

네이버 뉴스 댓글 크롤링

1. 전체 댓글의 갯수 가져오기
2. 사용자에게 수집할 댓글 수 입력받기
3. 댓글 수집하기
4. 파일로 저장하기

웹페이지 접근하기

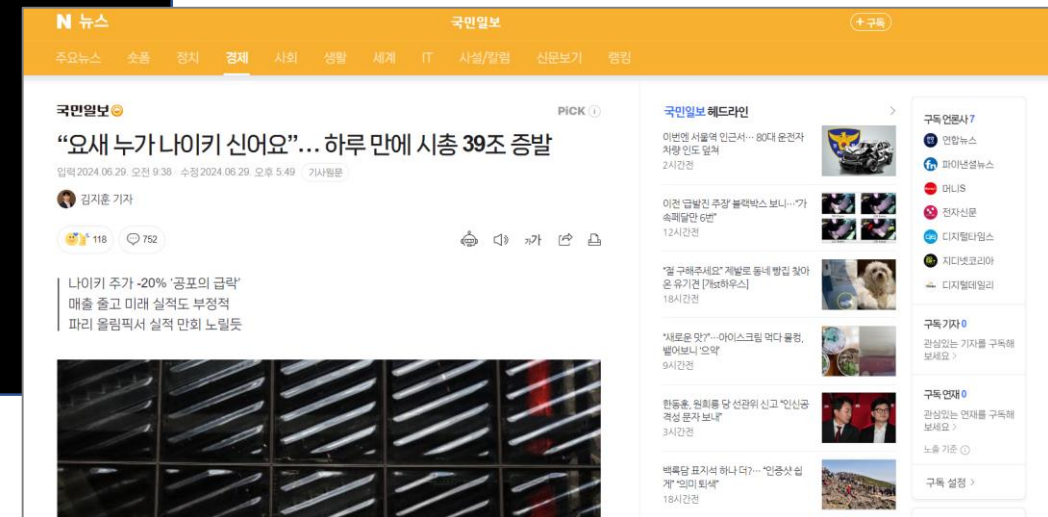
```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from bs4 import BeautifulSoup
import pandas as pd
import time
import math
```

웹드라이버 생성

```
driver = webdriver.Chrome()
```

웹페이지 접근

```
url = 'https://n.news.naver.com/article/005/0001707120'
driver.get(url)
driver.maximize_window()
```



전체 댓글 수 가져오기

국민일보

PICK

“요새 누가 나이키 신어요”... 하루 만에 시총 39조 증발

입력 2024.06.29. 오전 9:38 · 수정 2024.06.29. 오후 5:49 기사원문

 김지훈 기자

 118

 752

나이키 주가 -20% '공포의 급락'

매출 줄고 미래 실적도 부정적

파리 올림픽서 실적 만회 노릴듯

전체 댓글 수 가져오기

```
# 전체 댓글 수 가져오기  
review_total_cnt = int(driver.find_element(By.ID, 'comment_count').text)  
print(review_total_cnt)
```

752

수집할 댓글 수 입력받기

```
# 수집할 댓글 수 입력받기
message = f'총 {review_total_cnt}개의 댓글이 있습니다. 몇개의 댓글을 수집할까요?'
try:
    review_cnt = int(input(message))
    if review_cnt > review_total_cnt: # 수집할 댓글 수 > 전체 댓글 수라면
        review_cnt = review_total_cnt # 수집할 댓글수를 전체 댓글수로 지정
except:
    review_cnt = 20 # 오류 발생 시 수집할 댓글 수 20개로 지정

message = f'{review_cnt}개의 댓글을 수집하겠습니다.===== '
print(message)
```

댓글 화면으로 이동

국민일보

PICK

“요새 누가 나이키 신어요”... 하루 만에 시총 39조 증발

입력 2024.06.29. 오전 9:38 · 수정 2024.06.29. 오후 5:49 기사원문

 김지훈 기자

 118

 752

클릭


 가가  


나이키 주가 -20% ‘공포의 급락’
매출 줄고 미래 실적도 부정적
파리 올림픽서 실적 만회 노릴듯

댓글 화면으로 이동

```
# 댓글 수를 클릭하면 댓글 화면으로 이동  
driver.find_element(By.CSS_SELECTOR, '#comment_count').click()
```

더보기 클릭하여 수집할 개수만큼 화면에 댓글 보이기

 **aoir****** >
2024.06.29. 10:56
팔로우 ⋮
솔직히 메이커 값임. 이념들 인체공학적 편안함 같은건 신경도 안쓰는거 같은데. 그저께 나이키 슬리퍼 샀는데 역시나 그
지같은 착용감ㅏㅏ
답글
19 3

 **gusq****** >
2024.06.29. 16:26
팔로우 ⋮
동양인은 발볼 넓어서 나이키랑은 잘 안맞는듯
답글
17 2

더보기 ▾

한번에 20개씩 추가로 보여줌

더보기 클릭하여 수집할 개수만큼 화면에 댓글 보이기

```
# 더보기 클릭
# 더보기 클릭할 때마다 20개의 댓글이 추가로 보여짐
# 더보기 클릭 횟수는? ceil(수집할 댓글 갯수/20)
pagecnt = math.ceil(review_cnt/20)-1
for i in range(pagecnt):
    driver.find_element(By.CLASS_NAME, 'u_cbox_page_more').click()
    time.sleep(1)
```

데이터 추출

```
# 웹드라이버의 html 가져와 BeautifulSoup 객체 만들기
html = driver.page_source
soup = BeautifulSoup(html, 'html.parser')
```

데이터 추출

```
# BeautifulSoup 객체에서 데이터 추출하기
reviews = soup.select('.u_cbox_area')
review_list = []
for review in reviews:
    # 날짜
    date = review.select_one('.u_cbox_date').text
    # 댓글내용
    try:
        content = review.select_one('.u_cbox_contents').text
    except:
        content = '-----사용자에의해 삭제된 댓글입니다.-----'
    # 좋아요갯수
    try:
        good = review.select_one('.u_cbox_cnt_recomm').text
    except:
        good = 0
    # 싫어요갯수
    try:
        bad = review.select_one('.u_cbox_cnt_unrecomm').text
    except:
        bad = 0
    review_list.append([date, content, good, bad])
```

파일에 저장

데이터프레임 생성

```
file_name = 'review'
df = pd.DataFrame(review_list,
                  columns = ['일시', '댓글내용', '좋아요', '싫어요']).iloc[:review_cnt]
```

파일명 생성

```
import time
import openpyxl
now = time.localtime()
filename = f'{now.tm_year}{now.tm_mon:02d}{now.tm_mday:02d}_{now.tm_hour:02d}{now.tm_min:02d}{now.tm_sec:02d}'
filename = '뉴스댓글_'+filename
```

파일로 저장

```
df.to_csv(filename+'.csv', index=False)
df.to_excel(filename+'.xlsx', index=False)

# pd.read_csv(filename+'.csv')
# pd.read_excel(filename+'.xlsx')
```

하나의 코드로 합치기