

05. openAPI를 이용한 데이터 수집

목차

- ▶ 크롤링이란
- ▶ 서버-클라이언트 사이의 데이터 전달 방식에 대한 이해
- ▶ 데이터 수집을 위한 모듈 설치
- ▶ 모듈을 이용한 데이터 수집 과정
- ▶ openAPI를 이용한 데이터 수집 실습

기본 개념

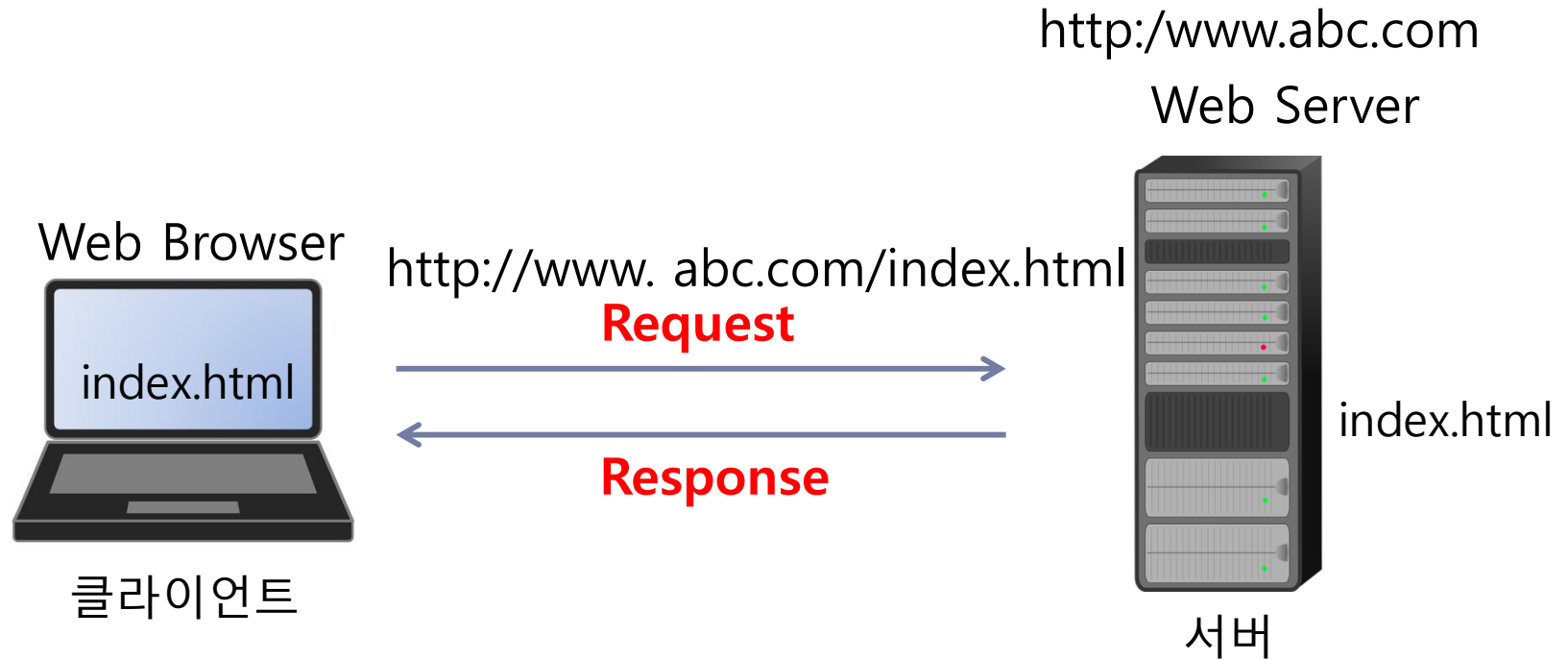
- ▶ 크롤링(Crawling)

- ▶ 웹에서 링크를 타고 다니며 웹 페이지들을 수집하는 것.

- ▶ 크롤러(Crawler)

- ▶ 웹에 접속해서 데이터를 받아오는 프로그램
 - ▶ 스파이더, 봇, 지능 에이전트

서버와 클라이언트



서버와 클라이언트

▶ URL(Uniform Resource Locator)

- ▶ 네트워크 상에서 자원을 요청하는 규약
- ▶ 데이터를 주고 받을 때 URL을 이용하여 데이터를 주고 받음

프로토콜://주소 또는 IP:포트번호/리소스경로?쿼리스트링

| | | | | |
|---|------------------------------|---|----------|-----------------|
| 인터넷 : http HTTPS, FTP, SFTP, SSH... | www. abc.com 196.255.3.12 | http : 80 HTTPS : 443 FTP : 20, 21 telnet : 23 SSH : 22 | /score/a | name=kim&kor=98 |
|---|------------------------------|---|----------|-----------------|

서버와 클라이언트

프로토콜://주소 또는 IP:포트번호/리소스경로?쿼리스트링

<https://openapi.naver.com/v1/search/encyc?query='빅데이터'&start=1&display=5>

빅데이터 처리의 과정

1. 웹 사이트에서 HTML 문서 가져오기



2. HTML 문서를 파이썬 객체로 변환, 정제, 파일로 저장



3. 데이터를 읽어들이며 목적에 맞게 처리



4. 처리 결과를 그래프로 나타내기

데이터
수집

분석
및
시각화

필요한 모듈

- ▶ 요청 모듈
 - 웹 데이터를 받아올 때 사용하는 모듈
 - ▶ requests
 - ▶ urllib
- ▶ 파싱 모듈
 - 받아 온 데이터를 파이썬 객체로 변환, 다룰 수 있게 함
 - ▶ bs4
- ▶ requests, bs4
 - 기본 내장 모듈이 아니므로 설치 후 사용해야 함.

pip으로 모듈 설치하기

- ▶ 터미널 창에서 입력
 - ▶ 설치

```
pip install 모듈명(또는 라이브러리명)
```

- ▶ 설치된 모듈 확인

```
pip list
```

- ▶ 주의!
pip을 이용한 모듈 설치에 운영체제에 따라 오류가 발생할 수 있음.

requests 모듈 설치

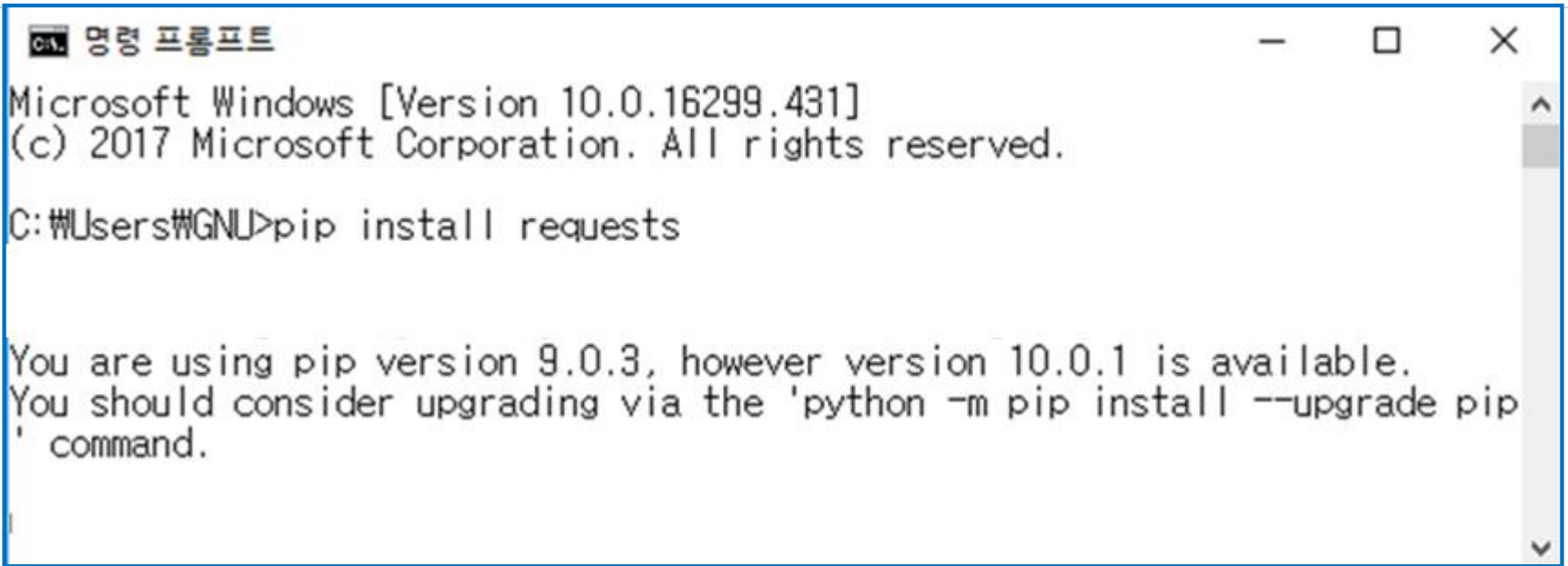
pip install requests

```
C:\Windows\system32\cmd.exe

C:\Users\EunKyoung>pip install requests
Collecting requests
  Downloading https://files.pythonhosted.org/packages/49/df/50aa1999ab9bde74656c2919d9c0c085fd2b3775fd3eca826012bef76d8c/requests-2.18.4-py2.py3-none-any.whl (88kB)
    100% |#####| 92kB 263kB/s
Collecting chardet<3.1.0,>=3.0.2 (from requests)
  Downloading https://files.pythonhosted.org/packages/bc/a9/01ffebfb562e4274b6487b4bb1ddec7ca55ec7510b22e4c51f14098443b8/chardet-3.0.4-py2.py3-none-any.whl (133kB)
    100% |#####| 143kB 450kB/s
Collecting idna<2.7,>=2.5 (from requests)
  Downloading https://files.pythonhosted.org/packages/27/cc/6dd9a3869f15c2edfab863b992838277279ce92663d334df9ecf5106f5c6/idna-2.6-py2.py3-none-any.whl (56kB)
    100% |#####| 61kB 330kB/s
Collecting urllib3<1.23,>=1.21.1 (from requests)
  Downloading https://files.pythonhosted.org/packages/63/cb/6965947c13a94236f6d4b8223e21beb4d576dc72e8130bd7880f600839b8/urllib3-1.22-py2.py3-none-any.whl (132kB)
    100% |#####| 133kB 272kB/s
Collecting certifi>=2017.4.17 (from requests)
  Downloading https://files.pythonhosted.org/packages/7c/e6/92ad559b7192d846975fc916b65f667c7b8c3a32bea7372340bfe9a15fa5/certifi-2018.4.16-py2.py3-none-any.whl (150kB)
    100% |#####| 153kB 289kB/s
Installing collected packages: chardet, idna, urllib3, certifi, requests
Successfully installed certifi-2018.4.16 chardet-3.0.4 idna-2.6 requests-2.18.4 urllib3-1.22
```

requests 모듈 설치

▶ 설치가 된 경우



```
C:\> 명령 프롬프트

Microsoft Windows [Version 10.0.16299.431]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\GNU>pip install requests

You are using pip version 9.0.3, however version 10.0.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip'
command.
```

모듈 불러오기

```
>>> import requests
Traceback (most recent call last):
  File "<pyshell#0>", line 1, in <module>
    import requests
ImportError: No module named 'requests'
>>> import requests
>>> | 오류가 없으면 정상적으로 import된 것임.
```

bs4 모듈 설치

▶ 설치 방법 1

```
pip install bs4
```

▶ 설치 방법 2

```
pip install beautifulsoup4
```

빅데이터 처리를 위한 파이썬 기초와 크롤링

1. 강의기간 : 2022. 8. 1.(월) ~ 8. 11.(2주)

2. 강의 시간 : 18:30 ~ 20:50

3. 담당강사 : 김은경

4. 강의실 : 경상국립대학교 교육정보전산원 1층 105호(4동 학술정보관)

5. 강의 목차

- [빅데이터 개요](#)

- [빅데이터 처리를 위한 파이썬 기초 1](#)

- [빅데이터 처리를 위한 파이썬 기초 2](#)

- [빅데이터 처리를 위한 파이썬 기초 3](#)

- [크롤링 기초 1](#)

- [크롤링 기초 2](#)

- [크롤링에서 시각화까지](#)

- [파이썬을 이용한 데이터 수집과 분석](#)

HTML 문자열 객체 생성

코드

```
#필요한 모듈 импорт
import requests
from bs4 import BeautifulSoup
import json
import os

#HTTP 요청
response = requests.get('https://madelen2016.github.io/class/')
print(response)

htmlObj = response.content
print(htmlObj)
```

실행결과

한글 : 바이너리 인코딩!

바이너리는 인코딩 설정에 상관없이
같은 문자는 똑같은 형태로 변환되므로
인코딩 때문에 문자가 깨지는 것을 방지

BeautifulSoup의 필요성

- ▶ **requests**의 요청의 결과는

- ▶ 단순한 html태그를 담고 있는 문자열이기 때문에 정보를 추출하기 어렵다

- ▶ **BeautifulSoup**는

- ▶ html 코드를 파이썬 객체 구조로 변환해 주는 **Parsing**을 맡고 있으며 제대로 된 의미 있는 정보를 추출할 수 있도록 해준다

BeautifulSoup을 이용한 파싱

코드

#필요한 모듈 импорт

```
import requests  
from bs4 import BeautifulSoup  
import json  
import os
```

```
response = requests.get(' https://madelen2016.github.io/class/')  
htmlObj = response.content
```

```
bsObj = BeautifulSoup( htmlObj , 'html.parser')  
print(bsObj)
```

#bsObj는 파싱을 거친 파이썬 객체

BeautifulSoup 파싱 결과

실행결과

```
<html>
<head>
<meta charset="utf-8"/>
<meta content="web crawling example page" name="description"/>
<title> GNU Bigdata Class - Web Crawling Example Page</title>
</head>
<body>
<style type="text/css">
    td.td-year{
        text-align : center;
        color : blue;
    }

    td.td-part{
        color : navy;
    }

    div.why{
        list-style-type: circle;
    }
    h1.notice{ color : red; font-weight : bold }
</style>
<h1><b>빅데이터 처리를 위한 파이썬 기초와 크롤링</b></h1>
<h3> 1. 강의기간 : 2022. 8. 1.(월) ~ 8. 11.(2주)</h3>
<h3> 2. 강의 시간 : 18:30 ~ 20:50 </h3>
<h3> 3. 담당강사 : 김은경 </h3>
<h3> 4. 강의실 : 경상국립대학교 교육정보전산원 1층 105호(4동 학술정보관)</h3>
<h3> 5. 강의 목차 </h3>
<a href="01 intro.pdf"> - 빅데이터 개요 </a> <br/>
<a href="02 python basic1.pdf"> - 빅데이터 처리를 위한 파이썬 기초 1 </a> <br/>
<a href="03 python basic2.pdf"> - 빅데이터 처리를 위한 파이썬 기초 2 </a> <br/>
<a href="04 python basic3.pdf"> - 빅데이터 처리를 위한 파이썬 기초 3 </a><br/>
```

파싱 결과에서 <A>태그 부분만 추출

빅데이터 처리를 위한 파이썬 기초와 크롤링

1. 강의기간 : 2022. 8. 1.(월) ~ 8. 11.(2주)

2. 강의 시간 : 18:30 ~ 20:50

3. 담당강사 : 김은경

4. 강의실 : 경상국립대학교 교육정보전산원 1층 105호(4동 학술정보관)

5. 강의 목차

- 빅데이터 개요

- 빅데이터 처리를 위한 파이썬 기초 1

- 빅데이터 처리를 위한 파이썬 기초 2

- 빅데이터 처리를 위한 파이썬 기초 3

- 크롤링 기초 1

- 크롤링 기초 2

- 크롤링에서 시각화까지

- 파이썬을 이용한 데이터 분석

새 탭에서 링크 열기

새 창에서 링크 열기

시크릿 창에서 링크 열기

다른 이름으로 링크 저장...

링크 주소 복사

복사

하이라이트 링크 복사

Google에서 '빅데이터 처리를 위한 파이썬 기초 1 - 빅데이터 처리를 위한 파이썬 기

인쇄

검사

또는 F12

파싱 결과에서 a태그 위치 분석

```
<h3> 5. 강의 목차 </h3>
<a href="01_intro.pdf"> - 빅데이터 개요 </a>
<br>
... <a href="02_python_basic1.pdf"> - 빅데이터 처리를 위한 파이썬 기초 1 </a> == $0
<br>
<a href="03_python_basic2.pdf"> - 빅데이터 처리를 위한 파이썬 기초 2 </a>
<br>
<a href="04_python_basic3.pdf"> - 빅데이터 처리를 위한 파이썬 기초 3 </a>
<br>
<a href="05_openAPI.pdf"> - 크롤링 기초 1 </a>
<br>
<a href="06_crawling1.pdf"> - 크롤링 기초 2 </a>
<br>
<a href="07_crawling2.pdf"> - 크롤링에서 시각화까지 </a>
<br>
<a href="08_crawling_analysis.pdf"> - 파이썬을 이용한 데이터 수집과 분석 </a>
<br>
```

html body a

화면 하단에 <a> 태그의 중첩 경로가 표시됨

select를 이용한 a태그 추출

코드

```
import requests
from bs4 import BeautifulSoup
import json
import os

response = requests.get('https://madelen2016.github.io/class/')

htmlObj = response.content
bsObj = BeautifulSoup(htmlObj, 'html.parser') #bsObj는 파싱을 거친 파이썬 객체

listObj= bsObj.select('body > a') #a 태그만 추출

print(listObj)
```

select 결과 : 리스트 객체

```
[<a href="01 intro.pdf"> - 빅데이터 개요 </a>, <a href="02 python basic1.pdf"> - 빅데이터 처리를 위한 파이썬 기초 1 </a>, <a href="03 python basic2.pdf"> - 빅데이터 처리를 위한 파이썬 기초 2 </a>, <a href="04 python basic3.pdf"> - 빅데이터 처리를 위한 파이썬 기초 3 </a>, <a href="05 openAPI.pdf"> - 크롤링 기초 1 </a>, <a href="06 crawling1.pdf"> - 크롤링 기초 2 </a>, <a href="07 crawling2.pdf"> - 크롤링에서 시각화까지 </a>, <a href="08 crawling_analysis.pdf"> - 파이썬을 이용한 데이터 수집과 분석 </a>]
```

추출한 정보를 딕셔너리에 담기 - 더욱 정제

코드

```
import requests
from bs4 import BeautifulSoup
import json
import os
```

```
response = requests.get('https://madelen2016.github.io/class/')
```

```
htmlObj = response.content
```

```
bsObj = BeautifulSoup(htmlObj, 'html.parser')
```

```
listObj = bsObj.select('body > a')
```

#bsObj는 파싱을 거친 파이썬 객체

#a 태그만 추출

```
dictObj= {}
```

```
for i in listObj :
```

```
    dictObj[i.text] = i.get('href')
```

```
print(dictObj)
```

왜 딕셔너리에 저장할까?

 - 빅데이터 개요

딕셔너리[키] = 값

- 빅데이터 개요

⋮

01 intro.pdf

⋮

딕셔너리 정보 출력

{ - 빅데이터 개요 ': '01 intro.pdf' - 빅데이터 처리를 위한 파이썬 기초 1 ': '02 python basic1.pdf' - 빅데이터 처리를 위한 파이썬 기초 2 ': '03 python basic2.pdf' - 빅데이터 처리를 위한 파이썬 기초 3 ': '04 python basic3.pdf' - 크롤링 기초 1 ': '05 openAPI.pdf' - 크롤링 기초 2 ': '06 crawling1.pdf' - 크롤링에서 시각화까지 ': '07 crawling2.pdf' - 파이썬을 이용한 데이터 수집과 분석 ': '08 crawling_analysis.pdf' }

- 빅데이터 개요 > 01 intro.pdf
- 빅데이터 처리를 위한 파이썬 기초 1 > 02 python basic1.pdf
- 빅데이터 처리를 위한 파이썬 기초 2 > 03 python basic2.pdf
- 빅데이터 처리를 위한 파이썬 기초 3 > 04 python basic3.pdf
- 크롤링 기초 1 > 05 openAPI.pdf
- 크롤링 기초 2 > 06 crawling1.pdf
- 크롤링에서 시각화까지 > 07 crawling2.pdf
- 파이썬을 이용한 데이터 수집과 분석 > 08 crawling_analysis.pdf

딕셔너리를 JSON 파일로 저장

코드

```
import requests
from bs4 import BeautifulSoup
import json
import os

response = requests.get('https://madelen2016.github.io/class/')

htmlObj = response.content
bsObj = BeautifulSoup(htmlObj, 'html.parser')      #bsObj는 파싱을 거친 파이썬 객체
listObj = bsObj.select('body > a')                #a 태그만 추출

dictObj= {}
for i in listObj :
    dictObj[i.text] = i.get('href')

with open(os.path.join('result.json'), 'w+') as json_file:
    json.dump(dictObj, json_file)
```

JSON

- ▶ 제이슨, JavaScript Object Notation
- ▶ 인터넷에서 자료를 주고 받을 때 그 자료를 표현하는 방법
- ▶ 속성-값 쌍으로 이루어진 데이터 오브젝트를 전달하기 위해 인간이 읽을 수 있는 텍스트를 사용하는 개방형 표준 포맷.
- ▶ 프로그래밍 언어나 플랫폼에 독립적.

▶ 예 :

```
{  
  "이름": "홍길동",  
  "나이": 30,  
  "주소": "진주시 가좌동",  
  "특기": ["농구", "축구"],  
  "가족관계": {"아버지": "홍판서", "어머니": "춘섬"},  
  "회사": "경상대학교"  
}
```

result.json 파일

```
{ " - Wube45Wub370Wuc774Wud130 Wuac1cWuc694 " : "01 intro.pdf", " -  
Wube45Wub370Wuc774Wud130 Wucc98Wub9acWub97c Wuc704Wud55c  
Wud30cWuc774Wuc36c Wuae30Wucd08 1 ": "02 python basic1.pdf", " -  
Wube45Wub370Wuc774Wud130 Wucc98Wub9acWub97c Wuc704Wud55c  
Wud30cWuc774Wuc36c Wuae30Wucd08 2 ": "03 python basic2.pdf", " -  
Wube45Wub370Wuc774Wud130 Wucc98Wub9acWub97c Wuc704Wud55c  
Wud30cWuc774Wuc36c Wuae30Wucd08 3 ": "04 python basic3.pdf", " -  
Wud06cWub864Wub9c1 Wuae30Wucd08 1 ": "05 openAPI.pdf", " -  
Wud06cWub864Wub9c1 Wuae30Wucd08 2 ": "06 crawling1.pdf", " -  
Wud06cWub864Wub9c1Wuc5d0Wuc11c  
Wuc2dcWuac01Wud654Wuae4cWuc9c0 ": "07 crawling2.pdf", " -  
Wud30cWuc774Wuc36cWuc744 Wuc774Wuc6a9Wud55c  
Wub370Wuc774Wud130 Wuc218Wuc9d1Wuacfc Wubd84Wuc11d ": "08  
crawling_analysis.pdf" }
```

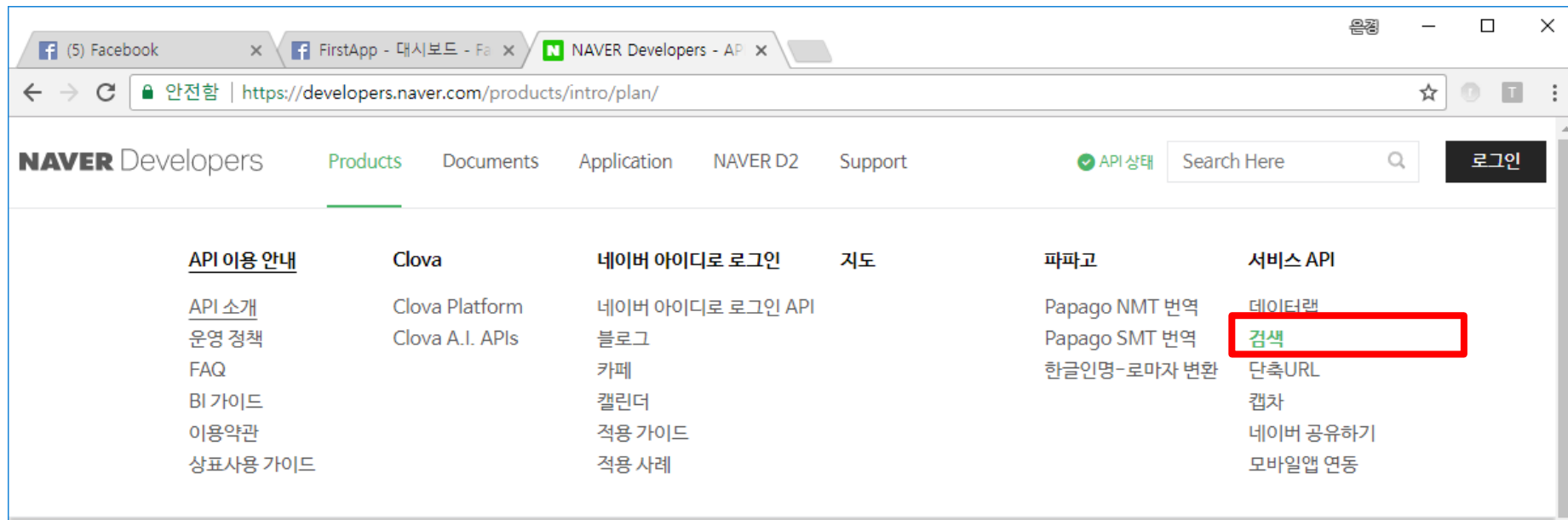
데이터 수집 방법

- ▶ openAPI를 이용
- ▶ 웹 페이지 구조 분석을 이용
- ▶ 공공데이터 이용 <https://www.data.go.kr/>

openAPI를 이용한 데이터 수집 네이버 API 사용하기

네이버 API 사용하기

- ▶ 네이버 개발자페이지로 이동
 - ▶ <https://developers.naver.com>
 - ▶ 상단의 메뉴 Products > 서비스 API > 검색 을 선택



네이버 API 사용하기

NAVER Developers - 검색 x

← → ↻ 안전함 | https://developers.naver.com/products/search/

NAVER Developers Products Documents Application NAVER D2 Support

API 상태 Search Here 로그인

API 이용 안내
Clova
네이버 아이디로 로그인
지도
파파고
서비스 API

데이터랩
검색
단축URL
캡차
네이버 공유하기
모바일앱 연동

네이버 검색 결과 콘텐츠

웹 서비스 또는 모바일 앱에 네이버 웹문서/블로그/뉴스/책/영화/카페글/지식iN/쇼핑/이미지/백과사전/전문자료 분야에 대한 검색 결과를 보여 줄 수 있습니다.

지역 검색

'OO역 맛집', 'OO동 술집'과 같은 검색 결과를 보여주고 싶을 때 사용하며, 네이버 지역 서비스에 등록된 각 지역별 업체 및 상호 검색결과를 보여줍니다.

검색 부가 기능

검색 부가 기능으로 특정 검색어에 대해 성인검색어 여부를 알려주는 기능과 검색창에 입력된 오타를 바로 잡아 주는 오타변환 기능을 제공합니다.

* 처리한도 : 25,000/일

오픈 API 이용 신청 개발 가이드 보기

네이버 API 사용하기

▶ 약관 동의에 체크 > 확인 버튼 클릭

NAVER Developers Products Documents Application NAVER D2 Support 🟢 API 상태 Search Here  

약관동의 > 계정 정보 등록 > 애플리케이션 등록

API 이용약관

가. 일반조건

1. 목적

NAVER API서비스 이용약관(이하 '본 약관'이라 합니다)은 네이버주식회사(이하 '회사'라 합니다)와 이용자 간에 회사가 제공하는 API 서비스의 제공 및 이용과 관련한 회사와 이용자의 권리, 의무 및 책임에 관한 제반 사항과 기타 필요한 사항을 구체적으로 규정함을 목적으로 합니다.

2. 용어의 정의

본 약관에서 사용하는 용어의 정의는 다음과 같습니다.

네이버 API 사용하기

약관동의

> 계정 정보 등록

> 애플리케이션 등록

계정 설정

| | |
|-------|---|
| ① 휴대폰 | <div>휴대폰 인증</div> <p>초기 등록을 위해서 개인 인증이 필요합니다. 인증은 10분간 유효 합니다.</p> |
| 이메일 | <div>els*****@naver.com</div> <div>이메일 정보 변경</div> |
| 회사 이름 | <div></div> <p><input checked="" type="checkbox"/> 소속 회사 이름은 제휴 접수 및 검토 등을 위해 사용되며, 서비스 이용 기간 동안 보관됩니다.</p> <p>&"'<> 문자는 허용하지 않습니다.</p> |

확인

취소

네이버 API 사용하기

약관동의

>

계정 정보 등록

>

애플리케이션 등록

이미 개발자 계정을 신청한 경험이 있다면 이 단계부터 시작됨.
애플리케이션 등록 (API 이용신청)

애플리케이션의 기본 정보를 등록하면, 좌측 [내 애플리케이션](#) 메뉴의 서브 메뉴에 등록하신 애플리케이션 이름으로 서브 메뉴가 만들어집니다.

| | |
|-----------|---|
| 애플리케이션 이름 | <input type="text" value="애플리케이션 이름"/> ⓘ <ul style="list-style-type: none">네이버 아이디로 로그인할 때 사용자에게 표시되는 이름이므로 가급적 10자 이내의 간결한 이름을 사용해주세요.40자 이내의 영문, 한글, 숫자, 공백문자, "-", "_" 만 입력 가능합니다. |
| 사용 API | <input type="text" value="선택하세요."/> ⓘ 사용할 API를 추가해 주세요. |

- [\[애플리케이션 이름\]](#) 설정을 확인해 주세요.
- [\[사용 API\]](#) 설정을 확인해 주세요.

등록하기

이전

네이버 API 사용하기

애플리케이션 등록 (API 이용신청)

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

| | |
|--------------------|---|
| 애플리케이션 이름 | <div>FirstApp ✓</div> <ul style="list-style-type: none">• 네이버 아이디로 로그인할 때 사용자에게 표시되는 이름이므로 가급적 10자 이내의 간결한 이름을 사용해주세요.• 40자 이내의 영문, 한글, 숫자, 공백문자, "-", "_" 만 입력 가능합니다. |
| 사용 API | <div>선택하세요. ▼ ✓</div> <div>검색 ✕</div> |
| 비로그인 오픈 API 서비스 환경 | <div>환경 추가 ▼</div> <div>WEB 설정 ✕ ^</div> <div>웹 서비스 URL (최대 10개)</div> <div>http://localhost + ✓</div> <ul style="list-style-type: none">• 텍스트 폼 우측 끝의 '+' 버튼을 누르면 행이 추가되며, '-' 버튼을 누르면 행이 삭제됩니다.• http와 https는 구분하지 않습니다.• www는 빼고 입력해 주세요. 예) http://naver.com• 서브 도메인이 있으면 대표 도메인명만 입력해 주세요. (예: http://naver.com)• 하이브리드 앱은 location.href 객체 출력 값을 입력하면 됩니다. (예: file://로컬 URL) |

네이버 API 사용하기

NAVER Developers

Products

Documents

Application

NAVER D2

Support

✓ API 상태

Search Here



Application

API 이용을 위해 애플리케이션을 등록하고 API 설정을 할 수 있습니다.

내 애플리케이션



FirstApp

FirstApp

개요

API 설정

멤버관리

로그인 통계

API 통계

Playground(Beta)

애플리케이션 등록

API 제휴 신청

계정 설정

애플리케이션 정보

Client ID

qOt_gq

Client Secret

.....

보기

크롤러에서 사용되므로 기억!!!
타인에게 노출되지 않도록!!!

네이버 서비스의 유형

▶ 비로그인 서비스

- ▶ HTTP 헤더에 Client ID와 Client Secret 값만 넣어서 요청
- ▶ 뉴스, 블로그, 카페 글은 비로그인 방식으로 조회, 저장 가능
- ▶ But, 제약 사항이 존재
 - 1. 하루 검색 기사 수 제한 있음
(API호출이 25,000회/일로 제한, 한번 호출에 최대 100개의 검색을 가지고 올 수 있으므로 250,000건의 데이터 가져올 수 있음)
 - 2. 검색 포인터의 문제
검색 시점에 따라 하나의 키워드의 몇 십만 건의 데이터 중 API로는 최대 1000개만 가져올 수 있음

네이버 검색 API

- ▶ API 공통 가이드

- ▶ <https://developers.naver.com/docs/common/openapiguide/>

- ▶ 블로그 검색

- ▶ <https://developers.naver.com/docs/search/blog/>

- ▶ 뉴스 검색

- ▶ <https://developers.naver.com/docs/search/news/>

비로그인 방식을 통한 뉴스 검색 - 코드

코드

```
import urllib.request

client_id = "여러분의 Client ID"
client_secret = "여러분의 Client Screte"
search_text = urllib.parse.quote("빅데이터")
page_start = 1
display = 5

url =                                     # 'news', 'blog', 'cafearticle' 'movie' 'encyc'
"https://openapi.naver.com/v1/search/news?query=%s&start=%s&display=%s" %(search_text, page_start, display )

reqObj = urllib.request.Request(url)
print(reqObj)
```


비로그인 방식을 통한 뉴스 검색 - 코드

코드

```
reqObj.add_header("X-Naver-Client-Id", client_id)
reqObj.add_header("X-Naver-Client-Secret", client_secret)

htmlObj = urllib.request.urlopen(reqObj)
print(htmlObj)

htmlObj_body = htmlObj.read()
print(htmlObj_body.decode('utf-8'))
```

비로그인 방식을 통한 뉴스 검색 -결과

<urllib.request.Request object at 0x000001E6A8EE0A90>

← reqObj

<http.client.HTTPResponse object at 0x000001E6A94DBA00>

← htmlObj

`{`
"lastBuildDate":"Sun, 31 Jul 2022 11:55:45 +0900",

"total":1168256,

"start":1,

"display":5,

"items":[

`{`

검색 결과를 파일로 저장하려면 어떻게 해야 할까?

"title":"정부, 文 부동산 정책 뒤집기 본격화...시장에선 부작용 속출",

"originallink":"http://www.m-i.kr/news/articleView.html?idxno=93889

7",

"link":"http://www.m-i.kr/news/articleView.html?idxno=938897",

"description":"부동산 빅데이터업체 아실에 따르면 31일 기준 서울 아파트

매물이 중부세 개정 발표 전날인 20일보다 2.0% 감소한 것으로 나타났다. 서초구는 4061건으로 5.5% 감소했고 광진구는 3.4%, 은평구는 2.7%, 성동구는 2.7% 매물이...

"pubDate":"Sun, 31 Jul 2022 11:44:00 +0900"

`}`

6&inflow 97",

id=102",

"title":"은평 혁신파크에 '세대공존형 주택' 공급...싱가포르서 공공주택 구상 밝...",

"originallink":"http://news.heraldcorp.com/view.php?ud=202207310000

"link":"https://n.news.naver.com/mnews/article/016/0002022518?s

"description":"임대주택 계획부터 빅데이터·사물인터넷 등 첨단기술 도입...

디지털 도시' 구현 오 시장의 '서울형... 빅데이터·사물인터넷·인공지능·블록체인 등 첨단기술을 적극적으로 도입해 에너지·물 소비와 폐기물 배출...

"pubDate":"Sun, 31 Jul 2022 11:17:00 +0900"

`}`

`]`

`}`