**Conversion of PDF to image:**

**Link reference:** <https://www.print-driver.com/stories/convert-pdf-to-jpg-via-the-command-line>

1. Download 2JPEG from the below URL:

<https://www.2jpeg.com/download>

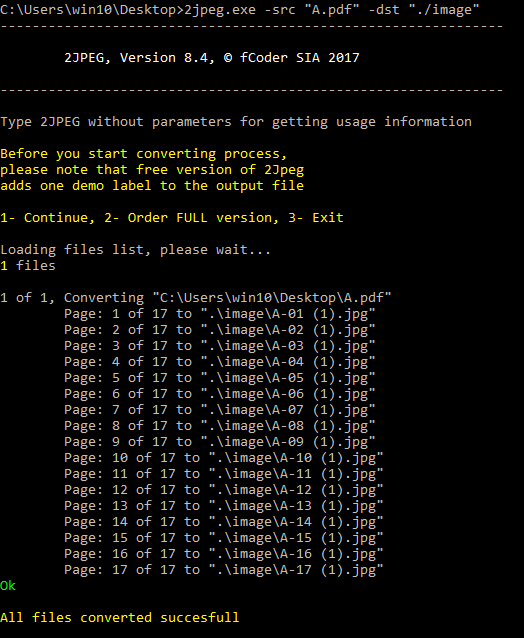
1. If the exe run requires admin rights, go to the below repo and download zip:

<https://github.com/SubhenduShekhar/2JPEG>

1. Unzip it to any location and set the path to environment variable.
2. Write the following command to generate the images:

2jpeg.exe -src "Source/File/Name.pdf" -dst "Destination/Folder/Path"

Example: 2jpeg.exe -src "A.pdf" -dst "./image"



1. Not all files will be generated perfectly. But we can get the details from other files also.

Extracting text from image:

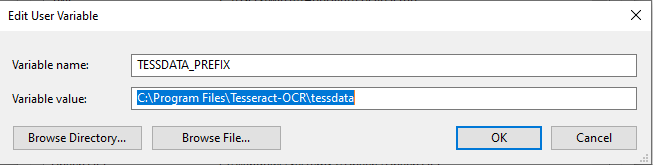
1. pip install pytesseract
2. Download the tesseract for python from the below repository:

<https://github.com/SubhenduShekhar/TesseractForPython>

1. Extact the zip to any location and set the environment variable as shown below:

Variable Name: TESSDATA\_PREFIX

Variable Value: <Path of Tesseract-OCR>\tessdata



1. Save the variable.
2. Write the following code to extract the text from file(say A-1.jpg):

import pytesseract

pytesseract.pytesseract.tesseract\_cmd = r'C:\Program Files\Tesseract-OCR\tesseract'

a = pytesseract.image\_to\_string(r'C:\Users\win10\Desktop\A-1.jpg')

f = open('textA1.txt','w')

f.write(a)

f.close()

You will get the text as in the image.