# ===== 뉴스 감성분석 데이터 넣기 =====

# train set

#구글 크롤링

from selenium import webdriver

from urllib.parse import quote\_plus

# 검색할 페이지 수 range(start, end)

# 참고 영상 <https://www.youtube.com/watch?v=3b_VMk3WlNY>

# 웹드라이버 각자 설정

driver = webdriver.Edge('./msedgedriver.exe')

pages = 3

a = ["삼성전자", "LG전자"]

titles = dict.fromkeys(a)

first\_flag = True

for i in a:

query\_word = i + " 주가 뉴스"

titles[i] = []

baseUrl = 'https://www.google.com/search?q='

url = baseUrl + quote\_plus(query\_word)

driver.get(url)

sleep(0.5)

# 최초 검색 시에만 기간을 지정하고 뉴스 탭을 클릭하도록 합니다

if first\_flag:

# 도구를 클릭합니다

driver\_obj = driver.find\_element\_by\_css\_selector("#hdtb-tls")

driver\_obj.click()

sleep(0.5)

# 모든 날짜를 클릭합니다

driver\_obj = driver.find\_element\_by\_css\_selector("#hdtbMenus > span:nth-child(3) > g-popup > div.rIbAWc > div")

driver\_obj.click()

sleep(0.5)

# 기간 설정을 클릭합니다

driver\_obj = driver.find\_element\_by\_css\_selector("#lb > div > g-menu > g-menu-item:nth-child(8) > div > div")

driver\_obj.click()

# 찾을 시작할 날짜를 지정합니다

# from date

driver\_obj = driver.find\_element\_by\_css\_selector("#OouJcb")

driver\_obj.click()

driver\_obj.send\_keys("20101201")

# 찾을 마지막을 날짜를 지정합니다

# to date

driver\_obj = driver.find\_element\_by\_css\_selector("#rzG2be")

driver\_obj.click()

driver\_obj.send\_keys("20101231")

sleep(0.5)

# 기간 입력 후 실행을 클릭합니다

driver\_obj = driver.find\_element\_by\_css\_selector("#T3kYXe > g-button")

driver\_obj.click()

sleep(0.5)

first\_flag = False

# 뉴스를 클릭합니다

driver\_obj = driver.find\_element\_by\_css\_selector("#hdtb-msb > div:nth-child(1) > div > div:nth-child(3) > a")

driver\_obj.click()

sleep(0.5)

# 크롤링 할 페이지 수에 대한 loop

for \_ in range(pages):

# 각 기사 title 에 대한 css 참조 샘플

# rso > div:nth-child(1) > g-card > div > div > a > div > div.iRPxbe > div.mCBkyc.JQe2Ld.nDgy9d

# rso > div:nth-child(2) > g-card > div > div > a > div > div.iRPxbe > div.mCBkyc.JQe2Ld.nDgy9d

# 기사가 있는 body 부분을 선택

driver\_obj = driver.find\_elements\_by\_css\_selector("#rso > div:nth-child(n) > g-card > div > div > a > div > div.iRPxbe > div.mCBkyc.JQe2Ld.nDgy9d")

# 기사 수 만큼 제목을 추출한 후 기업 딕셔너리에 리스트로 저장합니다

for j in driver\_obj:

titles[i].append(j.text)

# 다음 페이지 설정

nextpage = driver.find\_element\_by\_css\_selector('#pnnext')

# 다음 페이지 클릭

nextpage.click()

sleep(0.5)

print("데이터 수집 완료")

# article2 = dataframe(titles)

# article2.to\_excel('./구글크롤링.xlsx', encoding='utf-8')

driver.close()

titles["삼성전자"]

titles["LG전자"]