF

lector = PdfReader(pdf\_path)

texto\_completo = ""

# Extraer texto de todas las páginas

for pagina in lector.pages:

texto\_completo += pagina.extract\_text()

# Buscar fechas con formato dd/mm/yy o similar

patron\_fecha = r'\b\d{1,2} -\d{1,2} -\d{2,2}\b'

fechas = re.findall(patron\_fecha, texto\_completo)

return fechas

**Paso a Paso:**

1. **Definición de la Función**: La función extraer\_fechas toma un argumento pdf\_path, que es la ruta del archivo PDF.
2. **Lectura del PDF**: Utiliza PdfReader para abrir el archivo PDF especificado por pdf\_path y crea un objeto lector.
3. **Inicialización de Texto Completo**: Define una variable texto\_completo como una cadena vacía, que se usará para acumular el texto extraído de todas las páginas del PDF.
4. **Extracción de Texto**:
   * Recorre todas las páginas del PDF usando for pagina in lector.pages.
   * Usa pagina.extract\_text() para extraer el texto de cada página y lo agrega a texto\_completo.
5. **Definición del Patrón de Fecha**:
   * Define un patrón de expresión regular patron\_fecha para buscar fechas con el formato dd-mm-yy o similar. En este caso, r'\b\d{1,2} -\d{1,2} -\d{2,2}\b' busca números que coincidan con este formato.
6. **Búsqueda de Fechas**:
   * Usa re.findall(patron\_fecha, texto\_completo) para encontrar todas las coincidencias del patrón de fechas en el texto completo extraído del PDF y las almacena en una lista fechas.
7. **Devolver Fechas**: La función retorna la lista de fechas encontradas.

**Solicitar la Ruta del PDF**

python

ruta\_pdf = input("Introduce la ruta del archivo PDF: ")

* Solicita al usuario que introduzca la ruta del archivo PDF y guarda la entrada en la variable ruta\_pdf.

**Bloque try-except**

python

try:

fechas\_encontradas = extraer\_fechas(ruta\_pdf)

if fechas\_encontradas:

print("Fechas encontradas en el documento:")

for fecha in fechas\_encontradas:

print(fecha)

else:

print("No se encontraron fechas en el documento.")

except Exception as e:

print(f"Ocurrió un error: {e}")

**Paso a Paso:**

1. try **Bloque**:
   * Intenta llamar a la función extraer\_fechas con ruta\_pdf.
   * Si se encuentran fechas, las imprime una por una.
   * Si no se encuentran fechas, imprime un mensaje indicando que no se encontraron fechas.
2. except **Bloque**:
   * Captura cualquier excepción que ocurra durante la ejecución del bloque try.
   * Imprime un mensaje de error con la descripción de la excepción e.

**Posibles Excepciones**

1. FileNotFoundError: Si el archivo PDF no se encuentra en la ruta especificada.
2. PermissionError: Si no se tienen permisos para abrir el archivo PDF.
3. PdfReadError: Si hay un problema al leer el archivo PDF (archivo corrupto).
4. re.error: Si hay un problema con la expresión regular.
5. **Otras Excepciones**: Cualquier otro error inesperado que pueda ocurrir.