

파이썬 패키지 조사

1. Pandas

- 테이블 형태의 데이터 자료를 핸들링 하는데 많이 사용되는 패키지이다.
- 주로 엑셀이나 CSV형태의 파일을 불러와 사용하거나 output으로 내보낸다.
- 그 외에도 Data frame을 직접 만들 수도 있다.

2. numpy

- 수업시간에도 자주 사용되는 패키지로, 수치 해석이나 다차원 배열, 벡터 연산 등에 사용된다.
- 일반적으로 array처리를 하기 편하게 만들어 matrix연산이나 다차원 배열을 처리하기 좋다.
- 위 특성들에 따라 선형대수와 관련된 계산이나, 관련 알고리즘을 구성하는 코드에서 많이 사용된다.

3. Matplotlib

- 처리한 데이터들을 시각화 할 때 많이 사용되는 패키지이다.
- 차트 등의 그래프로 데이터 프레임의 정보들을 보여준다.

4. BeautifulSoup

- HTML 정보로부터 원하는 데이터를 가져오기 쉽게 parsing해주는 패키지이다.
- 주로 크롤링을 할 때, requests로 웹 페이지 정보를 불러오고 이를 parsing할 때 사용된다.

5. Requests

- 특정 웹사이트에 HTTP요청을 보내 HTML문서를 받아올 수 있는 패키지이다.
- 크롤링을 할 때 HTML정보를 받아오고 BeautifulSoup를 통해 데이터를 parsing한다.

6. Selenium

- 웹사이트 상에서 클릭 등의 이벤트를 자동으로 동작하게 만들어 주는 패키지이다.
- JS가 동적으로 만든 데이터를 크롤링 하거나, 반복적인 웹상의 업무를 자동화 하는 것이 가능하다.