

Python 라이브러리를 이용한 인공지능 비서 제작



개인 프로젝트 오성준

CONTENTS

01

라이브러리

02

기본 세팅

03

크롤링

1. 라이브러리

라이브러리

```
import requests  
from bs4 import BeautifulSoup
```

requests : Python에서 HTTP 요청을 보내는 라이브러리

BeautifulSoup : HTML을 파싱하여 데이터를 추출하는 라이브러리

파싱이란?

문자열데이터를 분석하고 분해하여 목적인 패턴에 맞게 문자열의 구조를 결정하는 것

1. 라이브러리

2. 기본 세팅

기본 세팅

```
def create_soup(url):  
    headers={'User-Agent': 'Mozilla/5.0 (Windows NT  
    res=requests.get(url,headers=headers)  
    soup=BeautifulSoup(res.text,'html.parser')  
    return soup
```

headers 안의 User-Agent 코드를 넣어주는 이유

일부 웹 사이트에서 스크래핑 요청을 거부하는 경우가 있기 때문에 이를 우회하기 위함
(일반적인 브라우저에서 요청하는 것과 같은 값을 가지게 하는 것)

가져온 HTML 문서를 BeautifulSoup 객체로 변환하여 반환
이때 'html.parser'를 사용하여 HTML 문서를 파싱하도록 지정

1. 라이브러리

2. 기본 세팅

3. 날씨 크롤링

무엇을 도와드릴까요? (날씨, 뉴스, 주식 / 종료):

날씨
어떤 도시의 날씨를 알고 싶으세요? 영화동

현재 온도 : 현재 온도9.8°
최저 기온 : 최저기온3°
최고 기온 : 최고기온17°
강수 확률 : 오전 20% 구름많음 오후 20% 구름많음

이상입니다.

날씨 크롤링

```
def get_weather(city):  
    url = f"https://search.naver.com/search.naver?where=#  
        nexearch&sm=top_hyt&fbm=1&ie=utf8&query={city}날씨"  
    soup=create_soup(url)  
    min_temp=soup.find('span',{'class':'lowest'}).get_text()  
    max_temp=soup.find('span',{'class':'highest'}).get_text()  
    rain=soup.find('div',{'class':'cell_weather'}).get_text().strip()  
    curr_temp=soup.find('div',{'class':'temperature_text'}).get_text().strip()  
    print("-----")  
    print("현재온도 : ", curr_temp)  
    print("최저 기온 : ", min_temp)  
    print("최고 기온 : ", max_temp)  
    print("강수 확률 : ", rain)  
    print("-----")  
    return "이상입니다."
```

#create_soup 함수를 사용하여, 해당 URL의 HTML 문서를 BeautifulSoup 객체로 변환

각각의 날씨 정보(최저 기온, 최고 기온, 강수 확률, 현재 온도)는 HTML 문서 내에서

CSS class를 가진 요소로 표시되어 있는데 soup 객체에서 이를 추출하여,각각의 변수에 저장

이 때 get_text() 메서드를 사용하여, 해당 요소의 텍스트 값을 가져온다.

1. 라이브러리

2. 기본 세팅

3. 날씨 크롤링

4. 뉴스 크롤링

무엇을 도와드릴까요? (날씨, 뉴스, 주식 / 종료):

뉴스

어떤 카테고리의 뉴스를 알고 싶으세요? (정치(100), 경제(101), 사회(102), 문화(103), 세계(104)): 101

1 롯데제과 → '롯데월드푸드'로 사명 변경[이코노 브리징]

2 롯데푸드 품고 글로벌 종합식품기업으로... 롯데제과, '롯데월드푸드'로 사명 변경

3 롯데제과, 56년만에 '제과' 떼고 '롯데월드푸드'로 사명 변경

4 롯데제과, 56년 만에 '롯데월드푸드'로 사명 변경 확정... 4월부터 적용

5 글로벌 기업 총수 100여명 부른 中... 반도체 협력·적극적 투자 요구할듯

뉴스 크롤링

```
def get_news(category):  
    url = f"https://news.naver.com/main/main.naver?mode=LSD&mid=shm&sid1={category}"  
    soup = create_soup(url)  
    news_list = soup.select("ul.cluster_list li a")  
    news_titles = [news.text for news in news_list]  
    return news_titles
```

네이버 뉴스에서는 ul 태그의 class 이름이 "cluster_list"인
태그 내에 li 태그의 하위 요소인 a 태그에 뉴스 제목이 포함

이를 select 메서드를 사용하여 추출하고, 뉴스 제목을 리스트 형태로 저장

1. 라이브러리

2. 기본 세팅

3. 날씨 크롤링

4. 뉴스 크롤링

5. 주식 크롤링

무엇을 도와드릴까요? (날씨, 뉴스, 주식 / 종료):

주식
주식 코드 번호를 입력해주세요. (예 : 삼성전자(005930))018880

해당 종목의 이름은 한온시스템입니다.

시가 총액 정보입니다. 현재 한온시스템의 시가총액은 4조4,893억원입니다.

PER 정보입니다. 현재 한온시스템의 PER은 29.41배 입니다.

배당 정보입니다. 현재 한온시스템의 배당수익률은 2022.12 기준 4.28%입니다.

현재 전문가들의 투자 의견은 3.50매수입니다.

동일 업종 평균 PER는 10.50이므로 동일 업종 평균보다 높습니다.

이상입니다.

주식 크롤링

```
def get_stock(code):  
    url = f"https://finance.naver.com/item/main.naver?code={code}"  
    soup = create_soup(url)  
    name = soup.select_one('h2 a').text.strip() #종목이름  
    name = name.replace('\\\\t', '').replace('\\\\n', '')  
    marketcap = soup.select_one('#_market_sum').text.strip() #시가총액  
    marketcap = marketcap.replace('\\\\t', '').replace('\\\\n', '')  
    per = soup.select_one('#_per').text.strip()  
    per = per.replace('\\\\t', '').replace('\\\\n', '')  
    divedends = soup.select_one('#_dvr').text.strip() #배당 수익률  
    divedends = divedends.replace('\\\\t', '').replace('\\\\n', '')  
    prof = soup.select_one('.rwidth .f_up').text.strip() #전문가 투자 의견  
    prof = prof.replace('\\\\t', '').replace('\\\\n', '')  
    s_per = soup.select_one('div~ div+ .gray .strong em').text.strip() #동일 업종 평균 per  
    s_per = s_per.replace('\\\\t', '').replace('\\\\n', '')
```

select_one 메서드를 사용하여 추출하고, 각 정보를 저장

각 정보에 대해, 문자열에서 탭 문자와 개행 문자를 제거하는 replace 메서드를 적용

1. 라이브러리

2. 기본 세팅

3. 날씨 크롤링

4. 뉴스 크롤링

5. 주식 크롤링

6. 비서 코드

무엇을 도와드릴까요? (날씨, 뉴스, 주식 / 종료):

안녕하세요! 개인 비서입니다.

비서 코드

```
def personal_assistant():
    print("-----")
    print("안녕하세요! 개인 비서입니다.")
    print("-----")
    while True:
        try:
            command = input("무엇을 도와드릴까요? (날씨, 뉴스, 주식 / 종료): \n\n")
            if command == "종료":
                print("감사합니다. 다음에 또 만나요!")
                break

            elif command == "날씨":
                city = input("어떤 도시의 날씨를 알고 싶으세요? ")
                print(get_weather(city))

            elif command == "뉴스":
                category = input("어떤 카테고리의 뉴스를 알고 싶으세요? (정치(100), 경제(200), 스포츠(300), 연예(400)) \n\n")
                idx = 1
                for i in get_news(category):
                    if i == "\n\n":
                        continue
                    elif i == "\n\n동영상기사\n\n":
                        continue
                    print(idx, " ", i)
                    print("- * 100")
                    idx+=1

            elif command == "주식":
                code = input("주식 코드 번호를 입력해주세요. (예 : 삼성전자(005930))")
                print(get_stock(code))

            else:
                print("잘못된 명령어입니다.")
        except:
            print("오류가 발생했습니다. 다시 입력해주세요.")
```




THANKS YOU

