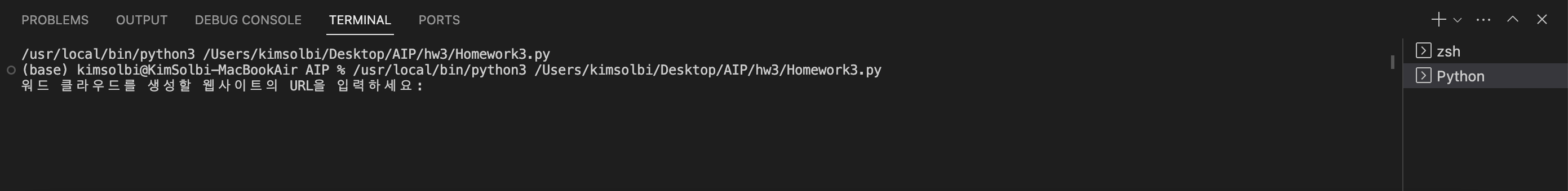
Homework 3 프로그램 소개 문서

<웹 크롤링을 기반으로 한 워드 클라우드 생성>

먼저 프로그램을 실행하면 "워드 클라우드를 생성할 웹사이트의 URL을 입력하세요:" 라는 메시지가 출력됩니다.



웹 페이지의 URL을 입력하면, 프로그램이 해당 웹 페이지의 본문을 분석하여 워드 클라우드를 생성해줍니다.

결과 화면은 다음과 같습니다.

다채로움, 패턴, 그래픽, 직사각형이(가) 표시된 사진

자동 생성된 설명

<코드 분석>

from selenium import webdriver

from selenium.webdriver.chrome.service import Service

from selenium.webdriver.chrome.options import Options

from webdriver\_manager.chrome import ChromeDriverManager

import wordcloud

import matplotlib.pyplot as plt

from selenium.webdriver.common.by import By

# 크롬 옵션 설정

chrome\_options = Options()

chrome\_options.add\_argument("--headless") # 헤드리스 모드 설정 / 크롬 창 따로 안 띄우도록 설정

# 웹 드라이버 설치 및 실행

service = Service(ChromeDriverManager().install())

driver = webdriver.Chrome(service=service, options=chrome\_options)

# URL 입력 받기

url = input("워드 클라우드를 생성할 웹사이트의 URL을 입력하세요: ")

driver.get(url)

# 페이지에서 본문 텍스트 가져오기

conan\_data = driver.find\_element(By.XPATH, '//\*[@id="JqFElvLbD"]/div[2]/div/div/div/div/div/div/div/div/div[1]/div[5]/div/div/div/div/div/div[5]/div/div/div/div/div/div[14]/div[1]/div/div[23]/div').text

# 필요없는 단어들 목록 만들기

s\_words = wordcloud.STOPWORDS.union({'네이버','카페','지부','온리전','공식','사이트는','셈','티켓값이','교보문고','감청의 권'})

# 워드 클라우드 생성

image = wordcloud.WordCloud(font\_path='/System/Library/Fonts/Supplemental/AppleGothic.ttf', width=1000, height=700, stopwords=s\_words).generate(conan\_data)

plt.figure(figsize=(40,30))

plt.imshow(image)

plt.show()

driver.quit()