과제1_0503.

네이버 로그인 후 첫번째 메일 제목 가져와서 출력하세요.(개인정보 관련 내용 은 제외)

```
In [8]:
    import time
 2
    import selenium
 3
   from selenium import webdriver
   from selenium.webdriver.common.by import By
   from selenium.webdriver.support.ui import WebDriverWait
   from selenium.webdriver.support import expected_conditions a
   path = 'c:/tool/chromedriver.exe'
8
9
   driver = webdriver.Chrome(path)
   driver.get('https://www.naver.com')
10
   driver.maximize_window()
11
12
   time.sleep(1)
13
   element = driver.find_element(By.CLASS_NAME, 'link_login')
   element.click()
   id = 
15
   pw = II
16
17
18
   # 'id'라는 아이디를 가진 요소의 값은, id 변수에 저장된 값으로
   # execute_script 함수를 사용하면 자바스크립트 코드를 실행
19
20
   # is는 getElementById('d')/value로 값을 가져올 수 있다.
21
22
   driver.execute_script("document.getElementById('id').value=\|
23
   time.sleep(1)
   driver.execute_script("document.getElementById('pw').value=₩
24
25
   time.sleep(1)
26
   element = driver.find_element(By.CLASS_NAME, 'btn_login')
27
28
   element.click()
   # print(driver.page source)
29
30
31
32
   # 이메일 가져오기
33
   driver.find_element(By.XPATH, '//*[@id="NM_FAVORITE"]/div[1]
34
35
   # 요소 로딩까지 웨이트 주기
36
   try:
37
       element = WebDriverWait(driver,5).until(
       EC.presence_of_element_located((By.XPATH, '//*[@id="mail.
38
39
40
   finally:
41
       pass
42
43
   # 첫 번째 메일 제목 가져오기
   driver.find_element(By.XPATH, '//*[@id="mail_list_wrap"]/ul/
44
C:\Users\h\AppData\Loca|\Temp\ipykernel_7700\h\19896
06467.py:9: DeprecationWarning: executable_path ha
s been deprecated, please pass in a Service object
  driver = webdriver.Chrome(path)
```

과제2_0503,

url 날 때 : 사 준 ws. haver.com/ Fash 출 haif-in-in? mode=LSD&mid=shm&sid1=100' (https://news.naver.com/main/main.nhn? mode=LSD&mid=shm&sid1=100') 사이트에서id가 'footer'인 한글만 출력하세요.

```
In [13]:
   from bs4 import BeautifulSoup
   import re
3
   import requests as rq
5
   url = 'https://news.naver.com/main/main.naver?mode=LSD&mid=s
7
   headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64
8
  reg = rg.get(url, headers=headers)
9
   bs = BeautifulSoup(reg.content. 'html.parser')
10
11
12
   # footer는 한 개 밖에 없음
   print(' '.join(re.findall('[가-힣]+', bs.find(id='footer').t
13
14
  # findall쓰면 for문 돌려야 함
15
  for i in bs.find_all(id='footer'):
16
      print(' '.join(re.findall('[가-힣]+', i.text)))
17
이용약관 서비스 안내 기사배열 원칙 책임자 유봉석
청소년 보호 책임자 이희만 개인정보처리방침 책임의
한계와 법적고지 뉴스 고객센터 본 콘텐츠의 저작권은
제공처 또는 네이버에 있으며 이를 무단 이용하는 경
우 저작권법 등에 따라 법적책임을 질 수 있습니다
이용약관 서비스 안내 기사배열 원칙 책임자 유봉석
청소년 보호 책임자 이희만 개인정보처리방침 책임의
한계와 법적고지 뉴스 고객센터 본 콘텐츠의 저작권은
제공처 또는 네이버에 있으며 이를 무단 이용하는 경
우 저작권법 등에 따라 법적책임을 질 수 있습니다
```

과제3_0503,

url = 'https://news.naver.com/main/main.nhn? mode=LSD&mid=shm&sid1=100' (https://news.naver.com/main/main.nhn? mode=LSD&mid=shm&sid1=100') 사이트에서p태그가 있는 데이터를 5개만 출력하세요

```
In [14]:
    from bs4 import BeautifulSoup
 2
    import requests as rq
 3
   url = 'https://news.naver.com/main/main.naver?mode=LSD&mid=s
 6
   headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64
8
   reg = rg.get(url, headers=headers)
9
   bs = BeautifulSoup(reg.content, 'html.parser')
10
   for i in bs.find_all('p', limit=2):
11
       print(i.text, '\mathbf{W}n')
12
AiRS추천으로 구성된 뉴스를 제공합니다.
 오후 5시~오후 6시까지 집계한 결과입니다.
```

과제4 0503.

url = '<u>https://news.daum.net/politics#1' (https://news.daum.net/politics#1')</u> 사 이트에서 '정치'를 출력하세요.

```
In [15]:

1 import requests as rq
from bs4 import BeautifulSoup

url = 'https://news.daum.net/politics#1'

headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64)

req = rq.get(url, headers=headers)
bs = BeautifulSoup(req.text, 'html.parser')

bs.select_one('#gnbContent > div > ul > li.on > a > span').te

'정치'
```

과제5 0503.

```
url = 'https://news.naver.com/main/main.naver?
mode=LSD&mid=shm&sid1=100'에서
(https://news.naver.com/main/main.naver?
mode=LSD&mid=shm&sid1=100'%EC%97%90%EC%84%9C) 뉴스 제목 10개를 가져오세요(사이트의 뉴스와 일치)
```

```
In [16]:
 1
2
   from selenium import webdriver
3
   from selenium.webdriver.common.by import By
5
   url = 'https://news.naver.com/main/main.naver?mode=LSD&mid=s
6
   driver = webdriver.Chrome('C:/tool/chromedriver.exe')
7
8
9
   driver.get('https://news.naver.com/main/main.naver?mode=LSD&
10
   headlines = driver.find elements(Bv.XPATH. '//*[@id="main co
11
12
13
   for headline in headlines[:10]:
14
      print(headline.text)
C:\Users\h\AppData\Loca|\Temp\ipykernel_7700\foral15287
58378.py:6: DeprecationWarning: executable_path ha
s been deprecated, please pass in a Service object
  driver = webdriver.Chrome('C:/tool/chromedriver.
exe')
비상장 주식 다단계 판매 적발 60대 이상 피해자, 5천
억 뜯겨
5천억원대 불법 다단계 투자업체 적발...피해자 4만6
 천 여 명
 장외주식 불법 다단계 판매로 수천억 손해 입힌 일당
 적 발
 "큰 돈 벌게 해드릴게"...피해액만 5천억, 60대 노
린 다단계 일당
엠폭스 확진 50명 넘었다…8일부터 고위험군도 예방접
 엠폭스 감염자 94%는 '2040 남성'…'익명 성접촉' 비
 중 높아
엠폭스 누적 52명 중 50명이 남성… "익명 성접촉 감
염 많아"
엠폭스 국내 환자 96%가 남성…"익명 성접촉 감염 많
OF"
송영길 전 대표 보좌관 소환· "돈 봉투 모른다"
 검찰, '돈봉투 의혹' 송영길 측근 보좌관 소환 조사
```

과제6 0503.

상기에서 출력된 데이터를 news.txt 파일로 저장하세요.

```
from selenium import webdriver
2
   from selenium.webdriver.common.by import By
3
   with open('news.txt', 'w') as f:
       f.write('')
5
6
7
   url = 'https://news.naver.com/main/main.naver?mode=LSD&mid=s
8
9
   driver = webdriver.Chrome('C:/tool/chromedriver.exe')
10
   driver.get('https://news.naver.com/main/main.naver?mode=LSD&
11
12
   headlines = driver.find_elements(By.XPATH, '//*[@id="main_co
13
14
15
   for headline in headlines[:10]:
       print(headline.text, file=open('news.txt', 'a', encoding
16
C:\Users\h\AppData\Loca|\Temp\jetapkernel_7700\35897
25264.py:9: DeprecationWarning: executable_path ha
s been deprecated, please pass in a Service object
  driver = webdriver.Chrome('C:/tool/chromedriver.
exe')
```

과제7 0503.

네이버 등 자주 사용하는 웹페이지에서 가치있다고 생각할 수 있는 데이터를 가져와서 활용할 수 있는 형태로 출력하세요.(최소 1000개 이상)

```
In [18]:
   # 네이버 증권 리서치 - 종목분석 리포트 (BeautifulSoup 사용)
 2
 3
   import re
   import requests as rq
   from bs4 import BeautifulSoup
 7
   url = 'https://finance.naver.com/research/company_list.naver
   headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64
8
9
   reg = rg.get(url, headers=headers)
10
11
12
   count = 0
13
   total\_count = 0
14
15
   print('<종목명 / 분석내용제목 / 분석증권사 / 작성일 / 조회수
   print('='*90. '\n')
16
17
18
   for i in range(1, 36):
       url = 'https://finance.naver.com/research/company_list.n
19
20
       reg = rg.get(url, headers=headers)
21
       bs = BeautifulSoup(req.text, 'html.parser')
       for j in bs.find('table').find_all('td'): # table 태그
22
          match = re.findall('[^\W]+', j.text)
23
                                          # 깔끔한 출력 위히
24
           if match:
              print(' '.join(match))
25
26
                                               # 데이터 개4
              count += 1
27
              if count % 5 == 0:
28
                  print()
29
30 print(f'총 {int(count/5)} 개의 종목 분석 데이터임', '₩n')
<종목명 / 분석내용제목 / 분석증권사 / 작성일 / 조
 회수 순임>
기업은행
선제적 충당금 적립으로 기대치 하회
DS투자증권
23 05 03
463
에코프로비엠
비싼 데는 이유가 있지만
DS투자증권
23 05 03
2166
```

```
In [21]:
   # 네이버 증권 리서치 - 종목분석 리포트 (selenium 사용)
   # 페이지를 35를 클릭해서 데이터를 가져오는 동적 행위를 반복히
 3
   from selenium import webdriver
5
   from selenium.webdriver.common.bv import Bv
 6
7
   path = 'c:/tool/chromedriver.exe'
8
9
   driver = webdriver.Chrome(path)
10
11
   count=0
12
   print('<종목명 / 분석내용제목 / 분석증권사 / 작성일 / 조회수
13
   print('='*90. '\n')
14
15
   for i in range(1,35):
16
17
       url = f'https://finance.naver.com/research/company_list.
18
       driver.get(url)
19
       texts = driver.find_elements(By.CLASS_NAME, 'type_1')
20
       for text in texts:
21
           for td in text.find_elements(By.TAG_NAME, 'td'):
              match = re.findall('[^\WW]+', td.text)
22
23
              if match:
                                                # 깔끔한 출
24
                  print(td.text)
25
                  count += 1
26
                  if count % 5 == 0:
27
                     print()
28
29 print(f'총 {int(count/5)} 개의 종목 분석 데이터임'. '₩n')
2070
 효성ITX
 꾸준한 실적, 저평가 매력
SK증권
23.03.29
 1679
 한온세스템
 1Q23 Preview: 아직은 냉기가 감도는 실적
 유진투자증권
23.03.29
 1407
 총 1020 개의 종목 분석 데이터임
```