



API를 활용한 데이터 수집

made by Jay Hong

API를 활용한 데이터 수집

1. API란?



1. API란?

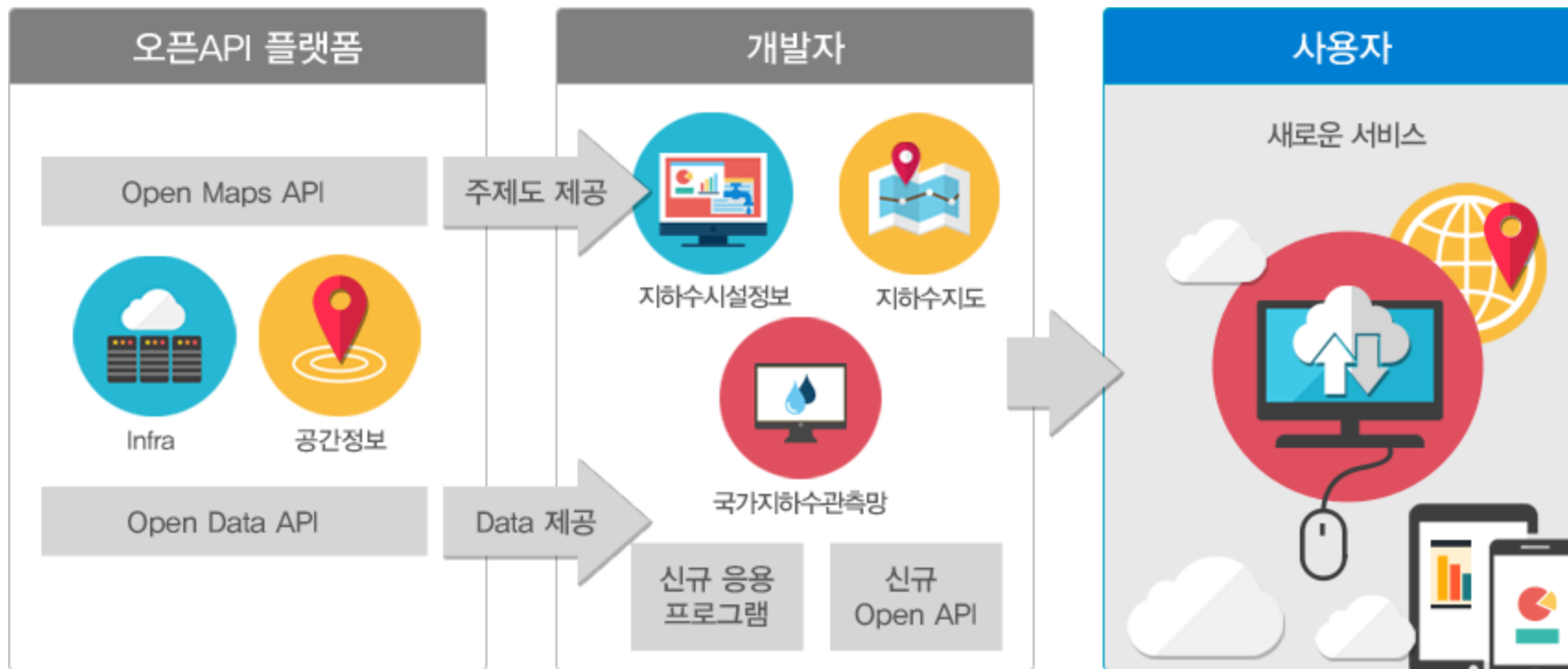
1. 오픈 API란?

2. XML

3. JSON

1) 오픈 API의 정의

- ✓ 서비스, 정보, 데이터 등을 어디서나 쉽게 이용할 수 있도록 개방된 API
- ✓ 다양한 Application을 만들 수 있도록 개발자와 사용자에게 공개



2) 오픈 API의 특징

- ✓ 데이터를 제어할 수 있는 간단하고 직관적인 Interface를 통해 사용자의 참여 유도
- ✓ 방화벽 문제로부터 자유로움
- ✓ 웹을 통한 데이터 수집 필요 X
- ✓ 서비스 제공자의 정책에 따라 데이터 제공 범위가 상이

사실상 기존의 Web Crawling은 거의 대부분의 사이트에선 법적인 접촉을 받는다.

하지만 오픈 API는 서비스 제공자와 사용자 사이의 합의를 통해 제공한다.

그 과정에서 데이터의 제공 범위는 달라질 수 있다.

3) 오픈 API로 데이터 수집

- ✓ 투표소 확인하기
- ✓ <https://www.data.go.kr/data/15000836/openapi.do>
- ✓ 해당 페이지에서, API 활용 신청
 - > 마이페이지 오픈 API
 - > 활용
 - > 일반 인증키 2개를 획득할 수 있다.

3) 오픈 API로 데이터 수집

✓ 다시 원래 페이지로 돌아와서 확인해보면, 이런 정보를 확인할 수 있다.

선거ID, 시도명, 구시군명을 입력받아 사전투표소명, 시도명, 위원회명,읍면동명, 오더순, 건물명, 주소, 층수등을 조회할 수 있는 사전투표소 정보 조회

- 활용승인 절차 개발단계 : 허용 / 운영단계 : 허용
- 신청가능 트랙픽 10,000 / 운영계정은 활용사례 등록시 신청하면 트랙픽 증가 가능
- 요청주소 <http://apis.data.go.kr/9760000/PolplcInfolnquireService2/getPrePolplcOtlmapTrnsportInfolnquire>
- 서비스URL <http://apis.data.go.kr/9760000/PolplcInfolnquireService2>

[활용신청](#)

요청변수(Request Parameter)

항목명(국문)	항목명(영문)	항목크기	항목구분	샘플데이터	항목설명
서비스키	ServiceKey	4	필수	-	공공데이터포털에서 받은 인증키
페이지 번호	pageNo	4	옵션	1	페이지번호
한 페이지 결과 수	numOfRows	4	옵션	10	한 페이지 결과 수
선거ID	sgId	10	필수	20220309	선거ID
시도명	sdName	40	필수	서울특별시	시도명
위원회명	wiwName	40	옵션	종로구	위원회명

✓ 옵션은 선택사항이며, 필수는 반드시 요청할 때 url과 함께 보내주어야 한다.

3) 오픈 API로 데이터 수집

✓ 다음과 같이 구성하면 된다.

<http://apis.data.go.kr/9760000/PolplcInfoInquireService2/getPrePolplcOtInmapTrnsportInfoInquire>?필수정보1=정보1%필수정보2=정보2

```
import requests
```

```
url = f'http://apis.data.go.kr/9760000/PolplcInfoInquireService2/getPrePolplcOtInmapTrnsportInfoInquire?serviceKey={encoding_key}&pageNo=1&nur  
res = requests.get(url)
```

```
res
```

```
<Response [200]>
```

```
res.text
```

```
'<?xml version="1.0" encoding="UTF-8"?><response>\n<header>\n<resultCode>INFO-00</resultCode>\n<resultMsg>NORMAL SERVICE</resultMsg>\n</header>\n<body>\n<items>\n<item>\n<num>1</num>\n<sgld>20220309</sgld>\n<psName>청운효자동제1투</psName>\n<sdName>서울특별시</sdName>\n<wiwName>종로구</wiwName>\n<emdName>청운효자동</emdName>\n<placeName>청운초등학교(2층, 체육관)</placeName>\n<addr>서울특별시 종로구 자하문로 105 (청운동)</addr>\n<floor>2층</floor>\n</item>\n<item>\n<num>2</num>\n<sgld>20220309</sgld>\n<psName>청운효자동제2투</psName>\n<sdName>서울특별시</sdName>\n<wiwName>종로구</wiwName>\n<emdName>청운효자동</emdName>\n<placeName>서울농학교(1층, 대강당)</placeName>\n<addr>서울특별시 종로구 필운대로 103 (신교동)</addr>\n<floor>1층</floor>\n</item>\n<item>\n<num>3</num>\n<sgld>20220309</sgld>\n<psName>청운효자동제3투</psName>\n<sdName>서울특별시</sdName>\n<wiwName>종로구</wiwName>\n<emdName>청운효자동</emdName>\n<placeName>청운효자동 자치회관(1층, 회의실)</placeName>\n<addr>서울특별시 종로구 자하문로19길 36 (옥인동)</addr>\n<floor>1층</floor>\n</item>\n<item>\n<num>4</num>\n<sgld>20220309</sgld>\n<psName>사직동제1투</psName>\n<sdName>서울특별시</sdName>\n<wiwName>종로구</wiwName>\n<emdName>사직동</emdName>\n<placeName>서울생활문화지원센터 체부(1층, 체부홀)</placeName>\n<addr>서울특별시 종로구 자하문로1나길 3-2 (체부동)</addr>\n<floor>1층</floor>\n</item>\n<item>\n<num>5</num>\n<sgld>20220309</sgld>\n<psName>사직동제2투</psName>
```


4) 최종 코드

```
encoding_key =
decoding_key =
```

```
import requests
```

```
page_No = 1
numOfRows = 10
sgld = 20220309 # 본투표
sdName = '서울특별시'
wiwName = '종로구'
```

```
url = f'http://apis.data.go.kr/9760000/PolplcInfoInquireService2/getPolplc0tInmapTrnsportInfoInquire#
?serviceKey={encoding_key}&#
&pageNo={page_No}&#
&numOfRows={numOfRows}&#
&sgld={sgld}&#
&sdName={sdName}&#
&wiwName={wiwName}'
res = requests.get(url)
```

```
res
```

```
<Response [200]>
```

```
res.text
```

```
'<?xml version="1.0" encoding="UTF-8"?><response>#<header>#<resultCode>INFO-00</resultCode>#<resultMsg>NORMAL SERVICE</resultMsg>#
</header>#<body>#<items>#<item>#<num>1</num>#<sgld>20220309</sgld>#<psName>청운효자동제1투</psName>#<sdName>서울특별시</sdName>#<wiwName>종로구</wiwName>#<emdName>청운효자동</emdName>#<placeName>청운초등학교(2층, 체육관)</placeName>#<addr>서울특별시 종로구 자하문로 105 (청운동)</addr>#<floor>2층</floor>#</item>#<item>#<num>2</num>#<sgld>20220309</sgld>#<psName>청운효자동제2투</psName>#<sdName>서울특별시</sdName>#<wiwName>종로구</wiwName>#<emdName>청운효자동</emdName>#<placeName>서울농학교(1층, 대강당)</placeName>#<addr>서울특별시 종로구 필운대로 103 (신교동)</addr>#<floor>1층</floor>#</item>#<item>#<num>3</num>#<sgld>20220309</sgld>#<psName>청운효자동제3투</psName>#<sdName>서울특별시</sdName>#<wiwName>종로구</wiwName>#<emdName>청운효자동</emdName>#<placeName>청운효자동 자치회관(1층, 회의실)</placeName>#</items>#</body>#</response>'
```

✓ response code가 200으로, 잘 수집 되었다.

5) 결과 해석

- ✓ 다시 data.go.kr의 해당 페이지를 접속하면,

XML 중앙선거관리위원회 사전 및 선거일 투표소 정보

사전투표소 및 선거일투표소등 투표소 정보를 제공하는 서비스 이다.

👍 0 🗨️ 0 📌 관심

[활용신청](#)[오류신고 및](#)[담당자 문의](#)

penAPI 정보

메타데이터 다운로드

분류체계	일반공공행정 - 국정운영	제공기관	중앙선거관리위원회
관리부서명	정보운영과	관리부서 전화번호	02-3294-1153
API 유형	REST	데이터포맷	XML
활용신청	169	키워드	투표,선거,투표소
등록	2018-06-04	수정	2022-02-04
심의유형	개발단계 : 허용 / 운영단계 : 허용		

- ✓ 데이터의 포맷이 “XML”임을 알 수 있다.
- ✓ res.text의 결과가 XML의 형태이므로, 이를 해석하기 위해 XML과 JSON에 대해서 이야기해보자.

1) XML이란?

- ✓ eXtensible Markup Language
 - ✓ HTML도 Markup Language로, HTML과 유사한 구조를 가진다.
 - ✓ 특징
 - ✓ 데이터의 구조와 의미를 설명하는데, 태그, 속성, 값으로 표현
 - ✓ 시작 태그와 종료 태그가 있고, 그 사이에 값이 있음
- ⇒ HTML과 거의 유사함

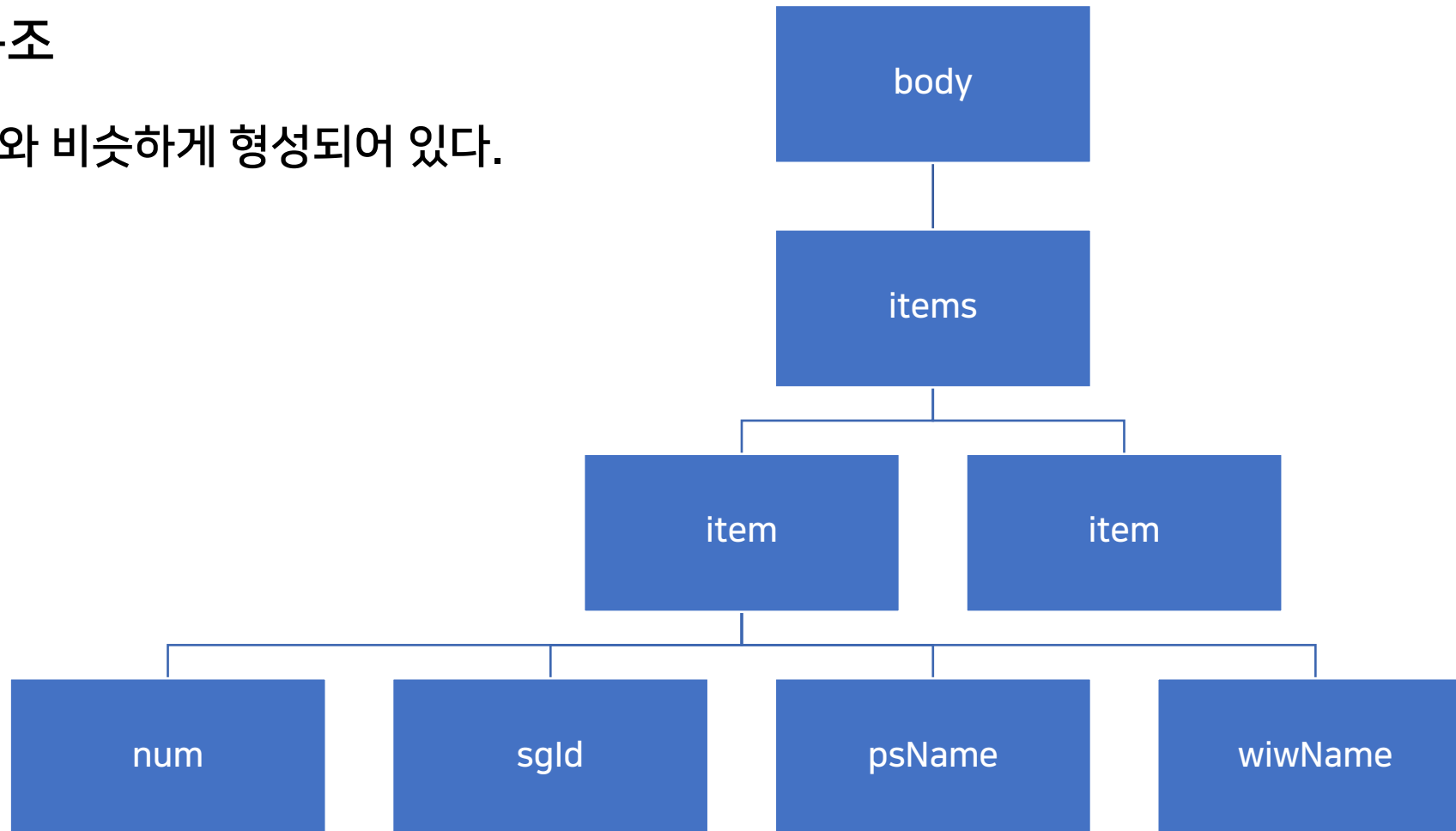
2) XML 구조

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<response>
  <header>
    <resultCode>INFO-00</resultCode>
    <resultMsg>NORMAL SERVICE</resultMsg>
  </header>
  <body>
    <items>
      <item>
        <num>1</num>
        <sgld>20220309</sgld>
        <psName>청운효자동제1투</psName>
        <sdName>서울특별시</sdName>
        <wiwName>종로구</wiwName>
        <emdName>청운효자동</emdName>
        <placeName>청운초등학교(2층, 체육관)</placeName>
        <addr>서울특별시 종로구 자하문로 105 (청운동)</addr>
        <floor>2층</floor>
      </item>
      <item>
        <num>2</num>
        <sgld>20220309</sgld>
        <psName>청운효자동제2투</psName>
        <sdName>서울특별시</sdName>
        <wiwName>종로구</wiwName>
        <emdName>청운효자동</emdName>
        <placeName>서울농학교(1층, 대강당)</placeName>
        <addr>서울특별시 종로구 필운대로 103 (신교동)</addr>
        <floor>1층</floor>
      </item>
    </items>
  </body>
</response>
```

2) XML 구조

✓ 나무와 비슷하게 형성되어 있다.



3) XML 해석

- ✓ Python에서는 이를 그냥 불러오면 해석하기 어렵다.

res.text

```
'<?xml version="1.0" encoding="UTF-8"?><response>\n<header>\n<resultCode>INFO-00</resultCode>\n<resultMsg>NORMAL SERVICE</resultMsg>\n</header>\n<body>\n<items>\n  <item>\n    <num>1</num>\n    <sgld>20220309</sgld>\n    <psName>청운효자동제1투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>청운효자동</emdName>\n    <placeName>청운초등학교(2층, 체육관)</placeName>\n    <addr>서울특별시 종로구 자하문로 105 (청운동)</addr>\n    <floor>2층</floor>\n  </item>\n  <item>\n    <num>2</num>\n    <sgld>20220309</sgld>\n    <psName>청운효자동제2투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>청운효자동</emdName>\n    <placeName>서울농학교(1층, 대강당)</placeName>\n    <addr>서울특별시 종로구 필운대로 103 (신교동) </addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>3</num>\n    <sgld>20220309</sgld>\n    <psName>청운효자동제3투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>청운효자동</emdName>\n    <placeName>청운효자동 자치회관(1층, 회의실) </placeName>\n    <addr>서울특별시 종로구 자하문로19길 36 (옥인동)</addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>4</num>\n    <sgld>20220309</sgld>\n    <psName>사직동제1투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>사직동</emdName>\n    <placeName>서울생활문화지원센터 체부(1층, 체부홀)</placeName>\n    <addr>서울특별시 종로구 자하문로1나길 3-2 (체부동) </addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>5</num>\n    <sgld>20220309</sgld>\n    <psName>사직동제2투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>사직동</emdName>\n    <placeName>광화문스페이스본 101동 북카페(1층)</placeName>\n    <addr>서울특별시 종로구 사직로8길 4 (사직동, 광화문풍림스페이스본)</addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>6</num>\n    <sgld>20220309</sgld>\n    <psName>삼청동투표소</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>삼청동</emdName>\n    <placeName>삼청동주민센터(1층, 민원실)</placeName>\n    <addr>서울특별시 종로구 삼청로 107 (삼청동)</addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>7</num>\n    <sgld>20220309</sgld>\n    <psName>무암동제1투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>무암동</emdName>\n    <placeName>무암동주민센터(1층, 민원실)</placeName>\n    <addr>서울특별시 종로구 창의문로 145 (무암동)</addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>8</num>\n    <sgld>20220309</sgld>\n    <psName>무암동제2투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>무암동</emdName>\n    <placeName>세검정초등학교(1층, 체육관)</placeName>\n    <addr>서울특별시 종로구 세검정로9길 1 (신영동) </addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>9</num>\n    <sgld>20220309</sgld>\n    <psName>무암동제3투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>무암동</emdName>\n    <placeName>상명대학교사범대학부속초등학교(1층, 도서관)</placeName>\n    <addr>서울특별시 종로구 홍지문2길 1 (홍지동) </addr>\n    <floor>1층</floor>\n  </item>\n  <item>\n    <num>10</num>\n    <sgld>20220309</sgld>\n    <psName>평창동제1투</psName>\n    <sdName>서울특별시</sdName>\n    <wiwName>종로구</wiwName>\n    <emdName>평창동</emdName>\n    <placeName>평창동주민센터(4층, 대강당)</placeName>\n    <addr>서울특별시 종로구 평창문화로 65 (평창동) </addr>\n    <floor>3층이상</floor>\n  </item>\n</items>\n<numOfRows>10</numOfRows>\n<pageNo>1</pageNo>\n<totalCount>43</totalCount>\n</body>\n</response>\n'
```

3) XML 해석

- ✓ 따라서 xmltodict library를 활용해 해석하자.
- ✓ xml의 구조상의 특징을 이용해 dictionary 형태로 바꿔주어 해석한다.

```
import xmltodict
```

```
xml_dict = xmltodict.parse(res.text)
```

```
xml_dict.get('response').get('body').get('items')
```

```
OrderedDict([('item',
OrderedDict([('num', '1'),
('sgld', '20220309'),
('psName', '청운효자동제1투'),
('sdName', '서울특별시'),
('wiwName', '종로구'),
('emdName', '청운효자동'),
('placeName', '청운초등학교(2층, 체육관)'),
('addr', '서울특별시 종로구 자하문로 105 (청운동)'),
('floor', '2층')]),
OrderedDict([('num', '2'),
('sgld', '20220309'),
('psName', '청운효자동제2투'),
('sdName', '서울특별시'),
('wiwName', '종로구'),
('emdName', '청운효자동'),
('placeName', '서울농학교(1층, 대강당)'),
('addr', '서울특별시 종로구 필운대로 103 (신교동)'),
('floor', '1층')]),
OrderedDict([('num', '3'),
('sgld', '20220309'),
('psName', '청운효자동제3투'),
('sdName', '서울특별시'),
('wiwName', '종로구'),
('emdName', '청운효자동'),
('placeName', '청운효자동 자치회관(1층, 회의실)'),
('addr', '서울특별시 종로구 자하문로19길 36 (옥인동)'),
('floor', '1층')])])])
```

4) XML의 현 상황

- ✓ XML은 과거 DB보다 좀 더 자유롭게 다루기 위해 만들어진 언어
- ✓ 이제는 JSON의 등장으로 중요성이 더 떨어짐
- ✓ 하지만 과거에 많이 쓰였기 때문에,
오래전에 만들어진 API나 기존 시스템에서는 여전히XML을 사용중

1) JSON

- ✓ Java Script Object Notation
- ✓ 자바 스크립트의 객체 표현 방식을 사용하여 데이터 교환의 표준으로 삼은 언어
- ✓ 특징
 - ✓ 사람, 기계 모두 이해하기 쉽다.
 - ✓ 데이터 용량이 적고 코드가 간결하여 XML을 대체하고 있다.
 - ✓ 최근의 웹은 JSON을 이용하여 정보를 교환한다.

```
import xmltodict

xml_dict = xmltodict.parse(res.text)

xml_dict.get('response').get('body').get('items')

OrderedDict([('item',
  [OrderedDict([('num', '1'),
    ('sgld', '20220309'),
    ('psName', '청운효자동제1투'),
    ('sdName', '서울특별시'),
    ('wiwName', '종로구'),
    ('emdName', '청운효자동'),
    ('placeName', '청운초등학교(2층, 체육관)'),
    ('addr', '서울특별시 종로구 자하문로 105 (청운동)'),
    ('floor', '2층')]),
  OrderedDict([('num', '2'),
    ('sgld', '20220309'),
    ('psName', '청운효자동제2투'),
    ('sdName', '서울특별시'),
    ('wiwName', '종로구'),
    ('emdName', '청운효자동'),
    ('placeName', '서울농학교(1층, 대강당)'),
    ('addr', '서울특별시 종로구 필운대로 103 (신교동)'),
    ('floor', '1층')]),
  OrderedDict([('num', '3'),
    ('sgld', '20220309'),
    ('psName', '청운효자동제3투'),
    ('sdName', '서울특별시'),
    ('wiwName', '종로구'),
    ('emdName', '청운효자동'),
    ('placeName', '청운효자동 자치회관(1층, 회의실)'),
    ('addr', '서울특별시 종로구 자하문로19길 36 (옥인동)'),
    ('floor', '1층')])])])
```

3) JSON의 활용


- ✓ 네이버 파파고 api
- ✓ <https://developers.naver.com/docs/papago/README.md>
- ✓ 네이버 개발자 -> 로그인 -> Application -> 내 application -> application 등록
-> 이름 설정 + Papago 번역 설정 -> 등록하기
- ✓ application의 Client ID와 Client Secret 확인

3) JSON의 활용

- ✓ 네이버 파파고 api
- ✓ <https://developers.naver.com/docs/papago/README.md>
- ✓ 네이버 개발자 -> 로그인 -> Application -> 내 application -> application 등록
-> 이름 설정 + Papago 번역 설정 -> 등록하기
- ✓ application의 Client ID와 Client Secret 확인

3) JSON의 활용

- ✓ 네이버 파파고는 다음의 파라미터를 전달해야 합니다.
- ✓ 기본적으로, client id와 client secret은 Header 의 형태로 전달해주어야 합니다.

파라미터 

파라미터	타입	필수 여부	설명
source	String	Y	원본 언어(source language)의 언어 코드
target	String	Y	목적 언어(target language)의 언어 코드
text	String	Y	번역할 텍스트. 1회 호출 시 최대 5,000자까지 번역할 수 있습니다.

3) JSON의 활용

✓ 최종 JSON으로 호출하는 코드

```
papago_id =
papago_key =
```

```
import urllib
```

```
s = 'ko' #시작 언어
t = 'en' #도착 언어
txt = 'API를 활용한 데이터 수집방법 - JSON으로 수집하기' # 번역할 text
txt = urllib.parse.quote(txt) # 컴퓨터의 언어로 encode
print(txt)

params = f"source={s}&target={t}&text={txt}" # 이 파라미터를 url의 끝에 붙이면 된다.
print(params)
```

```
API%EB%A5%BC%20%ED%99%9C%EC%9A%A9%ED%95%9C%20%EB%8D%B0%EC%9D%B4%ED%84%B0%20%EC%88%98%EC%A7%91%EB%B0%A9%EB%B2%95%20-%20JSON%EC%9C%BC%EB%A1%9C%20%EC%88%98%EC%A7%91%ED%95%98%EA%B8%B0
source=ko&target=en&text=API%EB%A5%BC%20%ED%99%9C%EC%9A%A9%ED%95%9C%20%EB%8D%B0%EC%9D%B4%ED%84%B0%20%EC%88%98%EC%A7%91%EB%B0%A9%EB%B2%95%20-%20JSON%EC%9C%BC%EB%A1%9C%20%EC%88%98%EC%A7%91%ED%95%98%EA%B8%B0
```

```
url = 'https://openapi.naver.com/v1/papago/n2mt'
print(url)
```

```
https://openapi.naver.com/v1/papago/n2mt
```

```
res = urllib.request.Request(url)
res.add_header("X-Naver-Client-Id", papago_id)
res.add_header("X-Naver-Client-Secret", papago_key)
response = urllib.request.urlopen(res, data=params.encode('utf-8'))
response.getcode()
```

3) JSON의 활용

- ✓ JSON 형태로 해석
- ✓ 딕셔너리 형태로 잘 불러와졌다.

```
import json
```

```
result = json.load(response)
```

```
result
```

```
{'message': {'@type': 'response',  
             '@service': 'naverservice.nmt.proxy',  
             '@version': '1.0.0',  
             'result': {'srcLangType': 'ko',  
                        'tarLangType': 'en',  
                        'translatedText': 'How to collect data using API - Collect with JSON',  
                        'engineType': 'N2MT',  
                        'pivot': None}}}
```

```
result_message = result.get('message').get('result').get('translatedText')  
result_message
```

'How to collect data using API - Collect with JSON'



실습

- ✓ Spotify developer 페이지에서 노래 데이터를 수집해보자.