**Extracting Text from Scanned Documents**

# Boto3 & AWS Textract

# 

# 

# 

# 

# 

# 

import tkinter as tk

from tkinter import filedialog #to browse any image file from folders of my desktop or system

from tkinter.filedialog import askopenfile

from PIL import Image, ImageTk

import boto3

my\_w=tk.Tk()

my\_w.geometry("450x400")

my\_w.title("AWS-Textract")

my\_font1=('times',18, 'bold')

l1=tk.Label(my\_w,text="Upload an Image",width=30,font=my\_font1)

l1.pack()

b1=tk.Button(my\_w,text='Upload File & See what it has!!!',width=30,command= lambda:upload\_file())

b1.pack()

def upload\_file():

    aws\_mag\_con=boto3.session.Session(profile\_name='demo\_user')

    boto3.client=aws\_mag\_con.client(service\_name='textract', region\_name='us-east-1')

    global img

    f\_types= [('Jpg Files', '\*.jpg')]

    filename=filedialog.askopenfilename(filetype=f\_types)

    img= Image.open(filename)

    # resizing

    img\_resized=img.resize((400,200))

    img=ImageTk.PhotoImage(img\_resized)

    imgbytes=get\_image\_byte(filename)

    b2=tk.Button(my\_w,image=img)

    b2.pack()

    response=boto3.client.detect\_document\_text(Document={'Bytes':imgbytes})

    for item in response['Blocks']:

         if item['BlockType']=='LINE':

              print(item['Text'])

def get\_image\_byte(filename):

    with open(filename, 'rb') as imgfile:

        return imgfile.read()

my\_w.mainloop()

# 

# 