

```
In [2]: from konlpy.tag import Hannanum
        from konlpy.tag import Kkma
        from konlpy.tag import Komoran
        from konlpy.tag import Mecab
        from konlpy.tag import Twitter
```

Hannanum

```
In [3]: text = '자연어처리는 재미있습니다. 그러나 한국어 분석은 쉽지않습니다.'
```

```
In [4]: hannanum = Hannanum()
```

C:\WPython\WAnaconda3\lib\site-packages\Wjpycore.py:210: UserWarning:

Deprecated: convertStrings was not specified when starting the JVM. The default behavior in JPyPy will be False starting in JPyPy 0.8. The recommended setting for new code is convertStrings=False. The legacy value of True was assumed for this session. If you are a user of an application that reported this warning, please file a ticket with the developer.

```
""")
```

```
In [5]: hannanum.analyze(text)
```

```
Out[5]: [[(['자연어처리', 'ncpa'), ('는', 'jxc')],
          [(['자연어', 'ncn'), ('처리', 'ncpa'), ('는', 'jxc')]],
          [(['재미있', 'pvg'), ('습니다', 'ef')],
          [(['재미', 'ncn'), ('있', 'xsmn'), ('습니다', 'ef')]],
          [(['.', 'sf'), [('.', 'sy')]],
          [],
          [(['그러나', 'maj')]],
          [(['한국어', 'ncn')]],
          [(['분석', 'ncpa'), ('은', 'jxc')], [(['분석', 'ncpa'), ('은', 'ncn')]],
          [(['쉽', 'paa'), ('지', 'ecx'), ('않', 'px'), ('습니다', 'ef')]],
          [(['.', 'sf'), [('.', 'sy')]]]
```

```
In [6]: hannanum.morphs(text)
```

```
Out[6]: ['자연어처리',
          '는',
          '재미있',
          '습니다',
          '.',
          '그러나',
          '한국어',
          '분석',
          '은',
          '쉽',
          '지',
          '않',
          '습니다',
          '.']
```

```
In [7]: hannanum.nouns(text)
```

```
Out[7]: ['자연어처리', '한국어', '분석']
```

```
In [8]: hannanum.pos(text)
```

```
Out[8]: [('자연어처리', 'N'),  
         ('는', 'J'),  
         ('재미있', 'P'),  
         ('습니다', 'E'),  
         ('.', 'S'),  
         ('그러나', 'M'),  
         ('한국어', 'N'),  
         ('분석', 'N'),  
         ('은', 'J'),  
         ('쉽', 'P'),  
         ('지', 'E'),  
         ('않', 'P'),  
         ('습니다', 'E'),  
         ('.', 'S')]
```

Kkma

```
In [9]: text = '''  
꼬꼬마 한글 형태소 분석기는 Java 라이브러리로써 jar 파일 형태로 배포한다.  
사전을 메모리에 적재하기 위한 충분한 힙메모리를 지정해주어야 한다.  
따라서 512MB이상의 메모리를 지정할 것을 권장한다.  
'''
```

```
In [10]: kkma = Kkma()
```

```
In [11]: kkma.morphs(text)
```

```
Out[11]: ['꼬꼬마',  
          '한글',  
          '형태소',  
          '분석기',  
          '는',  
          'Java',  
          '라이브러리',  
          '로써',  
          'jar',  
          '파일',  
          '형태',  
          '로',  
          '배포',  
          '하',  
          'ㄴ다',  
          '.',  
          '사전',  
          '을',  
          '메모리',  
          '에',  
          '적재',  
          '하',  
          '기',  
          '위하',  
          'ㄴ',  
          '충분',  
          '하',  
          'ㄴ',  
          '힘',  
          '메모리',  
          '를',  
          '지정',  
          '하',  
          '어',  
          '주',  
          '어야',  
          '하',  
          'ㄴ다',  
          '.',  
          '따르',  
          '아서',  
          '512',  
          '"',  
          'B',  
          '이상',  
          '의',  
          '메모리',  
          '를',  
          '지정',  
          '하',  
          'ㄴ',  
          '것',  
          '을',  
          '권장',  
          '하',  
          'ㄴ다',  
          '.']
```

```
In [12]: kkma.nouns(text)
```

```
Out[12]: ['꼬꼬마',  
          '한글',  
          '형태소',  
          '분석기',  
          '라이브러리',  
          '파일',  
          '형태',  
          '배포',  
          '사전',  
          '메모리',  
          '적재',  
          '힙',  
          '힙메모리',  
          '지정',  
          '512',  
          '이상',  
          '권장']
```

```
In [13]: kkma.pos(text)
```

```
Out[13]: [('꼬꼬마', 'NNG'),  
('한글', 'NNG'),  
('형태소', 'NNG'),  
('분석기', 'NNG'),  
('는', 'JX'),  
('Java', 'OL'),  
('라이브러리', 'NNG'),  
('로써', 'JKM'),  
('jar', 'OL'),  
('파일', 'NNG'),  
('형태', 'NNG'),  
('로', 'JKM'),  
('배포', 'NNG'),  
('하', 'XSV'),  
('ㄴ다', 'EFN'),  
('.', 'SF'),  
('사전', 'NNG'),  
('을', 'JKO'),  
('메모리', 'NNG'),  
('에', 'JKM'),  
('적재', 'NNG'),  
('하', 'XSV'),  
('기', 'ETN'),  
('위하', 'VV'),  
('ㄴ', 'ETD'),  
('충분', 'XR'),  
('하', 'XSA'),  
('ㄴ', 'ETD'),  
('힉', 'NNG'),  
('메모리', 'NNG'),  
('를', 'JKO'),  
('지정', 'NNG'),  
('하', 'XSV'),  
('어', 'ECS'),  
('주', 'VXV'),  
('어아', 'ECD'),  
('하', 'VV'),  
('ㄴ다', 'EFN'),  
('.', 'SF'),  
('따르', 'VV'),  
('아서', 'ECD'),  
('512', 'NR'),  
('"'', 'SW'),  
('B', 'OL'),  
('이상', 'NNG'),  
('의', 'JKG'),  
('메모리', 'NNG'),  
('를', 'JKO'),  
('지정', 'NNG'),  
('하', 'XSV'),  
('ㄹ', 'ETD'),  
('것', 'NNB'),  
('을', 'JKO'),  
('권장', 'NNG'),  
('하', 'XSV'),  
('ㄴ다', 'EFN'),  
('.', 'SF')]
```

```
In [14]: kkma.sentences(text)
```

```
Out[14]: ['꼬꼬마 한글 형태소 분석기는 Java 라이브러리로써 jar 파일 형태로 배포한다.',  
          '사전을 메모리에 적재하기 위한 충분한 힙 메모리를 지정해 주어야 한다.',  
          "따라서 512MB 이상의 메모리를 지정할 것을 권장한다."]
```

Komoran

```
In [15]: text = '코모란도 오픈소스가 되었어요. 근데 메모리를 너무 먹네요 ^^'
```

```
In [16]: komoran = Komoran()
```

Twitter

```
In [17]: last_message = '''  
안녕하세요!  
오늘이 마지막 수업이군요.  
짧은 과정이었지만 파이썬이라는 프로그래밍언어를 통해 데이터분석 입문의 디딤돌이 되고자  
노력했습니다.  
노력은 절대 배신하지 않는다는 말을 믿습니다.  
언제 어디서 무슨일을 하든지 항상 응원하겠습니다.  
그동안 수고 많으셨습니다.  
'''
```

```
In [18]: twitter = Twitter()
```

```
C:\WPython\Anaconda3\lib\site-packages\konlpy\tag Okt.py:16: UserWarning: "Twitter" has  
s changed to "Okt" since KoNLPy v0.4.5.  
warn("Twitter" has changed to "Okt" since KoNLPy v0.4.5.)
```

```
In [21]: twitter.morphs(last_message)
```

```
Out[21]: ['Wn',  
'안녕하세요',  
'!',  
'Wn',  
'오늘이',  
'마지막',  
'수업',  
'이군',  
'요',  
'.',  
'Wn',  
'짧은',  
'과정',  
'이었지만',  
'파이썬',  
'이라는',  
'프로그램',  
'언어',  
'를',  
'통해',  
'데이터',  
'분석',  
'입문',  
'의',  
'디딤돌',  
'이',  
'되고자',  
'노력',  
'했습니다',  
'.',  
'Wn',  
'노력',  
'은',  
'절대',  
'배신',  
'하지',  
'않는다는',  
'말',  
'을',  
'믿습니다',  
'.',  
'Wn',  
'언제',  
'어디서',  
'무슨',  
'일',  
'을',  
'하던지',  
'항상',  
'응원',  
'하겠습니다',  
'.',  
'Wn',  
'그동안',  
'수고',  
'많으셨습니다',  
'.',  
'Wn']
```

```
In [20]: twitter.nouns(last_message)
```

```
Out[20]: ['오늘이',  
          '마지막',  
          '수업',  
          '이군',  
          '과정',  
          '파이썬',  
          '프로그램',  
          '언어',  
          '통해',  
          '데이터',  
          '분석',  
          '입문',  
          '디딤돌',  
          '노력',  
          '노력',  
          '절대',  
          '배신',  
          '말',  
          '언제',  
          '일',  
          '항상',  
          '응원',  
          '그동안',  
          '수고']
```

```
In [22]: twitter.phrases(last_message)
```

```
Out[22]: ['오늘이',  
          '오늘이 마지막',  
          '오늘이 마지막 수업이군',  
          '짧은 과정',  
          '파이썬',  
          '프로그램언어',  
          '통해',  
          '통해 데이터분석',  
          '통해 데이터분석 입문',  
          '통해 데이터분석 입문의 디딤돌',  
          '노력',  
          '절대',  
          '절대 배신',  
          '언제',  
          '무슨일',  
          '항상',  
          '항상 응원',  
          '그동안',  
          '그동안 수고',  
          '마지막',  
          '수업',  
          '이군',  
          '과정',  
          '프로그램',  
          '언어',  
          '데이터',  
          '분석',  
          '입문',  
          '디딤돌',  
          '배신',  
          '응원',  
          '수고']
```



```
In [23]: twitter.pos(last_message)
```

```
Out[23]: [('Wn', 'Foreign'),
          ('안녕하세요', 'Adjective'),
          ('!', 'Punctuation'),
          ('Wn', 'Foreign'),
          ('오늘이', 'Noun'),
          ('마지막', 'Noun'),
          ('수업', 'Noun'),
          ('이군', 'Noun'),
          ('요', 'Josa'),
          ('.', 'Punctuation'),
          ('Wn', 'Foreign'),
          ('짧은', 'Adjective'),
          ('과정', 'Noun'),
          ('이었지만', 'Verb'),
          ('파이썬', 'Noun'),
          ('이라는', 'Josa'),
          ('프로그램', 'Noun'),
          ('언어', 'Noun'),
          ('를', 'Josa'),
          ('통해', 'Noun'),
          ('데이터', 'Noun'),
          ('분석', 'Noun'),
          ('입문', 'Noun'),
          ('의', 'Josa'),
          ('디딤돌', 'Noun'),
          ('이', 'Josa'),
          ('되고자', 'Verb'),
          ('노력', 'Noun'),
          ('했습니다', 'Verb'),
          ('.', 'Punctuation'),
          ('Wn', 'Foreign'),
          ('노력', 'Noun'),
          ('은', 'Josa'),
          ('절대', 'Noun'),
          ('배신', 'Noun'),
          ('하지', 'Verb'),
          ('않는다는', 'Verb'),
          ('말', 'Noun'),
          ('을', 'Josa'),
          ('믿습니다', 'Verb'),
          ('.', 'Punctuation'),
          ('Wn', 'Foreign'),
          ('언제', 'Noun'),
          ('어디서', 'Adverb'),
          ('무슨', 'Modifier'),
          ('일', 'Noun'),
          ('을', 'Josa'),
          ('하든지', 'Verb'),
          ('항상', 'Noun'),
          ('응원', 'Noun'),
          ('하겠습니다', 'Verb'),
          ('.', 'Punctuation'),
          ('Wn', 'Foreign'),
          ('그동안', 'Noun'),
          ('수고', 'Noun'),
          ('많으셨습니다', 'Adjective'),
          ('.', 'Punctuation'),
          ('Wn', 'Foreign')]
```

```
In [24]: # norm - If True, normalize tokens.
twitter.pos(last_message, norm=True)
```

```
Out[24]: [('Wn', 'Foreign'),
('안녕하세요', 'Adjective'),
('!', 'Punctuation'),
('Wn', 'Foreign'),
('오늘이', 'Noun'),
('마지막', 'Noun'),
('수업', 'Noun'),
('이군', 'Noun'),
('요', 'Josa'),
('.', 'Punctuation'),
('Wn', 'Foreign'),
('짧은', 'Adjective'),
('과정', 'Noun'),
('이었지만', 'Verb'),
('파이썬', 'Noun'),
('이라는', 'Josa'),
('프로그램', 'Noun'),
('언어', 'Noun'),
('를', 'Josa'),
('통해', 'Noun'),
('데이터', 'Noun'),
('분석', 'Noun'),
('입문', 'Noun'),
('의', 'Josa'),
('디딤돌', 'Noun'),
('이', 'Josa'),
('되고자', 'Verb'),
('노력', 'Noun'),
('했습니다', 'Verb'),
('.', 'Punctuation'),
('Wn', 'Foreign'),
('노력', 'Noun'),
('은', 'Josa'),
('절대', 'Noun'),
('배신', 'Noun'),
('하지', 'Verb'),
('않는다는', 'Verb'),
('말', 'Noun'),
('을', 'Josa'),
('믿습니다', 'Verb'),
('.', 'Punctuation'),
('Wn', 'Foreign'),
('언제', 'Noun'),
('어디서', 'Adverb'),
('무슨', 'Modifier'),
('일', 'Noun'),
('을', 'Josa'),
('하든지', 'Verb'),
('항상', 'Noun'),
('응원', 'Noun'),
('하겠습니다', 'Verb'),
('.', 'Punctuation'),
('Wn', 'Foreign'),
('그동안', 'Noun'),
('수고', 'Noun'),
('많으셨습니다', 'Adjective'),
('.', 'Punctuation'),
('Wn', 'Foreign')]
```

```
In [25]: # stem - If True, stem tokens.  
twitter.pos(last_message, norm=True, stem=True)
```

```
Out[25]: [('Wn', 'Foreign'),  
( '안녕하다', 'Adjective'),  
( '!', 'Punctuation'),  
( 'Wn', 'Foreign'),  
( '오늘이', 'Noun'),  
( '마지막', 'Noun'),  
( '수업', 'Noun'),  
( '이군', 'Noun'),  
( '요', 'Josa'),  
( '.', 'Punctuation'),  
( 'Wn', 'Foreign'),  
( '짧다', 'Adjective'),  
( '과정', 'Noun'),  
( '이다', 'Verb'),  
( '파이썬', 'Noun'),  
( '이라는', 'Josa'),  
( '프로그램', 'Noun'),  
( '언어', 'Noun'),  
( '를', 'Josa'),  
( '통해', 'Noun'),  
( '데이터', 'Noun'),  
( '분석', 'Noun'),  
( '입문', 'Noun'),  
( '의', 'Josa'),  
( '디딤돌', 'Noun'),  
( '이', 'Josa'),  
( '되다', 'Verb'),  
( '노력', 'Noun'),  
( '하다', 'Verb'),  
( '.', 'Punctuation'),  
( 'Wn', 'Foreign'),  
( '노력', 'Noun'),  
( '은', 'Josa'),  
( '절대', 'Noun'),  
( '배신', 'Noun'),  
( '하다', 'Verb'),  
( '않다', 'Verb'),  
( '말', 'Noun'),  
( '을', 'Josa'),  
( '믿다', 'Verb'),  
( '.', 'Punctuation'),  
( 'Wn', 'Foreign'),  
( '언제', 'Noun'),  
( '어디서', 'Adverb'),  
( '무슨', 'Modifier'),  
( '일', 'Noun'),  
( '을', 'Josa'),  
( '하다', 'Verb'),  
( '항상', 'Noun'),  
( '응원', 'Noun'),  
( '하다', 'Verb'),  
( '.', 'Punctuation'),  
( 'Wn', 'Foreign'),  
( '그동안', 'Noun'),  
( '수고', 'Noun'),  
( '많다', 'Adjective'),  
( '.', 'Punctuation'),  
( 'Wn', 'Foreign')]
```

In []: