Name - Suyash Karpe

Roll Number - 114

Topic - OCR(Mini Project)

In [17]:

```
import os
import glob
import pytesseract
import re
import pandas as pd


class UserInput():
    '''This class take a input from user about which file information they want Aadhar Card or pancard.'''
    def welcome_note(self):
        print('Welcome to OCR - Here you convert your image into text.')
        print()
    def user_input(self):
        print('Press "1" for csv file of Aadhar Card information.')
        print('Press "2" for csv file of Pan Card information.')
        print('Press "3" for csv file of Both card information.')
        print()

        self.choice = int(input('Enter your choice here = '))
        return self.choice

## Class for extracting data from aadhar card and pancard images :-

class Ocr():

    def Aadhar_card(self):
        '''This Method extract the data from aadhar card images.'''

        tesseract_path = r'C:\Program Files\Tesseract-OCR\tesseract.exe'
        pytesseract.pytesseract.tesseract_cmd = tesseract_path
        path = r'C:\Users\shrir\Desktop\velocity\python practice\My_Assignment\adhar_card_imges\**'
        aadhar_file_list = glob.glob(path + '**.jpg')

        self.aadhar_file_name = []
        pattern = '[\w]{2,15}[.][jpg]{3}'
        for i in  aadhar_file_list:
            new_string = re.search(pattern,i)
            self.aadhar_file_name.append(new_string.group())

        self.adhar_number = []
        self.adhar_dob = []

        for i in  aadhar_file_list:
            text = pytesseract.image_to_string(i)
            num = re.findall('\d{3,4}\s\d{3,4}\s\d{3,4}',text)
            self.adhar_number.append(num)
            dob = re.findall('\d{1,2}[-/]\d{1,2}[-/]\d{4}',text)
            self.adhar_dob.append(dob)
        return self.adhar_number, self.adhar_dob

    def Pan_card(self):
        '''This Method extract the data from pan card images.'''

        tesseract_path = r'C:\Program Files\Tesseract-OCR\tesseract.exe'
        pytesseract.pytesseract.tesseract_cmd = tesseract_path
        path = r'C:\Users\shrir\Desktop\velocity\python practice\My_Assignment\pan_card\**'
        pan_list = glob.glob(path + '**.jpg')

        self.new_pan_list = []
        pattern = '[\w]{2,15}[.][jpg]{3}'
        for i in pan_list:
            new_pan_string = re.search(pattern,i)
            self.new_pan_list.append(new_pan_string.group())

        self.pan_number = []
        self.pan_dob = []

        for i in pan_list:
            pan_text = pytesseract.image_to_string(i)
            num = re.findall('[A-Z]{5}\d{4}[A-Z]{1}',pan_text)
            self.pan_number.append(num)
            dob = re.findall('\d{1,2}[-/]\d{1,2}[-/]\d{4}',pan_text)
            self.pan_dob.append(dob)
        return self.pan_number,self.pan_dob

    def printing_dict(self):
        pan_card_dict = {'File Name' : self.new_pan_list,
                         'Pan Number' : self.pan_number,
                         'Date of Birth' : self.pan_dob }


        aadhar_card_dict = {'File Name' :self.aadhar_file_name,
                       'Aadhar Number' : self.adhar_number,
                       'Date of Birth' : self.adhar_dob }

        return pan_card_dict,aadhar_card_dict

    def methods_calling(self):
```

```
        self.Aadhar_card()
        self.Pan_card()
        return self.printing_dict()

## Class for converting information to csv files :-

class CsvFile():
    '''This class convert our information into CSV files.'''
    def __init__(self,adhar_dict,pan_dict):
        self.aadhar_info = adhar_dict
        self.pan_info = pan_dict
    def converting(self):
        df = pd.DataFrame(self.aadhar_info)
        df.to_csv('aadhar.csv',index = False)

        df = pd.DataFrame(self.pan_info)
        df.to_csv('pan.csv',index = False)

        return pd.read_csv("aadhar.csv"),pd.read_csv("pan.csv")


## Making object of class :-

intro = UserInput()
intro.welcome_note()
user_choice = intro.user_input()
obj = Ocr()
value = obj.methods_calling()
obj1 = CsvFile(value[0],value[1])
file = obj1.converting()

## printing of information according to user input :-


if user_choice == 1:
    print(f'Aadhar card information in the form of dictionary = \n\n{value[1]}')
    print()
    print(f'Content in Csv file of Aadhar card = \n\n{file[1]}' )
elif user_choice == 2:
    print(f'Pan card information in the form of dictionary = \n\n{value[0]}')
    print()
    print(f'Content in Csv file of Pan card = \n\n{file[0]}' )
else:
    print(f'Aadhar card information in the form of dictionary = \n\n{value[1]}')
    print()
    print(f'Content in Csv file of Aadhar card = \n\n{file[1]}' )
    print()
    print()
    print(f'Pan card information in the form of dictionary = \n\n{value[0]}')
    print()
    print(f'Content in Csv file of Pan card = \n\n{file[0]}' )

```

```
Welcome to OCR - Here you convert your image into text.

Press "1" for csv file of Aadhar Card information.
Press "2" for csv file of Pan Card information.
Press "3" for csv file of Both card information.

Enter your choice here = 3
Aadhar card information in the form of dictionary =

{'File Name': ['abhinav.jpg', 'akesh.jpg', 'alam1.jpg', 'anchal1.jpg', 'gomati1.jpg', 'harindra.jpg', 'kanika.jpg',
'karona.jpg', 'manoj1.jpg', 'maulik.jpg', 'muhammad1.jpg', 'bai.jpg', 'noor.jpg', 'radha.jpg', 'rajeev.jpg', 'shubha
m.jpg', 'sid.jpg', 'subharta.jpg', 'vangla.jpg'], 'Aadhar Number': [['557 2998 5205'], ['4421 9689 9245'], ['6123 27
55 2779'], ['2786 7508 6425'], [], ['5505 8787 4595'], ['2114 5270 9955'], ['9091 2919 3929'], ['4967 7765 4463'],
['8774 6726 6338'], [], [], ['9502 1517 6995'], [], [], ['4275 2080 8052'], ['3425 0653 1151'], ['9150 6575 7100'],
['6344 0324 0334']], 'Date of Birth': [['16/03/1991'], ['16/04/1995'], ['11/07/2009'], ['08/03/2012'], ['01/05/199
3'], [], ['11/09/1993'], ['01/01/2011'], ['04/08/1996'], [], ['10/10/1998'], ['05/07/1990'], [], ['05-06-1965'], [],
['05/03/1997'], ['28/05/2000'], ['02/03/2003'], ['01/01/1984']]}

Content in Csv file of Aadhar card =

        File Name       Aadhar Number   Date of Birth
0     abhinav.jpg   ['557 2998 5205']  ['16/03/1991']
1       akesh.jpg  ['4421 9689 9245']  ['16/04/1995']
2       alam1.jpg  ['6123 2755 2779']  ['11/07/2009']
3     anchal1.jpg  ['2786 7508 6425']  ['08/03/2012']
4     gomati1.jpg                  []  ['01/05/1993']
5    harindra.jpg  ['5505 8787 4595']              []
6      kanika.jpg  ['2114 5270 9955']  ['11/09/1993']
7      karona.jpg  ['9091 2919 3929']  ['01/01/2011']
8      manoj1.jpg  ['4967 7765 4463']  ['04/08/1996']
9      maulik.jpg  ['8774 6726 6338']              []
10  muhammad1.jpg                  []  ['10/10/1998']
11        bai.jpg                  []  ['05/07/1990']
12       noor.jpg  ['9502 1517 6995']              []
13      radha.jpg                  []  ['05-06-1965']
```

```
14     rajeev.jpg                  []              []
15    shubham.jpg  ['4275 2080 8052']  ['05/03/1997']
16        sid.jpg  ['3425 0653 1151']  ['28/05/2000']
17   subharta.jpg  ['9150 6575 7100']  ['02/03/2003']
18     vangla.jpg  ['6344 0324 0334']  ['01/01/1984']
```

Pan card information in the form of dictionary =

{'File Name': ['Aditya.jpg', 'Amit.jpg', 'Arjun_durga.jpg', 'Asharaf.jpg', 'DuraiSwami.jpg', 'Ketkee.jpg', 'Maharana
pratap.jpg', 'Monika.jpg', 'Naresh_das.jpg', 'Pramod.jpg', 'Prasanth.jpg', 'Premshankar.jpg', 'Shri1.jpg', 'Suresh.j
pg', 'Tanknath1.jpg', 'Vikas.jpg'], 'Pan Number': [['BODPM4264E'], ['CHIPS1802F'], ['CHKPD8490A'], ['BYOPA2085A'],
['BNZPM2501F'], ['CYIPS8165C'], ['BVEPP1809C'], ['EJAPS0276M'], [], ['ANRPM2537J'], ['AAICP8972D'], ['BJDPP6011M'],
[], ['AZHPN8387P'], ['BLAPT0864M'], ['BIOPS4195M']], 'Date of Birth': [['02/06/1976'], ['03/07/1983'], ['02/07/199
0'], ['10/10/1995'], ['16/07/1986'], ['04/12/1982'], ['10/07/1990'], ['31/10/1992'], ['09/12/1997'], ['03/04/1982'],
['05/10/2016'], ['09/07/1986'], ['21/03/2009'], ['24/10/1992'], ['01/01/1965'], []]}

Content in Csv file of Pan card =

```
             File Name      Pan Number   Date of Birth
0           Aditya.jpg  ['BODPM4264E']  ['02/06/1976']
1             Amit.jpg  ['CHIPS1802F']  ['03/07/1983']
2      Arjun_durga.jpg  ['CHKPD8490A']  ['02/07/1990']
3          Asharaf.jpg  ['BYOPA2085A']  ['10/10/1995']
4       DuraiSwami.jpg  ['BNZPM2501F']  ['16/07/1986']
5           Ketkee.jpg  ['CYIPS8165C']  ['04/12/1982']
6   Maharanapratap.jpg  ['BVEPP1809C']  ['10/07/1990']
7           Monika.jpg  ['EJAPS0276M']  ['31/10/1992']
8       Naresh_das.jpg              []  ['09/12/1997']
9           Pramod.jpg  ['ANRPM2537J']  ['03/04/1982']
10        Prasanth.jpg  ['AAICP8972D']  ['05/10/2016']
11     Premshankar.jpg  ['BJDPP6011M']  ['09/07/1986']
12           Shri1.jpg              []  ['21/03/2009']
13          Suresh.jpg  ['AZHPN8387P']  ['24/10/1992']
14       Tanknath1.jpg  ['BLAPT0864M']  ['01/01/1965']
15           Vikas.jpg  ['BIOPS4195M']              []
```

In [ ]:
```
1
```