

```

1 from selenium import webdriver
2 from selenium.webdriver.chrome.service import Service as ChromeService
3 from selenium.webdriver.chrome.options import Options as ChromeOptions
4 from selenium.webdriver.common.by import By
5 from webdriver_manager.chrome import ChromeDriverManager
6 from selenium.common.exceptions import NoSuchElementException,
  StaleElementReferenceException
7 import pandas as pd
8 import re
9 import time
10 import datetime
11
12 category = ['Finance', 'Securities', 'Industry', 'Venture_company', 'Real_estate', '
  Global_economy', 'Living_economy', 'General_economy']
13 sections = [259, 258, 261, 771, 260, 262, 310, 263]
14 options = ChromeOptions()
15 user_agent = 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like
  Gecko) Chrome/120.0.0.0 Safari/537.36'
16 options.add_argument('user_agent=' + user_agent)
17 options.add_argument('lang=ko_KR')
18
19 service = ChromeService(executable_path=ChromeDriverManager().install())
20 driver = webdriver.Chrome(service=service, options=options)
21
22 df_titles = pd.DataFrame()
23 re_title = re.compile('[^가-힣]')
24 for i, section in enumerate(sections):
25     url = 'https://news.naver.com/breakingnews/section/101/{}'.format(section)
26     titles = []
27     driver.get(url)
28     time.sleep(0.5)
29     for n in range(100):
30         try:
31             driver.find_element('xpath', '//*[@id="newsct"]/div[2]/div/div[2]/a').click()
32             time.sleep(0.2)
33         except:
34             print('driver.get', category[i], n)
35     title_tags = driver.find_elements(By.CLASS_NAME, 'sa_text_strong')
36     for title_tag in title_tags:
37         try:
38             title = re_title.sub(' ', title_tag.text)
39             titles.append(title)
40         except:
41             print('re compile miss')
42     df_section_title = pd.DataFrame(titles, columns=['titles'])
43     df_section_title['category'] = category[i]
44     df_section_title.to_csv('./crawling_data/data_{}.csv'.format(category[i]), index=False)
45     df_titles = pd.concat([df_titles, df_section_title], axis=0, ignore_index=True)
46
47 df_titles.to_csv('./crawling_data/naver_news_economy_data.csv', index=False)
48 print(df_titles.head())
49 print(df_titles["category"].value_counts())
50 print(df_titles.info())
51
52 driver.close()
53

```