Hello today is beautiful

thon – How To Extract Text From PDF

Before proceeding to main topic of this post, i will explain you some use cases where these type of PDF extraction required.

- One example is, you are using job portal where people used to upload their CV in PDF format. And when the recruiters researching for some kind of keywords like say a recruiters needs Hadoop developers, big data developers, python developers, java developers etc. So the keyword will be get matched with the skills what you have specified in the resume. This is again a processing so they extract data from your PDF document and they will matched with the keyword what the recruiter is searching for and then they will just give you your name, email or all those stuffs. So this is the use case.

Python provides many modules for PDF extraction but here we will see **PyPDF2** module. So let'e see how to extract text from PDF using this module.

PDF To Text Python – Extraction Text Using PyPDF2 module

- [PyPDF2](#) is a Pure-Python library built as a PDF toolkit. It is capable of:
  - extracting document information (title, author, …)
  - splitting documents page by page
  - merging documents page by page
  - cropping pages
  - merging multiple pages into a single page
  - encrypting and decrypting PDF files
  - and more!

So now we will see how to extract text from PDF using PyPDF2 module. Write the following code on your python IDE(check [best python IDEs](#)).

Installing PyPDF2

Run the following command on terminal to install PyPDF2.

```
pip install PyPDF2
```

Now you have to import PyPDF2 module. So write the following code.

```
import PyPDF2
```

Creating a PDF File Object