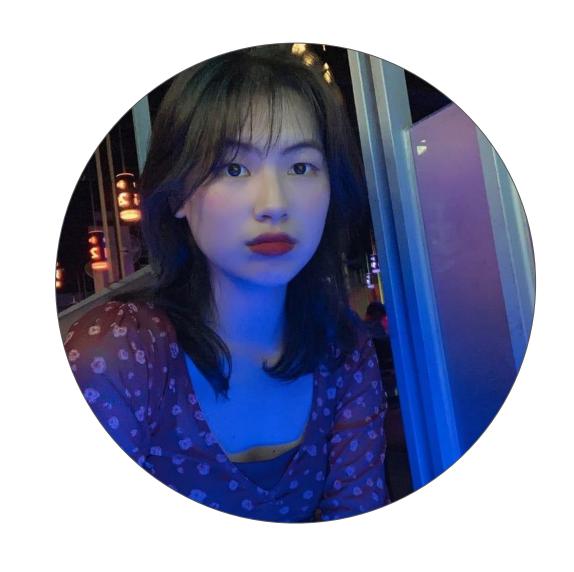
HUFS Missing Semester

웹 안의 스파이더맨, 크롤링





민선아 발표자/실습 진행자

영어통번역학부 / GBT학부 멋쟁이사자처럼



정지원 발표자/실습 진행자

독일어통번역학과 / 융복합소프트웨어 멋쟁이사자처럼



본 강의는 한국외국어대학교 AI교육원의 지원을 받아 제작되었음을 알려드립니다.

#0. 수료 조건



실습 참여도 + 실습 후 파일 제출 ②

실습 중 질문은 google meet의 **손들기** 🖐 를 사용해주세요. Slack Crawling 채널도 적극 활용해주세요! (상시 대기)

실습 파일 제출은 실습이 끝난 후, 아래 메일로 제출해주세요.

wony4094@likelion.org

sunaslight 1508@likelion.org



#1. 라이브러리

동적 크롤링 Urllib, Requests, Beautifulsoup4



urllib pip install urllib

- URL 관련된 작업을 수행하기 위해 필요한 파이썬 라이브러리
 - Request, parser, error와 같은 모듈을 포함한다.



#1.라이브러리

requests

pip install requests

HTTP 요청을 보내고 응답 받는 파이썬 라이브러리



#1.라이브러리

Beautifulsoup4

pip install beautifulsoup4

HTML, XML과 같은 문서에서 원하는 데이터를 가져오는 파이썬 라이브러리



Selenium

윈도우: pip install selenium Mac: pip3 install selenium

- 웹 사이트 테스트를 위한 도구
- 브라우저를 실제로 띄워서 사람처럼 동작하도록 만든다



#2. 그 외 속성

동적 크롤링 참고 method



명령어

Find_element_by_css_selector

CSS 선택자로 태그를 선택하는 명령어



#2. 그 외 속성

시크를 Window.scrollY

현재 스크롤된 높이를 알 수 있는 자바스크립트 명령어



실습에서 만나요! ⓒ