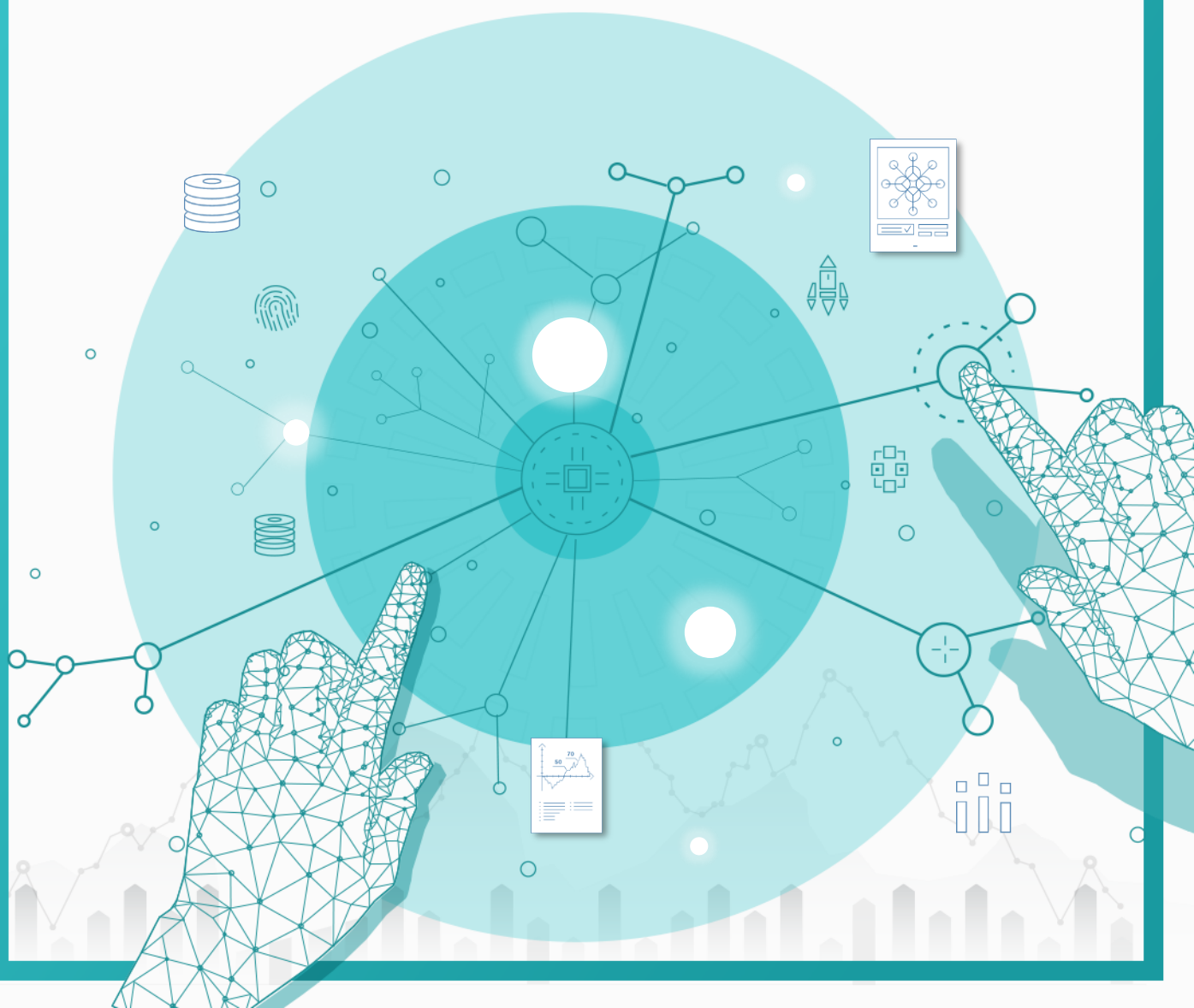




한국기술교육대학교  
온라인평생교육원

# 파이썬을 활용한 인공지능 자연어 처리(실습)

자연어 처리 형태소 분석



# 자연어 처리 형태소 분석

## 학습 목표

1. 영문·한글 형태소 분석을 수행할 수 있다.
2. 한글 형태소 분석을 위한 사전을 활용할 수 있다.

## 학습 내용

1. 영문·한글 형태소 분석
2. 형태소 분석 사전 활용

### 1. 영문·한글 형태소 분석

#### 1) 영문 형태소 분석

##### (1) 형태소 분석을 위한 전처리 수행

특수문자  
제거

토큰화

불용어  
제거

소문자화

##### (2) 자연어 처리를 위한 nltk | nltk를 이용한 영문 형태소 분석

- pos\_tag 함수 import 후 pos\_tag 함수를 이용하여 형태소 분석 수행

```
from nltk.tag import pos_tag

tagged_list = pos_tag(set(tokens))
print(tagged_list)

[('and', 'CC'), ('stay', 'VB'), ('words', 'NNS'), ('begin', 'VB'), ('graduate', 'NN')]
```

### 1. 영문·한글 형태소 분석

#### 2) 한글 형태소 분석

##### (1) 형태소 분석을 위한 전처리 수행

특수문자  
제거

토큰화

불용어  
제거

띄어쓰기  
처리

##### (2) KoNLPy 및 관련 패키지 설치

###### ▪ Java, JPyype 설치

```
!pip install JPyype1
```

###### ▪ KoNLPy 설치

```
!pip install KoNLPy
```

###### ▪ Linux용 Mecab 설치

```
!pip install JPyype1
!pip install konlpy
!sudo apt-get install curl git
!bash <<(curl -s https://raw.githubusercontent.com/konlpy/konlpy/master/s
```

### 1. 영문·한글 형태소 분석

#### 2) 한글 형태소 분석

##### (3) KoNLPy를 이용한 한글 형태소 분석

- 형태소 분석기 객체 생성
  - KoNLPy의 Okt, Kkma, Komoran, Hannanum, Mecab 등의 클래스를 이용하여 객체 생성
  - 각 형태소 클래스별 특성 이해 및 활용

```
from konlpy.tag import Okt, Kkma, Komoran, Hannanum, Mecab
# tagger = Kkma()
# tagger = Komoran()
# tagger = Hannanum()
# tagger = Okt()
tagger = Mecab()
```

1. 영문·한글 형태소 분석

2) 한글 형태소 분석

(3) KoNLPy를 이용한 한글 형태소 분석

▪ 형태소 분석 클래스별 품사 태깅 성능 비교

Hannanum	Kkma	Komoran	Mecab	Twitter
아버지가방에 들어가 / N	아버지 / NNG	아버지가방에 들어가신다 / NNP	아버지 / NNG	아버지 / Noun
이 / J	가방 / NNG		가 / JKS	가방 / Noun
시~다 / E	에 / JKM		방 / NNG	에 / Josa
	들어가 / VV		에 / JKB	들어가신 / Verb
	시 / EPH		들어가 / VV	다 / Eomi
	s다 / EFN		신다 / EP+EC	

▪ 형태소 분석 클래스별 수행시간 성능 비교

Hannanum	Kkma	Komoran
Kkma	5.6988 secs	35.7163 secs
Komoran	5.4866 secs	25.6008 secs
Hannanum	0.6591 secs	8.8251 secs
Okt(previous Twitter)	1.4870 secs	2.4714 secs
Mecab	0.0007 secs	0.2838 secs

※ 출처 : KoNLPy 형태소 분석 및 품사 태깅,  
<https://konlpy-ko.readthedocs.io/ko/v0.4.3/morph/>

### 1. 영문·한글 형태소 분석

#### 2) 한글 형태소 분석

##### (3) KoNLPy를 이용한 한글 형태소 분석

- morphs 함수를 이용한 형태소 분리

```
tag_list = tagger.morphs(cleaned_text)
print(tag_list)
```

```
['그', '밑', '에', '는', '계속', '배고픔', '을', '느끼', '세요',
```

- pos 함수를 이용한 형태소 분리 및 품사 태깅

```
pos_list = tagger.pos(cleaned_text)
print(pos_list)
```

```
[('그', 'MM'), ('밑', 'NNG'), ('에', 'JKB'), ('는', 'JX'),
```

- nouns 함수를 이용한 명사 추출

```
noun_list = tagger.nouns(cleaned_text)
print(noun_list)
```

```
['밑', '배고픔', '바보', '문구', '그', '마지막', '인사말', '배고픔']
```

### 2. 형태소 분석 사전 활용

#### 1) 신조어 형태소 분석

##### (1) 신조어 형태소 분석 방법

“최근 정부 주도 디지털뉴딜 사업인 **확장현실(XR)** 메타버스 프로젝트는 **증강현실(AR)** 플랫폼 분야의 주도 기업이 추진할 것으로 예상된다.”



확장현실(XR)

메타버스

증강현실(AR)

#### ▪ 메타버스 등과 같은 신조어 형태소 분석은?

##### ① 다양한 형태소 분석 방법 활용

- SoyNLP 및 ETRI 형태소 분석 Open API 등 활용

##### ② 형태소 분석기 사전 활용

- 형태소 분석기 사전에 신조어나 도메인 특화 용어 등록

##### ③ 모델 학습을 통해 신조어나 도메인 특화 용어 학습



### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (1) KoalaNLP 개요

- 한국어 형태소 및 구문 분석기 모음
  - KoNLPy, ETRI 형태소 분석, OpenAPI, Khaiii, Kiwi 등

##### (2) KoalaNLP 설치

- pip를 이용한 koalanlp 설치

```
!pip install koalanlp

Collecting koalanlp
  Downloading koalanlp-2.1.7-py3-none-any.whl (51 kB)
  |████████████████████████████████████████████████████████████████████████████████|
Collecting py4j~=0.10
  Downloading py4j-0.10.9.2-py2.py3-none-any.whl (198 kB)
  |████████████████████████████████████████████████████████████████████████████████|
Requirement already satisfied: requests~=2.22 in /usr/local/lib/python3.
Collecting kss~=2.5.1
  Downloading kss-2.5.1-py3-none-any.whl (65 kB)
  |████████████████████████████████████████████████████████████████████████████████|
```

2. 형태소 분석 사전 활용

2) KoalaNLP 패키지 활용

(3) KoalaNLP 지원 형태소 분석기

분석기별 접근키	설명	분석 범위
API.KMR	코모란 Wrapper	형태소
API.EUNJEON	은전한닢 Wrapper	형태소
API.ARIRANG	아리랑 Wrapper	형태소
API.RHINO	RHINO Wrapper	형태소
API.DAON	Daon Wrapper	형태소 2-3
API.KHAIIL	Khaiil Wrapper	형태소 2-4
API.UTAGGER	울산대 UTagger Wrapper	형태소, 문장분리
API.OKT	Open Korean Text Wrapper	형태소, 의존구문
API.KKMA	꼬꼬마 Wrapper	형태소, 의존구문
API.HNN	한나눔 Wrapper	문장분리, 형태소, 구문분석, 의존구문
API.ETRI	ETRI Open API Wrapper	형태소, 의존구문, 구문분석, 개체명, 의미역
API.KSS	KSS Wrapper	문장분리
API.KIWI	Kiwi Wrapper	형태소

※출처 : <https://pypi.org/project/koalanlp/>

### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (4) KoalaNLP 지원 형태소 분석기 사용

- import
  - 파이썬에서 사용하기 위한 모듈, 클래스, 함수 import
- initialize
  - 다운로드 및 자원 설정 초기화
- 형태소 분석
  - KoalaNLP 지원 형태소 분석기 사용
- finalize
  - 사용자원 반납 등 사용종료 처리

### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (5) 형태소 분석을 위한 코퍼스(Corpus)

```
text = """최근 정부 주도 디지털뉴딜 사업인 확장현실(XR)  
메타버스 프로젝트는 증강현실(AR) 플랫폼 분야의  
주도 기업이 추진할 것으로 예상된다."""
```

- 파이썬에서 사용하기 위한 모듈, 클래스, 함수 import

```
from koalanlp.Util import initialize, finalize  
from koalanlp.proc import *  
from koalanlp import API
```

- 다운로드 및 자원 설정 초기화

```
initialize(java_options="-Xmx4g -Dfile.encoding=utf-8",
```

### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (5) 형태소 분석을 위한 코퍼스(Corpus)

- KoalaNLP 지원 형태소 분석기 사용 - eunjeon

```
tagger = Tagger(API.EUNJEON)

tagged = tagger.tag(text)

print(tagged[0].singleLineString())
```

최근/NNG 정부/NNG 주도/NNG 디지털/NNG 뉴딜/NNG 사업  
/NNG+이/VCP+L/ETM

확장/NNG 현실/NNG (/SS XR/SL )/SS 메타/NNG 버스/NNG 프  
로젝트/NNG+는/JX

증강/NNG+현실/NNG

### 2. 형태소 분석 사전 활용


#### 2) KoalaNLP 패키지 활용

##### (5) 형태소 분석을 위한 코퍼스(Corpus)

- KoalaNLP 지원 형태소 분석기 사용 - ETRI
  - ETRI Key 발급 : 공공 인공지능 오픈 API·DATA 서비스 포털 접속 후 이메일 인증 및 키 발급 신청
  - [https://aiopen.etri.re.kr/key\\_main.php](https://aiopen.etri.re.kr/key_main.php)

Open API Key 발급 및 관리


Open API Key 발급 및 관리



키 발급 신청

API를 사용하기 위한 Key를 신청합니다.

키 발급 신청하기



키 관리

신청한 Key를 확인 및 비밀번호 재설정을 할 수 있습니다.

키 관리하기

- 이메일 입력 후 인증메일 받기
- 이메일로 수신된 인증키 입력 후 키 발급 신청

이메일 인증

\* 는 필수 항목입니다.

이메일 입력 \*

인증메일 받기

인증키 입력 \*

인증키 확인

인증메일 발송 후 받게 되시는 관리자 메일의 인증 키를

### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (5) 형태소 분석을 위한 코퍼스(Corpus)

- KoalaNLP 지원 형태소 분석기 사용 - ETRI

```
tagger = Tagger(API.ETRI, etri_key= '발급받은 키 입력')  
  
tagged = tagger.tag(text)  
  
print(tagged[0].singleLineString())
```

최근/NNG 정부/NNG 주도/NNG 디지털/NNG+뉴딜/NNG 사업  
/NNG+이/VCP+L/ETM  
확장/NNG+현실/NNG+(/SS+XR/SL+)/SS 메타버스/NNG 프로  
젝트/NNG+는/JX  
증강/NNG+현실/NNG+(/SS+AR/SL+)/SS

- eunjeon 형태소 분석기 사전 등록

```
from koalanlp.types import POS  
from koalanlp.proc import Dictionary  
  
dictionary = Dictionary(API.EUNJEON)  
  
dictionary.addUserDictionary(("메타버스", POS.NNG))  
dictionary.addUserDictionary(("확장현실", POS.NNG), ("메타버스", POS.NNG))  
  
print(dictionary.getItems())  
  
[( '메타버스', NNG), ( '확장현실', NNG)]
```

### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (5) 형태소 분석을 위한 코퍼스(Corpus)

- eunjeon 형태소 분석 수행

```
tagger = Tagger(API.EUNJEON)

tagged = tagger.tag(text)

print(tagged[0].singleLineString())
```

최근/NNG 정부/NNG 주도/NNG 디지털/NNG 뉴딜/NNG  
사업/NNG+이/VCP+ㄴ/ETM 확장현실/NNG (/SS XR/SL )/SS  
메타버스/NNG  
프로젝트/NNG+는/JX 증강/NNG+현실/NNG



### 2. 형태소 분석 사전 활용

#### 2) KoalaNLP 패키지 활용

##### (5) 형태소 분석을 위한 코퍼스(Corpus)

- 사용자원 반납 등 사용종료 처리를 위한 finalize

```
finalize()
```

```
True
```