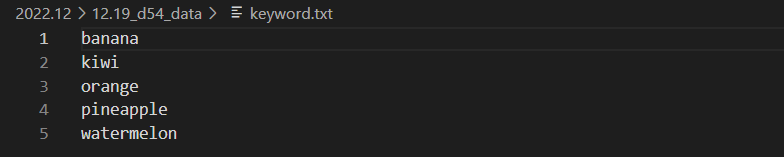
**ex01. 데이터 수집 (Web Crawling)**

- keyword.txt



- Code

import os, time

import pandas as pd

import urllib.request

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

from multiprocessing import Pool

# 파일 경로

folder\_path = './2022.12/12.18\_d54\_data'

chromedriver\_path = './2022.12/chromedriver.exe'

# 검색 키워드 호출

key = pd.read\_csv(folder\_path+'/keyword.txt', encoding= 'utf-8', names =['keyword'])

keyword = []

[keyword.append(key['keyword'][x]) for x in range(len(key))]

def create\_folder(directory):   # 폴더 생성

    try:

        if not os.path.exists(directory):

            os.makedirs(directory)

    except OSError:

        print("error : Creating directory ... " + directory)

def image\_download(keywords):

    create\_folder(folder\_path + '/' + keywords )

    # 크롬 드라이브 호출

    options = webdriver.ChromeOptions()

    options.add\_experimental\_option("detach", True)

    driver = webdriver.Chrome(chromedriver\_path, options= options)

    driver.implicitly\_wait(3)

    # 검색

    print('검색 >> ', keywords)

    driver.get("https://www.google.co.kr/imghp?h1=en")

    keyword = driver.find\_element\_by\_xpath(

        '/html/body/div[1]/div[3]/form/div[1]/div[1]/div[1]/div/div[2]/input')

    keyword.send\_keys(keywords)

    keyword.send\_keys(Keys.RETURN)

    # 스크롤 내리기 -> 결과 더보기 버튼 클릭

    elem = driver.find\_element\_by\_tag\_name('body')

    N = 20

    for i in range(N):

        elem.send\_keys(Keys.PAGE\_DOWN)

        time.sleep(0.2)

    try:

        driver.find\_element\_by\_xpath(

            '//\*[@id="islmp"]/div/div/div/div[2]/div[1]/div[2]/div[2]/input').click()

        for i in range(N):

            elem.send\_keys(Keys.PAGE\_DOWN)

            time.sleep(0.2)

    except:

        pass

    images = driver.find\_elements\_by\_css\_selector("img.rg\_i.Q4LuWd")

    print(keywords+' 찾은 이미지 개수:', len(images))

    links = []

    for i in range(1, len(images)):

        try:

            driver.find\_element\_by\_xpath(

                '//\*[@id="islrg"]/div[1]/div['+str(i)+']/a[1]/div[1]/img').click()

            links.append(driver.find\_element\_by\_xpath(

                '//\*[@id="Sva75c"]/div[2]/div/div[2]/div[2]/div[2]/c-wiz/div[2]/div[1]/div[1]/div[2]/div/a/img').get\_attribute('src'))

            # driver.find\_element\_by\_xpath('//\*[@id="Sva75c"]/div[2]/div/div[2]/div[2]/div[2]/c-wiz/div[2]/div[1]/div[1]/div[2]/div/a').click()

            print(keywords+' 링크 수집 중..... number :'+str(i)+'/'+str(len(images)))

        except:

            continue

    print("링크 수집 완료 !!!")

    forbidden = 0

    for index, i in enumerate(links):

        try:

            url = i

            start = time.time()

            urllib.request.urlretrieve(url, folder\_path + '/'+ keywords +'/' + keywords + "\_" + str(index) +".jpg" )

            print(str(index+1) + '/'+str(len(links)) + " " + keywords+ ' 다운로드 중...... Download time : ' + str(time.time() - start)[:5]+ ' 초' )

        except:

            forbidden += 1

            continue

    print(keywords + ' ---다운로드 완료---')

if \_\_name\_\_ == '\_\_main\_\_':

    pool = Pool(processes= 3)

    pool.map(image\_download, keyword)

**ex02. Cat & Dog 바운딩 박스 (cvTest)**

**ex03. 날씨Transform**

**ex04. Cutout**

**ex05. Mixup**

**ex06. Cutmix**