

# 텍스트 추출하기

## 사전 설치

```
pip install opencv-python
pip install pytesseract
```

```
In [4]: import cv2
import pytesseract

# 이미지 파일 경로 설정
image_path = 'googlesearch.png'

# 이미지 파일을 읽어옴
image = cv2.imread(image_path)

# 이미지를 그레이스케일로 변환
gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
```

```
In [5]: # 이미지에 있는 텍스트 추출
text = pytesseract.image_to_string(gray_image, lang='eng')
```

## 만약 아래와 같은 에러 메시지 발생시

TesseractNotFoundError: tesseract is not installed or it's not in your PATH. See README file for more information.

## 해결

- 01 테서렉트 다운로드 후, 설치
  - <https://github.com/UB-Mannheim/tesseract/wiki>
  - 2023/05 기준 다음 파일 설치
    - tesseract-ocr-w64-setup-5.3.1.20230401.exe (64 bit)
- 02 윈도우 환경변수 지정.
  - 시스템 환경 변수 편집 검색
  - 고급 - 환경변수 선택 - Path 선택 - 편집 버튼 선택
  - 새로 만들기 - 설치된 경로 등록
    - C:\Program Files\Tesseract-OCR

```
In [6]: # 추출된 텍스트 출력
print(text)
```

© pyimagesearch.com  
https://pyimagesearch.com > Blog

Installing Tesseract, PyTesseract, and Python OCR packages ...  
2021. 8. 16. - In this tutorial, you will: Learn how to install the Tesseract OCR engine on your machine; Learn how to create a Python virtual environment ...

x nanonets.com  
https://nanonets.com > blog > ocr-with-tesseract }

Tesseract OCR in Python with Pytesseract and OpenCV  
2023. 2. 27. - An in-depth tutorial on using Tesseract OCR, Pytesseract & OpenCV ...  
You can install the python wrapper for tesseract after this using pip.

projectpro.io  
https://www.projectpro.io » recipes » what-is-pytessera..  
What is the pytesseract python library and How do you install it  
2022. 8. 3. - 1. Install tesseract using windows installer available at: https://github.com/UB-Mannheim/tesseract/wiki - 2. Note the tesseract path from the ...

- 참고할 만한 URL
  - <https://ddolcat.tistory.com/954>