6주차 AI 과제

41830022 박범진

코드

from cgitb import text

import bs4

import urllib.request

import csv

import datetime

import time

import requests

'''

날씨 정보 웹 크롤링 프로젝트

'''

csvName = 'paju\_weather.csv' #csv파일 이름 정의

with open(csvName, 'w', newline='') as csvfp: #csv파일 쓰기로 열기.

csvWriter = csv.writer(csvfp) #csvwriter 선언

csvWriter.writerow(['연월일', '시분초','날씨','온도', '습도', '강수확률', '풍향']) #종류를 행으로 적음

while True:

URL = 'https://www.google.com/search?q=%ED%8C%8C%EC%A3%BC+%EB%82%A0%EC%94%A8' #파주시 구글 웹사이트

headers = { #헤더 작성.

"User-Agent": 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.149 Safari/537.36 OPR/67.0.3575.115'}

page = requests.get(URL, headers=headers) #구글의 경우 크롤링시 웹 보안이 걸려 헤더가 없으면 제대로 불러올 수 없다. 따라서 헤더를 따로 정의해주어야 한다.

soup = bs4.BeautifulSoup(page.content, 'html.parser', from\_encoding="utf8") #utf8형식으로 크롤링 된 데이터를 받음

weather = soup.find("div", {"class": "wob\_dcp"}).text #날씨 찾기

print(weather)

rain\_percent = soup.find('span', {'id': 'wob\_pp'}).text #강수 확률

print(rain\_percent)

temp = soup.find('span', {'id': 'wob\_tm'}).text #온도

print(temp)

humi = soup.find('span', {'id': 'wob\_hm'}).text #습도

print(humi)

wind = soup.find('span', {'id': 'wob\_ws'}).text #풍속

print(wind)

now = datetime.datetime.now() #현재 시간

date = now.strftime('%Y-%m-%d') #연월일 출력 형식 지정

hour = now.strftime('%H:%M:%S') #시분초 출력 형식 지정

weathers = [date, hour, temp, weather, humi, rain\_percent, wind]

#리스트에 크롤링한 데이터를 담음

with open(csvName, 'a', newline='') as csvfp: #추가 모드(\*append)로 열기

csvWriter = csv.writer(csvfp)

csvWriter.writerow(weathers) #weather 행을 작성

print(weathers)

#time.sleep(1) #1초

#1sec x 60 =1min

#60sec x 60 = 1hour

#3600sec x 2 = 2hour

time.sleep(7200) #2시간

