

빅데이터 수집시스템 개발



공공 DB 데이터와 SNS 수집 실습

학습목표

- Open API를 사용하여 공공 DB의 수집 프로그램을 구현할 수 있다.
- Rest API와 SNS API를 사용하여 SNS에 올려진 글을 수집하는 프로그램을 구현할 수 있다.

학습내용

- Open API를 활용한 공공 DB 수집
- Rest API와 SNS API를 활용한 SNS 글 수집

Open API를 활용한 공공 DB 수집

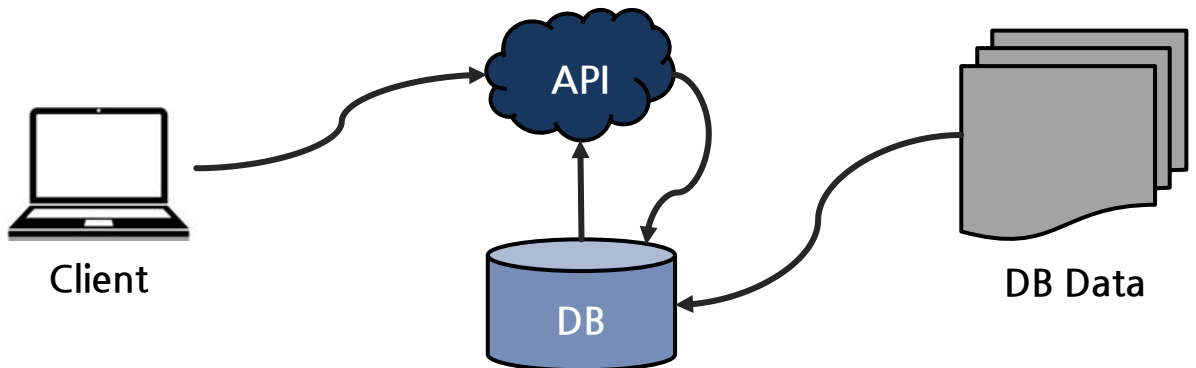


1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보 읽어오기

1 Open API와 Rest API

API

- Application Programming Interface의 약자
- 특정 프로그램을 만들기 위해 제공되는 모듈(함수 등)



Open API를 활용한 공공 DB 수집



1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보 읽어오기

1 Open API와 Rest API

Open API

- 누구나 사용할 수 있도록 공개된 API
- 주로 Rest API 기술을 많이 사용

Rest API

- Representational State Transfer API의 약자
- HTTP 프로토콜을 통해서 정보를 제공하는 함수
- 실질적인 API 사용은 정해진 구조의 URL 문자열 사용



일반적으로 XML, JSON의 형태로 응답 출력

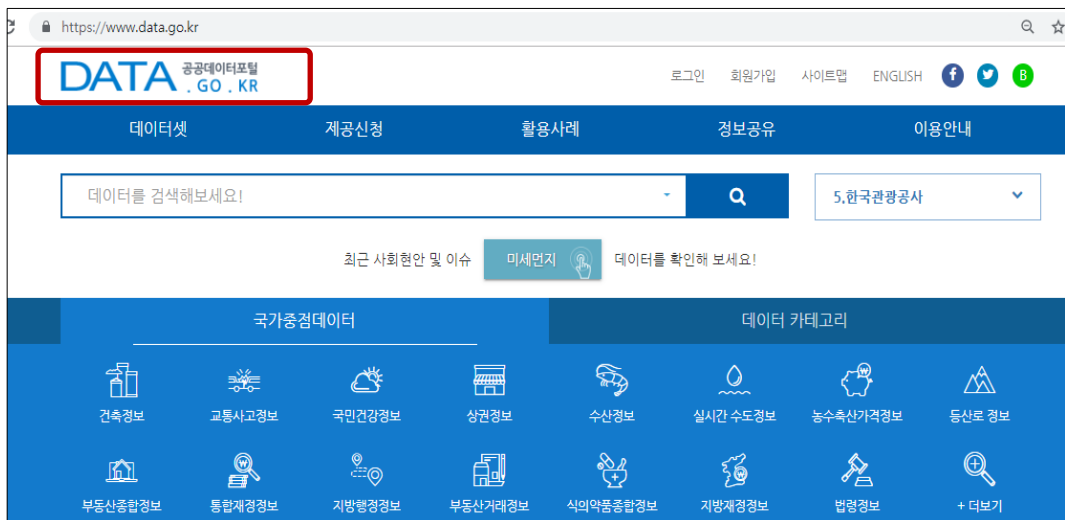
Open API를 활용한 공공 DB 수집

1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보
읽어오기

2 서울시 버스 정보와 버스 위치 정보

- 공공데이터포털 사이트 URL

<http://www.data.go.kr/>



[출처 : 공공데이터포털]

Open API를 활용한 공공 DB 수집

1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보 읽어오기

2 서울시 버스 정보와 버스 위치 정보

1 공공데이터포털 사이트에 로그인

버스 위치
정보 조회
서비스

노선 정보
조회 서비스

데이터
사용 요청

인증키를
받음

[승인] 버스위치정보조회 서비스

신청일:20 -01-09 [연장신청] 만료예정 :20 -01-09

서비스유형 :REST 분류 :교통및물류 > 도로 제공기관 :서울특별시

[승인] 노선정보조회 서비스

신청일:20 -01-09 [연장신청] 만료예정 :20 -01-09

서비스유형 :REST 분류 :교통및물류 > 도로 제공기관 :서울특별시

2 Open API 활용 가이드 제공 ➡ 이 문서를 참조하여 구현

기본정보					
서비스명	노선정보조회 서비스 상세설명				
서비스 유형	REST	일일트래픽	#####	평균응답속도(초)	53.599777777777774
심의여부	자동승인	신청유형	개발계정 연장신청	처리상태	승인
활용기간	2011-11-11 ~ 2011-01-09				

서비스정보	
일반 인증키 (UTF-8)	복사
End Point	
데이터포맷	XML
참고문서	서울특별시_노선정보조회_서비스_활용가이드_2 90110.docx

공공데이터·개방·공유·활용·체계·개발

Open API 활용가이드

[출처 : 공공데이터포털]

Open API를 활용한 공공 DB 수집

1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보
읽어오기

2 서울시 버스 정보와 버스 위치 정보

3 버스의 위치 정보를 받아오기

- 버스의 노선 정보 요청
➡ 버스의 라우트아이디<busRouteId>를 알아야 함
- 부여받은 인증키와 정보를 얻고자 하는 버스 번호를 입력
하여 URL 구성

<http://ws.bus.go.kr/api/rest/busRouteInfo/getBusRouteList?serviceKey=인증키&strSrch=버스번호>

- 버스 번호에 대한 노선 정보 요청

← → ↻ ⓘ 주의 요함 | ws.bus.go.kr/api/rest/busRouteInfo/getBusRouteList?serviceKey=%2Bjz

This XML file does not appear to have any style information

```
<ServiceResult>
  <comMsgHeader/>
  <msgHeader>
    <headerCd>0</headerCd>
    <headerMsg>정상적으로 처리되었습니다.</headerMsg>
    <itemCount>0</itemCount>
  </msgHeader>
  <msgBody>
    <itemList>
      <busRouteId>100100057</busRouteId>
      <busRouteNm>360</busRouteNm>
      <corpNm>BRT 02-404-8241</corpNm>
      <edStationNm>여의도환승센터</edStationNm>
      <firstBusTm>20190704040000</firstBusTm>
      <firstLowTm> </firstLowTm>
      <lastBusTm>20190704225000</lastBusTm>
      <lastBusYn> </lastBusYn>
      <lastLowTm>20150717222900</lastLowTm>
      <length>55.7</length>
      <routeType>3</routeType>
    </itemList>
  </msgBody>
</ServiceResult>
```

[출처 : 공공데이터포털]

Open API를 활용한 공공 DB 수집

1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보
읽어오기

2 서울시 버스 정보와 버스 위치 정보

3 버스의 위치 정보를 받아오기

- 버스의 위치 정보를 얻어오기 위한 URL 구성
- busRouteId에는 버스 노선 정보를 요청하고 받아온 XML 문서를 스크래핑하여 추출한 busRouteId를 사용

<http://ws.bus.go.kr/api/rest/busRouteInfo/getStaionByRoute?ServiceKey=인증키&busRouteId=100100057>

```

<?xml version="1.0" encoding="UTF-8" style="display:none" />
<ServiceResult>
  <comMsgHeader/>
  <msgHeader>
    <headerCd>0</headerCd>
    <headerMsg>정상적으로 처리되었습니다.</headerMsg>
    <itemCount>0</itemCount>
  </msgHeader>
  <msgBody>
    <itemList>
      <arsId>24655</arsId>
      <beginTm>04:00</beginTm>
      <busRouteId>100100057</busRouteId>
      <busRouteNm>360</busRouteNm>
      <direction>여의도환승센터</direction>
      <gpsX>127.1315385503</gpsX>
      <gpsY>37.4693007674</gpsY>
      <lastTm></lastTm>
      <posX>211634.62395161434</posX>
      <posY>441111.3499874971</posY>
      <routeType>3</routeType>
    </itemList>
  </msgBody>
</ServiceResult>
  
```

[출처 : 공공데이터포털]

Open API를 활용한 공공 DB 수집



1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보 읽어오기

3 소스

```
#파일명 : exam8_1.py
from bs4 import BeautifulSoup
import urllib.request as req

busNum = '360'
key = '인증키'
url1 =
'http://ws.bus.go.kr/api/rest/busRouteInfo/getBusRouteList?s
erviceKey='+key+'&strSrch='+busNum
savename = 'C:/Temp/businfo.xml'
req.urlretrieve(url1, savename)

xml = open(savename, 'r', encoding='utf-8').read()
soup = BeautifulSoup(xml, 'xml')
busRouteId = None
for itemList in soup.find_all('itemList') :
    busRouteId = itemList.find('busRouteId').string
    busRouteNm = itemList.find('busRouteNm').string
    if busRouteNm == busNum :
        break

url2 =
'http://ws.bus.go.kr/api/rest/busRouteInfo/getStaionByRoute
?ServiceKey='+key+'&busRouteId='+busRouteId

savename = 'C:/Temp/buspos.xml'
req.urlretrieve(url2, savename)

xml = open(savename, 'r', encoding='utf-8').read()
soup = BeautifulSoup(xml, 'xml')
```



Open API를 활용한 공공 DB 수집



1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보 읽어오기

3 소스

```
busPos = []
for itemList in soup.find_all('itemList') :
    gpsY = itemList.find('gpsY').string
    gpsX = itemList.find('gpsX').string

    busPos.append((gpsY, gpsX))

print('[ ' + busNum + '번 버스의 운행 위치 ]')
if len(busPos) != 0 :
    print(busNum + '번 버스 ' + str(len(busPos)) + '대 운행중...')
    for lat,lng in busPos :
        print(lat+', '+lng)
else :
    print('현재 운행중인 ' + busNum + '번 버스가 없어요...')
```



Open API를 활용한 공공 DB 수집



1 공공데이터포털 사이트에서 서울시 버스 정보와 버스 위치 정보 읽어오기

4 실행 결과 화면

```
PS C:\example\myvscode> & C:/Users/UNICO/Anaconda3/python.exe c:/example/myvscode/
unit8/exam8_1.py
[ 360번 버스의 운행 위치 ]
360번 버스 93대 운행중...
37.4693007674,127.1315385503
37.4701247674,127.1273151082
37.4790433896,127.1261237834
37.482125,127.124748
37.4883648963,127.1213627915
37.4943822012,127.1171609467
37.5002084826,127.1118384337
37.5034082968,127.1089442853
37.5068001738,127.1058262006
37.5128993133,127.0979457604
37.5122132454,127.0939144028
37.5117722886,127.0872805398
```

Open API를 활용한 공공 DB 수집

2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

1 서울열린데이터광장

서울
열린데이터
광장

- 열린시정 3.0에 의해 공공데이터를 민간에 개방·소통
- 공익성, 업무 효율성, 투명성을 높임
- 시민의 자발적 참여로 새로운 서비스와 공공의 가치를 창출할 수 있도록 하는 서비스



[출처 : 서울열린데이터광장]

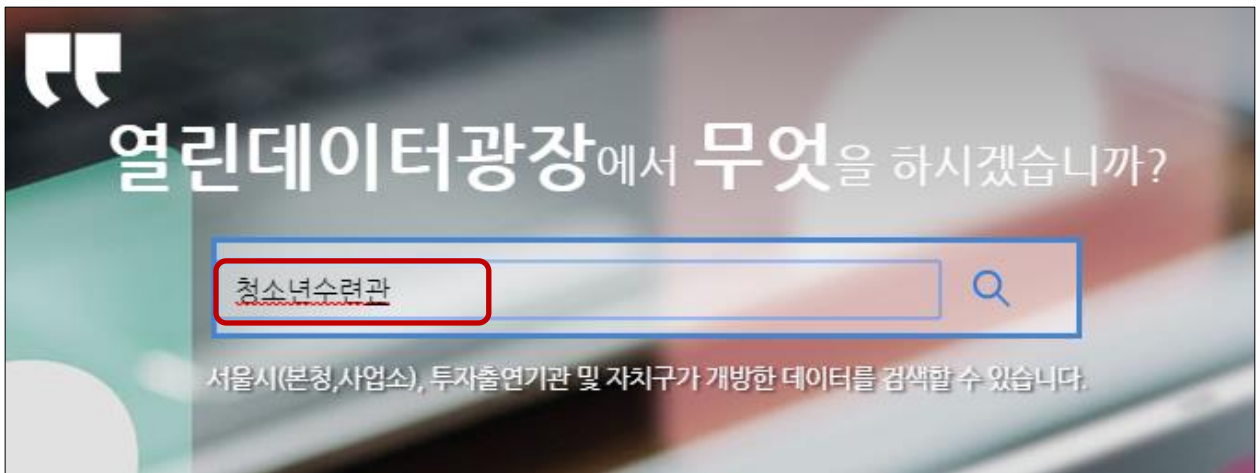
Open API를 활용한 공공 DB 수집



2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

2 서울시 청소년수련관 강좌 정보

- 1 서울시 청소년수련관 관련 정보를 얻어오기 위해 청소년수련관으로 검색



[출처 : 서울열린데이터광장]

2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

2 서울시 청소년 강좌정보를 얻기 위해 Open API에 대한 인증키 신청

[출처 : 서울열린데이터광장]

Open API를 활용한 공공 DB 수집



2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

2 서울시 청소년수련관 강좌 정보

3 인증키를 받기 위해 내용 작성

서비스 (사용) 환경	<input checked="" type="radio"/> 웹 사이트 개발 <input type="radio"/> 앱개발 (모바일 솔루션 등) <input type="radio"/> 연구 (논문 등) <input type="radio"/> 기타 참고자료
사용 URL (150자이내) *	<input type="text"/>
관리용 단체 / 기업 / 기관 대 표 이메일 *	<input type="text"/> @ <input type="text"/> <input type="text"/> 선택
활용용도 *	<input type="text"/>
내용 *	<div> 제공하여 주신 담당자 정보는 API중단, 서버점검 등에 대한 공지 및 정책 변경에 따른 안내를 위해 사용됩니다. </div> <div> <input type="text"/> </div> <div>0/200자</div>

[출처 : 서울열린데이터광장]

Open API를 활용한 공공 DB 수집



2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

2 서울시 청소년수련관 강좌 정보

4 요청 URL은 표의 내용을 참조하여 구성

<http://openapi.seoul.go.kr:8088/인증키/타입명/서비스명/시작인덱스/마지막인덱스/>

요청인자

변수명	타입	변수설명	값설명
KEY	String(필수)	인증키	OpenAPI 에서 발급된 인증키
TYPE	String(필수)	요청파일타입	xml : xml, xml파일 : xmlf, 엑셀파일 : xls, json파일 : json
SERVICE	String(필수)	서비스명	LampScpgmtb
START_INDEX	INTEGER(필수)	요청시작위치	정수 입력 (페이징 시작번호 입니다 : 데이터 행 시작번호)
END_INDEX	INTEGER(필수)	요청종료위치	정수 입력 (페이징 끝번호 입니다 : 데이터 행 끝번호)

Open API를 활용한 공공 DB 수집

2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

2 서울시 청소년수련관 강좌 정보

5 URL 문자열로 요청한 결과 화면

- XML 형식으로 100개의 정보를 요청하는 내용으로 구성

<http://openapi.seoul.go.kr:8088/인증키/xml/LampScpgmtb/1/100/>

```

→ C 주의 요함 | openapi.seoul.go.kr:8088/인증키/xml/LampScpgmtb/1/100/
his XML file does not appear to have any style information associated with it. The doc

<LampScpgmtb>
  <list_total_count>12309</list_total_count>
  ▼<RESULT>
    <CODE>INF0-000</CODE>
    <MESSAGE>정상 처리되었습니다</MESSAGE>
  </RESULT>
  ▼<row>
    <UP_ID>32</UP_ID>
    <UP_NM>도화청 소년문화의 집</UP_NM>
    <SALE_COM_CD>10</SALE_COM_CD>
    <SALE_COM_NM>교육문화사업수입</SALE_COM_NM>
    <BAS_CD>02</BAS_CD>
    <BAS_NM>교육문화사업</BAS_NM>
    <ITEM_CD>020006</ITEM_CD>
    <ITEM_NM>창의과학</ITEM_NM>
    <CLS_CD>000013</CLS_CD>
    <CLS_NM>수학</CLS_NM>
    <PGM_CD>021700040</PGM_CD>
    <PGM_NM>개념뿌리잡기 수학 D</PGM_NM>
    <FG_CD>02</FG_CD>
    <FG_NM>이 용료</FG_NM>
    <YYMM>201905</YYMM>
    <TARGET_CD>09</TARGET_CD>
    <TARGET_NM>초등생</TARGET_NM>
    <MM_QTY>9</MM_QTY>
  </row>

```

Open API를 활용한 공공 DB 수집



2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

3 소스



```
#파일명 : exam8_2.py
from bs4 import BeautifulSoup
import urllib.request as req

key = '인증키'
contentType = 'xml'
startIndex = '1'
endIndex = '100'
url =
'http://openapi.seoul.go.kr:8088/'+key+'/'+contentType+'/'
'LampScpgmtb/'+startIndex+'/'+endIndex+'/'
savename = 'C:/Temp/edu.xml'
req.urlretrieve(url, savename)

xml = open(savename, 'r', encoding='utf-8').read()
soup = BeautifulSoup(xml, 'xml')
pjList = []
for itemList in soup.find_all('row') :
    up_nm = itemList.find('UP_NM').string
    up_nm = '없음' if up_nm is None else up_nm
    pgm_nm = itemList.find('PGM_NM').string
    pgm_nm = '없음' if pgm_nm is None else pgm_nm
    target_nm = itemList.find('TARGET_NM').string
    target_nm = '없음' if target_nm is None else target_nm
    u_price = itemList.find('U_PRICE').string
    u_price = '없음' if u_price is None else u_price
    pjList.append((up_nm, pgm_nm, target_nm, u_price))
print('[ 서울 청소년수련관 강좌 리스트 ]')
for up_nm,pgm_nm,target_nm,u_price in pjList :
    print(up_nm+', '+pgm_nm+', '+target_nm+', '+str(u_price))
```

Open API를 활용한 공공 DB 수집



2 서울열린데이터광장에서 서울시 청소년수련관 강좌 정보 읽어오기

4 실행 결과 화면

```
PS C:\example\myvscode> & C:/Users/UNICO/Anaconda3/python.exe c:/example/myvscode/unit8/exam8_2.py
[ 서울 청소년 수련관 강좌 리스트 ]
구로청소년수련관,07시 수영(월 수 금)-초급,성인,60500
창동청소년센터,과)두뇌운동,치매예방 한자속독A(성인),구분없음,30000
창동청소년센터,과)바둑&체스D(6세~초등생),누구나,45000
창동청소년센터,과)바둑&체스H(6세~초등생),누구나,30000
창동청소년센터,과)어린이레고블럭A(4~5세),누구나,30000
금천청소년수련관,난초A상급-청소년,주3회수영-청소년,41000
창동청소년센터,과)어린이레고블럭B(6~7세),누구나,30000
서대문청소년수련관,속성 제빵자격증반,초6~성인(1개월),50000
구로청소년수련관,11시토요수영소수반-왕예은5월,초등생,70000
구로청소년수련관,꾸미다-특수분장사,구분없음,10000
구로청소년수련관,꾸미다-파티쉐,중고생,11000
구로청소년수련관,07시 수영(월 수 금)-중급,수영패키지,51500
```

Rest API와 SNS API를 활용한 SNS 글 수집

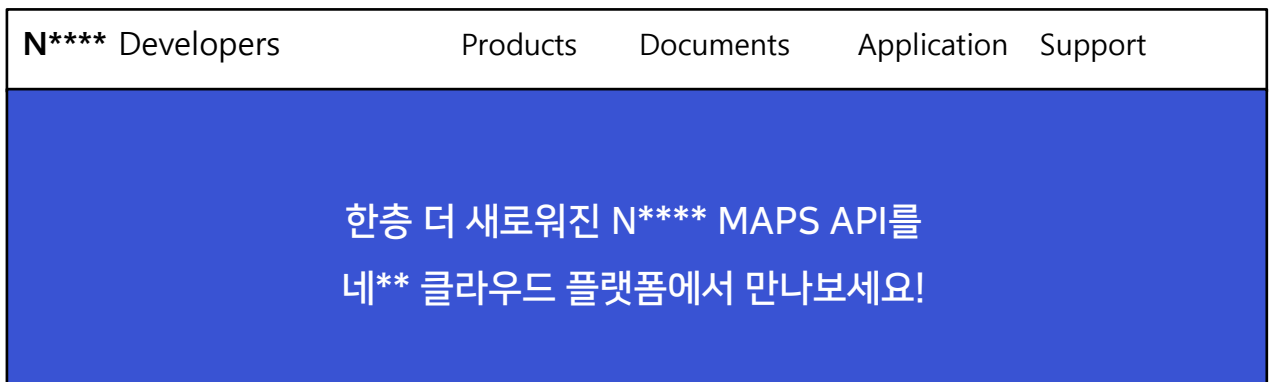


1 포털사이트 블로그 글 검색하여 읽어오기

1 포털사이트 블로그 글 검색

1 포털사이트 개발자 메인 웹 페이지에 방문하고 로그인

https://developers.n****.com/main/



Rest API와 SNS API를 활용한 SNS 글 수집

1 포털사이트 블로그 글 검색하여 읽어오기

1 포털사이트 블로그 글 검색

2 포털사이트의 검색 Open API 서비스

- 블로그, 뉴스, 책, 백과사전 등 다양한 분야의 내용을 검색할 수 있게 지원
- Documents 메뉴 > 검색 메뉴 선택

N**** Developers	Products	Documents	Application	Support	
API 공통 가이드	SDK&Tools Open API SDK 개발전용폰트/에디터 JavaScript 라이브러리 마크업 툴 스토리지/DBMS	Clova	아이디로 로그인	번역기	서비스 API 데이터랩 검색 단축 URL 이미지 캡처 공유하기

검색

- 블로그
- 뉴스
- 책
- 성인 검색어 판별
- 백과사전
- 영화
- 카페글
- 지식
- 지역
- 오타변환
- 웹문서
- 이미지
- 쇼핑
- 전문자료

Rest API와 SNS API를 활용한 SNS 글 수집



1 포털사이트 블로그 글 검색하여 읽어오기

1 포털사이트 블로그 글 검색

3 포털사이트 블로그 검색 결과를 출력해주는 REST API의 웹 화면

- 연두색의 오픈 API 이용 신청 링크 클릭

https://developers.n****.com/docs/search/blog/

N**** Developers	Products	Documents	Application Support
Document > 서비스 API > 검색 > 블로그			
검색 > 블로그			
API 공통 가이드 SDK&Tools Clova 아이디로 로그인 번역기	<p>*** 블로그 검색 결과를 출력해주는 REST API입니다. 비로그인 오픈 API이므로 GET으로 호출할 때 HTTP Header에 애플리케이션 등록 시 발급받은 Client ID와 Client Secret 값을 같이 전송해 주시면 활용 가능합니다.</p>		
	오픈 API 이용 신청 >		

Rest API와 SNS API를 활용한 SNS 글 수집

1 포털사이트 블로그 글 검색하여 읽어오기

1 포털사이트 블로그 글 검색

4 Client ID와 Client Secret를 부여받음

- 애플리케이션 이름, 사용 API, 비 로그인 오픈 API 신청에 대한 정보 입력

Application | API이용을 위해 애플리케이션을 등록하고 API 설정을 할 수 있습니다.

애플리케이션 등록 (API 이용 신청)

애플리케이션의 기본 정보를 등록하면, 좌측 내 애플리케이션 메뉴의 서브 메뉴에 등록하신 애플리케이션 이름으로 서브 메뉴가 만들어집니다.

애플리케이션 이름

애플리케이션 이름



- 아이디로 로그인 할때 사용자에게 표시되는 이름이므로 가급적 10자 이내의 간결한 이름을 사용해 주세요.
- 40자 이내의 영문, 한글, 숫자, 공백문자, “, ”, _만 입력 가능합니다.

Rest API와 SNS API를 활용한 SNS 글 수집

1 포털사이트 블로그 글 검색하여 읽어오기

1 포털사이트 블로그 글 검색

5 '오픈 API 테스트'라는 애플리케이션 이름으로 Client ID, Client Secret를 부여받은 화면

오픈 API 테스트	
개요	API 설정
네아로 검 수상태	멤버관리
로그인 통계	
애플리케이션 정보	
Client ID	iG***2***e*****d
Client Secret	● ● ● ● ● ● ● ● ● ●

6 포털사이트 블로그 검색 Rest API URL 문자열, Query 문자열 구성 정보

메서드	인증	요청 URL	출력 포맷
GET	-	https://openapi.n****.com/v1/search/blog.xml	XML
GET	-	https://openapi.n****.com/v1/search/blog.json	JSON

요청 변수	타입	필수 여부	기본값	설명
query	String	Y	-	- 검색을 원하는 문자열, UTF-8로 인코딩
display	Integer	N	10(기본값), 100(최대)	- 검색 결과 출력 건수 지정

Rest API와 SNS API를 활용한 SNS 글 수집



1 포털사이트 블로그 글 검색하여 읽어오기

2 소스

```
#파일명 : exam8_3.py
import urllib.request
import json
client_key = '부여받은 Client ID'
client_secret = '부여받은 Client Secret'
query = '치맥'
encText = urllib.parse.quote_plus(query)

num = 100
n****_url =
'https://openapi.n****.com/v1/search/blog.json?query=' +
encText + '&display=' + str(num)
request = urllib.request.Request(naver_url)
request.add_header("X-N****-Client-Id",client_key)
request.add_header("X-N****-Client-Secret",client_secret)
response = urllib.request.urlopen(request)
rescode = response.getcode()

if(rescode == 200):
    response_body = response.read()
    dataList = json.loads(response_body)
    count = 1
    print([' + query + '에 대한 포털사이트 블로그 글 ]')
    for data in dataList['items'] :
        print (str(count) + ' : ' + data['title'])
        print ('[' + data['description'] + ']')
        count += 1
else:
    print('오류 코드 : ' + rescode)
```



Rest API와 SNS API를 활용한 SNS 글 수집



1 포털사이트 블로그 글 검색하여 읽어오기

3 실행 결과 화면

```
PS C:\example\myvscode> & C:/Users/UNICO/Anaconda3/python.exe c:/example/myvscode/unit8/exam8_3.py
```

[치맥에 대한 네이버 블로그 글]

1 : 분위기에 빠진 강남역 치맥

[강남역 치맥 먹었는데 소소하지만 한잔 하면서 이야기거리도 풀어놓을 수 있었네요 강남역에서... 강남역 치맥 맛보면서 경기 봐도 좋을것 같더라구요 아니나 다를까 가게 한쪽 맥주 숙성하는 모습도 직접...]

2 : 광고호수공원 치맥하러 크래프트 다녔어요 :)

[날씨 좋을 때마다 거의 맨날 나가는 광고호수공원에 아기랑 놀상 가다가 오랜만에 아기 없이 광고호수공원으로 치맥을 위해!!! 오로지 치맥을 위해!!!!!!!!!!!!!! ! 다녔습니다 :) 역시 치맥은 낮에 먹는 게 제 맛이죠...]

100 : [10월 11일] 저녁 한강 따릉이&&치맥

[따릉이 중간 치맥 치맥 후 셀카 실제로 이렇게 어둡지는 않음. 내 체력 참 안좋구나 싶었다. 한 12km 넘어가면서부터는 너무 힘들었던 것 같다. 근데 달리다보니 동작역에서 잠실온거 짱신기... 아 그리고...]

Rest API와 SNS API를 활용한 SNS 글 수집



2 포털사이트 뉴스 글 읽어오기

1 포털사이트 뉴스 글

1 포털사이트 뉴스 검색 결과를 출력해주는 REST API의 웹 화면

- 블로그 검색에서 부여받은 Client ID, Client Secret 공통 사용 가능

N**** Developers		Products	Documents	Application Support
Document > 서비스 API > 검색 > 뉴스				
API 공통 가이드 SDK&Tools Clova 아이디로 로그인 번역기 서비스 ADI	검색 > 뉴스 *** 뉴스 검색 결과를 출력해주는 REST API입니다. 비로그인 오픈 API이므로 GET 으로 호출할 때 HTTP Header에 애플리케이션 등록 시 발급받은 Client ID와 Client Secret 값을 같이 전송해 주시면 활용 가능합니다.			
	오픈 API 이용 신청 >			

Rest API와 SNS API를 활용한 SNS 글 수집



2 포털사이트 뉴스 글 읽어오기

1 포털사이트 뉴스 글

2 포털사이트 뉴스 검색 Rest API URL 문자열, Query 문자열 구성 정보

메서드	인증	요청 URL	출력 포맷
GET	-	https://openapi.n****.com/v1/search/news.xml	XML
GET	-	https://openapi.n****.com/v1/search/news.json	JSON

요청 변수	타입	필수 여부	기본값	설명
query	String	Y	-	- 검색을 원하는 문자열, UTF-8로 인코딩
display	Integer	N	10(기본값),100(최대)	- 검색 결과 출력 건수 지정

Rest API와 SNS API를 활용한 SNS 글 수집



2 포털사이트 뉴스 글 읽어오기

2 소스

```
#파일명 : exam8_4.py
import urllib.request
import json
client_key = '부여받은 Client ID'
client_secret = '부여받은 Client Secret'
query = '치맥'
encText = urllib.parse.quote_plus(query)

num = 100
naver_url =
'https://openapi.n****.com/v1/search/news.json?query=' +
encText + '&display=' + str(num)
request = urllib.request.Request(naver_url)
request.add_header("X-N****-Client-Id",client_key)
request.add_header("X-N****-Client-Secret",client_secret)
response = urllib.request.urlopen(request)
rescode = response.getcode()

if(rescode == 200):
    response_body = response.read()
    dataList = json.loads(response_body)
    count = 1
    print([' + query + '에 대한 포털사이트 뉴스 글 ]')
    for data in dataList['items'] :
        print (str(count) + ' : ' + data['title'])
        print ('[' + data['description'] + ']')
        count += 1
else:
    print('오류 코드 : ' + rescode)
```



Rest API와 SNS API를 활용한 SNS 글 수집



2 포털사이트 뉴스 글 읽어오기

3 실행 결과 화면

```
PS C:\example\myvscode> & C:/Users/UNICO/Anaconda3/python.exe c:/example/myvscode/
unit8/exam8_4.py
[치맥에 대한 사이트 뉴스 글 ]
1 : 좋아하는 치킨 부위로 알아보는 연애 성향
[재미로 보는 좋아하는 치킨 부위에 따른 연애 성향. #치킨 #<b>치맥</b> #치콜 #치쏘 #
닭가슴살 #닭다리 #닭날개 #닭목 #닭살 #후라이드치킨 #양념치킨 #반반무많이 기획 : 김
영기 기자 1 디자인 : 이영환 비주얼다이브 디지털뉴스팀]
2 : [포커스] 대한민국, 유튜브 거품에 갇히다
[‘<b>치맥</b>’을 처음 먹어본 영국인들의 반응’ 동영상도 1400만번 이상 시청됐다. 사람
들은 영국남자의 동영상을 보며 내가 살고 있는 세계가 타인에게 어떻게 보여지는지 관
찰하기를 즐긴다. 좋은 반응이 나오면 기분... ]
```

Rest API와 SNS API를 활용한 SNS 글 수집

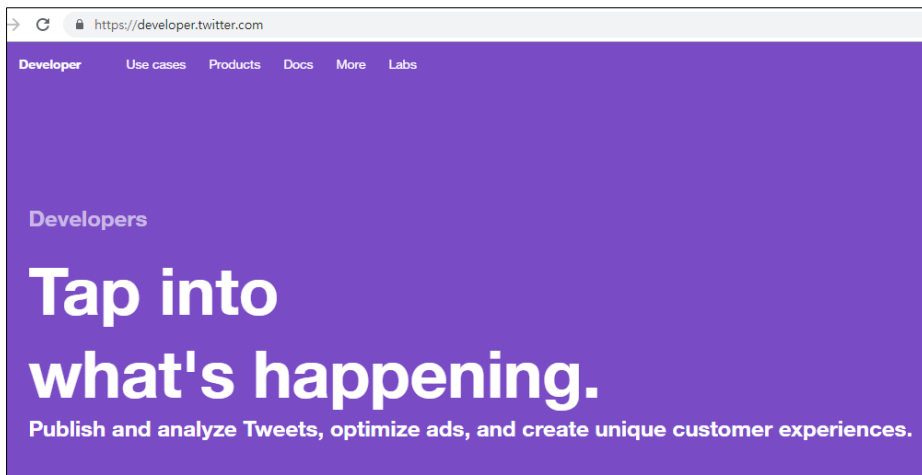


3 트위터 글을 검색하여 읽어오기

1 트위터 글 검색

1 트위터 글 검색

<https://developer.twitter.com/>



[출처 : twitter]

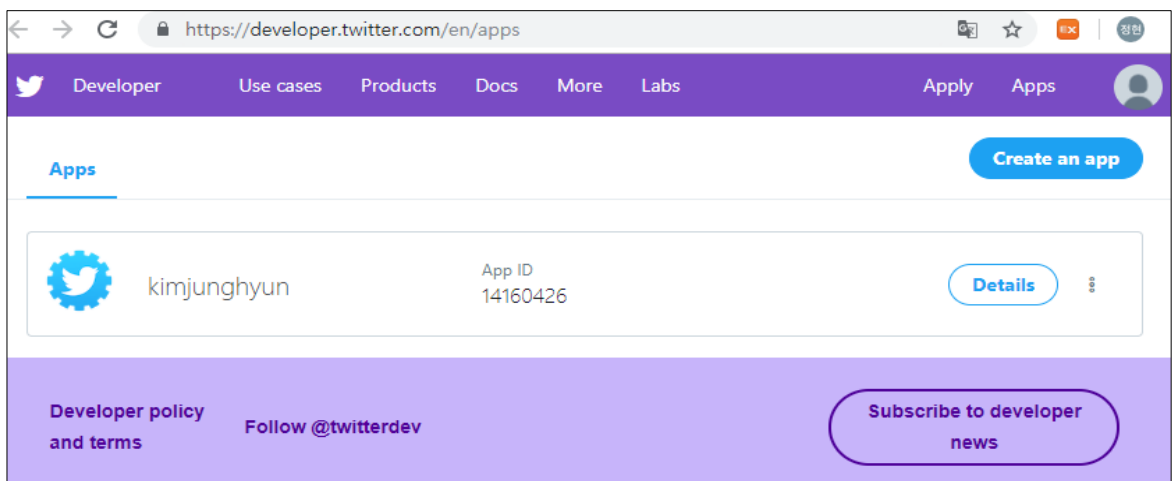
Rest API와 SNS API를 활용한 SNS 글 수집

3 트위터 글을 검색하여 읽어오기

1 트위터 글 검색

2 SNS에 개발자로 회원가입 후 로그인

- 이 후 Create an app 링크를 클릭하여 App(애플리케이션) 생성



[출처 : twitter]

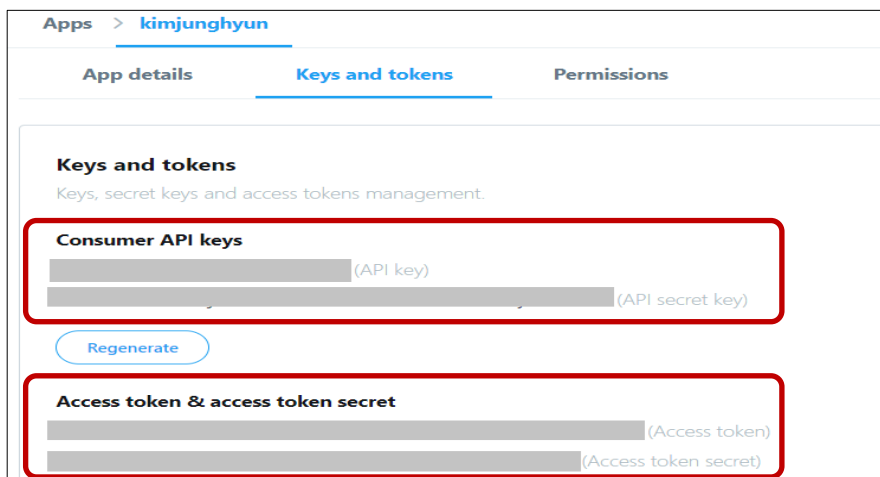
Rest API와 SNS API를 활용한 SNS 글 수집

3 트위터 글을 검색하여 읽어오기

1 트위터 글 검색

3 App 생성 및 등록

- Consumer API Key, Access token & Access token secret를 부여받음



[출처 : twitter]

Rest API와 SNS API를 활용한 SNS 글 수집



3 트위터 글을 검색하여 읽어오기

1 트위터 글 검색

4 Tweepy

Tweepy

- SNS Open API로의 접근을 쉽게 구현할 수 있게 지원하는 파이썬 라이브러리
- SNS를 크롤링할 때 일반적으로 Tweepy 모듈을 많이 사용
- 직접 크롤링을 하는 것보다 훨씬 간단하고 빠름

Rest API와 SNS API를 활용한 SNS 글 수집



3 트위터 글을 검색하여 읽어오기

1 트위터 글 검색

4 Tweepy

- SNS 검색의 정책 변화

Premium
search

Enterprise
search

Standard
search

Search Tweets

[Overview](#) [Quick Start](#) [Guides](#) [FAQ](#) [API Reference](#)

Overview contents ^

[Premium search](#)

[Standard search](#)

[Enterprise search](#)

비용이 들지 않지만 7일 내의 자료만 얻을 수 있음

Rest API와 SNS API를 활용한 SNS 글 수집

3 트위터 글을 검색하여 읽어오기

1 트위터 글 검색

4 Tweepy

- Tweepy 설치 : Anaconda Prompt 창에서 **pip install tweepy** 명령 수행

```
Anaconda Prompt

(base) C:\Users\UNICO>pip install tweepy
Collecting tweepy
  Downloading https://files.pythonhosted.org/packages/36/1b/2bd38043d22ade352fc3d3902cf30ce0e2f4bf285be3b304a2782a767aec/tweepy-3.8.0-py2.py3-none-any.whl
Requirement already satisfied: PySocks>=1.5.7 in c:\users\unico\anaconda3\lib\site-packages (from tweepy) (1.6.8)
Requirement already satisfied: six>=1.10.0 in c:\users\unico\anaconda3\lib\site-packages (from tweepy) (1.12.0)
Requirement already satisfied: requests>=2.11.1 in c:\users\unico\anaconda3\lib\site-packages (from tweepy) (2.21.0)
Collecting requests-oauthlib>=0.7.0 (from tweepy)
  Downloading https://files.pythonhosted.org/packages/a3/12/b92740d845ab62ea4edf04d2f4164d82532b5a0b03836d4d4e71c6f3d379/requests_oauthlib-1.3.0-py2.py3-none-any.whl
Requirement already satisfied: idna<2.9,>=2.5 in c:\users\unico\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy) (2.8)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\unico\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy) (2019.9.11)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in c:\users\unico\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy) (1.24.1)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\unico\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy) (3.0.4)
Collecting oauthlib>=3.0.0 (from requests-oauthlib>=0.7.0->tweepy)
  Downloading https://files.pythonhosted.org/packages/05/57/ce2e7a8fa7c0afb54a0581b14a65b56e62b5759dbc98e80627142b8a3704/oauthlib-3.1.0-py2.py3-none-any.whl (147kB)
    97% |#####| 143kB 693kB/s eta 0:00:01
    100% |#####| 153kB 2.5MB/s
Installing collected packages: oauthlib, requests-oauthlib, tweepy
Successfully installed oauthlib-3.1.0 requests-oauthlib-1.3.0 tweepy-3.8.0

(base) C:\Users\UNICO>
```

Rest API와 SNS API를 활용한 SNS 글 수집



3 트위터 글을 검색하여 읽어오기

2 Tweepy의 주요 API

1 OAuth 1a 인증

```
auth = tweepy.OAuthHandler(api_key, api_secret)
auth.set_access_token(access_token,
access_token_secret)
```

2 SNS에 대한 모든 요청을 처리하는 래퍼 객체 생성

```
api = tweepy.API(auth)
```

3 지정된 쿼리와 일치하는 트윗을 반환

```
api.search(q=keyword, count=100)
```

- q : 최대 500자의 검색어 문자열 사용 가능
- count : 페이지 당 반환할 트윗 수(최대 100, 기본값 15)

4 Tweepy API 도큐멘테이션

<https://tweepy.readthedocs.io/en/latest/>

Rest API와 SNS API를 활용한 SNS 글 수집



3 트위터 글을 검색하여 읽어오기

3 소스

```
#파일명 : exam8_5.py
import tweepy

api_key = "API 키"
api_secret = "API 비밀키"
access_token = "액세스 토큰"
access_token_secret = "액세스 토큰 암호"
auth = tweepy.OAuthHandler(api_key, api_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth)

keyword = '치맥'
search = []
cnt = 1
while(cnt <= 10):
    tweets = api.search(q=keyword, count=100)
    for tweet in tweets :
        search.append(tweet)
    cnt += 1

data = {}
i = 1
print('[ ' + keyword + '에 대한 트윗 글 ]')
for tweet in search:
    data['text'] = tweet.text
    print(i, " : ", data)
    i += 1
```



Rest API와 SNS API를 활용한 SNS 글 수집



3 트위터 글을 검색하여 읽어오기

4 실행 결과 화면

```
PS C:\example\myvscode> & C:/Users/UNICO/Anaconda3/python.exe c:/example/myvscode/unit8/exam8_5.py
```

```
[치맥에 대한 트윗 글]
```

```
1 : {'text': 'RT @artist_nuest: 부암동카페에서 치(즈케이크)+맥(주,아메리카노)\n\n◆◆가 찾은 힙플, 핵인싸의 치맥 \n# 핵인싸음식인기 차트 https://t.co/P*****N'}
2 : {'text': 'RT @artist_nuest: 부암동카페에서 치(즈케이크)+맥(주,아메리카노)\n\n◆◆가 찾은 힙플, 핵인싸의 치맥 \n# 핵인싸음식인기 차트 https://t.co/P*****N'}
3 : {'text': '솔직히 난 아직 치맥 이해안가 치킨엔 콜라지 !'}
4 : {'text': 'RT @artist_nuest: 부암동카페에서 치(즈케이크)+맥(주,아메리카노)\n\n◆◆가 찾은 힙플, 핵인싸의 치맥 \n# 핵인싸음식인기 차트 https://t.co/P*****N'}
5 : {'text': 'RT @artist_nuest: 부암동카페에서 치(즈케이크)+맥(주,아메리카노)\n\n◆◆가 찾은 힙플, 핵인싸의 치맥 \n# 핵인싸음식인기 차트 https://t.co/P*****N'}
6 : {'text': '우리 한강가자! 한강데이트! 치맥!'}
```

```
1000 : {'text': '홍차장은 하루 일과 마치고 맥주 한잔씩 꼭 하는거 잘어울림 김과장은 맨날 치맥타령하면서 사실은 안주빨만 세우는 거면 좋겠다 치킨맛집 리스트 뽑아다니는 김과장. 포크들고 열심히 치킨 먹는 김과장 맞은편에서 꿀... https://t.co/6*****b'}
```

학습정리

1. Open API를 활용한 공공 DB 수집



- Rest API는 Representational State Transfer API의 약자
 - HTTP 프로토콜을 통해서 정보를 제공하는 함수
 - 실질적인 API 사용은 정해진 구조의 URL 문자열 사용
- 공공데이터포털, 서울열린데이터광장 등 공공 DB를 수집하기 위해서는 회원가입, Open API 인증키 발급 등을 필수적으로 수행
- 공공 DB 사이트들이 제공하는 데이터 서비스 Open API는 Rest API로 전달해야 하는 Query 문자열의 구조와 원하는 출력형식(XML, JSON)에 따라 URL 문자열을 구성

학습정리

2. Rest API와 SNS API를 활용한 SNS 글 수집



- 애플리케이션 이름, 사용 API 그리고 비 로그인 Open API 신청에 대한 정보를 입력한 후 Client ID와 Client Secret를 부여받음
- 부여받은 Client ID와 Client Secret으로 포털사이트 블로그 검색, 포털사이트 뉴스 검색 등 검색 관련 모든 기능 처리
- Tweepy란 트위터 API으로의 접근을 쉽게 구현할 수 있게 지원하는 파이썬 라이브러리로 추가 설치를 해야 사용 가능