

2025-04-15(화)

DataScience-master.zip 문서 참고

- [\[Markdown\] md 아이콘 사용하기 #markdown icon](#)
- **Python**
 - [Python 계단밧기](#)
 - [04. 삼항 연산자\(Ternary operators\)](#)
 - [\[Python\] 삼항 연산자 \(Tenary Operator\)](#)
 - [\[pandas\] 리스트를 데이터 프레임으로 변환하는 법](#)
 - [python trim string](#)
 - **clear()**
 - [Python에서 clear\(\) 사용해 터미널 지우는 법](#)
 - [python console 지우기](#)

```
>>> import os
>>> os.system('cls') # window
>>> os.system('clear') # linux, mac
```

- [\[Python\] 프로그램 재실행시 terminal clear하는 방법](#)

```
import os



def clear(): # clear 대신 함수명은 자유롭게 할 수 있다.
    os.system("cls") # 윈도우: clear() 함수가 호출되면 Terminal을
clear 시킨다.
    os.system("clear") # linux, mac: clear() 함수가 호출되면
Terminal을 clear 시킨다.

clear() # 함수 호출
```

Web Scraping

- 실습코드
 1. [webscraping.py](#)
 2. [launcher.py](#)
- ★★ 파이썬 패키지 인덱스(PyPI) : 파이썬 패키지 관리 시스템
 - ★ Q_01_08. PIP(Python Package Index) 란 무엇인가 (wikidocs)
 - [Q_01_09. PIP 명령문 사용하기](#)

- [tqdm](#)
 - Fast, Extensible Progress Meter
 - [함께해요 파이썬 생태계](#) ♦ [wikidocs](#)
 - [tqdm: 진행률 표시 도구](#)
- [requests](#)
 - [API Reference and User Guide](#)
 - Requests is a simple, yet elegant, HTTP library.
 - [HTTP 상태 코드\(resp.status_code\)](#)
 - [200 OK](#)
 - [400 Bad Request](#)
 - [404 Not Found](#)
 - [500 Internal Server Error](#)
 - 하이퍼텍스트 전송 프로토콜 (HTTP) 500 Internal Server Error 서버 에러 응답 코드는 요청을 처리하는 과정에서 서버가 예상하지 못한 상황에 놓였다는 것을 나타냅니다.
 - ★ [requests 라이브러리의 get 메서드의 headers 파라미터 중 **user-agent** 값을 넣어주면 해결 됨](#)
 - [How can I view HTTP headers in Google Chrome?](#)
 - [함께해요 파이썬 생태계](#) ♦ [wikidocs](#)
 - [3. 웹 크롤링](#)
 - [01. selenium 4](#)
 - [02. selenium 3](#)
 - [03. requests](#)
- [DOMParser](#)
 - [DOMParser](#)
 - [dom-parser](#) ♦ [nodejs](#)
 - [python xml.dom.minidom을 이용한 xml 파싱](#)
 - [xml.dom.minidom — Minimal DOM implementation](#)
 - [How to Parse HTML With Python \(Using The Top 6 Parsers\)](#)
- [beautifulsoup4](#)
 - [Beautiful Soup Documentation](#)
 - Beautiful Soup is a library that makes it easy to scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.
- [selenium](#)
 - [API Docs](#)
 - Python language bindings for Selenium WebDriver.
 - The selenium package is used to automate web browser interaction from Python.
- [XML-Sitemaps.com Sitemap Generator](#)
 - [사이트맵 \(Sitemap\)을 알아보기](#)

- [문과생도 할 수 있는] 파이썬 업무 자동화 일잘러 되기 + 챗GPT
 - 08. 웹 스크래핑  wikidocs
- ★ python으로 web scraping
 - 웹 크롤링(Web Crawling)
 - 자동화된 웹 크롤러로 검색 엔진을 통해 웹에서 필요한것을 찾아 자료들을 다운받고 인덱스라는 목차를 만들고 내가 원하는 정보를 메타정보를 기준으로 추출하여 분리하는 작업.
 - 자동화 봇(웹 크롤러)이 웹을 돌아다니면서 정보를 얻는 것
 - 웹 스크래핑(web scraping)
 - 웹 사이트 상에서 원하는 부분에 위치한 데이터를 추출하여 수집하는 기술
 - 원하는 특정한 정보를 웹에서 추출하는 것
 - HTTP를 통해 웹 사이트의 내용을 긁어다 원하는 형태로 가공하는 것으로 크롤링도 스크래핑에 포함된다고 보는 견해도 있다.
- ★ 비전공자를 위한 파이썬 자동화 완벽 가이드  wikidocs
 - 2.5 사이트 정보 가져오기 - requests 사용법
 - [Python] request 모듈 간단 사용법
 - 2.6 사이트 정보 추출하기 - beautifulsoup 사용법 (1)
 - 2.7 사이트 정보 추출하기 - beautifulsoup 사용법 (2)
 - 2.8 사이트 자동화하기 - selenium 사용법(1)
- ★ 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
- 웹크롤링 - BeautifulSoup에서 find와 select 사용하기
- ★★ Python Crawling The Coding
 - 01 bs4 모듈 BeautifulSoup 객체

디버깅 방법

- pdb
 - pdb — 파이썬 디버거
 - pdb — The Python Debugger
 - 파이썬 - 디버깅편(Pdb, breakpoint())
 - [내장 함수] 파이썬 내장 함수 breakpoint
 - [Python] 파이썬 디버거 (pdb)
 - ★ 점프 투 파이썬 - 라이브러리 예제 편
 - 110 만든 코드를 디버깅하려면? — pdb
 - pdb는 파이썬 코드를 디버깅할 때 사용하는 모듈이다. 이 모듈을 사용하면 중단점(breakpoint) 설정, 소스 리스팅, 변수 치환, 콜 스택 검사 등의 다양한 기능을 사용할 수 있다.
 - 실용 파이썬 프로그래밍: 프로그래밍 유경험자를 위한 강좌

■ 8.3 디버깅

■ breakpoint()

■ 파이썬 디버거

■ 디버거 진입(파이썬 3.7 이상)

```
def some_function():  
    ...  
    breakpoint()      # 디버거 진입(파이썬 3.7 이상)  
    ...
```

■ 3.7 이전버전

```
import pdb  
...  
pdb.set_trace()      # `breakpoint()` 대신 이것을  
사용하라  
...
```

■ 디버거에서 실행하기

■ bash % python3 -m pdb someprogram.py

• Logger

- ★ [logging — Logging facility for Python](#)
 - ★★ [\[Python\] Logging 사용법](#)
 - ★ [\[Python\] Logger 사용법 - 잡다한 AI 관련 글들 - 티스토리](#)
-