

교육 과정 소개서.

파이썬을 활용한 실전 웹크롤링과 자동화 CAMP



코스요약

코스명	파이썬을 활용한 실전 웹크롤링과 자동화 CAMP
기간	2020.09.12 - 2020.11.07 (휴강 10/3)
일정	매주 토요일 10:00 - 13:00
장소	패스트캠퍼스 강남강의장
준비물	개인 노트북
담당자	02-501-9396 / help-gathering@fastcampus.co.kr
수강료	800,000
상세페이지 url	fastcampus.co.kr/data_camp_wcr

코스목표

실전에서 오류원인을 빠르게 찾기위한 웹 이해, 이미지/영상 데이터 크롤링과 서버자동화 훈련을 통해 업무 효율을 높일 수 있습니다.

코스정보

쉽지 않은 웹크롤링, 웹 구조 이해와 크롤링 훈련이 필요합니다. 책에서는 배울 수 없었던 웹구조 이해부터 매주 2~3개의 사이트 크롤링 훈련 및 복습을 위한 퀴즈까지. 이제는 오류 없이 자유자재로 데이터를 뽑아내고 효율적으로 업무에 활용해 보세요.



코스특징

파이썬 문법 리뷰 및 웹 서비스 이해

본격적인 크롤링에 앞서, 파이썬 기본문법을 빠르게 리뷰하고, 실제 웹 페이지를 살펴보며 체계적으로 웹구조를 머릿속에 그려봅니다.

웹 페이지 분석 후, 원하는 데이터 크롤링

API로 데이터를 주고 받는 json 포맷과, 웹페이지 화면을 구성하는 HTML 포맷에서의 데이터 크롤링 방법을 학습합니다.

브라우저 제어 후, 이미지/파일 데이터 크롤링

Requests 패키지를 활용해 크롤링이 안되는 웹 페이지를 크롤링해보고, 로그인이 필요한 사이트, 이미지/영상 크롤링 방법도 익혀봅니다.

크롤링 자동화 /서버 활용법 마스터

크롤링 후 데이터 저장 및 개인PC없이 서버를 이용해 크롤링하는 방법, 주기적으로 크롤링하는 자동화 방법에 대해 학습합니다.



커리큘럼

<Part 1. 막막한 데이터 분석, 어떻게 진행해야 할까?>

- 1회차 ● **Python 문법 리뷰**
 - [이론] Jupyter Notebook 사용법, 컴퓨터 구조 및 프로그램 동작 원리, Python 문법 리뷰 : 변수선언, 자료형, 연산자, 조건문, 반복문, 함수, 클래스, 모듈과 패키지, 예외처리, 입출력, 정규표현식
 - [실습] 윤년 계산, 로또 번호 출력 프로그램, 스타크래프트 마린 매딕 객체 만들기, 문자열에서 전화번호 및 이메일 패턴으로 찾아내기

- 2회차 ● **Pandas패키지 학습/웹 서비스에 대한 이해와 웹 페이지 구조 이해**
 - [이론] 데이터 프레임 생성groupby, merge, pivot
 - [실습] 타이타닉 테이블 데이터 처리

<Part 2. 웹페이지 분석과 데이터 크롤링부터 시작해볼까요?>

- 3회차 ● **API로 데이터를 주고 받는 Json 포맷의 크롤링**
 - [이론] 웹 서비스 기초, 크롬 개발자도구 사용법, GET, POST 방식, 서버에 데이터 요청시 설정되는 header를 변경하여 크롤링
 - [실습] JSON 포맷의 네이버, 다음 주가 데이터 크롤링 및 시각화

- 4회차 ● **웹 페이지 화면을 구성하는 HTML 포맷의 크롤링**
 - [이론] HTML, CSS-Selector, 이미지 데이터 수집
 - [실습] 네이버 영화 평점 데이터 크롤링, 네이버 연관검색어 수집, 뉴스 기사 데이터 수집, G마켓 상품 데이터 및 이미지 수집



커리큘럼

- 5회차 ● **서비스의 API를 이용한 데이터 수집 및 자동화**
- [이론] 서비스 API의 사용법, OAuth 개념
- [실습] 네이버 번역 API 사용, 카카오 주소를 위도 경도로 변경하는 API 사용, 카카오톡으로 메시지 보내기, DART(기업공시) 데이터 크롤링
- 6회차 ● **Selenium을 이용한 브라우저 자동화**
- [이론] 웹페이지에서 원하는 데이터를 선택하는 방법인 xpath 학습, 멀티 태스킹을 이용한 빠른 크롤링 자동으로 빠르게 웹 크롤링을 해주는 scrapy 프레임워크의 사용 방법
- [실습] selenium 설치, 국민은행 금융상품 데이터 크롤링
- 7회차 ● **아마존 클라우드 서비스(AWS)를 활용한 데이터 수집 및 저장**
- [이론] 클라우드 서버 컴퓨터 AWS EC2의 사용법, crontab 스케줄러의 사용, MongoDB 데이터베이스 학습, AWS의 Cloud Watch, Lambda 서비스의 사용법
- [실습] AWS EC2 클라우드 서버 만들기, 뉴스 기사 데이터를 크롤링 후 mongodb에 저장
- 8주차 ● **개인 PC없이 서버에서 크롤링 자동화**
- [이론] slack의 incoming webhook, outgoing webhook 학습, 웹 서버 프레임워크 flask 학습, 이메일 전송 SMTP 서버의 이해
- [실습] - AWS EC2 클라우드 서버 만들기, PC 없이 서버에서 crontab 스케줄러를 이용한 파이썬 코드의 실행 번역 챗봇 만들기, 뉴스 기사를 크롤링한 데이터를 csv 파일로 저장 후 파일첨부로 이메일로 전송



강사소개



“안녕하세요, 박두진입니다. 단순 크롤링 방법을 가르쳐 드리기보단, 수집하고자 하는 특정 웹 사이트 혹은 데이터가 있을 때 디테일한 접근 방법을 가르쳐 드리려고 합니다. 피드백을 통해 실무에서 문제가 발생하더라도 스스로 원인을 찾을 수 있도록 안내해 드릴 계획입니다.

박두진

[약력]

시지온 웹서비스개발팀 팀장

소셜 댓글 라이브리 개발

국민은행 소셜 댓글 서비스 개발

댓글 통계 관리 서비스 개발

JLK-INSPECTION 인공지능팀 과장

스마트 팩토리 모니터링 서비스 개발

분류 AI 개발 및 서비스화

KOTRA K-Global 발표

N3N 웹 서비스팀 선임연구원

사물인터넷(IoT) 데이터 시각화

솔루션 위즈아이(Wizeye) 개발

베라노스 책임연구원

GBI(Goal-based Investment) 기반의 금융 솔루션 개발

삼성자산운용 GBI 시뮬레이터 개발

패스트캠퍼스

데이터 사이언스 스쿨 7기, 8기, Plus 프로그래밍 파트 강의 (540시간)



수강환경

패스트캠퍼스 강남강의장



❖ 강의에 따라 강의장이 변경될 수 있습니다.