

파이썬 스터디 활동보고서 DAY 7

기본 활동 정보

날짜 : 2018년 7월 20일 (금)

시간 : 6시 30분 - 9시

장소 : 국민대학교 북악관 카페테리아

참여 인원 : 4명

이성재 20132651 경영학 전공

우현웅 20153195 소프트웨어 전공

최승호 20142772 소프트웨어 전공

김선필 20143038 소프트웨어 전공

1. 스터디 과제 검사

강의 : Edwith 파이썬을 이용한 웹 스크래핑 챕터11 들어오기

과제 : 없음

2. 강의 리뷰

새롭게 알게된 내용들 정리

파트 1 : 정규식을 이용한 패턴 찾기

> import re 를 사용해서 정규식 관련 메소드를 활용할 수 있다.

> rstrip() 은 가장 오른쪽 공백을 제거하는 메소드이다. rstrip 내에 특정 문자를 작성하면, 가장 오른쪽의 특정 문자를 모두 제거한다. rstrip 은 오른쪽이며, lstrip 은 왼쪽에 대해 적용하는 메소드이다.

> re.search(regex, string) 는 regex 정규식을 string 에 대해 검사하여, 해당하는 문자가 있다면 True, 없다면 False 를 반환하는 메소드이다.

> open() 은 현재 wd 에 대해 특정 파일을 읽어오는데 쓰이는 메소드이다.

> \s 는 공백문자를, \S 는 비 공백문자를 의미한다.

> \$ 는 문자열의 끝을 의미한다.

> [a-z0-9] // [가-힣] 등은 특정 문자열

파트 2 : 정규식을 이용한 패턴 추출

> re.findall(regex, string) 은 string 에 대해 regex 를 만족하는 부분문자열을 모두 찾아서 리스트로 반환.

> 탐욕적 매칭 : 하나의 문자열에 대해 여러가지 패턴이 매칭될 경우, 가장 긴 문자열을 선택하여 반환한다.

> 비탐욕적 매칭 : 하나의 문자열에 대해 여러가지 패턴이 매칭될 경우, 가장 짧은 문자열을 선택하여 반환한다.

> 정규식은 기본적으로 탐욕적 매칭이며, 비탐욕적 매칭을 사용하기 위해서는 + 나 * 뒤에 ? 를 붙여야 한다.

> 소괄호 () 를 정규식 내에서 사용하면, 정규식으로 추출된 문자열의 특정 부분만 잘라내어 반환한다.

파트 3 : 정규식을 활용하는 다양한 방법

정규식 없이 문자열 추출하는 방법 : 빨찌

> @ 의 인덱스와 공백의 인덱스를 찾아서, 해당 길이만큼 추출한다.

> Split 메소드를 두 번 사용하여 추출한다.

정규 표현식을 활용하여 문자열 추출

> 대괄호 [] 내부의 ^ 헛 표시는 헛 표시 이후의 모든 문자열에 대해, 해당 문자열을 제외한 모든 문자열의 집합.

> 정규 표현식을 사용하여 추출된 자료는 모두 문자열이므로, 숫자로 사용하기 위해서는 형변환이 필요하다.

3. 다음시간 강의 / 과제

다음시간 : 7월 25일 (수)

그 이후의 스터디 : 7월 27일 / 7월 31일 / 8월 2일 / 8월 8일 / 8월 10일 ...

1) edwith 파이썬 웹 크롤링 챕터 11 파트 1 ~ 3 듣고 실습하기



< 2018년 7월 18일 스터디 DAY6 >