

PDF EXTRACTOR

Introduction:

All of you must be familiar with what PDFs are. In-fact, they are one of the most important and widely used digital media. PDF stands for **Portable Document Format**. It uses **.pdf** extension. It is used to present and exchange documents reliably, independent of software, hardware, or operating system.

Invented by **Adobe**, PDF is now an open standard maintained by the International Organization for Standardization (ISO). PDFs can contain links and buttons, form fields, audio, video, and business logic.

The project PDF Extractor will help people to extract PDF file into simple text even the Snapshot of any hand written page can be extracted which may help in getting it converted into text document for further use and also this application can be used for rotating ,merging, splitting the PDF as well .

This is designed in one language Python as front-end and the back-end.

Existing System:

The existing system lack proper GUI (Graphical User Interface) and that system can only be executed by trained programmer or the person who have knowledge of programming , no other user can able to use the system.

Proposed System:

The proposed system is having many advantages over the existing system. It requires less overhead and is very efficient. The new proposed PDF EXTRACTOR system will eradicate the issues that they are facing now and this software is created in such a manner that it can be used by anyone , even a person who have just a little knowledge of working of computer. It also gives us a path to wherever we want to save to file which lacks in the previous system as well as the option to select the desired file.

Salient features of the proposed System:

- To be able to extract text from PDF files.
- To be able to extract text from image as well.
- To make searching a keyword easier .
- The system has been developed using Python. So that all features are added in this system. At the time of handling this system, it allows multi-user environment facility.
- The burden of manual work is reduced because of the we can now able to use search even in the snapshot of any image in the PDF file.
- The system is user friendly and built in Window environment.

Requirement Analysis and Specification:**End User(s):**

There is only one end user for this system which is administrator.

Module(s):

1. Target PDF File Explorer
2. Converted PDF File Explorer
3. Selecting the number of pages
4. Selecting between a Text PDF or Snapshot PDF
5. Drop box for more selections
6. Opening the path of converted file

Technology Used:

3

Programming Language: Python

Front-End Tool: PyCharm 19.2

GUI Tool: Appjar 2.4

Operating System: Windows 10

Approval/Comments of the Mentor:

