

# 파이썬 크롤링 - 1

API를 이용한 크롤링

# 목차

01 네이버 API를 이용한 크롤링

02 공공데이터 API 기반 크롤링

# 01. 네이버 API를 이용한 크롤링

## ■ 크롤링이란

### ■ 크롤링

- 웹에서 데이터를 수집하는 작업
- 크롤러 또는 스파이더라는 프로그램으로 웹 사이트에서 데이터를 추출

### • 웹 API

- 웹 API는 일반적으로 HTTP 통신을 사용하는데 사용
- 지도, 검색, 추가, 환율 등 다양한 정보를 가지고 있는 웹 사이트의 기능을 외부에서 쉽게 사용할 수 있도록 사용 절차와 규약을 정의한 것

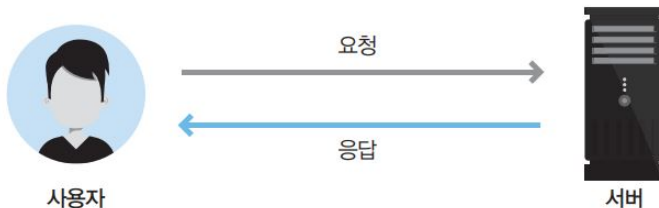


그림 5-1 웹 API를 이용한 HTTP 요청과 응답

표 5-1 웹 API 제공자

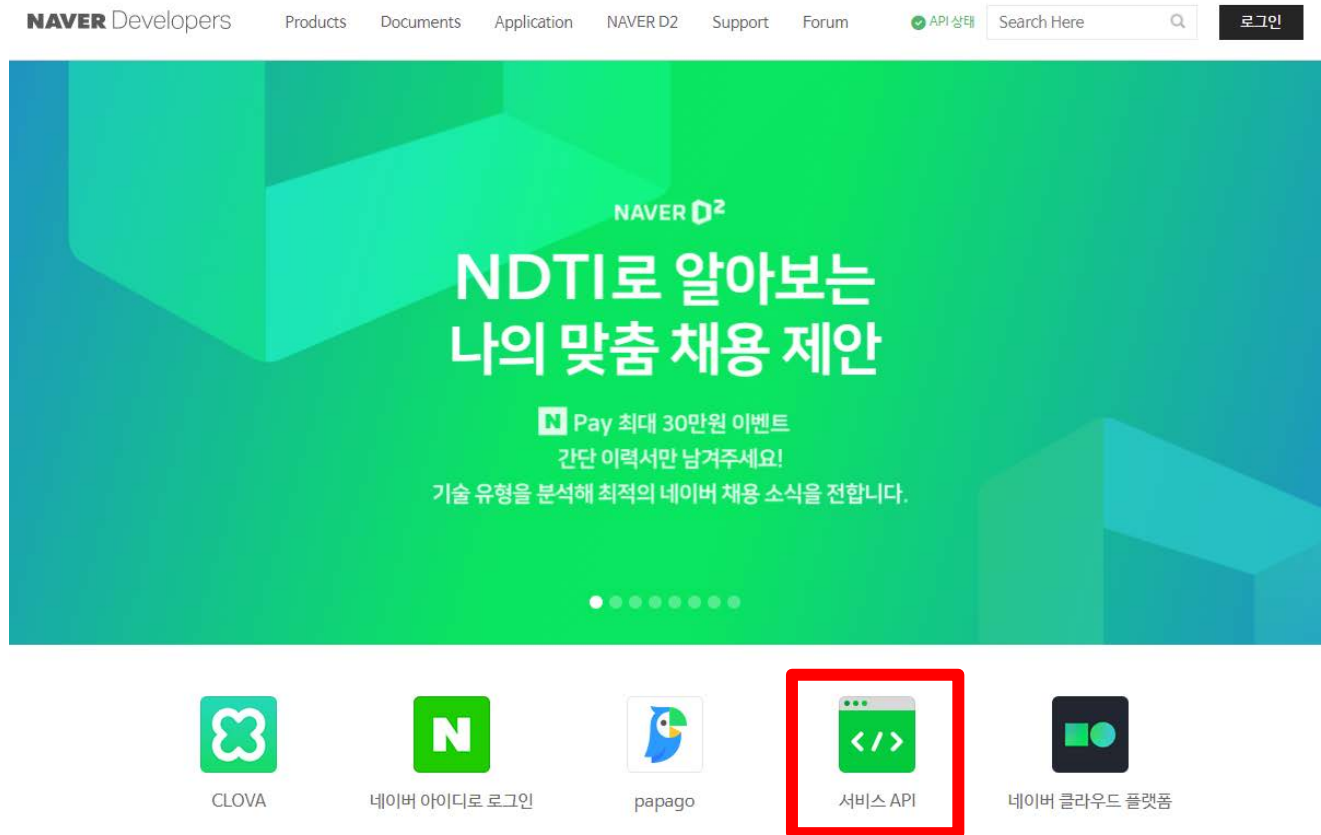
| 종류               | 주소   |
|------------------|--|
| 네이버 개발자 센터       | <a href="https://developers.naver.com">https://developers.naver.com</a>  |
| 카카오 앱 개발 플랫폼 서비스 | <a href="https://developers.kakao.com">https://developers.kakao.com</a>  |
| 페이스북 개발자 센터      | <a href="https://developers.facebook.com">https://developers.facebook.com</a>  |
| 트위터 개발자 센터       | <a href="https://developer.twitter.com">https://developer.twitter.com</a>  |
| 공공데이터포털          | <a href="https://www.data.go.kr">https://www.data.go.kr</a>  |
| 세계 날씨            | <a href="http://openweathermap.org">http://openweathermap.org</a>  |
| 유료/무료 API 스토어    | <a href="http://mashup.or.kr">http://mashup.or.kr</a><br><a href="http://www.apistore.co.kr/api/apiList.do">http://www.apistore.co.kr/api/apiList.do</a> |

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 1. 네이버 개발자 센터 접속하기

- 네이버 개발자 센터([https:// developers.naver.com](https://developers.naver.com))에 접속



# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 2. 오픈 API 이용 신청하기

NAVER Developers

ProductsDocumentsApplicationNAVER D2SupportForum

API 상태

Search Here

로그인

데이터랩

검색

단축URL

캡차

네이버 공유하기

모바일앱 연동

네이버 오픈메인

통합검색어 트렌드

통합검색어 트렌드는 네이버 통합검색에서 발생하는 검색어를 연령별, 성별, 기기별(PC, 모바일)로 세분화해서 조회할 수 있는 API입니다.

분석하고 싶은 주제군을 설정합니다

궁금한 주제어를 설정하고, 하위 주제어에 해당하는 검색어를 쉼표(,)로 구분해 입력합니다. 입력한 단어의 검색 추이를 하나로 합산해 해당 주제가 네이버에서 얼마나 검색되는지 조회할 수 있습니다.

예) 주제어 캠핑 : 캠핑, Camping, 캠핑용품, 겨울캠핑, 캠핑장, 글캠핑, 오토캠핑, 캠핑가, 텐트, 캠핑요리

세분화된 정보를 확인할 수 있습니다

설정된 주제군에 대해 기간을 일간, 주간, 월간 단위로 조회할 수 있습니다. 그리고 5살 단위로 연령을 세분화해 조회할 수 있습니다. 연령을 5살 간격으로 설정한 이유는 연령 이외에 직업군 유추도 어느 정도 가능할 것이라 보기 때문입니다. 성별도 남성, 여성으로 세분화해 조회할 수 있습니다. 마지막으로 PC와 모바일을 각각 구분해서 조회할 수 있습니다.

상대적 값으로만 제공됩니다

검색어 트렌드는 요청된 기간 중 검색 횟수가 가장 높은 시점을 100으로 두고 나머지는 상대적 값으로 제공하고 있습니다. 검색 횟수의 절대값 제공은 아직 고려하고 있지 않습니다. 상대적 값만으로도 특정 주제, 검색어의 트렌드를 파악하는 데는 충분할 것입니다.

오픈 API 이용 신청

네이버 로그인  
(본인 아이디)

쇼핑인사이드

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 2. 오픈 API 이용 신청하기

NAVER Developers

ProductsDocumentsApplicationNAVER D2SupportForum

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.

☒ 이용약관에 동의합니다.

확인

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 2. 오픈 API 이용 신청하기

NAVER Developers

ProductsDocumentsApplicationNAVER D2SupportForum

API 상태

Search Here

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

약관동의

>

계정 정보 등록

>

애플리케이션 등록

계정 설정

① 휴대폰

휴대폰 인증

초기 등록을 위해서 개인 인증이 필요합니다.  
인증은 10분간 유효 합니다.

이메일

이메일 정보 변경

회사 이름 (선택)

☐ 소속 회사 이름은 제휴 접수 및 검토 등을 위해 사용되며, 서비스 이용 기간 동안 보관 됩니다.

&"'() 문자는 허용하지 않습니다.

휴대폰 인증 후 확인

확인

취소

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 2. 오픈 API 이용 신청하기

**NAVER** Developers Products Documents Application NAVER D2 Support Forum API 상태 Search Here

약관동의 > 계정 정보 등록 > **애플리케이션 등록**

### 애플리케이션 등록 (API 이용신청)

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

|             |   |
|-------------|---|
| 애플리케이션 이름 ⇄ | <div><div><b>애플리케이션 이름 :<br/>nvBig 입력</b></div><div>nvBig </div></div> <ul style="list-style-type: none"><li>• 네이버 아이디로 로그인할 때 사용자에게 표시되는 이름이므로 서비스 브랜드를 대표할 수 있는 이름으로 가급적 10자 이내로 간결하게 설정해주세요.</li><li>• 40자 이내의 영문, 한글, 숫자, 공백문자, "-", "_" 만 입력 가능합니다.</li></ul>  |
| 사용 API ⇄    | <div><div><b>사용API :<br/>[검색] 선택</b></div><div><div>선택하세요. </div><div><div>검색</div><div>네이버 (네이버 아이디로 로그인)</div><div>단축 URL</div><div>데이터랩 (검색어트렌드)</div><div>데이터랩 (쇼핑인사이드)</div><div>카페</div><div>캘린더</div><div>캡차 (음성)</div><div>캡차 (이미지)</div><div>한글인명-로마자변환</div><div>Clova Face Recognition</div><div>Papago 번역</div><div>Papago 언어감지</div></div></div></div> |

• [사용 API] 설정을 확인해 주세요.



# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 2. 오픈 API 이용 신청하기

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

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

|                    |   |
|--------------------|---|
| 애플리케이션 이름          | <input type="text" value="nvBig"/> <span>✓</span><br><ul style="list-style-type: none"><li>네이버 아이디로 로그인할 때 사용자에게 표시되는 이름이므로 서비스 브랜드를 대표할 수 있는 이름으로 가급적 10자 이내로 간결하게 설정해주세요.</li><li>40자 이내의 영문, 한글, 숫자, 공백문자, "-", "_" 만 입력 가능합니다.</li></ul>  |
| 사용 API             | <div>선택하세요. <span>✓</span></div> <div>검색 <span>×</span></div>   |
| 비로그인 오픈 API 서비스 환경 | <div>환경 추가 <span>▼</span></div> <div>WEB 설정 <span>×</span></div> <div>웹 서비스 URL (최대 10개)<br/><input type="text" value="http://localhost"/> <span>✓</span><br/><ul style="list-style-type: none"><li>텍스트 폼 우측 끝의 '+' 버튼을 누르면 추가됩니다.</li><li>http와 https는 구분하지 않습니다.</li><li>www는 빼고 입력해 주세요. 예) http://naver.com</li><li>서브 도메인이 있으면 대표 도메인명만 입력해 주세요. (예: http://naver.com)</li><li>하이브리드 앱은 location.href 객체 출력 값을 입력하면 됩니다. (예: file:///로컬 URI)</li></ul></div> |

환경추가 : WEB 설정

비로그인 오픈 API : http://localhost 입력

등록하기

이전

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 2. 오픈 API 이용 신청하기

Client ID 복사 후 Documents 로 이동

NAVER Developers Products **Documents** Application NAVER D2 Support Forum API 상태 Search Here

내 애플리케이션

nvBig

애플리케이션 등록  
API 제휴 신청  
계정 설정

nvBig

|    |        |      |        |        |                   |
|----|--------|------|--------|--------|-------------------|
| 개요 | API 설정 | 멤버관리 | 로그인 통계 | API 통계 | Playground (Beta) |
|----|--------|------|--------|--------|-------------------|

애플리케이션 정보

|               |                                      |  |
|---------------|--------------------------------------|--|
| Client ID     | Client ID와 Client Secret<br>메모장에 복사! |  |
| Client Secret | API 호출량/일일허용량                        |  |

API 호출 안내

지도 API 인증실패나 네이버 로그인 이용 제한이 걸렸다면 [API 설정] 탭에서 URL 관련 설정을 수정하시면 정상 이용 가능합니다 !!!

일일 호출량 25000

비로그인 오픈 API 당일 사용량

|    |         |
|----|---------|
| 검색 | 0/25000 |
|----|---------|

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 5. 검색 API 이용 안내 페이지 확인하기

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

## Documents | 네이버 오픈 API를 이용해 창의적인 애플리케이션을 제작해 보세요.

Documents > [API공통가이드](#)

API 공통 가이드  
네이버 오픈API종류  
사전 준비 사항  
내 애플리케이션  
용어 정리  
샘플 코드  
오류 코드

API 공통 가이드  
CLOVA  
네이버 아이디로 로그인  
파파고  
**서비스 API**

### API 공통 가이드

네이버 오픈API는 네이버 플랫폼의 기능을 외부 개발자가 쉽게 이용할 수 있게 웹 또는 SDK 형태로 공개한 기술들입니다. 네이버 오픈API로 사용할 수 있는 기술에는 네이버 아이디로 로그인과 지도, 검색이 있으며, Clova의 음성 인식 기술과 음성 합성 기술, 얼굴 인식 기술, pigo의 기계 번역 기술 등이 있습니다.

### API 공통 가이드 개요

API 공통 가이드는 네이버 오픈API를 사용해 클라이언트 애플리케이션을 개발할 때 미리 알아 두어야 하는 내용을 설명합니다.

- 최종 수정일: 2017년 8월 16일

이 문서의 내용은 언제든지 변경될 수 있습니다.

### API 공통 가이드 구성

API 공통 가이드의 내용은 다음과 같습니다.

- 네이버 오픈API 종류
  - 로그인 방식 오픈 API
  - 비로그인 방식 오픈 API

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 5. 검색 API 이용 안내 페이지 확인하기

**NAVER** Developers

ProductsDocumentsApplicationNAVER D2SupportForum

API 상태

Search Here

통합 검색어 트렌드

통합 검색어 트렌드 개요

통합 검색어 트렌드 API 레퍼런스

통합 검색어 트렌드 API 구현 예제

- Java

- PHP

- Node.js

**- Python**

- C#

쇼핑인사이트

Python

```
#!/usr/bin/env python
#-*- coding: utf-8 -*-
import os
import sys
import urllib.request
client_id = "YOUR_CLIENT_ID"
client_secret = "YOUR_CLIENT_SECRET"
url = "https://openapi.naver.com/v1/datalab/search";
body = '{"startDate":"2017-01-01","endDate":"2017-04-30","timeUnit":"month","keywordGroups":[{"groupName":"한글"}]';

request = urllib.request.Request(url)
request.add_header("X-Naver-Client-Id",client_id)
request.add_header("X-Naver-Client-Secret",client_secret)
request.add_header("Content-Type","application/json")
response = urllib.request.urlopen(request, data=body.encode("utf-8"))
rescode = response.getcode()
if(rescode==200):
    response_body = response.read()
    print(response_body.decode('utf-8'))
else:
    print("Error Code:" + rescode)
```

[GitHub에서 보기](#)

# 01. 네이버 API를 이용한 크롤링

## ■ 네이버 개발자 가입

### 5. 검색 API 이용 안내 페이지 확인하기

NAVER Developers

ProductsDocumentsApplicationNAVER D2SupportForum

API 상태

Search Here

통합 검색어 트렌드

통합 검색어 트렌드 개요

통합 검색어 트렌드 API 레퍼런스

통합 검색어 트렌드 API 구현 예제

- Java

- PHP

- Node.js

- Python

- C#

소핑인사이트

통합 검색어 트렌드 API 레퍼런스

네이버 통합 검색어 트렌드 조회

오류 코드

네이버 통합 검색어 트렌드 조회

설명

그룹으로 묶은 검색어에 대한 네이버 통합검색에서 검색 후 데이터를 JSON 형식으로 반환합니다.

요청 URL

https://openapi.naver.com/v1/data/lab/search

프로토콜

HTTPS

HTTP 메서드

POST

파라미터

파라미터를 JSON 형식으로 전달합니다.

| 파라미터      | 타입     | 필수 여부 | 설명   |
|-----------|--------|-------|--|
| startDate | string | Y     | 조회기간 시작 날짜(yyyy-mm-dd 형식). 2016년 1월 1일부터 조회할 수 있습니다. |
| endDate   | string | Y     | 조회기간 종료 날짜(yyyy-mm-dd 형식)                            |