

파이썬X웹크롤링 업무자동화 2주차 과제

반드시 2주차의 모든 동영상 강의 시청 후 스터디 방법에 대해서 pdf 파일을 참고해 이해하시기 바랍니다. 스터디 방법에 대한 내용은 깃허브(<http://github.com/etilelab/webcrawling>)의 위키를 참고해주시기 바랍니다.

소스코드 작성 후 깃허브->이슈->이슈작성에서 파일을 드래그 앤 드랍하면 소스코드 파일이 첨부됩니다.

과제 제출날짜는 **10월15일(일) 오후 11시50분** 까지입니다. 이후 제출되는 과제는 미제출로 간주합니다.

문제에 대한 이해를 높이기 위해 유튜브 문제소개영상을 준비중입니다. 단체톡방을 참고해주세요.

1. 사용자로부터 키워드(ex : 안철수)와 크롤링할 뉴스기사 갯수를 입력받고, 네이버 뉴스 검색 페이지에서 네이버뉴스링크를 크롤링해오시오. (Hint : <https://search.naver.com/search.naver?ie=utf8&where=news&query=안철수&start=19>)

2. 문제1번에서 크롤링해온 네이버뉴스링크(ex : <http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=001&aid=0009573718>)에 접속해 아이디,덧글달린시각,덧글내용을 크롤링해오시오. 크롤링 한 후 csv파일로 저장하시오. (이때 csv파일명은 키워드+현재시각으로 하시오. 각 열은 키워드,아이디,시각,덧글내용으로 설정하시오)

3. 문제2에서 저장한 csv파일을 불러와 덧글의 단어를 워드클라우드를 사용해 시각화하라