

## Stopwords

it is necessary to remove word tokens that are not very meaningful. Here, 'not very meaningful' refers to words that appear frequently but are not very helpful in analyzing the data. For example, words like I, my, me, over, particles (조사), suffixes (접미사) appear frequently in sentences, but often contribute little to actual semantic analysis. These words are called stop words, and NLTK predefines more than 100 such English words as stop words within the package."

### Code:

```
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from konlpy.tag import Okt

stop_words_list = stopwords.words('english')
print('no_stopwords :', len(stop_words_list))
print('불용어 10개 출력 :', stop_words_list[:10])
```

### Output:

```
불용어 개수 : 179
불용어 10개 출력 : ['i', 'me', 'my', 'myself', 'we', 'our',
'ours', 'ourselves', 'you', "you're"]
```