

# 빅데이터를 활용한 SNS 인기 여행지

프로젝트명	대한민국 구석구석 SNS 인기 여행지 데이터 수집, 저장 및 RPA 활용
이름	김민준

ver 0.9.1  
2022. 05. 21

## 1. 개요

- ☐ 프로젝트 이름 : 대한민국 구석구석 SNS 인기 여행지 데이터 수집, 저장 및 RPA 활용
- ☐ 프로젝트 기간 : 2022.04.18 ~ 2022.05.13
- ☐ 프로젝트 인원: 김민준

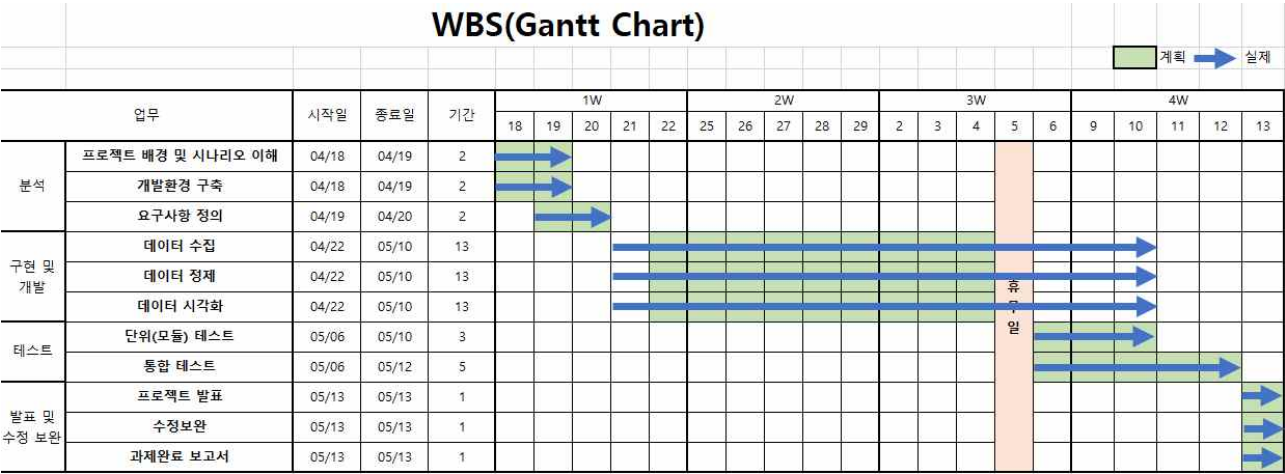
## 2. 추진 목표

- ☐ SNS 인기 여행지 데이터 수집, 저장 및 RPA 활용
  - 대한민국 구석구석의 인스타그램 인기 여행지 데이터 수집 및 저장
  - 수집 데이터를 저장 및 RPA 활용(이메일 전송, 카카오톡 메시지 전달)

## 3. 과제 수행 범위

- ☐ 대한민국 구석구석
  - 인스타그램 인기 여행지 TOP 11 수집
    - 여행지명, 지역, 상세정보, 해시태그, 여행지별 이미지 3장
- ☐ 인스타그램
  - 여행지명에 따른 게시물 개수 수집
    - 인기 여행지일수록 게시물 개수 증가
- ☐ 네이버 날씨
  - 네이버 날씨 크롤링(오늘, 내일, 모레)

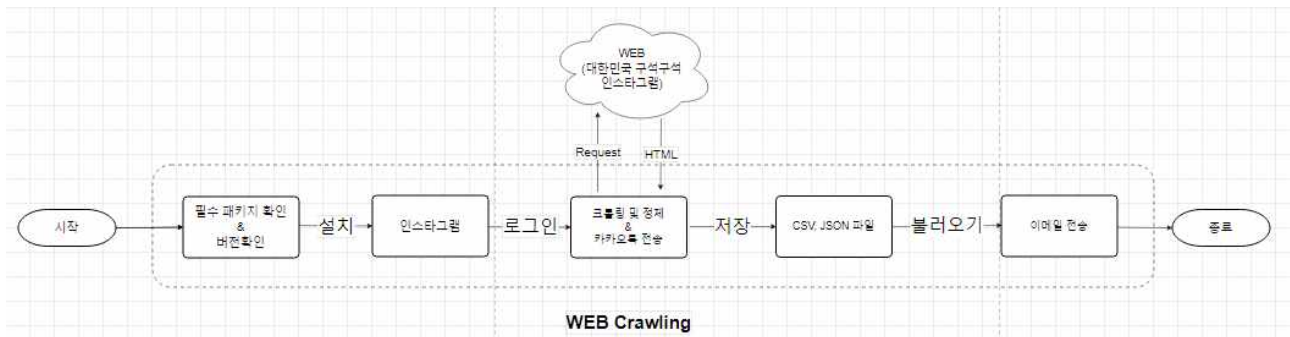
4. WBS (Work Breakdown Structure)



<그림 1.1 WBS>

## 1. 요구사항 목록

- ☐ 인스타그램 게시물 수집
  - 여행지 명에 맞는 게시물 수집
    - 여행지 명이 불명확할 경우 공백처리
  
- ☐ CSV, JSON 파일저장
  - 수집한 데이터 파일 저장
    - 업무용 혹은 웹 데이터 활용
  
- ☐ 파일 이메일 전송
  - 이메일 자동화
  
- ☐ 카카오톡 메시지 전송
  - 카카오톡 자동화



<그림 3.1 프로세스 설계>

□ 필수 패키지 확인 & 버전 확인

- 데이터 수집에 필요한 패키지 설치 및 최신 버전 확인

□ 인스타그램

- 데이터 수집 전 인스타그램 로그인

□ 크롤링 및 정제 & 카카오톡 전송

- 대한민국 구석구석 수집 및 여행지명에 맞는 인스타그램 수집, 수집 내용 카카오톡 피드 형식 전송

□ CSV, JSON 파일저장

- 데이터 수집 내용 CSV 파일과 JSON 파일저장

□ 이메일 전송

- CSV 파일과 JSON 파일을 이메일 전송

## 1. 데이터 저장

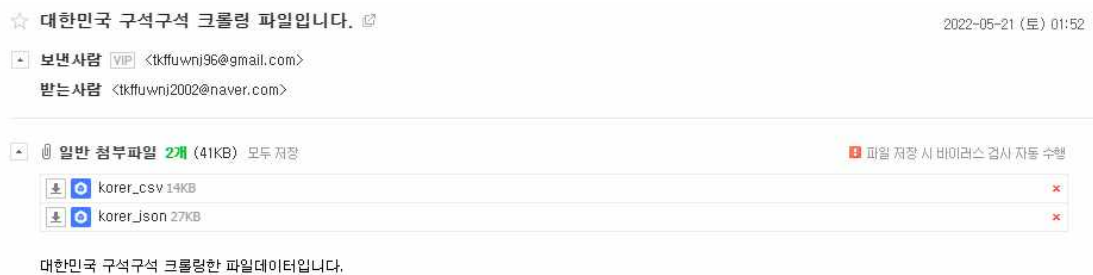
□ csv,json,jpg

korea.txt	2022-05-16 오전 7:04	텍스트 문서
korer_csv.csv	2022-05-16 오전 7:04	Microsoft Ex
korer_json.json	2022-05-16 오전 7:05	JSON 원본 I
경복궁0.jpg	2019-10-25 오후 2:01	JPG 파일
경복궁1.jpg	2015-05-20 오후 3:52	JPG 파일
경복궁2.jpg	2010-08-15 오후 4:56	JPG 파일
군위화산산성전망...	2019-07-22 오후 1:05	JPG 파일
군위화산산성전망...	2019-11-06 오전 9:10	JPG 파일
군위화산산성전망...	2018-11-04 오전 7:28	JPG 파일
농암종택0.jpg	2014-08-28 오후 1:50	JPG 파일
농암종택1.jpg	2014-08-28 오후 1:49	JPG 파일
농암종택2.jpg	2014-08-28 오후 1:54	JPG 파일

<그림 4.1 데이터 저장 결과>

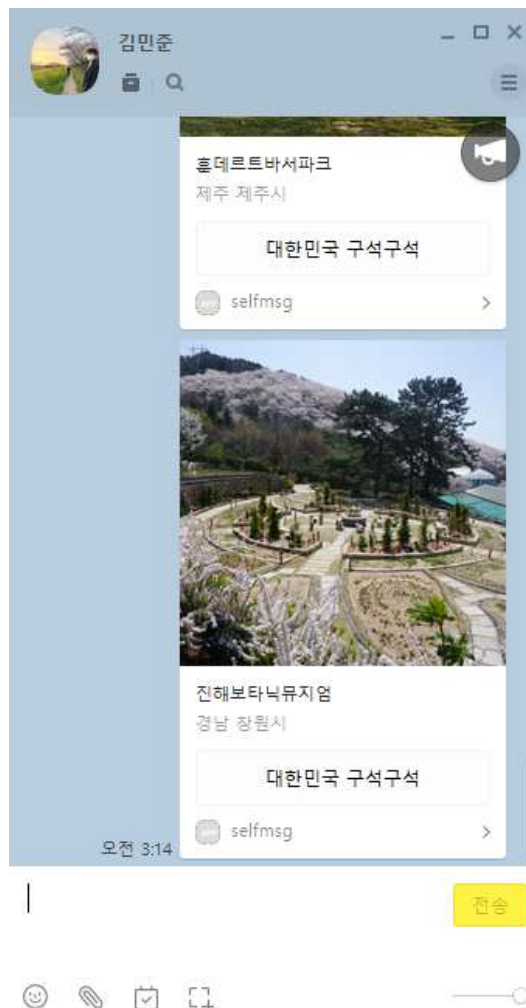
## 2. RPA 활용 - 이메일, 카카오톡

□ 이메일 결과



<그림 4.2 이메일 결과>

## □ 카카오톡 결과



<그림 4.3 카카오톡 결과>

## 3. 콘솔 창

### □ 대한민국 구석구석 및 인스타그램

```
=====
여행지: 진해보타닉뮤지엄 / 경남 창원시
검색태그: 진해보타닉뮤지엄 / 인스타그램 게시물 수: 10,674 건
상세정보: 진해보타닉뮤지엄은 뛰어난 자연환경속에 드림파크 생태숲과 어우러지며, 뒤로는 천자봉이 병풍처럼 둘러싸여있고 앞쪽으로는 진해 앞바다가 한눈에 내려다보이는 장복산의 산중턱에 위치한 경상남도 제1호 사립수목원이다. 식물에 대한 관심과 애정이 남달랐고, 정성들여 가꾼 정원을 여러 사람들과 함께 공유하고자 현재에 이르게 되었다.교목은 160종 (600주), 관목은 240종 (2,500주), 야생화 (다년생 초화) 600여종, (150,000본)이 식재되어 1월부터 12월까지 나무와 꽃들이 꽃을 피우거나, 단풍이 들며 열매를 맺어 아름다운 풍경을 연출한다.이렇게 식물들이 사계절 순서에 맞춰 개화하게 하는 기법으로 연출된 정원은 진해보타닉뮤지엄이 유일하다
URL : https://korean.visitkorea.or.kr/detail/ms_detail.do?cotid=744e1ab4-2079-4761-9c6b-72bd62956dd7&big_category=A01&mid_category=A0101&big_area=36
해시태그: #가족과함께#관광지#남녀노소#사진찍기좋은곳#사진촬영#사진촬영명소#산책하기좋은곳#아이와함께#연인과함께#자연좋은곳#진해보타닉뮤지엄
=====
```

<그림 4.4 데이터 수집 콘솔 창 결과>

## □ CSV, JSON 저장

```
=====
CSV 파일로 저장을 시작합니다
=====
c:\Users\jmk\Desktop\portfolio_kmj_ver0.4_20220520\portfolio\complete_code\korea_crawling.py:180: FutureWarning: The error
and will be removed in a future version.

save_data.csv_save() # save_data 모듈의 csv 저장 함수호출
=====
CSV 파일로 저장을 완료했습니다.
=====
JSON 파일로 저장을 시작합니다
=====
JSON 파일로 저장을 완료했습니다.
=====
```

<그림 4.5 데이터 저장 콘솔 창 결과>

## □ 네이버 날씨

```
=====
오늘 날씨 ( 5 / 18 )
=====
오늘 오전: 구름많음 / 오늘 오후: 구름많음
최저 기온: 14° / 최고 기온: 27°
오전 강수확률: 0% / 오후 강수확률: 0%
=====
내일 날씨 ( 5 / 19 )
=====
내일 오전: 맑음 / 내일 오후: 구름많음
최저 기온: 14° / 최고 기온: 27°
오전 강수확률: 0% / 오후 강수확률: 20%
=====
모래 날씨 ( 5 / 20 )
=====
모래 오전: 구름많음 / 모래 오후: 맑음
최저 기온: 14° / 최고 기온: 27°
오전 강수확률: 30% / 오후 강수확률: 0%
=====
네이버 날씨 바로가기 > https://weather.naver.com
=====
```

<그림 4.6 네이버 날씨 콘솔 창 결과>

## □ 메일 및 카카오톡

```
=====
메일을 보내는 중입니다.
=====
메일을 보내기를 완료했습니다.
=====

=====
카카오톡 메시지를 보냅니다.
=====
카카오톡 메시지 전송 성공하였습니다.
=====
```

<그림 4.7 이메일 및 카카오톡 콘솔 창 결과>