

빅데이터 수집시스템 개발_최현진

2021-06-11

```

1 import os
2 import requests as req
3 from bs4 import BeautifulSoup as bs
4 from datetime import datetime
5
6 # 페이지 요청하기
7 response = req.get('https://issue.zum.com/')
8 # print(response.text)
9
10 # 페이지 파싱하기
11 dom = bs(response.text, 'html.parser')
12 divs = dom.select('#issueKeywordList > li > div.cont')
13
14 # 디렉터리 생성
15 dir = './test2/{:%Y-%m-%d}'.format(datetime.now())
16
17 if not os.path.exists(dir):
18     os.makedirs(dir)
19
20 # 파일 생성
21 fname = "{:%Y-%m-%d-%H-%M.txt}".format(datetime.now())
22 file = open(dir+'/'+fname, mode='w', encoding='utf-8')
23
24 # 파일 저장
25 for div in divs:
26     rank = div.find(class_='num').text
27     word = div.find(class_='word').text
28     file.write('%s,%s\n' % (rank[:-1], word))
29

```

```

30 file.close()
31
32 print('검색어 수집 완료')
33
34 file.close()
35 print('검색어 프로그램 종료')

```

2.아래 crawling작업을 수행하시오

1)네이버 실시간 검색어 수집

```

1 SHELL=/bin/bash
2 PATH=/sbin:/bin:/usr/sbin:/usr/bin
3 MAILTO=root
4
5 # For details see man 4 crontabs
6
7 # Example of job definition:
8 # .----- minute (0 - 59)
9 # | .----- hour (0 - 23)
10 # | | .----- day of month (1 - 31)
11 # | | | .----- month (1 - 12) OR jan,feb,mar,apr ...
12 # | | | | .---- day of week (0 - 6) (Sunday=0 or 7) OR sun,mon,tue,wed,thu,fri,sat
13 # | | | | |
14 # * * * * * user-name command to be executed
15 * * * * * root python3 /root/weather.py

```

-- INSERT --

15,2

All

2.아래 crawling작업을 수행하시오

1분마다 한번 crawler 실행

```

1 import os
2
3 import requests as req
4 from bs4 import BeautifulSoup as bs
5 from datetime import datetime
6
7 # 페이지 이동
8
9 response = req.get('https://www.weather.go.kr/w/obs-climate/land/city-obs.do')
10
11 # 페이지 파싱
12 dom = bs(response.text, 'html.parser')
13 trs = dom.select('#weather_table > tbody > tr')
14
15
16 #디렉토리 생성
17 dir = "./test1/{:%Y-%m-%d}".format(datetime.now())
18
19 if not os.path.exists(dir):
20     # 폴더가 없다면
21     os.makedirs(dir)
22
23 # 파일 생성
24 fname = "{:%Y-%m-%d-%H-%M.txt}".format(datetime.now())
25 file = open(dir+'/'+fname, mode='w', encoding='utf-8')
26
27 file.write('지역, 현재일기, 시정, 운량, 중하운량, 현재기온, 이슬점온도, 불쾌지수, 일강수, 습도, 풍향, 풍속, 해면기압\n')
28
29 #print(trs)

```

3) 아래 crawling 작업을 수행하시오

1) 전국 날씨 데이터 수집


```
1 SHELL=/bin/bash
2 PATH=/sbin:/bin:/usr/sbin:/usr/bin
3 MAILTO=root
4
5 # For details see man 4 crontabs
6
7 # Example of job definition:
8 # .----- minute (0 - 59)
9 # | .----- hour (0 - 23)
10 # | | .----- day of month (1 - 31)
11 # | | | .----- month (1 - 12) OR jan,feb,mar,apr ...
12 # | | | | .---- day of week (0 - 6) (Sunday=0 or 7) OR sun,mon,tue,wed,thu,fri,sat
13 # | | | | |
14 # * * * * * user-name command to be executed
15 1 * * * * root python3 /root/weather.py
```

15,3

All

3. 아래 crawling작업을 수행하시오

2) 1시간 마다 한번 Crawler 실행