

한글 형태소 분석기를 설치해 봅시다

핀인사이트
데이터분석가 김현진

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)

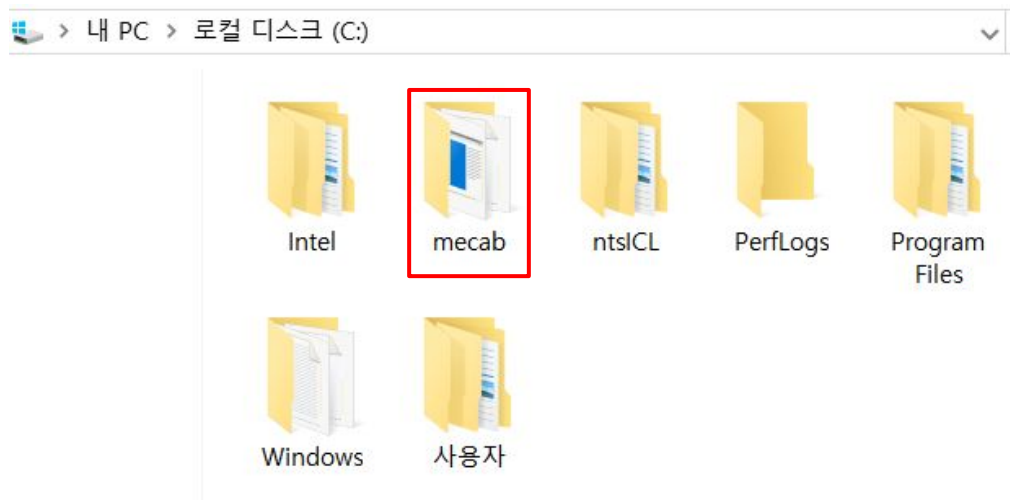
MeCab 설치하기

Window에 설치하기

설치 방법

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

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



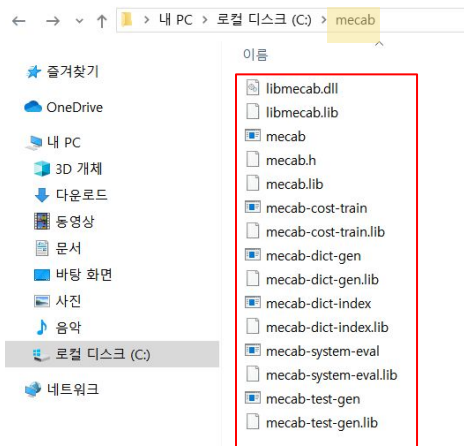
설치 방법

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)



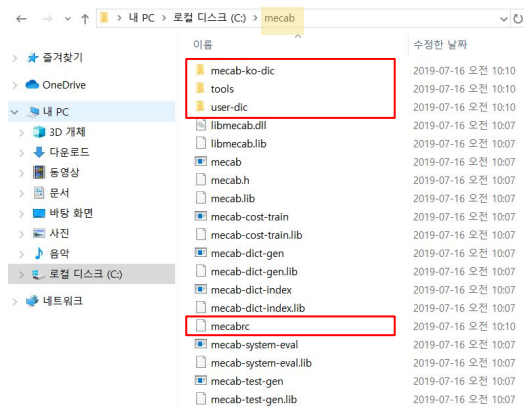
설치 방법

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

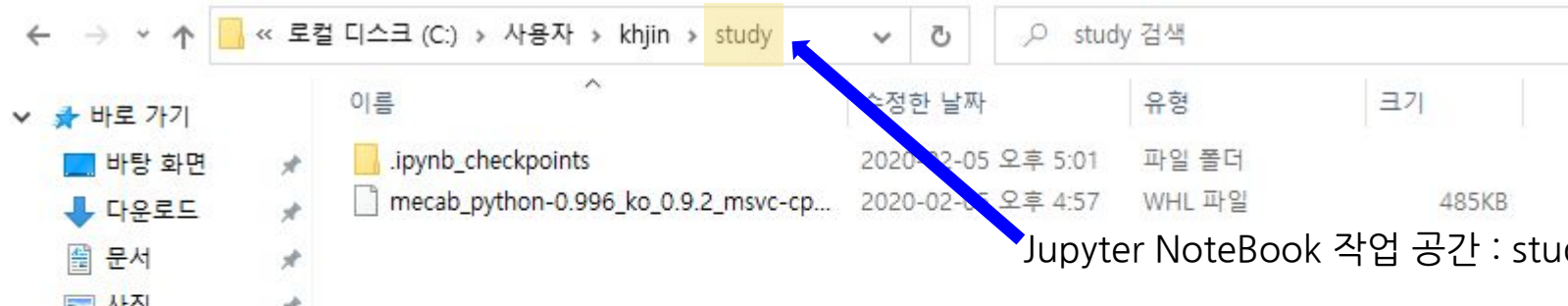
4. python wheel 설치

(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

5. 주피터 노트북의 파이썬 파일을 하나 만들어 설치

> !pip install mecab_python-0.996_ko_0.9.2_msvc-cp37-cp37m-win_amd64.whl

```
In [1]: !pip install mecab_python-0.996_ko_0.9.2_msvc-cp37-cp37m-win_amd64.whl
Processing c:\Users\Wkhjin\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,*,*,*,*,*,*
미 JKS,*,F,미,*,*,*,*
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,미/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

Java SE Downloads



Java Platform, Standard Edition








Java SE 14

Java SE 14.0.2 is the latest release for the Java SE Platform

- Documentation
- Installation Instructions
- Release Notes
- Oracle License
 - Binary License
 - Documentation License
- Java SE Licensing Information User Manual
 - Includes Third Party Licenses
- Certified System Configurations
- Readme

Oracle JDK

-  **JDK Download**
-  Documentation Download

Product / File Description	File Size	Download
Linux Debian Package	15793 MB	 jdk-14.0.2_linux-x64_bin.deb
Linux RPM Package	165.06 MB	 jdk-14.0.2_linux-x64_bin.rpm
Linux Compressed Archive	182.06 MB	 jdk-14.0.2_linux-x64_bin.tar.gz
macOS Installer	176.37 MB	 jdk-14.0.2_osx-x64_bin.dmg
macOS Compressed Archive	176.79 MB	 jdk-14.0.2_osx-x64_bin.tar.gz
Windows x64 Installer	162.11 MB	 jdk-14.0.2_windows-x64_bin.exe
Windows x64 Compressed Archive	181.56 MB	 jdk-14.0.2_windows-x64_bin.zip

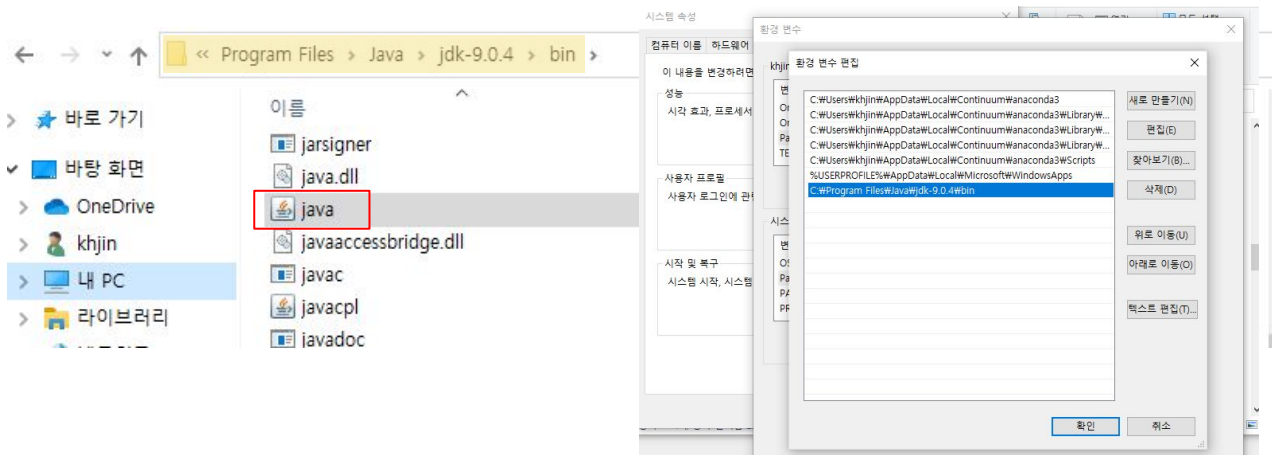
설치방법

1-2. JDK 환경변수 추가

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

현재 java가 설치되어 있는 폴더 경로 추가

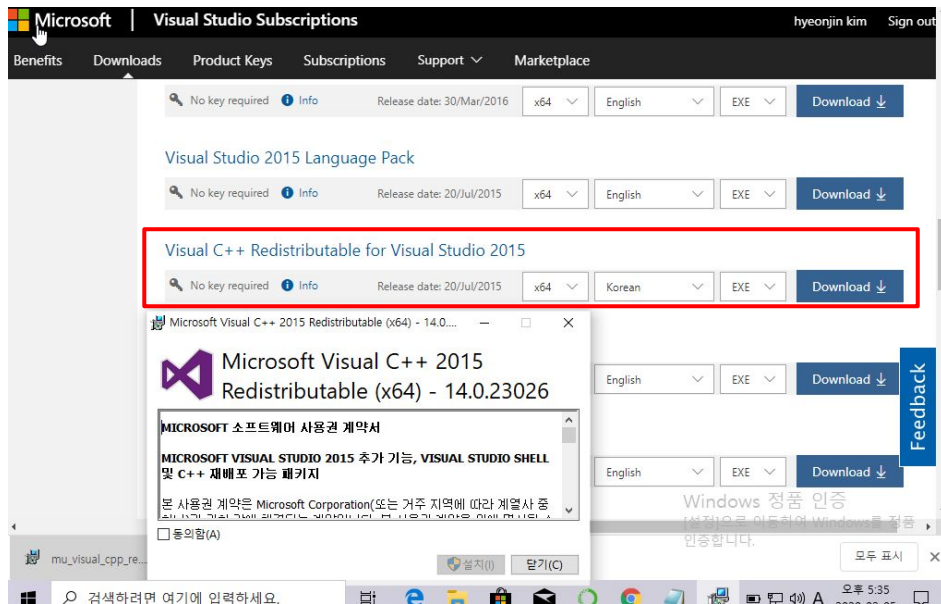
(참고) java 환경변수 추가 후, 아나콘다를 모두 종료하고 재부팅 해주어야 함



설치 방법

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

> 15년도 버전 설치



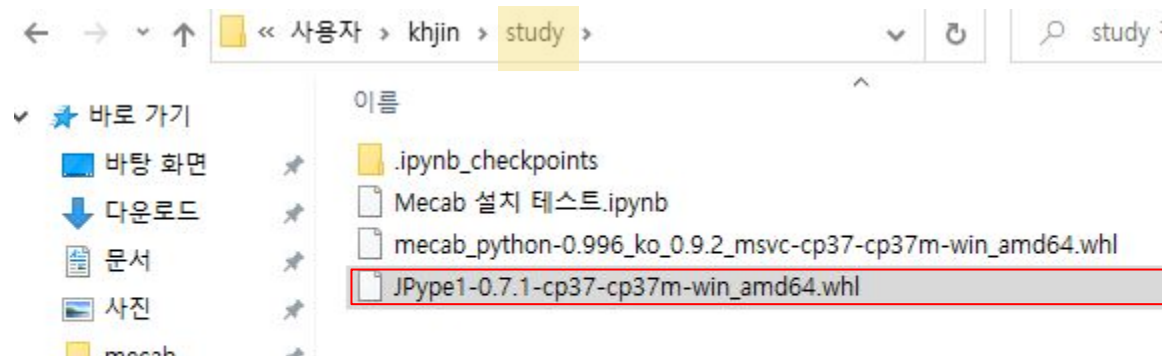
microsoft 계정에 로그인해야 다운받을수 있음.

설치 방법

3. JPype1 다운 : <https://www.lfd.uci.edu/~gohlke/pythonlibs/#jpype>

> 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