

DATA CRAWLING

이재근

WEB SCRAPING WITH PYTHON

USING BEAUTIFULSOUP

GETTING DATA BY API

USING TWITTER API

REQUIREMENTS

- tweepy
- ~~- tweet-preprocessor~~
- selenium
- bs4
- word_cloud
- Chromedriver.exe → **모듈이 아님.**

WEB SCRAPING WITH PYTHON USING BEAUTIFULSOUP

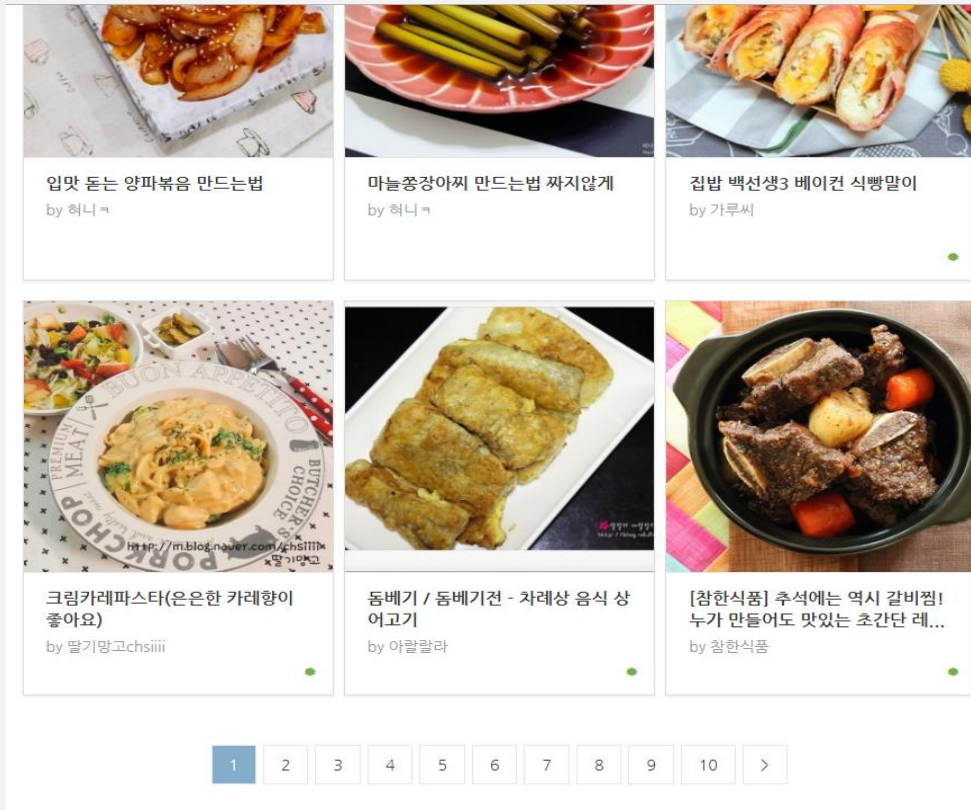
개념 & 원리

WEB SCRAPING 이란? (1)

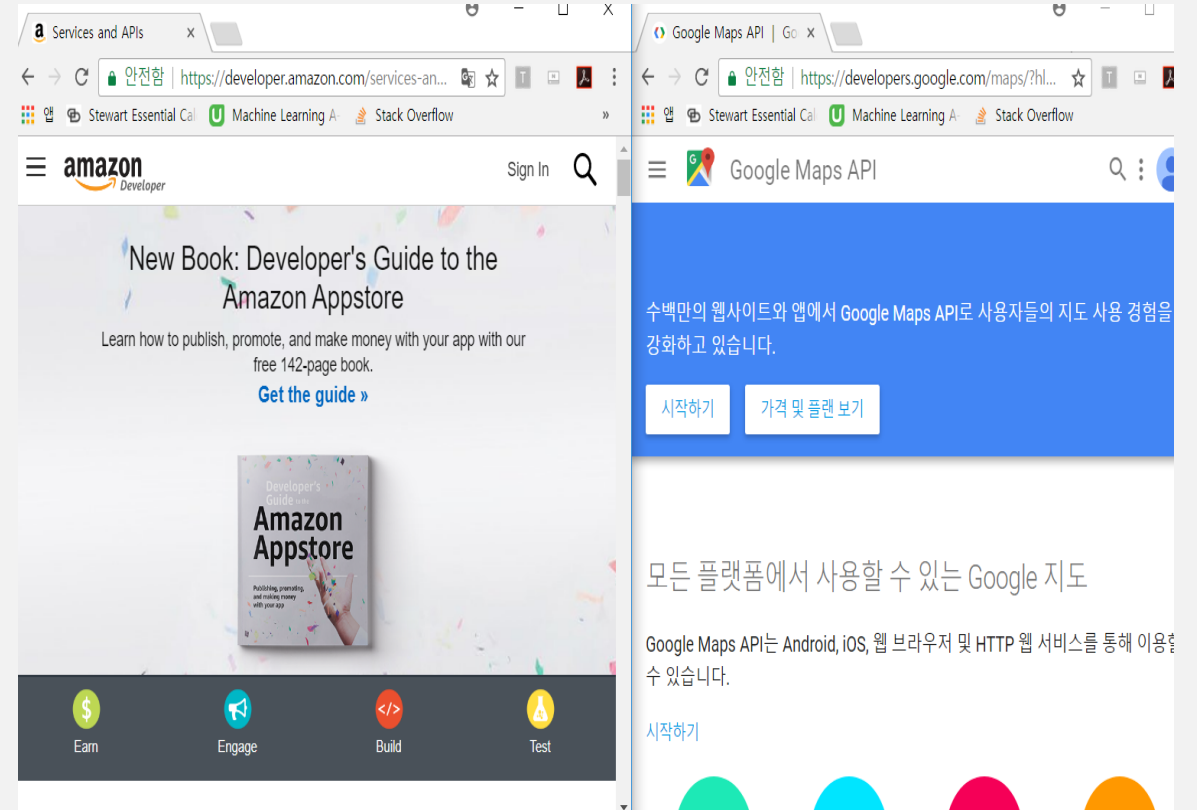
- 웹사이트에서 직접 데이터를 추출 하는 행위
- Data Crawling 이라고도 합니다
- 데이터가 주어지지 않았을 때,
스스로 구하기 위해서



WEB SCRAPING 이란? (2)

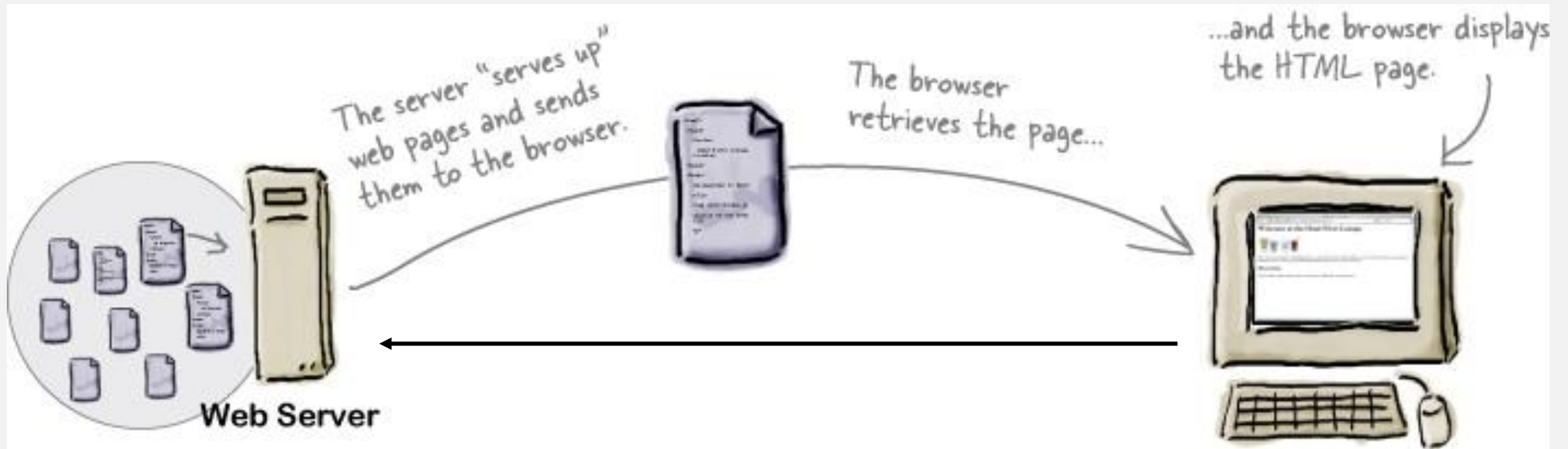


Raw HTML from web browser



Using API to get company data

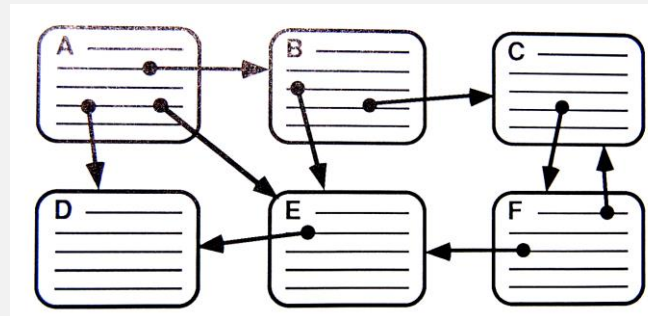
WEB SCRAPING의 원리(1)



서버

HTML 페이지
Hyper Text Markup Language

웹 브라우저
Internet Explorer,
Chrome, Safari, Firefox



Hyper Text
외부 링크로 연결시켜주는 텍스트

WEB SCRAPING의 원리(2)


← → ↻ | 🔒 안전함 | https://ko.wikipedia.org/wiki/HTML#.EC.B5.9C.EC.B4.88_EA.B7.9C.EA.B2.A9 ☆ ⋮

📄 앱 📄 Machine Learning A 📄 Stack Overflow 📄 연세대학교 커리어연 📄 LFD Python Extension Pa 📄 IPython Notebook을 📄 IPython Notebook을 📄 통계방법 문의 합니다 📄 r 📄 Python handbook 📄 yellowid-flask/view.p

로그인하지 않음 토큰 기여 계정 만들기 로그인

div#content.mw-body | 748.8 × 12919.9

읽기 편집 역사 보기 위키백과 검색 🔍



위키백과
우리 모두의 백과사전

대문
사용자 모임
요즘 화제
최근 바뀐
모든 문서 보기
임의 문서로
도움말
기부

도구

여기를 가리키는 문서
가리키는 글의 최근 바뀐
파일 올리기
특수 문서 목록
고유 링크
문서 정보
위키데이터 항목
이 문서 인용하기

인쇄/내보내기
책 만들기
PDF로 다운로드

비활용 관리자의 회수에 대한 지칭 토큰이 열리고 있습니다 [숨기기]

HTML

위키백과, 우리 모두의 백과사전.

HTML은 **하이퍼텍스트 마크업 언어**(HyperText Markup Language, 문화어: 초본문표식달기언어, 하이퍼본문표식달기언어)라는 의미의 웹 페이지를 위한 지배적인 마크업 언어다. HTML은 제목, 단락, 목록 등과 같은 본문을 위한 구조적 의미를 나타내는 것뿐만 아니라 링크, 인용과 그 밖의 항목으로 구조적 문서를 만들 수 있는 방법을 제공한다. 그리고 이미지와 객체를 내장하고 대화형 양식을 생성하는 데 사용될 수 있다. HTML은 웹 페이지 콘텐츠 안의 꺾쇠 괄호에 둘러싸인 "태그"로 되어있는 **HTML 요소** 형태로 작성한다. HTML은 웹 브라우저와 같은 HTML 처리 장치의 행동에 영향을 주는 자바스크립트와 본문과 그 밖의 항목의 외관과 배치를 정의하는 CSS 같은 스크립트를 포함하거나 불러올 수 있다. HTML과 CSS 표준의 공동 책임자인 W3C는 명확하고 표상적인 마크업을 위하여 CSS의 사용을 권장한다.^[1]

HTML	
(HyperText Markup Language)	
개발자	W3C와 WHATWG
최근 버전	HTML5🔗
최근 버전 출시일	2014년 10월 28일
미리보기 버전	HTML 5.1🔗 (초안)
미리보기 버전 출시일	2015년 10월 8일
주요 구현체	TEXT/HTML
영향을 받은 언어	SGML
영향을 준 언어	XHTML
웹사이트	http://www.w3.org/🔗

HTML

```
<!DOCTYPE html>
<html>
<head>
</head>
<body>
</body>
</html>
```

<!-- created 2010-01-01 -->

Elements Console Sources Network Performance Memory >> ⚠ 1 ⋮ ✕

```
<!DOCTYPE html>
<html class="client-js ve-available" lang="ko" dir="ltr">
</html>
<body class="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-subject page-HTML rootpage-HTML vector-nav-directionality skin-vector action-view">
  <div id="mw-page-base" class="noprint"></div>
  <div id="mw-head-base" class="noprint"></div>
  <div id="content" class="mw-body" role="main">
    <div id="mw-notification-area" class="mw-notification-area mw-notification-area-layout" style="display: none;"></div>
    <a id="top"></a>
    <div id="siteNotice" class="mw-body-content"></div>
    <div class="mw-indicators mw-body-content">
    </div>
    <h1 id="firstHeading" class="firstHeading" lang="ko">HTML</h1>
    <div id="bodyContent" class="mw-body-content"></div>
  </div>
  <div id="mw-navigation">
    <h2>둘러보기 메뉴</h2>
    <div id="mw-head"></div>
    <div id="mw-panel">
      <div id="p-logo" role="banner">
        <a class="mw-wiki-logo" href="/wiki/%EC%9C%84%ED%82%A4%EB%B0%B1%EA%B3%BC%EB%8C%80%EB%AC%B8" title="대문으로 가기"></a>
      </div>
      <div class="portal" role="navigation" id="p-navigation" aria-labelledby="p-navigation-label">
        <h3 id="p-navigation-label">둘러보기</h3>
        <div class="body"></div>
      </div>
      <div class="portal" role="navigation" id="p-tb" aria-labelledby="p-tb-label"></div>
      <div class="portal" role="navigation" id="p-coll-print_export" aria-labelledby="p-coll-print_export-label"></div>
      <div class="portal" role="navigation" id="p-wikibase-otherprojects" aria-labelledby="p-wikibase-otherprojects-label"></div>
      <div class="portal" role="navigation" id="p-lang" aria-labelledby="p-lang-label"></div>
```


기본적인 WEB SCRAPING 하는 방법

1. 원하는 웹페이지를 불러온다 - requests
2. 웹페이지의 HTML 소스코드를 가져온다 - requests
3. 필요한 부분만을 선별 적으로 가져온다(parsing) - BeautifulSoup
4. 이를 반복적으로 수행 할 수 있도록 코드를 짠다 - Selenium

크롤링 할 때 유의해야 할 점

- 크롤링 할 웹사이트의 구조를 파악한다
- 특정 사이트는 bot의 접근을 제한한다. (Daum, Naver 등등)
→ DOS 방지, 정보 보호

구조 파악

 comic.naver.com/webtoon/list.nhn?titleId=20853&weekday=tue&page=4



1079. 현 집 줄게

★★★★★ 9.94

2017.03.20



1078. 효(孝) 크러쉬

★★★★★ 9.93

2017.03.13



1077. 캠핑 부자

★★★★★ 9.96

2017.03.06



1076. 거인의 삶

★★★★★ 9.89

2017.02.27



1075. 집이야

★★★★★ 9.94

2017.02.20



1074. 평행 세계

★★★★★ 9.95

2017.02.13



1073. 주인공

★★★★★ 9.94

2017.02.06



1072. 이어서 하자

★★★★★ 9.94

2017.01.30



1071. 소리나는 인형

★★★★★ 9.95

2017.01.23

<이전

1

2

3

4

5

6

7

8

9

10

다음>

구조 파악

① comic.naver.com/webtoon/detail.nhn?titleId=20853&no=1114&weekday=tue

백구3(rnfl**)**

만하 내용에 광고가 없었으면 좋겠다 단순 소품 그림이면 몰라도... 내 확대 해석일지 모르겠지만
에 웃을 못 입고 그로인한 내용이 진행된 것 같다재미는 있지만 작품 내용에 영향주는 광고는 지양
서 더 감사할 것 같다.

2일 전 | 신고

만화 내용에는 광고가 없었으면 좋겠다 단순 소품 그림이면 몰라도... 내 확대 해석일지 모르겠지만
에 옷을 못 입고 그로인한 내용이 진행된 것 같다재미는 있지만 작품 내용에 영향주는 광고는 지양
서 더 감사할 것 같다.

2일 전 | 신고

(zker****)
아빠 좀 그만 괴롭혀라 ㅋㅋㅋ
2일 전 | 신고

아빠 좀 그만 괴롭혀라 ㅋㅋㅋ

2일 전 | 신고

스텔라(stel***)
오늘꺼 = 참신하고 재밌다 뭐 들 그랬지만ㅋㅋㅋㅋㅋㅋㅋㅋㅋ오늘도 힐링하고갑니다 재중
고 알았지만 감독 뭉칠거리는거 너무웃김ㅋㅋㅋㅋㅋ
2일 전 | 신고

오늘까 ㄹㅇ참신하고 재밌다 워늘 그랬지만ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ오늘도 힐링하고갑니다 재충
고 알았지만 감독 움찔거리는거 너무웃김 ㅋㅋㅋㅋㅋ

2일 전 | 신고

오오츠츠키 테우치(exo3****)

마음의 소리 유튜브에서 무료로 볼수있는거 맞아요..

2일 전 | 신고

마음의 소리 유튜브에서 무료로 볼수있는거 맞아요..

2일 전 | 신고

랑(anne****)
1시간 동안 페브리즈 맨 뿌리고 있어 ㅋㅋㅋㅋㅋㅋ
2일 전 | 신고

1시간 동안 페브리즈 맨 뿌리고 있어 ㅋㅋㅋㅋㅋㅋ

2일 전 | 신고

그미(2lov****)
점점 광고가되는가봄

점점 광고가되는가봄

접근 제한



Daum첫화면 · 고객센터



서비스 이용에 불편을 드려 죄송합니다.
일시적으로 서비스 이용이 제한되었습니다.

비정상적인 접근이 발견되어 이용이 제한되었습니다.
다시 정상적인 서비스 이용을 위해서는 아래 그림문자를 입력하셔야 합니다.
그림문자를 정확하게 입력한 후에도 계속 검색결과가 보이지 않는다면 [고객센터](#)로
문의해 주시길 바랍니다.
감사합니다.



[새로고침](#)

보이는 순서대로 숫자 및 문자를 모두 입력해주세요.

[확인](#)

간단 크롤링

```
import requests
from bs4 import BeautifulSoup

url = requests.get("http://corners.gmarket.co.kr/Bestsellers?viewType=G&groupCode=G068")

html = url.text
soup = BeautifulSoup(html, 'html.parser')

print(soup)
```

```
<p class="no100" id="no100">100</p>
<div class="thumb">
<a href="http://item2.gmarket.co.kr/Item/DetailView/Item.aspx?goodscode=970192620
" onclick="pdsClickLog('2000000680', 'Item', {'ASN': 100, 'goodsCode': '970192620'
});"></a>
</div>
<!--div class="goods-view">
<a href="
http://minishop.gmarket.co.kr/bestpickup"><span class="view">판매자 다른상품 보기<
```

```
In [10]: soup.findAll('a',class_='itemname')|
```

```
Out[10]: [<a class="itemname" href="" id="topPlusItemName0"></a>,
  <a class="itemname" href="" id="topPlusItemName1"></a>,
  <a class="itemname" href="" id="topPlusItemName2"></a>,
  <a class="itemname" href="" id="topPlusItemName3"></a>,
  <a class="itemname" href="" id="topPlusItemName4"></a>,
  <a class="itemname" href="http://item2.gmarket.co.kr/Item/DetailView/Item.aspx?g
oodscode=1128657797" onclick="pdsClickLog('200000680', 'Item', {'ASN': 1, 'goodsC
ode': '1128657797'});">로지텍 무선마우스 M238 카카오에디션/라이언 (Ryan)</a>,
  <a class="itemname" href="http://item2.gmarket.co.kr/Item/DetailView/Item.aspx?g
oodscode=1129412432" onclick="pdsClickLog('200000680', 'Item', {'ASN': 2, 'goodsC
ode': '1129412432'});">로지텍 무선마우스 M238 카카오에디션/어피치(Apeach)</a>,
  <a class="itemname" href="http://item2.gmarket.co.kr/Item/DetailView/Item.aspx?g
oodscode=1129415109" onclick="pdsClickLog('200000680', 'Item', {'ASN': 3, 'goodsC
ode': '1129415109'});">로지텍 무선마우스 M238 카카오에디션/무지 (Muzi)</a>,
  <a class="itemname" href="http://item2.gmarket.co.kr/Item/DetailView/Item.aspx?g
oodscode=909553255" onclick="pdsClickLog('200000680', 'Item', {'ASN': 4, 'goodsCo
de': '909553255'});">삼성전자 86 PC4 19200 데스크탑 메모리</a>,
  <a class="itemname" href="http://item2.gmarket.co.kr/Item/DetailView/Item.aspx?g
oodscode=776015847" onclick="pdsClickLog('200000680', 'Item', {'ASN': 5, 'goodsCo
```

```
In [13]: items = soup.findAll('a',class_='itemname') # list 형태임

for item in items:
    print(item.get_text())
```

로지텍 무선마우스 M238 카카오에디션/라이언 (Ryan)
로지텍 무선마우스 M238 카카오에디션/어피치(Apeach)
로지텍 무선마우스 M238 카카오에디션/무지 (Muzi)
삼성전자 8G PC4 19200 데스크탑 메모리
[로지텍]10%추가할인+키스킨 로지텍코리아 MK235 무선콤보
[아이피타임]오늘출발 IPTIME A2003ns-MU 공유기/와이파이/무선
무료배송 키보드/최저가/브랜드정품/삼성/게이밍/세트
ABKO B510U PRO Virtual 7.1CH RGB 게이밍 진동헤드셋

WEB SCRAPING WITH PYTHON USING BEAUTIFULSOUP

OHMYCOMPANY 데이터 크롤링

OHMYCOMPANY

OHMYCOMPANY

프로젝트 보기

프로젝트 신청

가이드



사회적기업 특화

중권형

로그인

회원가입

한국치즈의 원조고장 전북임실에서 온 임실농부

전주하면 초코파이. 임실농부하면 치즈초코파이!!



89,541명 4,074,115,019원



농식품전문관 관광전문관

Oh my choice!



임실농부가 만든 치즈초코파이



래놀의 신문 화분이야기



[광주] 소중한 인연을 만들어가는
여행



[남해] 웃어라 꽃섬남해로의 여행

목표: 모든 프로젝트의 이름, 내용, 규모 등 상세 정보 가져오기!

프로젝트 보기
프로젝트 신청
가이드

Oh! My Choice

산속에서 황금을 찾아라! 허니플러스

시흥양봉협동조합
사용의달달한하네들 천연벌꿀

14% 280,000원

[영주]소백산골 다문화새댁들의 `제2고향만들기`

영주시다문화회랑공동체
`모두의꿈' 함께할수있는공간을니룹니다

100% 3,020,000원

[포항]폐지수거노인 리어카

포항시 사회적기업 협의회
폐지수거노인에게세리어카와광고비를지원합니다

109% 3,290,000원

들기름 세계화를 위한 하이브리드 라벨

에버그린에버블루협동조합
코리아올리브오일로홍보들기름의고유성을세계화

5% 109,000원

What's Popular

희미해지지 않게 묶다, 노란리본 이음팔찌

온프로젝트
우리의마음이희미해지지않게 두번매노란프로젝트

[포항]무연분묘 벌초 및 독거노인 집수리 봉사

대명장례협동조합
무연분묘벌초및독거노인집수리봉사

[청도]청도에서 온 국악 청년의 이야기

온누리국악예술인협동조합
청도토박이청년국악인들이들려주는청도팔경

[영천] 흙과 불의 흔적, 그 전통을 빛다.

흙과인 협동조합
사라져가는 지역전통문화를채움이다.

See all >


웹페이지 구조 탐색

→ 안전함 https://www.ohmycompany.com/project/prjView.php?bbs_code=won_project&seq=1590

Stewart Essential Cal Stack Overflow 연세대학교 커리어엔 Building Spark - Spa LFD Python Extension Pa IPython Notebook IPython Notebook Jupyter Notel

OHMYCOMPANY 프로젝트 보기 프로젝트 신청 가이드

산속에서 황금을 찾아라! 허니플러스



280,000원
목표액 2,000,000원 중 14%모집

4명
참여자수

D-23
마감일 2017.07.23

참여하기

시홍양행협동조합
message profile

shhoney.modoo.at
07077697173


공유수 5 페이스북 트위터

→ 안전함 https://www.ohmycompany.com/project/prjView.php?bbs_code=won_project&seq=1487

Stewart Essential Cal Stack Overflow 연세대학교 커리어엔 Building Spark - Spa LFD Python Extension Pa IPython Notebook IPython Notebook Jupyter

OHMYCOMPANY 프로젝트 보기 프로젝트 신청 가이드

[포항] 폐지수거노인 리어카



3,290,000원
목표액 3,000,000원 중 109%모집

20명
참여자수

D-5
마감일 2017.07.05

참여하기


포항시 사회적기업협의회
message profile

https://www.ohmycompany.com/project/prjView.php?bbs_code=won_project&seq=1444

al Cal Stack Overflow 연세대학교 커리어엔 Building Spark - Spa LFD Python Extension Pa IPython Notebook IPython Notebook Jupyter

프로젝트 보기 프로젝트 신청 가이드

생각 그 이상의 휠체어



177,000원
목표액 3,000,000원 중 6%모집

7명
참여자수

D-15
마감일 2017.07.15

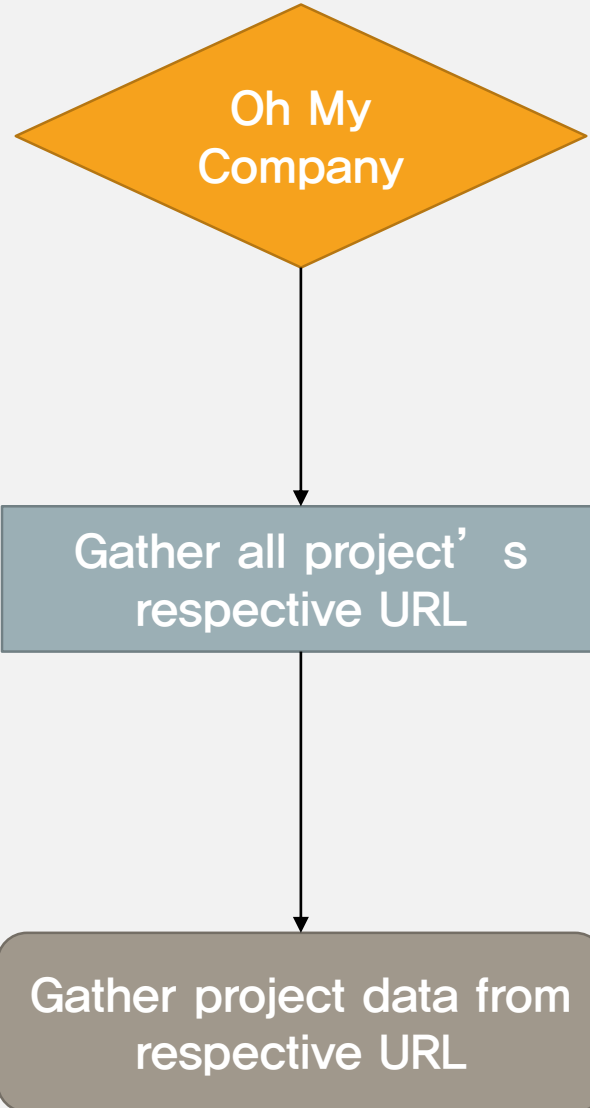
참여하기

(주)휠라인
message profile

www.wheel-line.com
www.facebook.com/주휠라인-Wheel-...
0317341674

신제품 | 파북공동체 공유수 0 페이스북 트위터

코딩 순서도



개별 URL 가져오기

중권형 로그인 회원가입

의 모든 프로젝트

교육/출판

신제품

문화예술

지역사회

테크

파트너십 펀딩대회

#여행

#위안부

#경북

#따복공동체

#지속가능경영재단



a#link | 205.45 x 19.2

연잎으로 담근 장, 맛보실래요?

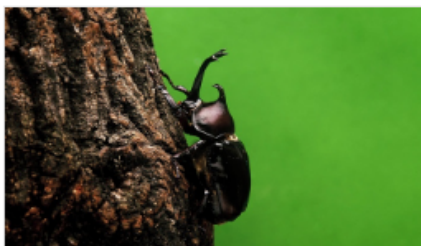


시종시장독대협동조합

콩으로만드는FUN한세상

16%

336,000원



꽃벵이를 아시나요?



협동조합 시종호도회

모두에게유익한공병이랍니다

19%

















390,000원

```
</ul>
</div>-->
<div class="prj_box">
  <ul class="prj_box_list clearF" id="updates">
    <!-- 제일 왼쪽 contents 마진왼쪽 삭제 class first 추가-->
    <li class="prj_box_item fst">...</li>
    <li class="prj_box_item ">...</li>
    <li class="prj_box_item ">
      <div class="prj_figure">
        <div class="irap">...</div>
        <div class="figureCaption">
          <h3 class="tit_h3">
            <a href=" ../project/prjView.php?
              bbs_code=von_project&seq=1583" id="link">연잎으
              로 담근 장, 맛보실래요?</a>
          </h3>
          <p class="by">...</p>
          <div class="txt_wrap">...</div>
          <div class="bottom">...</div>
          <!-- //bottom-->
        </div>
      </div>
      <!-- 공유 아이콘 시작-->
      <script type="text/javascript">...</script>
      <div class="item_share">...</div>
      <!-- 공유 아이콘 끝-->
    </li>
    <li class="prj_box_item ">...</li>
    <li class="prj_box_item fst">...</li>
    <li class="prj_box_item ">...</li>
    <li class="prj_box_item ">...</li>
    <li class="prj_box_item ">...</li>
    <li class="prj_box_item fst">...</li>
    <li class="prj_box_item ">...</li>
    <li class="prj_box_item ">...</li>
  </ul>
</div>
```

개별 URL 가져오기

This button prevents us from scraping all project data...

[프로젝트 보기](#) [프로젝트 신청](#) [가이드](#)

 <p>단편영화 <고래> 제작 프로젝트</p> <p> 손지은</p> <p>해마(해마)인하는 모든 것들에 대하여</p> <p>75% 마감 450,000원</p>	 <p>[고령]세련되게 맛있다, 대가야 참기름</p> <p> 참고소한다신평동조합</p> <p>대가야전통의맛을 소개합니다.</p> <p>97% 970,000원</p>	 <p>[영양]자연먹고 곤드레 만드레</p> <p> 영농조합법인소청</p> <p>중장년층건강식품물은송다아트어른!아는바른역거리</p> <p>22% 440,000원</p>	 <p>[문경]점촌의 뽕뽕한마켓 점뽕</p> <p> 복합문화공간 사이</p> <p>지역사회에점은활력을불어넣는버룩시장</p> <p>28% 853,000원</p>
 <p>[성주]하하수미와 함께 떠나는 별마실 나들이</p> <p> 이수미</p> <p>별마실(성주)로떠나는신나는시골여행</p> <p>49% 1,470,000원</p>	 <p>[영주]소백산골 다문화새댁들의 '제2고향 만들기'</p> <p> 영주시다문화희망공동체</p> <p>'모두'의꿈함께할수있는공간을나눔합니다</p> <p>117% 3,520,000원</p>	 <p>[청송]청송의 밤과 예술은 길다. 마을예술영화관</p> <p> 농업회사법인주식회사바오바트</p> <p>마을영화관만들기프로젝트</p> <p>25% 775,000원</p>	 <p>[군위]농산물 공동재배와 이웃나눔 사랑</p> <p> 군위군 군위읍 새마을지도자회/부녀자회</p> <p>공유지대부 농작물 공동재배를 통한 이웃나눔 실천</p> <p>101% 3,040,000원</p>

[더보기](#)

개별 URL 가져오기

```
from bs4 import BeautifulSoup
from selenium import webdriver
```

Use BeautifulSoup to parse url

Use Selenium Webdriver to automate clicking “see more” button



```
driver = webdriver.Chrome('/Users/agdal/Desktop/chromedriver')
driver.set_window_size(1120, 800)
driver.get("https://www.ohmycompany.com/project/prjList.php?l_kind=&contest=&sort=new")
```

```
a = 2
```

```
for i in range(1,29):
    driver.find_element_by_id(str(a)).click()
    driver.implicitly_wait(60)
    a = a+1
```

Automating chromedriver to click “see more” button



```
html = driver.page_source
```

```
soup = BeautifulSoup(html, "html.parser")
links = soup.select("h3 > a[href]")
```

Crawling via BeautifulSoup

Using <h3><href> tags to find each project's url



```
f = open('omcprojects.txt', 'w')
```

```
for link in links:
    curLink = link.get('href')
    f.write("www.ohmycompany.com/project" + curLink + "\n")
```

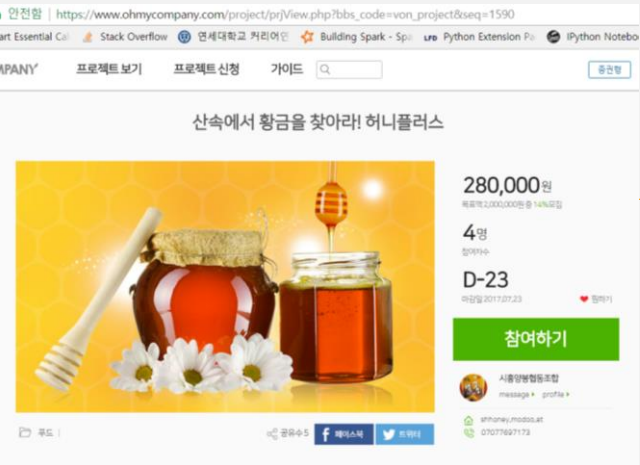
Writing a text file to save links



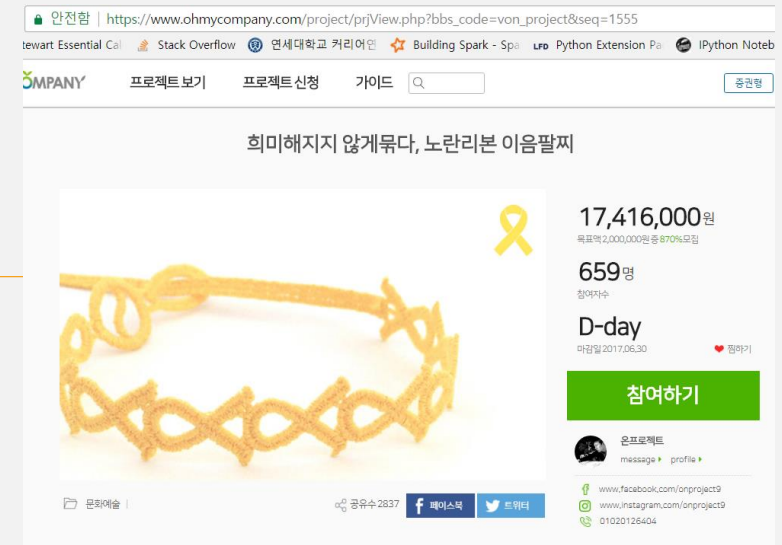
```
f.close()
```


개별 URL 가져오기 - 결과

Omcprojects.txt



http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1590
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1589
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1583
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1576
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1572
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1566
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1562
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1557
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1556
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1545
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1540
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1537
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1532
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1528
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1526
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1519
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1484
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1483
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1444
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1438
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1422
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1555
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1559
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1490
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1551
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1553
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1542
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1533
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1530
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1522
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1513
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1511
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1509
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1507
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1504
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1500
http://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1499



데이터 크롤링

Now that we have each project' s page with data, crawling code for individual page is required

Because each project' s page has similar structure, we may code crawling for 1 url page and use “for” loop to crawl the rest

데이터 크롤링

Using Google Chrome Developer Option,
you can find where your information is and what kind of html tag it has

산속에서 황금을 찾아라! 허니플러스



.number | 126.88 x 37.33

280,000 원

목표액 2,000,000원 중 14%모집

4명

참여자수

D-23

마감일 2017.07.23

♥ 찜하기

참여하기



시흥양봉협동조합

message ▶ profile ▶

shhoney.modoo.at

07077697173

공유수 5

f 페이스북

트위터

```
▼<div id="subContainer" style="padding:0;">
  ▼<div id="subContents">
    <!-- 서브 상세 내용-->
    ▼<div id="detailSub">
      ▼<div class="detail-view-bg">
        ▶<h3 class="h3-tit">...</h3>
        ▼<div class="detail-view">
          ▶<div class="leftConts">...</div>
          <!--//leftConts-->
          ▼<div class="rightConts">
            <!-- 투자자 시작-->
            ▼<div class="spend boxCorner">
              <h3 class="visible">참여자</h3>
              ▼<ul class="s_list">
                ▼<li class="s_item">
                  ▼<p class="money">
                    <strong class="number">280,000</strong>
                    <span class="unit">원</span>
                  </p>
                  ▶<p class="txt">...</p>
                </li>
                ▶<li class="s_item">...</li>
                ▶<li class="s_item">...</li>
              </ul>
              ▶<p class="invested">...</p>
            </div>
            <!-- 투자자 끝-->
```

데이터 크롤링

```
import requests
from bs4 import BeautifulSoup

line = 'https://www.ohmycompany.com/project/prjView.php?bbs_code=von_project&seq=1576'
source_code = requests.get(line)
plain_text = source_code.text
soup = BeautifulSoup(plain_text, 'lxml')
div = soup.findAll('em')
comp = []

for k in div:
    x = k.get_text()
    y = (str.split(x))
    comp.append(y)

compname = separator.join(comp[2])
print(compname)
```

← Project Url

←

- Parsing with BeautifulSoup
- Finding tag
- Making an empty list

←

- Getting texts only, inside tag
- Splitting and appending the text into empty list

←

- Finding desired information from the list
- Checking whether you have the right data or not

데이터 크롤링

```
import requests
from bs4 import BeautifulSoup
import csv
```

```
tsvfile = open('omc_data.tsv', 'w', encoding='utf-8', newline='')
```

← Creating a tsv format file using csv module

```
with open('omcprojects.txt') as f:
    lines = f.read().splitlines()
```

← Opening omcprojects text file that contains URLs of all projects

```
for i in range(0,895):
```

← Using “for” loop to create a variable with URL and repeat the code below

```
    url = lines[i]
```

```
    source_code = requests.get(url)
    plain_text = source_code.text
    soup = BeautifulSoup(plain_text, 'html.parser')
    pars = soup.findAll('p')
```

```
    data = []
```

```
    for line in pars:
        raw = line.get_text()
        clean = (str.split(raw))
        data.append(clean)
```

```
    category = data[0][0].replace('|', '')
    current = data[2][0]
    goal = data[3][1]
    participant = data[4][0]
    due = data[7][1].replace('썹하기', '')
```

← Crawling for 1 URL page

```
    wr = csv.writer(tsvfile, delimiter='Wt')
    wr.writerow([category, project, compname, current, goal, participant, due])
```

← Writing the data into tsv format & closing file

```
tsvfile.close()
```

결과

Data Crawling Complete!

omc_data - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

푸드	산속에서	황금을 찾아라!	허니플러스	시흥양봉협동조합	280,000	2,000,000원	4	2017.07.23																								
지역사회	하하하!	영계씨	(주)비알인	토크	540,000	2,000,000원	7	2017.07.23																								
푸드	영민이	로	당근 장,	맛보실래요?	시흥시장독대협동조합	336,000	2,000,000원	4	2017.07.23																							
일자리	꽃병이	를	아시나요?	협동조합	시흥효도회	390,000	2,000,000원	5	2017.07.23																							
교육/출판	영문법을	이해하며,	생각	넓히기	와이넷	77,000	1,000,000원	2	2017.07.10																							
사회이슈	따복	영민이	디저트	판매	수량만큼	기부가 되는	카페	한즈곳	84,000	2,000,000원	4	2017.07.15																				
지역사회	따복	영민이	마음을	공유하는	휴먼서비스,	희망이	한마음	희망나눔센터	100,000	2,000,000원	2	2017.07.15																				
사회이슈	따복	영민이	위기의	아이들에게	꿈과	희망을	원예라이프	코칭연구소	100,000	2,000,000원	4	2017.07.15																				
지역사회	따복	영민이	당신의	꿈,	꿈(honey)	이	함께해요	협동조합	함박꽃웃음	100,000	2,000,000원	4	2017.07.15																			
푸드	따복	영민이	막	내려도	맛있는	드립백	커피	농부들의	카페장터	120,000	2,000,000원	4	2017.07.15																			
푸드	따복	영민이	건강하고	맛있게	발효시킨	우리	축산물	화성시발효식품협동조합	1,020,000	2,000,000원	16	2017.07.15																				
문화예술	따복	영민이	환경을	살리는	핸드메이드	나누리	창작공방	협동조합	190,000	3,000,000원	5	2017.09.15																				
푸드	따복	영민이	병아리	에서	젖소,	꿈을	실현하다	농업회사법인(주)꿈목장	2,630,000	2,000,000원	102	2017.07.15																				
푸드	따복	영민이	동면지로	행복한	마을	만들기	농업회사법인	동탄지마울주식회사	110,000	2,000,000원	5	2017.07.15																				
환경	재생	따복	과일	상징	아이	안심	과일	세척제	(주)에코바이오	4,438,000	2,000,000원	9	2017.07.15																			
사회이슈	따복	영민이	건전한	반려동물	문화	만들기	드로잉	왈츠	바우앤뮤	협동조합	100,000	2,000,000원	4	2017.07.15																		
환경	재생	따복	자연을	신다,	천연	염색	오가니	코튼	양말	(주)아트앤크래프트	1,150,000	2,000,000원	20	2017.07.15																		
푸드	따복	영민이	경력	단절	여성들,	바리스타	로	다시	일어서다	오앤오커피협동조합	360,000	2,000,000원	12	2017.07.15																		
신제품	따복	영민이	생각	그	이상의	유희체어	(주)유희라인	177,000	3,000,000원	7	2017.07.15																					
푸드	따복	영민이	즐거움	세계화	를	위한	하이드리드	라벨	에버그린	에버블루	협동조합	109,000	2,000,000원	4	2017.07.15																	
교육/출판	따복	영민이	엄마가	선택	하는	보드게임	<로컬푸드한끼>	(주)오즈하우스	725,000	3,000,000원	15	2017.07.15																				
문화예술	따복	영민이	희미	해지	지	않게	된다,	노란리본	이음팔찌	온프로젝트	17,130,000	2,000,000원	851	2017.06.30																		
신제품	따복	영민이	색	다	를	추출	기기,	보체티	모카팟	맥앤코리아	199,400	1,000,000원	3	2017.07.19																		
문화예술	따복	영민이	[정주]	전통	매듭	놀이	유객주	(주)아트세상	124,000	3,000,000원	7	2017.07.05																				
문화예술	따복	영민이	단편영화	<고래>	제작	프로젝트	손지은	450,000	600,000원	22	2017.06.16																					
지역사회	영민이	영민이	[고령]	세련	되게	맞았다,	대가야	참기름	참고소한	다산협동조합	970,000	1,000,000원	28	2017.07.05																		
푸드	경북	영민이	[영양]	자	영양	먹고	고드레	만드레	영농조합법인	소청	440,000	2,000,000원	13	2017.07.05																		
지역사회	영민이	영민이	[문경]	점촌	의	뽕뽕한	마켓	점빵	복합문화공간	사이	853,000	3,000,000원	79	2017.07.05																		
지역사회	영민이	영민이	[상주]	하하수미	와	함께	떠나는	별마실	나들이	이수미	1,470,000	3,000,000원	13	2017.07.04																		
지역사회	영민이	영민이	[영주]	소백산	골	다문화	새마을의	'제2고향	만들기'	영주시다문화희망공공체	3,520,000	3,000,000원	50	2017.07.05																		
지역사회	영민이	영민이	[청송]	청송	의	방과	예술은	길다,	마을예술	영화관	동업회사법인	주식회사	바오바트	775,000	3,000,000원	17	2017.07.05															
지역사회	영민이	영민이	[군위]	농산물	공중	재배와	이웃	나눔	상생	군위읍	새마을지도자회/두녀자회	3,040,000	3,000,000원	50	2017.07.05																	
푸드	경북	영민이	[영덕]계	은	농부	의	수제	차	쌀	파이	영덕	인랑리	친환경	쌀	작목반	2,520,000	2,000,000원	72	2017.07.05													
교육/출판	경북	영민이	[영산]	치매	예방	다이어리,	나이	누릴	두꺼비	학 교	협동조합	2,670,000	1,000,000원	47	2017.07.05																	
푸드	경북	영민이	[문경]	회	양산	회	낙락	뽕튀기	회	양산	마을	1,860,000	5,500,000원	71	2017.07.05																	
문화예술	영민이	영민이	[울릉]	죽도	를	지키는	또	다른	방법!	꽃섬공방	292,000	3,000,000원	9	2017.07.05																		
지역사회	영민이	영민이	[영천]	휴과	불	의	흔적,	그	전통을	빛다,	협동조합	7,090,000	3,000,000원	119	2017.07.05																	
일자리	경북	영민이	[구미]	아빠!	희망	을	두드리는	목	공체	협	함께	가요	희망	찾는	마을	목공	소	협동조합	854,000	3,000,000원	29	2017.07.05										
문화예술	영민이	영민이	[창도]	창도	에서	온	국악	청년	의	이야기	온누리	국악	예술인	협동조합	7,930,000	3,000,000원	111	2017.07.05														
지역사회	영민이	영민이	[포항]	무연	문묘	발초	및	특가	노인	집수리	봉사	대명	장래	협동조합	7,940,000	3,000,000원	114	2017.07.05														
지역사회	영민이	영민이	[영주]	홍삼	이	머문	인건	스카프	풍기	고려	홍삼	협	동	조합	950,000	3,000,000원	22	2017.07.05														
지역사회	영민이	영민이	[임천]	호두와	자	산	꿀	어르신	의	따뜻한	동행	자	산	꿀	주	민	협	의	체	4,650,000	3,000,000원	125	2017.07.05									
지역사회	영민이	영민이	[인동]	할매!	우리	마	을	을	무	력	해요!	그	럼	애	문	화	마	을	협	의	회	4,485,000	3,000,000원	176	2017.07.05							
사회이슈	영민이	영민이	[포항]	폐지	수거	노인	리어	카	포항	시	사회	적	기	업	협	의	회	3,290,000	3,000,000원	20	2017.07.05											
지역사회	영민이	영민이	[안동]	청년	들과	떠나는	450년	전	과	거	여행!	AD	문	화	경	영	기	획	5,715,000	3,000,000원	280	2017.07.05										
문화예술	영민이	영민이	[울진]	울진	더	새	롭게	협	동	조합	울림이	1,330,000	1,000,000원	46	2017.07.05																	
지역사회	영민이	영민이	[의성]	목	화	의	꽃	말	은	어	머니	의	사	랑	입	니	다,	황	원	영	(금성	목	화	체	협	사	업	단)	1,250,000	3,000,000원	41	2017.07.05
일자리	경북	영민이	[봉화]	표	고	버섯	농사	짓는	청년	들	사	회	복	지	법	인	중	화	하	늘	6,880,000	3,000,000원	147	2017.07.05								
푸드	경북	영민이	[칠곡]	폴드	브루	한	잔	에	아	몬드	봉	봉	초	콜	렛	한	합	알	맹	이	카	페	1,720,000	1,500,000원	63	2017.07.05						
지역사회	영민이	영민이	[상주]	미녀	농부	와	할	매	의	문	화	협	스	터	원	표	영	농	소	합	법	인	600,000	3,000,000원	10	2017.07.05						
지역사회	영민이	영민이	[예천]	어	御!	홍	시,	방	타	이	고	시	장	간	아	이	스	홍	시	감	무	자	영	농	소	합	법	인	6,727,500	3,000,000원	100	2017.07.05
교육/출판	프로젝트	101	꿈	탈	을	프로	젝트	써	앗	캠	프	1,650,000	3,000,000원	59	2017.06.30																	
지역사회	영민이	영민이	[상주]	허니	포	레	스트	의	수	제	소	시	지,	햄,	베	이	컨	허니	포	레	스트	528,000	1,000,000원	12	2017.07.05							
사회이슈	영민이	영민이	K	히	어	로	즈	소	화	기	파	이	어	마	커	스	(주)	1,439,000	1,000,000원	19	2017.06.20											

GETTING DATA BY API USING TWITTER API

API의 개념과 원리

API?

Application Interface:

응용프로그램 (어플)에서 사용할 수 있도록 운영 체제나 프로그래밍 언어가 제공하는 기능을 제어할 수 있게 만든 인터페이스를 지칭한다.

쉽게 말해, API를 제공하는 프로그램의 기능을 다른 프로그램에서 사용을 가능하게 만드는 인터페이스

ex) 카카오톡으로 로그인 (카톡 로그인 플러그인 API)
포켓몬 고 (구글 지도 API)

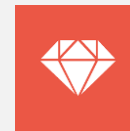
API의 기본

API key

API secret key

API token

API secret token



Log in

ACCESS TOKEN?

Access Token

“It is an object that describes the security context of a process or thread, usually used with password and username”

```
https://login.skype.com/COPY_ALL_THIS_URL?  
#access_token=CAAAAPJmB8ZBwBACHhatOID6U8BbotaR5rZAJzCigqyYCxBZC06v9yY1pMjzqm5RHgbZCQ0qpkMHByyX8tRUI8KaNU1c58HyZUyg5WoG0OgRkj4zhzWNZAST  
cM2yfXCznpCTPnWYhR16qH74mCx2wv5ZBdNjeyuhF4hXuA9sCukdO9j9av9pToM&expires_in=0
```



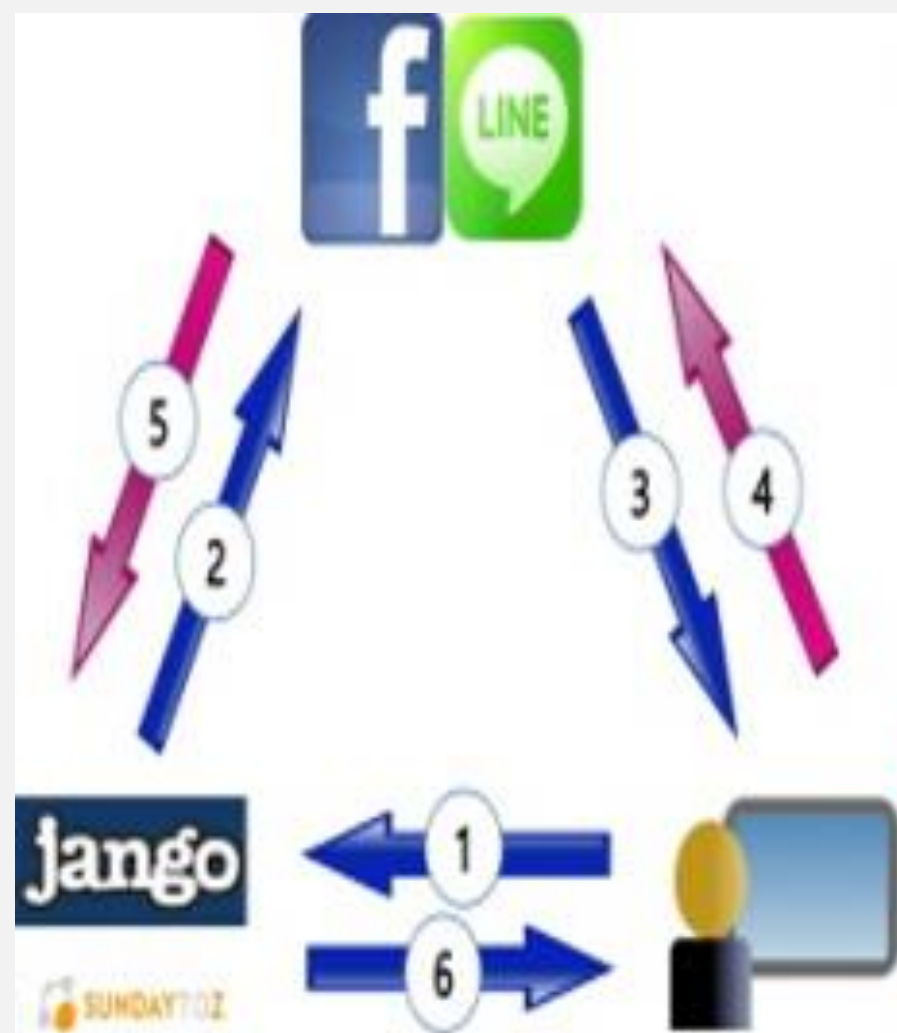
The Code in Red Color is Your Access token

Usage

Twitter, Google, Github, KakaoTalk, and Facebook API에서 사용되며, 플랫폼을 통합, 연결시켜주며 로그인과 유사한 기능을 제공한다.



KakaoTalk



1. 사용자가 웹사이트나 모바일 앱 서비스에 접근
2. 웹사이트나 앱 개발사에서 페이스북, 네이버 등에 로그인 정보를 담고 있는 액세스 토큰 요청
3. 사용자에게 계정과 패스워드를 요청
4. 사용자가 계정 및 비밀번호 입력
5. 각 개발사에 액세스 토큰 제공
6. 서비스 이용



개발가이드

iOS 레퍼런스

Android 개발가이드

Android 레퍼런스

JavaScript 개발가이드

JavaScript 레퍼런스

REST API 개발가이드

시작하기

사용자 가이드

REST API는 HTTP 요청을 보낼 수 있는 환경이라면 어디에서든 이용할 수 있습니다. 다음은 REST API를 사용할 수 있는 환경입니다.

- 모바일/PC 웹 환경에서 Javascript를 활용
- 다양한 환경(Java, Ruby, Python 등)의 웹 서버에서 활용
- iOS, Android 등 다양한 모바일 환경에서 활용

iOS, Android, Javascript의 경우 개발을 좀 더 쉽고 편리하게 할 수 있는 [Kakao SDK](#)를 제공합니다.

개발자 웹사이트에서는 REST API를 개발하고 디버깅 할 수 있는 다양한 툴을 제공하며, 본 문서에 대해 더 자세히 알아보실 수 있습니다.

curl이 설치되어 있지 않은 환경의 경우 [curl 다운로드](#)를 통해 설치가 가능합니다.

앱 생성



Search Tweets

Basics

Accounts and users

Tweets

Post, retrieve and engage with Tweets

Get Tweet timelines

Curate a collection of Tweets

Optimize Tweets with Cards

[Search Tweets](#)

Filter realtime Tweets

Sample realtime Tweets

Get batch historical Tweets

Tweet compliance

Tweet data dictionaries

Rules and filtering

Premium enrichments

Tweet updates

Direct Messages

[Overview](#)[Guides](#)[API Reference](#)

API Reference contents ^

[30-Day Search API](#)[GET saved_searches/show/:id](#)[Full-Archive Search API](#)[POST saved_searches/create](#)[GET search/tweets](#)[POST saved_searches/destroy/:id](#)[GET saved_searches/list](#)

GET search/tweets

Returns a collection of relevant [Tweets](#) matching a specified query.

Please note that Twitter's search service and, by extension, the Search API is not meant to be an exhaustive source of Tweets. Not all Tweets will be indexed or made available via the search interface.

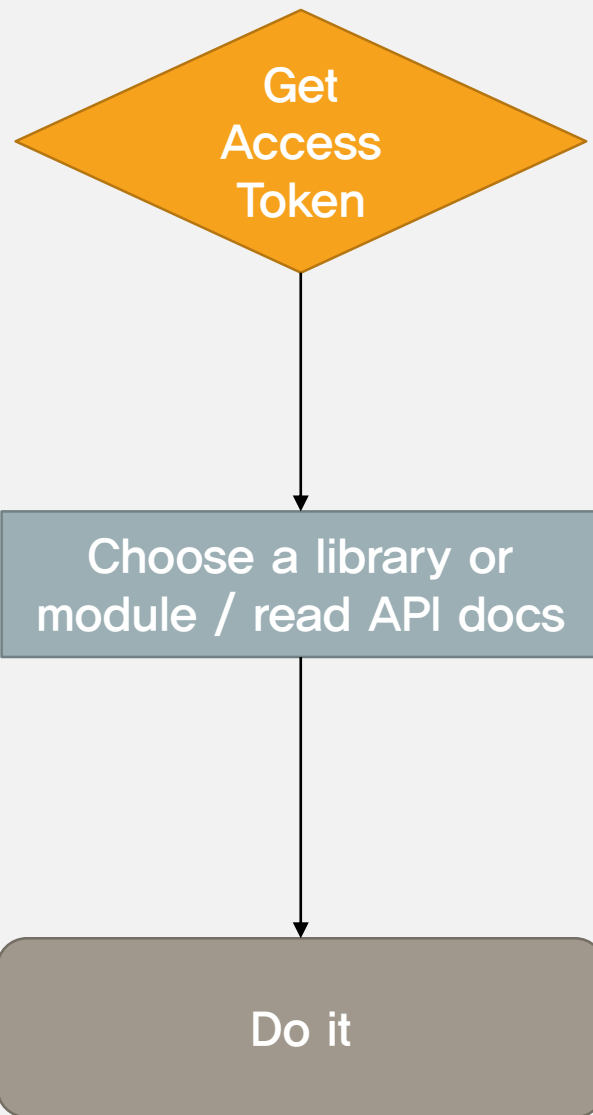
In API v1.1, the response format of the Search API has been improved to return [Tweet objects](#) more similar to the objects you'll find across the REST API and platform. However, perspectival attributes (fields that pertain to the perspective of the authenticating user) are not currently supported on this endpoint.


To learn how to use [Twitter Search](#) effectively, consult our guide on [How to build a query](#). See [Working with Timelines](#) to learn best practices for navigating results by `since_id` and `max_id`.

GETTING DATA BY API USING TWITTER API

Twitter API 사용

순서도



 Application Management

Twitter Apps

You don't currently have any Twitter Apps.

Create New App

Create an application

Application Details

Name *

issu_study

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens.

Description *

for crawling study

Your application description, which will be shown in user-facing authorization screens. Between 140 and 512 characters.

Website *

https://www.google.com

Your application's publicly accessible home page, where users can go to download, make use of, or learn more about your application. This URL will be used for attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify a callback URL.

issu_study

Details

Settings

Keys and Access Tokens

Permissions

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) nLuJVJ66BWkcSypddBR52ie10

Consumer Secret (API Secret) coetQC5qRdB2IzVMYIXb2A6IlUeUth3Qj9ohNX002TCYNOfy8

Access Level Read and write (modify app permissions)

Owner agdal1125

Owner ID 829564938377117697

Application Actions

Regenerate Consumer Key and Secret

Change App Permissions

Your Access Token

You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. The application's current permission level.

Token Actions

Create my access token


```
# -*- coding: utf-8 -*-
```

```
import tweepy
```

```
# API 인증 요청
```

```
consumer_key = ""
```

```
consumer_secret = ""
```

```
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
```

```
# Access 토큰 요청
```

```
access_token = ""
```

```
access_token_secret = ""
```

```
auth.set_access_token(access_token, access_token_secret)
```

```
# Twitter API 생성
```

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

```
# 검색 내용 설정하기
```

```
location = "%s,%s,%s" % ("35.95", "128.25", "1000km")
```

```
keyword = "쿠팡 OR 로켓페이"
```

```
# 검색한 내용 파일로 저장
```

```
wfile = open(os.getcwd() + "/twitter.txt", mode="w")
```

```
cursor = tweepy.Cursor(api.search,  
                        q = keyword,  
                        since='2015-01-01',  
                        count = 100,  
                        geocode = location,  
                        include_entities = True)  
  
for i, tweet in enumerate(cursor.items()):  
    print("{}:{}".format(i, tweet.text))  
    wfile.write(tweet.text + '\n')  
  
wfile.close()
```

PRINT 결과물

0:엘리스 생리대 사보았다. 쿠팡 로켓배송으로 샀는데 가격도 한국거랑 차이 없음 ㅋㅋㅋㅋㅋㅋ https://t.co/u5qpMO34O1

1:@T5vAp ㄹㅇ SMG하나있을때 쿠팡배달받는기분이자너

2:@madokim0412 쿠팡으로 장보는 중

3:쿠팡.위메프;하이진물병 https://t.co/GerLs72EYf

4:쿠팡에서 폰케이스 살려고 둘러본중에 개웃긴 후기

볼ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ https://t.co/P6wMvh0SLF

5:음 쿠팡맨에게 배송완료 문자가 왔는데 몇일전에 받아본 기저귀가 또있다 .. ?
뭐지 ... ?

6:쿠팡 시켰는데... 왜 내 택배 뜯집으로 갔어... https://t.co/ilMDIWmPln

7:나만의명품보틀

쿠팡.위메프검색;시크링보틀 https://t.co/Q7Kbj0OXIC

8:[브랜드평판] 오픈마켓 브랜드 10월 빅데이터 분석...1위 G마켓, 2위 티몬,
3위 쿠팡 : https://t.co/53tv4iU7q0

9:RT @purengom: 그런데 손도 못쓰는게 감당할 사람도 업체도 없음. 사가와
규빈은 이미 5년전에 때려쳤고 야마토도 때려쳤으니... 오죽하면 쿠팡마냥 직접
배달하는 것도 검토들어갔다는 기사가 나오고.

10:쿠팡에서 딱 이만원어치 주문 하는게 미안하긴 하지만, 이만원에서 모자라는
그 몇천원 더 채워넣는 것도 자꾸 들었다놔다 하게 되는 지갑 상태, 마음 상태

11:쿠팡 로켓배송 시킬게있어서 최소금액 맞추려고 보고있는 중인데

취미품목에 있는거보고 눈을 의심했다...ㄹㅇ https://t.co/sGB2e1W6Uc

12:쿠팡맨왔다!

13:쿠팡맨 사랑해요 한진택배 시발

14:아악 ,,쿠팡맨 아저시.. 빨리와요...

15:[IT 이직] 쿠팡 회사 어떤가요??

https://t.co/w5BrytGShX

모두의 뉴스앱 https://t.co/nTa5kYEz4C

16:@rosetw48 쿠팡맨은 이렇게 좋은걸 준적이 없는데ㅍㅍㅍㅍ 감사합니다

GETTING DATA BY API USING TWITTER API

Word Cloud 만들기

데이터 전처리

```
# 데이터 클렌징 (Data Cleansing)

import preprocessor as p          # tweet-preprocessor 라이브러리
from collections import Counter
import os

f = open("/home/jaekeun/twitter.txt", "r", encoding="utf-8") # 파일 열기

g = open(os.getcwd() + "/twitter_clean.txt", mode="w", encoding="utf-8")

for line in f:                    #파일 안에 있는 각 줄마다 같은 명령 반복
    #print(line)
    p.set_options(p.OPT.URL, p.OPT.EMOJI, p.OPT.MENTION)
    one_twt = p.clean(line)
    #print(one_twt)
    one_twt.replace("#", "")
    one_twt.replace("...", "")
    one_twt.replace("@", "")
    one_twt.replace(";", "")
    one_twt.replace(".", "")

    g.write(one_twt + '\n')

g.close()
```

Option Name	Option Short Code
URL	p.OPT.URL
Mention	p.OPT.MENTION
Hashtag	p.OPT.HASHTAG
Reserved Words	p.OPT.RESERVED
Emoji	p.OPT.EMOJI
Smiley	p.OPT.SMILEY
Number	p.OPT.NUMBER

WORDCLOUD 생성하기

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

df = open("/home/jaekeun/twitter_clean.txt", "r", encoding="utf-8")

wordcloud = WordCloud(
    background_color='black',
    width=1200,
    height=1000
).generate(df.read())

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```