PDF Search & Summary Tool using Python

# Problem Statement

Searching for specific keywords across multiple PDF documents manually is a time-consuming and tedious task. This project solves that problem by allowing the user to search through a folder full of PDFs for any keyword, and then generates a report showing where that keyword was found.

# Tools / Libraries Used

- Python: Programming language used for development  
- pdfplumber : Used to extract text from PDF documents  
- os: Used to navigate files and folders

# Solution Explanation

The tool is a Python script that:  
1. Takes the folder path containing PDF files.  
2. Asks the user to enter a keyword.  
3. Reads each page of every PDF file in the folder.  
4. Searches for the entered keyword line-by-line.  
5. When found, records the file name, page number, and sentence into a report file named 'summary\_report.txt'.

# Features

- User can search for any word inside all PDF files.  
- Searches every page of each file in the selected folder.  
- Output is saved neatly in a .txt report file.  
- Works well with study notes, academic papers, and more.

# Sample Output

PDF File | Page Number | Line with Keyword  
----------------------------------------------------  
DBMSNOTES(UNIT-2).pdf | Page 3 | Attributes define the properties of an entity in a database.  
DBMS\_UT-3\_.pdf | Page 2 | Key attributes uniquely identify a record.

# Folder Structure

gdk-pdf\  
├── pdf\_search\_tool.py -- Python script  
├── summary\_report.txt -- Output report file  
├── DBMSNOTES(UNIT-2).pdf --Sample PDF  
├── DBMS\_UT-3\_.pdf --Sample PDF  
├── DBMS(UT-5).pdf --Sample PDF

# Conclusion

This tool makes it easy to search for important terms in large collections of PDFs, saving time and effort. It is especially helpful for students, researchers, and professionals dealing with lots of documents.

Done by : Jabu Meghana

GKB Internship Program – July 2025