

Project Guidelines

Instructor: Eunil Park (eunilpark@skku.edu)



SUNG KYUN KWAN
UNIVERSITY



Project Guidelines

1. Final Project

개요

- 제목: 웹페이지 크롤러 만들기
- 내용: 수업시간에 배웠던 BeautifulSoup를 활용하여, 실제 웹페이지를 크롤링하고 그 결과를 파일에 저장하는 프로젝트를 수행함.(셀레늄 등과 같은 기타 라이브러리도 활용 가능)
- 제출물
 - 실행 예와 선정한 웹사이트명, 분류 정의 등을 포함한 간단한 리포트 1부(자유 양식)
 - 대상 크롤러 파일 1개 (XX.py 형태)
 - 출력결과물 파일 1개 (.csv 형태)
- 제출 기한 (5월 27일 저녁 11시 59분까지)

01. Final Project

Site List

순번	내용	url
1	잡코리아 회사별 합격자들 평균 스펙	http://www.jobkorea.co.kr/starter/spec/
2	뉴욕타임스	https://www.nytimes.com/section/universal/ko
3	학교 장학 페이지 정보	https://www.skku.edu/skku/campus/skk_comm/notice06.do
4	와인 구매 후기 사이트 vivino	https://www.vivino.com/
5	트게더 우울증 커뮤니티	https://tgd.kr/depressive
6	k 리그 포털 경기 별 매치 요약	http://data.kleague.com/
7	모두닥 병원 별 리뷰	https://www.mododoc.com/q/?search_query=%EC%B9%98%EA%B3%BC
8	마켓컬리 베스트 상품	https://www.kurly.com/shop/goods/goods_list.php?category=029
9	Snopes Fact check 포스트	https://www.snopes.com/fact-check/
10	Buzzfeed 사이트 뉴스기사	https://www.buzzfeed.com
11	성대 뉴스	https://www.skku.edu/skku/campus/skk_comm/news.do
12	잡플래닛 네이버 기업 리뷰 1~5페이지	https://www.jobplanet.co.kr/companies/42217/reviews/%EB%84%A4%EC%9D%B4%EB%B2%84
13	잡플래닛 데이터 분석 공고	https://www.jobplanet.co.kr/job_postings/search?rs_act=browse&rs_con=job&rs_element=job_postings&occupation_level2_ids%5B%5D=11613

개 요

- 주요 기능
 1. 학생이 고른 웹페이지를 입력받을 수 있어야 함.
 2. 입력 받은 웹페이지를 본인 나름대로 정한 고유한 정의를 통해 분류하고 이를 CSV파일로 저장하여 export 할 수 있어야 함.
 3. 출력 CSV파일의 이름은 현재 날짜와 시간이 포함된 CSV파일이어야 함
예: 현재 시간이 2022년 5월 9일 오후 3시 20분일 경우 – 202205091520.csv
 4. 완료 후에는 크롤링 결과와 소요 시간을 안내해주어야 함.
- 주의 사항
 1. 해당 크롤러의 경우, 각 line이 어떠한 역할과 동작을 하는 지 주석을 쓸 것.

01. Final Project

실행 예

웹사이트 주소: <https://movie.naver.com/movie/sdb/rank/rmovie.nhn>
'202005091520.csv' 파일에 크롤링 결과가 저장되었습니다.

소요 시간: 5초

시작시간: 2022년 5월 09일 15시 20분 52초

완료시간: 2022년 5월 09일 15시 20분 57초

계속하시겠습니까? 네

웹사이트 주소: <https://movie.naver.com/movie/sdb/rank/rmovie.nhn>
'202005091521.csv' 파일에 크롤링 결과가 저장되었습니다.

소요 시간: 5초

시작시간: 2022년 5월 09일 15시 21분 9초

완료시간: 2022년 5월 09일 15시 20분 14초

계속하시겠습니까? 아니오

안녕히 가십시오

CSV 예

순위,영화명,변동방향,폭
1,프리즌 이스케이프,na,o
2,더 플랫폼,na,o
3,위대한 쇼맨,up,1
4,언더워커,up,3
5,카페 벨에포크,down,2

The screenshot shows the Naver Movie Ranking page. The left sidebar contains navigation links: 영화출, 상영작·예정작, 영화랭킹 (selected), 랭킹, 디렉토리, 예매, 평점·리뷰, 다운로드, and 인디극장. The main content area is titled '영화랭킹' and shows a table of movie rankings. The table has columns for '순위' (Rank), '영화명' (Movie Name), and '변동폭' (Change). The table is filtered by '평점순 (현재상영영화)' and '2020.05.28'. The movies listed are: 1. 프리즌 이스케이프, 2. 더 플랫폼, 3. 위대한 쇼맨, 4. 언더워커, 5. 카페 벨에포크, 6. 범탈, 7. #살아있다, 8. 초미의 관심사, 9. 날씨의 아이, and 10. 그집.

순위	영화명	변동폭
1	프리즌 이스케이프	- 0
2	더 플랫폼	- 0
3	위대한 쇼맨	↑ 1
4	언더워커	↑ 3
5	카페 벨에포크	↓ 2
6	범탈	↓ 1
7	#살아있다	↓ 1
8	초미의 관심사	↑ 1
9	날씨의 아이	↓ 1
10	그집	↑ 5



Thanks

Project Guidelines

Instructor: Eunil Park (eunilpark@skku.edu)

