

학습 정리

팀	이진조	구성원	이규진, 조현진
---	-----	-----	----------

일정	발제자	주제
06/04	조현진	파이썬 & 데이터베이스 연동(TinyDB, SQLite3, MySQL & MairaDB)

주요 내용 요약

섹션3 : 파이썬 고급 스크래핑

requests 통신 실습(로그인 처리) 고급(2) - 위시켓(Wishket)

- 보안 토큰(csrf Token), Fake User-agent, Header Payload 처리
 - csrf : 파라미터
 - csrf token : 임의로 계속 바뀐다.
- fake user agent 설치
: pip install fake-useragent
- 위시켓(wishket) 사이트 로그인 처리 후 정보 가져오기

```
import sys
import io
from bs4 import BeautifulSoup
import requests
from fake_useragent import UserAgent
```

```
sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding = 'utf-8')
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding = 'utf-8')
```

요청 URL

```
URL = 'https://www.wishket.com/accounts/login/'
```

Fake User-Agent 생성

```
ua = UserAgent()
```

```
# print(ua.ie)
```

```
# print(ua.chrome)
```

```
# print(ua.random)
```

```

with requests.Session() as s:
    # Login 정보 Payload
    LOGIN_INFO = {
        'identification': 'ID',
        'password': 'PW',
        'csrfmiddlewaretoken': s.cookies['csrftoken']
    }

    print('token', s.cookies['csrftoken'])
    # 요청
    response = s.post(URL, data=LOGIN_INFO,
headers={'User-Agent':str(ua.chrome),
'Referer':'https://www.wishket.com/accounts/login/'})
    # HTML 결과 확인
    # print('response', response.text)
    if response.status_code == 200 and response.ok:
        soup = BeautifulSoup(response.text, 'html.parser')
        projectList = soup.select("table.table-responsive > tbody > tr")
        for i in projectList:
            print(i.find('th').string, i.find('td').text)

```

웹 브라우저 없는 스크래핑 및 파싱 실습(1)

- Selenium 개념 및 설치(크롬, 파이어폭스)
 - Selenium + 크롬
 - Selenium + 파이어폭스(권장)
 - Selenium + PhantomJS
 - Selenium 설치 : pip install selenium
- PhantomJS 개념 및 설치
 - : pip install selenium
- Selenium을 이용한 사이트 이미지 캡처 테스트 및 예제 작성 (전체)


```

import sys
import io
from selenium import webdriver

```

```

sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding = 'utf-8')
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding = 'utf-8')

```

```

driver = webdriver.PhantomJS('c:/section3/webdriver/Phantomjs/Phantomjs')

```

```

driver.implicitly_wait(5)

```

```

driver.get("https://google.com")

```

```

driver.save_screenshot("c:/website1.png")

```

```

driver.implicitly_wait(5)

```

```
driver.get('https://www.daum.net')
```

```
driver.save_screenshot("c:/website2.png")
```

```
driver.quit()
```

```
print('스크린샷 완료')
```

- Selenium을 이용한 인프런(Inflern) 로그인 자동화 처리하기

```
import sys
```

```
import io
```

```
import time
```

```
from selenium import webdriver
```

```
from selenium.webdriver.chrome.options import Options
```

```
sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding = 'utf-8')
```

```
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding = 'utf-8')
```

```
chrome_options = Options()
```

```
#chrome_options.add_argument("--headless")
```

```
chrome_options.add_argument('--log-level=3')
```

```
#driver = webdriver.Chrome(chrome_options=chrome_options,  
executable_path=r'C:/section3/webdriver/chrome/chromedriver')
```

```
driver = webdriver.Chrome('C:/section3/webdriver/chrome/chromedriver')
```

```
driver.set_window_size(1920, 1280)
```

```
driver.implicitly_wait(3)
```

```
driver.get('https://www.inflern.com/wp-login.php?redirect_to=https%3A%2F%2Fw  
ww.inflern.com%2F')
```

```
time.sleep(5)
```

```
driver.implicitly_wait(3)
```

```
driver.find_element_by_name("log").send_keys("id")
```

```
driver.implicitly_wait(3)
```

```
driver.find_element_by_name('pwd').send_keys('pw')
```

```
driver.implicitly_wait(3)
```

```
# 로그인
```

```
driver.find_element_by_xpath('//*[@id="wp-submit"]').click()
```

- Selenium & Chrome or Firefox Headless Mode 최종 세팅

웹 브라우저 없는 스크래핑 및 파싱 실습(2)

- Selenium을 이용한 네이버 카페 자동으로 글쓰기

```
import sys
```

```

import io
import time
from selenium import webdriver
from selenium.webdriver.chrome.options import Options

sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding = 'utf-8')
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding = 'utf-8')

class NcafeWriteAtt:
    # 초기화 실행(webdriver 설정)
    def __init__(self):
        chrome_options = Options()
        chrome_options.add_argument("--headless") # CLI (User-agent)
        self.driver =
webdriver.Chrome(chrome_options=chrome_options,executable_path="c:sections
3/webdriver/chrome/chromedriver")
        self.driver.implicitly_wait(5)

    # 네이버 카페 로그인 && 출석 체크
    def writeAttendCheck(self):
        self.driver.get('https://nid.naver.com/nidlogin.login')
        self.driver.find_element_by_name("id").send_keys('네이버 ID')
        self.driver.find_element_by_name("pw").send_keys('비밀번호')

self.driver.find_element_by_xpath('//*[@id="frmNIDLogin"]/fieldset/input').click()
        self.driver.implicitly_wait(30)
        self.driver.get('글을 작성하고자하는 url')
        self.driver.implicitly_wait(30)
        self.driver.switch_to_frame('cafe_main')
        self.driver.find_element_by_id('cmtinput').send_keys('반갑습니다!!')
        self.driver.find_element_by_xpath('글 입력버튼 xpath')
        time.sleep(3)

    # 소멸자
    def __del__(self):
        # self.driver.close() # 현재 실행 포커스 된 영역을 종료
        self.driver.quit() # selenium 전체 프로그램 종료
        print("Removed driver Object")

    # 실행
    if __name__ == '__main__':
        # 객체 생성
        a = NcafeWriteAtt()
        # 시작 시간
        start_time = time.time()
        # 프로그램 실행
        a.writeAttendCheck()
        # 종료 시간 출력
        print("---Total %s seconds ---" % (time.time() - start_time))
        # 객체 소멸

```

del a

- 네이버 로그인 처리 후 회원 리스트 정보 가져오기

```
import sys
import io
import time
```

```
from selenium import webdriver
from selenium.webdriver.chrome.options import Options
from bs4 import BeautifulSoup
```

```
sys.stdout = io.TextIOWrapper(sys.stdout.detach(), encoding = 'utf-8')
sys.stderr = io.TextIOWrapper(sys.stderr.detach(), encoding = 'utf-8')
```

```
class NcafeMemberExr:
```

```
    # 초기화 실행(webdriver 설정)
```

```
    def __init__(self):
        chrome_options = Options()
        chrome_options.add_argument("--headless") # CLI (User-agent)
        self.driver =
```

```
webdriver.Chrome(chrome_options=chrome_options,executable_path="c:sections
3/webdriver/chrome/chromedriver")
        self.driver.implicitly_wait(5)
```

```
    # 네이버 카페 회원 1페이지 정보 리스트 추출
```

```
    def getMemberlist(self):
        self.driver.get('https://nid.naver.com/nidlogin.login')
        self.driver.find_element_by_name("id").send_keys('네이버 ID')
        self.driver.find_element_by_name("pw").send_keys('비밀번호')
```

```
self.driver.find_element_by_xpath('//*[@id="frmNIDLogin"]/fieldset/input').click()
        self.driver.get('회원정보가 있는 URL')
        self.driver.implicitly_wait(30)
        self.driver.switch_to_frame('cafe_main')
        self.driver.implicitly_wait(5)
        self.driver.switch_to_frame('innerframe')
        soup = BeautifulSoup(self.driver.page_source,'html.parser')
        print(soup.prettify)
        return soup.select('div.mem_wrap > div.mem_list div.ellipsis.m-tcol-c')
```

```
    # 네이버 회원 리스트 출력 및 저장
```

```
    def printMemberList(self,list):
        f = open("c:/memberList.txt",'wt')
        for i in list:
            f.write(i.string.strip()+"\n")
            print(i.string.strip())
        f.close()
```

```
    # 소멸자
```

```
    def __del__(self):
        # self.driver.close() # 현재 실행 포커스 된 영역을 종료
```

```
self.driver.quit() # selenium 전체 프로그램 종료  
print("Removed driver Object")
```

```
# 실행  
if __name__ == '__main__':  
    # 객체 생성  
    a = NcafeMemberExr()  
    # 시작 시간  
    start_time = time.time()  
    # 프로그램 실행  
    a.printMemberList(a.getMemberlist())  
    # 종료 시간 출력  
    print("---Total %s seconds ---" % (time.time() - start_time))  
    # 객체 소멸  
    del a
```

- Class Module 실습 작성
- 자주 사용하는 Selenium 레퍼런스(Reference) 문서 훑어보기
<http://selenium-python.readthedocs.io/index.html>