

분노의 질주 분석해 보기

```
In [1]: import nltk
import matplotlib.pyplot as plt
import platform
import numpy as np
```

```
In [2]: # from konlpy.tag import Twitter # 0.45 버전부터 변경.
from konlpy.tag import Okt
```

```
In [3]: ### 한글 폰트 설정
from matplotlib import font_manager, rc
path = "C:/Windows/Fonts/malgun.ttf"
if platform.system() == "Windows":
    font_name = font_manager.FontProperties(fname=path).get_name()
    rc('font', family=font_name)
elif platform.system()=="Darwin":
    rc('font', family='AppleGothic')
else:
    print("Unknown System")
```

```
In [10]: doc_ko = open("data/15_TheExtreme.txt").read()
```

```
In [11]: doc_ko
```

```
Out[11]: '"x"Wn"1" " 분노의 질주 시리즈중에서 제일 별루"Wn"2" " 스케일 큰 시끄러운 액션이 난무하는데도 이렇게까지 지루할수 있다
니....."Wn"3" " 시~원 하게 잘 본 영화. 다음 시리즈에서는 여자 주인공의 비중이 더 높아졌으면 하는 바램!"Wn"4" " 반지닥기, 자살닥
이, 고무닥이, 정의닥이...로 이어지는 한심한 DC 시리즈 Wn레지던트 이블 시리즈 Wn그리고 이 영화 분노의 질주 시리즈 Wn공통점은 시
리즈가 거듭될수록 돈은 많이 들지만 재미는 없어지고 WnCG는 딱질되지만 실감나는 장면은 더 없어지도 뻔히 가짜라는게 드러나는 영화
들 Wn그러나Wn아무리 엉터리로 만들고, 자국에서 망해도 Wn미국 블록버스터라면 맹목적으로 보는 중국애들 땀에 Wn아무리 쓰레기 영화라
도 본전 건지는 것은 물론 상당히 많은 돈을 버니... Wn이런 쓰레기들이 매년 양산된다. Wn물론, 중국애들도 할말은 있을 거다 Wn공산당
이 검열하는 자국영화보다는 낫다고... Wn하지만 우리들은 다른 전세계의 재미있는 영화를 볼 선택의 자유가 있잖아! Wn왜 이런 쓰레기
영화를 보는 거지?"Wn"5" " "Wn"6" " 그냥 액션만 보면 멋진데Wn스토리는 주인공이 전여친한테 싸지른Wn애새끼 구하러 간다며 아빠행세
하면서Wn그 덕분에 지동료들 다 버리고 미쳐 날뛰는 내용"Wn"7" " "Wn"8" " 아래는 다들 평점 알바들인가부네.. 이런 개 쓰레기 영화가
평점이 이리 높다니"Wn"9" " "Wn"10" " "Wn"11" " "Wn"12" " "Wn"13" " "Wn"14" " "Wn"15" " 스케일은 점점 더 커지지만, 액션은 멍청할
정도로 어이가없음 과유불급"Wn"16" " 이 시리즈로 이렇게 길게 간다는게 신기.. 새로운 건 없지만 달리는 걸 좋아하시는 분이라면
"Wn"17" " "Wn"18" " "Wn"19" " "Wn"20" " 대머리들은 TV물로 찍고,Wn감독은 CG실에서 이어붙히고."Wn"21" " "Wn"22" " "Wn"23" " "Wn"2
4" " "Wn"25" " "Wn"26" " "Wn"27" " "Wn"28" " "Wn"29" " 이제는 제목만 남은 분노의 질주"Wn"30" " "Wn"31" " 해도 너무하네"Wn"32" "
"Wn"33" " "Wn"34" " 전형적인 헐리우드 액션물 강 생각없이 봐야 좋은 영화"Wn"35" " 이렇게 뻔하기도 힘들다 노잼 노볼거리 어떻게
이런 평점이 나오냐"Wn"36" " 재밌어요"Wn"37" " 시리즈로 8편까지 질질 간거면 지겹고 루즈할만 한데Wn분노의 질주는 회가 갈수록 더
박진감 넘치고 더 스케일이 커지니 어쩔 줄냐... Wn통쾌한 액션과 스케일로 시간 가는지 모르고 봤네..Wn그냥 이정도 영화를 만드는 헐
리웃이 부러울 따름..."Wn"38" " 보는 재미"Wn"39" " "Wn"40" " 내용이야 뭐 어따랴.....Wn저렇게 달리는 데....."Wn"41" " 열받아서 이
야심한 시간에 평가 남긴다. 이런걸로 8.5 받으면 다음 영화 평점은 사기다.Wn"Wn"42" " 1편보다 못한듯,,,분노의 질주가 대박인
데,,, "Wn"43" " "Wn"44" " 좀 과장해서 대사의 반이 가쪽 어찌고 억겨웠다 억지 스러움"Wn"45" " 배우들때에 5점준다..돈은 붓고 싶었는
데 이정도면...한편 무라더기 가다... 그 후속편 시리즈로 으악으악 하... "Wn"46" " 뻔한 스토리에 액션은 다해사 시키면 노리 거가다
```

```
In [12]: # t = Twitter()  
t = Okt()  
tokens_ko = t.nouns(doc_ko)  
tokens_ko
```

```
Out[12]: ['분노',  
'질주',  
'시리즈',  
'제일',  
'별루',  
'스케일',  
'액션',  
'난무',  
'시',  
'원',  
'영화',  
'다음',  
'시리즈',  
'여자',  
'주인공',  
'비중',  
'더',  
'바람',  
'반지',  
'기차']
```

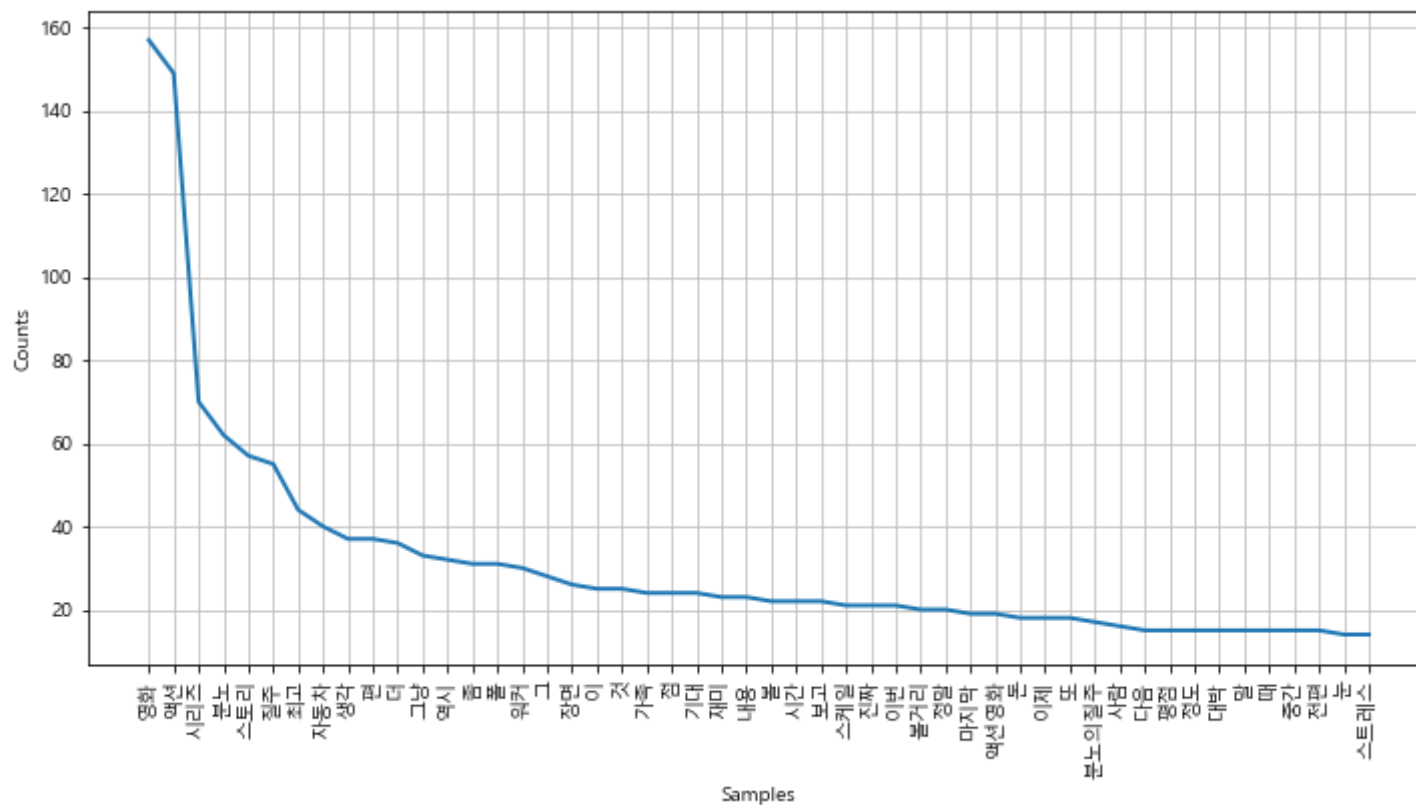
```
In [13]: ko = nltk.Text(tokens_ko, name="분노의 질주")
print(len(ko.tokens))      # tokens의 수
print(len(set(ko.tokens))) # 유일한 tokens의 수
ko.vocab()                 # 단어 빈도
```

```

'적': 3,
'다시': 7,
'세계': 2,
'기상': 1,
'외한': 1,
'차로': 4,
'수': 9,
'여러가지': 1,
'중합셋트': 1,
'답': 3,
'차': 13,
'부쉬': 1,
'광고': 1,
'어차피': 1,
'언제': 5,
'압도': 2,
'중간': 15,
'브라이언': 12,
'상기': 1,
'정기': 10

```

```
In [14]: plt.figure(figsize=(12,6))
ko.plot(50)
plt.show()
```



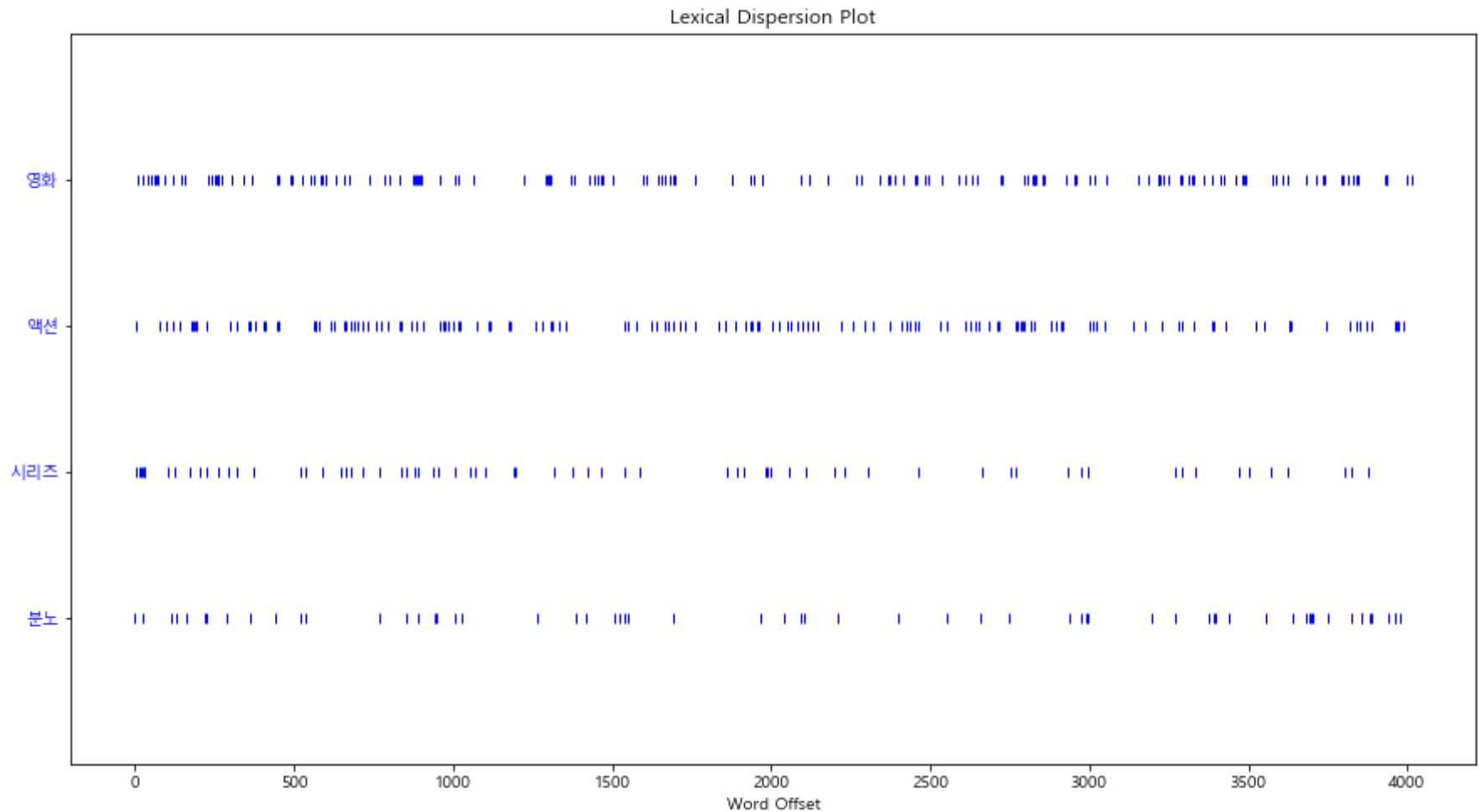
```
In [15]: ### 불용어 처리
stop_words = ['더', '말', '때', '눈', '또', '이', '그', '좀', '점', '것', '돈', '이제']

#ko = [each_word for each_word in ko if each_word not in stop_words]
new_ko = []
for one_word in ko:
    if one_word not in stop_words:
        new_ko.append(one_word)
```

Sample	Counts
1	158
2	148
3	70
4	65
5	60
6	55
7	55
8	45
9	40
10	38
11	38
12	33
13	32
14	31
15	30
16	25
17	24
18	23
19	23
20	21
21	21
22	21
23	21
24	21
25	21
26	20
27	20
28	20
29	19
30	18
31	15
32	15
33	15
34	15
35	15
36	15
37	15
38	15
39	12
40	11

6/10

```
In [18]: ### 원하는 단어의 문서내 위치를 알 수 있음.  
plt.figure(figsize=(15,8))  
new_ko.dispersion_plot(["영화", "액션", "시리즈", "분노"])
```



```
In [27]: ### 기타 참조 - konlpy.pdf 참조
### new_ko.concordance("영화") # 주변부 단어 확인
### new_ko.collocations() # 어떤 단어들이 연어로 사용되었을까?
new_ko.collocations()
```

제이슨 스타덤; 드웨인존슨 제이슨; 어벤저스 히어로; 제이슨 스타덤; 스토리 주인공; 스토리 볼거리; 분노의질주 시리즈;
시리즈 자동차

워드 클라우드 적용

```
In [31]: from wordcloud import WordCloud, STOPWORDS

import numpy as np
from PIL import Image

stopwords = set(STOPWORDS) # 불용어 처리
alice_mask = np.array(Image.open("img/Draw_car1.png"))
```

```
In [32]: data = new_ko.vocab().most_common(150)
```



```
path = "C:/Windows/Fonts/malgun.ttf"
wc = WordCloud(font_path=path, relative_scaling=0.2,
               background_color="white", mask=alice_mask).generate_from_frequencies(dict(data))

plt.figure(figsize=(12,8))
plt.imshow(wc)
plt.axis("off")
plt.show()
```



In []:

