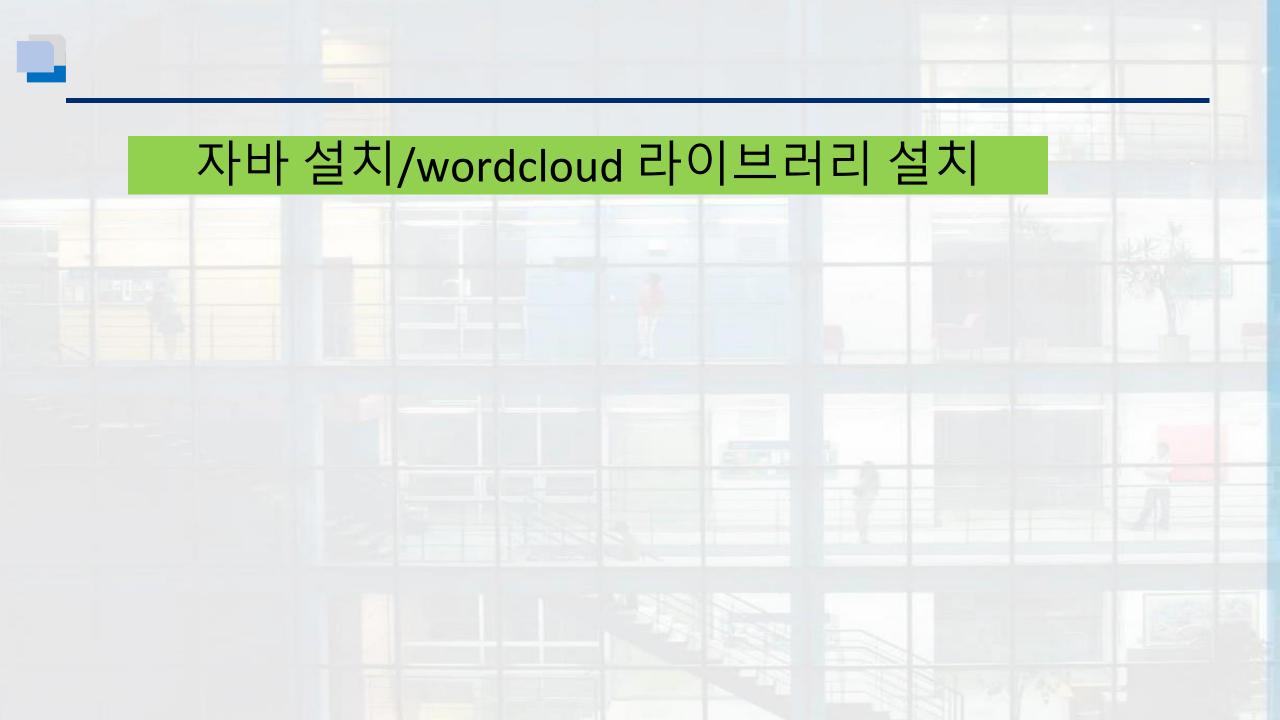




## 오늘의 학습

## 학습목표

- 자바설치/wordcloud 설치
- txt 파일 읽고 쓰기
- 워드클라우드
- 형태소 분석기
- 워드클라우드 코딩없이 만들기

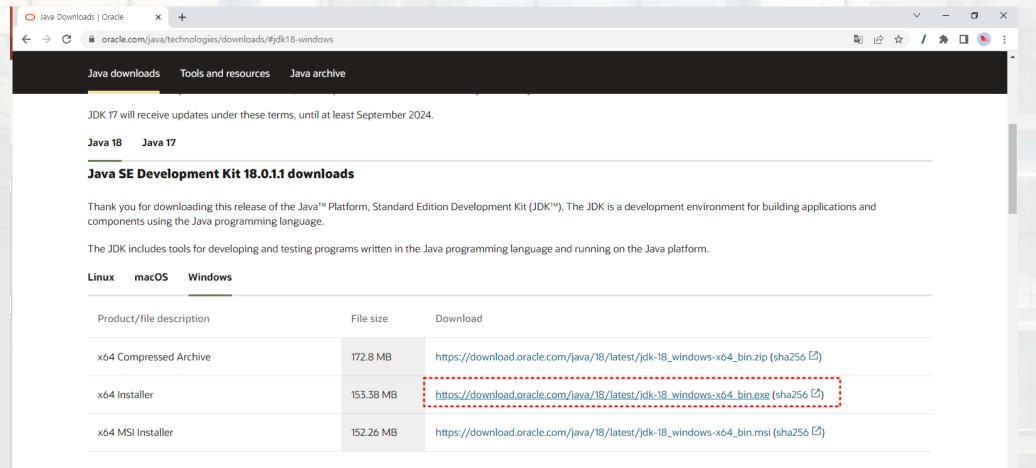




## 자바 설치하기\_windows

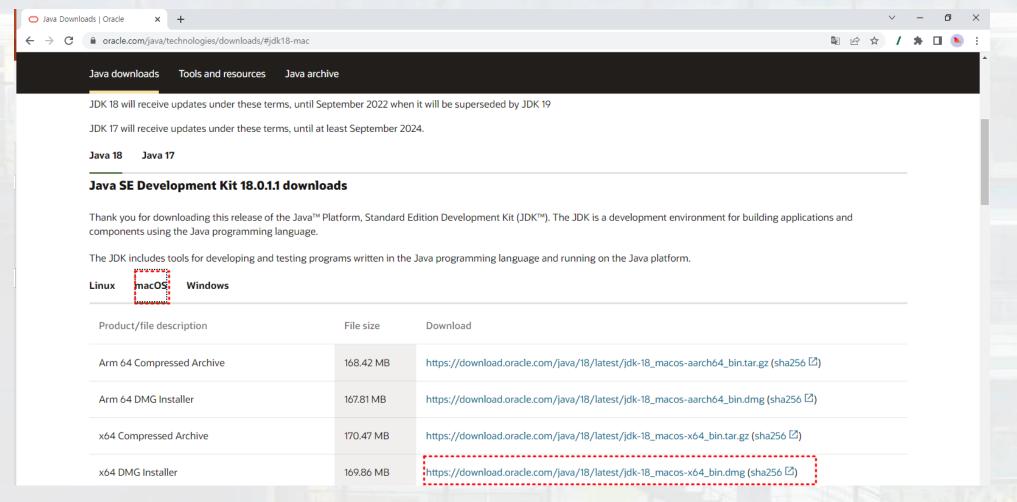
#### JDK설치하기

## https://www.oracle.com/java/technologies/downloads/#jdk18-windows



## 자바 설치하기\_maxOS [설치참고https://jsikim1.tistory.com/191

## JDK설치하기 **버전 1.7이상의 자바가 설치되어 있어야 함**



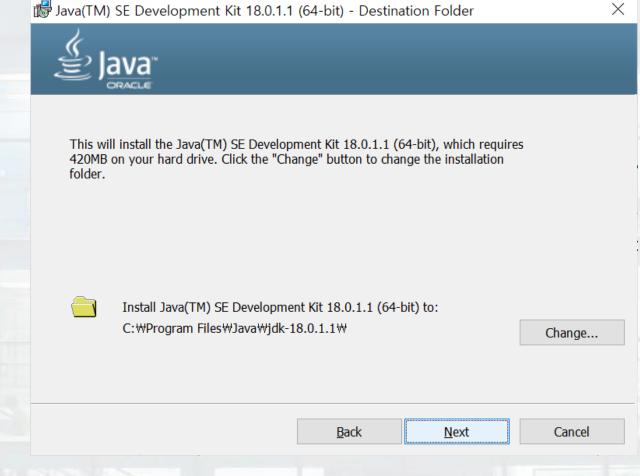


## 자바 설치



Cancel

Next >





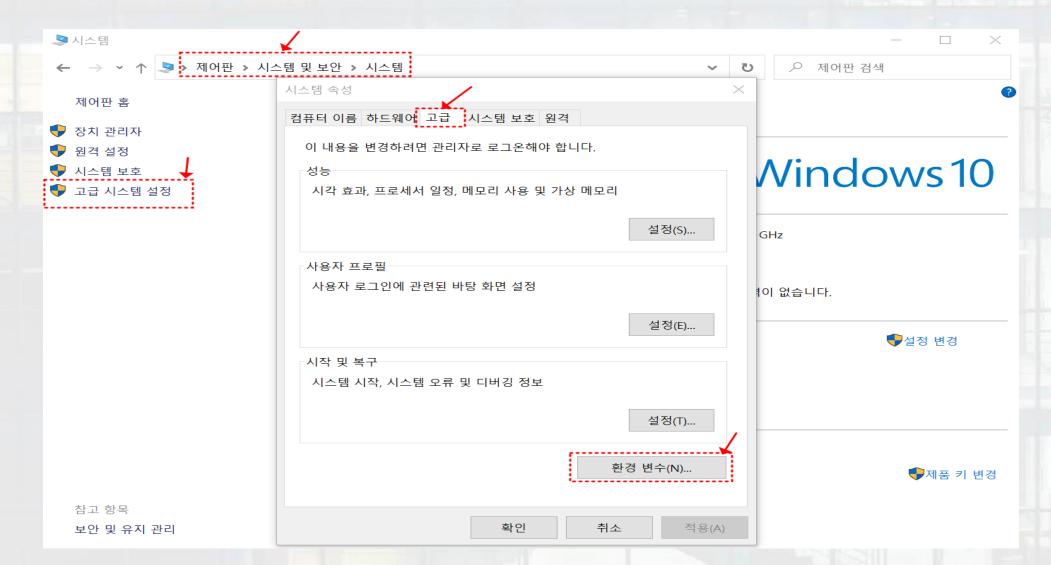
## 자바 설치



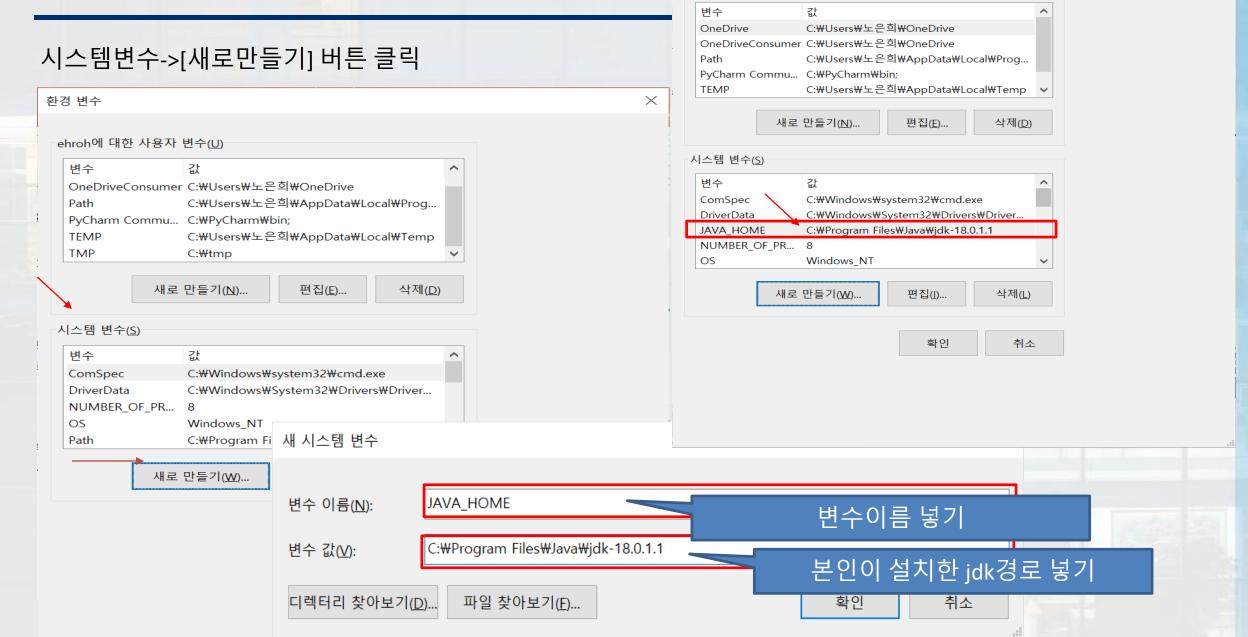


## 환경변수 설치하기

#### [제어판]->[시스템및 보안]->[시스템]->고급 시스템 설정->고급 ->환경변수

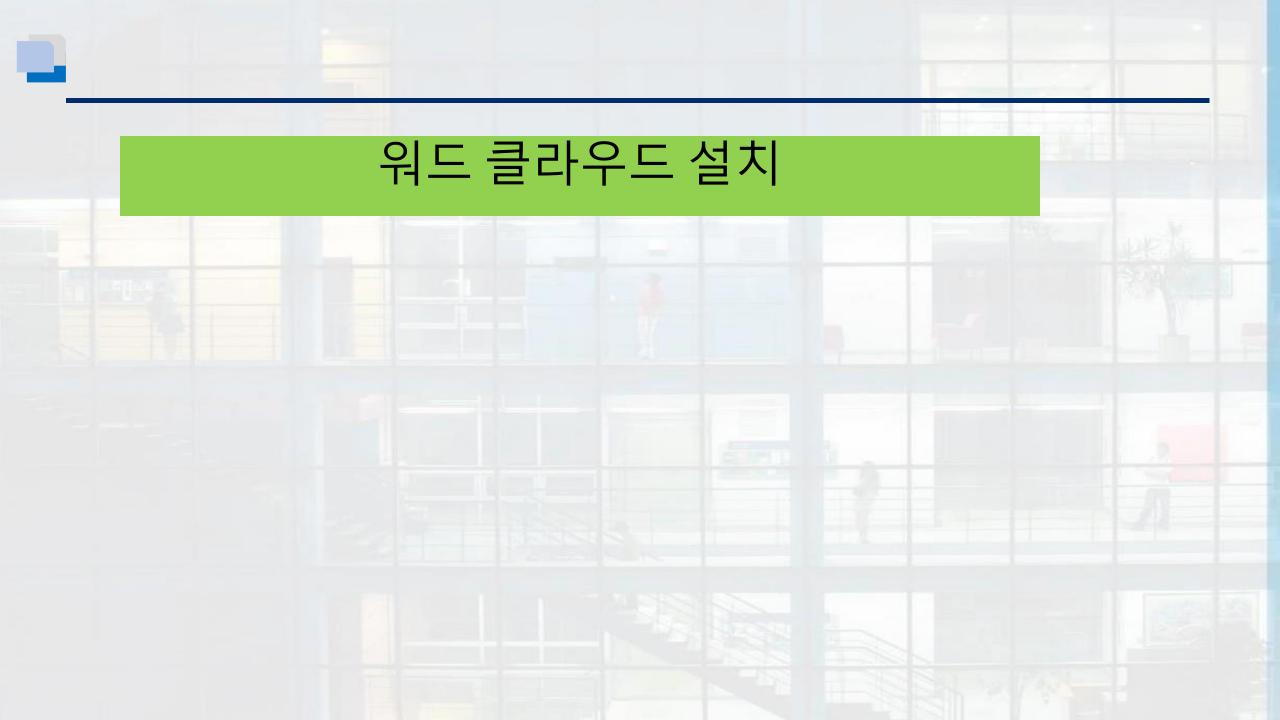


## 환경변수 설치하기



환경 변수

ehroh에 대한 사용자 변수(U)





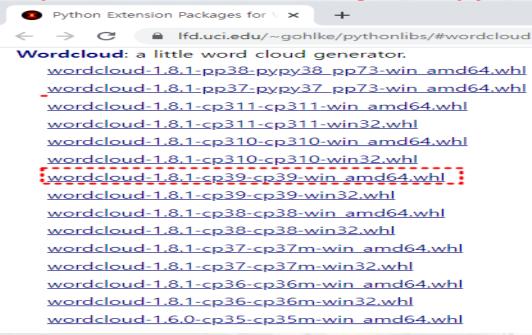
## wordcloud 설치

#### pip install wordcloud 만약 에러가 발생하면

```
| Associate Powershell Prompt (Anacondas) | Associated Prompt
```

아래 링크로 들어가서 wordcloud whl 파일을 설치된 파이썬 버전 확인 후 다운받기

## https://www.lfd.uci.edu/~gohlke/pythonlibs/#wordcloud

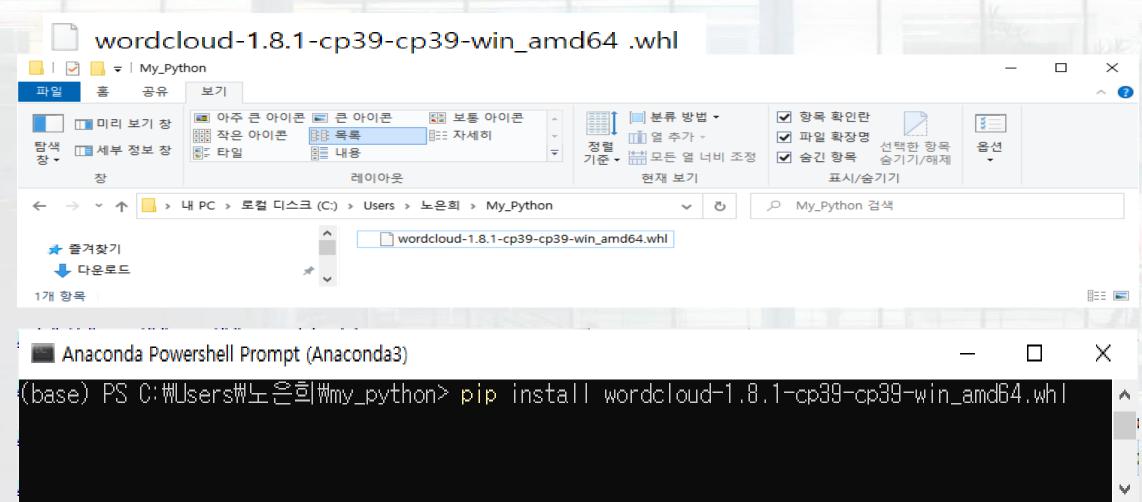


```
■ Anaconda Powershell Prompt (Anaconda3) - □ X

(base) PS C:\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Use
```

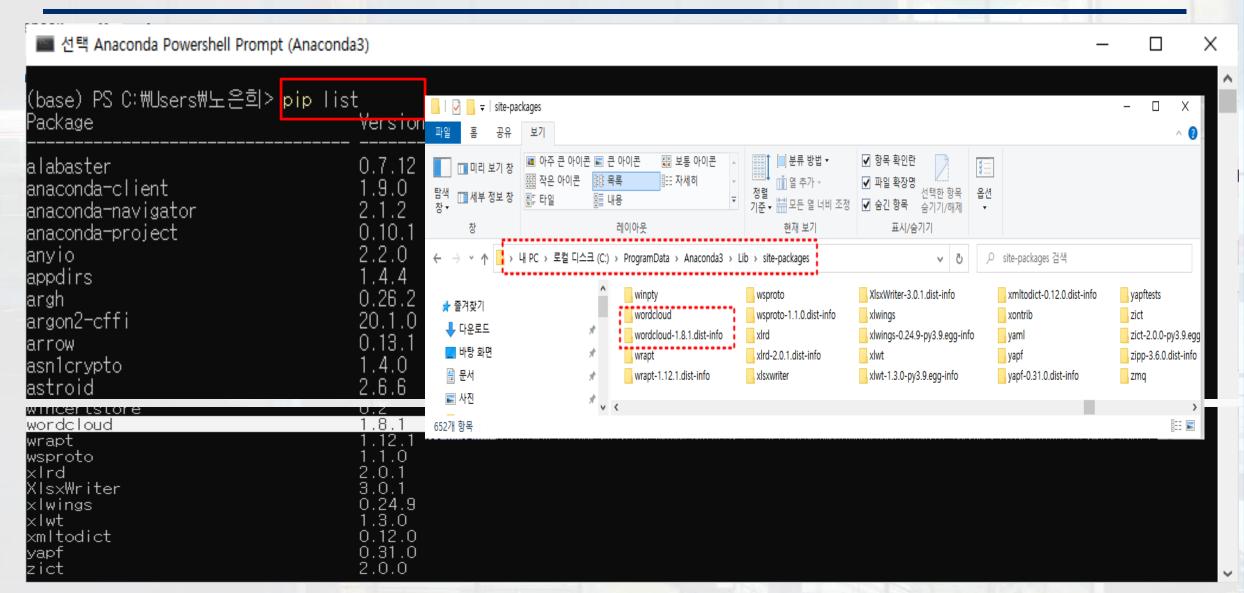
## wordcloud 설치

다운받은 폴더에서(whl파일이 있는 경로) cmd 창에 'pip install 파일명'을 해주면 된다. pip install wordcloud-1.8.1-cp39-cp39-win\_amd64.whl

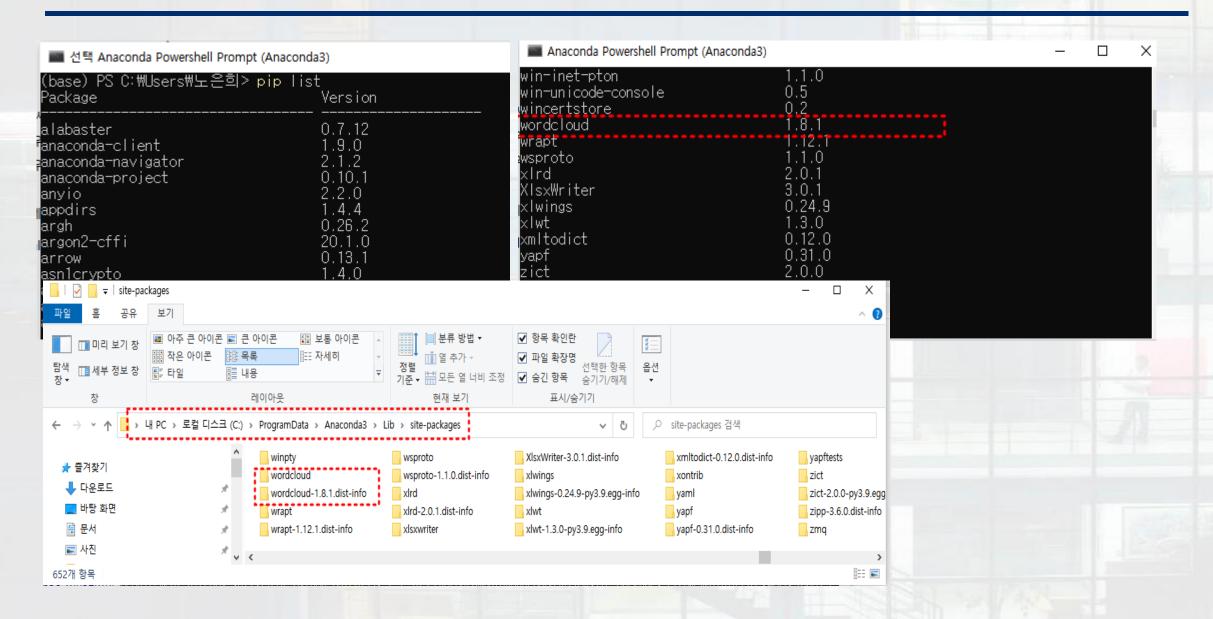


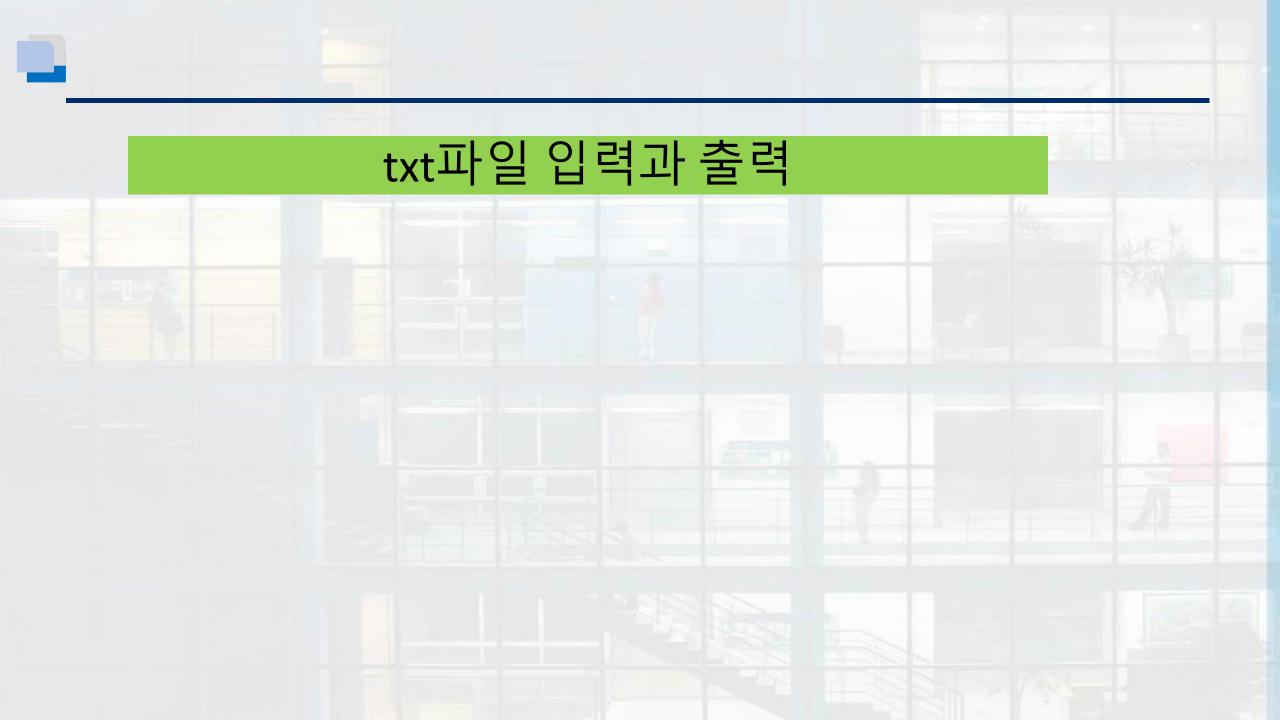


#### wordcloud 라이브러리가 설치된 모습



## wordcloud 설치 확인하기







#### 파일 입출력 과정

파일에 저장된 데이터를 읽기 위해서는 아래 그림과 같은 단계를 거칩니다. 먼저 파일을 열고 파일이 열리면 파일에서 데이터를 읽거나 쓸 수 있습니다. 작업이 끝나면 파일을 닫아야 합니 다.



#### l단계 => 파일 열기

파일을 열려면 open() 함수에서 파일명을 입력하고 읽기 또는 쓰기 인지를 지정해야 합니다. r은 read이고 w는 write입니다.

#### 2단계 => 파일에서 데이터를 읽거나 쓰기

파일에 데이터를 쓰거나 파일로부터 데이터를 읽어올 수 있는 상태입니다.

#### 3단계 => 파일 닫기

작업이 끝나면 파일을 닫아야 합니다. 파일을 닫을때는 1단계에서 open() 함수로 연 변수명을 닫아줍니다.

#### 변수명.close()

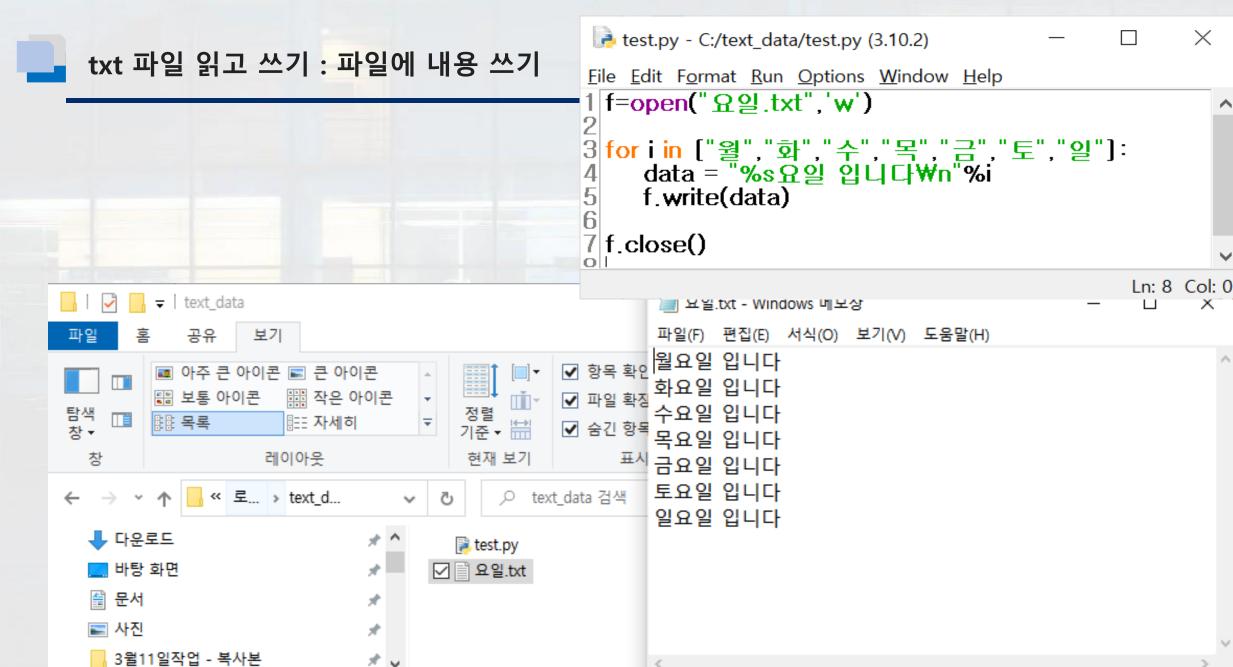


## txt 파일 읽고 쓰기

```
f = open("파일명.txt", "r")
파일개체 파일명 파일 여는 모드개체
임기
변수명 = open("파일명","r")
쓰기
변수명 = open("파일명","w")
```

#### 파일의 열기 모드

모드	설명	파일이 없을때			
생략	읽기모드				
r	읽기모드(read), 기본값으로 파일의 처음부터 읽기	에러			
w	쓰기모드(write), 파일의 처음부터 쓰기, 기존 파일이	생성			
vv	있으면 덮어쓰기, 기존내용이 사라짐	ÖÖ			
r <sup>+</sup>	읽기/쓰기 겸용모드로 파일의 처음부터 읽기/쓰기	에러			
а	추가모드(append)로 기존의 파일이 있으면 끝에	생성			
а	이어서 쓰기	ÖÖ			
t	텍스트 모드, 텍스트 파일 처리				
b	이진모드, 이진파일을 처리				



Ln 1, Col 1

Windows (CRLF)

ANSI

1개 항목 선택함 105바이트



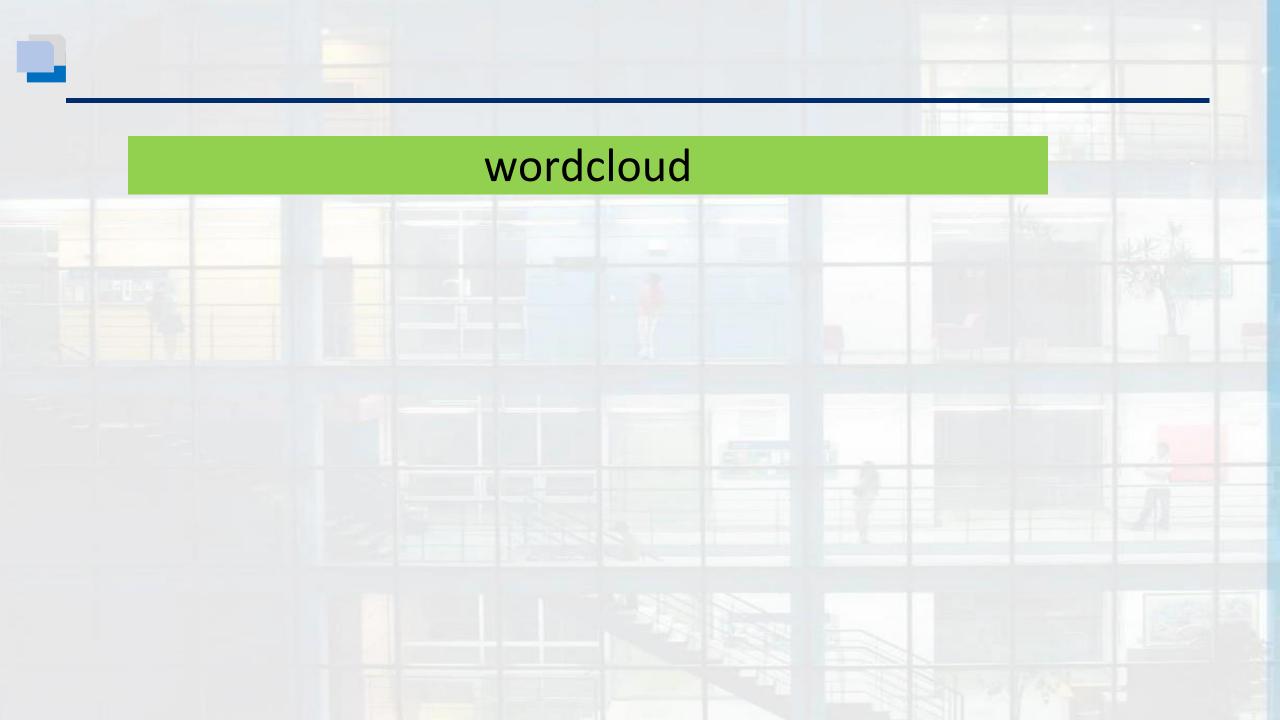
## txt 파일 읽고 쓰기 : 외부파일 불러오기

```
readline(): 한 줄씩 읽기
                                                                        \times
 file_open.py - C:/text_data/file_open.py (3.10....
<u>File Edit Format Run Options Window Help</u>
1 f = open("요일2.txt", 'r')
2 line = f.readline()
4 print(line)
5 f.close()
                                                                Ln: 6 Col: 0
                                                                                                                   X
 IDLE Shell 3.10.2
<u>File Edit Shell Debug Options Window Help</u>
\rangle\rangle\rangle
                                      RESTART: C:/text_data/file_open.py =======
      월요일 입니다
                                                                                                            Ln: 29 Col: 0
```



## txt 파일 읽고 쓰기 : 외부파일 불러오기

#### Readlines: 여러줄 읽기 X file\_open2.py - C:/text\_data/file\_open2.py (3.10.2) IDLE Shell 3.10.2 File Edit Format Run Options Window Help File Edit Shell Debug Options Window Help ▲1 f = open("요일2.txt", 'r') >>> |3|lines = f.readlines() ======== RESTART: C:/t ext\_data/file\_open2.py ==== 5 for line in lines: print(line) 월요일 입니다 8 f.close() 화요일 입니다 수요일 입니다 목요일 입니다 금요일 입니다 토요일 입니다 일요일 입니다 월요일 입니다 화장이 하나나 Ln: 65 Col: 0 Ln: 9 Col: 0





## 워드클라우드 필요한 라이브러리

#### 주요 모듈

-BeautifulSoup : 크롤링

-requests: url get request

-Twitter : 형태소 분석

-Counter: 형태소 count

-WordCloud : 워드클라우드 생성

-matplotlib : 워드클라우드 출력

#### wordcloud

-> 설치하기기 pip install wordcloud



## 의 워드 클라우드(Word Cloud)

워드 클라우드(Word Cloud)는 단어들을 분석해 중요도 빈도수 인기도 등을 고려해 단어들을 시각적으로 배치하는 것

중요도를 나타내기 위해 글자 크기, 굵기, 색 등의 형태를 바꿔 나타낸다.

• font\_path : 한글 폰트의 경로

• background\_color : 배경 색 지정

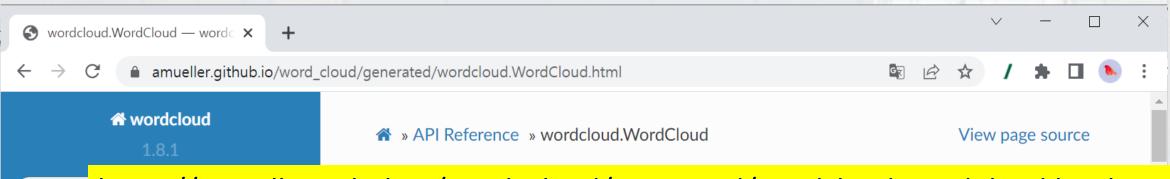
width : 가로폭 지정

• height : 세로폭 지정

max\_words: 이미지에 넣을 최대 word 수를 지정

• max\_font\_size : 이미지에 넣을 최대 폰트 크기를 지정

## wordcloud.WordCloud



https://amueller.github.io/word\_cloud/generated/wordcloud.WordCloud.html

#### wordcloud.WordCloud

#### **USER DOCUMENTATION**

□ API Reference

#### wordcloud.WordCloud

wordcloud.ImageColorGenerator wordcloud.random\_color\_func wordcloud.get single color func

#### **Command Line Interface**

**Gallery of Examples** 

Changelog

class wordcloud.WordCloud(font\_path=None, width=400, height=200, margin=2, ranks\_only=None, prefer\_horizontal=0.9, mask=None, scale=1, color\_func=None, max\_words=200, min\_font\_size=4, stopwords=None, random\_state=None, background\_color='black', max\_font\_size=None, font\_step=1, mode='RGB', relative\_scaling='auto', regexp=None, collocations=True, colormap=None, normalize\_plurals=True, contour\_width=0, contour\_color='black', repeat=False, include\_numbers=False, min\_word\_length=0, collocation\_threshold=30)

Word cloud object for generating and drawing.

Parameters: font\_path : string

Font path to the font that will be used (OTF or TTF). Defaults to DroidSansMono path on a Linux machine. If you are on another OS or don't have this font, you need to adjust this path.

width · int (default=400)

#### wordcloud 실습

```
from wordcloud import WordCloud import matplotlib.pyplot as plt

f = open('test.txt','rt', encoding='UTF-8').read() #택스트 파일 떨어주기

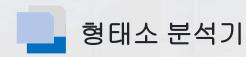
wordcloud = WordCloud(font_path = 'C:/Windows/Fonts/malgun.ttf', #한글픈트 반드시 지정하기
background_color="black",
width = 1000,
height = 1000,
max_words = 1000,
max_font_size = 10000)

wordcloud = wordcloud.generate(f)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```



# 형태소 분석 프로그램 사용



형태소 분석기의 각 장점과 단점에 따라 프로젝트에서 사용할 형태소 분석기를 사용해야 한다. Konlpy(코모란, 한나눔, kkma, mecab, twitter)에서 선택하여 사용 하는 것을 추천한다. Konlpy 안에서 각 형태소의 장단점

	코모란	한나눔	Kkma	Mecab	Twitter
장점	자소 분리 가능, 오탈자 분석 가능, 고유명사 분석 가능	로딩시간이 빠른편	띄어쓰기 오류에 덜 민감	새로운 사전 추가 가능, (twitter 공개전) 띄어쓰기에서 가장 좋은 성능, 속도, 정확도	띄어쓰기 성능이 가장 좋다, stemming 가능, 이모티콘, 해쉬태그 같은 인터넷 텍스트에 강함 , 비속어, 비표준어도 분석 가능
한계점	로딩 속도가 길다, 띄어쓰기 없는 문장분석에 취약하다	띄어쓰기 없는 문장 분석에 매우 취약, 정제된 언어가 사용되지 않는 문서에 대한 형태소 분석 정확도가 높지 않은 문제점	분석 시간 오래 걸림, 정제된 언어가 사용되지 않는 문서에 대한 형태소 분석 정확도가 높지 않은 문제점	미등록어 처리 문제, 동음의어 처리 문제	미등록어 처리 문제, 동음의어 처리문제, 분석 범주 적은편

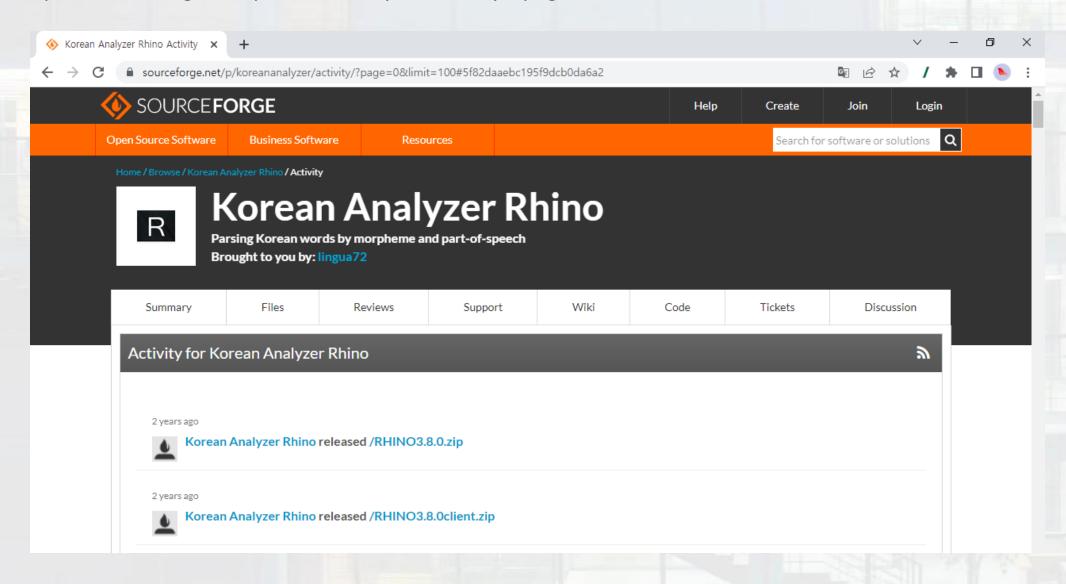
<출처>https://velog.io/@metterian/한국어-형태소-분석기POS-분석-3편.-형태소-분석기-비교

형태소(形態素, 영어: morpheme)는 언어학에서 (일반적인 정의를 따르면) 일정한 의미가 있는 가장 작은 말의 단위



#### Real Hangul INput Object

https://sourceforge.net/p/koreananalyzer/activity/?page=0&limit=100#5f82daaebc195f9dcb0da6a2

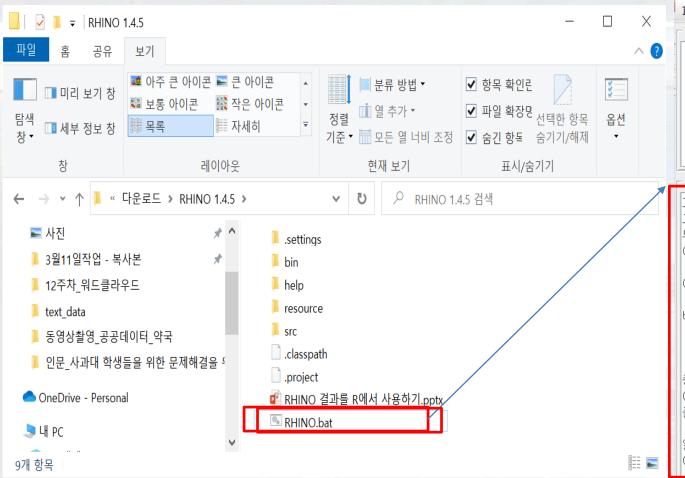


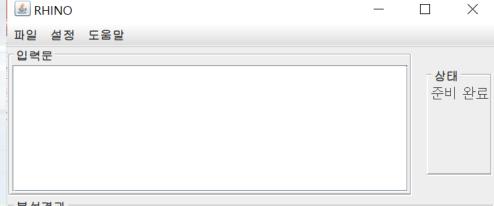


#### 한글 형태소 분석기 RHINO

#### 한글 형태소 분석기 RHINO

Korean Morphological Analyzer, RHINO





기자, 간담회, 사진, 황금빛, 기자, 글로벌, 메타버스", 플랫폼, 의, 아바타, 하체, 그간, 아바타, 메타, 메타버스", 플랫폼, 호라이즌, 아바타, 하체, 하체, 이유, 무엇, 일, 공동, 창업자, 이진, 최고, 제품, 책임자, 일, 오전, ", 열린, 기자, 간담회, 이유, 웹, 기반, 메타버스", 플랫폼, 이용자, 자유로, 자기, 표현, 욕구, 팬, 믹, 감염, 병, 유행, 이후, 웹, 기반, 메타버스", 플랫폼, 이용자, 요구, 온라인, 이용자, 관계, 소통, 경험, 필요, 해, 메타버스", 무엇, 필요, 때, 인터넷, 진화, 과정, 정보, 소통, 도구, 경험, 액, 티비, 티, 활동, ", 수단, 변환, 미션, 밥, 인터넷, 진화, ", 비전, 설명, ", 그동안, 하체, 실사, 기반, 아바타, 이용자, 이유, 출시, 당시, 가상현실, 증강, 현실, 기반, 협업, 도구, 플랫폼, 기기, 가상, 세계, 회의, 서비", 하체, 기기, 상체, 시야, 사람, 실제, 의자, 회의, 하체, 문제, 기기, 서비", 이용, 기기, 보편, 속도, 서비", 이용, 지난해, 웹, 버전, 기존, 접근, 자신, 개성, 이용자, 요구, ", 문화, 예술, 콘텐츠, 중심, 메타버스", 플랫폼, 창작자, 자신, 작품, 차원, 가상, 갤러리, 용도, 주제, 공간, 이벤트, 파티, 창작자, 입장, 자신, 아바타, 메타버스", 욕구, 결국, ", 메타버스", 아바타, 플랫폼, "타트, 레디, 플레이어, 미, 파트너십, 바탕, 전신, 아바타, 지원, 기능, 이날, 이번 파트너십, 계기, 이용자, 프로필, 피, ", 신체, 모양, 복장, 헤어, "타일, 가운데, 아바타, 얼굴, 기존, 실제, 사진, 기반, 옵션, 레디, 플레이어, 미, 일러"트, 기반, 캐릭터, 옵션, 아바타, 자신, 신체, 문화, 정체, 자유, 지속, 방침, 문화, 전통, 의상, 향후, 아바타, 아이템

## VOYANT사용하여 워드클라우드 만들기



# 워드클라우드 생성기르 이용하여 워드클라우드 만들기

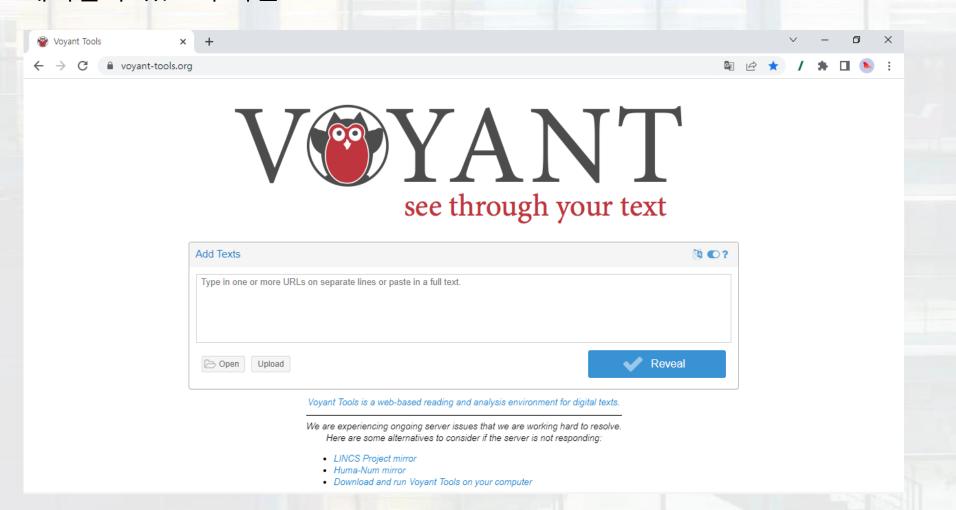


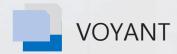


## ■ 보이언트 툴(Voyant Tools)

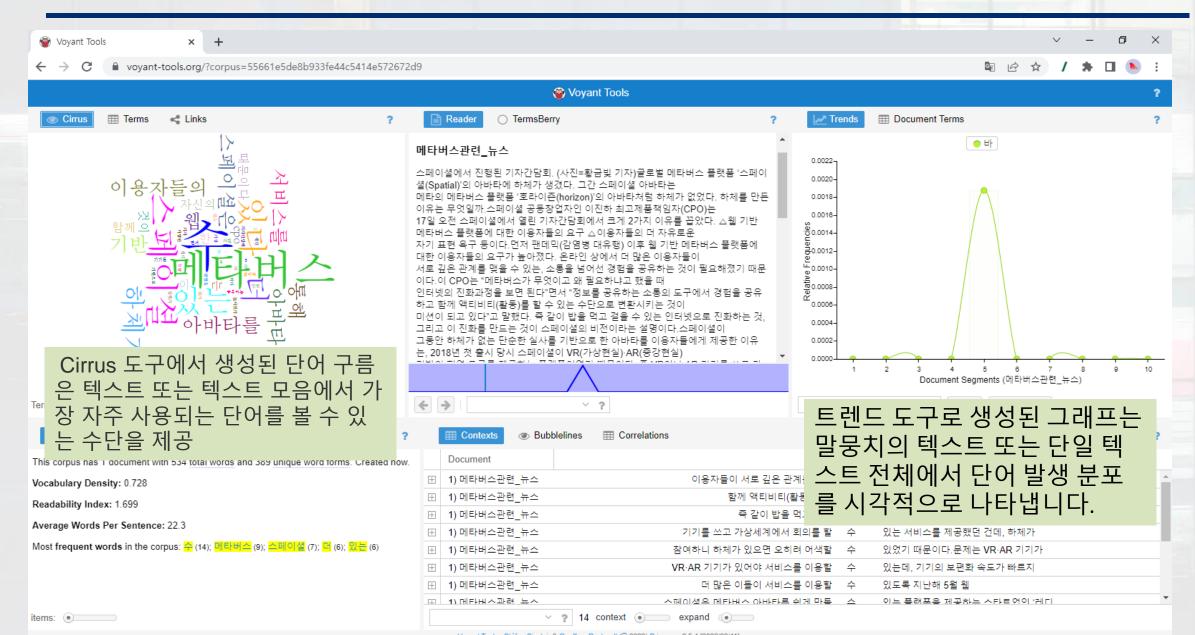
https://voyant-tools.org/

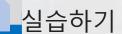
Voyant Tools 는 텍스트 분석을 수행하기 위한 오픈 소스 웹 기반 응용 프로그램 디지털 인문학 분야 의 학자 뿐 아니라 학생과 일반 대중 이 텍스트 또는 말뭉치를 학술적으로 읽고 해석할 수 있도록 지원



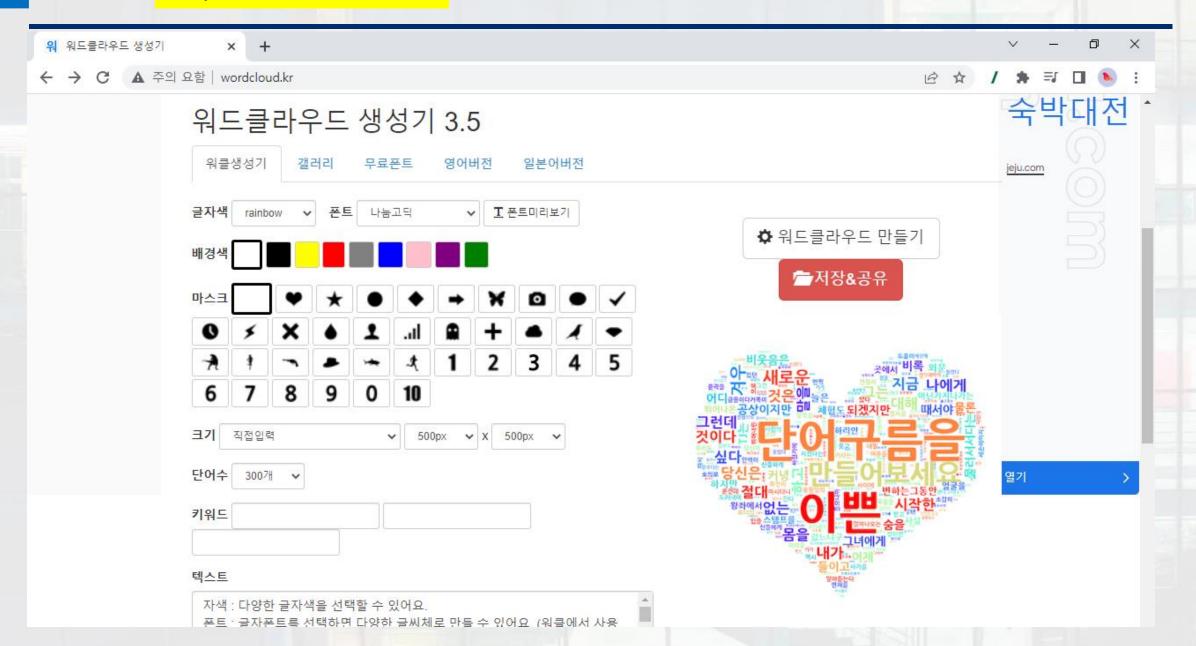


#### https://voyant-tools.org/





## http://wordcloud.kr/



## 실습하기 http://wordcloud.kr/





기자, 간담회, 사진, 황금빛, 기자, 글로벌, 메타버스", 플랫폼, 의, 아바타, 하체, 그간, 아바타, 메타, 메타버스", 플랫폼, 호라이즌, 아바타, 하체, 하체, 이유, 무엇, 일, 공동, 창업자, 이진, 최고, 제품, 책임자, 일, 오전, ", 열린, 기자, 간담회, 이유, 웹, 기반, 메타버스", 플랫폼, 이용자, 요구, 이용자, 자유로, 자기, 표현, 욕구, 팬, 믹, 감염, 병, 유행, 이후, 웹, 기반, 메타버스", 플랫폼, 이용자, 요구, 온라인, 이용자, 관계, 소통, 경험, 필요, 해, 메타버스", 무엇, 필요, 때, 인터넷, 진화, 과정, 정보, 소통, 도구, 경험, 액, 티비, 티, 활동, ", 수단, 변환, 미션, 밥, 인터넷, 진화, ", 비전, 설명, ", 그동안, 하체, 실사, 기반, 아바타, 이용자, 이유, 출시, 당시, 가상현실, 증강, 현실, 기반, 협업, 도구, 플랫폼, 기기, 가상, 세계, 회의, 서비", 하체, 기기, 상체, 시야, 사람, 실제, 의자, 회의, 하체, 문제, 기기, 서비", 이용, 기기, 보편, 속도, 서비", 이용, 지난해, 웹, 버전, 기존, 접근, 자신, 개성, 이용자, 요구, ", 문화, 예술, 콘텐츠, 중심, 메타버스", 플랫폼, 창작자, 자신, 작품, 차원, 가상, 갤러리, 용도, 주제, 공간, 이벤트, 파티, 창작자, 입장, 자신, 아바타, 메타버스", 욕구, 결국, ", 메타버스", 아바타, 플랫폼, "타트, 레디, 플레이어, 미, 파트너십, 바탕, 전신, 아바타, 지원, 기능, 이날, 이번, 파트너십, 계기, 이용자, 프로필, 피, ", 신체, 모양, 복장, 헤어, "타일, 가운데, 아바타, 얼굴, 기존, 실제, 사진, 기반, 옵션, 레디, 플레이어, 미, 일러"트, 기반, 캐릭터, 옵션, 아바타, 자신, 신체, 문화, 정체, 자유, 지속, 방침, 문화, 전통, 의상, 향후, 아바타, 아이템, 창작자, 기능, 예정, ", 아바타, 기능, 사진, 황금빛, 기자, ", 웹, 기반, 메타버스", 플랫폼, 서비", 본격, 이상, 가상, 공간, ", 이용자, 지난해, 때, 배, 가량, 인간, 방식, 보편, 다양, 플랫폼, 이용자, 커뮤니티, 저변, 이날, 기자, 간담회, 참석, 메타버스", 체험, 기자, 이동, 방법, 자동, 화질, 간담회, 참석자, 공간, 때, 속도, ", 이, 화질, 속도, 경우, 컴퓨터, 와이파이, 상황, 조금, ", 이용자, 중심, 서비", 개선, 지속

#### TIP>> 엑셀에서 텍스트 수정하기

