

Crawling

DATASCIENCELAB 2019 SUMMER

이주경 임지현 조유나

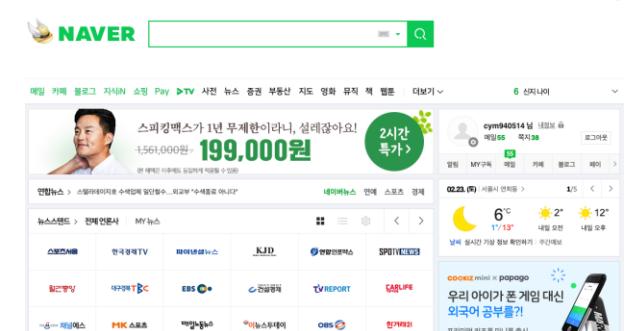
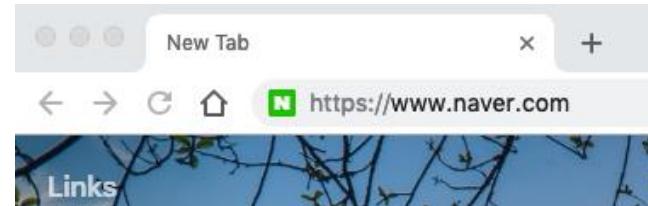


STEP 01 · INTRODUCTION

Web Crawling

» Web Crawling 이란?

- 웹 사이트에서 원하는 정보를 수집하는 과정



Get!



STEP 01 · INTRODUCTION

Web Crawling

» Webpage

- 웹 상에 있는 개개의 문서
- 텍스트, 그림, 소리, 동영상 등을 표현 가능
- 대부분 HTML이라는 언어로 이루어져 있음
 - HTML(hypertext markup language): 웹 문서를 작성하는 일종의 프로그래밍 언어
- HTTP를 통해 HTML 문서를 전송하거나 받아온다.

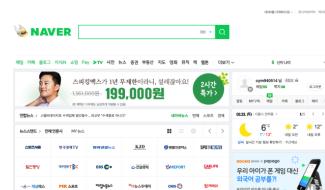


STEP 01 · INTRODUCTION

Web Crawling

» Basics of Web communication

네이버에 가고 싶어



Get!

[https://www.naver.com 가져와!](https://www.naver.com)

```
html code snippet showing various HTML elements like <div>, <img>, <script>, and <style>.
```

Html 코드



HTTP server

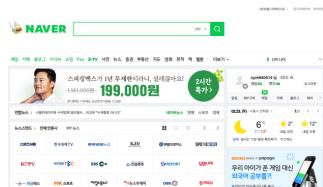
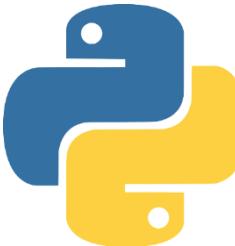


STEP 01 · INTRODUCTION

Web Crawling

» Basics of Web communication

네이버에 가고 싶어



Get!

[https://www.naver.com 가져와!](https://www.naver.com)

```
HTTP/1.1 200 OK
Content-Type: text/html; charset=UTF-8
Content-Length: 123456789
...
<html>
<head>
<title>NAVER</title>
</head>
<body>
...
</body>
</html>
```



HTTP server

Html 코드



STEP 01 · INTRODUCTION

Web Crawling

» 데이터를 수집하는 과정

- 오픈 APIs(Application Programming Interface) 활용
 - Twitter, Facebook, Naver, etc.
- Web scraping (직접 coding)
 - API를 제공하지 않는 사이트
 - API로 수집할 수 없는 데이터
 - Python, R, Java, etc.
 - Ethics are critical.



STEP 02 ... OPEN APIs

Open APIs

» 데이터를 수집하는 과정

- 오픈 APIs(Application Programming Interface) 활용
 - Twitter, Facebook, Naver, etc.
- Web scraping (직접 coding)
 - API를 제공하지 않는 사이트
 - API로 수집할 수 없는 데이터
 - Python, R, Java, etc.
 - Ethics are critical.



STEP 02 ... OPEN APIs

Open APIs

» API (Application Programming Interface)

- 응용 프로그램에서 사용할 수 있도록, 운영 체제나 프로그래밍 언어가 제공하는 기능을 제어 할 수 있게 만든 인터페이스
- 즉, 어떠한 응용 프로그램에서 데이터를 주고 받기 위한 방법을 의미함.
- 오픈 API/비공개 API로 나누어지며 특정 사이트에서 특정 데이터를 공유할 경우 ,어떠한 방식 으로 정보를 요청하고 받을 수 있는지에 대한 규격들을 API라고 부름



STEP 02 ... OPEN APIs

Open APIs

» Open API

- Open API: 누구나 사용할 수 있도록 공개된 API
- Open API를 사용하는 이유?
 - : 개발시 들어가는 시간을 줄이고 비용을 절감할 수 있어 더욱 양질의 앱을 개발할 수 있기 때문
- 대부분 무료로 제공되지만,
 - 호출수에 제한을 두고 있고, 초과사용을 원할 경우, 별도 비용을 받기도 함



STEP 02 ... OPEN APIs

Open APIs

» Open API 사용방법

- 오픈 API 서비스를 이용하기 위해서는 해당 사이트에서 제공하는 **공개키를 발급** 받아야 함
- 승인된 공개키가 있어야만 오픈 API에서 제공되는 데이터를 모두 이용할 수 있음
- 자세한 설명은 이후 직접 api를 사용해보면서 알아보기로 한다.



STEP 03 ... Selenium

Web Crawling

» 데이터를 수집하는 과정

- 오픈 APIs(Application Programming Interface) 활용
 - Twitter, Facebook, Naver, etc.
- Web scraping (직접 coding)
 - API를 제공하지 않는 사이트
 - API로 수집할 수 없는 데이터
 - Python, R, Java, etc.
 - Ethics are critical.



STEP 03 ... Selenium

Web Crawling

» Source codes

- 웹 페이지는 HTML, XML, Javascript 등의 컴퓨터 프로그래밍 언어로 작성된 source code로 구성되어 있음
- 웹 페이지의 Source code 확인 방법
: 마우스 우클릭 → ‘view source codes’ 또는 ‘페이지 소스 보기’

뒤로(B)	Alt+왼쪽 화살표
앞으로(F)	Alt+오른쪽 화살표
새로고침(R)	Ctrl+R
다른 이름으로 저장(A)...	Ctrl+S
인쇄(P)	Ctrl+P
전송(C)...	
한국어(으)로 번역(T)	
페이지 소스 보기(V)	Ctrl+U
검사(N)	Ctrl+Shift+I



STEP 03 ... Selenium

Web Crawling

» Source codes

- Source code는 Tag으로 구성되어 있음
 - : <tag_name> content </tag_name>
- 각각의 Tag는 다른 역할을 수행
 - ex) content : content를 진하게(bold) 할 때 사용하는 태그
- 우리가 모든 태그의 역할을 알아야 할 필요는 없음
- Tag에는 우리가 얻고자 하는 정보들이 담겨있음 (주로 text 형식)



STEP 03 ... Selenium

Web Crawling

» Crawling 과정

- 우리가 얻고자 하는 정보를 얻기 위한 과정
 - 1) 정보를 얻고자 하는 웹 페이지에서 Source Codes를 다운로드 받는다.
 - 2) 우리가 수집하고자 하는 정보를 담고 있는 Tag에 접근한다.
 - 3) 정보를 수집한다.
 - 4) 수집한 정보를 변수 혹은 파일로 저장한다.



STEP 03 ... Selenium

Web Crawling

» Crawling 패키지의 역할

Role	Related module
Web communication => download the (raw) data (i.e., source codes)	Requests/urllib, Selenium
Information extraction => extract the information from the source codes and saves it	BeautifulSoup



STEP 03 ... Selenium

Web Crawling

» Crawling 패키지 비교



VS





STEP 03 ... Selenium

Web Crawling

» Crawling 패키지 비교

1) Selenium

- 브라우저를 컨트롤 해서 데이터를 얻는 방법
- 브라우저를 직접 띄우기 때문에 css, image등 필요하지 않은 정보들도 다운로드 받음
→ 속도가 느림
- 동적 페이지도 크롤링 가능 (javascript 실행 가능)

2) Requests/urllib

- HTTP request를 날려서 데이터를 얻는 방법
- 속도가 빠르다
- Javascripts 실행이 불가능함



STEP 03 ... Selenium

Selenium

» Selenium이란?

- 사람이 브라우저를 직접 동작하는 것처럼, 브라우저의 동작을 자동화 해주는 프로그램
- requests를 통해서 데이터를 수집할 수 없는 경우 사용
- Selenium의 webdriver라는 API를 통해 firefox, chrome 브라우저를 제어
- pip install selenium
`from selenium import webdriver`



S T E P 0 3 ... S e l e n i u m

Selenium

» Selenium이 정보를 수집하는 과정

- 1) 새로운 브라우저 창을 실행한다.
- 2) 사람의 동작 (i.e., clicking) 을 따라 한다.
- 3) 웹 페이지의 Source code를 실시간으로 다운로드 한다.



STEP 03 ... Selenium

Selenium

» Chrome Driver

- Selenium의 webdriver API가 브라우저를 직접 제어할 수 있도록 도와주는 API
- 크롬, 파이어폭스, 사파리 등 다양한 브라우저에서 드라이버를 제공
- Chrome Driver 다운로드 링크
<https://sites.google.com/a/chromium.org/chromedriver/downloads>
→ 본인의 크롬 버전과 맞는 버전 선택 필요
- Chrome Driver 다운받은 경로 꼭 기억하기!! (어디에 다운 받는지는 상관 없음)



STEP 03 ... Selenium

Selenium

» Selenium 실습

- 목표: 멜론 사이트 멜론차트 TOP100 결과를 크롤링하여 데이터 프레임으로 저장해보자!

1	—		안녕 풀 킴	호텔 델루나 OST Part.10	♡ 43,939				
2	—		기억해줘요 내 모든 날과 그때... 거미	호텔 델루나 OST Part.7	♡ 65,610				
3	—		오늘도 빛나는 너에게 (To You... 마크툽 (MAKTUB))	Red Moon : To You My L...	♡ 95,574				
4	—		그대라는 시 태연 (TAEYEON)	호텔 델루나 OST Part.3	♡ 112,278				
5	—		내 목소리 들리니 벤	호텔 델루나 OST Part.9	♡ 29,657				
6	—		니 소식 송하예	니 소식	♡ 122,747				
7	—		술이 문제야 장혜진, 윤민수(바이브)	술이 문제야	♡ 116,835				



STEP 03 ... Selenium

Selenium

» Selenium 실습

- 목표: 멜론 사이트 멜론차트 TOP100 결과를 크롤링하여 데이터 프레임으로 저장해보자!

	song_name	artist	album_name	date	genre	like	review
0	안녕	풀김	호텔 렐루나 OST Part.10	2019.08.12	발라드, 국내드라마	40,112	219
1	기억해줘요 내 모든 날과 그때를	거미	호텔 렐루나 OST Part.7	2019.08.04	발라드, 국내드라마	64,061	263
2	오늘도 빛나는 너에게 (To You My Light) (Feat.이라온)	마크툽 (MAKTUB)	Red Moon : To You My Light	2019.06.09	R&B/Soul, 임디 음악	94,070	406
3	그대라는 시	태연 (TAEYEON)	호텔 렐루나 OST Part.3	2019.07.21	발라드, 국내드라마	111,085	866
4	내 목소리 들리니	벤	호텔 렐루나 OST Part.9	2019.08.11	발라드, 국내드라마	28,144	182
5	니 소식	송하예	니 소식	2019.05.11	발라드	122,012	769
6	포장마차	황인욱	포장마차	2019.06.20	발라드	63,705	304
7	술이 문제야	장혜진, 윤민수(바이브)	술이 문제야	2019.06.18	발라드	116,418	780



STEP 03 ... Selenium

Selenium

» Selenium 실습

1. Melon 페이지 접속
2. 멜론차트 클릭
→ ‘멜론차트’를 클릭하기 위한 ‘멜론차트’ 버튼 element정보



STEP 03 ... Selenium

Selenium

» Selenium 실습

뒤로(B)	Alt+왼쪽 화살표
앞으로(F)	Alt+오른쪽 화살표
새로고침(R)	Ctrl+R
다른 이름으로 저장(A)...	Ctrl+S
인쇄(P)	Ctrl+P
전송(C)...	
한국어(으)로 번역(T)	
페이지 소스 보기(V)	Ctrl+U
검사(N)	Ctrl+Shift+I



The screenshot shows the Chrome DevTools Elements tab open. The code block displays the HTML structure of a menu, specifically a dropdown menu for 'Melon Chart'. A yellow box highlights the 'Elements' tab icon. The code includes comments in Korean explaining the structure, such as '메인 컨텐츠 가운데 정렬을 위한 pr_none 클래스 추가' and '140314_gnb마크업 수정'. The 'menu_bg menu01' class is highlighted in blue, indicating it's selected or being inspected.

```
<!-- 메인 컨텐츠 가운데 정렬을 위한 pr_none 클래스 추가 -->
▶<div id="gnb" class="clfix">...</div>
<!-- 140314_gnb마크업 수정 -->
▼<div id="gnb_menu">
  ▼<ul>
    ▼<li class="nth1">
      ▼<a href="/chart/index.htm" class="cur_menu mlog" data=
        "LOG_PRT_CODE=1&MENU_PRT_CODE=0&MENU_ID_LV1=&
        CLICK_AREA_PRT_CODE=R01&ACTION_AF_CLICK=V1">
        <span class="cur_status none">현재 선택된
        메뉴-</span>
      ...
      <span class="menu_bg menu01">멜론차트
      </span> == $0
      </a>
    </li>
    ▶<li class="nth2">...</li>
    ▶<li class="nth3">...</li>
    ▶<li class="nth4">...</li>
    ▶<li class="nth5">...</li>
    ▶<li class="nth6">...</li>
```



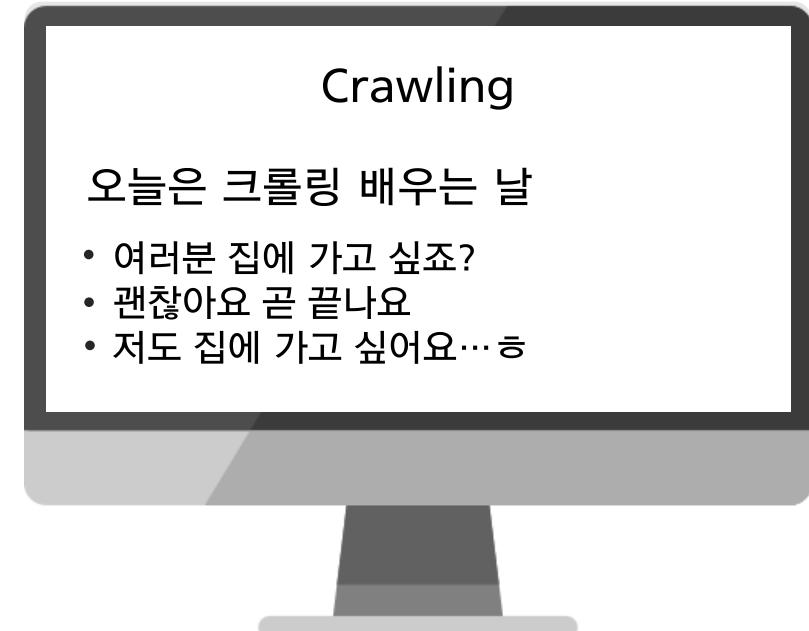
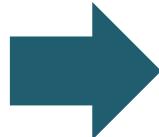
STEP 03 ... Selenium

Selenium

» HTML_tag

1. html 파일은 tag를 단위로 구성되어 있다.

```
<html>
  <head>
    <title>Crawling</title>
  </head>
  <body>
    <h1>오늘은 크롤링 배우는 날</h1>
    <ul>
      <li>여러분 집에 가고 싶죠?</li>
      <li>괜찮아요 곧 끝나요</li>
      <li>저도 집에 가고 싶어요…哼</li>
    </ul>
  </body>
</html>
```





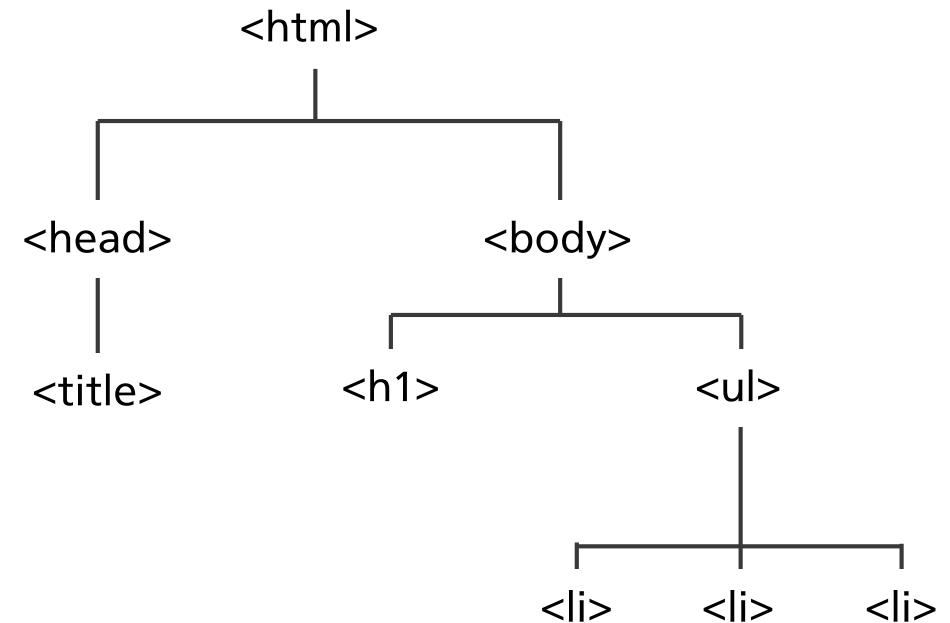
STEP 03 ... Selenium

Selenium

» HTML_tag

2. tag는 tree 구조를 갖는다.

```
<html>
  <head>
    <title>Crawling</title>
  </head>
  <body>
    <h1>오늘은 크롤링 배우는 날</h1>
    <ul>
      <li>여러분 집에 가고 싶죠?</li>
      <li>괜찮아요 곧 끝나요</li>
      <li>저도 집에 가고 싶어요…哼</li>
    </ul>
  </body>
</html>
```





STEP 03 ... Selenium

Selenium

» HTML_tag

3. 각 tag 별로 기능이 다르다. (그 기능을 모두 알아야 할 필요는 없음)

- 단락, 소제목, 역 지정 tag

<p> : 단락을 나타낼 때 사용하는 태그

<h1> ~ <h6> : 소제목을 나타낼 때 사용하는 태그

<div> : 역을 지정할 때 사용하는 태그(공간 단위)

 : 역을 지정할 때 사용하는 태그(줄 단위)

- 목록(리스트) tag

 : unordered list, 순서 없는 목록

 : ordered list, 순서 있는 목록

 : 항목



STEP 03 ... Selenium

Selenium

» HTML_tag

3. 각 tag 별로 기능이 다르다. (그 기능을 모두 알아야 할 필요는 없음)

- 표 관련 tag

<table> : 표

<thead> : 표의 머리 부분 (table head)

<tbody> : 표의 몸통 부분 (table body)

<tr> : 표의 각 행 (table row)

<td> : 표의 데이터 부분 (table data)

<th> : 열 제목, 행 제목 등의 부분 (table header)

The diagram illustrates the structure of an HTML table. It starts with a yellow-bordered <thead> tag at the top, which contains three black th cells. Below it is a red-bordered <tr> tag, which contains three white td cells. This pattern repeats for four rows, each defined by a blue-bordered <tr> tag, containing three white td cells each. The entire structure is enclosed in a dark blue <tbody> tag.

th	th	th
td	td	td



STEP 03 ... Selenium

Selenium

» HTML_tag

4. tag은 그 기능을 더 잘 구현하기 위해 attribute(속성)를 가진다.

- ex) <a> : Text에 링크를 달아주는 태그,
 - 태그 단독으로는 사용하지 않고, href 속성과 함께 사용
 - 구글
- 보통 class와 id 속성으로 태그를 구분하는 경우가 많음



STEP 03 ... Selenium

Selenium

» Locating Elements

- 하나의 페이지 안에서 태그나 속성을 통해 element를 찾을 수 있는 방법

- *find_element_by_id*
- *find_element_by_name*
- *find_element_by_xpath*
- *find_element_by_link_text*
- *find_element_by_partial_link_text*
- *find_element_by_tag_name*
- *find_element_by_class_name*
- *find_element_by_css_selector*

[단수 element]

- *find_elements_by_name*
- *find_elements_by_xpath*
- *find_elements_by_link_text*
- *find_elements_by_partial_link_text*
- *find_elements_by_tag_name*
- *find_elements_by_class_name*
- *find_elements_by_css_selector*

[복수 element]

→ list로 결과 반환

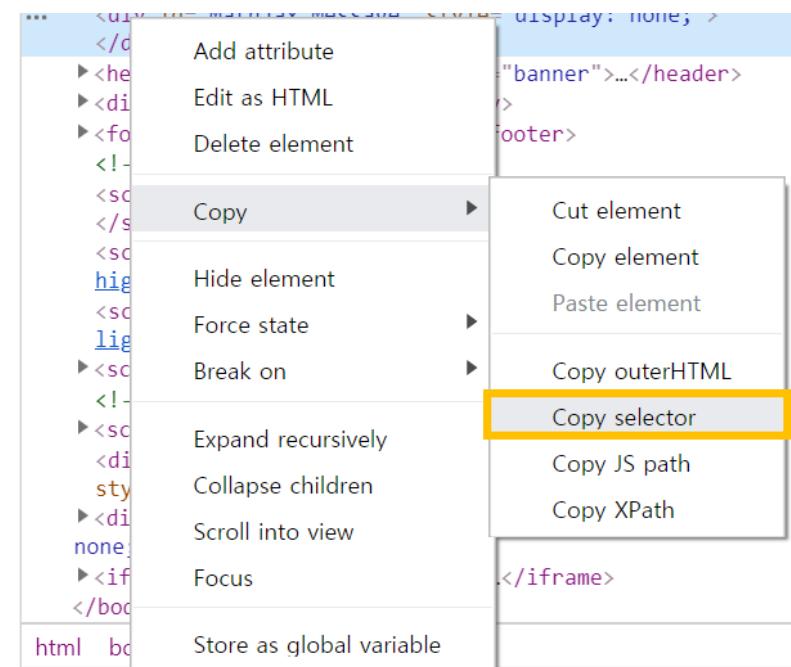


STEP 03 ... Selenium

Selenium

» find_element_by_css_selector

- CSS Selector란?
 - CSS 스타일로 구성된 웹 페이지에서 특정 요소들을 선택해주는 요소
 - CSS: 시각적인 디자인과 레이아웃을 표현하는 언어





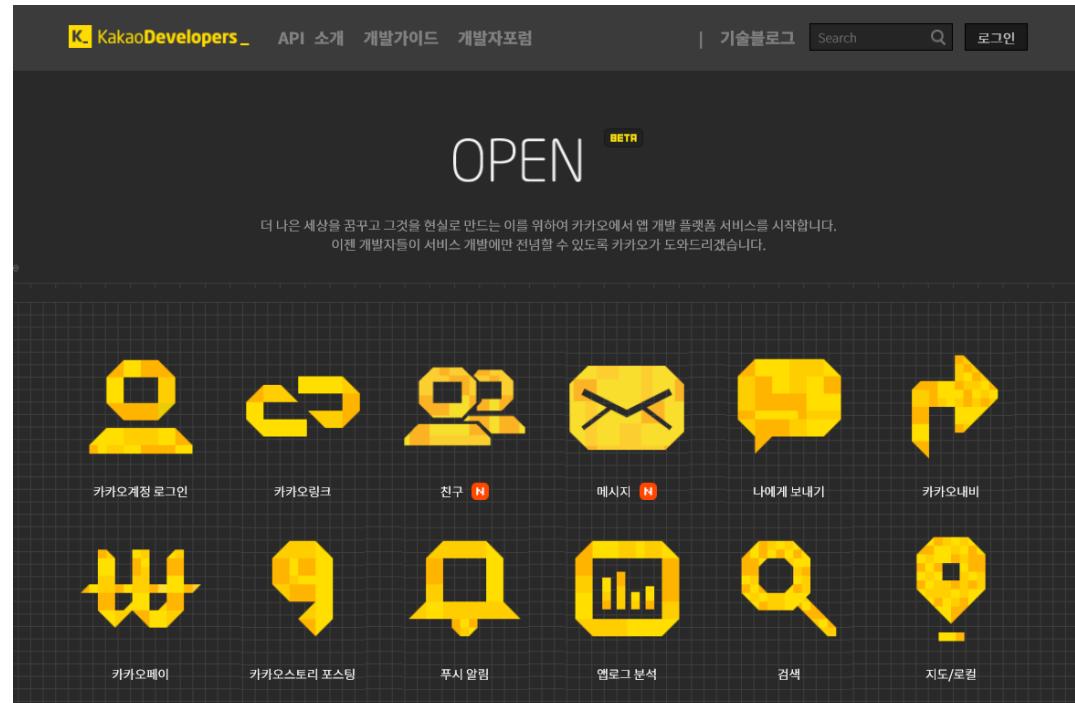
S T E P 0 4 ... Example

Open APIs

» Open APIs 실습

- 목표: 카카오 API를 이용하여 크롤링을 해보자!

준비물: 카카오 개발사이트에서 제공하고 있는 사용자 토큰





S T E P 0 4 ... Example

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

1. Chrome을 이용하여 카카오 개발사이트에 로그인 (본인의 카카오 계정으로)
→ <https://developers.kakao.com>

카카오 계정

자주쓰는 이메일을 등록하여 다양한 카카오 서비스를 이용해보세요!

카카오계정은?
모든 카카오 서비스와 카카오 연계 서비스의 로그인에 사용되는 이메일 형식의 아이디입니다.

카카오톡 정보 복원
기기나 번호 변경 시 기존 카카오톡에 설정한 정보를 복구할 수 있어요.
- 친구목록, 선물받거나 구입한 아이템, 카카오톡 아이디, 프로필 이미지, 상태 메시지

카카오톡과 연결
1개의 카카오톡에는 1개의 카카오 계정만 연결할 수 있어요

kakao

이메일 또는 전화번호 ?

비밀번호

로그인 상태 유지

로그인

[회원가입](#) [카카오계정 찾기](#) | [비밀번호 재설정](#)



S T E P 0 4 ... Example

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

2. 앱만들기 → 앱이름(ex. 이름) & 회사명 입력 후, 앱만들기 클릭

The screenshot shows the 'App Creation' section of the Kakao Developers website. On the left, there's a sidebar with a logo and a dropdown menu set to 'App Creation'. The main area has a title 'App 만들기'. It contains three input fields: 'Icon' (with a placeholder 'APP' and an 'Upload' button), 'App Name' (with a note that it's used for OAuth logins), and 'Company Name' (with a note that it's used for OAuth logins and can affect service usage). A large 'Create App' button is at the bottom.

내애플리케이션

앱 만들기

앱 만들기

아이콘 APP Upload 데스크탑에서는 아이콘 이미지 파일을 끌어서 넣어도 됩니다.

앱 이름

회사명

입력된 앱 이름으로 카카오 로그인 창에 출력됩니다.

입력된 회사명으로 카카오 로그인 창에 출력됩니다.
잘못된 정보를 입력할 경우, 서비스 이용에 제한될 수 있습니다.

앱 만들기

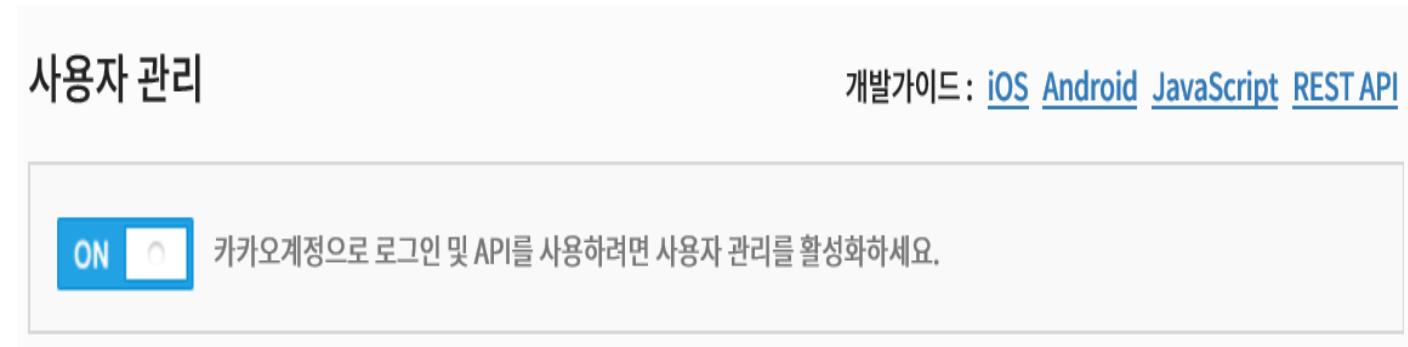


S T E P 0 4 ... E x a m p l e

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

3. 앱 생성 후, [설정] → [사용자 관리] → 사용자 관리를 'ON'으로 활성화





S T E P 0 4 ... Example

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

4. [사용자 관리]의 [로그인 동의항목]에서 수집목적(ex. 크롤링 실습)을 입력 후 저장

The screenshot shows the 'Login Consent' section of the Kakao Developers site. It includes a description of what consent means, tabs for 'Personal Information Protection' and 'Recent Permission Management', and a 'Profile Information (Nickname/Profile Photo)' section where the 'Collection Purpose' field contains '크롤링실습' (Crawling Practice), which is highlighted with a red border. A note at the bottom says '수집목적을 입력해야 합니다.' (A collection purpose must be entered).

로그인 동의항목

로그인을 통해 앱(서비스) 사용자에게 보여주는 동의항목을 설정합니다.
연결 시 최소한의 동의를 받고, 특정 기능 사용 시 추가로 동의를 받을 수 있습니다.
연결 시 동의항목은 필수항목과 선택항목으로 나뉩니다.
필수항목은 사용자가 반드시 동의해야만 가입을 완료할 수 있습니다.

개인정보 보호항목 접근권한 관리항목

● 연결 시 필수 ? ● 연결 시 선택 ? ○ 이용 중 사용 ?

ON 프로필 정보(닉네임/프로필 사진) profile

수집조건 ○ 연결 시 필수 ○ 연결 시 선택 ○ 이용 중 사용

수집목적 크롤링실습 X

수집목적을 입력해야 합니다.

관련 API 카카오스토리 프로필 요청, 카카오톡 프로필 요청, 사용자 정보 요청, 앱 연결



S T E P 0 4 ... E x a m p l e

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

5. [사용자 관리]의 [로그인 Redirect URL]을 아래와 같이 설정 후 저장
(= jupyter notebook 주소)

로그인 Redirect URI

• 카카오 로그인에서 사용할 OAuth Redirect URI를 설정합니다. REST API로 개발하는 경우 필수로 설정하여야 합니다.
• 최대 10개까지 설정할 수 있습니다.
• 추가 URI 등록은 데브톡으로 문의주세요.

|



S T E P 0 4 ... Example

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

6. [앱 정보] → [앱 키 표시] → 자신의 'REST API키' 확인

The screenshot shows a user interface for managing app keys. At the top, there's a header with '앱 정보' and a '설정' link. Below that, a section titled '앱 키' has a '설정' link. The main content lists several key types: '네이티브 앱 키', 'REST API 키', 'JavaScript 키', 'Admin 키', and '설정된 플랫폼'. The 'REST API 키' row is highlighted with a thick yellow border. Each key entry includes a question mark icon for help.



S T E P 0 4 ... Example

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

7. 토큰을 발급받을 수 있는 코드 받기

- [https://kauth.kakao.com/oauth/authorize?client_id=자신의 REST API키
&redirect_uri=플랫폼 사이트 도메인/redirectpath&response_type=code](https://kauth.kakao.com/oauth/authorize?client_id=자신의 REST API키&redirect_uri=플랫폼 사이트 도메인/redirectpath&response_type=code)
→ ‘플랫폼 사이트 도메인/redirect path’은 5번에서 지정한 <http://localhost:8888/tree>로 설정
- 위 주소로 이동 → [동의하고 계속하기] 클릭
- 아래 그림의 ‘code = ~’ 뒤에 부분 코드 내용 확인하기!





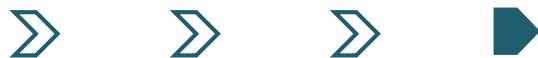
Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

8. 사용자 토큰 받아오기

← → C kauth.kakao.com/oauth/token?grant_type=authorization_code&client_id=eec1bd774d43730235be098ca3c1b945&redirect_uri=http://localhost:8888/tree&code... ☆

```
{"access_token": "mu--IA4DxfD34y0Zz2dJ645nkz8DndSTjWorTAo9dVwAAAFsqnRK7g", "token_type": "bearer", "refresh_token": "6Ch4wKGZUlqvI1_tM_q_b2VMIndwipS4unYI9wo9dVwAAAFsqnRK7A", "expires_in": 21599, "scope": "profile", "refresh_token_expires_in": 5183999}
```



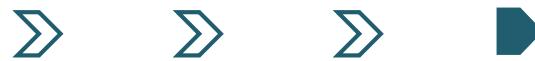
S T E P 0 4 ... Example

Open APIs

» 카카오 개발사이트에서 제공하는 사용자 토큰 받기

9. 사용자 토큰 갱신

- 8번까지 실행 후, 바로 사용할 것이라면 이 과정을 건너 뛰어도 무방
- 위에서 받은 토큰은 access_token으로 발급 받은 후 12 ~ 24시간 동안 유효
(정책에 따라 변동 가능)
- refresh token은 한달 간 유효
→ refresh token 만료가 1주일 이내로 남은 시점에서 사용자 토큰 갱신 요청을 하면,
갱신된 access token과 갱신된 refresh token이 함께 반환됨.
- https://kauth.kakao.com/oauth/token?grant_type=refresh_token&client_id=본인의 rest api 키&refresh_token=8번에서 받았던 refresh token입력



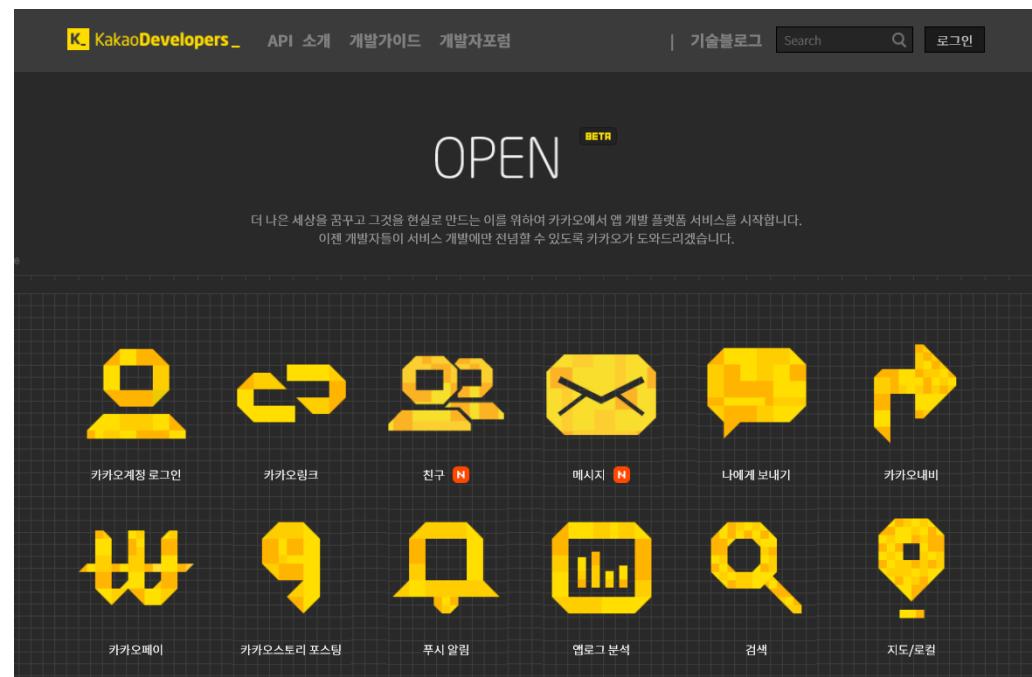
S T E P 0 4 ... Example

Open APIs 실습

» 카카오 API를 이용하여 키워드로 장소 검색하기

- 카카오 API 단점: 개수 제한 있음

→ <https://developers.kakao.com/docs/restapi/local#키워드로-장소-검색>





S T E P 0 4 ... E x a m p l e

Open APIs 실습

» 카카오 API를 이용하여 키워드로 장소 검색하기

키워드로장소검색 URL

질의어에 매칭된 장소 검색 결과를 지정된 정렬 기준에 따라 제공합니다. 현재 위치 좌표, 반경 제한, 정렬 옵션, 페이지 등의 기능을 통해 원하는 결과를 요청 할 수 있습니다. 각 장소는 이름, 주소, 좌표, 카테고리 등의 기본 정보와 다양한 부가정보, 다음 지도의 장소 상세 페이지로 연결되는 URL 정보도 제공됩니다.

[Request]

```
GET /v2/local/search/keyword.{format} HTTP/1.1
Host: dapi.kakao.com
Authorization: KakaoAK {app_key}
```

애키를 헤더에 담아 GET으로 요청합니다. 원하는 검색어와 함께 결과 형식 파라미터의 값을 선택적으로 추가할 수 있습니다.

키	설명	필수	타입
query	검색을 원하는 질의어	0	String



S T E P 0 4 ... Example

Open APIs 실습

» 카카오 API를 이용하여 키워드로 장소 검색하기

앱키를 헤더에 담아 GET으로 요청합니다. 원하는 검색어와 함께 결과 형식 파라미터의 값을 선택적으로 추가할 수 있습니다.

키	설명	필수	타입
query	검색을 원하는 질의어	O	String
category_group_code	카테고리 그룹 코드. 결과를 카테고리로 필터링을 원하는 경우 사용.	X	아래 category_group_code 테이블 참고
x	중심 좌표의 X값 혹은 longitude. 특정 지역을 중심으로 검색하려고 할 경우 radius와 함께 사용 가능.	X	String
y	중심 좌표의 Y값 혹은 latitude. 특정 지역을 중심으로 검색하려고 할 경우 radius와 함께 사용 가능.	X	String
radius	중심 좌표부터의 반경거리. 특정 지역을 중심으로 검색하려고 할 경우 중심좌표로 쓰일 x,y와 함께 사용. 단위 meter	X	0~20000 Integer



S T E P 0 4 ... E x a m p l e

Selenium

» Selenium 실습

- 목표: 멜론 사이트에서 멜론차트 TOP100 결과를 크롤링하여 데이터 프레임으로 저장해보자!

1	—		안녕 풀 킴	호텔 델루나 OST Part.10	♡ 43,939	   
2	—		기억해줘요 내 모든 날과 그때... 거미	호텔 델루나 OST Part.7	♡ 65,610	   
3	—		오늘도 빛나는 너에게 (To You...) 마크툽 (MAKTUB)	Red Moon : To You My L...	♡ 95,574	   
4	—		그대라는 시 태연 (TAEYEON)	호텔 델루나 OST Part.3	♡ 112,278	   
5	—		내 목소리 들리니 벤	호텔 델루나 OST Part.9	♡ 29,657	   
6	—		니 소식 송하예	니 소식	♡ 122,747	   
7	—		술이 문제야 장혜진, 윤민수(바이브)	술이 문제야	♡ 116,835	   