

Convertir una imagen a texto en Python (usando la biblioteca Pytesseract)

Para convertir una imagen a texto en Python, puedes usar la biblioteca Pytesseract, que es una envoltura de Python para el motor OCR (Reconocimiento Óptico de Caracteres) Tesseract, desarrollado por Google. Este motor extrae texto de imágenes y lo convierte en texto.

Pasos a seguir:

1. **Instalar Tesseract OCR:**
2. **Instalar las dependencias de Python:**
 - Instala pytesseract y Pillow (para manejar imágenes) usando pip:
`pip install pytesseract pillow`

Código para convertir una imagen a texto:

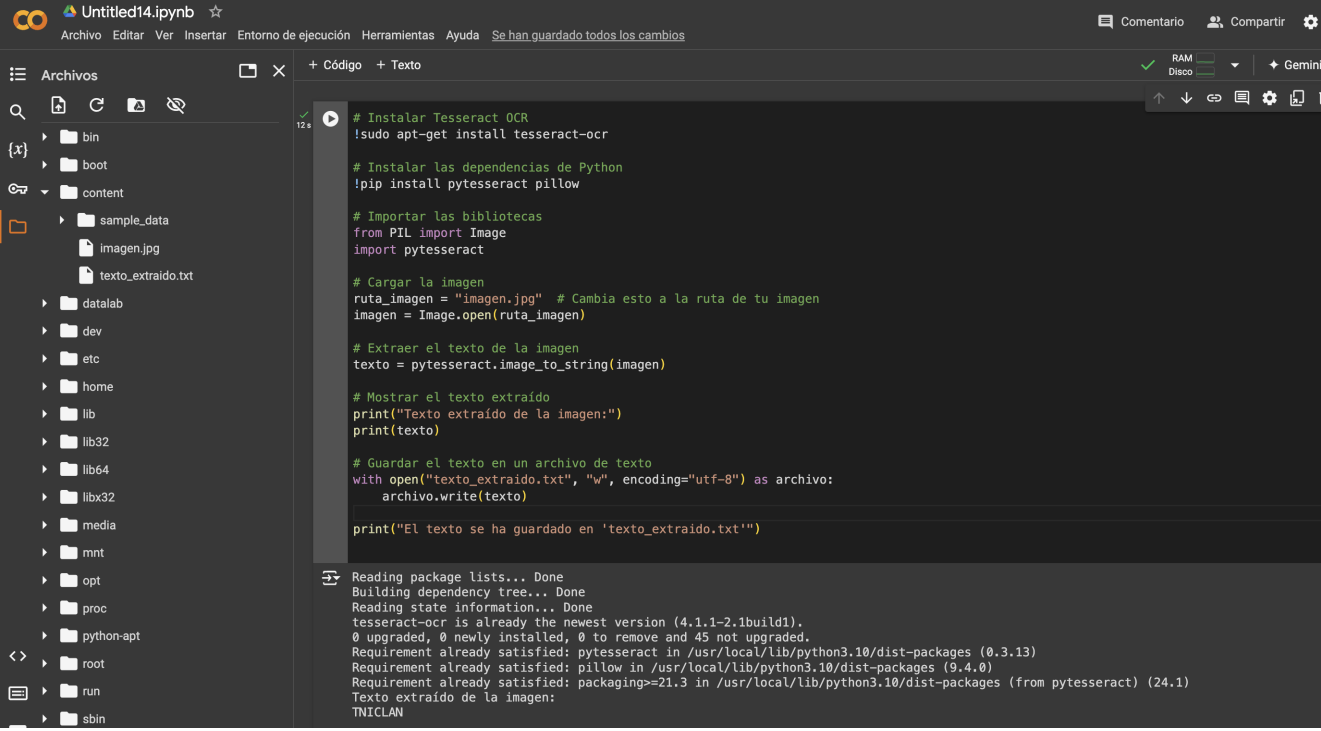
[crayon-6a9d4efef32b9969421833/]

Explicación del código:

1. **Importar las bibliotecas:**
2. **Especificar la ruta de Tesseract (opcional):**
3. **Cargar la imagen:**
4. **Extraer el texto:**
5. **Guardar el texto:**

Ejemplo de uso:

TNCLXY



```
# Instalar Tesseract OCR
!sudo apt-get install tesseract-ocr

# Instalar las dependencias de Python
!pip install pytesseract pillow

# Importar las bibliotecas
from PIL import Image
import pytesseract

# Cargar la imagen
ruta_imagen = "imagen.jpg" # Cambia esto a la ruta de tu imagen
imagen = Image.open(ruta_imagen)

# Extraer el texto de la imagen
texto = pytesseract.image_to_string(imagen)

# Mostrar el texto extraído
print("Texto extraído de la imagen:")
print(texto)

# Guardar el texto en un archivo de texto
with open("texto_extraido.txt", "w", encoding="utf-8") as archivo:
    archivo.write(texto)

print("El texto se ha guardado en 'texto_extraido.txt'")
```

Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
tesseract-ocr is already the newest version (4.1.1-2.1build1).
0 upgraded, 0 newly installed, 0 to remove and 45 not upgraded.
Requirement already satisfied: pytesseract in /usr/local/lib/python3.10/dist-packages (0.3.13)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (9.4.0)
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.10/dist-packages (from pytesseract) (24.1)
Texto extraído de la imagen:
TNCLAN