

2025-04-16(수)

Web scraping 2일차

- Naming Conventions and Best Practices for Class and Object Names in Python
- ★★ 파이썬 패키지 인덱스(PyPI) : 파이썬 패키지 관리 시스템
 - ★ Q_01_08. PIP(Python Package Index) 란 무엇인가 (wikidocs)
 - Q_01_09. PIP 명령문 사용하기
 - streamlit
 - A faster way to build and share data apps.
 - Streamlit documentation
 - selenium
 - API Docs
 - Python language bindings for Selenium WebDriver.
 - The selenium package is used to automate web browser interaction from Python.
 - ★ [Google chrome closes immediately after being launched with

```
options = webdriver.ChromeOptions()
options.add_experimental_option("detach", True)
driver = webdriver.Chrome(options=options)
```

- SQLAlchemy
 - SQLAlchemy 2.0 Documentation
 - ★ [Python] Sqlalchemy 구조 ORM, Core 에 대해
- Python
 - [Python] 생성자(**init**)와 소멸자(**del**)
 - 7.5. 특별한 메서드들
 - **init** 메서드 (초기화)
 - **del** 메서드 (소멸자)
 - **repr** 메서드 (프린팅)
 - **add** 메서드 (덧셈)
 - **lt** 메서드 (비교)
- ★ 파이썬으로 배우는 알고리즘 트레이딩 (개정판-2쇄)
 - 17. 추가 데이터 저장하기 (revision)
 - 1) SQLite

- 2) Pandas와 SQLite
 - 1) DataFrame 객체를 SQLite DB에 저장하기
 - 2) SQLite DB에서 테이블 로드하기
 - 3) Pandas를 이용한 주가 데이터 저장
 - Markdown 문법
 - 마크다운 - 표(테이블) 만들기
 - ★ 웹크롤링 - BeautifulSoup에서 find와 select 사용하기
 - Webcrawling을 해보자 🤖
 - Requests
 - BeautifulSoup
 - Selenium
 - ★★ Selenium 옵션과 이벤트 처리 코드 모음
 - ★ Python Crawling The Coding
 - 01 bs4 모듈 BeautifulSoup 객체
 - 02 BeautifulSoup - 기본 요소 사용
 - 03 BeautifulSoup - find(), find_all()
 - 04 BeautifulSoup - select_one(), select()
 - ★★ 사장님 몰래 하는 파이썬 업무자동화(부제 : 들키면 일 많아짐)
 - 3) 동적/정적 웹 페이지
 - 4) 정적 웹크롤링 - request / beautifulsoup
 - 6) 동적 웹크롤링 - selenium 소개 및 기초사용법
 - 7) 동적 웹크롤링 - selenium으로 화면 조작하기
 - 8) 동적 웹크롤링 - 텍스트, 하이퍼링크, 이미지 가져오기
 - ★ SQLAlchemy Query를 Pandas DataFrame로 만들기
 - SQLAlchemy를 활용하여 MySQL에 주가데이터 저장하기
 - Pandas to_sql() 메소드: 효율적인 SQL 작성을 위한 팁
-

api 명세

- spring swagger

[python] 공공 데이터 포털 API 불러오기

- [Python 실습 준비] 오픈API 활용하기(1) 오픈 API 요청키 생성
- [Python 실습 준비] 공공 데이터 오픈API 활용하기(1) 불러오기 및 저장하기
- [python] 공공 데이터 포털 API 불러오기

- **Crawling_[공공데이터 포털의 OPEN API]**

- [Python] 공공데이터 포털의 OPEN API 사용 방법
- [Python] 공공데이터 포털의 OPEN API 사용 방법(2)
- [python] 공공데이터 OPEN API의 xml 을 DataFrame으로 변환하기(feat. 코로나 확진자 수)

- 공공데이터포털 오픈 API 사용법 : 인증키 생성, XML 파일 접속

- 파이썬 - 공공데이터 포털 OpenAPI 사용방법 및 예제

```
import requests

Servicekey = "발급받은 일반인증키(Decoding) 입력"
base_url = "https://api.odcloud.kr/api"
url = "/15081901/v1/uddi:e86f20c4-44de-4839-bf4a-e53c23d361e2"
params = { 'serviceKey': Servicekey,
           'page':1,
           'perPage':10,
           'returnType': "json"
         }

response = requests.get(base_url + url, params=params)
response.json()
```

공공 데이터 포털 API > 기상청_단기예보 ((구)_동네예보) 조회서비스

- 기상청_단기예보 ((구)_동네예보) 조회서비스 (XML, JSON)

활용승인 절차 개발단계 : 자동승인 / 운영단계 : 자동승인
 신청가능 트래픽 개발계정 : 10,000 / 운영계정 : 활용사례 등록시 신청하면 트래픽 증가 가능
 요청주소
 http://apis.data.go.kr/1360000/VilageFcstInfoService_2.0/getUltraSrtNcst
 서비스URL http://apis.data.go.kr/1360000/VilageFcstInfoService_2.0

- 상세기능 목록

번호	API명(국문)	상세기능명(영문)	상세기능명(국문)
1	단기예보 조회서비스	getUltraSrtNcst	초단기실황조회
2	단기예보 조회서비스	getUltraSrtFcst	초단기예보조회
3	단기예보 조회서비스	getVilageFcst	단기예보조회
4	단기예보 조회서비스	getFcstVersion	예보버전조회

- 요청변수(Request Parameter)

항목명(국문)	항목명(영문)	항목크기	항목구분	샘플데이터	항목설명
서비스키	ServiceKey	4	필수	-	공공데이터포털에서 받은 인증키
페이지 번호	pageNo	4	필수	1	페이지번호
한 페이지 결과 수	numOfRows	4	필수	1000	한 페이지 결과 수
응답자료형식	dataType	4	옵션	XML	요청자료형식(XML/JSON) Default: XML
발표일자	base_date	8	필수	20210628	'21년 6월 28일 발표
발표시각	base_time	4	필수	0600	06시 발표(정시단위)
예보지점 X 좌표	nx	2	필수	55	예보지점의 X 좌표값
예보지점 Y 좌표	ny	2	필수	127	예보지점의 Y 좌표값

Streamlit

- [\[Streamlit\] 개발자 환경 구축: VS Code의 가상 환경에 Streamlit 설치하기](#)

selenium

- 실습코드
 - [launcher.py](#) - crawling sel & webscraping launcher
 - [crawling.py](#) - 공공 데이터 API 데이터 조회
 - [crawlingsel.py](#) - selenium 을 이용한 crawling
 - [webscraping.py](#) - beautifulsoup4 을 이용한 Web scraping
- ★★ Selenium 옵션과 이벤트 처리 코드 모음
- [\[selenium\] webdriver.ChromeOptions 활용](#)
- [\[코드잇\] Selenium 기능 더 알아보기\(select요소, 옵션, iframe, headless모드, robots.txt\)](#)
- ★ Google chrome closes immediately after being launched with selenium
- ★ [\[PYTHON\] 크롤링 - 크롤링 종료 \(close, quit, dispose 차이 비교\)](#)

함수	크롬 브라우저 종료	크롬 드라이버 종료	특징
close()	현재 실행중인 1개만 종료	X	1. 활성화 된 창만 닫기 때문에 현재 창을 닫고 새로운 창을 열어 재사용 가능 2. 드라이버가 종료된게 아니라 메모리를 계속 사용 3. 완전 종료 시 사용하면 안됨
quit()	두 종료	O	1. Selenium 전체 종료 2. 완전 종료시 사용
dispose()	모두 종료	O	1. quit()와 동일