

MeCab 설치하기

MeCab

Window에 설치하기

MeCab이란?

일본어 형태소 분석 엔진

➤ 은전한닢 프로젝트에서 MeCab 엔진이 일본어와 한국어의 유사점으로 한글 분석도 동작하는 것을 확인하고 한국어 형태소 분석기를 개발했다.

➤ “검색에서 쓸만한 오픈소스 한국어 형태소 분석기를 만들자!” 라는 생각으로 시작된 프로젝트로

- 검색 엔진 색인기에서 쓸 수 있는 적당한 품질과 속도
- 자유로운 라이선스

➤ 이 두가지에 초점을 맞추어 개발된 프로젝트



다양한 언어에 대응이 가능하다.
(python, java, c#, ruby, perl)

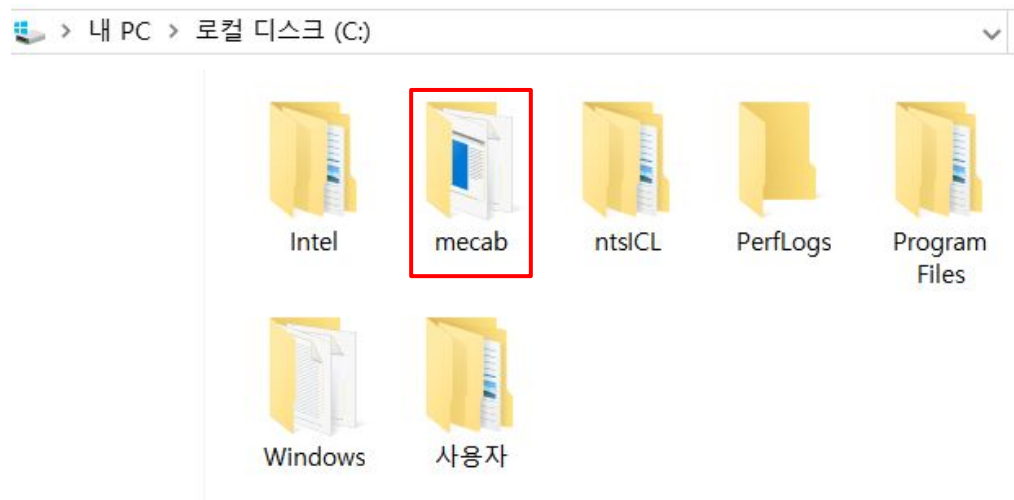
MeCab 설치하기

Window에 설치하기

설치 방법

참고 사이트 > <https://cleancode-ws.tistory.com/97>

1. c:/mecab 폴더 생성 (c 드라이브에 mecab 폴더를 만든다.)



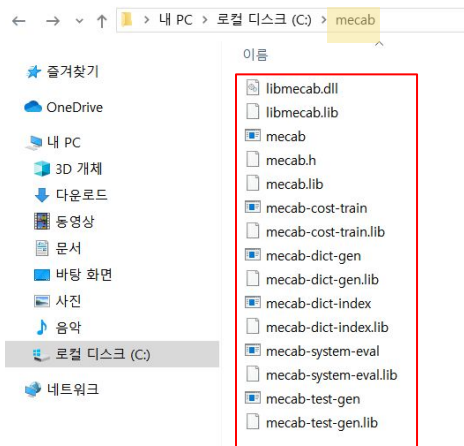
설치 방법

2. mecab-ko-msvc 설치

(참고) ‘mecab-ko-msvc’는 C기반으로 만들어진 mecab이 윈도우에서 실행될 수 있게 한다.

<https://github.com/Pusnow/mecab-ko-msvc/releases/tag/release-0.9.2-msvc-3>

> mecab-ko-msvc-x64 : 다운로드 후 압축 풀어서 mecab 폴더에 넣는다. (컴퓨터 사양에 따라 선택)



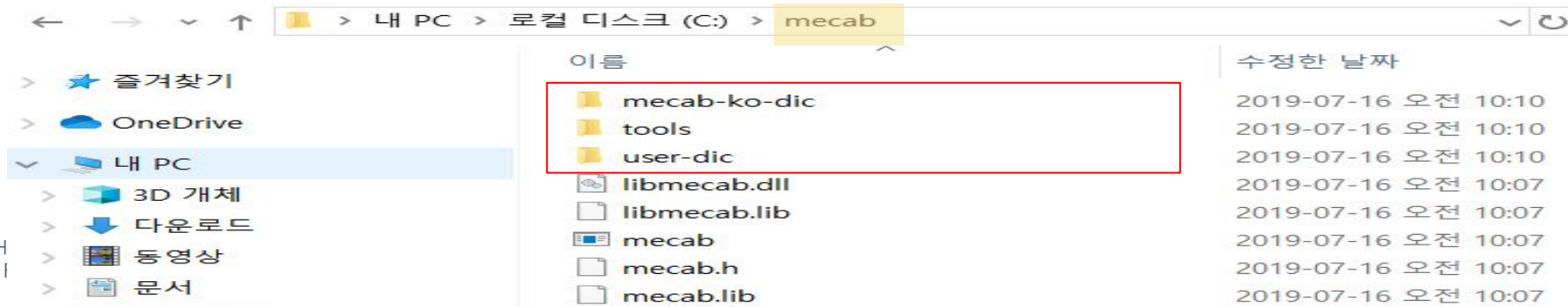
설치 방법

3. mecab-ko-dic-msvc 설치

(참고) 'mecab-ko-dic' 기본 한글 사전

<https://github.com/Pusnow/mecab-ko-dic-msvc/releases/tag/mecab-ko-dic-2.1.1-20180720-msvc>

> mecab-ko-dic-msvc.zip 다운로드 후 압축 풀어서 mecab 폴더에 넣는다.



설치 방법

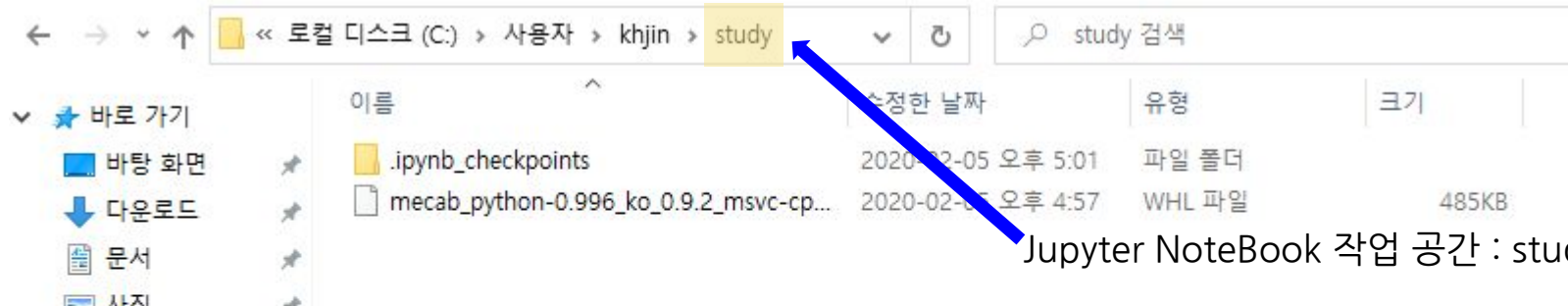
4. python wheel 설치

(참고) 'python whl' : 파이썬 패키지를 windows 환경에서 설치하기 위한 패키지 설치파일

https://github.com/Pusnow/mecab-python-msvc/releases/tag/mecab_python-0.996_ko_0.9.2_msvc-2

> 각자의 파이썬 버전에 맞는 파일 다운 후

현재 주피터 노트북이 실행되고 있는 폴더에 넣는다.



Jupyter Notebook 작업 공간 : study

설치 방법

```
In [1]: !pip install mecab_python-0.996_ko-0.9.2_msvc-cp37-cp37m-win_amd64.whl

Processing c:\#users\#khjin\#study\#mecab_python-0.996_ko-0.9.2_msvc-cp37-cp37m-win_amd64.whl
Installing collected packages: mecab-python
Successfully installed mecab-python-0.996-ko-0.9.2-msvc
```

```
In [2]: # 설치 확인
import MeCab
m = MeCab.Tagger()
out = m.parse("MeCab이 잘 설치 되었는지 확인중입니다.")
print(out)

MeCab SL,*,*,*,*,*,*,*
이 JKS,*,F,이,*,*,*,*
잘 MAG,*,T,잘,*,*,*,*
설치 NNG,행위,F,설치,*,*,*,*
되 XSV,*,F,되,*,*,*,*
었 EP,*,T,었,*,*,*,*
는지 EC,*,F,는지,*,*,*,*
확인 NNG,행위,T,확인,*,*,*,*
중 NNB,*,T,중,*,*,*,*
입니다 VCP+EF,*,F,입니다,Inflect,VCP,EF,이/VCP/**입니다/EF/*
. SF,*,*,*,*,*,*,*
EOS
```

KoNLPy 설치하기

KoNLPy

Window에 설치하기

KoNKPyl란?

- 코엔엘파이
- 한국어 정보처리를 위한 파이썬 패키지
- 다양한 형태소 분석, 태깅 라이브러리를 파이썬에서 쉽게 사용할 수 있도록 모아 놓은 패키지
- 한국어 말뭉치 제공 kolaw (대한민국 헌법 말뭉치)
 kobill (국회법안 말뭉치)

KoNKPy란?

다양한 형태소 분석, 태깅 라이브러리를 파이썬에서 쉽게 사용할 수 있도록 모아 놓은 패키지

- Hannanum(한나눔) : KAIST Semantic Web Research Center에서 개발
<http://semanticweb.kaist.ac.kr/hannanum/>
- Kkma(꼬꼬마) : 서울대학교 IDS(Intelligent Data Systems) 연구실에서 개발
<http://kkma.snu.ac.kr/>
- Komoran(코모란) : Shineware에서 개발
<https://github.com/shin285/KOMORAN>
- Mecab(메카브) : 일본어용 형태소 분석기를 한국어를 사용할 수 있도록 수정
<https://bitbucket.org/eunjeon/mecab-ko>
- Open Korean Text (Okt): 오픈 소스 한국어 분석기 . 과거 트위터 형태소 분석기
<https://github.com/open-korean-text/open-korean-text>
- Twitter(트위터) :

빠른 분석이 중요할 때 : 트위터

정확한 품사 정보가 필요할 때 : 꼬꼬마

정확성, 시간 모두 중요할 때 : 코모란

KoNLPy 설치하기

Window에 설치하기

KoNKPY 사용방법

다양한 형태소 분석, 태깅 라이브러리를 파이썬에서 쉽게 사용할 수 있도록 모아 놓은 패키지

```
from konlpy.tag import Hannanum, Kkma, Komoran, Mecab, Okt, Twitter
hannanum = Hannanum()
kkma = Kkma()
komoran = Komoran()
mecab = Mecab()
okt = Okt()
twitter = Twitter()
```

위의 패키지가 공통적으로 사용할 수 있는 클래스

- nouns : 명사 추출
- morphs : 형태소 추출
- pos : 품사 부착

설치 방법

참고 사이트 > <https://cleancode-ws.tistory.com/95?category=848890>

1-1. JDK 설치하기 : <https://www.oracle.com/java/technologies/javase-downloads.html>

> JDK download

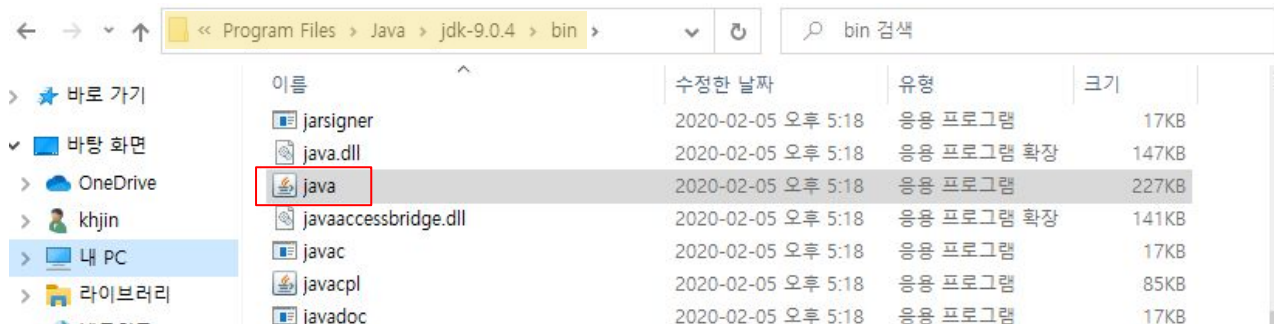
설치방법

1-2. JDK 환경변수 추가

> 내컴퓨터 > 속성 > 고급 시스템 설정 > 환경변수 > 시스템 변수(S) > Path

현재 java가 설치되어 있는 폴더 경로 추가

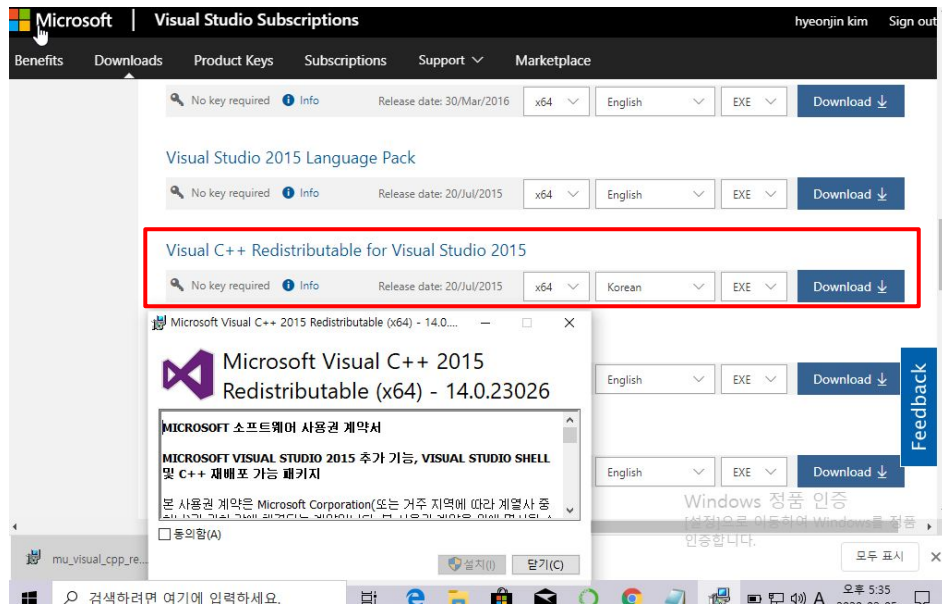
(참고) java 환경변수 추가 후, 아나콘다를 모두 종료하고 재부팅 해주어야 함



설치 방법

2. Visual studio 설치 : <https://visualstudio.microsoft.com/ko/vs/older-downloads/>

> 15년도 버전 설치

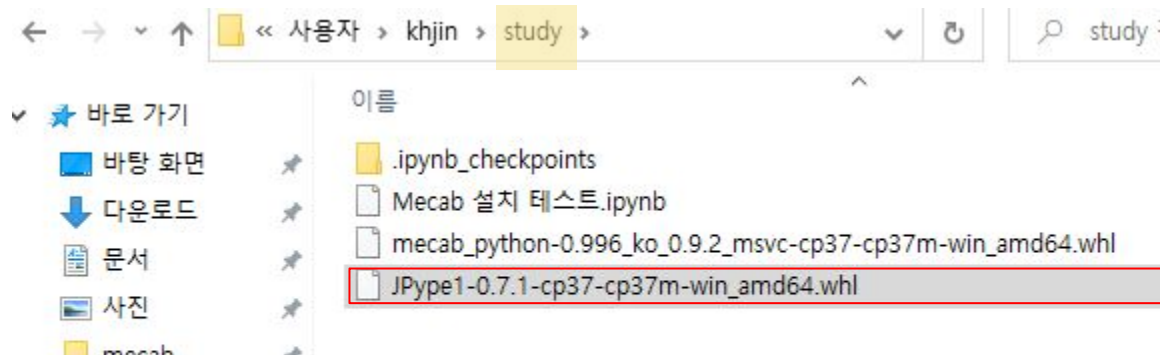


설치 방법

3. JPyte1 다운 : <https://www.lfd.uci.edu/~gohlke/pythonlibs/#jpyte>

> 각자의 파이썬 버전에 맞는 파일을 다운 받은 후

현재 주피터 노트북이 실행되고 있는 폴더에 넣는다.



설치 방법

4. JPyep1 설치

쥬피터 노트북의 파이썬 파일을 하나 만들어 설치

> !pip install JPyep1-0.7.1-cp37-cp37m-win_amd64.whl

JPyep1 설치

```
In[4]: !pip install JPyep1-0.7.1-cp37-cp37m-win_amd64.whl  
  
Processing c:\users\khjin\study\jpyep1-0.7.1-cp37-cp37m-win_amd64.whl  
Installing collected packages: JPyep1  
Successfully installed JPyep1-0.7.1
```

설치 방법

5. KoNLPy 설치

쥬피터 노트북의 파이썬 파일을 하나 만들어 설치

> !pip install konlpy

```
In [5]: !pip install konlpy
```

```
Collecting konlpy
  Downloading konlpy-0.5.2-py2.py3-none-any.whl (19.4 MB)
Collecting BeautifulSoup4==4.6.0
  Downloading BeautifulSoup4-4.6.0-py3-none-any.whl (86 kB)
Collecting numpy>=1.6
  Downloading numpy-1.18.1-cp37-cp37m-win_amd64.whl (12.8 MB)
Requirement already satisfied: JPype1>=0.7.0 in c:\users\khjin\appdata\local\continuum\anaconda3\envs\test\lib\site-packages (from konlpy) (0.7.1)
```



설치 방법

6. KoNLPy 설치 확인

```
In [1]: # 설치 확인
from konlpy.tag import Kkma
Kkma_pos = Kkma()
K_nouns = Kkma_pos.nouns("지금부터 코앤엘파이 한국어 형태소 분석기 설치를 확인해 봅니다.")
print(K_nouns)

['지금', '코', '코앤엘파이', '앤', '엘', '파이', '한국어', '형태소', '분석기', '설치', '확인']
```

여기까지 완료 되었으면, 영문, 한글 전처리 실습을 시작해 봅시다

MeCab과 KoNLPy 설치 완료

eKoNLPy 설치하기

eKoNLPy 설치하기

Window에 설치하기

eKoNLPy란?

eKoNLPy란 한국어를 지원하는 경제분석을 위한 한국어 NLP library이다.

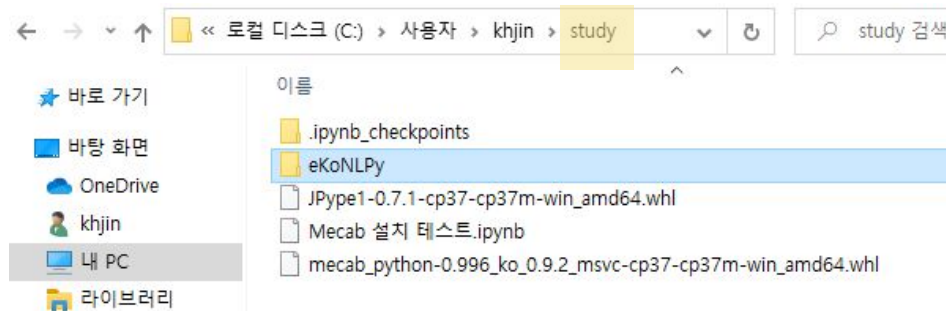
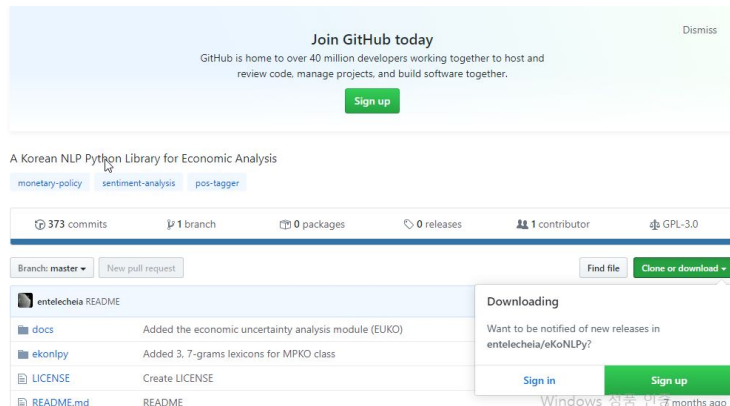
- KoNLPy의 MeCab tagger를 기반으로 경제관련 전문용어, 금융기관, 기업명 등을 하나의 명사로 분류하도록 후처리 기능을 추가한 라이브러리이다.
- 통화정책(Monetary Policy)의 어조(Hawkish/Dovish)를 판단할 수 있는 Sentiment Analysis 기능이 포함되어 있으며 경제문서의 주제를 분류할 수 있는 Topic Analysis기능이 포함되어 있다.

(참고) <https://github.com/entelecheia/eKoNLPy>

설치 방법

1. eKoNLPy 다운 받기 : <https://github.com/entelecheia/eKoNLPy>

> 다운 받은 후, 압축 풀고 현재 사용하는 জু피터 노트북 폴더에 넣기



Page 10 of 10

> KoNLPy : 0.4.4

gensim : 3.1.0

scipy : 0.19.1

numpy : 1.13

설치 방법

3. Anaconda Prompt에서 ekonlpy 설치하기

Anaconda Prompt > cd study > cd ekonlpy > pip install .

Anaconda Prompt (anaconda3)

```
(base) C:\Users\khjin>conda activate test  
(test) C:\Users\khjin>cd study  
(test) C:\Users\khjin\study>cd ekonlpy  
(test) C:\Users\khjin\study\ekonlpy>pip install .
```

만약 Anaconda의 작업 환경이 base가 아니라면
현재 작업 중인 환경으로 이동 후 설치한다.

> base에서 작업하고 있으면 이 단계는
생략 해도 좋다.

> 만약 이렇게 했을때 설치가 안된다면, “pip install . --upgrade”를 한다.

설치 방법

4. eKoNLPy 설치 확인

```
In [1]: from ekoNLPy.tag import Mecab
mecab = Mecab()
mecab.pos('금통위는 따라서 물가안정과 병행, 경기상황에 유의하는 금리정책을 펼쳐나가기로 했다고 밝혔')

Out[1]: [('금통위', 'NNG'),
          ('은', 'JX'),
          ('따라서', 'MAJ'),
          ('물가', 'NNG'),
          ('안정', 'NNG'),
          ('과', 'JC'),
          ('병행', 'NNG'),
          ('.', 'SC'),
          ('경기', 'NNG'),
          ('상황', 'NNG'),
          ('에', 'JKB'),
          ('유의', 'NNG'),
          ('하', 'XSV'),
          ('은', 'ETM'),
          ('심리', 'NNG'),
          ('금리정책', 'NNG'),
          ('을', 'JKO'),
          ('펼쳐', 'VV'),
          ('나', 'VX'),
          ('기', 'ETN'),
          ('로', 'JKB'),
          ('했', 'VV'),
          ('다고', 'EC'),
          ('밝혔', 'VV'),
          ('다', 'EF'),
          ('.', 'SF')]
```

오류를 해결해 봅시다

오류 해결 1

1. 접근 경로 오류

```
-----
RuntimeError                                Traceback (most recent call last)
~\Anaconda3\envs\khj\lib\site-packages\konlpy\tagger\_mecab.py in __init__(self, dicpath)
    107         try:
--> 108             self.tagger = Tagger('-d %s' % dicpath)
    109             self.tagset = utils.read_json('%s/data/tagset/mecab.json' % utils.installpath)

~\Anaconda3\envs\khj\lib\site-packages\MeCab.py in __init__(self, *args)
    308     def __init__(self, *args):
--> 309         this = _MeCab.new_Tagger(*args)
    310         try: self.this.append(this)
```

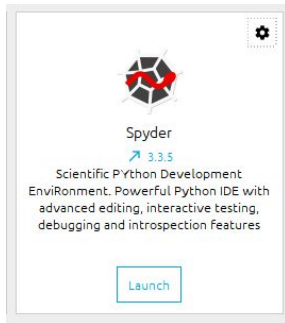
Exception: The MeCab dictionary does not exist at "/usr/local/lib/mecab/dic/mecab-ko-dic". Is the dictionary correctly installed?
 You can also try entering the dictionary path when initializing the Mecab class: "Mecab('/some/dic/path')"

➤ 해당 경로로 찾아 들어가 mecab dictionary가 설치된 경로로 변경해 준다.

오류 해결 1

해결방법

- 1) 아나콘다의 Spyder를 사용하면 편리하게 변경 할 수 있다. Spyder를 실행하여 오류가 발생한 경로의 파일을 열어준다.



오류 해결 1

해결방법

- 2) MeCab dict를 처음 여는 path를 찾아서 r'C:\mecab\mecab-ko-dic'로 변경

```
104
105 def __init__(self, dicpath=r'C:\mecab\mecab-ko-dic'):
106     self.dicpath = dicpath
107     try:
108         self.tagger = Tagger('-d %s' % dicpath)
109         self.tagset = utils.read_json('%s/data/tagset/mecab.json' % utils.installpath)
110     except RuntimeError:
111         raise Exception('The MeCab dictionary does not exist at "%s". Is the dictionary correctly insta
112     except NameError:
113         raise Exception('Install MeCab in order to use it: http://konlpy.org/en/latest/install/')
114
```

오류 해결 2

2. 인코딩 오류 >> encoding=utf-8

1) 오류가 난 곳의 경로를 확인하고, spyder에서 해당 파일을 열어 수정한다.

```
~\AppData\Local\Continuum\anaconda3\envs\test\lib\site-packages\ekonlpy\utils.py in load_vocab(file_path, delimiter)
    35     if os.path.isfile(file_path):
    36         with open(file_path) as f:
--> 37             for i, line in enumerate(f):
    38                 if delimiter in line:
    39                     w = line.strip().split(delimiter)

UnicodeDecodeError: 'cp949' codec can't decode byte 0x80 in position 3: illegal multibyte sequence
```

(참고) 경로

site-packages > ekonlpy > utils.py

오류 해결 2

2) 앞의 경로로 들어가서 encoding = 'utf-8'을 추가 한다.

```
33 def load_vocab(file_path, delimiter=','):
34     vocab = {}
35     if os.path.isfile(file_path):
36         with open(file_path, encoding='utf-8') as f:
37             for i, line in enumerate(f):
38                 if delimiter in line:
39                     w = line.strip().split(delimiter)
40                     vocab[w[0].lower().replace(' ', '')] = w[1].lower().replace(
41                 else:
42                     save_vocab(vocab, file_path)
43                 # print('Loaded the file: {}, No. of words: {}'.format(file_path, len(vocab))
44                 vocab = OrderedDict((k, v) for k, v in sorted(vocab.items(), key=lambda x: x
45                 return vocab
```

저장 후, 설치 확인 코드를 다시 실행해 본다.
(단, 아나콘다를 모두 종료 후 다시 실행한다.)

ekonlpy 설치 완료

이제 프로젝트를 시작해 봅시다

Colab에서 MeCab 설치하기

Colab이란?

설치방법 1

아래의 코드를 입력하면 된다

```
!sudo apt-get install g++ openjdk-7-jdk # Install Java 1.7+
!sudo apt-get install python-dev; pip install konlpy      # Python 2.x
!sudo apt-get install python3-dev; pip3 install konlpy    # Python 3.x
!sudo apt-get install curl
!bash <(curl -s https://raw.githubusercontent.com/konlpy/konlpy/master/scripts/mecab.sh)
```

설치 완료!

<설치 확인>

```
from konlpy.tag import Mecab
mecab = Mecab()
```