

2024.10.19. 2장

◆ API

- 프로그램 간 데이터 전달 규칙
- 수동 반복작업 대체 → 공개 API 활용 → 데이터 수집 자동화 가능

◆ HTTP

- 웹 데이터 주고받는 프로토콜
- 웹페이지, 이미지 등 HTTP를 통해 전송됨
- URL: "http" 또는 보안 강화된 "https" 시작

◆ JSON

- 최근 많이 사용하는 데이터 포맷
- 가독성 우수, 변환 용이 (HTML, XML 대비)
- 웹 기반 API에서 대중적 사용

◆ XML

- JSON보다 장황하지만 계층적인 구조
- xml 패키지 사용 → 엘리먼트 탐색
- pandas.read_xml() → 데이터프레임으로 변환 가능

※ 핵심 함수/메서드 정리

- ◆ json.dumps(): 객체 → JSON 문자열 변환
- ◆ json.loads(): JSON 문자열 → 객체 변환
- ◆ pandas.read_json(): JSON 문자열 → Series/DataFrame 변환
- ◆ xml.etree.ElementTree.fromstring(): XML 문자열 → Element 객체 변환
- ◆ xml.etree.ElementTree.Element.findtext(): 첫 자식 엘리먼트의 텍스트 반환
- ◆ xml.etree.ElementTree.Element.findall(): 지정한 태그의 모든 자식 엘리먼트 반환
- ◆ requests.get(): URL로 HTTP GET 요청
- ◆ requests.Response.json(): 응답받은 JSON을 객체로 변환