**<데이터베이스 과제1,2>**

**<과제1: Target url을 crawling 할 수 있는 crawler 만들기>**

1. html 저장

- url 및 filename 입력 시 html로 저장하는 코드 (교수님 스켈레톤 코드 형식)

- 대상페이지: ‘https://www.luck-d.com/’ (럭키 드로우)

- 페이지 선정 이유: 신발 및 옷에 관심 有

2. html 저장(count)

- url, filename, count 입력 시 해당 수만큼 html로 저장하는 코드(교수님 스켈레톤 코드 형식)

- 대상페이지: "https://www.luck-d.com/release-raffle/” (럭키 드로우 래플 페이지)

- 페이지 선정 이유: 기간이 끝나지 않은 현재 진행중인 래플 현황 및 디테일 사진 저장

- 소스 코드 설명 :

1.beautifulsoup을 이용하여 target url 파싱 후, 현재 진행중인 raffle 이미지 .find()를 이용하여 탐색 (해당 사이트에선 class=”gallery\_horizontal”)

2.해당 url 리스트에 저장

3.순서대로 base\_url + sub\_url 들을 count수만큼 저장

4.해당 사이트에선 하이퍼링크로 저장되어있기에, href. 탐색

5.html 파일 로 저장

**<과제2: Target url에 login한page를 crawling할 수 있는 crawler 만들기>**

- target\_url, filename, id, password 입력 시 로그인 한 페이지를 크롤링 하여 저장하는 코드(교수님 스켈레톤 코드 형식)

- 테스트 페이지: 네이버(로그인)

- 페이지 선정 이유: KREAM 또는 LUCK-D로 하고싶었지만, id, pw 칸을 찾아도 자동으로 입력되지 않기에, 튜토리얼이 존재하는 네이버로 선정

- 직면한 문제: 실행 시 로그인 및 개인정보 입력되지만, 자동로그인 방지로 로그인이 완전하게 되진않음 + html파일 저장돼도 본래의 네이버페이지와 다른 형식

- 소스 코드 설명 :

1. 교수님 스켈레톤 코드 참고하여 과제1과 같이 crawl\_login 함수 작성

2. 힌트인 selenium 이용하여 크롬 브라우저를 통한 webdriver 받기

3. find\_element\_by\_name 함수를 이용하여 ‘id’ 및 ‘pw’의 키워드에 해당하는 키들을 자동 입력

4. 개발자 도구 들어가서 login 창의 xpath 받기

5. 저장될 파일이름,target\_url,아이디,비밀번호 입력 후 실행.