

C:_

파이썬을 이용한
**개인비서
만들기**

이태선

목차


Contents 1.	-----	함수 create_soup
Contents 2.	-----	함수 scrap_headlines
Contents 3.	-----	함수 crawl_school_violence_questions
Contents 4.	-----	함수 scrap_to_excel
Contents 5.	-----	함수 scripted_data_word_count

Contents 1

사용한 코드 | 함수 create_soup

```
import requests
from bs4 import BeautifulSoup

def create_soup(url):
    headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)
    AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Safari/537.36"}
    res = requests.get(url, headers=headers)
    soup = BeautifulSoup(res.text, "html.parser")
    return soup
url="https://news.naver.com/main/main.naver?mode=LSD&mid=shm&sid1=100"
soup=create_soup(url)
ul=soup.find("ul",{"class":"section_list_ranking_press_rankingList"})
lis=ul.findAll("li")
index=1
for li in lis:
    atag=li.find("a",{"class":"list_tit nclicks('rig.renws2')"})
    aurl=atag.get("href")
    title=atag.get_text()
    print(index,title)
    print(aurl)
    index += 1
```



```
1 '전두환 비리' 손자 폭로전 ... 이순자씨 "할미 품으로 와라"
https://n.news.naver.com/mnews/ranking/article/005/0001594053?ntype=RANKING&sid=001
2 "초등생과 부적절한 관계" ... '올해의 교사상' 30대 女교사 ...
https://n.news.naver.com/mnews/ranking/article/009/0005102211?ntype=RANKING&sid=001
3 "완전 왕따, 미운털 박힌 것 같고" ... 극단선택 내몰린 어린 ...
https://n.news.naver.com/mnews/ranking/article/057/0001728995?ntype=RANKING&sid=001
4 KBS 재방료 밀린 돈 10억 ... 배우 측 "미제껏 주다 갑 ...
https://n.news.naver.com/mnews/ranking/article/023/0003752073?ntype=RANKING&sid=001
5 삼성전자 어쩌나 ... 美 반도체공장 건설비 10조 불어났다
https://n.news.naver.com/mnews/ranking/article/015/0004821421?ntype=RANKING&sid=001
```

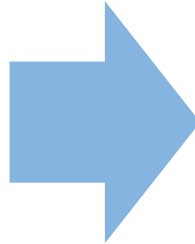
Contents 2

사용한 코드 | 함수 scrap_headlines

```
def scrap_headlines():  
    url="https://search.naver.com/search.naver?where=news&ie=utf8&sm=nws_hly&query=%ED%95%99%EA%B5%90%ED%8F%AD%EB%A0%A5"  
    soup=create_soup(url)  
    ul=soup.find("ul",{"class":"list_news"})  
    lis=ul.findAll("li")
```

```
    index=1  
    for li in lis:  
        atag = li.find("a", {"class": "news_tit"})  
        if atag is not None:  
            aurl = atag.get("href")  
            title = atag.get_text()  
            print(index, title)  
            print(aurl)  
            index += 1  
            print()
```

```
scrap_headlines()
```



- 1 학교폭력 피해로 대면 상담한 초·중학생 늘었다
<http://www.kookje.co.kr/news2011/asp/newsbody.asp?code=0300&key=20230315.99099004368>
- 2 전북교육청, 지역 극단 10곳과 학교폭력 예방 맞손
<https://www.yna.co.kr/view/AKR20230315074100055?input=1195m>
- 3 "학교폭력 피해자, 장기간 우울·불안... 전문조직 보강해야"(종합)
http://www.newsis.com/view/?id=NISX20230314_0002226668&cID=10205&pID=10200
- 4 윤혜진 "나는 학교 폭력 피해자... 딸 유명인 딸이라"(왓챠티비)
<http://www.sportsseoul.com/news/read/1206202?ref=naver>
- 5 이주호 "인성교육, 학교폭력 대응하는 데 근간 돼야"
<https://www.joongang.co.kr/article/25147429>
- 6 울산교육청, 학교폭력 화해분쟁조정지원단 운영
<https://www.news1.kr/articles/4982473>
- 7 이주호, 학교폭력 대책 위해 정신건강 전문가 간담회 참석
<http://www.fnnews.com/news/202303141155211301>
- 8 제주시 학교 폭력 증가세... 지난해 180건 심의
<https://news.kbs.co.kr/news/view.do?ncd=7627936&ref=A>
- 9 "춘지 안준다고 패던 선생님" ... '교사 폭력' 고발 확산
<https://economist.co.kr/article/view/ecn202303140014>
- 10 정부 학교폭력대책위, 연 2회 원칙인데 2년 연속 1회만
http://tbs.seoul.kr/news/newsView.do?typ_800=68&idx_800=3492061&seq_800=20484762


Contents 3

사용한 코드 | 함수 crawl_school_violence_questions

```
import requests
from bs4 import BeautifulSoup

def crawl_school_violence_questions():
    url = "https://kin.naver.com/search/list.nhn?query=%ED%95%99%EA%B5%90%ED%8F%AD%EB%A0%A5"
    res = requests.get(url)
    res.raise_for_status()
    soup = BeautifulSoup(res.text, "html.parser")
    question_selector = "ul.basic1 > li"
    question_list = soup.select(question_selector)
    question_dict = {}
    for question in question_list:
        title = question.select_one("a").text.strip()
        link = question.select_one("a")["href"]
        question_dict[title] = link
    return question_dict
question_dict = crawl_school_violence_questions()

for title in sorted(question_dict.keys()):
    print(title, ":", question_dict[title])
```



```
학원학교폭력변호사 비용이 궁금합니다 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=6020602&docId=441675861&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=28&search_sort=0&spq=0
학교폭력 대상 개정 이유 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=60208&docId=441693212&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=9&search_sort=0&spq=0
학교폭력 따돌림 고소문의 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=60206&docId=439916951&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=7&search_sort=0&spq=0
학교폭력 방관자에 대해 : https://kin.naver.com/qna/detail.naver?d1id=11&dirId=110307&docId=440565176&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=6&search_sort=0&spq=0
학교폭력 변호사 필요할까요? : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=60208&docId=440462796&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=38&search_sort=0&spq=0
학교폭력 조치 관련 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=60208&docId=441114717&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=8&search_sort=0&spq=0
학교폭력 형사신고 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=6020602&docId=438203314&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=48&search_sort=0&spq=0
학교폭력변호사 상담 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=6020602&docId=438830787&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=10&search_sort=0&spq=0
학교폭력변호사 선임 비용 문의 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=60208&docId=441265190&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=58&search_sort=0&spq=0
학교폭력처벌 관련 도와주세요 : https://kin.naver.com/qna/detail.naver?d1id=6&dirId=6020602&docId=440487940&qB=7ZWZ6rWQ7Y+t66Cl&enc=utf8&ion=kin&rank=1&search_sort=0&spq=0
```

Contents 4

사용한 코드 | 함수 scrap_to_excel

```
import pandas as pd
```

```
def scrap_to_excel(keyword, page):
```

```
    result = []
```

```
    for i in range(1, page+1):
```

```
        url = "https://search.daum.net/search?nil_suggest=btn&w=news&DA=PGD&cluster=y&q=" + keyword + "&p=" + str(i)
```

```
        soup = create_soup(url)
```

```
        ul = soup.find("ul", {"class": "list_news"})
```

```
        lis = ul.findAll("li")
```

```
        for li in lis:
```

```
            atag = li.find("a", {"class": "tit_main fn_tit_u"})
```

```
            aurl = atag.get("href")
```

```
            title = atag.get_text()
```

```
            p_class = li.find("p", {"class": "desc"})
```

```
            body = p_class.get_text()
```

```
            span_class = li.find("span", {"class": "f_nb"})
```

```
            date = span_class.get_text()
```

```
            result.append([title, date, aurl, body])
```

```
df = pd.DataFrame(result, columns=['title', 'date', 'url', 'body'])
```

```
df.to_excel(keyword+'.xlsx', index=False)
```

```
print("\n저장 완료\n 저장 위치는 \"D:\\Anaconda\" 입니다.\n")
```

```
return df
```

```
df = scrap_to_excel("학교폭력", 10)
```

"저장 완료" 저장 위치는 "D:\Anaconda" 입니다.

사용한 코드 | 함수 scrap_to_excel

[illegible]

Contents 5

사용한 코드 | 함수 scripted_data_word_count

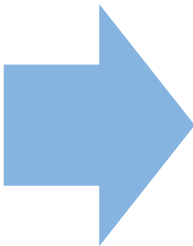
```
import pandas as pd
from collections import Counter
from konlpy.tag import Okt

def scripted_data_word_count(keyword):
    result = []

    df = pd.read_excel(keyword+".xlsx")

    okt = Okt()
    text_list = df['body'].tolist()
    text = '\n'.join(text_list)
    nouns = [word for word in okt.nouns(text) if len(word) >= 2 and keyword != word]
    count = Counter(nouns)
    count_top = [(i[0], i[1]) for i in count.most_common(10)]
    result = print(*[f'{k+1}. {i[0]} : {i[1]}' for k, i in enumerate(count.most_common(10))], sep='\n')
    return result

scripted_data_word_count("학교폭력")
```



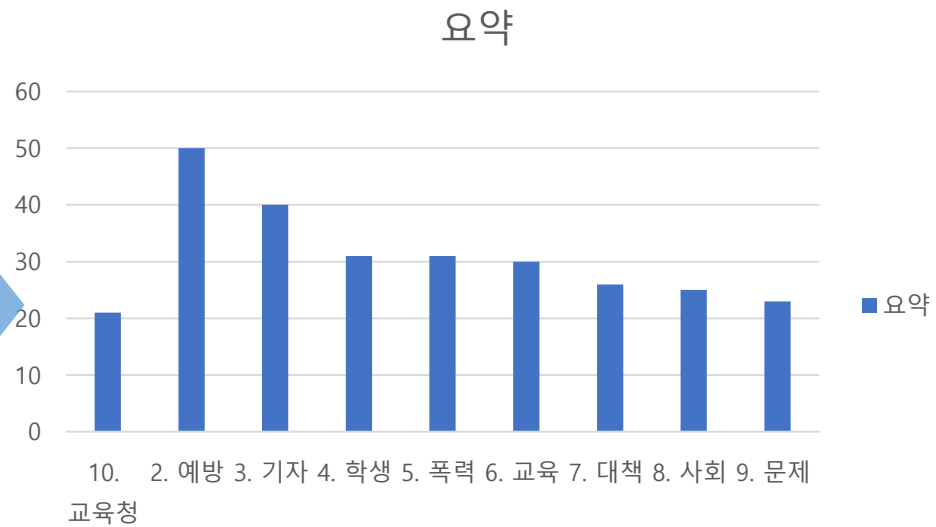
1.	학교	: 57
2.	예방	: 50
3.	기자	: 40
4.	학생	: 31
5.	폭력	: 31
6.	교육	: 30
7.	대책	: 26
8.	사회	: 25
9.	문제	: 23
10.	교육청	: 21

Contents 5

사용한 코드 | 함수 scripted_data_word_count

```
1. 학교 : 57  
2. 예방 : 50  
3. 기자 : 40  
4. 학생 : 31  
5. 폭력 : 31  
6. 교육 : 30  
7. 대책 : 26  
8. 사회 : 25  
9. 문제 : 23  
10. 교육청 : 21
```

A	B
1. 학교	57
2. 예방	50
3. 기자	40
4. 학생	31
5. 폭력	31
6. 교육	30
7. 대책	26
8. 사회	25
9. 문제	23
10. 교육청	21



C:_

감사합니다.