

2025년 서울 민간기업 맞춤형 매력일자리 사업

# JupyterNotebook, Crawling, WordCloud

- 2025.05.08 -

# Crawling + WordCloud

# 크롤링 실습(TEXT)

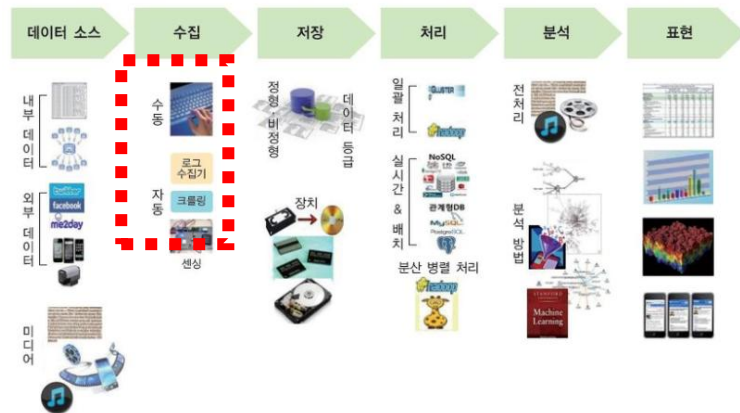
## - 네이버 VIEW -



# 네이버의 지식 IN 크롤링 예제(파이썬)



# 네이버의 지식 IN 크롤링 예제(파이썬)



```

web_crawling.py X
D:\> python_exam > Django > web_crawling.py > ...
1 import requests
2 from bs4 import BeautifulSoup
3
4 url = 'https://kin.naver.com/search/list.nhn?query=%ED%8C%8C%EC%9D%B4%EC%8D%AC'
5
6 response = requests.get(url)
7
8 if response.status_code == 200:
9     html = response.text
10    soup = BeautifulSoup(html, 'html.parser')
11    ul = soup.select_one('ul.basic1')
12    titles = ul.select('li > dl > dt > a')
13    for title in titles:
14        print(title.get_text())
15 else:
16    print(response.status_code)
17
18

```

파이썬 크롤링 프로그램(소스)

```

Anaconda Prompt (anaconda3)
(django) D:\python_exam\ Django > python web_crawling.py
파이썬 기초 질문입니다.
파이썬 인강으로 공부 해보려고해요
파이썬의 한계
파이썬 기초 질문
파이썬 온라인 강의 퀄리티 좋은 곳 알
파이썬으로 어떻게 하냐요?
파이썬 인강 어디서 볼까요?
파이썬 Folium 사용방법 질문
파이썬 문제
파이썬 질문입니다.(채택 바로 해드립니다)
(django) D:\python_exam\ Django >

```

파이썬 크롤링 실행(데이터 취합)



The screenshot shows the Naver Knowledge In search results for the query '파이썬' (Python). The results are listed in a table with columns for question type, date, and content. A red dashed box highlights the search results, and a red arrow points from the Anaconda Prompt output to this box.

네이버 지식인 정보

# 데이터수집 – anaconda 설치

```
# conda create -n p310_crawling python=3.10
```

Python 3.10 설치

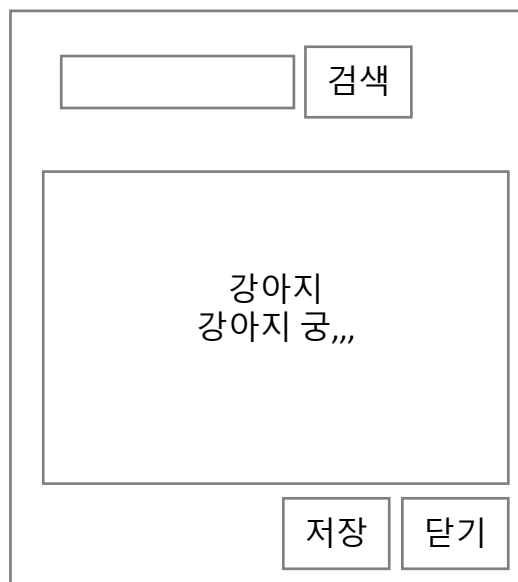
VSCODE에서 실행해보기

```
# conda activate p310_crawling
```

# 데이터수집 - anaconda 설치



```
메인
main.py
ui/ui_main.ui
```



크롤링  
crawling.py  
ui/ui\_crawling.ui



WORDCLOUD  
wordcloud.py  
ui/ui wordcloud.ui

# 데이터수집 - 크롤링 라이브러리 설치

- > pip install beautifulsoup4
- > pip install selenium
- > pip install requests

위의 3가지의 라이브러리를 설치하면 크롤링을 위한 기본 준비완료



# 데이터수집 - 크롤링 : 네이버 VIEW 크롤링

```
import csv
from urllib.request import urlopen
from bs4 import BeautifulSoup as bs
from urllib.parse import quote_plus

# 검색어를 입력한다.
search = input('검색어를 입력하세요 : ')

url = f'https://m.search.naver.com/search.naver?where=m_view
&sm=mtb_jum&query={quote_plus(search)}'
```

# 데이터수집 - 크롤링 : 네이버 VIEW 크롤링

```
html = urlopen(url).read()
soup = bs(html, "html.parser")

total = soup.select('.api_txt_lines.total_tit')
searchList = []

for i in total:
    temp = []
    temp.append(i.text) # 제목
    temp.append(i.attrs['href']) # 링크
    searchList.append(temp)
```

# 데이터수집 - 크롤링 : 네이버 VIEW 크롤링

```
# 엑셀에서 열때 utf-8 표준이면 깨짐현상이 일어남
f = open(f'{search}.csv', 'w', encoding = 'cp949', newline='')
csvWriter = csv.writer(f)
for i in searchList:
    # 한줄씩 써 내려감
    csvWriter.writerow(i)
f.close()

print('완료 되었습니다.')
```

# 크롤링 실습(TEXT)

## - 네이버 지식인 -

# 데이터수집 - 크롤링 : 네이버 지식인 크롤링

```
import csv
from re import search
import requests
from bs4 import BeautifulSoup
from urllib.parse import quote_plus

searchList = []

search = input('검색어를 입력하세요 : ')
```



# 데이터수집 - 크롤링 : 네이버 지식인 크롤링

```
for i in range(1, 101):  
    url = f'https://kin.naver.com/search/list.nhn?query={quote_plus(search)}&pa  
  
    response = requests.get(url)  
  
    if response.status_code == 200:  
        print( "===== " + str(i) + " =====")  
  
        html = response.text  
        soup = BeautifulSoup(html, 'html.parser')
```

# 데이터수집 - 크롤링 : 네이버 지식인 크롤링

```
ul = soup.select_one('ul.basic1')
titles = ul.select('li > dl > dt > a')
for title in titles:
    temp = []
    temp.append(title.text)
    temp.append(title.attrs['href'])
    searchList.append(temp)

    print(title.get_text())
else :
    print(response.status_code)
```

# 데이터수집 - 크롤링 : 네이버 지식인 크롤링

```
### csv 파일로 저장
# newline = " 한줄로 내리기
f = open(f'{search}.csv', 'w', encoding='cp949', newline='')
csvWriter = csv.writer(f)

for i in searchList:
    csvWriter.writerow(i)
f.close()
```

2025년 서울 민간기업 맞춤형 매력일자리 사업

# JupyterNotebook, Crawling, WordCloud

- 2025.05.08 -



# wordcloud



# wordcloud

<https://pypi.org/project/wordcloud/>

## word\_cloud

A little word cloud generator in Python. Read more about it on the [blog post](#) or the [website](#).

The code is tested against Python 2.7, 3.4, 3.5, 3.6 and 3.7.

## Installation

If you are using pip:

```
pip install wordcloud
```

If you are using conda, you can install from the `conda-forge` channel:

```
conda install -c conda-forge wordcloud
```

# 라이브러리 설치

<가상환경>

] 가상환경 계정 : p37\_crawling

] pip install matplotlib

] pip install pandas

] conda install -c conda-forge wordcloud

# wordCloud 실습

# 워드클라우드 실습

## 라이브러리 불러오기

```
In [6]: 1 import pandas as pd
        2 import numpy as np
        3 import re
        4 import matplotlib.pyplot as plt
        5
        6 %matplotlib inline
        7 # 시각화 결과가 선명하게 표시되도록
        8 %config InlineBackend.figure_format = 'retina'
```

## 시각화를 위한 한글폰트 설정

```
In [7]: 1 # Window 한글폰트 설정
        2 plt.rc("font", family="Malgun Gothic")
        3 # Mac 한글폰트 설정
        4 # plt.rc("font", family="AppleGothic")
        5 plt.rc('axes', unicode_minus=False)
```

# 워드클라우드 실습

## 크롤링 파일 로드

```
In [8]: 1 # df_crawl 에 크롤링 파일을 불러온다.  
2 df_crawl = pd.read_csv("cat.csv", encoding='cp949')  
3 df_crawl.shape
```

Out[8]: (1000, 2)

```
In [9]: 1 df_crawl.head()
```

Out[9]:

	text	link
0	고양이모래 고민됩니다	https://kin.naver.com/qna/detail.naver?d1id=3&...
1	고양이 알레르기?	https://kin.naver.com/qna/detail.naver?d1id=7&...
2	고양이에 대해 알려주세요.	https://kin.naver.com/qna/detail.naver?d1id=13...
3	고양이 알러지 증상	https://kin.naver.com/qna/detail.naver?d1id=8&...
4	고양이모래 꼭 필요한가요?	https://kin.naver.com/qna/detail.naver?d1id=11...



# 워드클라우드 실습

```
In [10]: 1 df = df_crawl.drop(['link'], axis='columns')  
        2 df.shape
```

```
Out[10]: (1000, 1)
```

```
In [11]: 1 df.head()
```

```
Out[11]:
```

	text
0	고양이모래 고민됩니다
1	고양이 알레르기?
2	고양이에 대해 알려주세요.
3	고양이 알러지 증상
4	고양이모래 꼭 필요한가요?

# 워드클라우드 실습

```
In [12]: 1 # tail 로 미리보기 합니다.  
        2 df.tail()
```

Out[12]:

	text
995	고양이가 외로운것 같아요
996	수원고양이분양시 유의점?
997	새끼고양이 키우는 법...
998	고양이가 아파요
999	새끼 고양이 키우기

# 워드클라우드 실습

## DataFrame을 리스트로 변환

```
In [13]: 1 list = df['text'].astype(str).tolist()  
        2 list
```

```
'키우는 고양이가 다치는 꿈',  
'뱅갈고양이분양 알아보려면?',  
'고양이 입양하려는데',  
'고양이에 관하여',  
'고양이 질문 100내공',  
'고양이 이사',  
'고양이 꿈해몽',  
'부산고양이분양 알아봐요',  
'고양이 키울때 비용&물품',  
'고양이 마약방식 추천',  
'꿈해몽 고양이꿈',  
'고양이에 대한것',  
'고양이들도 더위를 타나요?',  
'고양이를 키우고 싶은데',  
'고양이 펫로스? 자11111살',  
'고양이에 관하여',  
'고양이질문',  
'새끼 고양이를 데려왔어요',  
'고양이 꼬리 불편할까요',  
'아기고양이가 섬모를 물어오'
```

# 워드클라우드 실습

In [14]:

```
1 ## String.join()으로 리스트의 모든 요소들을 하나의 문자열로 변환
2 word_str = ' '.join(s for s in list)
3 word_str
```

Out[14]:

'고양이모래 고민됩니다 고양이 알레르기? 고양이에 대해 알려주세요. 고양이 알려지 증상 고양이모래 꼭 필요한가요? 고양이에 관하여 고양이 합사 2일차 입니다 고양이 합사 고양이가 죽는 꿈 고양이 간식 x됨 제가 고양이를 불렀어요 저는 고양이가 될거예요!! 고양이 질문 고양이 카페 키우는 고양이가 다치는 꿈 뱅갈고양이분양 알아보려면? 고양이 입양하려는데 고양이에 관하여 고양이 질문 100내공 고양이 이사 고양이 꿈해몽 부산고양이분양 알아봐요 고양이 키울때 비용&물품 고양이 마약방식 추천 꿈해몽 고양이꿈 고양이에 대한것 고양이들도 더위를 타나요? 고양이를 키우고 싶는데 고양이 펫로스? 자11111살 고양이에 관하여 고양이질문 새끼 고양이를 데려왔어요 고양이 꼬리 불편할까요 아기고양이가 성묘를 물어요 고양이 진드기? 벼룩? 고양이키우기 분리불안 고양이? 고양이 서열정리 고양이 합사 아기고양이 처음 키우는 집사 고양이 스트레스 사망 고양이 입양 고양이 강아지가 고양이한테 집착해요 아기 고양이에 대해서 고양이한테 고양이소리 고양이 펜스 설치하고 키우기 고양이 질문 고양이 화장실 가성비 새끼 고양이 키울때 주의할점! 고양이 요로결석 고양이 훈육 이거 맞나요? 모란장 고양이 집주인과 고양이 문제 고양이 초보집사 (삼고양이) 고양이 유해동물 고양이 출산 새끼고양이 고양이에 관하여 고양이 키즈 고양이 잇는 방법 고양이 합사 고양이 외로움 고양이 꿈 해몽 부탁드려요 노묘와 아기고양이 합사 질문 고양이 키우는거 고양이들의 이상행동 고양이꿈 질문 같은 고양이 일까요? 고양이와할아버지꿈. 고양이 구조(급함) 고양이 점프력이 높은 이유 고양이사료 어떻게 있어요? 고양이 모래교체 질문 고양이 알레르기 있는데 고양이들이 안보여요 고양이를 키우고 싶어요. 고양이와 뱀꿈. 고양이한테 물리는꿈 유아 고양이 알레르기 증상 새끼고양이 아가고양이 눈 (내공 100) 꿈해몽 부탁드려요 고양이꿈 고양이간식 성분좋은 제품 고양이가 말하는꿈 고양이가 떠났어요 고양이 알머리컷 가능한가요 고양이 게임 추천 고양이 산책 질문 고양이 두마리 고양이랑강아지 고양이 꿈해몽 부탁드려요 고양이 키우는 문제 새끼 고양이에 대해서 토끼 vs 새끼고양이 고양이 알레르기 고양이 외로움 관련 질문 마눌고양이도 고양이겠조??? 고양이 점프력이 높은 이유 고양이 키우기 설득 고양이 꿈 해몽 고양이 합사 고양이 합사 2일차입니다. 고양이 물품 뭐뭐 사야될까요? 고양이 죽음 고양이사료추천 부탁드려요! 고양이 털갈이 고양이들 끼리 잘 지내는 법 고양이 키우기 고양이가 울어요 고양이간식 제품별 비교 고양이 울음소리 고양이정보좀 고양이의 모든 것! 고양이 음식 마인크래프트 고양이 고양이 키우는 법 고양이 견종 고양이 고양이 합사 새끼 고양이 꿈해몽 고양이 질문 고양이 지내는 법 고양이 뱀을 사야할까요? 고양이

## 불필요한 문자제거

In [15]:

```

1 import re
2
3 def clean_text(inputString):
4     text_rmv = re.sub('[-+=, #/\?:^.@*%\`~·!|\'|\(\)\[\]\`\'...>\"\"\\\'\'\\.']', ' ', inputString)
5     return text_rmv
6
7 word_str = clean_text(word_str)
8 word_str

```

Out[15]: '고양이모래 고민됩니다 고양이 알레르기 고양이에 대해 알려주세요 고양이 알러지 증상 고양이모래 꼭 필요한가요 고양이에 관하여 고양이 합사 2일차 입니다 고양이 합사 고양이가 죽는 꿈 고양이 간식 x됨 제가 고양이를 불렀어요 저는 고양이가 될거 예요 고양이 질문 고양이 카페 키우는 고양이가 다치는 꿈 뱀갈고양이분양 알아보려면 고양이 입양하려는데 고양이에 관하여 고양이 질문 100내공 고양이 이사 고양이 꿈해몽 부산고양이분양 알아봐요 고양이 키울때 비용&물품 고양이 마약방식 추천 꿈해 몽 고양이꿈 고양이에 대한것 고양이들도 더위를 타나요 고양이를 키우고 싶은데 고양이 펫로스 자11111살 고양이에 관하여 고양이질문 새끼 고양이를 데려왔어요 고양이 꼬리 불편할까요 아기고양이가 성묘를 물어요 고양이 진드기 벼룩 고양이키우기 분리불안 고양이 고양이 서열정리 고양이 합사 아기고양이 처음 키우는 집사 고양이 스트레스 사망 고양이 입양 고양이 강아지 가 고양이한테 집착해요 아기 고양이에 대해서 고양이한테 고양이소리 고양이 펜스 설치하고 키우기 고양이 질문 고양이 화장실 가성비 새끼 고양이 키울때 주의할점 고양이 요로결석 고양이 훈육 이거 맞나요 모란장 고양이 집주인과 고양이 문제 고양이 초보집사 삼고양이 고양이 유해동물 고양이 출산 새끼고양이 고양이에 관하여 고양이 키즈 고양이 잇는 방법 고양이 합사 고양 이 외로움 고양이 꿈 해몽 부탁드려요 노묘와 아기고양이 합사 질문 고양이 키우는거 고양이들의 이상행동 고양이꿈 질문 같은 고양이 일까요 고양이와할아버지꿈 고양이 구조 급함 고양이 점프력이 높은 이유 고양이사료 어떻게 있어요 고양이 모래교체 질문 고양이 알레르기 있는데 고양이들이 안보여요 고양이를 키우고 싶어요 고양이와 뱀꿈 고양이한테 물리는꿈 유아 고양이 알레르기 증상 새끼고양이 아가고양이 눈 내공 100 꿈해몽 부탁드려요 고양이꿈 고양이간식 성분좋은 제품 고양이가 말하는꿈 고양이가 떠났어요 고양이 알머리컷 가능한가요 고양이 게임 추천 고양이 산책 질문 고양이 두마리 고양이랑강아지 고양이 꿈해몽 부탁드려요 고양이 키우는 문제 새끼 고양이에 대해서 토끼 vs 새끼고양이 고양이 알레르기 고양이 외로움 관련 질문 마눌고양이 도 고양이겠조 고양이 점프력이 높은 이유 고양이 키우기 설득 고양이 꿈 해몽 고양이 합사 고양이 합사 2일차입니다 고양 이 물품 뭐뭐 사야될까요 고양이 죽음 고양이사료추천 부탁드려요 고양이 털갈이 고양이들 끼리 잘 지내는 법 고양이 키우기 고양이가 울어요 고양이간식 제품별 비교 고양이 울음소리 고양이정보초음 고양이의 모든 것 고양이 음식 마인크래프트 고양이 고 양이 키우는 연필의 결속 고양이 고양이 합사 새끼 고양이 꿈해몽 고양이 고양이 질문 고양이 장에비 신발 고양이알고 고양이 고양이



# 워드클라우드 실습

## wordCloud 사용하기

In [19]:

```
1 stopwords = ['고양이', '고양이를', '질문', '고양이가']
2 # displayWordCloud 라는 함수를 만들어 재사용합니다.
3 from wordcloud import WordCloud
4
5 wc1 = WordCloud(
6     font_path = '/System/Library/Fonts/AppleSDGothicNeo.ttc',
7     # font_path=r"C:\Windows\Fonts\malgun.ttf",
8     stopwords = stopwords,
9     background_color = 'white',
10    width = 1024, height = 768,
11    random_state=42)
12
13 wc1.generate(word_str)
14
15 plt.figure(figsize = (15 , 10))
16 plt.imshow(wc1)
17 plt.axis("off")
18 plt.show()
```

# 워드클라우드 실습

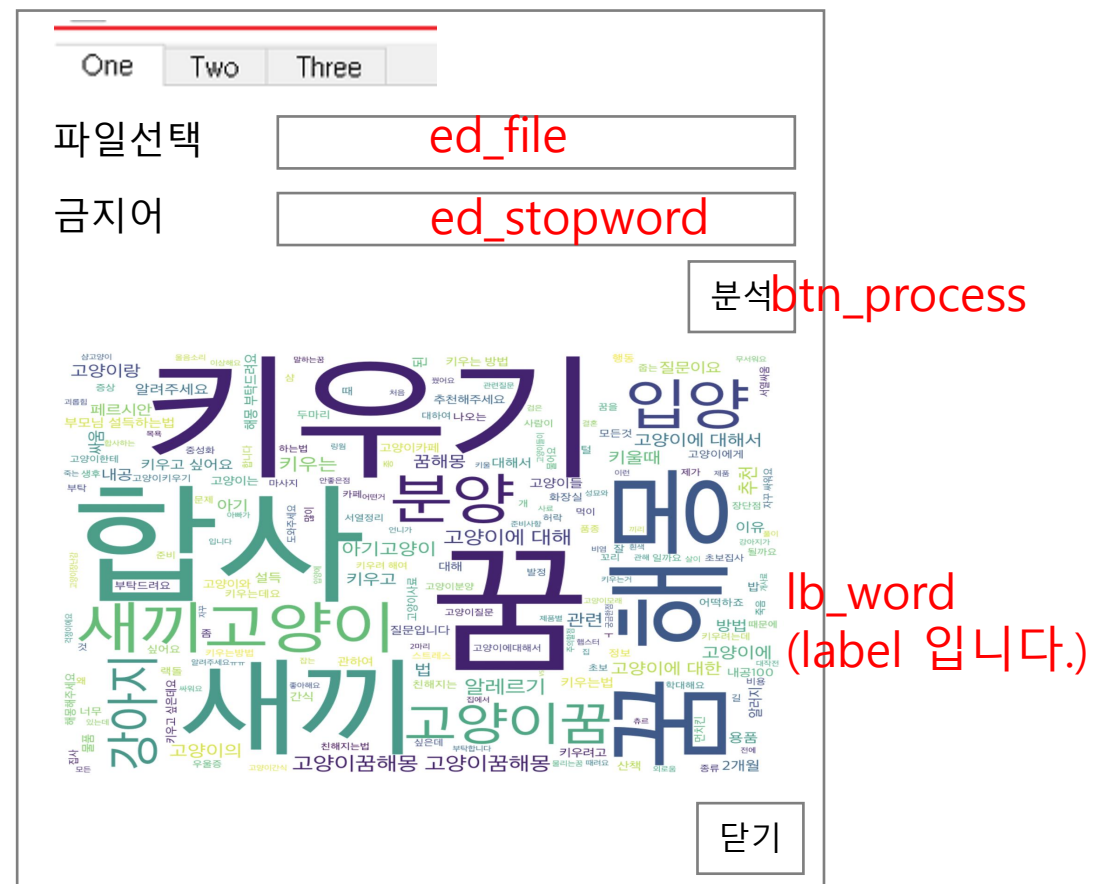
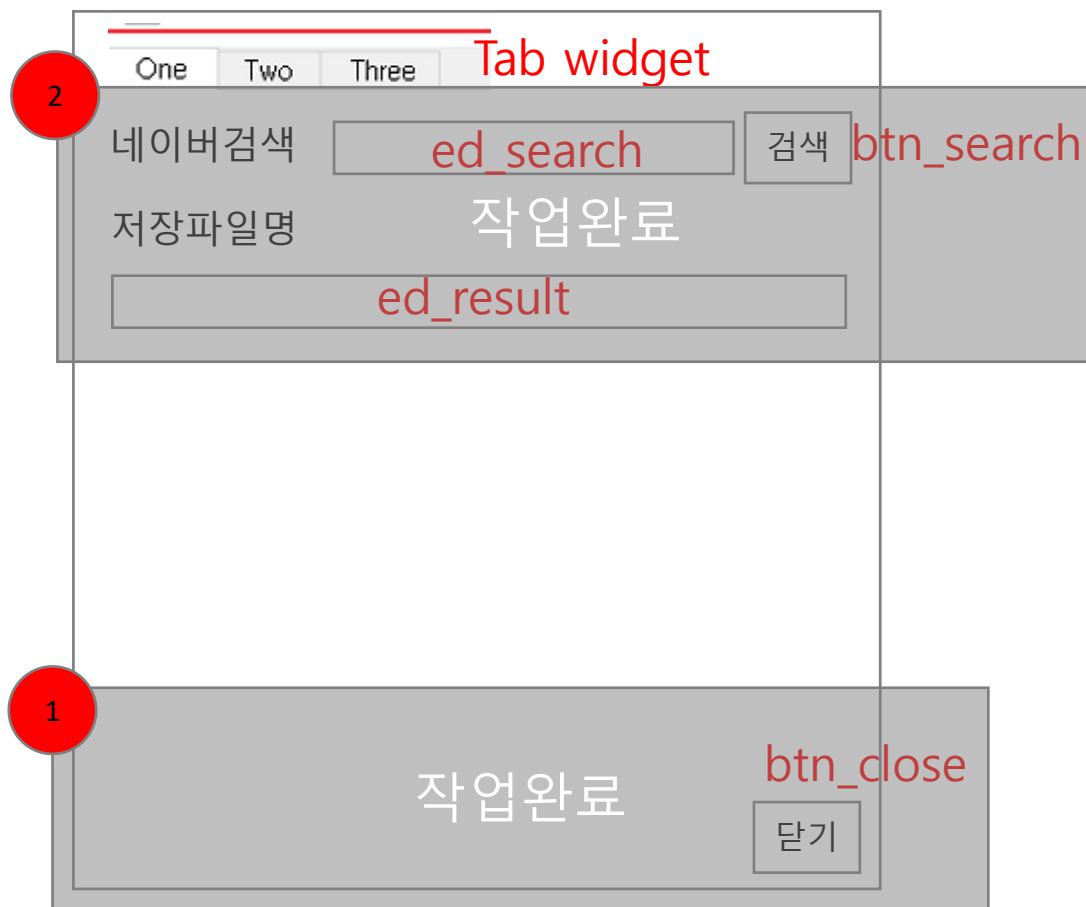
```
In [22]: 1 wcl.to_file('cloud.png')
```

```
Out[22]: <wordcloud.wordcloud.WordCloud at 0x145c738e0>
```

```
In [ ]: 1
```



# 크롤링 프로그램 완성



```
wordcloud_main.py
ui/wordcloud_main.ui
```



2025년 서울 민간기업 맞춤형 매력일자리 사업

# JupyterNotebook, Crawling, WordCloud

- 2025.05.08 -