

In [1]:

```
import os
os.environ['JAVA_HOME'] = r'C:\Program Files\Java\jdk-18.0.2\bin\server'
print(os.environ.get('JAVA_HOME'))
#자바환경변수 설정
```

C:\Program Files\Java\jdk-18.0.2\bin\server

In [2]:

```
pip install --upgrade pip
```

Requirement already satisfied: pip in c:\Users\User\Anaconda3\lib\site-packages (23.0)

Note: you may need to restart the kernel to use updated packages.

In [3]:

```
pip install JPy1-1.4.0-cp310-cp310-win_amd64.whl
```

Note: you may need to restart the kernel to use updated packages.

WARNING: Requirement 'JPy1\2011.4.0\2011cp310\2011cp310\2011win_amd64.whl' looks like a filename, but the file does not exist
ERROR: JPy1\2011.4.0\2011cp310\2011cp310\2011win_amd64.whl is not a valid wheel filename.

In [4]:

```
pip install mecab_python-0.996_ko_0.9.2_msvc-cp37-cp37m-win_amd64.whl
```

Note: you may need to restart the kernel to use updated packages.

WARNING: Requirement 'mecab_python-0.996_ko_0.9.2_msvc-cp37-cp37m-win_amd64.whl' looks like a filename, but the file does not exist
ERROR: mecab_python-0.996_ko_0.9.2_msvc-cp37-cp37m-win_amd64.whl is not a supported wheel on this platform.

In [5]:

```
pip install customized_konlpy #패키지 설치
```

Note: you may need to restart the kernel to use updated packages.

ERROR: Invalid requirement: '#패키지'

In [6]:

```
pip install PyKomoran
```

Requirement already satisfied: PyKomoran in c:\Users\User\Anaconda3\lib\site-packages (0.1.6.post1)

Requirement already satisfied: py4j==0.10.9.2 in c:\Users\User\Anaconda3\lib\site-packages (from PyKomoran) (0.10.9.2)

Note: you may need to restart the kernel to use updated packages.

In [7]:

```
import ckonlpy.tag
from ckonlpy.tag import *

twitter=Twitter()

twitter.add_dictionary(['지원서', '봉사단'], 'Noun') #텍스트 분석할 때 특정 단어 리스트 추가하기

import re
import networkx as nx
import pandas as pd
import numpy as np
import datetime
from collections import Counter
from datetime import datetime
import datetime
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

C:\Users\Wuser\Anaconda3\lib\site-packages\Wkonlpy\Wtag\W_0kt.py:17: UserWarning: "Twitter" has changed to "0kt" since KoNLPy v0.4.5.
warn('"Twitter" has changed to "0kt" since KoNLPy v0.4.5.')

In [8]:

```
df = pd.read_csv('2020년 키워드 분석 프로그램 샘플중복제거.csv', encoding='cp949')
# csv 파일 불러와서 df로 저장
df
```

Out[8]:

| 대분류 | 대응부서 | 접수채널 | 답변부 세내용 | 문의세 부내용 | 일 (Day) | 월 (Month) | 연 (Year) | 텍스트 병합 소거 | 텍스트2 | 텍스트1 | 고유화 시리얼 | Unnamed:0 |
|-----|----------|------|------------|---|------------|--------------|-------------|-----------------|------|------|------------|-----------|
| 0 | 글로벌인재총괄실 | 전화 | 0 | Q. 해양수산부 / 해외파견 사업을 진행하는데 되어, 코이카 봉사단의 경우 어떻게 진... | 31.0 | 7.0 | 2020.0 | Q | NaN | Q. | 2A | 0 |
| 1 | 글로벌인재총괄실 | 전화 | 0 | Q. 수료, 졸업예정인 경우 지원가능여부의 문의 \\nA. 졸업예정자 지원가능하 나... | 26.0 | 8.0 | 2020.0 | Q | NaN | Q. | 3A | 20 |
| 2 | 글로벌인재총괄실 | 전화 | 0 | Q. UNV 단원으로 활동중이며 유수연 직원 연락처 문의 \\nA. 코이카 글로벌인재 기... | 27.0 | 8.0 | 2020.0 | Q | NaN | Q. | 4A | 35 |

| 대분류 | 대응부서 | 접수채널 | 답변세부내용 | 문의세부내용 | 일 (Day) | 월 (Month) | 연 (Year) | 텍스트병합소거 | 텍스트2 | 텍스트1 | 고유화시리얼 | Unnamed: 0 | |
|-------------------|-------------------|------------|-----------------------|--|---------|-----------|----------|---------|------|-------|--------|------------|------|
| 0 | 글인재본부 - 글로벌 인재총괄실 | 전화 | 0 | Q. 지원 국가 선택가능 여부, 생활비/급여 지원 문의\r\nA. 수요 내에서 선... | 28.0 | 8.0 | 2020.0 | Q | NaN | Q. | 5A | 45 | 3 |
| 0 | 본부 - 조달계약 1팀 | 전화 | 14년도 가입여유 중 협업체 임을 안내 | 협업체 등 되는지 확인차 문의 | 11.0 | 6.0 | 2020.0 | 협업체 | NaN | 협업체 | 550A | 59 | 4 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 개발협력종은 일자리 (개인고객) | 글인재본부 - 해외봉사활동관리팀 | 민원 (국민신문고) | ○ 증명서 신청차 완료 | ○ 증명요청\r\n○ 증명서 | 31.0 | 7.0 | 2020.0 | 경력증명서 | NaN | 경력증명서 | 13693B | 117374 | 2938 |

| Unnamed: 0 | 고유화 시리얼 | 텍스트 1 | 텍스트 2 | 텍스트 병합 소거 | 연 (Year) | 월 (Month) | 일 (Day) | 문의세 부내용 | 변 세 내 용 | 접 수 채 널 | 대 응 부 서 | 대 분 류 |
|------------|------------|----------|----------------------------|-----------------|-------------|--------------|------------|---|---|----------------------------|---|---|
| 2939 | 117377 | 13694B | 해 외 지 원 사 업 | NaN | 2020.0 | 8.0 | 2.0 | <ul style="list-style-type: none"> ○ KOICA 해외사업 진행문의 | <ul style="list-style-type: none"> ○ 이코나19 대응을 위한 별 력 업 전 및 사 관 리 지 침 | 민원 (국 민 신 문 고) | 본 부 「 국 별 사 업 총 괄 팀 | 코 로 나 19 고 충 처 리 |
| 2940 | 117401 | 13695B | 대 중 국 | NaN | 2020.0 | 8.0 | 4.0 | <ul style="list-style-type: none"> ○ 대 중 국 코 로 나 구 호 물 품 원 과 관 련 여 가 요 청 | <ul style="list-style-type: none"> ○ 화 선 적 류 및 수 검 보 고 는 인 의 영 업 상 비 밀 및 개 인 정 보 | 민원 (국 민 신 문 고) | 본 부 「 다 자 협 력 인 도 지 원 실 | 코 로 나 19 고 충 처 리 |
| 2941 | 117424 | 13696B | 단 원 활 동 | NaN | 2020.0 | 8.0 | 10.0 | <ul style="list-style-type: none"> ○ 단 원 활 동 평 가 점 수 문 의 | <ul style="list-style-type: none"> ○ 평 가 점 수 공 개 절 차 내 외 완 료 | 민원 (국 민 신 문 고) | 글 인 재 본 부 「 해 외 봉 사 활 동 관 리 팀 | 개 발 협 력 중 은 일 자 리 (개 인 고 객) |

| Unnamed: 0 | 고유화 시리얼 | 텍스트 1 | 텍스트 2 | 텍스트 병합 소거 | 연 (Year) | 월 (Month) | 일 (Day) | 문의세 부내용 | 답변 세부 내용 | 접수 채널 | 대응 부서 | 대분류 |
|------------|------------|----------|----------|-----------------|-------------|--------------|------------|------------|--|---|-------------------------------|--|
| 2942 | 117429 | 13697B | 3차 | NaN | 3차 | 2020.0 | 8.0 | 25.0 | ○ 3차 국제개 발협력 기본계 획에 대하여 국제개 발협력 시민사 회포럼 와 국 제개발 \n ... | ○ 의 견 서 는 3차 기본 계획 에 반영 되고 실현 될 수 있 도 록 국 무 조 정 실 차 원 의 3 차 \n... | 민원 (국 민 신 문 고) | 본 부 「 기 획 조 정 실 개 발 협 력 사 업 상 담 (기 관 고 객) |

2943 rows × 16 columns

In []:

In [9]:

```
dff = df[['문의세부내용', '답변세부내용', '대응부서']].fillna("") # 해당 컬럼만 추출하여 dff로 저장 ,
dff
```

Out[9]:

| | 문의세부내용 | 답변세부내용 | 대응부서 |
|------|---|--|-------------------------|
| 0 | Q. 해양수산부 / 해외파견 사업을 진행학 데 되어, 코이카 봉사단의 경우 어떻게 진... | 0 | 글인재본부_글 로별인재총괄 실 |
| 1 | Q. 수료, 졸업예정인 경우 지원 가능여부 문의\r\nA. 졸업예정자 지원 가능하나... | 0 | 글인재본부_글 로별인재총괄 실 |
| 2 | Q. UNV 단원으로 활동중이며 유수연 직 원 연락처 문의\r\nA. 코이카 글로벌인재 기... | 0 | 글인재본부_글 로별인재총괄 실 |
| 3 | Q. 지원 시 국가 선택가능여부, 생활비/급 여 지원 문의\r\nA. 수요 내에서 선... | 0 | 글인재본부_글 로별인재총괄 실 |
| 4 | 협력업체 등록 되어있는지 확인차 문의 | 14년도 가입하여 이용중인 협력업체임을 안내 | 본부_조달계약 1팀 |
| ... | ... | ... | ... |
| 2938 | ○ 경력증명서 요청 | ○ 증명서 신청절차 안내 완료 | 글인재본부_해 외봉사활동관 리팀 |
| 2939 | ○ KOICA 해외지원사업 진행 관련 문의 | ○ 코이카는 ‘코로나19 대응을 위한 국별 협력사업 안전 및 사업관리 지침’을 \n ... | 본부_국별사업 총괄팀 |
| 2940 | ○ 대중국 코로나 구호물품 지원과 관련 하여 추가 자료 요청 | ○ 화물선적서류 및 검수보고서는 법인 의 영업상 비밀 및 개인정보\r\n(단가, ... | 본부_다자협력 인도지원실 |
| 2941 | ○ 단원활동 평가 점수 문의 | ○ 평가점수 공개절차 안내 완료 | 글인재본부_해 외봉사활동관 리팀 |
| 2942 | ○ 3차 국제개발협력 기본계획에 대하여 국제개발협력시민사회포럼과 국제개발\r\n... | ○ 의견서는 3차 기본계획에 반영되고 실 현될 수 있도록 국무조정실 차원의 3차 \n... | 본부_기획조정 실 |

2943 rows × 3 columns

In [10]:

```
dff.sort_values(by=['대응부서'], ascending=False)
# 대응부서 별 내림차순 정렬,
#기본(ascending=True)은 오름차순 ascending
#안 적으면 기본으로 오름차순으로 정렬
```

Out[10]:

| | 문의세부내용 | 답변세부내용 | 대응부서 |
|------|---|---|----------------------|
| 2507 | 코로나 사태로 현지와 본부 모두 많은 변동 과 긴급상황에 노고가 많으십니다. 건강하 시... | 0 | 지역실_중 동북아프리 카실 |
| 208 | "이라크 쿠르드지방정부 공공조달 현대화 사업 PMC용역"사업의 공식하도급 업체인 아이... | 안녕하십니까? 코로나 사태로 여러모로 어 려움이 많으실 것이라 생각합니다.?귀하께 서 ... | 지역실_중 동북아프리 카실 |
| 981 | 1. 행정처리의 간소화 / 2. 심사위원 선발 등 전문성 고려 (특히 현장 경험의 ... | 0 | 지역실_중 동북아프리 카실 |
| 2865 | 현지사무소의 위기관리 능력이 우수하다고 판단됨 | 0 | 지역실_중 동북아프리 카실 |
| 2139 | 예상치 못한 상황발생 시 (원칙보단) 유연 하고 신속한 대처를 우선으로 고려해 주기 바람 | 0 | 지역실_중 동북아프리 카실 |
| ... | ... | ... | ... |
| 2496 | 커뮤니케이션을 원활하게 하기 위해 봉사 자들이 언어를 배울 수 있도록 지원 부탁드 려요. | 0 | WFK(솔로 문군도) |
| 1797 | 모든 봉사자들은 영어로 말하고 쓸 줄 알아 야 합니다. | 0 | WFK(솔로 문군도) |
| 2160 | 우리 학교로 파견을 나온 봉사자들은 무척 도움이 되었습니다. | 0 | WFK(부탄) |
| 1337 | KOICA의 서비스는 무척 효율적이면서 효 과적이었습니다. 서비스 제공자들은 전문 적이... | 0 | WFK(부탄) |
| 1561 | 국가 사무소는 더 강화되어야 하며 국가 상 황/필요를 고려한 후에 지원 분야를 확대 해... | 0 | WFK(부탄) |

2943 rows × 3 columns

In [11]:

```
dff['대응부서'].value_counts() # 개수 카운트
```

Out[11]:

```
글인재본부_해외봉사활동관리팀    649
본부_연수사업실                    246
본부_ODA교육팀                    193
본부_국민소통센터                178
글인재본부_해외봉사모집팀          149
...
사무소(부산)                      2
본부_기후감염병대응팀              1
사무소(알제리)                    1
본부_감사실                        1
본부_기획조정실                    1
Name: 대응부서, Length: 83, dtype: int64
```

In [12]:

```
stopwords=pd.read_csv('C:/Users/user/코이카_불용어.csv',encoding='cp949')
stop_words_list=[i for i in stopwords['문의']] # 리스트 불러오기
len(stop_words_list) # 길이
```

Out[12]:

530

In [13]:

```
stopwords=pd.read_csv('코이카_불용어.csv', encoding='cp949', header=None)
stopwords
```

Out[13]:

| | 0 |
|-----|-----|
| 0 | 문의 |
| 1 | 질의 |
| 2 | 민원 |
| 3 | 또한 |
| 4 | 다시 |
| ... | ... |
| 526 | 라며 |
| 527 | 주신 |
| 528 | 먼저 |
| 529 | 다만 |
| 530 | 국민 |

531 rows × 1 columns

In [14]:

```
minwon_contents = dff.get('문의세부내용') # get 해당 컬럼 추출
```

In [15]:

```
minwon_answers=dff.get('답변세부내용')
```

In [16]:

```
minwon_contents
```

Out[16]:

```
0      Q. 해양수산부 / 해외파견 사업을 진행학데 되어, 코이카 봉사단의 경우 어떻게
진...
1      Q. 수수료, 졸업예정인 경우 지원 가능여부 문의WrWnA. 졸업예정자 지원 가능하
나...
2      Q. UNV 단원으로 활동중이며 유수연 직원 연락처 문의WnA. 코이카 글로벌인재
기...
3      Q. 지원 시 국가 선택가능여부, 생활비/급여 지원 문의WrWnA. 소요 내에서 선...
4      협력업체 등록 되어있는지 확인차 문의
```

```
2938      ...      O 경력증명서 요청
2939      O KOICA 해외지원사업 진행 관련 문의
2940      O 대중국 코로나 구호물품 지원과 관련하여 추가 자료 요청
2941      O 단원활동 평가 점수 문의
2942      O 3차 국제개발협력 기본계획에 대하여 국제개발협력시민사회포럼와 국제개발Wn
```

```
...
Name: 문의세부내용, Length: 2943, dtype: object
```

In [17]:

```
minwon_answers
```

Out[17]:

```
0      0
1      0
2      0
3      0
4      14년도 가입하여 이용중인 협력업체임을 안내
...
2938      O 증명서 신청절차 안내 완료
2939      O 코이카는 ‘코로나19 대응을 위한 국별협력사업 안전 및 사업관리 지침’을 Wn
...
2940      O 화물선적서류 및 검수보고서는 법인의 영업상 비밀 및 개인정보Wn (단가,
...
2941      O 평가점수 공개절차 안내 완료
2942      O 의견서는 3차 기본계획에 반영되고 실현될 수 있도록 국무조정실 차원의 3차
Wn...
```

```
Name: 답변세부내용, Length: 2943, dtype: object
```

In [18]:

```
minwon=pd.concat([minwon_contents,minwon_answers],axis=0) #concat 합치기
```

In [19]:

```
minwon
```

Out[19]:

```
0      Q. 해양수산부 / 해외파견 사업을 진행학데 되어, 코이카 봉사단의 경우 어떻게
진...
1      Q. 수료, 졸업예정인 경우 지원 가능여부 문의WrWnA. 졸업예정자 지원 가능하
나...
2      Q. UNV 단원으로 활동중이며 유수연 직원 연락처 문의WnA. 코이카 글로벌인재
기...
3      Q. 지원 시 국가 선택가능여부, 생활비/급여 지원 문의WrWnA. 수요 내에서 선...
4      협력업체 등록 되어있는지 확인차 문의

...
2938      ○ 증명서 신청절차 안내 완료
2939      ○ 코이카는 ‘코로나19 대응을 위한 국별협력사업 안전 및 사업관리 지침’을 Wn
...
2940      ○ 화물선적서류 및 검수보고서는 법인의 영업상 비밀 및 개인정보Wn (단가,
...
2941      ○ 평가점수 공개절차 안내 완료
2942      ○ 의견서는 3차 기본계획에 반영되고 실현될 수 있도록 국무조정실 차원의 3차
Wn...
Length: 5886, dtype: object
```

In [20]:

```
start_time = datetime.datetime.now() #시간 측정 시작

# 형태소 분석
from re import match # 전처리 위해서 정규표현식 관련 re 패키지 import

result=[] #result 리스트에 담기

for sent in minwon_contents:
    nouns = []
    for noun in twitter.nouns(sent):
        # 단어 전처리 : 2음절 이상, 수사 제외
        if len(str(noun)) >= 2 and not(match('^[0-9]', noun)):
            nouns.append(noun) # append() 추가하는 것 즉 nouns 리스트에 명사 추가
    result.append(list(set(nouns)))

result

result1=sum(result, []) #[]안의 리스트 합치기
result1

nouns2=[]
for noun in result1:
    # 불용어 제외하기
    if str(noun) not in stop_words_list:
        nouns2.append(noun)

end_time = datetime.datetime.now() #시간 측정 끝
elapsed_time = end_time - start_time #시간 측정
elapsed_time
```

Out[20]:

```
datetime.timedelta(seconds=32, microseconds=959188)
```

In [21]:

```
nouns2
```

Out[21]:

```
['봉사단',  
'선발',  
'계약',  
'신규',  
'전형',  
'중단',  
'파견',  
'해양수산부',  
'대학생',  
'단원',  
'형태',  
'태임',  
'잠정',  
'일시',  
'귀국',  
'전문가',  
'파견',  
'지원']
```

In [22]:

```
from collections import Counter  
num=10  
nouns_counter = Counter(nouns2) # 카운트하기  
minwon_top_nouns=dict(nouns_counter.most_common(num)) # num값 만큼 추출 개수를 내림차순 상위 30개  
minwon_top_nouns
```

Out[22]:

```
{'지원': 521,  
'단원': 416,  
'파견': 369,  
'코로나': 357,  
'봉사단': 343,  
'교육': 333,  
'귀국': 329,  
'프로그램': 315,  
'직원': 238,  
'제공': 222}
```

In [23]:

```
pd.DataFrame([minwon_top_nouns]) #데이터 프레임으로 만든다
```

Out[23]:

| | 지원 | 단원 | 파견 | 코로나 | 봉사단 | 교육 | 귀국 | 프로그램 | 직원 | 제공 |
|---|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|
| 0 | 521 | 416 | 369 | 357 | 343 | 333 | 329 | 315 | 238 | 222 |

In [24]:

```
df2=pd.DataFrame([nouns_counter])
df2.rename(index={0: "20년"}, inplace=True) # 첫 행 이름을 2021년으로 설정하기
df3=df2.sort_values(by="20년", axis=1, ascending=False) # 2021년 내림차순으로 정렬

df3.T[:10]
```

Out [24]:

| | 20년 |
|------|-----|
| 지원 | 521 |
| 단원 | 416 |
| 파견 | 369 |
| 코로나 | 357 |
| 봉사단 | 343 |
| 교육 | 333 |
| 귀국 | 329 |
| 프로그램 | 315 |
| 직원 | 238 |
| 제공 | 222 |

In [25]:

```
df2.to_csv('2020년 키워드 분석 프로그램 샘플 빈도수 최종.csv',encoding='cp949') # csv파일로 저장
```

In [26]:

```
df_2020 = pd.read_csv('2020년 키워드 분석 프로그램 샘플 빈도수 최종.csv', encoding='cp949')
df_2021 = pd.read_csv('2021년 키워드 분석 프로그램 샘플 빈도수 최종.csv', encoding='cp949')
df_2022 = pd.read_csv('2022년 키워드 분석 프로그램 샘플 빈도수 최종.csv', encoding='cp949')

print(df_2020)
print('-'*120)
print(df_2021)
print('-'*120)
print(df_2022)
```

```
Unnamed: 0    봉사단    선발    계약    신규    전형    중단    파견    해양수산부    대학생    ...
이탈리아    연합체    소문    W
0          20년    343    76    130    24    12    29    369          1    19    ...          1    1    1
```

```
매출    난민캠프    로힝아    바다    영유아보육법    원아    의견서
0    1          1    1    1          1    1    1
```

[1 rows x 7513 columns]

```
Unnamed: 0    지원    제안    서류    증명서    프로그램    교육    참여    면접    일정    ...
폴무원    한문    탁틴    W
0          21년    512    397    361    320    304    292    284    275    274    ...          1    1    1
```

```
격무    도아    력입니    구청    이호    호직    중일
0    1    1    1    1    1    1    1
```

[1 rows x 6264 columns]

```
Unnamed: 0    지원    봉사단    교육    증명서    서류    모집    발급    파견    면접    ...
박기환    국임    민협    W
0          22년    1151    846    516    497    420    417    416    410    374    ...          1    1    1
```

```
임금    강도    컨설팅비    상하좌우    윤수    사결    교보
0    1    1          1          1    1    1    1
```

[1 rows x 6897 columns]

In [27]:

df_2020

Out[27]:

```

      Unnamed: 0    봉사단    선발    계약    신규    전형    중단    파견    해양수산부    대학생    ...    이탈리아    연합체    소문    매출    난민캠프    로힝아    바다    영유아보육법    원아    의견서
0          20년    343    76    130    24    12    29    369          1    19    ...          1    1    1    1    1    1    1    1    1
```

1 rows x 7513 columns

In [28]:

```
df_2021
```

Out[28]:

| Unnamed: 0 | 지원 | 제안 | 서류 | 증명서 | 프로그램 | 교육 | 참여 | 면접 | 일정 | ... | 풀무원 | 한문 | 탁틴 | 격무 | 도아 | 력입니 | 구청 | 이호 |
|------------|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|----|----|----|----|-----|----|----|
| 0 | 21년 | 512 | 397 | 361 | 320 | 304 | 292 | 284 | 275 | 274 | ... | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

1 rows × 6264 columns

In [29]:

```
df_2022
```

Out[29]:

| Unnamed: 0 | 지원 | 봉사단 | 교육 | 증명서 | 서류 | 모집 | 발급 | 파견 | 면접 | ... | 박기환 | 국임 | 민협 | 임금 | 강도 | 컨설팅비 | 상하좌우 | 윤수 |
|------------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|----|------|------|----|
| 0 | 22년 | 1151 | 846 | 516 | 497 | 420 | 417 | 416 | 410 | 374 | ... | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

1 rows × 6897 columns

In [30]:

```
dff_2020=df_2020.set_index("Unnamed: 0")# 인덱스를 첫 컬럼으로 설정
dff_2021=df_2021.set_index("Unnamed: 0")
dff_2022=df_2022.set_index("Unnamed: 0")
```

In [31]:

```
dff_2020
```

Out[31]:

| Unnamed: 0 | 봉사단 | 선발 | 계약 | 신규 | 전형 | 중단 | 파견 | 해양수산부 | 대학생 | 단원 | ... | 이탈리아 | 연합체 | 소문 | 매출 | 난민캠프 | 로힝야 | 바다 | 영유아보육법 | 원아 |
|------------|-----|----|-----|----|----|----|-----|-------|-----|-----|-----|------|-----|----|----|------|-----|----|--------|----|
| 20년 | 343 | 76 | 130 | 24 | 12 | 29 | 369 | 1 | 19 | 416 | ... | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

1 rows × 7512 columns

In [32]:

```
dff_2021
```

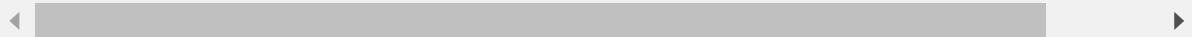
Out[32]:

| | 지원 | 제안 | 서류 | 증명서 | 프로그램 | 교육 | 참여 | 면접 | 일정 | 코로나 | ... | 풀무원 | 한문 | 탁틴 | 격무 | 도아 | 력인니 | 구청 |
|--|----|----|----|-----|------|----|----|----|----|-----|-----|-----|----|----|----|----|-----|----|
|--|----|----|----|-----|------|----|----|----|----|-----|-----|-----|----|----|----|----|-----|----|

Unnamed:
0

| | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|---|---|---|---|---|---|
| 21년 | 512 | 397 | 361 | 320 | 304 | 292 | 284 | 275 | 274 | 270 | ... | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|---|---|---|---|---|---|

1 rows × 6263 columns



In [33]:

```
dff_2022
```

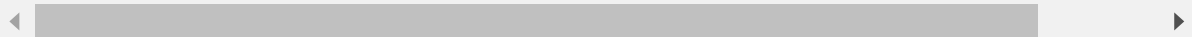
Out[33]:

| | 지원 | 봉사단 | 교육 | 증명서 | 서류 | 모집 | 발급 | 파견 | 면접 | 일정 | ... | 박기환 | 국임 | 민협 | 임금 | 강도 | 컨설팅비 | 상하좌우 |
|--|----|-----|----|-----|----|----|----|----|----|----|-----|-----|----|----|----|----|------|------|
|--|----|-----|----|-----|----|----|----|----|----|----|-----|-----|----|----|----|----|------|------|

Unnamed:
0

| | | | | | | | | | | | | | | | | | | |
|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|---|---|---|---|---|---|
| 22년 | 1151 | 846 | 516 | 497 | 420 | 417 | 416 | 410 | 374 | 364 | ... | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|---|---|---|---|---|---|

1 rows × 6896 columns



In [34]:

```
dff_add=pd.concat([dff_2020, dff_2021]) # 두 데이터프레임 합치기  
dff_add1=pd.concat([dff_2021, dff_2022])
```

In [35]:

```
dft=dff_add.T # 행과 열을 바꾸기  
dft1=dff_add1.T
```


In [36]:

```
dft
```

Out[36]:

| Unnamed: 0 | 20년 | 21년 |
|------------|-------|-------|
| 봉사단 | 343.0 | 235.0 |
| 선발 | 76.0 | 120.0 |
| 계약 | 130.0 | 98.0 |
| 신규 | 24.0 | 70.0 |
| 전형 | 12.0 | 89.0 |
| ... | ... | ... |
| 탁틴 | NaN | 1.0 |
| 력입니 | NaN | 1.0 |
| 이호 | NaN | 1.0 |
| 호직 | NaN | 1.0 |
| 중일 | NaN | 1.0 |

9812 rows × 2 columns

In [37]:

```
dft1
```

Out[37]:

| Unnamed: 0 | 21년 | 22년 |
|------------|-------|--------|
| 지원 | 512.0 | 1151.0 |
| 제안 | 397.0 | 315.0 |
| 서류 | 361.0 | 420.0 |
| 증명서 | 320.0 | 497.0 |
| 프로그램 | 304.0 | 162.0 |
| ... | ... | ... |
| 국임 | NaN | 1.0 |
| 컨설팅비 | NaN | 1.0 |
| 상하좌우 | NaN | 1.0 |
| 윤수 | NaN | 1.0 |
| 교보 | NaN | 1.0 |

9403 rows × 2 columns

In [38]:

```
dft_sum=dft.sum(axis='columns') # 컬럼 합계
dft_sum.name='총합계' # 그 이름을 총합계로 설정
dft_vf=pd.concat([dft['20년'], dft['21년'], dft_sum], axis=1)
# dft 파일의 2020년, 2021년 컬럼과 총합계 컬럼을 한 파일로 병합
dft_result=dft_vf.sort_values('총합계', ascending=False)
# 총합계 컬럼 내림차순으로 정렬
dft_result
```

Out[38]:

| | 20년 | 21년 | 총합계 |
|------|-------|-------|--------|
| 지원 | 521.0 | 512.0 | 1033.0 |
| 코로나 | 357.0 | 270.0 | 627.0 |
| 교육 | 333.0 | 292.0 | 625.0 |
| 프로그램 | 315.0 | 304.0 | 619.0 |
| 파견 | 369.0 | 239.0 | 608.0 |
| ... | ... | ... | ... |
| 신건 | 1.0 | NaN | 1.0 |
| 허물 | 1.0 | NaN | 1.0 |
| 난관 | 1.0 | NaN | 1.0 |
| 모되 | 1.0 | NaN | 1.0 |
| 중일 | NaN | 1.0 | 1.0 |

9812 rows × 3 columns

In [39]:

```
dft1_sum=dft.sum(axis='columns') # 컬럼 합계
dft1_sum.name='총합계' # 그 이름을 총합계로 설정
dft1_vf=pd.concat([dft1['21년'], dft1['22년'], dft_sum], axis=1)
# dft 파일의 2020년, 2021년 컬럼과 총합계 컬럼을 한 파일로 병합
dft1_result=dft1_vf.sort_values('총합계', ascending=False)
# 총합계 컬럼 내림차순으로 정렬
dft1_result
```

Out[39]:

| | 21년 | 22년 | 총합계 |
|------|-------|--------|--------|
| 지원 | 512.0 | 1151.0 | 1033.0 |
| 코로나 | 270.0 | 220.0 | 627.0 |
| 교육 | 292.0 | 516.0 | 625.0 |
| 프로그램 | 304.0 | 162.0 | 619.0 |
| 파견 | 239.0 | 410.0 | 608.0 |
| ... | ... | ... | ... |
| 박기환 | NaN | 1.0 | NaN |
| 국임 | NaN | 1.0 | NaN |
| 컨설팅비 | NaN | 1.0 | NaN |
| 상하좌우 | NaN | 1.0 | NaN |
| 윤수 | NaN | 1.0 | NaN |

12033 rows × 3 columns

In [40]:

```
dft_result.to_csv('20,21년 키워드 분석 프로그램 샘플 빈도수 최종.csv',encoding='cp949')# 저장
dft1_result.to_csv('21,22년 키워드 분석 프로그램 샘플 빈도수 최종.csv',encoding='cp949')
```

In [41]:

```
dft_30=dft_result.head(30) # 상위 30개 추출  
dft_30
```

Out[41]:

| | 20년 | 21년 | 총합계 |
|------|-------|-------|--------|
| 지원 | 521.0 | 512.0 | 1033.0 |
| 코로나 | 357.0 | 270.0 | 627.0 |
| 교육 | 333.0 | 292.0 | 625.0 |
| 프로그램 | 315.0 | 304.0 | 619.0 |
| 파견 | 369.0 | 239.0 | 608.0 |
| 봉사단 | 343.0 | 235.0 | 578.0 |
| 제안 | 132.0 | 397.0 | 529.0 |
| 단원 | 416.0 | 77.0 | 493.0 |
| 서류 | 95.0 | 361.0 | 456.0 |
| 참여 | 165.0 | 284.0 | 449.0 |
| 기간 | 207.0 | 239.0 | 446.0 |
| 증명서 | 109.0 | 320.0 | 429.0 |
| 직원 | 238.0 | 179.0 | 417.0 |
| 귀국 | 329.0 | 65.0 | 394.0 |
| 개발 | 205.0 | 175.0 | 380.0 |
| 일정 | 98.0 | 274.0 | 372.0 |
| 국제 | 129.0 | 241.0 | 370.0 |
| 수행 | 151.0 | 216.0 | 367.0 |
| 제공 | 222.0 | 140.0 | 362.0 |
| 발급 | 103.0 | 257.0 | 360.0 |
| 경력 | 118.0 | 221.0 | 339.0 |
| 파트너 | 108.0 | 224.0 | 332.0 |
| 경험 | 206.0 | 124.0 | 330.0 |
| 예산 | 59.0 | 265.0 | 324.0 |
| 변경 | 76.0 | 242.0 | 318.0 |
| 계획 | 138.0 | 170.0 | 308.0 |
| 자료 | 68.0 | 231.0 | 299.0 |
| 면접 | 16.0 | 275.0 | 291.0 |
| 계속 | 174.0 | 97.0 | 271.0 |
| 시스템 | 75.0 | 190.0 | 265.0 |

In [42]:

```
dftt1_30=dftt1_result.head(30) # 상위 30개 추출  
dftt1_30
```

Out[42]:

| | 21년 | 22년 | 총합계 |
|------|-------|--------|--------|
| 지원 | 512.0 | 1151.0 | 1033.0 |
| 코로나 | 270.0 | 220.0 | 627.0 |
| 교육 | 292.0 | 516.0 | 625.0 |
| 프로그램 | 304.0 | 162.0 | 619.0 |
| 파견 | 239.0 | 410.0 | 608.0 |
| 봉사단 | 235.0 | 846.0 | 578.0 |
| 제안 | 397.0 | 315.0 | 529.0 |
| 단원 | 77.0 | 294.0 | 493.0 |
| 서류 | 361.0 | 420.0 | 456.0 |
| 참여 | 284.0 | 207.0 | 449.0 |
| 기간 | 239.0 | 233.0 | 446.0 |
| 증명서 | 320.0 | 497.0 | 429.0 |
| 직원 | 179.0 | 83.0 | 417.0 |
| 귀국 | 65.0 | 138.0 | 394.0 |
| 개발 | 175.0 | 180.0 | 380.0 |
| 일정 | 274.0 | 364.0 | 372.0 |
| 국제 | 241.0 | 121.0 | 370.0 |
| 수행 | 216.0 | 91.0 | 367.0 |
| 제공 | 140.0 | 51.0 | 362.0 |
| 발급 | 257.0 | 416.0 | 360.0 |
| 경력 | 221.0 | 300.0 | 339.0 |
| 파트너 | 224.0 | 124.0 | 332.0 |
| 경험 | 124.0 | 108.0 | 330.0 |
| 예산 | 265.0 | 202.0 | 324.0 |
| 변경 | 242.0 | 205.0 | 318.0 |
| 계획 | 170.0 | 224.0 | 308.0 |
| 자료 | 231.0 | 152.0 | 299.0 |
| 면접 | 275.0 | 374.0 | 291.0 |
| 계속 | 97.0 | 115.0 | 271.0 |
| 시스템 | 190.0 | 240.0 | 265.0 |

In [43]:

```
rate=(dft_30['21년']-dft_30['20년'])/dft_30['20년']*100 #증감률 계산
rate.name='21-20년증감률' #rate의 컬럼이름 설정

rate1=(dft1_30['22년']-dft1_30['21년'])/dft1_30['21년']*100
rate1.name='22-21년증감률'
```

In [44]:

```
result=pd.concat([dft_30['20년'],dft_30['21년'], rate], axis=1) #concat 이어붙이기
result
```

Out[44]:

| | 20년 | 21년 | 21-20년증감률 |
|------|-------|-------|-------------|
| 지원 | 521.0 | 512.0 | -1.727447 |
| 코로나 | 357.0 | 270.0 | -24.369748 |
| 교육 | 333.0 | 292.0 | -12.312312 |
| 프로그램 | 315.0 | 304.0 | -3.492063 |
| 파견 | 369.0 | 239.0 | -35.230352 |
| 봉사단 | 343.0 | 235.0 | -31.486880 |
| 제안 | 132.0 | 397.0 | 200.757576 |
| 단원 | 416.0 | 77.0 | -81.490385 |
| 서류 | 95.0 | 361.0 | 280.000000 |
| 참여 | 165.0 | 284.0 | 72.121212 |
| 기간 | 207.0 | 239.0 | 15.458937 |
| 증명서 | 109.0 | 320.0 | 193.577982 |
| 직원 | 238.0 | 179.0 | -24.789916 |
| 귀국 | 329.0 | 65.0 | -80.243161 |
| 개발 | 205.0 | 175.0 | -14.634146 |
| 일정 | 98.0 | 274.0 | 179.591837 |
| 국제 | 129.0 | 241.0 | 86.821705 |
| 수행 | 151.0 | 216.0 | 43.046358 |
| 제공 | 222.0 | 140.0 | -36.936937 |
| 발급 | 103.0 | 257.0 | 149.514563 |
| 경력 | 118.0 | 221.0 | 87.288136 |
| 파트너 | 108.0 | 224.0 | 107.407407 |
| 경험 | 206.0 | 124.0 | -39.805825 |
| 예산 | 59.0 | 265.0 | 349.152542 |
| 변경 | 76.0 | 242.0 | 218.421053 |
| 계획 | 138.0 | 170.0 | 23.188406 |
| 자료 | 68.0 | 231.0 | 239.705882 |
| 면접 | 16.0 | 275.0 | 1618.750000 |
| 계속 | 174.0 | 97.0 | -44.252874 |
| 시스템 | 75.0 | 190.0 | 153.333333 |

In [45]:

```
result1=pd.concat([dft1_30['21년'],dft1_30['22년'], rate1], axis=1)
result1
```

Out[45]:

| | 21년 | 22년 | 22-21년증감률 |
|------|-------|--------|------------|
| 지원 | 512.0 | 1151.0 | 124.804688 |
| 코로나 | 270.0 | 220.0 | -18.518519 |
| 교육 | 292.0 | 516.0 | 76.712329 |
| 프로그램 | 304.0 | 162.0 | -46.710526 |
| 파견 | 239.0 | 410.0 | 71.548117 |
| 봉사단 | 235.0 | 846.0 | 260.000000 |
| 제안 | 397.0 | 315.0 | -20.654912 |
| 단원 | 77.0 | 294.0 | 281.818182 |
| 서류 | 361.0 | 420.0 | 16.343490 |
| 참여 | 284.0 | 207.0 | -27.112676 |
| 기간 | 239.0 | 233.0 | -2.510460 |
| 증명서 | 320.0 | 497.0 | 55.312500 |
| 직원 | 179.0 | 83.0 | -53.631285 |
| 귀국 | 65.0 | 138.0 | 112.307692 |
| 개발 | 175.0 | 180.0 | 2.857143 |
| 일정 | 274.0 | 364.0 | 32.846715 |
| 국제 | 241.0 | 121.0 | -49.792531 |
| 수행 | 216.0 | 91.0 | -57.870370 |
| 제공 | 140.0 | 51.0 | -63.571429 |
| 발급 | 257.0 | 416.0 | 61.867704 |
| 경력 | 221.0 | 300.0 | 35.746606 |
| 파트너 | 224.0 | 124.0 | -44.642857 |
| 경험 | 124.0 | 108.0 | -12.903226 |
| 예산 | 265.0 | 202.0 | -23.773585 |
| 변경 | 242.0 | 205.0 | -15.289256 |
| 계획 | 170.0 | 224.0 | 31.764706 |
| 자료 | 231.0 | 152.0 | -34.199134 |
| 면접 | 275.0 | 374.0 | 36.000000 |
| 계속 | 97.0 | 115.0 | 18.556701 |
| 시스템 | 190.0 | 240.0 | 26.315789 |

In [46]:

```
result.to_csv('20,21년 키워드 분석 프로그램 샘플 증감률 최종.csv', encoding='cp949') #저장  
result1.to_csv('21,22년 키워드 분석 프로그램 샘플 증감률 최종.csv', encoding='cp949')
```