

Part1. 워드클라우드 마스크

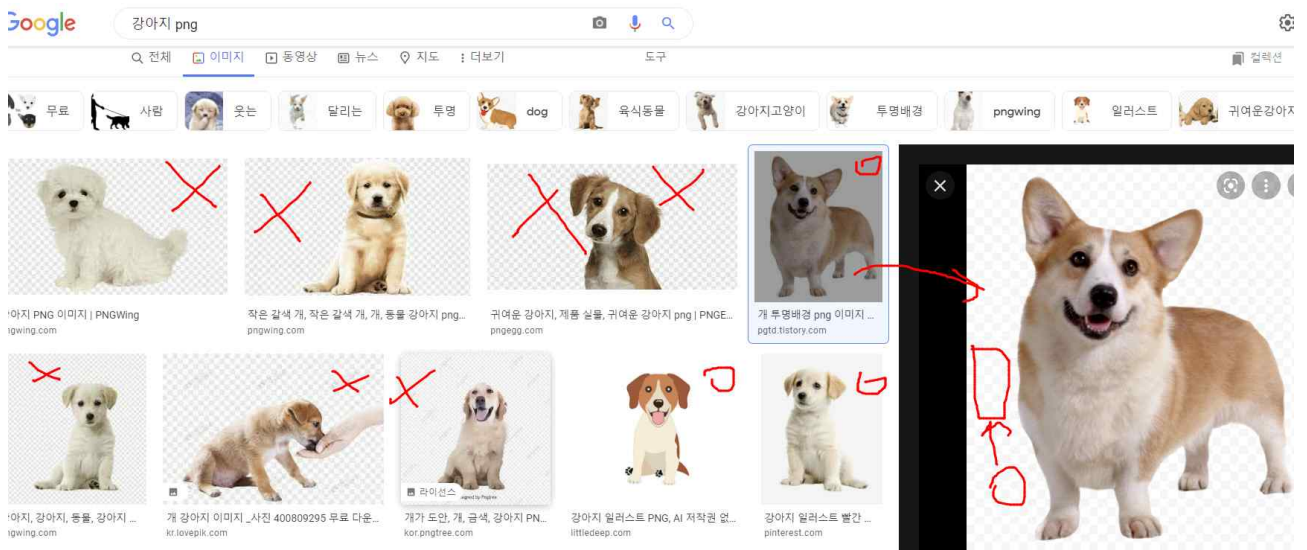
이미지 확장자중 png 파일에 대하여 검색

워드클라우드에 필요한 png 마스크 파일은 구글이미지검색에서 '이미지명 png'로 검색함.

'png'로 검색된 자료중 배경이 흰색인 이미지를 클릭하여서 바둑판식 배경이 나오면 이 바둑판식 배경이 투명하게 저장되는 자료임

이 이미지는 파워포인트등에서 불러서 이미지를 모두 검정색으로 변경할수 있음. (밝기0)

또는 '이미지명 픽토그램'으로 검색하여도 됨.



```
mask_=np.array(Image.open('dog.png'))
wc=WordCloud(font_path='./NanumGothic.ttf',
background_color=color,
width=1000,height=1000,
max_words=100,
max_font_size=300,
mask=mask_)          #mask_ 에 할당된 이미지 파일 작업
```



- for를 이용하여 여러개의 이미지를 동시에 작업할수 있음.

-

Part2. StopWord란?

1. 파이썬 불용어

STOPWORDS는 wordcloud 패키지에서 제공하는 함수로서 갖고 있는 데이터에서 유의미한 단어 토큰만을 선별하기 위해서는 큰 의미가 없는 단어 토큰을 제거하는 작업이 필요합니다. 여기서 큰 의미가 없다는 것은 자주 등장하지만 분석을 하는 것에 있어서는 큰 도움이 되지 않는 단어들을 말합니다. 예를 들면, I, my, me, over, 조사, 접미사 같은 단어들은 문장에서는 자주 등장하지만 실제 의미 분석을 하는 데는 거의 기여하는 바가 없는 경우가 있습니다. 이러한 단어들을 불용어(stopword)라고 함.

- [1] 파이썬 워드클라우드에서 제공하는 기본 stopwords_를 지정함

```
mask_=np.array(Image.open('star.png'))
wc=WordCloud(font_path='./NanumGothic.ttf',
              background_color=color,
              width=1000,height=1000,
              max_words=100,
              max_font_size=300,
              mask=mask_,
              stopwords=stopwords_)
wordFre=wc.generate(lines)
```

- [2] 사용자가 stopword리스트를 추가함
stopwords_.add(i)

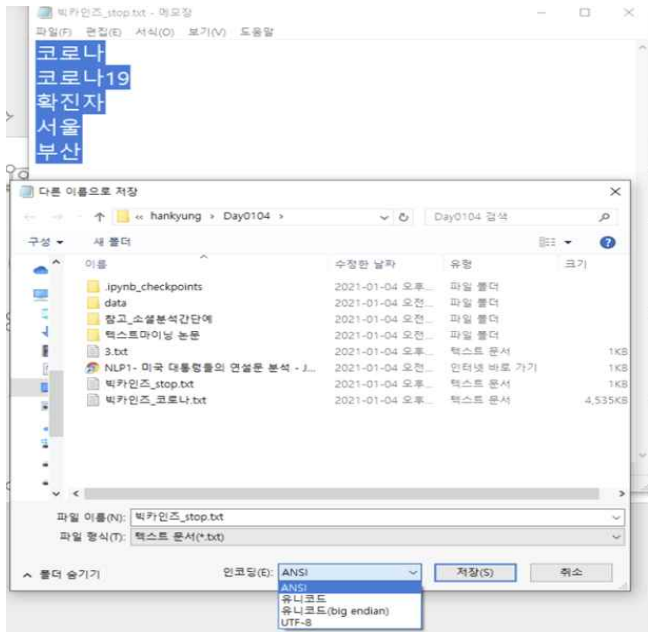
```
1 stopwords_=set(STOPWORDS)
2 stopwords_
3 type(stopwords_)
4
5 li=['now','starry','listen','say','know']
6 for i in li:
7     stopwords_.add(i)
8
```

```
1 myWord('white')
```

2. 불용어 모음집 외부 파일

불용어가 대량일 때 또는 불용어 추가 삭제를 해야할 경우가 있을때는 이 불용어를 메모장에 한단어넣고 enter해서 불용어를 정리한 메모장을 저장해서 읽어옴.

빅카인즈_stop.txt자료를 한줄씩 읽어들여서 마지막의 \n을 기준으로 자료를 나눕니다.



불용어(stopword)가 들어갈
새로운 메모장 파일을 만들

저장할때 ANSI로 저장하면 읽을때 'euc-kr'

단 ANSI는 저장시 특수문자등 저장안되는
기호가 있을수 있음

utf-8, utf-16은 전세계적으로 모든 문자를 표현할
수 있는 비트가 확장된 문자 표현 체계임

1. `f = open('빅카인즈_stop.txt','r',encoding='euc-kr')`
2. 읽을때 어떻 읽을것인가? (2-2가 정답=> `tmp=f.readlines()`)
 - 2-1 : `read()`해서 한개의문장으로 읽음
이자료는 한개의문장에서 `코로나Wn코로나19Wn`
`tmp[0] => 코 / tmp[1] =>로`
 - 2-2 : `readlines()`해서 리스트로 줄단위로 읽어서 처리
`코로나Wn`
`코로나19W`
`tmp[0] => 코로나Wn / tmp[1] =>코로나19Wn`

```
f=open('빅카인즈_stop.txt','r',encoding='euc-kr')
tmp=f.readlines()
f.close
print(tmp) : type(tmp)
```

for i in tmp:

- 3-1. `re=[]` 생성하고 `코로나Wn` 자료를 `W`단위로 나누어서
0번째위치한 글자를 `re.append`한다
`i.split('Wn')[0]`

- 3-2. `tmp[0]` 위치에 있는 '`코로나Wn`' 자료를
`Wn`단위로 나누어서 0번째에 있는 '`코로나`' 글자를
원래위치에(`tmp[0]`)에 할당

```
### 방법1 (3-1방법)
re=[]
for i in tmp:
    re.append(i.split('Wn')[0])
re
```

```
### 방법1 (3-2방법)
cnt=0
for i in tmp:
    cnt+=1
    tmp[cnt]=i.split('Wn')[0]
tmp
```