

# Lifebloom

About R, Python, SAS, Machine Learning, Data Mining and miscellaneous things

[홈](#) [Profile](#) [Contact](#) [R](#) [Python](#) [Visualization](#) [misc](#)

## Python을 활용한 텍스트 마이닝 1.모듈설치

이번 포스팅에서는 Python을 활용하여 웹페이지에서 텍스트를 추출하고 이를 워드클라우드로 나타내고 텍스트들간의 유사도를 활용해 관련성을 구해보겠습니다.

파이썬 텍스트 마이닝은 간단한 편이지만 몇 가지 모듈을 설치해야 합니다.

웹 크롤링을 위해 BeautifulSoup4, 한국어 텍스트 분석을 위한 KoNLPy, 워드 클라우드를 위한 pygame, simplejson, pytagcloud 가 필요합니다.

```
>pip install beautifulsoup4
```

cmd라인에서 다음과 같이 입력하여 먼저 beautifulsoup4 모듈을 설치하시고

<http://www.lfd.uci.edu/~gohlke/pythonlibs/#jpype> 로 가서 자신의 세팅에 맞는 JPype1-0.x.x-cp36-...whl 파일을 받으시면 됩니다. 글 작성 기준 버전은 0.6.2인데 최신버전으로 받으시면 됩니다. 참고로 KoNLPy를 설치하시기 전에 Java가 설치되어있는지 확인하시고 없다면 설치해주시면 됩니다.

```
>pip install JPype1 ..... .whl
```

## Search



## Recently posted

[Recommendation System 1.](#) 5월 13, 2018

[Regression with Machine Learning 4. Regularization for sparsity\(희소 학습\)](#) 4월 8, 2018

[경사 하강법\(Gradient Descent\)](#) 3월 27, 2018

[Regression with Machine Learning 3. Constrained Least Squares\(제약 최소제곱\)](#) 3월 18, 2018

[Regression with Machine Learning 2. Stochastic Gradient Descent\(확률적 경사법\)](#) 3월 18, 2018

## Posts

[2018년 5월](#) (1)

[2018년 4월](#) (1)



cmd라인에서 파일을 받은 경로로 이동하신 뒤(cd 폴더명), 받은 파일의 이름을 install 뒤에 입력하시면 모듈이 설치됩니다.

```
>pip install pygame
>pip install simplejson
>pip install pytagcloud
```

cmd라인에 3개의 모듈을 설치하시면 됩니다.

다음 포스팅에서는 설치된 모듈을 통해 웹페이지에서 크롤링한 텍스트를 시각화 해보겠습니다.

✎ kis0403   📅 7월 4, 2017   📁 installation   💬 댓글 한 개  
✎ 편집

← 2017 날씨 빅데이터 콘테스트 1. 데이터 전처리

Python을 활용한 텍스트 마이닝 2. 텍스트 분석-데이터 수집 →

## "Python을 활용한 텍스트 마이닝 1. 모듈설치"에 대한 1개의 생각

ender info \_\_\_\_\_ by CleanTalk  
mon@naver.com | 203.230.106.85 | Mark as spam



✎ giraffemon  
📅 11월 9, 2017 11:17 오전  
📍 고유주소  
✎ 고치기

잘배우고있습니다! 텍스트마이닝 관련한 포스트들을 배우고 저도 포스팅을 해보려하는데 참고하고 출처를 남겨도 될까요?

↩️ 응답

2018년 3월 (5)

2018년 1월 (1)

2017년 9월 (2)

2017년 8월 (6)

2017년 7월 (12)

2017년 6월 (7)

Etc

사이트 관리

로그아웃

글 RSS

댓글 RSS

WordPress.org

category

misc (2)

Python (17)

coding with python (1)

installation (5)

Neural Network (1)

Text Mining (10)

R (14)

2017 Weather Contest (4)

machine learning (5)

Packages & Base (2)

Recommendation System (2)

Visualization (2)



Manage

# 답글 남기기

[kis0403로\(으로\) 로그인 함.](#) [로그아웃?](#)

댓글

댓글 달기

