**AIM: Extract text from an image using Optical Character Recognition (OCR) with Tesseract and perform text summarization.**

**DESCRIPTION :** Optical Character Recognition (OCR) is a technology that extracts text from images or scanned documents. Tesseract is a popular open-source OCR engine. To use Tesseract in Python, you can use the pytesseract wrapper along with the Pillow library for image processing. Before starting, you need to install Tesseract on your system. You can download the installer from the official Tesseract GitHub page or use a package manager like apt or brew on Linux or macOS. After installing Tesseract, you also need to install the required Python libraries.

**SOFTWARES TO INSTALL :**

https://github.com/UB-Mannheim/tesseract/wiki

CLICK ON tesseract-ocr-w64-setup-5.3.3.20231005.exe (64 bit)

Download Software and Install

**COMMAND PROMPT INSTALL :**

pip install pytesseract

pip install Pillow

**CODING :**

from PIL import Image

import pytesseract

# Set the path to the Tesseract executable (update this based on your installation)

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

def perform\_ocr(image\_path):

# Open the image using Pillow

image = Image.open(image\_path)

# Perform OCR using Tesseract

text = pytesseract.image\_to\_string(image)

return text

if \_\_name\_\_ == "\_\_main\_\_":

# Replace 'path/to/your/image.png' with the actual path to your image file

image\_path = r'C:\SUREKHA/News.png'

result = perform\_ocr(image\_path)

print("OCR Result:")

print(result)

**OUTPUT :**

Save below file with name News.png and give location path in code

image\_path = r'C:\SUREKHA/News.png'

****

C:\Users\civilsys51\PycharmProjects\pythonProject\.venv\Scripts\python.exe C:\Users\civilsys51\PycharmProjects\pythonProject\Tesser.py

OCR Result:

\_DAILY \_NEWS\_

NEWS |

Process finished with exit code 0

**RESULT :**

Thus, we sextracted text from an image using Optical Character Recognition (OCR) with Tesseract and perform text summarization.