**Documentation**

Hello welcome to my project. This project is about analyzing the content in a pdf format question paper and printing the difficulty of the question paper based on its bloom level.

Here I have used few modules and they are:

1.Pypdf – pypdf is a open source python external module which is capable of doing operations on a pdf, I have imported pdfReader module which is inside Pypdf module which is capable of reading the contents in a PDF

2. Termcolor – I have imported colored module which is inside termcolor. Colored is used to change the color of the program output It is used to make the appearance of the program output better.

In this program I have used green color to print if the question paper is easy, Yellow if the question paper is medium difficult and red If the question paper is hard.

At the very first step I have taken a question paper as sample and it is in written PDF form using pdfReader module I have read the contents of the PDF file and written the contents of the file into another file which contains the whole question paper.

Now the problem is the whole question paper is written as it is into a text file, Now we need to analyze the data and selectively select the data which we need first I observed the sequence of the question paper which is set in our SSIT college

At the very first in Header the college name and subject name and code with duration of the exam would be given but we don’t need that data

The question starts with an integer number, The question and its bloom level is what we need to extract the questions from whole question paper I have again used File handling in python and wrote the complete line which starts with an integer. To find I had to typecast the first character in each line to integer if it successfully converted into a string the I wrote the content into another file.

But another problem arises here we don’t need question number and question our main concentration is on blooms level of each question for that again I used the sequence which is usually followed. In first column the marks(M) of each question would be printed after the course outcome(C) would be printed and in the end bloom level (B) would be printed.

Later after extracting the questions again using file handling I extracted only integer values which would be question number, marks of each question, course outcome and blooms level.

Now using the above file I added each bloom level and I found out the percentage of the bloom level and based on the percentage I’m printing the difficulty of the question paper given.

Disadvantages:

* The PDF has to be written converted PDF
* If the PDF is image converted this program would not work