

<Orbi 과외시장 데이터 분석>

독어교육과 2014-19498 정은주/ 인류학과 2016-19715 송지은

- 1. selenium chrome driver를 이용, orbi 페이지에 아이디와 비밀번호를 전송하여 로그인
- 2. 서울 지역의 학생 정보 선택, 모든 페이지(2019.12.15 기준 371페이지)를 자동으로 돌며 BeautifulSoup을 이용하여 각 페이지 크롤링
- 3. 학생 성별/ 지역(거주구)/ 학년 및 계열(문과/이과/기타)를 기준으로 나타나는 분포 양상을 데이터 분석하여 시각화하였습니다.

*** selenium의 경우 chromedriver 설치여부와 경로에 따라 실행이 되지 않을 수도 있습니다.

- 송지은: selenium 이용한 driver 제어, BeautifulSoup 이용한 웹크롤링 및 데이터 분석 알고리즘 설계
- 정은주: dictionary를 이용한 데이터 정렬, Pandas와 Matplotlib을 이용한 데이터 시각화

Orbi 과외시장

준회원

송지은 님

정회원등록

쪽지함

로그아웃

관심등록

이용안내

과외시장 홈

과외학생찾기

과외선생님찾기

선생님 자유게시판

마이페이지

과외학생 검색

검색어

검색어를 입력해 보세요.

희망지역

☒ 서울

☐ 경기

☐ 부산

☐ 대구

☐ 인천

☐ 광주

☐ 대전

☐ 울산

☐ 세종

☐ 강원

☐ 충북

☐ 충남

☐ 전북

☐ 전남

☐ 경북

☐ 경남

☐ 제주

희망과목

☐ 전체

☐ 수학

☐ 영어

☐ 국어

☐ 과탐

☐ 사탐

☐ 논술 (인문)

☐ 논술 (자연)

☐ 기타

학생 성별

전체

검색하기

강00	남	서울 노원구	재수/N수생 이과	수학, 영어, 국어, 과탐 과외	★
정00	여	서울 강북구	고1 문과	기타 과외	★
강00	남	서울 양천구	고3 이과	영어 과외	★
정00	여	서울 송파구	고1 문과	수학, 영어, 국어, 사탐 과외	★
장00	여	서울 강남구	대학생 문과	수학, 영어 과외	★
이00	남	서울 광진구	대학생 문과	기타 과외	★
강00	여	서울 서초구	중1 이과	수학, 과탐 과외	★
하00	여	서울 강남구	재수/N수생 문과	국어 과외	★
주00	남	서울 구로구	고3 이과	기타 과외	★
한0	남	서울 강북구	재수/N수생 이과	수학 과외	★
김00	여	서울 노원구	고1	수학, 과탐 과외	★

In [15]:

```
from bs4 import BeautifulSoup
from selenium import webdriver
import time
```

```
url = "https://login.orbi.kr/login/tutor?url=/oauth/authorize%3Fresponse_type%3Dcode%26client_id%3Dtutor%26redirect_uri%3Dhttps%253A%252F%252Ftutor.orbi.kr%252Flogin%252Fcomplete%26scope%3Dall%26state%3Dhttps%253A%252F%252Ftutor.orbi.kr%252Fsearch%252Fstudents"
```

```

driver = webdriver.Chrome("/Users/user/Desktop/chromedriver")

# chromedriver가 설치된 경로
driver.implicitly_wait(3)
driver.get(url)

driver.find_element_by_name("username").send_keys("wldms5764")
driver.find_element_by_name("password").send_keys("ekzm7732!")
driver.find_element_by_class_name("submit").click()

find_students = driver.find_element_by_xpath("//span[@class='menu-icon menu-icon-find-student']")
find_students.click()

login_button = driver.find_element_by_xpath("//a[@class='btn-login orbi']")
login_button.click()

seoul = driver.find_element_by_xpath("//input[@value='11']")
seoul.click()

search_button = driver.find_element_by_xpath("//input[@class='search-button']")
search_button.click()

areas = []
gender_list = []
grade_route_list = []

def web_crawling():
    html = driver.page_source
    soup = BeautifulSoup(html, "html.parser")

    student_table = soup.find("table", class_="teacher-list")
    student_info = student_table.find_all("tr")

    for info in student_info[1:]:
        data = info.find_all("td")
        gender_list.append(data[1])

    for info in student_info[1:]:
        data = info.find_all("td")
        grade_route_list.append(data[3])

    areas_data = soup.find_all("span", class_="area")
    areas = list(areas_data)
    return areas

# for grade route in grade route list:

```

```
# print(grade_route.text)

areas = web_crawling()

fp = open("orbi_areas.txt", 'w')
for area in areas:
    fp.write(area.text)
    fp.write("\n")
fp.close()

fp = open("orbi_gender.txt", 'w')
for gender in gender_list:
    fp.write(gender.text)
    fp.write("\n")
fp.close()

fp = open("orbi_grade_route.txt", 'w')
for grade_route in grade_route_list:
    fp.write(grade_route.text)
    fp.write("\n")
fp.close()

found_element = True
while found_element:
    areas = []
    gender_list = []
    grade_route_list = []

    try:
        next_button = driver.find_element_by_xpath("//a[@class
='btn next']")
        next_button.click()
        time.sleep(1)

        areas = web_crawling()

        fp = open("orbi_areas.txt", 'a')
        for area in areas:
            fp.write(area.text)
            fp.write("\n")
        fp.close()

        fp = open("orbi_gender.txt", 'a')
        for gender in gender_list:
            fp.write(gender.text)
            fp.write("\n")
        fp.close()
```

```
fp = open("orbi_grade_route.txt", 'a')

for grade_route in grade_route_list:
    fp.write(grade_route.text)
    fp.write("\n")
fp.close()

except Exception:
    found_element = False
```

In []: