워드 클라우드 (Word Cloud)

1. 워드 클라우드(Word Cloud)

- 핵심 단어를 시각화하는 기법
- 문서의 키워드, 개념 등을 직관적으로 파악할 수 있게 핵심 단어를 시각적으로 돋보이게 하는 기법

1-1, wordcloud 라이브러리 설치 및 임포트

```
!pip install wordcloud from wordcloud import WordCloud
```

```
# alice.txt 파일 읽어오기
text = open('/content/drive/MyDrive/KDT/Python/2. 데이터분석/alice.txt').read()
text
```

1-2. generate()

• 단어별 출현 빈도수를 비율로 반환하는 객체를 생성

```
wordcloud = WordCloud().generate(text)
wordcloud # 객체 주소값

>>>
<wordcloud.wordcloud.WordCloud at 0x7f16548619f0>
# words_ : 객체의 비율의 정보가 담긴 딕셔너리를 반환
wordcloud.words_
```

1-3. 시각화

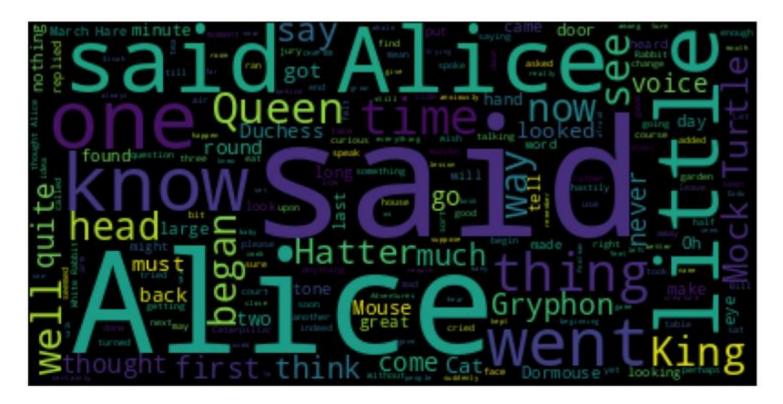
```
import matplotlib.pyplot as plt

plt.figure(figsize=(15,10))

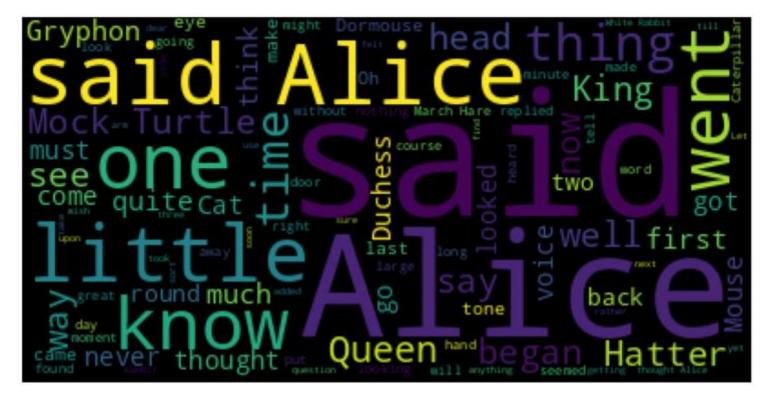
plt.imshow(wordcloud) # wordcloud 객체를 넣으면 워드클라우드 형태의 그래프 생성

plt.axis('off') #눈금 삭제

plt.show()
```



```
# max_words: 워드 클라우드에 표시되는 단어의 개수를 설정
wordcloud = WordCloud(max_words=100).generate(text) # 단어 100개만 표시
plt.figure(figsize=(15,10))
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```



1-4. 불용어 설정

• 워드클라우드로 작성할 글에서 제외할 단어를 설정

```
from wordcloud import STOPWORDS
STOPWORDS # 불용머로 등록된 단어들
type(STOPWORDS)
```

>>> set # 중복값x

✓ 불용어 추가하기

```
# add(): 불용어 추가하기
STOP#ORDS.add('said')
print(STOP#ORDS)

>>> {"i'm", 'after', 'few', 'most', 'since', "you're", 'shall', "i'll", 'would', 'com', 'otherwise', "she'd", "won't", "we'll", 'yourself', 'himself', "why's", 'during', 'what', "that's", 'not', 'how', 'me', "hadn't", "i'd", "there's", 'same', 'else', "he's", 'such', 'ourselves', "they'll", 'more', "you've", 'i', 'was', 'could', "she's", 'its', 'http', 'while', 'herself', 'the', 'by', 'in', 'therefore', 'be', 'through', "i've", 'did', "hasn't", 'has', 'it', 'also', 'why', "they'd", 'my',...}
```

☑ 불용어를 제외하고 핵심 단어만 워드클라우드로 시각화하기

```
wordcloud = WordCloud(max_words=100,
font_path='/usr/share/fonts/truetype/nanum/NanumGothicExtraBold.ttf',
stopwords=STOPWORDS).generate(text) # 'said' 사라짐
plt.figure(figsize=(15,10))
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```

