

네이버 기사 크롤링

- <http://news.naver.com/main/main.nhn?mode=LSD&mid=shm&sid1=105>
(<http://news.naver.com/main/main.nhn?mode=LSD&mid=shm&sid1=105>).
- thread 사용

[1]:

```
1 from selenium import webdriver
```

[2]:

```
1 from selenium.webdriver.common.by import By
```

[3]:

```
1 article_list = []
2
3 def get_article(page):
4     driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
5     driver.get("http://news.naver.com/main/main.nhn?mode=LSD&mid=shm&sid1=105")
6     articles = driver.find_elements(By.CSS_SELECTOR, '#section_body li')
7
8     for article in articles:
9         title = article.find_element(By.CSS_SELECTOR, 'dt:not(.photo)')
10        article_list.append(title)
11
12    print("end :", page)
13
14    driver.quit()
```

[4]:

```
1 %%time
2 for page in range(1, 5):
3     get_article(page)
```

C:\Users\student\AppData\Local\Temp\ipykernel_12844\100575472.py:4: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
 driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")

end : 1

end : 2

end : 3

end : 4

CPU times: total: 203 ms

Wall time: 15.7 s

[5]:

```
1 len(article_list), article_list[:30]
```

```
(80,
[' “부광약품, 자회사 신약 개발 진전 기대” ',
'"없어 못파는 천연물 기반 감기약, 아이월드 올매출 300억 거
뜬"',
' [사이언스카페] 동물의 고통, 인공지능이 표정 보고 알아낸다',
' 아이폰14 프로 · 프로맥스 가격 100달러 오를 듯',
' [넷글리뷰] “e심 다운로드 비용, 말도 안 된다” ',
'"2만달러 못 뚫는 '비트코인'...상승세 이어가는 '이더리움"',
' 체장염 신약후보 1차 지표 확보못한 ‘SCM생명과학’ ...개발 지속
가능성은?',
'"이재용 부회장의 첫 '빅딜'...하만 대신 ARM을 택했다면",
'"큰 거 온다"...W'아이폰14W'와 함께 W'에어팟 프로2W' 공개 전
망',
' 삼성전자, 2분기 세계 프리미엄폰 점유율 19% 2위... 1위는 애플 5
7%',
' 94兆 콘솔 시장 잡아라... 글로벌 공략 시동 건 K게임',
' 갤럭시Z폴드5 S펜 탑재 고민하는 삼성... “기술혁신이 필요하
다” ',
' [누리호를 만드는 기업들](17) 컨테이너에서 지내며 발사대 개발
한 현대중공업',
' 투자 촉한기인데..반년만에 기업가치 1兆→2兆 경총 똬 이 회사',
' 클라우드 보안 서비스 뜨는 이유',
' 역대급 힌남노에 부산 위치한 데이터센터들 비상상황 돌입',
' 삼성 가전, 에너지효율도 ‘원톱’ ... 세탁기 1년 써도 전기료 360
0원',
' “실제 게이머 집에 TV 놓고 의견 반영... 자유로운 휘어짐 · 높낮
이 디테일 탄생” ',
' [미래기술25]"나노기술은 이미 현실..범용기술로 과학산업 기
반"',
' 반도체 다운사이클 본격화... 전문가 77% “현재 위기다” ',
' “佛 배경의 피노키오 모험... 한국색 뺀 글로벌게임 통했다” ',
'" '절치부심' 삼성전자, 엑시노스로 종합반도체 기업 속도내나",
'"카카오게임즈, 성난 '말뚱 아빠들' 만난다",
'"태풍 끝날 때까지 재택근무 부탁"...IT업계, 줄줄이 출근 금지',
' ‘명절때 전 안부쳐도된다’ 차례상 표준안에... “왜 이제서야”
온라인 엇갈린 반응',
'"이더리움, 업그레이드 후 두개로 쪼개질 가능성 크다"',
'"마라도 CCTV 볼 수 있다"...힌남노 정보 찾아 W'네 · 카W'에 몰린
다',
' “대만의 위협” ...삼성 · LG 양강체제 ‘흔들’ ',
' 테슬라 주가, 주식분할에도 왜 지지부진할까...가장 큰 악재는?',
' “여배우는 아이폰만 쓰는줄 알았는데” 박은빈 손에 포착된 ‘이
것’ '])
```

pandas 사용

[6]:

```

1 import threading
2 import pandas as pd
3
4 df = pd.DataFrame(columns=["title"])

```

[7]:

```

1 def get_article(page):
2     driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
3     driver.get(
4         "http://news.naver.com/main/main.nhn?mode=LSD&mid=shm&sid1=10
5     articles = driver.find_elements(By.CSS_SELECTOR, '#section_body li
6
7     for article in articles:
8         title = article.find_element(By.CSS_SELECTOR, 'dt:not(.photo)
9         df.loc[len(df)] = {
10             "title": title,
11         }
12
13     driver.quit()

```

[8]:

```

1 %%time
2 for page in range(1, 5):
3     get_article(page)

```

C:\Users\student\AppData\Local\Temp\ipykernel_12844\3566123820.py:2: DeprecationWarning: executable_path has been deprecated, please pass in a Service object

```
driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
```

C:\Users\student\AppData\Local\Temp\ipykernel_12844\3566123820.py:2: DeprecationWarning: executable_path has been deprecated, please pass in a Service object

```
driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
```

C:\Users\student\AppData\Local\Temp\ipykernel_12844\3566123820.py:2: DeprecationWarning: executable_path has been deprecated, please pass in a Service object

```
driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
```

C:\Users\student\AppData\Local\Temp\ipykernel_12844\3566123820.py:2: DeprecationWarning: executable_path has been deprecated, please pass in a Service object

```
driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
```

CPU times: total: 344 ms

Wall time: 15.8 s

[9]:

```
1 df.tail()
```

	title
75	아바타 옷 만들어 판매..."메타버서 돈 버네"
76	MMORPG DNA 입은 '아키월드', 기존 P2E와 무엇이 다르나
77	삼성전자 또 해킹 당해...美 고객정보 유출
78	'강제 퇴근' '출근 금지'...'한남노' 상륙에 재택 서두르는 IT업계
79	[취재수첩] 美 IRA로 드러난 '칩4'와 '팹4' 속 숨은 위험

[]:

```
1
```