

Crawling project

---

# 코로나 알림 서비스 제공

---

이경무 4조  
이지홍

# INDEX

---

**01**

**개요**

**02**

**수집 데이터**

**03**

**진행 사항**

**04**

**회고**

---

COVID-19  
GETTYIMAGES ILLUSTRATION



COVID-19  
GETTYIMAGES ILLUSTRATION



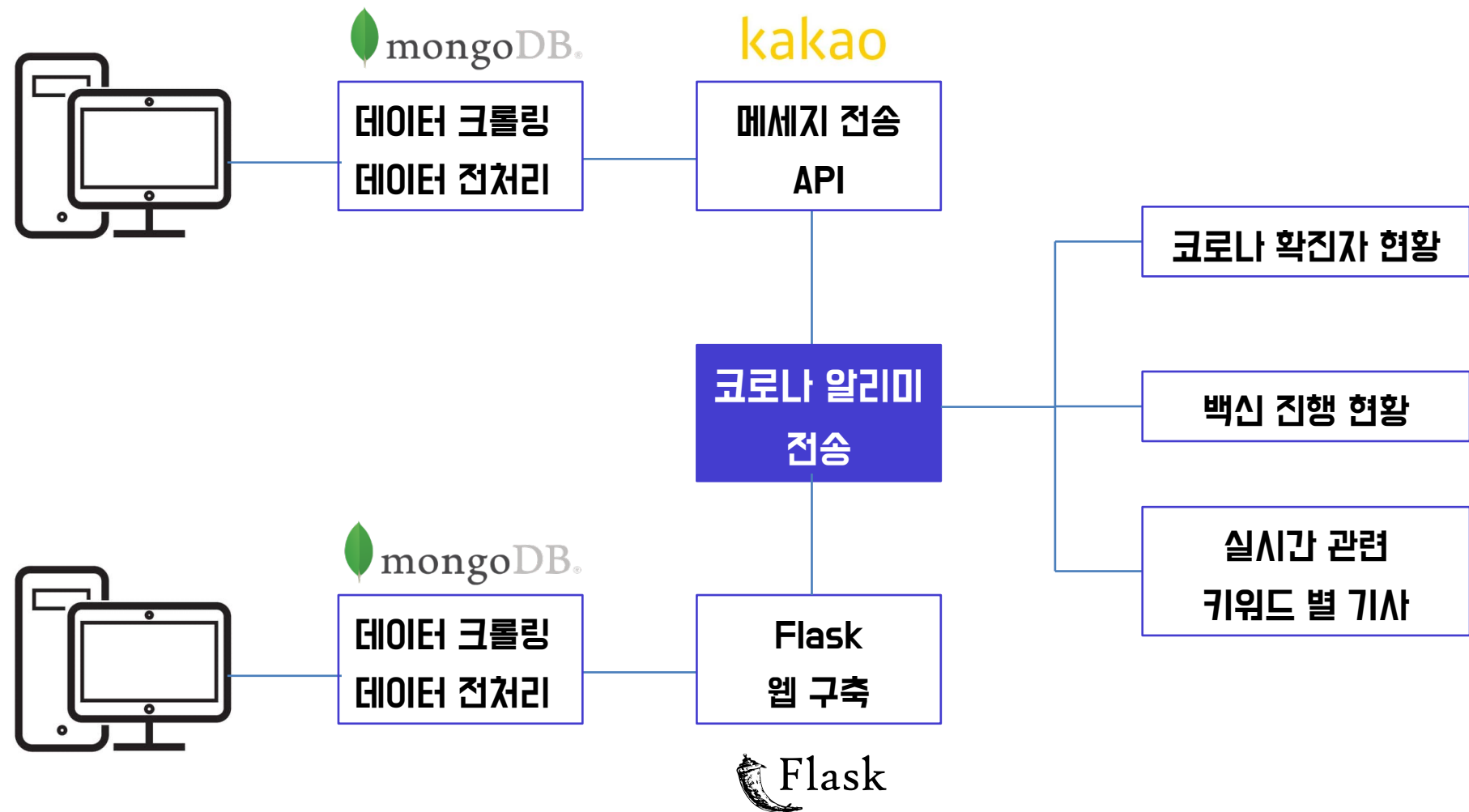
이미지 출처 : 게티이미지뱅크

- 코로나 시국의 장기화 인한 둔감을 극복하고 다시 경각심을 가지자는 취지
- 매일 변동되는 코로나 상황을 매일 아침 카카오톡 메시지를 이용하여 전달
- 당일 코로나 확진자 , 거리 두기 현황, 전일 백신 접종 현황, 관련 키워드 별 기사 정보 제공



kakao

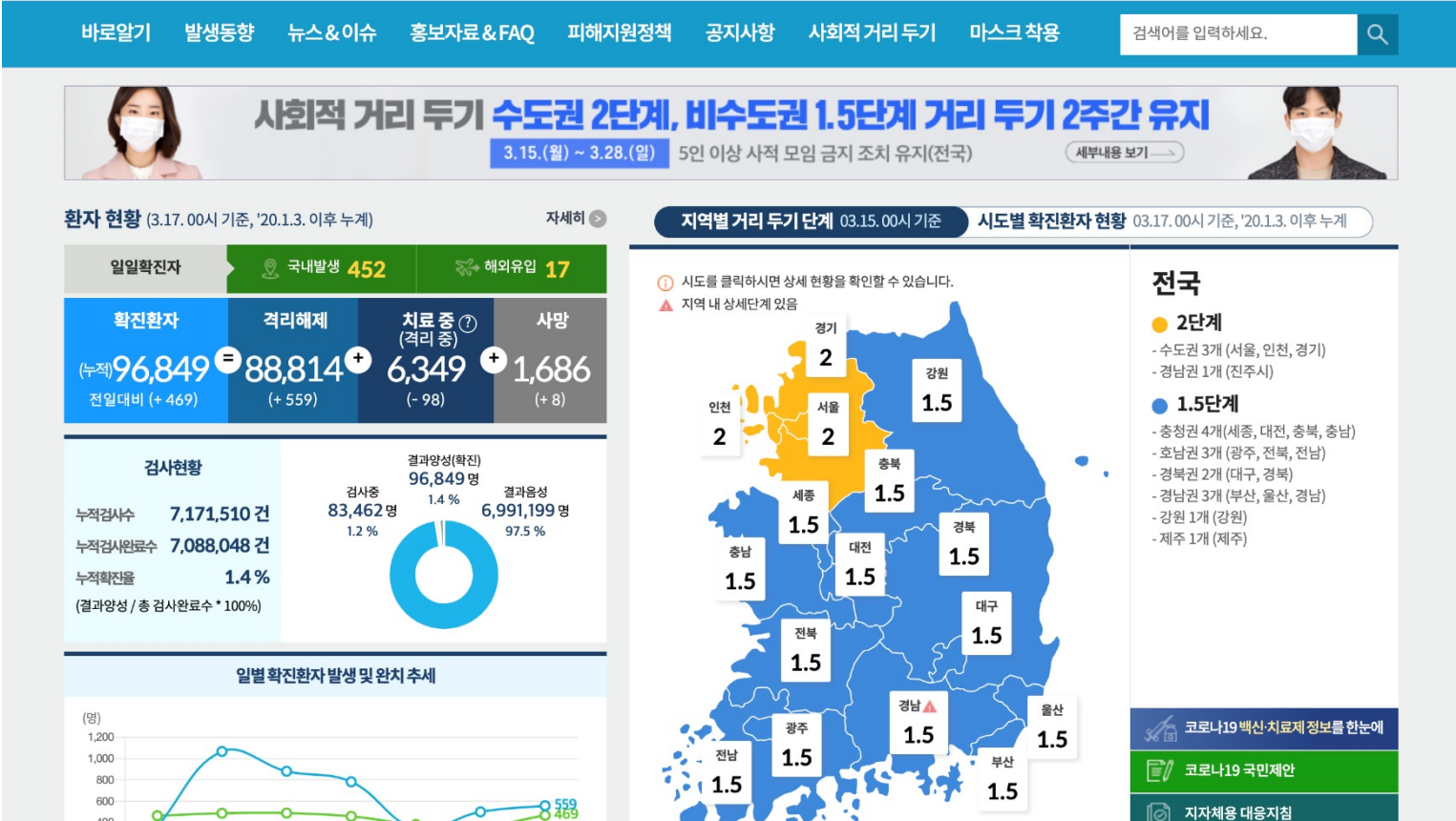




1. 코로나 바이러스 감염증 : <http://ncov.mohw.go.kr/>

코로나바이러스감염증-19(COVID-19)

코로나19 백신 및 예방접종 중앙사고수습본부 중앙방역대책본부 English 中文




2. 코로나19 예방접종 : <http://ncv.kdca.go.kr/>

코로나19 예방접종

코로나바이러스감염증-19 □ 중앙사고수습본부 □ 중앙방역대책본부 □ f y g t


바로알기 예방접종 현황 뉴스&팩트체크 허위조작정보신고 홍보자료 자주묻는 질문 알림·서식



**사회적 거리 두기 수도권 2단계, 비수도권 1.5단계 거리 두기 2주간 유지**

3.15.(월) ~ 3.28.(일) 5인 이상 사적 모임 금지 조치 유지(전국)

[세부내용 보기](#)



예방접종실적 총 현황 (3.17., 0시 기준) 자세히 보기

회차별 접종 현황

| 1회차 당일 누적 | 1회차 전일 실적 | 1회차 전일 누적  |
|-----------|-----------|------------|
| 621,734명  | = 18,283명 | + 603,451명 |

| 2회차 당일 누적 | 2회차 전일 실적 | 2회차 전일 누적 |
|-----------|-----------|-----------|
| 0명        | = 0명      | + 0명      |

예방접종실적은 접종기록 등록 지연 등에 따라 일부 변경될 수 있습니다.

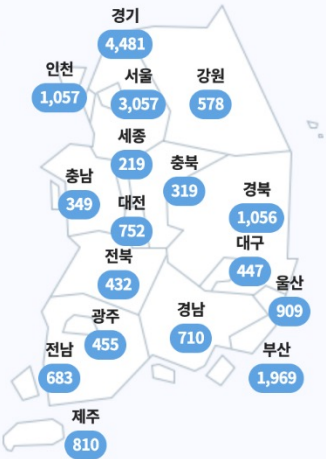
시도별 접종 현황

1차 접종 2차 접종

**1차 접종**

- 실적: 18,283명

-누계: 621,734명



### 3. 코로나 관련 키워드

**N | 코로나 백신**

통합 뉴스 지식IN 이미지 VIEW 동영상 어학사전 지식백과 쇼핑 ... 검색옵션 ^

정렬 1일 영역 유형 언론사 기자명 옵션유지 상세검색

• 관련도순 • 최신순 • 오래된순 검색결과 자동고침

**N** 원하는 기사를 빨리찾는 TIP! 뉴스검색 가이드

PICK 해당 언론사가 주요기사로 직접 선정한 기사입니다.

**연합뉴스 PICK** | 9시간 전 | 네이버뉴스  
 EMA "AZ백신 이익, 위험성보다 커...안전하고 효과적"(종합3보)  
 잇따라 예방 차원에서 특정 제조 단위 혹은 전체 아스트라제네카 백신에 대해 접  
 중을 일시 중단했다. 세계보건기구(WHO)도 17일 아스트라제네카가 코로나19 백...

유럽의약품청 "아스트라제네카 백신 ... 국민일보 PICK | 9시간 전 | 네이버뉴스  
 유럽의약품청 "AZ백신 혈전 위험 중... 강원일보 PICK | 5시간 전 | 네이버뉴스  
 EMA "AZ백신 이익, 위험성보다 커..... 연합뉴스 PICK | 10시간 전 | 네이버뉴스  
 관련뉴스 4건 전체보기 >

**한국경제 PICK** | 4시간 전 | 네이버뉴스  
 "AZ 백신, 안전하고 혈전 무관"...국내 접종에도 영향 끼칠까  
 유럽의약품청(EMA)이 아스트라제네카(AZ)사의 신종 코로나바이러스 감염증(코로  
 나19) 백신 접종과 혈전 생성 간에 관련성이 없고, 안전성이 확보됐다는 결론을 ...

유럽의약품청 'AZ백신-혈전 무관' 결... 연합뉴스 PICK | 5시간 전 | 네이버뉴스  
 유럽의약품청 "AZ백신-혈전 무관"... 헬스조선 PICK | 3시간 전 | 네이버뉴스  
 "AZ백신, 혈전 생길 수 있지만 맞... 파이낸셜뉴스 PICK | 5시간 전 | 네이버뉴스  
 EMA "백신혈전 무관" 결론... 정부 "... 세계일보 PICK | 2시간 전 | 네이버뉴스  
 관련뉴스 7건 전체보기 >

**서울경제** | 1시간 전 | 네이버뉴스  
 정부 "백신 접종 후 발열 등 가벼운 증상 발생시 응급실 방문 자...  
 한편 정부는 코로나19 백신 접종 후 '백신 휴가'를 주는 것과 관련해 본격화 하기  
 위해 10일 보건복지부 누리집을 통해 신묘한이력 연다 등 바자어 "백신 휴가 제

백신 :

<https://search.naver.com/search.naver?where=news&query=코로나%20국내%20백신&pd=4>

거리두기 :

<https://search.naver.com/search.naver?where=news&query=코로나%20사회적거리두기&pd=4>

확진 :

<https://search.naver.com/search.naver?where=news&query=코로나%20수도권%20확진&pd=4>

재난지원금 :

<https://search.naver.com/search.naver?where=news&query=코로나%20정부%20재난지원금&pd=4>



### items.py 작성

```
%writefile covid_vaccine/covid_vaccine/items.py
import scrapy

class CovidVaccineItem(scrapy.Item):
    title = scrapy.Field()
    link = scrapy.Field()
    date = scrapy.Field()

Overwriting covid_vaccine/covid_vaccine/items.py
```

### spider.py 작성

```
%writefile covid_vaccine/covid_vaccine/spiders/spider.py

import scrapy
from datetime import datetime
from covid_vaccine.items import CovidVaccineItem

class CovidSpider(scrapy.Spider):
    name = 'CovidVaccine'
    allow_domain = ['https://www.naver.com/']

    def start_requests(self):
        yield scrapy.Request("https://search.naver.com/search.naver?where=news&query=코로나%20국내%20백신&pd=4", callback=self.parse_keyword)

    def parse_keyword(self, response):
        item = CovidVaccineItem()
        now = datetime.now()
        #item['date'] = "%s년 %s월 %s일 %s시 %s분" % (now.year, now.month, now.day, now.hour, now.minute)
        item['date'] = "%s.%s.%s" % (now.year, now.month, now.day, now.hour, now.minute)
        vaccine_title = response.xpath('//*[@id="main_pack"]/section/div/div[3]/ul/li/div[1]/div/a/@title').extract()
        vaccine_link = response.xpath('//*[@id="main_pack"]/section/div/div[3]/ul/li/div[1]/div/a/@href').extract()
        for i in range(len(vaccine_title)):
            item['title'] = vaccine_title[i]
            item['link'] = vaccine_link[i]
        yield item

Overwriting covid_vaccine/covid_vaccine/spiders/spider.py
```

### mongodb.py 작성

```
%writefile covid_vaccine/covid_vaccine/mongodb.py
import pymongo

client = pymongo.MongoClient("mongodb://15.164.101.134:27017")
db = client.keyword
collection = db.keyword_vaccine

Overwriting covid_vaccine/covid_vaccine/mongodb.py
```

### pipelines.py 작성

```
%writefile covid_vaccine/covid_vaccine/pipelines.py
from itemadapter import ItemAdapter
from pymongo import collection

class CovidVaccinePipeline:
    def process_item(self, item, spider):
        data = {'title': item['title'], 'link': item['link'], 'date': item['date']}
        collection.insert(data)
        return item

Overwriting covid_vaccine/covid_vaccine/pipelines.py
```

```
class Spider(scrapy.Spider):
    name = "covid_19_total_backup"
    start_urls = ["http://ncov.mohw.go.kr/", "http://ncv.kdca.go.kr/"]

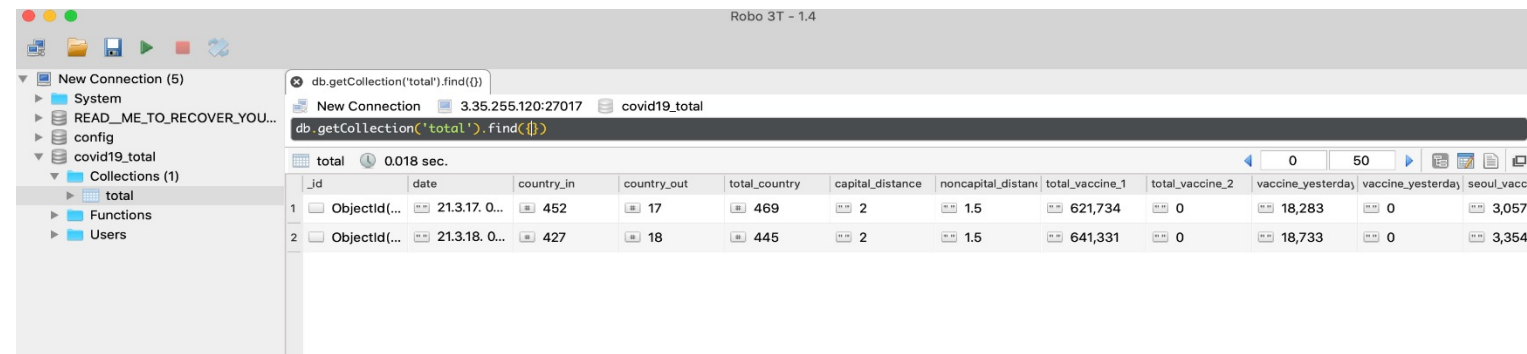
    def start_requests(self):
        yield scrapy.Request("http://ncov.mohw.go.kr/", callback=self.parse_content)

    def parse_content(self, response):
        item = Covid19TotalBackupItem()
        date = response.xpath('/html/body/div/div[5]/div[2]/div/div[1]/div[1]/h2/a/span[1]/text()').extract()
        date = list(date)[0].split(" ")
        date = date[0:3]
        date = " ".join(date)
        item['date'] = date.replace(" ", "21.").replace(" ", "")
        item['country_in'] = list(map(int, response.xpath('/html/body/div/div[5]/div[2]/div/div[1]/div[1]/div/ul/li[1]/span[2]/text()').extract()))
        item['country_out'] = list(map(int, response.xpath('/html/body/div/div[5]/div[2]/div/div[1]/div[1]/div/ul/li[2]/span[2]/text()').extract()))
        item['total_country'] = [item['country_in'], item['country_out']]
        item['capital_distance'] = response.xpath('//*[@id="main_maplayout"]/button[9]/span[2]/text()').extract()[0]
        item['noncapital_distance'] = response.xpath('//*[@id="main_maplayout"]/button[17]/span[2]/text()').extract()[0]
        meta = {'item': item}
        yield scrapy.Request('http://ncv.kdca.go.kr/', meta=meta, callback=self.parse)

    def parse(self, response):
        item = Covid19TotalBackupItem(response.meta['item'])
        take_yesterday_1 = response.xpath('//*[@id="content"]/div[2]/div/div/div[1]/div[1]/ul/li[2]/div/p/span/text()').extract()
        item['take_yesterday_1'] = list(take_yesterday_1)[0]
        take_yesterday_2 = response.xpath('//*[@id="content"]/div[2]/div/div/div[1]/div[2]/ul/li[2]/div/p/span/text()').extract()
        item['take_yesterday_2'] = list(take_yesterday_2)[0]
        total_1 = response.xpath('//*[@id="content"]/div[2]/div/div/div[1]/div[1]/ul/li[1]/div/p/span[1]/text()').extract()
        item['total_1'] = list(total_1)[0]
        total_2 = response.xpath('//*[@id="content"]/div[2]/div/div/div[1]/div[2]/ul/li[1]/div/p/span[1]/text()').extract()
        item['total_2'] = list(total_2)[0]
        seoul_vaccine = response.xpath('//*[@id="content"]/div[2]/div/div/div[2]/div/div[1]/div[2]/div/p[1]/span[2]/text()').extract()
        item['seoul_vaccine'] = list(seoul_vaccine)[0]
        gyeonggi_vaccine = response.xpath('//*[@id="content"]/div[2]/div/div/div[2]/div/div[1]/div[2]/div/p[9]/span[2]/text()').extract()
        item['gyeonggi_vaccine'] = list(gyeonggi_vaccine)[0]
        incheon_vaccine = response.xpath('//*[@id="content"]/div[2]/div/div/div[2]/div/div[1]/div[2]/div/p[4]/span[2]/text()').extract()
        item['incheon_vaccine'] = list(incheon_vaccine)[0]
        yield item
```

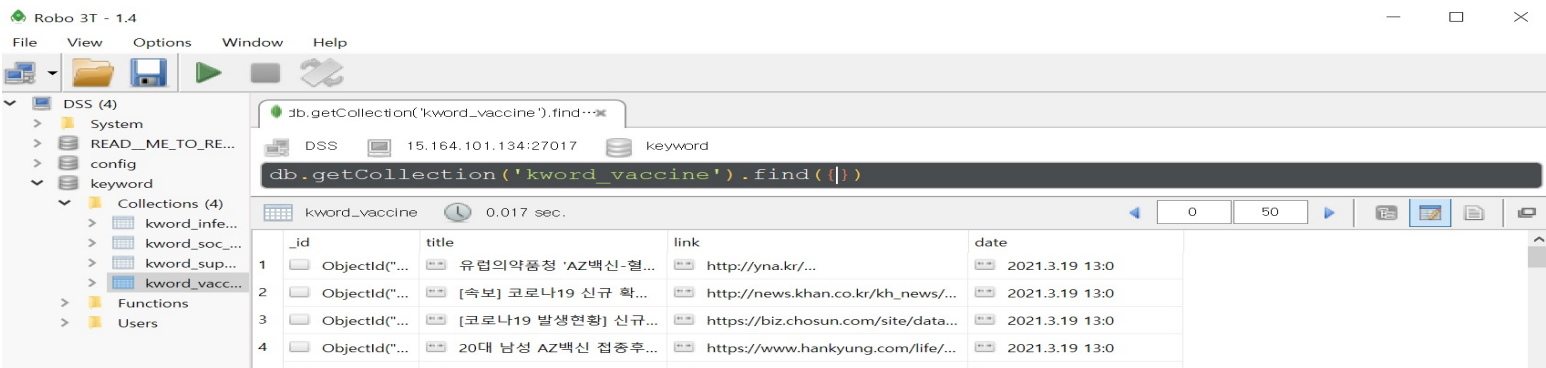
# 1. 코로나 관련 데이터 베이스

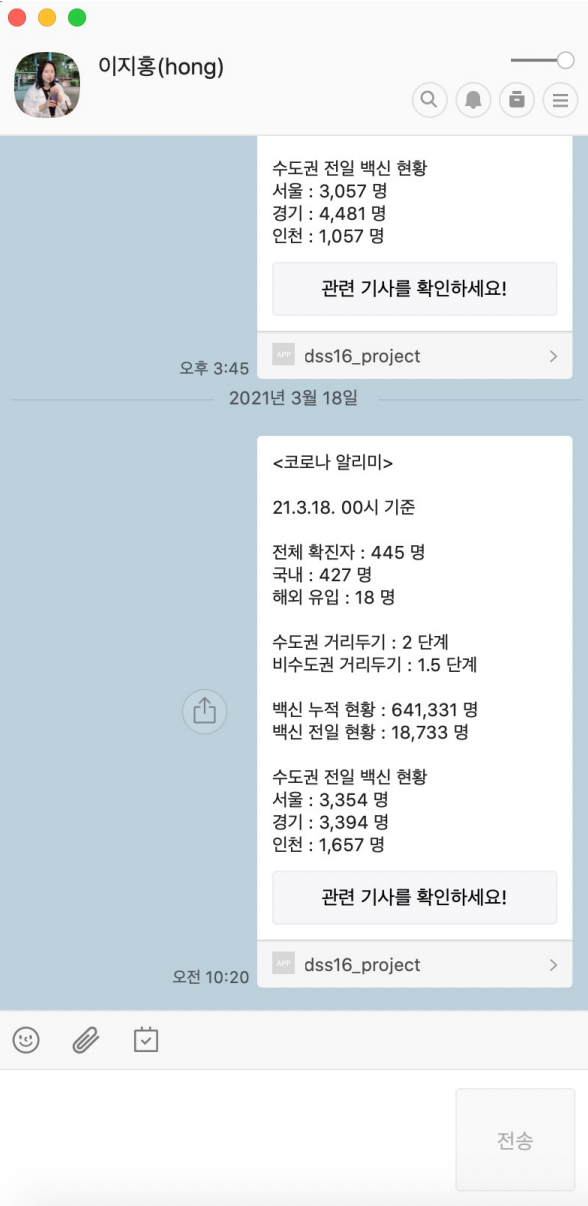
- 총 15개 컬럼 / lambda 사용하여 매일 아침 9시에 크롤링
- 당일 확진자 수, 거리두기 현황, 백신 누적 현황, 수도권 현황 데이터 수집



# 2. 코로나 관련 키워드 별 기사 데이터 베이스

- 4 collection 3 column
- 코로나 연관 키워드('백신', '거리두기', '확진', '재난지원금')





**Oauth Code**  
Selenium 이용



**메세지 보내기 API 사용**  
발급된 code 이용해 Access token 발급



**Pipeline.py 파일 생성**  
크롤링 된 데이터 첨부 및 링크 연결



**shell script 파일 생성**

링크 : <http://15.164.101.134/>

키워드별 최신 기사

백신 사회적 거리두기 국내 확진 재난 지원금

| # | Date                   | Title  | Link  |
|---|------------------------|--|---|
| 0 | 2021년 3월 19일<br>13시 6분 | 국내 코로나19 확진자 사흘째 400명대... 방역당국<br>거리두기 개편 고심 | <a href="http://www.kookje.co.kr/news2011/asp/newsbody.asp?code=0300&amp;key=20210319.99099006066">http://www.kookje.co.kr/news2011/asp/newsbody.asp?code=0300&amp;key=20210319.99099006066</a> |
| 1 | 2021년 3월 19일<br>13시 6분 | '서울형 작은학교' 신입생 19%... 코로나19 거리두기<br>적합       | <a href="https://www.nocutnews.co.kr/news/5518662">https://www.nocutnews.co.kr/news/5518662</a>   |
| 2 | 2021년 3월 19일<br>13시 6분 | 거제시, 19일부터 '사회적 거리두기 2단계'로                   | <a href="http://news.kmib.co.kr/article/view.asp?arcid=0015649681&amp;code=61122023&amp;cp=nv">http://news.kmib.co.kr/article/view.asp?arcid=0015649681&amp;code=61122023&amp;cp=nv</a>         |
| 3 | 2021년 3월 19일<br>13시 6분 | 속초시 거리두기 2단계 격상... 연쇄감염 '비상'                 | <a href="https://www.news1.kr/articles/74246672">https://www.news1.kr/articles/74246672</a>   |
| 4 | 2021년 3월 19일<br>13시 6분 | 유형업소발 확진 속출... 거제시 19일부터 거리두기<br>2단계로        | <a href="http://yna.kr/AKR20210318162300052?did=1195m">http://yna.kr/AKR20210318162300052?did=1195m</a>   |
| 5 | 2021년 3월 19일<br>13시 6분 | 거제시, 사회적 거리두기 2단계 격상... 조선소 폐쇄<br>조치 권고      | <a href="http://www.news1.com/view/?id=NISX20210318_0001375842&amp;clD=10812&amp;plD=10800">http://www.news1.com/view/?id=NISX20210318_0001375842&amp;clD=10812&amp;plD=10800</a>               |
| 6 | 2021년 3월 19일<br>13시 6분 | '코로나19 거리두기 장기화에 유튜브브드 '룩방'이 대<br>세"         | <a href="http://goodnews1.com/news/news_view.asp?seq=110132">http://goodnews1.com/news/news_view.asp?seq=110132</a>   |
| 7 | 2021년 3월 19일<br>13시 6분 | 주 초반 300명→중반 400명대... 거리두기 개편 차질<br>우려       | <a href="https://www.ytn.co.kr/_ln/0103_202103182021073584">https://www.ytn.co.kr/_ln/0103_202103182021073584</a>   |
| 8 | 2021년 3월 19일<br>13시 6분 | 석유공사, 사회적 거리두기 속 나눔활동 시작                     | <a href="http://www.gnetimes.co.kr/news/articleView.html?idxno=62136">http://www.gnetimes.co.kr/news/articleView.html?idxno=62136</a>   |
| 9 | 2021년 3월 19일           | 김제시 사회적 거리두기 2단계 격상                          | <a href="http://www.silksong.com/news/articleView.html?idxno=3104450">http://www.silksong.com/news/articleView.html?idxno=3104450</a>   |

Scrapy 크롤링 모듈 작성



Mongodb 데이터 적재, crontab 설정



Flask decorator / index.html 작성



웹페이지 구현

---

## 1. 힘들었던 점

- OAuth code 발급 문제
- Scrapy 모듈 작성 flask 구현과 jquery, decorator 작성

---

## 2. 아쉬운 점 및 보완 사항

- 확진자 및 백신 데이터베이스 내 row가 매일 한 줄 씩 사라짐
- Crontab으로 실행 시 .sh 파일이 OAuth code 발급까지 동작하고 메세지 전송이 안되는 문제
- 웹 페이지 디테일 설정

---

## 3. 앞으로의 계획

- 코로나 기간 내 이동과 생활 전반에 도움이 될 서비스를 구축 및 관리 하는 것이 목표
  - 현재 내가 있는 위치에 확진자 동선 알림 내용을 추가하여 확장 하고자 함
-

**Crawling  
Project**

**감사합니다**

---