



1조 발표

# [Chicken Store Data]

I love chicken, you love chicken, We love chicken.

## 수집

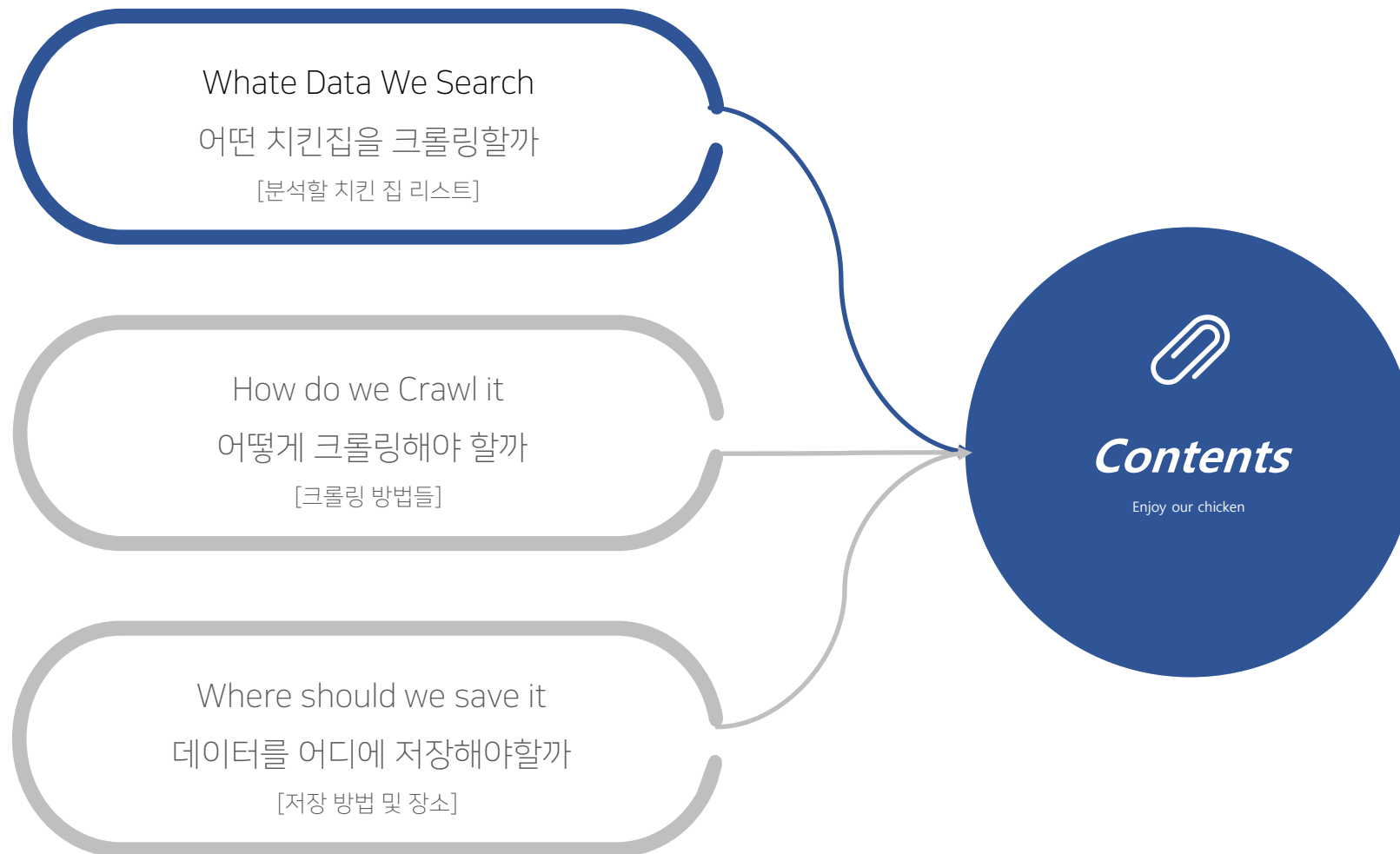
---



## 01

### CONTENTS

I love chicken, you love chicken, We love chicken

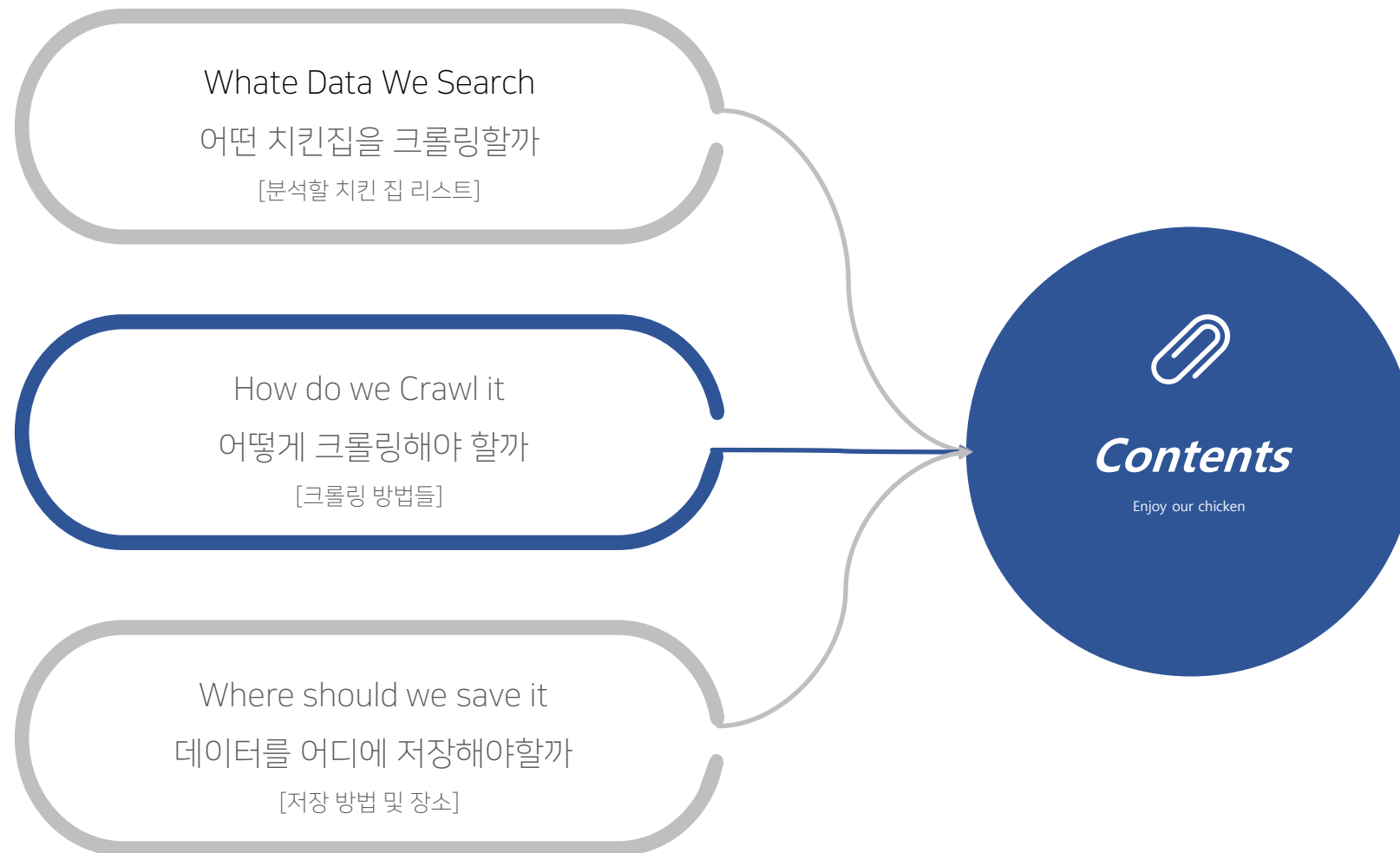




## 01

### CONTENTS

I love chicken, you love chicken, We love chicken

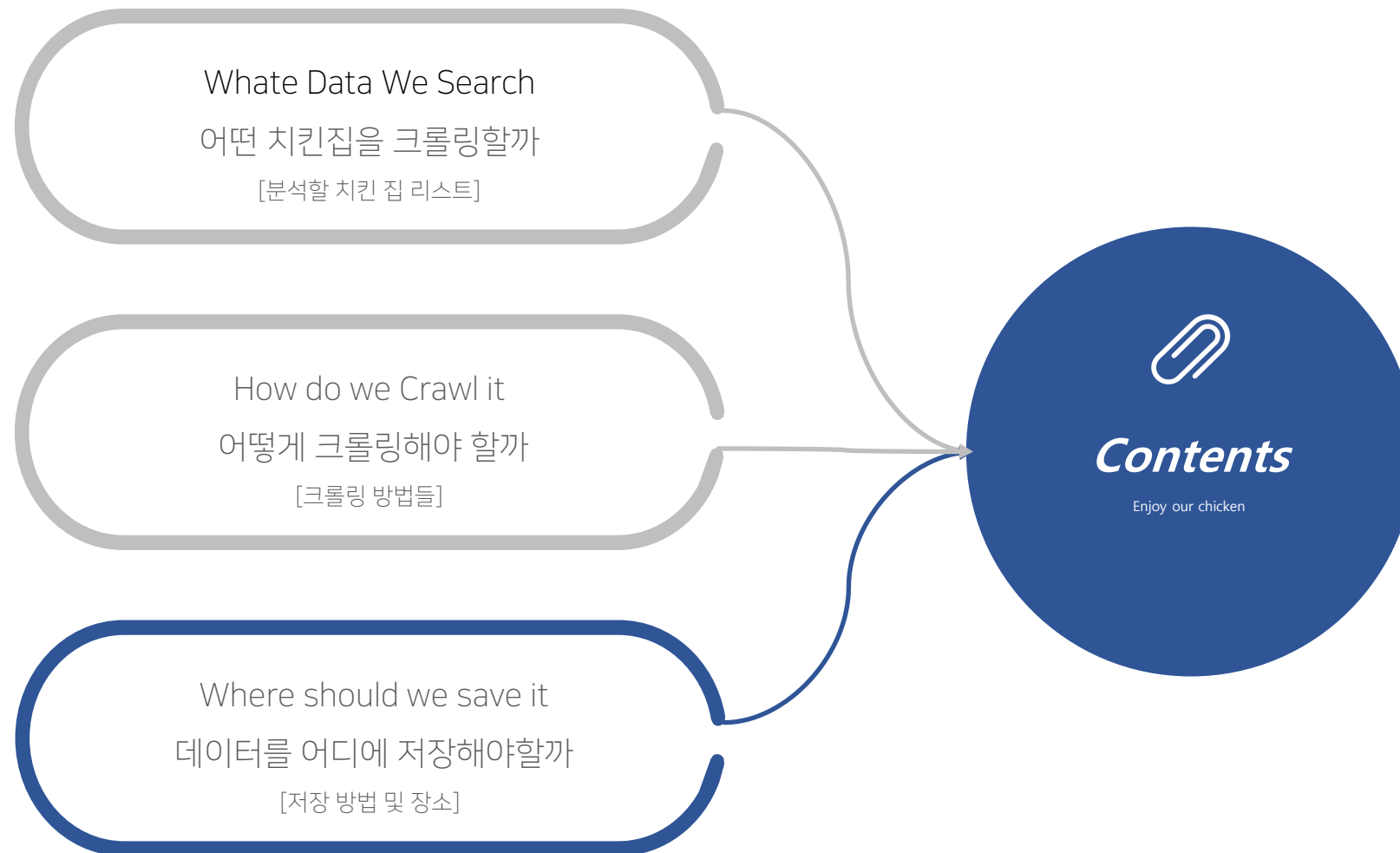




## 01

### CONTENTS

I love chicken, you love chicken, We love chicken





## 02

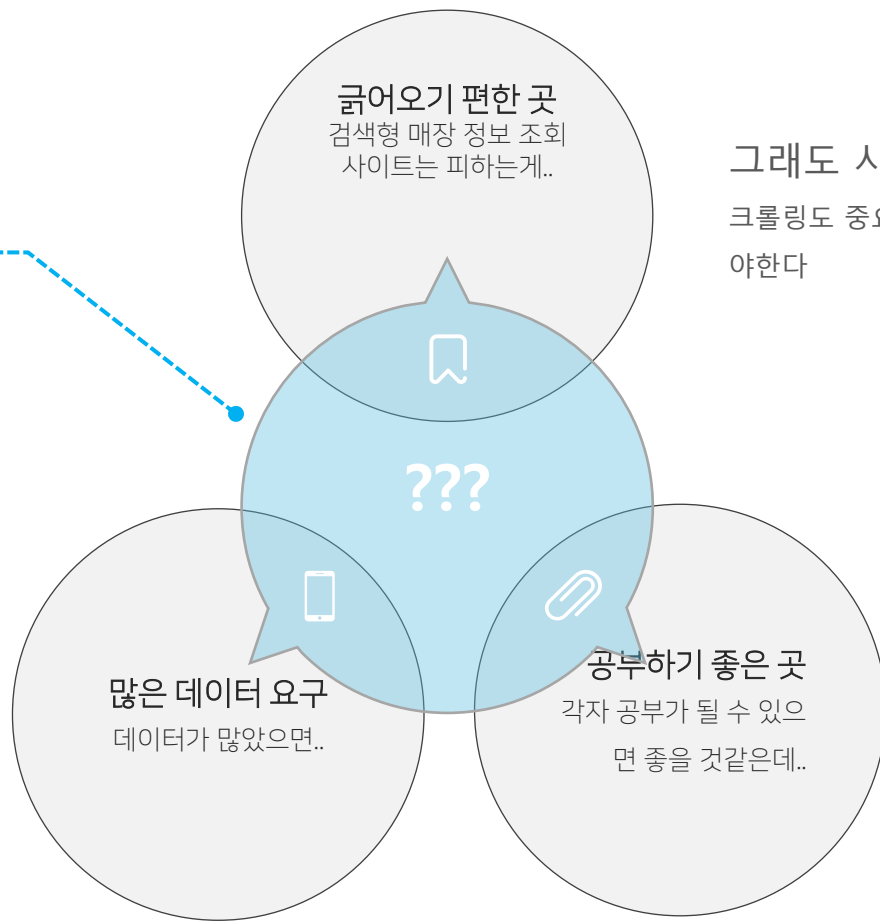
What Data We Search

I love chicken, you love chicken, We love chicken

어디를 해야할까..

좋아하는 치킨집이 다 다르다  
후보지도 많고 모두가 재미있게 할 수 있는 방  
법이 없을까...

컴퓨터만이 할 수 있다  
데이터 양이 적으면 굳이 컴퓨터에게 말  
길 필요가 없지 않을까.. 최대한 많이 긁자



그래도 시간 너무 쓰지 않도록  
크롤링도 중요하지만 역량에 맞는 곳을 찾아  
야한다

결국은 실력만이 남는다  
모두가 가볍게 넘어가지 않고 모두의 역  
량을 증진시킬 수 있도록



## 02

### What Data We Search

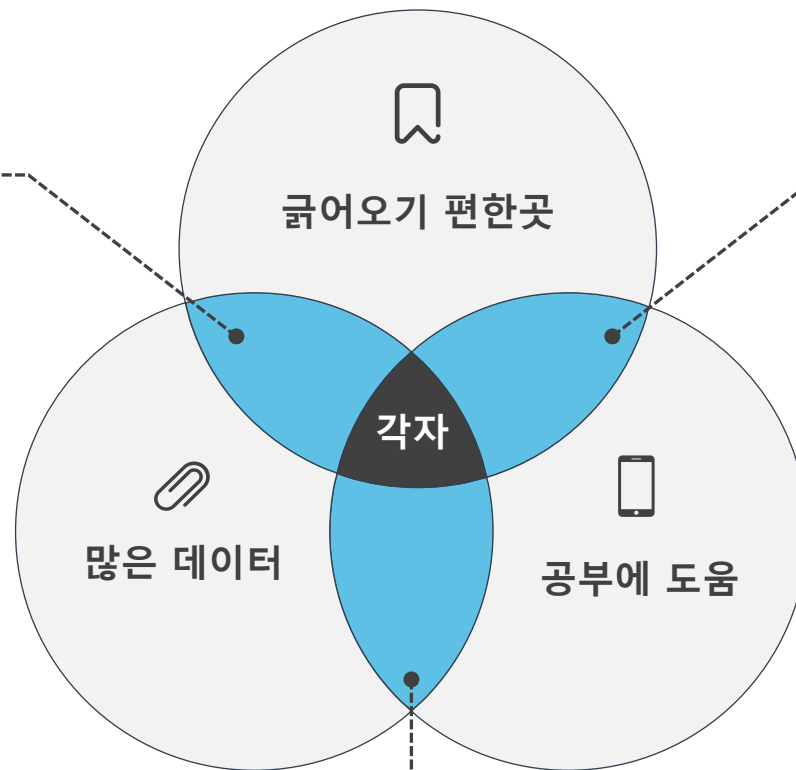
I love chicken, you love chicken, We love chicken

#### 사이트 조사

각자 사이트를 조사해서 적합한  
곳을 선별하자

#### 크롤링 한번씩은...

적어도 한번씩은 크롤링 해보고  
자기가 만든 데이터를 가공까지 해봐야 과정  
을 이해하기 쉬울 것이다



#### 많은 데이터, 많은 예외

데이터가 많아질수록 공부하기 좋  
을 것이다

데이터가 많을 수록 예외처리에 신경써야한다



## 02

## What Data We Search

I love chicken, you love chicken, We love chicken

## 사이트 조사

각자 사이트를 조사해서 적합한  
곳을 선별하자

클리어하기 편함

## 크롤링 한번씩은...

적어도 한번씩은 크롤링 해보고  
저장된 데이터를 가공까지 해봐야 과정  
을 이해하기 쉬울 것이다

각자

많은 데이터

공부에 도움

많은 데이터, 많은 예외

데이터가 많아질수록 공부하기 좋  
을 것이다

데이터가 많을 수록 예외처리에 신경써야한다





# Chicken Store Crawling

I love chicken, you love chicken, We love chicken.

## 페리카나 크롤링 적합성 판단 1

-링크-

[https://pelicana.co.kr/store/stroe\\_search.html?page=1](https://pelicana.co.kr/store/stroe_search.html?page=1)[https://pelicana.co.kr/store/stroe\\_search.html?page=1](https://pelicana.co.kr/store/stroe_search.html?page=1)

get 방식 페이지처리, 아주 좋다.

매장리스트

매장명	주소	전화번호	상세정보
가평역전점	경기도 가평군 가평읍 가화로 54	031-582-5242	<a href="#">상세정보</a>
가평점	경기도 가평군 가평읍 가화로 125-1	031-582-4419	<a href="#">상세정보</a>
가을점	경상북도 영주시 창진로 59	054-634-2500	<a href="#">상세정보</a>
각화주공점	광주광역시 북구 군왕로207번길 16 107호	062-264-8529	<a href="#">상세정보</a>
간성점	강원도 고성군 간성읍 간성로 45	033-681-3330	<a href="#">상세정보</a>
간현점	강원도 원주시 지정면 간현로 188	033-731-0887	<a href="#">상세정보</a>
갈곶원동점	경기도 오산시 오산로 68(갈곶동)	031-372-3573 031-372-3574	<a href="#">상세정보</a>
갈마점	대전광역시 서구 갈마로 60(갈마동)	042-531-9962	<a href="#">상세정보</a>
갈산점	인천광역시 부평구 굴포로 42	032-522-9292	<a href="#">상세정보</a>
갈산점(이천)	경기도 이천시 갈산로 81(갈산동, 102)	031-638-8320	<a href="#">상세정보</a>

« < 1 2 3 4 5 6 7 8 9 10 > »

## 페리카나 크롤링 적합성 판단 2

-태그 구성이 깔끔한가-

```

<tbody>
  <tr>
    <td class="t_center">가평역전점</td> == $0
    <td>경기도 가평군 가평읍 가화로 54</td>
    <td class="t_center"> 031-582-5242</td>
    <td class="t_center">...</td>
  </tr>
  <tr>...</tr>
  <tr>...</tr>
  <tr>...</tr>
  <tr>...</tr>

```

깔끔한 테이블 역시 좋다  
페리카나 너로 정했다.





## 03

How Do we Crawl it

I love chicken, you love chicken, We love chicken

나중에 원활하게 데이터를 사용하기 위해서..

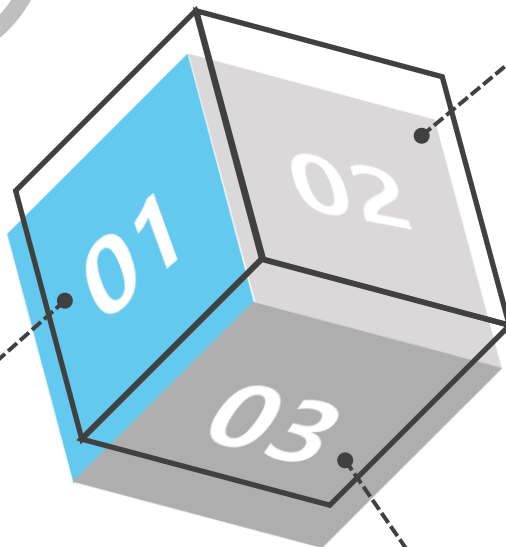
크롤링할 데이터의 속성을 다같이 약속하였다.



### 01. 가맹점 이름

매장 정보 중 가맹점이름

변수명 : name



### 02. 가맹점 위치(주소)

가맹점의 위치 정보

변수명 : address



### 03. 가맹점 전화번호

가맹점의 전화 번호

변수명 : tel



# How Do we Crawl it

I love chicken, you love chicken, We love chicken.

```
import multiprocessing
import parmap
import numpy as np
import pericana
import requests
from bs4 import BeautifulSoup
import time
import urllib.request
from selenium.webdriver import Chrome
import re
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.common.keys import Keys
import datetime as dt
import os
```

```
num_cores = multiprocessing.cpu_count() # 12
link_list = pericana.page_make(109)

splited_data = np.array_split(link_list, num_cores)
splited_data = [x.tolist() for x in splited_data]

result = parmap.map(pericana.store_list, splited_data, pm_pbar=True, pm_processes=num_cores)
```

[illegible]

```
import requests
from bs4 import BeautifulSoup
import time
import urllib.request
from selenium.webdriver import Chrome
import re
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.common.keys import Keys
import datetime as dt
import pandas as pd

def page_make(max_num):
    dial = []
    for i in range(1,max_num+1):
        addr = 'https://pelicana.co.kr/store/stroe_search.html?page='+str(i)
        dial.append(addr)
    return dial

def store_list(lists):
    store_pd2 = []
    path = './webdriver/chromedriver'
    delay = 2
    driver= Chrome(path)
    driver.implicitly_wait(delay)
    for l in lists:
        link = l
        driver.get(link)
        tr_list = driver.find_elements_by_xpath('//*[@id="contents"]/table/tbody/tr')
        for i in range(1,len(tr_list)+1):
            name = driver.find_element_by_xpath('//*[@id="contents"]/table/tbody/tr['+str(i)+']/td[1]').text
            addr = driver.find_element_by_xpath('//*[@id="contents"]/table/tbody/tr['+str(i)+']/td[2]').text
            if addr == "":
                break
            tel = driver.find_element_by_xpath('//*[@id="contents"]/table/tbody/tr['+str(i)+']/td[3]').text
            tel = tel.split('☎')
            tel = tel[0]
            store_pd2.append([name,addr,tel])
    return store_pd2
```



## 3. Pandas Dataframe 변환 및 csv 저장

```

import pandas as pd

last = []
for re in result:
    for r in re:
        last.append(r)
list_df = pd.DataFrame(last)
list_df.to_csv("./pericana_data/data.csv", header=True, index=False, encoding='utf-8-sig')

```

	A	B	C
1	0		1 2
2	가경동점	충청북도 청주시흥덕구 풍산로 103(북대동)	043-233-4091
3	가남점	경기도 여주시 가남면 태평중앙1길 8-6	031-883-5746
4	가래비점	경기도 양주시 광적면 가래비길 24	031-855-6126
5	가수원점	대전광역시 서구 별곡로 1353번길 63	042-541-1350
6	가양동점	서울특별시 강서구 양천로57길 10-11, 101-1호 (가양동)	02-3663-3700
7	가오점	대전광역시 동구 온여송로 77번길 17	042-284-0391
8	가운점	경기도 남양주시 가운로2길 63 가운동삼원프라자	105 031-557-9222
9	가장점	대전광역시 서구 갈마로 298	042-532-1140
10	가정지구1	인천광역시 서구 봉오재2로 37 104호	032-567-5885
11	가좌3점	인천광역시 서구 가좌로156번길 23-1	032-575-8999
12	가평역전점	경기도 가평군 가평읍 가화로 54	031-582-5242
13	가평점	경기도 가평군 가평읍 가화로 125-1	031-582-4419
14	가흥점	경상북도 영주시 창진로 59	054-634-2500
15	각화주공점	광주광역시 북구 군왕로207번길 16 107호	062-264-8529
16	간성점	강원도 고성군 간성읍 간성로 45	033-681-3330
17	간현점	강원도 원주시 지정면 간현로 188	033-731-0887
18	갈곶원동점	경기도 오산시 오산로 68(갈곶동)	031-372-3573
19	갈마점	대전광역시 서구 갈마로 60(갈마동)	042-531-9962
20	갈산점	인천광역시 부평구 굴포로 42	032-522-9292