

# 많이 찾는 요리 레시피

Web Crawling

Team : Build Up

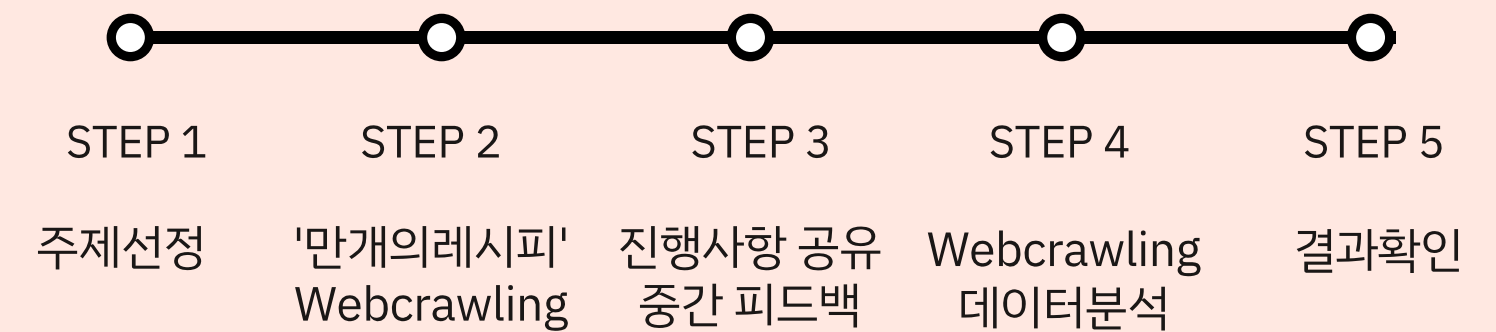
박경빈, 방울, 최유림, 한설아





# 사람들이 많이 찾는 레시피

인간생활의 세 가지 기본 요소라고 불리는 의식주에서도 볼 수 있듯  
우리의 삶에 있어서 빼놓을 수 없는 식(食)을 주제로 하여  
사람들이 많이 찾는 레시피를 레시피 사이트에서 Crawling 하여  
데이터를 분석해보고자 함.





# 레시피 사이트 정보수집 (1)

## 1. 요리 레시피 사이트 웹 크롤링을 통한 1차 분석

- ✓ '만개의레시피' 사이트 웹 크롤링
- ✓ 주로 사용되는 '식재료' 도출
- ✓ 그래프와 워드클라우드로 시각화



```
import urllib.request
from bs4 import BeautifulSoup as bs
import re
import json, csv

def toJson(recipe_dict):
    with open('recipe.json', 'w', encoding='utf-8') as file :
        json.dump(recipe_dict, file, ensure_ascii=False, indent='    ')

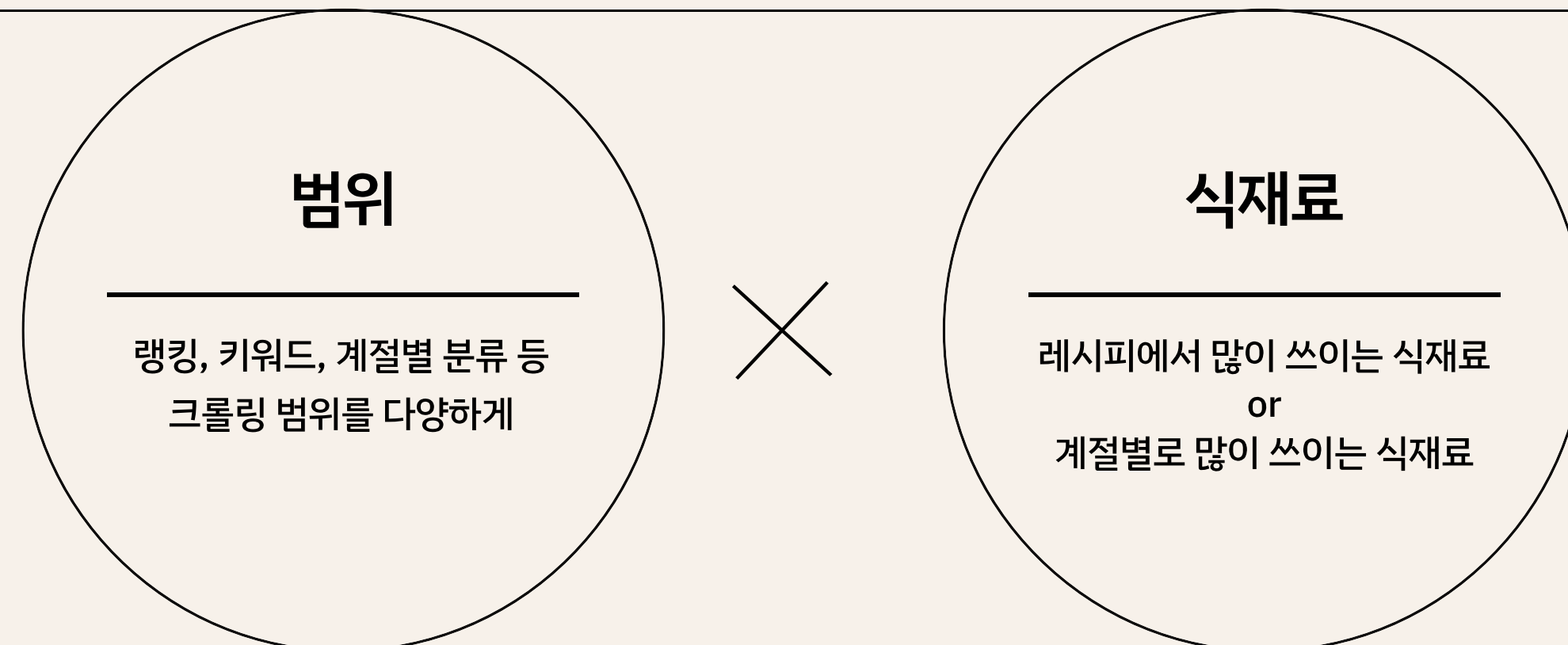
def toCSV(recipe_list):
    with open('ingredients.csv', 'w', encoding='utf-8', newline='') as file:
        csvfile = csv.writer(file)
        for row in recipe_list:
            csvfile.writerow(row)

def url_func(n, m):
    num_range = range(n, m)
    url = "https://www.10000recipe.com/profile/recipe.html?uid=gdubu33&page="
    url_list = []
    for num in num_range:
        req = urllib.request.Request(url + str(num))
        code = urllib.request.urlopen(url).read()
        soup = bs(code, "html.parser")
        try:
            res = soup.find(class_ = 'cont_list st3')
            for i in res.find_all('a'):
                url_tmp = i.get('href')
                url_list.append(url_tmp)
        except (AttributeError):
            pass
    return url_list
```

# 레시피 사이트 정보수집 (1)

## 2. 크롤링 1차 분석 결과

- ✔ 특정 식재료보다는 육수나 소스에 기본 재료로 많이 사용되는 '마늘', '다시마', '간장'과 같은 결과값이 크게 나왔다.
- ✔ 추출값에서 불용어가 많아 제거되는 값이 많았다.
- ✔ 해당하는 크롤링 결과값으로는 유의미한 결과를 얻기가 어려웠다.
- ✔ 페이지의 범위나 키워드에 따른 다른 목록값을 받아 결과를 도출해보자.
- ✔ 넓은 범위의 데이터를 수집 후 '다시마', '마늘'과 같은 양념 재료 외에 메인식재료에 대한 트렌드를 알아보자.



## 레시피 사이트 정보수집(2)

### 3. 요리 레시피 사이트 웹 크롤링을 통한 2차 분석

- ✓ 개별 질문과 피드백을 통한 코드 수정
- ✓ 랭킹(일간 TOP100)의 요리 리스트의 '요리명' - '요리자' - '조회수' - 'URL' 추출
- ✓ 해당 레시피의 '등록일'까지는 피드백을 받아 추출하였으나  
등록일 기준으로 (3~5월, 6~8월, 9~11월, 12~2월) 계절별 분석을 하는 것에는 실패함.

```
#.등록일
dates = []
for one in all_url:
    print(one)
    url = one
    page = urlopen(url)
    soup = BeautifulSoup(page, 'html.parser')
    print(soup.title)
    print(soup.find("p", class_='view_notice_date').find('b').text)
```

<https://www.10000recipe.com/recipe/5832651>  
<title>[참치 갈뎡장] 밥도둑 참치 갈뎡장으로 밥한공기 뚫얏</title>  
등록일 : 2015-08-24

<https://www.10000recipe.com/recipe/6946882>  
<title>스팸순두부찌개 톡하고 완전 밥도둑!!</title>  
등록일 : 2020-07-17

<https://www.10000recipe.com/recipe/6902902>  
<title>진정한 밥도둑-갈릭은 배추김치로 등갈비김치찜 만드느법~~~</title>  
등록일 : 2018-12-26

[illegible]



# 프로젝트를 통해 배운 점

## 프로젝트를 통해 배운 점

- 웹 크롤링과 데이터 수집, 분석 등을 위해서는 Base(기초학습)이 중요함.
- 직접 코드를 입력, 수정해보면서 BeautifulSoup이나 urllib, pandas 등 적용이 익숙해짐.
- 소스코드를 활용(수정)하거나 새로 구상하여 결과 도출하는 것이 어려웠음.
- 데이터수집이 실제로 다양하게 활용되고 있으며, 데이터에 대한 효율성 극대화함.

## 더 분석하게 된다면

- (미완성한 주제) 계절별로 많이 쓰이는 식재료에 대한 데이터 분석
- 우크라이나침공, 상하이봉쇄 등 경제 상황으로 인한 식량 가격 상승의 상관관계 분석
- 우리나라 메인 식재료로 육류가 곡물 소비를 추월했다는 결과에 대한 식재료 소비량 분석



Team Build Up

# Thank you!

감사합니다.

Team **Build Up**

박경빈, 방울, 최유림, 한설아

