

# Word Cloud

October 1, 2020

아래 패키지 설치

- wordcloud

- matplotlib

## 0.0.1 KR-WordRank

graph ranking 알고리즘을 이용하여 단어를 추출하기 위해 제안된 방법 > substring graph 를 만든 뒤  
> graph ranking 알고리즘을 이용하여 단어스러운 substring 찾기

```
[1]: from krwordrank.word import KRWordRank

texts = open('lalaland.rtf').read()
wordrank_extractor = KRWordRank(
    min_count = 5, # 단어의 최소 출현 빈도수 (그래프 생성 시)
    max_length = 10, # 단어의 최대 길이
    verbose = True
)

beta = 0.85 # PageRank의 decaying factor beta
max_iter = 10

keywords, rank, graph = wordrank_extractor.extract(texts, beta, max_iter)

for word, r in sorted(keywords.items(), key=lambda x:x[1], reverse=True)[:30]:
    print('%8s:\t%.4f' % (word, r))
```

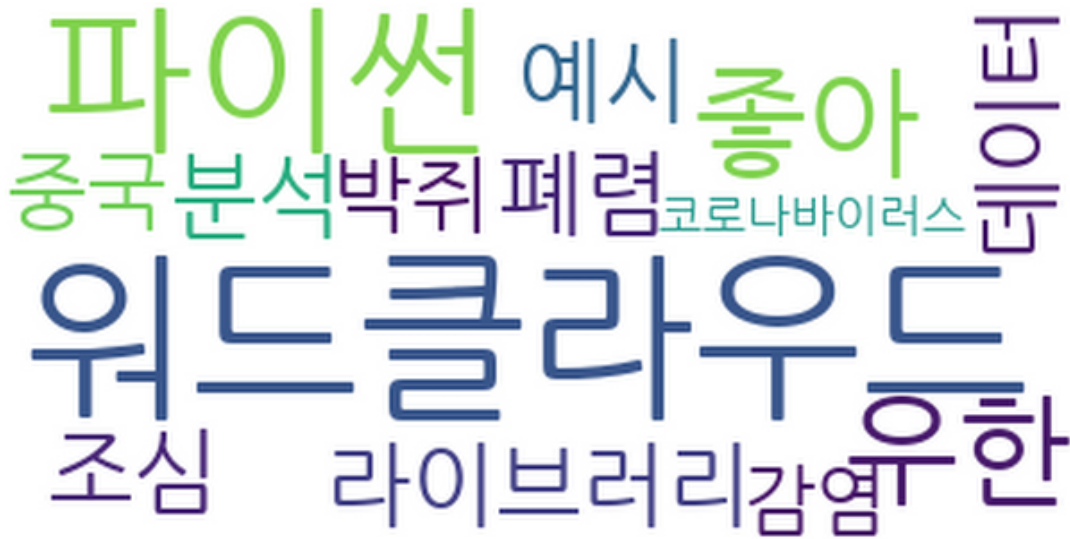
```
scan vocabs ...
num vocabs = 85
done = 1 Early stopped.
```

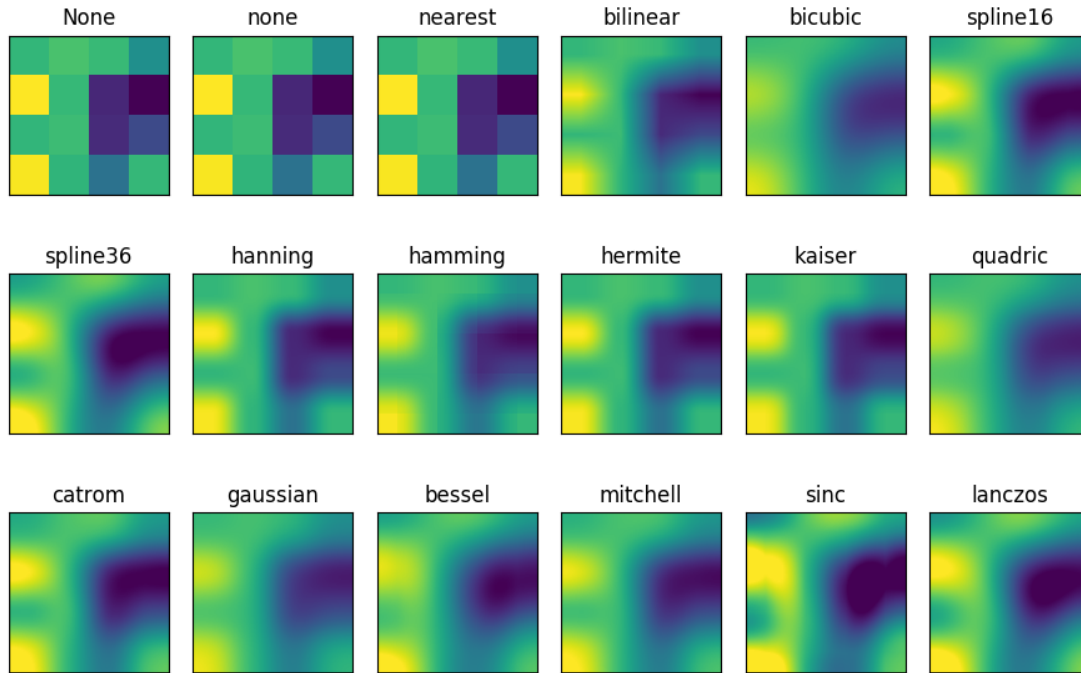
```
[2]: from wordcloud import WordCloud
import matplotlib.pyplot as plt

text = "파이썬 워드클라우드 파이썬 좋아 워드클라우드 파이썬 라이브러리 좋아 파이썬 워드클라우드 예시 워드클라우드 우한 폐렴 조심 데이터 분석 우한 워드클라우드 중국 박쥐 감염 코로나바이러스"
```

```
wordcloud = WordCloud(font_path='/Library/Fonts/NanumGothic.ttf',  
→background_color='white').generate(text)
```

```
[3]: plt.figure(figsize=(11,11)) #이미지 사이즈 지정  
plt.imshow(wordcloud, interpolation='lanczos') #이미지의 부드러기 정도  
plt.axis('off') #x y 축 숫자 제거  
plt.show()
```





참고 [https://matplotlib.org/3.1.1/gallery/images\\_contours\\_and\\_fields/interpolation\\_methods.html](https://matplotlib.org/3.1.1/gallery/images_contours_and_fields/interpolation_methods.html)

```
[4]: from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt

stopwords = set(STOPWORDS)
stopwords.add('워드클라우드')
stopwords.add('파이썬')

wordcloud = WordCloud(font_path='/Library/Fonts/NanumGothic.
→ttf',stopwords=stopwords,background_color='white').generate(text)
plt.figure(figsize=(11,11)) #이미지 사이즈 지정
plt.imshow(wordcloud, interpolation='lanczos') #이미지의 부드러기 정도
plt.axis('off') #x y 축 숫자 제거
plt.show()
```



```
[7]: import numpy as np
      from PIL import Image

      text = open('alice.txt').read()
      alice_mask= np.array(Image.open('alice.png'))

      stopword=set(STOPWORDS)
      stopword.add('said')
```

```
[8]: plt.figure(figsize=(8,8))
      plt.imshow(alice_mask,cmap=plt.cm.gray, interpolation='bilinear')
      plt.axis('off')
      plt.show
```

```
[8]: <function matplotlib.pyplot.show(close=None, block=None)>
```



```
[9]: wc = WordCloud(background_color='white',max_words=2000,mask=alice_mask,
    ↳ stopwords= stopwords)
    wc = wc.generate(text)
```

```
[10]: plt.figure(figsize=(12,12))
    plt.imshow(wc,interpolation='bilinear')
    plt.axis('off')
    plt.show()
```



```
[11]: import numpy as np
from PIL import Image

text = open('alice.txt').read()
alice_coloring= np.array(Image.open('alice_color.png'))

stopword=set(STOPWORDS)
stopword.add('said')

wc = WordCloud(background_color='white',max_words=2000,mask=alice_coloring,
↳stopwords= stopwords)
wc = wc.generate(text)
```

```
[12]: from wordcloud import ImageColorGenerator
image_colors=ImageColorGenerator(alice_coloring)
#image_colors.default_color = [0,0,0]
print(image_colors.image.shape)
```

```
(800, 600, 4)
```

```
[13]: plt.figure(figsize=(8,8))
plt.imshow(alice_coloring,cmap=plt.cm.gray, interpolation='bilinear')
plt.axis('off')
plt.show
```

```
[13]: <function matplotlib.pyplot.show(close=None, block=None)>
```



```
[14]: plt.figure(figsize=(12,12))  
plt.imshow(wc.recolor(color_func = image_colors), interpolation='bilinear')  
plt.axis('off')  
plt.show()
```





- [https://github.com/amueller/word\\_cloud](https://github.com/amueller/word_cloud)
- <https://github.com/lovit/KR-WordRank>

[ ]: