# 식의약용 자생식물 데이터 분석과 전처리

- Python을 이용하는 사전 데이터 분석 -

김규호

# 내용

- 데이터 파일 구성 탐색
- 데이터 내용 파악
- 데이터의 논리적 구조

# I. 데이터 파일 구성 탐색

사용한 프로그램

https://github.com/ekyuho/Plant/blob/main/plant_big_data.ipynb

# 순서

1. 전체 파일 리스트, 갯수확인
2. 폴더 경로 및 이름 규칙 파악
3. 원천데이타와 라벨링 데이터 구분 식별
4. 라벨링 데이터와 원천데이타의 각각 리스트 구성
5. 라벨링 데이터와 원천데이타의 1:1 매칭관계 파악
6. 라벨링 데이터로부터 원천 이미지화일을 연결시키는 도구 완성

# 1. 전체 파일: 리스트 확인

```
import os
mypath="E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터"

# os.walk는 디렉토리 전체를 리스팅 해줍니다.
path = os.walk(mypath)

# root 라는 변수를 통해 전체 디렉토리명을 구할 수 있습니다.
for root, directories, files in path:
    print(root)
```

# 1. 전체 파일: 개수 확인

```
import os
mypath="E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터"
path = os.walk(mypath)

# 이번에는 각 디렉토리의 화일 갯수를 살펴봅니다
for root, directories, files in path:
    files = os.listdir(root)
    print(root, '화일갯수=',len(files))
```

# 2. 폴더 및 화일이름 규칙확인

```
import os
import re
mypath="E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터"
path = os.walk(mypath)

# 디렉토리의 명칭이  ...데이터\05\009 형태인 것만 골라냅니다.
pattern = r'(\\\d{2}\\\d{3})$'
for root, directories, files in path:
    if re.search(pattern, root):
        print(root, len(os.listdir(root)))
```

# 3. 원천데이타와 라벨링 데이터 구분 식별

```
import os
import re
mypath="E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터"
path = os.walk(mypath)

# 라벨과 원천데이타의 화일갯수를 확인합니다.

pattern = r'(\\\d{2}\\\d{3})$'
n_label=0
n_image=0
for root, directories, files in path:
    n_files = len(os.listdir(root))
    if re.search(pattern, root):
        if '원천데이터' in root:
            n_image += n_files
        if '라벨링데이터' in root:
            n_label += n_files
        print(root, n_files)
print('원천데이타', n_image, '라벨링데이타', n_label)
```

# 4. 라벨링 데이터와 원천데이타의 각각 리스트 구성

```
import os
import re
mypath="E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터"
path = os.walk(mypath)

#원천데이터와 라벨데이터명을 리스트로 구성합니다.

pattern = r'(\\\d{2}\\\d{3})$'
n_label=0
n_image=0
label=[]
image=[]
for root, directories, files in path:
    if re.search(pattern, root):
        for file in os.listdir(root):
            fullpath = '{}\\{}'.format(root, file)
            if '원천데이터' in fullpath:
                n_source += 1
                image.append(fullpath)
            if '라벨링데이터' in fullpath:
                n_label += 1
                label.append(fullpath)

print('원천데이타', len(image), '라벨링데이타', len(label))
```

# 5. 라벨링 데이터와 원천데이타의 1:1 매칭관계 파악

```
import json

alabel = label[0]
with open(alabel, encoding='utf8') as f:
    j=json.load(f)
    print(json.dumps(j, indent=4, ensure_ascii=False))
```

{
"info": {
"datast_nm": "식의약용 자생식물 분석 데이터",
"datast_detail": "식의약용 자생식물 60종별로 식물종분류를 판별(인식)하기 위한 식물의 부위별(꽃, 열매, 잎앞면, 잎뒷면) AI 학습데이터",
"wd_plnt_idntfr": "009"
},
"images": {
"image_file_id": "CR009_02_50014a",
"image_file_nm": "CR009_02_50014a.jpg",
"file_stre_cours": "₩₩식의약용 자생식물 분석 데이터₩₩02₩₩009₩₩CR009_02_50014a.jpg",
"rsoltn": "332, 332",
"region_nm": "서귀포시 안덕면 서광리",
"region_type": "평지",
"plant_part": "꽃",
"grwh_step_nm": " ",
"image_file_frmat": "JPG",
"image_potogrf_dt": "2021-06-05 10:27:43"
},
"annotations": {
"antn_id": 3242973,
"antn_ty": "POLYGON",
"object_class_lclas_code": "02",
"object_class_mlsft_code": "009",
"object_class_sclas_code": "02",
"object_class_lclas_nm": "초본",
"object_class_mlsft_nm": "메밀",
"object_class_sclas_nm": "꽃",
"pyn_xcrdnt":
"61,54,55,59,66,45,32,26,21,22,29,22,14,11,6,7,12,16,21,27,38,48,58,71,90,103,112,122,127,122,117,108,100,97,99,108,120,128,138,155,151,150,156,166,171,182,192,203,213,219,224,229,236,242,252,266,283,296,300,299,290,282,269,258,269,278,283,287,290,284,277,260,247,251,252,253,262,271,290,307,314,320,322,313,299,290,282,284,290,293,288,277,272,260,251,238,233,223,217,210,202,194,186,185,184,185,185,184,186,188,182,178,173,165,158,143,130,116,107,98,90,85,81,72,67",
"pyn_ycrdnt":
"64,71,83,94,107,104,109,117,128,140,157,162,169,176,182,188,187,188,191,195,198,200,201,201,194,191,187,185,189,194,198,203,209,215,221,233,240,244,249,250,262,276,304,317,322,319,316,305,293,286,280,275,281,286,289,295,295,291,276,263,241,233,226,221,214,206,193,179,168,155,153,153,155,146,142,134,133,132,127,121,116,111,100,93,91,90,89,81,77,72,64,59,46,33,28,32,38,44,48,38,28,22,27,35,42,50,59,66,79,100,111,120,131,146,157,150,142,136,129,121,112,106,83,68,64"
},
"plants": {
"wd_plnt_nm": "메밀",
"scientific_nm": "Fagopyrum esculentum Moench, 1794.",
"woody_herbal": "2",
"edible_yn": "Y",
"edible_part": "잎, 줄기, 꽃, 씨앗",
"acquisition_term": "06월",
"efficacy": "항염증|면역증진|항산화",
"efficacy_ncl": "8.9|9.0|0.0",
"nutrient": "15.1|15.8|15.2"
},
"ingredient": {
"irdnt_nm": "6-Hydroxykaempferol-3-O-glucoside|Kaempferol 3-O-β-D-glucopyranoside|1-O-Caffeoyquinic acid|Cnidimol F|Procyanidin B8|Rhamnetin|1-Galloyl-glucose|Quinquenoside I |5,7,8,3-Tetrahydroxy-3,4-dimethoxy flavone|3,5-Dihydroxybenzoic acid",
"irdnt_ncl": "0.6|0.4|-0.3|0.5|-0.4|0.0|0.0|-0.2|0.4|0.5",
"irdnt_chmcls_strct": " ",
"irdnt_anals_instlm": "제주산학융합원"
},
"licenses": {
"lcnse_nm": "CC-BY-SA"
}
}

01.원천데이터₩식의약용 자생식물 분석 데이터₩02₩009₩CR009_02_50014a.jpg

Google

online json parser

All | Books | Images | News | Videos | More

About 5,270,000 results (0.48 seconds)

http://json.parser.online.fr

## Json Parser Online

Analyze your **JSON** string as you type with an **online** Javascript **parser**, featuring tree view and syntax highlighting. Processing is done locally: no data send ...

Try out Beta! · Donation in Bitcoins · FAQ

You've visited this page many times. Last visit: 1/24/22

https://jsonformatter.org › json-parser

## Best JSON Parser Online

**JSON Parser Online** helps to parse, view, analyze JSON data in Tree View. It's a pretty simple and easy way to parse JSON data and share it with others. This ...

**Json Parser Online** You like it? Support it! Donate

Try out Beta! Samples ▾ Options ▾


```
{
    "info": {
        "datast_nm": "식의약용 자생식물 분석 데이터",
        "datast_detail": "식의약용 자생식물 60종별로 식물종분류를 판별(인식)하기 위한 식물의 부위별(꽃, 열매, 잎앞면, 잎뒷면) AI 학습데이터",
        "wd_plnt_idntfr": "009"
    },
    "images": {
        "image_file_id": "CR009_05_52927",
        "image_file_nm": "CR009_05_52927.jpg",
        "file_stre_cours": "\\식의약용 자생식물 분석 데이터\\05\\009\\CR009_05_52927.jpg",
        "rsoltn": "1823, 1823",
        "region_nm": "성산",
        "region_type": "평지",
        "plant_part": "잎-뒷면",
        "grwh_step_nm": " ",
        "image_file_frmat": "JPG",
        "image_potogrf_dt": "2021-10-31 17:12:26"
    },
    "annotations": {
        "antn_id": 3580832,
        "antn_ty": "POLYGON",
        "object_class_lclas_code": "02",
        "object_class_mlsft_code": "009",
        "object_class_sclas_code": "05",
        "object_class_lclas_nm": "초본",
        "object_class_mlsft_nm": "메밀",
        "object_class_sclas_nm": "잎-뒷면",
        "pyn_xcrdnt": "275,428,525,616,752,862,912,947,1056,1169,1224,1224,1285,1324,1332,1366,1375,1377,1519,1522,1373,1366,1371,1387,1363,1327,1302,1276,1125,1082,1077,1043,1013,1004,994,969,976,962,923,898,871,797,761,706,676,543,486,410,284",
        "pyn_ycrdnt": "788,766,754,761,715,704,637,548,467,449,472,491,516,523,566,690,750,876,891,930,924,940,1024,1192,1270,1421,1446,1434,1336,1322,1306,1283,1279,1265,1283,1254,1249,1238,1228,1233,1221,1123,1068,1031,1002,924,889,869,811"
    },
    "plants": {
        "wd_plnt_nm": "메밀",
        "scientific_nm": "Fagopyrum esculentum Moench, 1794.",
        "woody_herbal": "2",
        "edible_yn": "Y",
        "edible_part": "잎, 줄기, 꽃, 씨앗",
        "acquisition_term": "10월",
        "efficacy": "항염증|면역증진|항산화",
        "efficacy_ncl": "8.9|9.0|0.0",
        "nutrient": "15.1|15.8|15.2"
    },
    "ingredient": {
        "irdnt_nm": "6-Hydroxykaempferol-3-O-glucoside|Kaempferol 3-O-β-D-glucopyranoside|1-O-Caffeoyquinic acid|Cnidimol F|Procyanidin B8|Rhamnetin|1-Galloyl-glucose|Quinquenoside I |5,7,8,3-Tetrahydroxy-3,4-dimethoxy flavone|3,5-Dihydroxybenzoic acid",
        "irdnt_ncl": "0.6|0.4|-0.3|0.5|-0.4|0.0|0.0|-0.2|0.4|0.5",
        "irdnt_chmcls_strct": " ",
```


String parse | JS eval


```
{
  "info": {
    "datast_nm":"식의약용 자생식물 분석 데이터",
    "datast_detail":"식의약용 자생식물 60종별로 식물종분류를 판별(인식)하기 위한 식물의 부위별(꽃, 열매, 잎앞면, 잎뒷면) AI 학습데이터",
    "wd_plnt_idntfr":"009"
  },
  "images": {
    "image_file_id":"CR009_05_52927",
    "image_file_nm":"CR009_05_52927.jpg",
    "file_stre_cours":"\\식의약용 자생식물 분석 데이터\\05\\009\\CR009_05_52927.jpg",
    "rsoltn":"1823, 1823",
    "region_nm":"성산",
    "region_type":"평지",
    "plant_part":"잎-뒷면",
    "grwh_step_nm":" ",
    "image_file_frmat":"JPG",
    "image_potogrf_dt":"2021-10-31 17:12:26"
  },
  "annotations": {
    "antn_id":3580832,
    "antn_ty":"POLYGON",
    "object_class_lclas_code":"02",
    "object_class_mlsft_code":"009",
    "object_class_sclas_code":"05",
    "object_class_lclas_nm":"초본",
    "object_class_mlsft_nm":"메밀",
    "object_class_sclas_nm":"잎-뒷면",
    "pyn_xcrdnt":"275,428,525,616,752,862,912,947,1056,1169,1224,1224,1285,1324,1332,1366,1375,1377,1519,1522,1373,1366,1371,1387,1363,1327,1302,1276,1125,1082,1077,1043,1013,1004,994,969,976,962,923,898,871,797,761,706,676,543,486,410,284",
    "pyn_ycrdnt":"788,766,754,761,715,704,637,548,467,449,472,491,516,523,566,690,750,876,891,930,924,940,1024,1192,1270,1421,1446,1434,1336,1322,1306,1283,1279,1265,1283,1254,1249,1238,1228,1233,1221,1123,1068,1031,1002,924,889,869,811"
  },
  "plants": {
    "wd_plnt_nm":"메밀",
    "scientific_nm":"Fagopyrum esculentum Moench, 1794.",
    "woody_herbal":"2",
    "edible_yn":"Y",
    "edible_part":"잎, 줄기, 꽃, 씨앗",
    "acquisition_term":"10월",
    "efficacy":"항염증|면역증진|항산화",
    "efficacy_ncl":"8.9|9.0|0.0",
    "nutrient":"15.1|15.8|15.2"
  },
  "ingredient": {
    "irdnt_nm":"6-Hydroxykaempferol-3-O-
```

제공데이터₩02.라벨링데이터₩식의약용 자생식물 분석 데이터₩02₩009

| 이름 | 수정한 날짜 | 유형 |
|---|---|---|
| CR009_02_50014a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50015.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50017a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50018a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50019.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50021.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50023.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50044d.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50044h.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50071.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50077.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50085a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50086a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50087.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50088a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50088e.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50088z.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50096.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50102.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50103.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50117.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50117a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50122a.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50137.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50230b.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50343.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50345.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50350.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50354.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |
| CR009_02_50355.json | 2022-01-13 오후 5:42 | JSON 원본 파일 |

제공데이터₩01.원천데이터₩식의약용 자생식물 분석 데이터₩02₩009

| 이름 | 날짜 | 유형 |
|---|---|---|
| CR009_02_50014a.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50015.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50017a.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50018a.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50019.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50021.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50023.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50044d.jpg | 2021-06-05 오전 10:33 | JPG 파일 |
| CR009_02_50044h.jpg | 2021-06-05 오전 10:33 | JPG 파일 |
| CR009_02_50071.jpg | 2021-06-05 오전 10:25 | JPG 파일 |
| CR009_02_50077.jpg | 2021-06-05 오전 10:26 | JPG 파일 |
| CR009_02_50085a.jpg | 2021-06-05 오전 10:26 | JPG 파일 |
| CR009_02_50086a.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50087.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50088a.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50088e.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50088z.jpg | 2021-06-05 오전 10:27 | JPG 파일 |
| CR009_02_50096.jpg | 2021-06-05 오전 10:30 | JPG 파일 |
| CR009_02_50102.jpg | 2021-06-05 오전 10:31 | JPG 파일 |
| CR009_02_50103.jpg | 2021-06-05 오전 10:31 | JPG 파일 |
| CR009_02_50117.jpg | 2021-06-05 오전 10:32 | JPG 파일 |
| CR009_02_50117a.jpg | 2021-06-05 오전 10:32 | JPG 파일 |
| CR009_02_50122a.jpg | 2021-06-05 오전 10:33 | JPG 파일 |
| CR009_02_50137.jpg | 2021-06-05 오전 10:36 | JPG 파일 |
| CR009_02_50230b.jpg | 2021-06-30 오후 3:14 | JPG 파일 |
| CR009_02_50343.jpg | 2021-06-14 오전 9:26 | JPG 파일 |
| CR009_02_50345.jpg | 2021-06-14 오전 9:27 | JPG 파일 |
| CR009_02_50350.jpg | 2021-06-14 오전 9:27 | JPG 파일 |
| CR009_02_50354.jpg | 2021-06-14 오전 9:29 | JPG 파일 |
| CR009_02_50355.jpg | 2021-06-14 오전 9:29 | JPG 파일 |

# 5. JSON 라벨링 데이터와 이미지 파일 명칭 규칙파악

```
import json

alabel = label[0]
with open(alabel, encoding='utf8') as f:
    j=json.load(f)
    print(j["images"]["file_stre_cours"])
```

```
import json

root='E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터\\01.원천데이터'
alabel = label[0]
with open(alabel, encoding='utf8') as f:
    j=json.load(f)
    imagefile = '{}{}'.format(root,j["images"]["file_stre_cours"])
    exist = os.path.isfile(imagefile)
    print(imagefile, exist)
```

# 5. 전체 JSON레이블링 파일 내용 검증

```
import json

root='E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터\\01.원천데이터'
i=0
for alabel in label:
    with open(alabel, encoding='utf8') as f:
        j=json.load(f)
        imagefile = '{}{}'.format(root,j["images"]["file_stre_cours"])
        exist = os.path.isfile(imagefile)
        if not exist: print(i, "ERROR", imagefile, exist)
        if i%100==0: print(i, imagefile, exist)
        i+=1
```

# 5. 전체 JSON레이블링 파일 내용 검증 (예외처리)

```
import json

root='E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터\\01.원천데이터'
i=0
e=0
for alabel in label:
    with open(alabel, encoding='utf8') as f:
        try:
            j=json.load(f)
        except:
            print(e, i, alabel)
            e+=1
            continue
        imagefile = '{}{}'.format(root,j["images"]["file_stre_cours"])
        exist = os.path.isfile(imagefile)
        if not exist: print(i, "ERROR", imagefile, exist)
        #if i%100==0: print(i, imagefile, exist)
        i+=1
print("processed {} files. error={}. total {} files.".format(i, e, i+e))
```

# 6. 라벨링 데이터로부터 원천 이미지화일을 연결시키는 도구 완성

```
import matplotlib.pyplot as plt
import matplotlib.image as img
import json

alabel = label[9000]
with open(alabel, encoding='utf8') as f:
    j=json.load(f)

root='E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터\\01.원천데이터'
imagefile=root+j["images"]["file_stre_cours"]
im=img.imread(imagefile)
print(imagefile, im.shape)
plt.figure()
plt.imshow(im)
plt.colorbar()
plt.grid(False)
plt.show()

print(json.dumps(j, indent=4, ensure_ascii=False))
```

# 6. 라벨링 데이터로부터 원천 이미지화일을 읽은 함수 완성

```
import matplotlib.pyplot as plt
import matplotlib.image as img
import json

def look(no):
    alabel = label[no]
    with open(alabel, encoding='utf8') as f:
        j=json.load(f)

    root='E:\\.shortcut-targets-by-id\\1H5ozpQqq7a9WMbr3XnAMqCFq0-QoGUhr\\220114제공데이터\\01.원천데이터'
    imagefile=root+j["images"]["file_stre_cours"]
    im=img.imread(imagefile)
    print(imagefile, im.shape)
    plt.figure()
    plt.imshow(im)
    plt.colorbar()
    plt.grid(False)
    plt.show()

    print(json.dumps(j, indent=4, ensure_ascii=False))
```

```
look(9000)
```

# II. 데이터 내용 파악

# JSON 레이블 구성


```
{
    "info": {
        "datast_nm": "식의약용 자생식물 분석 데이터",
        "datast_detail": "식의약용 자생식물 60종별로 식물종분류를 판별(인식)하기 위한 식물의 부위별(꽃, 열매, 잎앞면, 잎뒷면) AI 학습데이터",
        "wd_plnt_idntfr": "051"
    },
    "images": {
        "image_file_id": "CR051_02_51036",
        "image_file_nm": "CR051_02_51036.jpg",
        "file_stre_cours": "₩₩식의약용 자생식물 분석 데이터₩₩02₩₩051₩₩CR051_02_51036.jpg",
        "rsoltn": "994, 994",
        "region_nm": "선덕사",
        "region_type": "평지",
        "plant_part": "꽃",
        "grwh_step_nm": " ",
        "image_file_frmat": "JPG",
        "image_potogrf_dt": "2021-10-27 12:24:40"
    },
    "plants": {
        "wd_plnt_nm": "한라꽃향유",
        "scientific_nm": "Elsholtzia hallasanensis Y.N.Lee, 2000.",
        "woody_herbal": "2",
        "edible_yn": "N",
        "edible_part": "N/A",
        "acquisition_term": "10월",
        "efficacy": "항염증|면역증진|항산화",
        "efficacy_ncl": "0.0|0.0|0.0",
        "nutrient": "15.2|15.6|15.6"
    },
    "ingredient": {
        "irdnt_nm": "Escin IVd|Chlorogenic acid|Kaempferol-3-O-β-D-glucuronide|Kaempferol-3-O-β-D-glucopyranoside|Genistin|Sanleng acid|Luteolin 7-beta-neohesperidoside|(E,E)-9-Oxooctadeca-10,12-dienoic acid|1-O-Caffeoyquinic acid|Apigenol",
        "irdnt_ncl": "0.2|-0.4|0.0|0.1|0.6|0.6|0.7|0.3|0.4|0.6",
        "irdnt_chmcls_strct": " ",
        "irdnt_anals_instlm": "제주산학융합원"
    },
    "licenses": {
        "lcnse_nm": "CC-BY-SA"
    }
```

# JSON 레이블 구성 (계속)


```
 "annotations": {
        "antn_id": 3600691,
        "antn_ty": "POLYGON",
        "object_class_lclas_code": "02",
        "object_class_mlsft_code": "051",
        "object_class_sclas_code": "02",
        "object_class_lclas_nm": "초본",
        "object_class_mlsft_nm": "한라꽃향유",
        "object_class_sclas_nm": "꽃",
        "pyn_xcrdnt":
"515,501,489,480,466,456,461,461,449,431,420,419,418,414,398,370,352,347,348,354,363,349,34
3,345,347,350,356,351,352,355,360,368,359,349,348,351,355,344,337,328,328,328,324,319,318,3
17,322,327,346,344,363,360,351,349,340,333,329,333,341,359,350,346,349,351,360,376,394,376,
364,357,361,377,385,380,374,371,376,384,393,399,411,398,394,390,389,396,402,409,414,432,424
,418,419,427,437,434,436,437,444,453,461,473,476,474,474,477,480,478,481,489,494,500,510,51
7,518,523,527,535,539,546,558,564,571,574,574,575,583,590,593,599,600,598,593,602,615,616,6
12,596,606,617,623,627,624,617,599,612,618,624,621,617,609,606,612,620,627,636,641,630,624,
614,597,604,609,613,611,607,600,595,602,614,625,631,629,618,597,610,621,620,614,605,600,592
,608,618,625,624,615,605,620,627,632,629,623,611,601,591,607,626,639,639,649,651,646,638,61
2,629,635,631,627,620,606,599,592,597,603,620,624,624,619,606,597,591,597,606,608,604,597,5
85,565,553,529",
        "pyn_ycrdnt":
"808,817,817,814,816,818,832,844,856,857,853,844,836,824,818,789,775,764,751,741,738,726,71
3,708,701,695,690,685,673,663,655,650,643,634,624,615,613,610,604,601,596,589,590,589,584,5
69,563,557,557,550,550,541,536,524,518,505,491,484,479,472,459,449,432,422,418,418,414,400,
386,370,368,386,384,374,366,356,346,343,342,340,339,324,315,304,286,290,283,278,275,274,259
,246,238,243,254,241,227,221,218,215,215,222,220,208,195,182,176,161,153,152,154,151,151,15
3,141,127,113,112,117,115,111,113,120,133,141,158,161,168,169,172,188,202,218,218,226,238,2
48,267,267,270,277,287,294,314,325,328,334,342,354,363,364,369,374,370,360,358,362,374,382,
387,392,399,401,410,422,429,434,437,445,444,441,441,448,456,460,466,472,483,495,502,504,508
,509,504,503,514,519,522,525,529,539,555,564,572,572,571,584,578,571,563,564,573,581,582,59
4,602,616,626,635,643,645,641,643,648,658,664,676,694,706,700,705,711,718,727,734,748,765,7
66,769,774,784"
    },
}
```

# 데이터의 종수 분석

```
# 전체 JSON 화일을 읽어들여 List로 만들어 둡니다.
i=0
all=[]
for alabel in label:
    if not alabel.endswith("json"): continue
    with open(alabel, encoding='utf8') as f:
        j=json.load(f)
        all.append(j)
        i+=1
print('total', i, 'files')
```

```
# JSON내의 필드 분포를 하나하나 살펴보기위한 함수를 만듭니다.
def examine(field1, field2):
    s={}
    key='{}.{}'.format(field1, field2)
    for j in all:
        val=j[field1][field2]
        if not key in s: s[key]={}
        if not val in s[key]: s[key][val]=0
        s[key][val] +=1
    return s
```

```
for f1 in all[0]:
    for f2 in all[0][f1]:
        print('{}.{}'.format(f1, f2), len(examine(f1,f2)['{}.{}'.format(f1,f2)]))
```

```
info.datast_nm 1
info.datast_detail 1
info.wd_plnt_idntfr 15
```

```
licenses.lcnse_nm 1
```

```
ingredient.irdnt_nm 14
ingredient.irdnt_ncl 14
ingredient.irdnt_chmcls_strct 1
ingredient.irdnt_anals_instlm 1
```

```
images.image_file_id 10127
images.image_file_nm 10127
images.file_stre_cours 10127
images.rsoltn 2004
images.region_nm 33
images.region_type 2
images.plant_part 4
images.grwh_step_nm 1
images.image_file_frmat 1
images.image_potogrf_dt 7143
```

```
annotations.antn_id 10127
annotations.antn_ty 1
annotations.object_class_lclas_code 2
annotations.object_class_mlsft_code 15
annotations.object_class_sclas_code 4
annotations.object_class_lclas_nm 2
annotations.object_class_mlsft_nm 15
annotations.object_class_sclas_nm 4
annotations.pyn_xcrdnt 10101
annotations.pyn_ycrdnt 10099
```

```
plants.wd_plnt_nm 15
plants.scientific_nm 15
plants.woody_herbal 2
plants.edible_yn 2
plants.edible_part 7
plants.acquisition_term 7
plants.efficacy 1
plants.efficacy_ncl 11
plants.nutrient 13
```

```
"info": {
    "datast_nm": "식의약용 자생식물 분석 데이터",
    "datast_detail": "식의약용 자생식물 60종별로 식물종분류를 판별(인식)하기 위한 식물의 부위별(꽃, 열매, 잎앞면, 잎뒷면) AI 학습데이터",
    "wd_plnt_idntfr": "051"
},
```

```python
def show(key):
    k=key.split('.')
    for x in sorted(examine(k[0],k[1])[key].items(), key=lambda x: x[1], reverse=True):
        print(x)
```

```python
show("info.wd_plnt_idntfr")
```

```
('040', 1159)
('046', 1133)
('009', 1089)
('041', 591)
('043', 572)
(160, 568)
('048', 567)
('022', 567)
('001', 565)
(153, 564)
('020', 563)
('002', 559)
('011', 547)
('051', 543)
('004', 540)
```

```
"images": {
    "image_file_id": "CR051_02_51036",
    "image_file_nm": "CR051_02_51036.jpg",
    "file_stre_cours": "\\식의약용 자생식물 분석 데이터\\02\\051\\CR051_02_51036.jpg",
    "rsoltn": "994, 994",
    "region_nm": "선덕사",
    "region_type": "평지",
    "plant_part": "꽃",
    "grwh_step_nm": " ",
    "image_file_frmat": "JPG",
    "image_potogrf_dt": "2021-10-27 12:24:40"
},
```


```
show("images.region_nm")
```

```
('한라수목원', 3123)
('농산물원종장', 1287)
('제주대학교', 827)
('제주특별자치도 제주시 용담동', 770)
('만장굴', 547)
('선덕사', 543)
('성산', 540)
('판포리', 530)
('동광리', 382)
('제주특별자치도 제주시 노형동', 376)
('용담해안도로', 368)
('연동', 225)
('제주특별자치도 제주시 오등동', 208)
('한라생태숲', 105)
('무수천', 76)
('제주특별자치도 서귀포시 성산읍', 61)
('제주특별자치도 제주시 조천읍', 51)
('제주특별자치도 제주시 영평동', 31)
('서귀포시 안덕면 서광리', 24)
('제주특별자치도 제주시 삼양동', 17)
('한라수목원 자연생태학습관', 11)
('첨단입구 교차로', 6)
('제주대 공과대학3호관', 4)
('제주시 아라일동', 4)
('제주특별자치도 제주시 구좌읍 월정리', 2)
('월정투명카약주차장', 2)
('제주시 오등동', 1)
('한라산 어리목 주차장', 1)
('한라생태숲 목련총림', 1)
('안덕계곡', 1)
('휘닉스', 1)
('제주특별자치도 제주시 한림읍 협재리', 1)
('섭지코지', 1)
```

```
"images": {
    "image_file_id": "CR051_02_51036",
    "image_file_nm": "CR051_02_51036.jpg",
    "file_stre_cours": "\\식의약용 자생식물 분석 데이터\\02\\051\\CR051_02_51036.jpg",
    "rsoltn": "994, 994",
    "region_nm": "선덕사",
    "region_type": "평지",
    "plant_part": "꽃",
    "grwh_step_nm": " ",
    "image_file_frmat": "JPG",
    "image_potogrf_dt": "2021-10-27 12:24:40"
},
```


```
show("images.region_type")
```

```
('평지', 9072)
('경사지', 1055)
```

```
show("images.plant_part")
```

```
('열매', 5092)
('꽃', 3955)
('잎-뒷면', 540)
('잎-앞면', 540)
```

```
,,
"annotations": {
    "antn_id": 3600691,
    "antn_ty": "POLYGON",
    "object_class_lclas_code": "02",
    "object_class_mlsft_code": "051",
    "object_class_sclas_code": "02",
    "object_class_lclas_nm": "초본",
    "object_class_mlsft_nm": "한라꽃향유",
    "object_class_sclas_nm": "꽃",
```


```
show("annotations.object_class_lclas_code")
```

```
('01', 7361)
('02', 2766)
```

```
show("annotations.object_class_lclas_nm")
```

```
('목본', 7361)
('초본', 2766)
```

```
show("annotations.object_class_sclas_code")
```

```
('03', 5092)
('02', 3955)
('05', 540)
('04', 540)
```

```
show("annotations.object_class_sclas_nm")
```

```
('열매', 5092)
('꽃', 3955)
('잎-뒷면', 540)
('잎-앞면', 540)
```

```
show("annotations.object_class_mlsft_code")
```

```
('040', 1159)
('046', 1133)
('009', 1089)
('041', 591)
('043', 572)
(160, 568)
('048', 567)
('022', 567)
('001', 565)
(153, 564)
('020', 563)
('002', 559)
('011', 547)
('051', 543)
('004', 540)
```

```
show("annotations.object_class_mlsft_nm")
```

```
('순비기나무', 1159)
('황근', 1133)
('메밀', 1089)
('참꽃나무', 591)
('참가시나무', 572)
('꽝꽝나무', 568)
('해국', 567)
('큰조롱', 567)
('까마귀쪽나무', 565)
('백량금', 564)
('돈나무', 563)
('좁은잎천선과', 559)
('구실잣밤나무', 547)
('한라꽃향유', 543)
('참식나무', 540)
```

```
},
"plants": {
    "wd_plnt_nm": "한라꽃향유",
    "scientific_nm": "Elsholtzia hallasanensis Y.N.Lee, 2000.",
    "woody_herbal": "2",
    "edible_yn": "N",
    "edible_part": "N/A",
    "acquisition_term": "10월",
    "efficacy": "항염증|면역증진|항산화",
    "efficacy_ncl": "0.0|0.0|0.0",
    "nutrient": "15.2|15.6|15.6"
},
```


```
show("plants.wd_plnt_nm")
```

```
('순비기나무', 1159)
('황근', 1133)
('메밀', 1089)
('참꽃나무', 591)
('참가시나무', 572)
('꽝꽝나무', 568)
('해국', 567)
('큰조롱', 567)
('까마귀쪽나무', 565)
('백량금', 564)
('돈나무', 563)
('좁은잎천선과', 559)
('구실잣밤나무', 547)
('한라꽃향유', 543)
('참식나무', 540)
```

```
show("plants.scientific_nm")
```

```
('Vitex rotundifolia L. f., 1781.', 1159)
('Hibiscus hamabo Siebold & Zucc., 1841.', 1133)
('Fagopyrum esculentum Moench, 1794.', 1089)
('Rhododendron weyrichii Maxim., 1871.', 591)
('Quercus salicina Blume, 1850.', 572)
('Ilex crenata', 568)
('Aster spathulifolius Maxim', 567)
('Cynanchum wilfordii (Maxim.) Hemsl., 1889.', 567)
('Litsea japonica (Thunb.) Juss., 1801.', 565)
('Ardisia crenata', 564)
('Pittosporum tobira (Thunb.) W. T. Aiton, 1811.', 563)
('Ficus erecta var. sieboldii (Miq.) King, 1888.', 559)
('Castanopsis sieboldii (Makino) Hatus. ex T. Yamaz. & Mashiba, 1971.', 547)
('Elsholtzia hallasanensis Y.N.Lee, 2000.', 543)
('Neolitsea sericea (Blume) Koidz., 1926.', 540)
```

```
"plants": {
    "wd_plnt_nm": "한라꽃향유",
    "scientific_nm": "Elsholtzia hallasanensis Y.N.Lee, 2000.",
    "woody_herbal": "2",
    "edible_yn": "N",
    "edible_part": "N/A",
    "acquisition_term": "10월",
    "efficacy": "항염증|면역증진|항산화",
    "efficacy_ncl": "0.0|0.0|0.0",
    "nutrient": "15.2|15.6|15.6"
},
```


```
show("plants.woody_herbal")
```

```
('1', 7361)
('2', 2766)
```

```
show("plants.edible_yn")
```

```
('N', 5093)
('Y', 5034)
```

```
show("plants.edible_part")
```

```
('N/A', 5093)
('열매', 1684)
('잎, 줄기, 꽃, 씨앗', 1089)
('뿌리, 줄기(제한 사용)', 568)
('잎', 567)
('뿌리(물추출물에 한함)', 567)
('잎, 열매', 559)
```

```
show("plants.acquisition_term")
```

```
('10월', 3935)
('11월', 2816)
('08월', 2789)
('07월', 414)
('06월', 123)
('12월', 45)
('05월', 5)
```

```
show("plants.efficacy_ncl")
```

```
('0.0|0.0|0.0', 1701)
('분석 진행중', 1699)
('67.0|3.6|72.4', 1159)
('41.1|13.3|0.0', 1133)
('8.9|9.0|0.0', 1089)
('46.4|17.1|82.6', 572)
('72.0|28.7|71.4', 565)
('88.7|9.7|52.1', 563)
('13.4|13.8|51.6', 559)
('57.8|18.2|82.4', 547)
('41.8|11.3|74.3', 540)
```

```
show("plants.nutrient")
```

```
('분석 진행중', 1699)
('15.3|15.2|15.0', 1159)
('15.0|15.4|15.6', 1133)
('15.1|15.8|15.2', 1089)
('15.2|15.2|15.4', 591)
('15.4|15.2|15.5', 572)
('15.1|15.6|16.0', 567)
('16.2|15.9|16.0', 565)
('15.3|15.7|15.6', 563)
('16.1|15.5|16.2', 559)
('15.3|15.2|16.7', 547)
('15.2|15.6|15.6', 543)
('15.6|16.2|15.5', 540)
```

"ingredient": {
"irdnt_nm": "Escin IVd|Chlorogenic acid|Kaempferol-3-O-β-D-glucuronide|Kaempferol-3-O-β-D-glucopyranoside|Genistin|Sanleng acid|Luteolin 7-beta-neohesperidoside|(E,E)-9-Oxooctadeca-10,12-dienoic acid|1-O-Caffeoyquinic acid|Apigenol",
"irdnt_ncl": "0.2|-0.4|0.0|0.1|0.6|0.6|0.7|0.3|0.4|0.6",
"irdnt_chmcls_strct": " ",
"irdnt_anals_instlm": "제주산학융합원"
},

```
show("ingredient.irdnt_nm")
```

```
('Agnuside|Oxypaeoniflorin|Trifolin|Kaempferol 3-O-β-D-glucuronide|Cimicifugic acid B|Apigenin-7-O-acetyl-β-D-glucoside|20(S)-Ginsenoside Rh2|5,6,4-Trihydroxy-7,8-dimethoxyflavone', 1159)
('Procyanidin B1|(E,E)-9-oxooctadeca-10,12-dienoic acid|Quinquenoside I|Tianshic acid|Kaempferol-3-Glucoside-2-p-coumaroyl|6-Hydroxykaempferol-3-O-glucoside|Vernolic acid|Acrinidioionoside|Decaffeoylverbascoside|Glucosyringic acid', 1133)
('분석 진행중', 1132)
('6-Hydroxykaempferol-3-O-glucoside|Kaempferol 3-O-β-D-glucopyranoside|1-O-Caffeoyquinic\xa0acid|Cnidimol F|Procyanidin\xa0B8|Rhamnetin|1-Galloyl-glucose|Quinquenoside I |5,7,8,3-Tetrahydroxy-3,4-dimethoxy flavone|3,5-Dihydroxybenzoic acid', 1089)
("6-Hydroxykaempferol-3-O-glucoside|21-O-Methyl\xa0toosendanopentaol|20(S)-Ginsenoside\xa0Rh2\xa0(Ginsenoside\xa0Rh2)|3β-O-trans-p-Coumaroyl\xa0alphitolic\xa0acid|Chlorogenic\xa0acid|1-O-Caffeoyquinic\xa0acid|Kaempferol-3-Glucoside-2'-p-coumaroyl|Procyanidin\xa0B8|Quercetin-3-O-xyloside|Neocomplanoside", 591)
('Madecassoside|Ellagic acid|Quinic acid|2,3-(S)-hexahydroxydiphenoyl-D-glucose|3-O-Methylellagic acid|Ethyl caffeate|3,3-Di-O-methylellagic acid|Tianshic acid|2,3-(S)-hexahydroxydiphenoyl-D-glucose|Yamogenin acetate', 572)
('apigenin-7-O-galactopyranoside|Hyperin|Rutin|Quinic\xa0acid|Quercetin-3-O-α-L-rhamnoside|Apigenol|Quercetin-3'-O-glucoside_1|Kaempferol', 567)
('Polygoacetophenoside|Kaempferol\xa03-O-β-D-glucopyranoside|Quercetin-3-O-xyloside|Quercetin-3-O-α-D-glucuronide|Kaempferol-3-O-β-D-glucuronide|Catechin-3-O-gallate|Bruceine\xa0H|Procyanidin\xa0B1|Sanleng\xa0acid|1-Galloyl-glucose', 567)
('Kaempferol-7-rhamnoside|chlorogenic acid|Questin|Chrysosplenetin B|Myricetin 3-O-glucoside|(E,E)-9-Oxooctadeca-10,12-dienoic acid|6-Hydroxykaempferol 3-glucoside|Luteolin-7-beta-neohesperidoside|lithospermic acid B|Trifolin', 565)
('Chlorogenic acid|Methyl chlorogenate|3-O-trans-Coumaroylquinic acid|1-O-Caffeoyl-β-D-glucopyranose|akebia saponin E|Lablaboside D|Vernolic acid|Methyl chlorogenate|(E,E)-9-Oxooctadeca-10,12-dienoic acid|Shikimic acid', 563)
('Cnidimol F|6-Hydroxykaempferol 3-glucoside|(E,E)-9-Oxooctadeca-10,12-dienoic\xa0acid|Sanleng acid|1-O-Caffeoyl-β-D-glucopyranose|morroniside|3-O-trans-p-Coumaroyl alphitolic acid|Curculigoside B|Kaempferol 3,7-diglucoside|Procyanidin\xa0B8', 559)
('1,2,3,6-Tetra-O-galloyl-D-glucose|quercetin-3-O-glucoside|1,2,6-tris-O-galloyl-β-D-glucose|Quercetin-3-O-α-D-glucuronide|Mudanpioside E|3,6-Di-O-Galloyl-β-D-glucose|1-Galloylglucose|Sesamoside|Secologanoside|Methyl 11α-hydroxytormentate', 547)
('Escin\xa0Ⅳd|Chlorogenic\xa0acid|Kaempferol-3-O-β-D-glucuronide|Kaempferol-3-O-β-D-glucopyranoside|Genistin|Sanleng\xa0acid|Luteolin\xa07-beta-neohesperidoside|(E,E)-9-Oxooctadeca-10,12-dienoic\xa0acid|1-O-Caffeoyquinic\xa0acid|Apigenol', 543)
('Dihydrokaempferol-5-O-β-D-glucopyranoside|Trifolin|6-Hydroxykaempferol 3-glucoside|Procyanidin B1|Kaempferol 3-O-(6"-O-p-coumaroyl)glucoside|1,6,7-Trihydroxy-2,3-dimethoxyxanthone|Gaillardin|Darendoside A|Cimidahurinine|Vernolic acid', 540)
```

```
show("ingredient.irdnt_ncl")
```

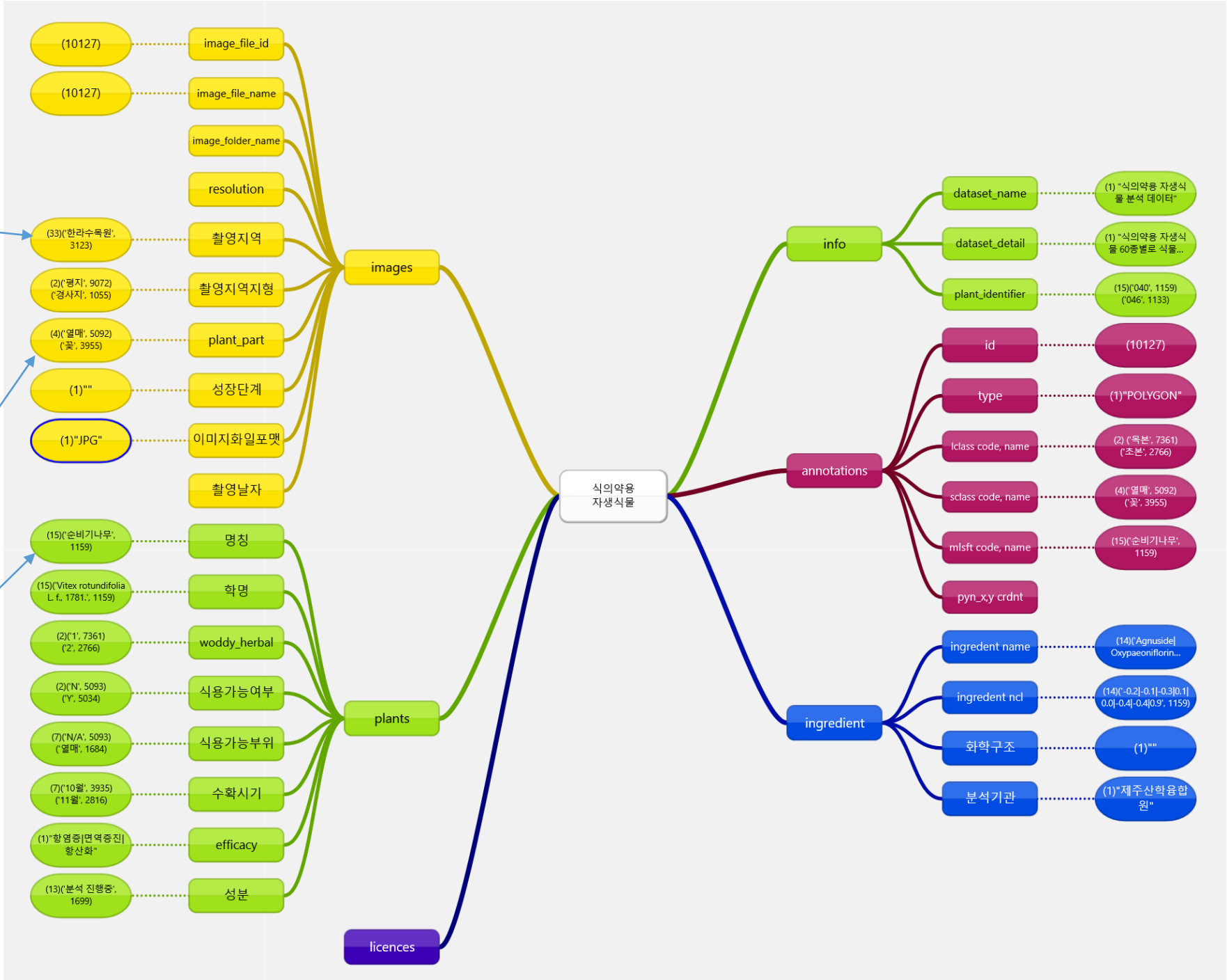
```
('-0.2|-0.1|-0.3|0.1|0.0|-0.4|-0.4|0.9', 1159)
('-0.1|-0.1|0.7|0.9|-0.4|-0.4|0.6|-0.5|0.3|-0.5', 1133)
('분석 진행중', 1132)
('0.6|0.4|-0.3|0.5|-0.4|0.0|0.0|-0.2|0.4|0.5', 1089)
('0.2|1.0|0.2|-0.5|-0.3|0.1|0.9|0.2|0.0|0.2', 591)
('-1.4|0.2|-0.1|-0.2|-0.7|0.7|0.5|0.3|-0.7|1.7', 572)
('0.0|-0.2|-0.4|-0.7|-0.3|0.5|-0.2|0.2', 567)
('2.8|0.4|0.4|-0.3|0.2|0.0|0.6|0.0|0.2|0.6', 567)
('-0.1|-1.9|-0.5|-0.3|-0.1|0.5|0.2|0.8|0.3|1.0', 565)
('0.5|0.3|0.8|1.3|0.6|0.4|1.6|0.7|1.7|3.5', 563)
('0.4|0.9|0.3|0.2|0.9|0.1|0.6|0.0|-0.6|0.4', 559)
('-0.6|-0.5|-0.6|-0.7|-0.7|-0.9|-0.7|-0.5|-0.5|0.6', 547)
('0.2|-0.4|0.0|0.1|0.6|0.6|0.7|0.3|0.4|0.6', 543)
('0.5|0.7|0.0|-0.6|-0.1|0.3|0.8|0.4|0.3|0.0', 540)
```

# III. 데이터의 논리적구조

```
('한라수목원', 3123)
('농산물원종장', 1287)
('제주대학교', 827)
('제주특별자치도 제주시 용담동', 770)
('만장굴', 547)
('선덕사', 543)
('성산', 540)
('판포리', 530)
('동광리', 382)
('제주특별자치도 제주시 노형동', 376)
('용담해안도로', 368)
('연동', 225)
('제주특별자치도 제주시 오등동', 208)
('한라생태숲', 105)
('무수천', 76)
('제주특별자치도 서귀포시 성산읍', 61)
('제주특별자치도 제주시 조천읍', 51)
('제주특별자치도 제주시 영평동', 31)
('서귀포시 안덕면 서광리', 24)
('제주특별자치도 제주시 삼양동', 17)
('한라수목원 자연생태학습관', 11)
('첨단입구 교차로', 6)
('제주대 공과대학3호관', 4)
('제주시 아라일동', 4)
('제주특별자치도 제주시 구좌읍 월정리', 2)
('월정투명카약주차장', 2)
('제주시 오등동', 1)
('한라산 어리목 주차장', 1)
('한라생태숲 목련총림', 1)
('안덕계곡', 1)
('휘닉스', 1)
('제주특별자치도 제주시 한림읍 협재리', 1)
('섭지코지', 1)
```

```
('열매', 5092)
('꽃', 3955)
('잎-뒷면', 540)
('잎-앞면', 540)
```

```
('순비기나무', 1159)
('황근', 1133)
('메밀', 1089)
('참꽃나무', 591)
('참가시나무', 572)
('짱짱나무', 568)
('해국', 567)
('큰조롱', 567)
('까마귀쪽나무', 565)
('백량금', 564)
('돈나무', 563)
('좁은잎천선과', 559)
('구실잣밤나무', 547)
('한라꽃향유', 543)
('참식나무', 540)
```



# 성분데이타 전체의 종류 파악

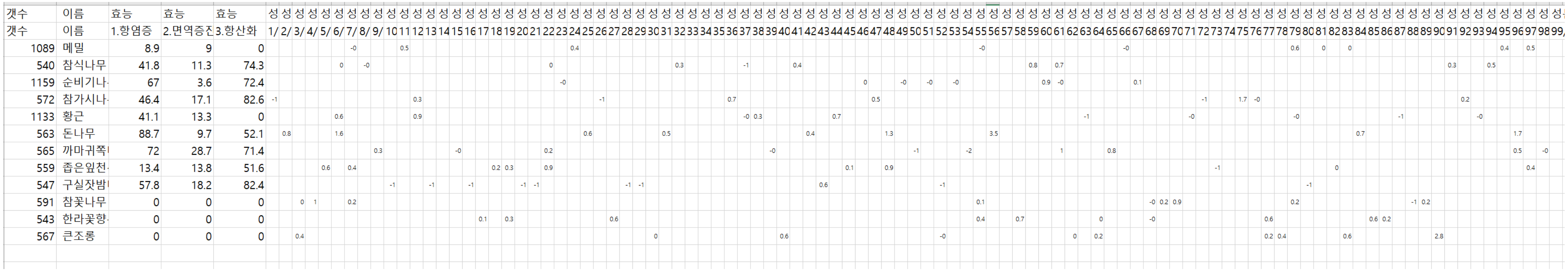
```python
# 성분데이타 전체의 종수를 파악합니다.
ingredient=set()
numbers={}
for item in all:
    t1 = item["ingredient"]["irdnt_nm"].split('|')
    c=len(t1)
    if not c in numbers: numbers[c]=0
    numbers[c]+=1
    if len(t1) == 1: continue    # 분석진행중
    ingredient = ingredient | set(t1)
print(numbers)
ingr=list(ingredient)
print(len(ingr), ingr)
```

```
{10: 7269, 8: 1726, 1: 1132}
99 ['Madecassoside', '3-O-trans-Coumaroylquinic acid', 'Quercetin-3-O-xyloside', '21-O-Methyl₩xa0toosendanopentaol', '3-O-trans-p-Coumaroyl alphitolic acid', 'Vernolic acid', 'Procyanidin₩xa0B8', 'Kaempferol 3-O-(6''-O-p-coumaroyl)glucoside', 'lithospermic acid B', '1-Galloylglucose', '3,5-Dihydroxybenzoic acid', 'Tianshic acid', '3,6-Di-O-Galloyl-β-D-glucose', 'Hyperin', 'Myricetin 3-O-glucoside', 'quercetin-3-O-glucoside', 'Kaempferol-3-O-β-D-glucopyranoside', 'Sanleng acid', '(E,E)-9-Oxooctadeca-10,12-dienoic₩xa0acid', 'Mudanpioside E', 'Sesamoside', '6-Hydroxykaempferol 3-glucoside', '20(S)-Ginsenoside Rh2', 'Kaempferol 3-O-β-D-glucopyranoside', 'akebia saponin E', '3-O-Methylellagic acid', 'Genistin', '1,2,6-tris-O-galloyl-β-D-glucose', '1,2,3,6-Tetra-O-galloyl-D-glucose', 'Catechin-3-O-gallate', 'Chlorogenic acid', 'Cimidahurinine', 'Kaempferol', 'Quercetin-3′ -O-glucoside_1', 'apigenin-7-O-galactopyranoside', 'Ethyl caffeate', 'Procyanidin B1', 'Decaffeoylverbascoside', 'Kaempferol-7-rhamnoside', 'Bruceine₩xa0H', 'Darendoside A', 'Lablaboside D', 'Methyl 11α-hydroxytormentate', 'Quinquenoside I', 'morronoside', 'Cimicifugic acid B', '3,3-Di-O-methylellagic acid', '1-O-Caffeoyl-β-D-glucopyranose', 'Oxypaeoniflorin', 'Questin', 'Agnuside', 'Quercetin-3-O-α-D-glucuronide', 'Apigenin-7-O-acetyl-β-D-glucoside', 'chlorogenic acid', '1-O-Caffeoyquinic₩xa0acid', 'Shikimic acid', 'Quinic₩xa0acid', 'Luteolin₩xa07-beta-neohesperidoside', 'Gaillardin', '5,6,4-Trihydroxy-7,8-dimethoxyflavone', 'Trifolin', 'Procyanidin₩xa0B1', 'Glucosyringic acid', 'Kaempferol-3-O-β-D-glucuronide', 'Luteolin-7-beta-neohesperidoside', 'Quinquenoside I ', 'Kaempferol 3-O-β-D-glucuronide', 'Chlorogenic₩xa0acid', '20(S)-Ginsenoside₩xa0Rh2₩xa0(Ginsenoside₩xa0Rh2)', "Kaempferol-3-Glucoside-2'-p-coumaroyl", 'Kaempferol-3-Glucoside-2-p-coumaroyl', '2,3-(S)-hexahydroxydiphenoyl-D-glucose', 'Kaempferol 3,7-diglucoside', 'Quercetin-3-O-α-L-rhamnoside', 'Yamogenin acetate', 'Quinic acid', 'Sanleng₩xa0acid', 'Kaempferol₩xa03-O-β-D-glucopyranoside', '6-Hydroxykaempferol-3-O-glucoside', 'Secologanoside', 'Rhamnetin', 'Curculigoside B', '1-Galloyl-glucose', 'Methyl chlorogenate', 'Apigenol', 'Escin₩xa0IVd', 'Acrinidioionoside', '3β-O-trans-p-Coumaroyl₩xa0alphitolic₩xa0acid', 'Neocomplanoside', 'Polygoacetophenoside', '1,6,7-Trihydroxy-2,3-dimethoxyxanthone', 'Ellagic acid', '(E,E)-9-oxooctadeca-10,12-dienoic acid', 'Dihydrokaempferol-5-O-β-D-glucopyranoside', '5,7,8,3-Tetrahydroxy-3,4-dimethoxy flavone', '(E,E)-9-Oxooctadeca-10,12-dienoic acid', 'Cnidimol F', 'Chrysosplenetin B', 'Rutin']
```

# JSON 라벨링 데이터에서 성분사항 추출

```
# 전체 JSON 화일에서 성분관련한 부분을 추출합니다.
unique={}
with open("plant_data.csv", "w", encoding='utf-8-sig') as f:
    n=0
    title="이름,효능1/항염증,효능2/면역증진,효능3/항산화"
    for k in range(len(ingr)):
        title += ',"성분{}/{}"'.format(k+1,ingr[k])
    #print(title)
    print(title,file=f)
    for item in all:
        a = item["annotations"]
        p = item["plants"]
        i = item["ingredient"]
        pe=p["efficacy_ncl"].split('|')
        if len(pe) != 3 and pe[0]=='분석 진행중':
            continue
        inm=i["irdnt_nm"].split('|')
        icl=i["irdnt_ncl"].split('|')
        idict = dict(zip(inm,icl))
        #print(idict)
        data='{},{},{},{}'.format(a["object_class_mlsft_nm"],pe[0],pe[1],pe[2])
        for k in range(len(ingr)):
            data += ','+idict.get(ingr[k],"")
        n +=1
        if n%1000==0: print(n, data)
        print(data,file=f)
        if not data in unique: unique[data]=1
        else: unique[data]+=1
        #if n>10: break
print("₩n Unique Ingredient")
with open("plant_unique_ingredient.csv", "w", encoding='utf-8-sig') as f:
    print('갯수,',title,file=f)
    for x in unique:
        print(unique[x],',', x, file=f)
```

# 자생식물별 효능과 성분표

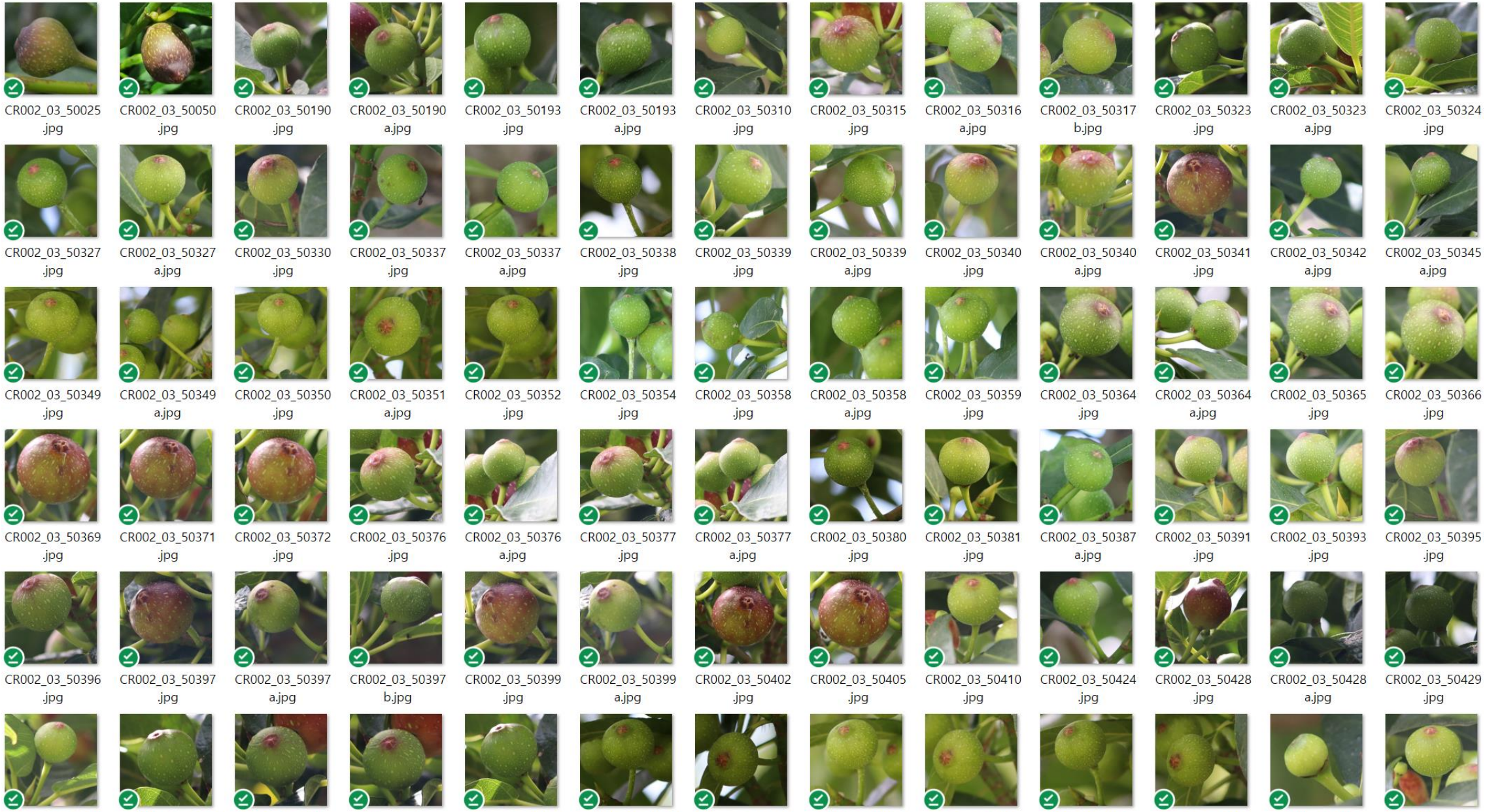


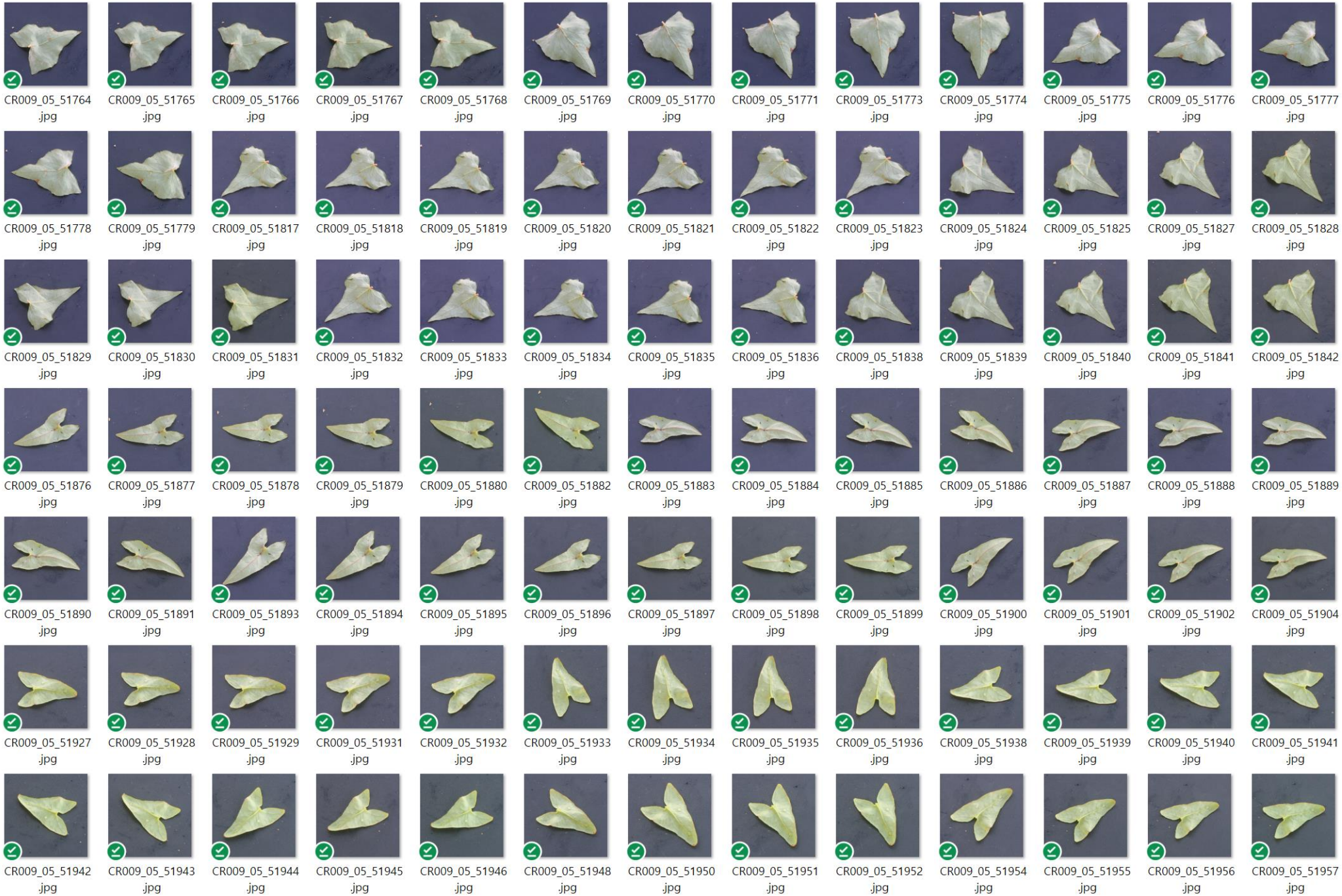
| 갯수 | 이름 | 효능 | 효능 | 효능 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 갯수 | 이름 | 1.항염증 | 2.면역증진 | 3.항산화 | 1/ | 2/ | 3/ | 4/ | 5/ | 6/ | 7/ | 8/ | 9/ | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
| 1089 | 메밀 | 8.9 | 9 | 0 | | | | | | | -0 | | | | 0.5 | | | | | | | | | | | | | 0.4 | | | | | | | | | |
| 540 | 참식나무 | 41.8 | 11.3 | 74.3 | | | | | | 0 | | -0 | | | | | | | | | | | | | | 0 | | | | | | | | | | 0.3 | |
| 1159 | 순비기나 | 67 | 3.6 | 72.4 | | | | | | | | | | | | | | | | | | | | | | | -0 | | | | | | | | | | |
| 572 | 참가시나 | 46.4 | 17.1 | 82.6 | -1 | | | | | | | | | | | 0.3 | | | | | | | | | | | | | | -1 | | | | | | | |
| 1133 | 황근 | 41.1 | 13.3 | 0 | | | | | | 0.6 | | | | | | 0.9 | | | | | | | | | | | | | | | | | | | | | |
| 563 | 돈나무 | 88.7 | 9.7 | 52.1 | | 0.8 | | | | 1.6 | | | | | | | | | | | | | | | | | | | 0.6 | | | | | | 0.5 | | |
| 565 | 까마귀쪽 | 72 | 28.7 | 71.4 | | | | | | | | | 0.3 | | | | | | -0 | | | | | | | 0.2 | | | | | | | | | | | |
| 559 | 좁은잎천 | 13.4 | 13.8 | 51.6 | | | | | 0.6 | | 0.4 | | | | | | | | | | | 0.2 | 0.3 | | | 0.9 | | | | | | | | | | | |
| 547 | 구실잣밤 | 57.8 | 18.2 | 82.4 | | | | | | | | | | -1 | | | -1 | | | -1 | | | | -1 | -1 | | | | | | | -1 | -1 | | | | |
| 591 | 참꽃나무 | 0 | 0 | 0 | | | 0 | 1 | | | 0.2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 543 | 한라꽃향 | 0 | 0 | 0 | | | | | | | | | | | | | | | | | 0.1 | | 0.3 | | | | | | | | 0.6 | | | | | | |
| 567 | 큰조롱 | 0 | 0 | 0 | | | 0.4 | | | | | | | | | | | | | | | | | | | | | | | | | | | 0 | | | |

| 갯수 | 이름 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 갯수 | 이름 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
| 1089 | 메밀 | | | | | | | | | | | | | | | | | | | | | | -0 | | | | | | | | | | | -0 |
| 540 | 참식나무 | | | | -1 | | | | 0.4 | | | | | | | | | | | | | | | | | | 0.8 | | 0.7 | | | | | |
| 1159 | 순비기나 | | | | | | | | | | | | | 0 | | | -0 | | -0 | | -0 | | | | | | | 0.9 | -0 | | | | | |
| 572 | 참가시나 | | | 0.7 | | | | | | | | | | | 0.5 | | | | | | | | | | | | | | | | | | | |
| 1133 | 황근 | | | | -0 | 0.3 | | | | | | 0.7 | | | | | | | | | | | | | | | | | | | -1 | | | |
| 563 | 돈나무 | | | | | | | | | 0.4 | | | | | | 1.3 | | | | | | | | 3.5 | | | | | | | | | | |
| 565 | 까마귀쪽 | | | | | | -0 | | | | | | | | | | | -1 | | | | -2 | | | | | | | 1 | | | | 0.8 | |
| 559 | 좁은잎천 | | | | | | | | | | | | 0.1 | | | 0.9 | | | | | | | | | | | | | | | | | | |
| 547 | 구실잣밤 | | | | | | | | | | 0.6 | | | | | | | | | -1 | | | | | | | | | | | | | | |
| 591 | 참꽃나무 | | | | | | | | | | | | | | | | | | | | | | 0.1 | | | | | | | | | | | |
| 543 | 한라꽃향 | | | | | | | | | | | | | | | | | | | | | | 0.4 | | | 0.7 | | | | | | 0 | | |
| 567 | 큰조롱 | | | | | | | 0.6 | | | | | | | | | | | | -0 | | | | | | | | | | 0 | | 0.2 | | |

| 갯수 | 이름 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 | 성 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 갯수 | 이름 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99, |
| 1089 | 메밀 | | | | | | | | | | | | | 0.6 | | 0 | | 0 | | | | | | | | | | | | 0.4 | | 0.5 | | |
| 540 | 참식나무 | | | | | | | | | | | | | | | | | | | | | | | | | 0.3 | | | 0.5 | | | | | |
| 1159 | 순비기나 | 0.1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 572 | 참가시나 | | | | | | -1 | | | 1.7 | -0 | | | | | | | | | | | | | | | | 0.2 | | | | | | | |
| 1133 | 황근 | | | | | -0 | | | | | | | | -0 | | | | | | | | -1 | | | | | | -0 | | | | | | |
| 563 | 돈나무 | | | | | | | | | | | | | | | | | | 0.7 | | | | | | | | | | | | 1.7 | | | |
| 565 | 까마귀쪽 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 0.5 | | -0 | |
| 559 | 좁은잎천 | | | | | | | -1 | | | | | | | | | 0 | | | | | | | | | | | | | | | 0.4 | | |
| 547 | 구실잣밤 | | | | | | | | | | | | | | -1 | | | | | | | | | | | | | | | | | | | |
| 591 | 참꽃나무 | | -0 | 0.2 | 0.9 | | | | | | | | | 0.2 | | | | | | | | | -1 | 0.2 | | | | | | | | | | |
| 543 | 한라꽃향 | | -0 | | | | | | | | | 0.6 | | | | | | | | 0.6 | 0.2 | | | | | | | | | | | | | |
| 567 | 큰조롱 | | | | | | | | | | | 0.2 | 0.4 | | | | | 0.6 | | | | | | | 2.8 | | | | | | | | | |


CR009_02_50014a.jpg
CR009_02_50015.jpg
CR009_02_50017a.jpg
CR009_02_50018a.jpg
CR009_02_50019.jpg
CR009_02_50021.jpg
CR009_02_50023.jpg
CR009_02_50044d.jpg
CR009_02_50044h.jpg
CR009_02_50071.jpg
CR009_02_50077.jpg
CR009_02_50085a.jpg
CR009_02_50086a.jpg
CR009_02_50087.jpg
CR009_02_50088a.jpg
CR009_02_50088e.jpg
CR009_02_50088z.jpg
CR009_02_50096.jpg
CR009_02_50102.jpg
CR009_02_50103.jpg
CR009_02_50117.jpg
CR009_02_50117a.jpg
CR009_02_50122a.jpg
CR009_02_50137.jpg
CR009_02_50230b.jpg
CR009_02_50343.jpg
CR009_02_50345.jpg
CR009_02_50350.jpg
CR009_02_50354.jpg
CR009_02_50355.jpg
CR009_02_50356.jpg
CR009_02_50357.jpg
CR009_02_50358.jpg
CR009_02_50361.jpg
CR009_02_50363.jpg
CR009_02_50366.jpg
CR009_02_50369.jpg
CR009_02_50374.jpg
CR009_02_50376.jpg
CR009_02_50377.jpg
CR009_02_50378.jpg
CR009_02_50380.jpg
CR009_02_50381.jpg
CR009_02_50385.jpg
CR009_02_50386.jpg
CR009_02_50387.jpg
CR009_02_50389.jpg
CR009_02_50393.jpg
CR009_02_50394.jpg
CR009_02_50395.jpg
CR009_02_50398.jpg
CR009_02_50399.jpg
CR009_02_50401.jpg
CR009_02_50404.jpg
CR009_02_50407.jpg
CR009_02_50409.jpg
CR009_02_50414.jpg
CR009_02_50416.jpg
CR009_02_50417.jpg
CR009_02_50424.jpg
CR009_02_50425.jpg
CR009_02_50427.jpg
CR009_02_50428.jpg
CR009_02_50429.jpg
CR009_02_50433.jpg
CR009_02_50435.jpg
CR009_02_50444.jpg
CR009_02_50446.jpg
CR009_02_50448.jpg
CR009_02_50451.jpg
CR009_02_50452.jpg
CR009_02_50454.jpg
CR009_02_50457.jpg
CR009_02_50458.jpg
CR009_02_50466.jpg
CR009_02_50467.jpg
CR009_02_50468.jpg
CR009_02_50469.jpg
CR009_02_50470.jpg
CR009_02_50471.jpg
CR009_02_50475.jpg
CR009_02_50476.jpg
CR009_02_50479.jpg
CR009_02_50489.jpg
CR009_02_50490.jpg
CR009_02_50491.jpg
CR009_02_50492.jpg
CR009_02_50493.jpg
CR009_02_50494.jpg
CR009_02_50495.jpg
CR009_02_50496.jpg


CR002_03_50025.jpg
CR002_03_50050.jpg
CR002_03_50190.jpg
CR002_03_50190a.jpg
CR002_03_50193.jpg
CR002_03_50193a.jpg
CR002_03_50310.jpg
CR002_03_50315.jpg
CR002_03_50316a.jpg
CR002_03_50317b.jpg
CR002_03_50323.jpg
CR002_03_50323a.jpg
CR002_03_50324.jpg
CR002_03_50327.jpg
CR002_03_50327a.jpg
CR002_03_50330.jpg
CR002_03_50337.jpg
CR002_03_50337a.jpg
CR002_03_50338.jpg
CR002_03_50339.jpg
CR002_03_50339a.jpg
CR002_03_50340.jpg
CR002_03_50340a.jpg
CR002_03_50341.jpg
CR002_03_50342a.jpg
CR002_03_50345a.jpg
CR002_03_50349.jpg
CR002_03_50349a.jpg
CR002_03_50350.jpg
CR002_03_50351a.jpg
CR002_03_50352.jpg
CR002_03_50354.jpg
CR002_03_50358.jpg
CR002_03_50358a.jpg
CR002_03_50359.jpg
CR002_03_50364.jpg
CR002_03_50364a.jpg
CR002_03_50365.jpg
CR002_03_50366.jpg
CR002_03_50369.jpg
CR002_03_50371.jpg
CR002_03_50372.jpg
CR002_03_50376.jpg
CR002_03_50376a.jpg
CR002_03_50377.jpg
CR002_03_50377a.jpg
CR002_03_50380.jpg
CR002_03_50381.jpg
CR002_03_50387a.jpg
CR002_03_50391.jpg
CR002_03_50393.jpg
CR002_03_50395.jpg
CR002_03_50396.jpg
CR002_03_50397.jpg
CR002_03_50397a.jpg
CR002_03_50397b.jpg
CR002_03_50399.jpg
CR002_03_50399a.jpg
CR002_03_50402.jpg
CR002_03_50405.jpg
CR002_03_50410.jpg
CR002_03_50424.jpg
CR002_03_50428.jpg
CR002_03_50428a.jpg
CR002_03_50429.jpg
CR002_03_50430.jpg
CR002_03_50431.jpg
CR002_03_50431a.jpg
CR002_03_50433.jpg
CR002_03_50433a.jpg
CR002_03_50441a.jpg
CR002_03_50441b.jpg
CR002_03_50442.jpg
CR002_03_50442a.jpg
CR002_03_50444.jpg
CR002_03_50444a.jpg
CR002_03_50446.jpg
CR002_03_50447.jpg


CR009_05_51764.jpg
CR009_05_51765.jpg
CR009_05_51766.jpg
CR009_05_51767.jpg
CR009_05_51768.jpg
CR009_05_51769.jpg
CR009_05_51770.jpg
CR009_05_51771.jpg
CR009_05_51773.jpg
CR009_05_51774.jpg
CR009_05_51775.jpg
CR009_05_51776.jpg
CR009_05_51777.jpg
CR009_05_51778.jpg
CR009_05_51779.jpg
CR009_05_51817.jpg
CR009_05_51818.jpg
CR009_05_51819.jpg
CR009_05_51820.jpg
CR009_05_51821.jpg
CR009_05_51822.jpg
CR009_05_51823.jpg
CR009_05_51824.jpg
CR009_05_51825.jpg
CR009_05_51827.jpg
CR009_05_51828.jpg
CR009_05_51829.jpg
CR009_05_51830.jpg
CR009_05_51831.jpg
CR009_05_51832.jpg
CR009_05_51833.jpg
CR009_05_51834.jpg
CR009_05_51835.jpg
CR009_05_51836.jpg
CR009_05_51838.jpg
CR009_05_51839.jpg
CR009_05_51840.jpg
CR009_05_51841.jpg
CR009_05_51842.jpg
CR009_05_51876.jpg
CR009_05_51877.jpg
CR009_05_51878.jpg
CR009_05_51879.jpg
CR009_05_51880.jpg
CR009_05_51882.jpg
CR009_05_51883.jpg
CR009_05_51884.jpg
CR009_05_51885.jpg
CR009_05_51886.jpg
CR009_05_51887.jpg
CR009_05_51888.jpg
CR009_05_51889.jpg
CR009_05_51890.jpg
CR009_05_51891.jpg
CR009_05_51893.jpg
CR009_05_51894.jpg
CR009_05_51895.jpg
CR009_05_51896.jpg
CR009_05_51897.jpg
CR009_05_51898.jpg
CR009_05_51899.jpg
CR009_05_51900.jpg
CR009_05_51901.jpg
CR009_05_51902.jpg
CR009_05_51904.jpg
CR009_05_51927.jpg
CR009_05_51928.jpg
CR009_05_51929.jpg
CR009_05_51931.jpg
CR009_05_51932.jpg
CR009_05_51933.jpg
CR009_05_51934.jpg
CR009_05_51935.jpg
CR009_05_51936.jpg
CR009_05_51938.jpg
CR009_05_51939.jpg
CR009_05_51940.jpg
CR009_05_51941.jpg
CR009_05_51942.jpg
CR009_05_51943.jpg
CR009_05_51944.jpg
CR009_05_51945.jpg
CR009_05_51946.jpg
CR009_05_51948.jpg
CR009_05_51950.jpg
CR009_05_51951.jpg
CR009_05_51952.jpg
CR009_05_51954.jpg
CR009_05_51955.jpg
CR009_05_51956.jpg
CR009_05_51957.jpg

# Summary

- 식의약용 자생식물 빅데이타 파일의 분석
- 10127개의 이미지화일과 레이블데이타 파일 처리
- 파일 구조 파악, 레이블데이와 매칭 확인
- 정보의 논리적 구조 파악
- Python Program
  - https://github.com/ekyuho/Plant/blob/main/plant_big_data.ipynb