

REPORT

빅데이터 수집 시스템 개발
정기 수행평가

과 목 명	통합구현
능 력 단 위	빅데이터 수집 시스템 개발
성 명	장경준

수행평가 실습 보고서			
능력단위 요소	빅데이터 수집 시스템 개발	작 성 자	장경준
평가방법	서술형	작성일자	
단 계 명	구현	문서번호	#1

1.SDK(System Development Kit) 또는 개발 도구 정보

구분	내용
SDK(개발도구) 명칭	PyCharm
SDK(개발도구) 버전	PyCharm Community Edition 2020.3
SDK(개발도구)설명 (주요기능의 설명)	<ul style="list-style-type: none"> ● IntelliJ 플랫폼만의 수준높은 코드 자동완성 기능을 제공 ● irtualEnv, Anaconda 등 패키지 격리 환경을 네이티브로 지원. ●
SDK(개발도구) 사용방법 (가격, 절차 등)	오픈소스(커뮤니티 에디션) 프로페셔널 에디션(유료, \$8.9/월 또는 \$89/년)

2. 평가문항 풀이

2.1. 2번 문제

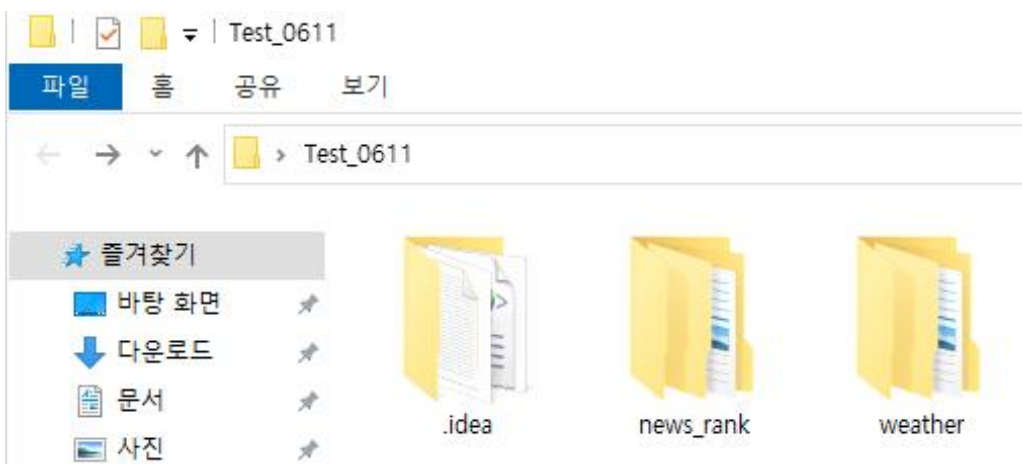
네이버 인기 검색어가 없으므로 다음 뉴스 랭킹으로 작성.
selenium 이용

-코드 작성 (selector 추출 부분 제외)

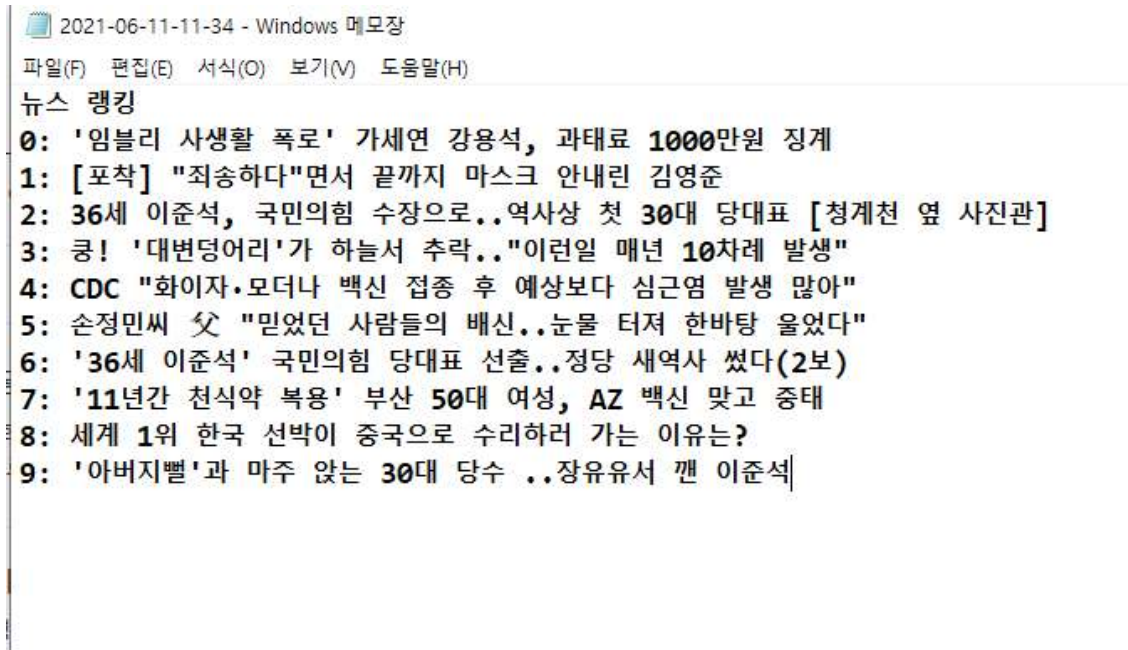
```
Test_0611 - test2.py
test2.py x test3.py x
9 import os
10 # 크롬 가상브라우저 실행
11 browser = webdriver.Chrome('./chromedriver.exe')
12 # 다음 뉴스 이동
13 browser.get('https://news.daum.net/')
14
15 # 랭킹 클릭
16 btn_rank = browser.find_element_by_css_selector('#kakao6nb > div > ul > li:nth-child(8) > a')
17 btn_rank.click()
18
19 # 디렉터리 생성
20 dir = "./news_rank/{:%Y-%m-%d}".format(datetime.now())
21
22 if not os.path.exists(dir):
23     # 해당 디렉터리가 존재하지 않으면
24     os.makedirs(dir)
25
26 # 데이터 파일 생성
27 fname = "{:%Y-%m-%d-%H-%M.txt}".format(datetime.now())
28 file = open(dir+'/' + fname, mode='w', encoding='utf-8')
29 # 데이터 파일 저장
30 # 타이틀 출력
31 news_tits = browser.find_elements_by_css_selector('#mArticle > div.rank_news > ul.list_news2 > li > div.con')
32
33 file.write('뉴스 랭킹\n')
34 for i in range(10):
35     file.write("{}: {}\n".format(i, news_tits[i].text))
36
37 print("뉴스 랭킹 수집 완료..")
38 file.close()
39
40 for i in range(10)
```

Run: test2 x
C:\Users\user\AppData\Local\Programs\Python\Python37\python.exe C:/Users/user/Desktop/Test_0611/test2.py
뉴스 랭킹 수집 완료..
Process finished with exit code 0

- 폴더 생성 결과



- 자료 수집 결과



2.2. 3번 문제

request 와 BeautifulSoup 이용.

- 코드 작성(selector 부분 제외)

The image shows a screenshot of an IDE (likely PyCharm) with a Python script named `test3.py` and a terminal window below it.

Python Script (`test3.py`):

```
6 from selenium import webdriver
7 from datetime import datetime
8 import os
9 # 가상 브라우저 실행
10 browser = webdriver.Chrome('./chromedriver.exe')
11 # 페이지 이동
12 browser.get('https://www.weather.go.kr/w/obs-climate/land/city-obs.do')
13 # 페이지 파싱
14 trs = browser.find_elements_by_css_selector('#weather_table > tbody > tr')
15
16 # 디렉터리 생성
17 dir_ = "./weather/{:%Y-%m-%d}".format(datetime.now())
18
19 if not os.path.exists(dir_):
20     # 해당 디렉터리가 존재하지 않으면
21     os.makedirs(dir_)
22
23 # 데이터 파일 생성
24 fname = "{:%Y-%m-%d-%H-%M.txt}".format(datetime.now())
25 file = open(dir_+'/' + fname, mode='w', encoding='utf-8')
26 file.write('지역,기온\n')
27 for tr in trs:
28     local = tr.find_element_by_css_selector('td:nth-child(1) > a').text
29     temp = tr.find_element_by_css_selector('td:nth-child(6)').text
30
31     print('{},{}'.format(local, temp))
32
33 file.close()
34 # 가상 브라우저 종료
35 browser.close()
```

Terminal Output:

```
Run: test3
↑
↓
함경군,22.6
합천,20.3
해남,25.5
홍성,22.2
홍천,21.6
흑산도,21.6
Process finished with exit code 0
```

Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.

- 자료 수집 결과

- 크론탭 등록

```
root@Server101:/home/server101
* * * * * python3 /home/server101/test2.py
* * * * * python3 /home/server101/test3.py
```

- 크론탭 등록 결과 및 데몬 시작

```
[root@Server101 server101]# systemctl start crond
[root@Server101 server101]# crontab -l
-bash: crontab: command not found
[root@Server101 server101]# crontab -l
* * * * * python3 /home/server101/test2.py
* * * * * python3 /home/server101/test3.py
```