[Python] python 에서 Seleium을 통한 동적 크롤링 - 4 — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2020-10-27 오전 8:32

URL: https://continuous-development.tistory.com/115?category=736681

Python

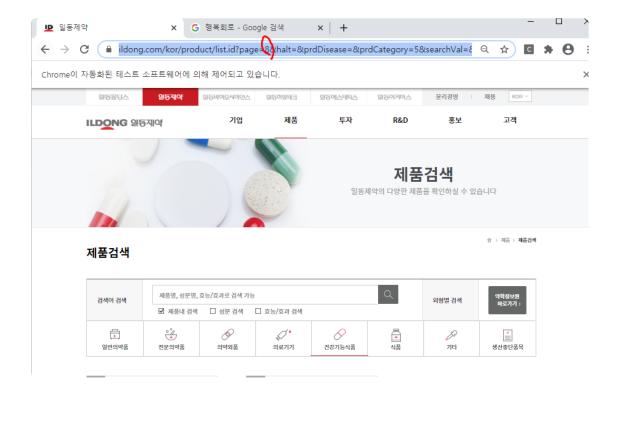
[Python] python 에서 Seleium을 통한 동적 크롤링 - 4

2020. 10. 8. 17:11 수정 삭제 공개

일동제약 상세 페이지 크<u>롤</u>링 하기

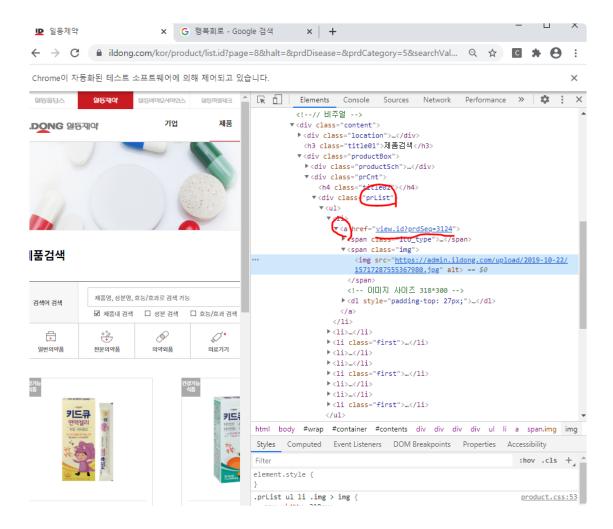
In [1]: from selenium import webdriver
import time
path = './chromedriver.exe'
driver = webdriver.Chrome(path)

기본적인 함수를 넣고 확인해보고 페이지에 따라 이동하는 것을 구하기 위해서 아래와 같이 페이지가 바뀌게끔 url을 만든다.

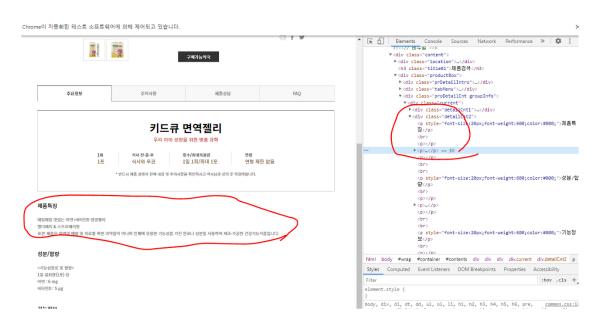


In [4]: for page in range(1,10):
 url = 'https://www.ildong.com/kor/product/list.id?page='+str(page)+'&halt=&prdDisease=&prdCategory=5&searchVal=&searchOption=0'
 driver.get(url)

해당 페이지에 상세페이지로 이동하여 클릭하고 안에 있는 데이터를 긁어오고 나오려 한다.



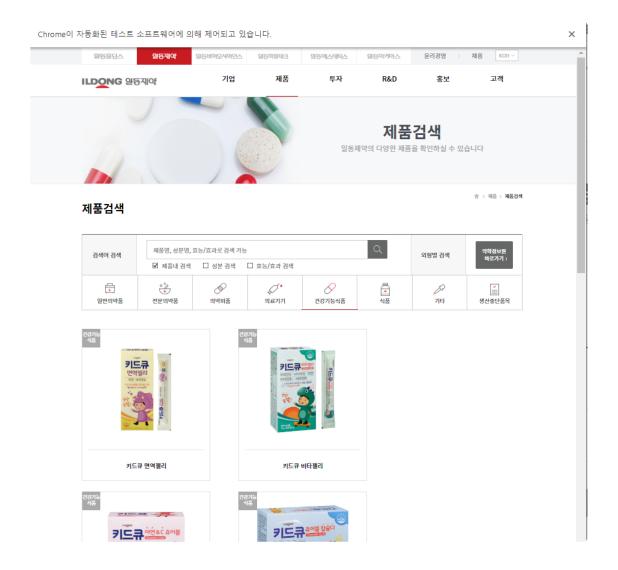
그러기 위해선 해당 링크를 타고 가기 위해 어디로 가야 하는지 경로를 확인하고



상세 페이지에서 필요한 부분이 어떤것인지 확인한다.

```
In [57]: details_list = []
for page in range(1,9):
    url = 'https://www.ildong.com/kor/product/list.id?page='+str(page)+'&halt=&prdDisease=&prdCategory=5&search
    driver.get(url)
    prList = driver.find_element_by_class_name('prList')
    a_tag = prList.find_elements_by_tag_name('a') # 해당 페이지로 이동하기 위한 a tag를 찾는다.
    for idx in range(0,9):
        a_tag[idx].click() # 이 a_tag를 index 순서대로 클릭하고
        detail = driver.find_element_by_class_name('detailCnt2') # 내가 필요로 하는 값을 가져온다.
        driver.back() # 기존될 클리어(초기화)한다.
        prList = driver.find_element_by_class_name('prList') # 그다음 위와 같이 다시 값을 가져오는 작업을 한다.
        a_tag = prList.find_element_by_tag_name('prList') # 그다음 위와 같이 다시 값을 가져오는 작업을 한다.
        a_tag = prList.find_element_by_tag_name('a')
        time.sleep(2) # 2초 간격으로 진행한다.
```

이걸 진행하면 자동화 프로그램에서 깜빡거리면서 페이지가 바뀌고 크롤 링하는 모습을 볼 수 있다.



'Python' 카테고리의 다른 글□

[Python] Numpy를 통한 배열 indexing(Boolen indexing, fancy indexing) 🗆

[Python] Numpy에 대한 기초 정리와 사용법 정리□

[Python] python 에서 Seleium을 통한 동적 크롤링 - 4□

[Python] python 에서 Seleium을 통한 동적 크롤링 - 3□

[Python] python 에서 Seleium을 통한 동적 크롤링 - 2□

[Python] python 에서 Seleium을 통한 동적 크롤링 - 1□

python Seleium

python Seleium 크롤링

python 동적 크롤링



나무늘보스

혼자 끄적끄적하는 블로그 입니다.