



YOUTUBE 댓글 수집 시각화 개발환경구축 가이드

Selenium, Konlpy, WebDriver_Manager, Pandas, BeautifulSoup4



대우직업능력개발원 4팀 TEAM 임대 임서인 황대명

START



목차



1. 개발환경구축 가이드의 목적
2. 프로그램 버전
3. 필요한 라이브러리 버전
4. 프로그램 폴더 요약
5. 프로그램 실행 방법
6. 개발환경 구축
7. 실행 및 결과 화면

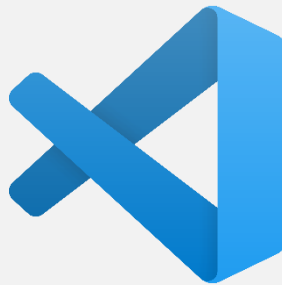


exe 파일을 실행하면 되는 일반 사용자와 달리, 개발자들은 코드를 통해 파일을 실행할 때 여러 **라이브러리를 다운로드**하고 적용해야 합니다. 그러나 이를 충분히 고려하지 않으면 실행 과정에서 **예기치 못한 문제**가 발생할 수 있습니다.

따라서 이 가이드 문서는 개발자들이 **개발 환경 설정을 간편하게 확인할 수** 있도록 제공되며, 이를 통해 개발 과정에서 발생할 수 있는 **문제를 미리 예방하고 실행을 원활하게** 진행할 수 있도록 도와줍니다.



Python
3.10.10



IDE : VS code
v-1.76.0



JDK 19
v-19.0.1



Selenium

*포터블 프레임워크 v-4.8.2

동적으로 생성되는 사이트의 데이터를
크롤링 할 때 매우 유용하게 사용

Webdriver Manager

라이브러리 v-3.8.5

크롬 드라이버를 크롬 브라우저
버전에 맞게 자동으로 다운로드

Wordcloud

라이브러리 v-1.8.1

텍스트데이터를 가지고 워드클라우드를
생성하기 쉽게 도와줌

Konlpy

라이브러리 v-0.6.0

한국어 정보처리를 위한
파이썬 패키지

bs4

라이브러리 v-0.0.1

HTML정보로 부터 데이터를 가져오기
쉽게, 데이터 별로 나누어 줌

Pandas

라이브러리 v-1.5.3

쉽고 직관적인 분류된 데이터로 작업할 수
있도록 데이터 구조를 제공

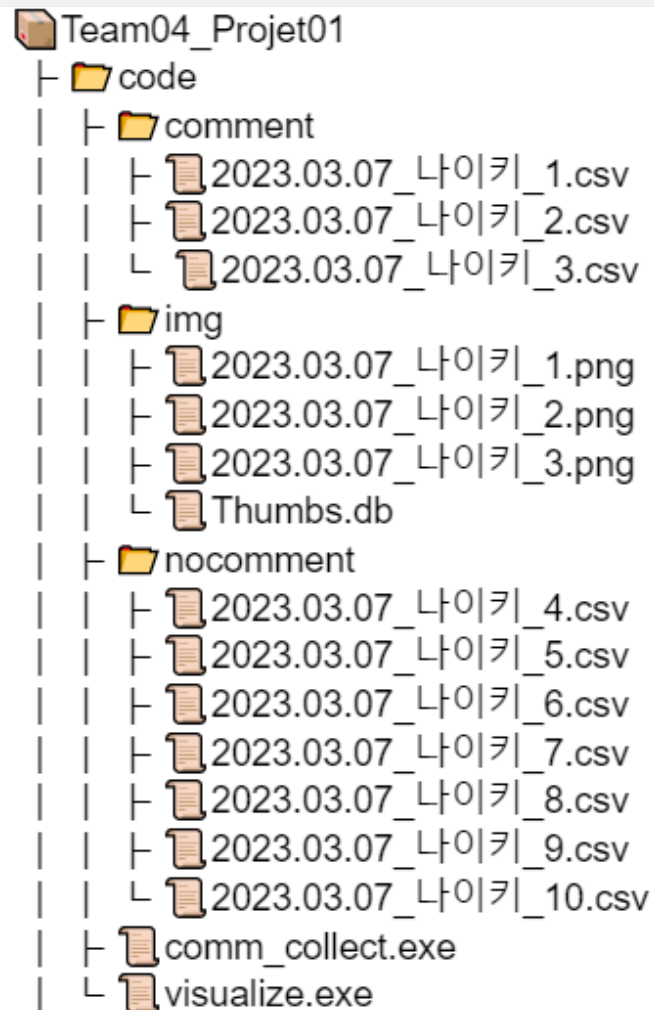
Matplotlib

라이브러리 v-3.7.1

크롬 드라이버를 크롬 브라우저
버전에 맞게 자동으로 다운로드

하이퍼링크 클릭 시
해당 라이브러리 참고사이트 접속

*포터블 프레임워크 : 특정 플랫폼이나 운영체제에서 독립적으로 실행될 수 있는 소프트웨어 프레임워크



code :

프로그램을 개발할 때 사용되거나 결과를 담는 폴더

└─ comment :

영상정보와 댓글이 담긴 csv의 저장소

└─ img :

comment폴더의 파일들이 시각화 된 결과물

└─ nocomment :

영상정보가 담긴 csv의 저장소

└─ comm_collect.exe :

댓글 수집 프로그램 실행파일

└─ visualize.exe :

시각화 프로그램 실행파일

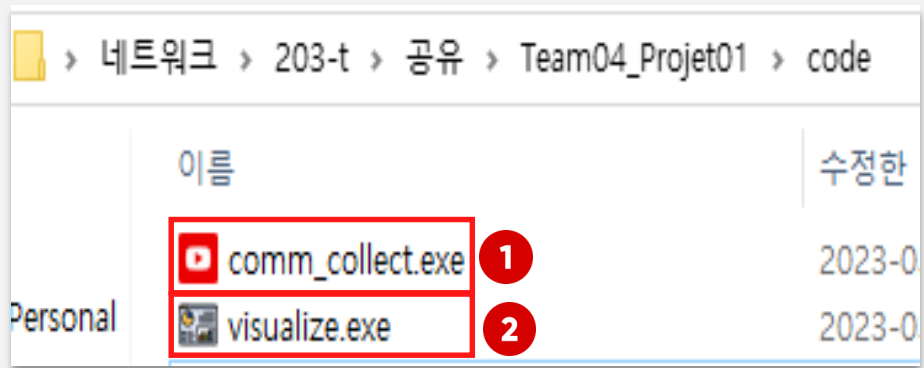


```
Team04_Projet01
└─ env
   └─ 모듈
      └─ wordcloud-1.8.1-cp311-cp311-win_amd64.whl
      └─ jdk-19_windows-x64_bin.exe
      └─ python-3.11.2-amd64.exe
```

code :

프로그램을 개발할 때 사용되거나 결과를 담는 폴더

```
└─ env
   └─ 개발환경 구축을 위해 필요한 파일들이 담긴 폴더
      └─ 모듈
         └─ wordcloud-1.8.1-cp311-
            cp311-win_amd64.whl
         └─ jdk-19_windows-x64_bin.exe
         └─ python-3.11.2-amd64.exe
```



1 *comm_collect.exe*

(댓글 수집 프로그램)

2 *visualize.exe*

(시각화 프로그램)

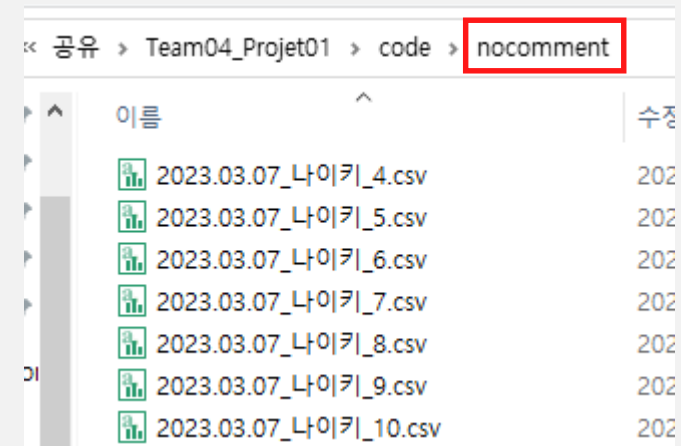
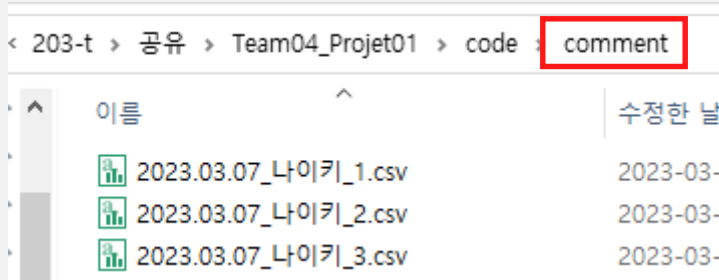
위 순서대로 실행한다.



1

comm_collect.exe

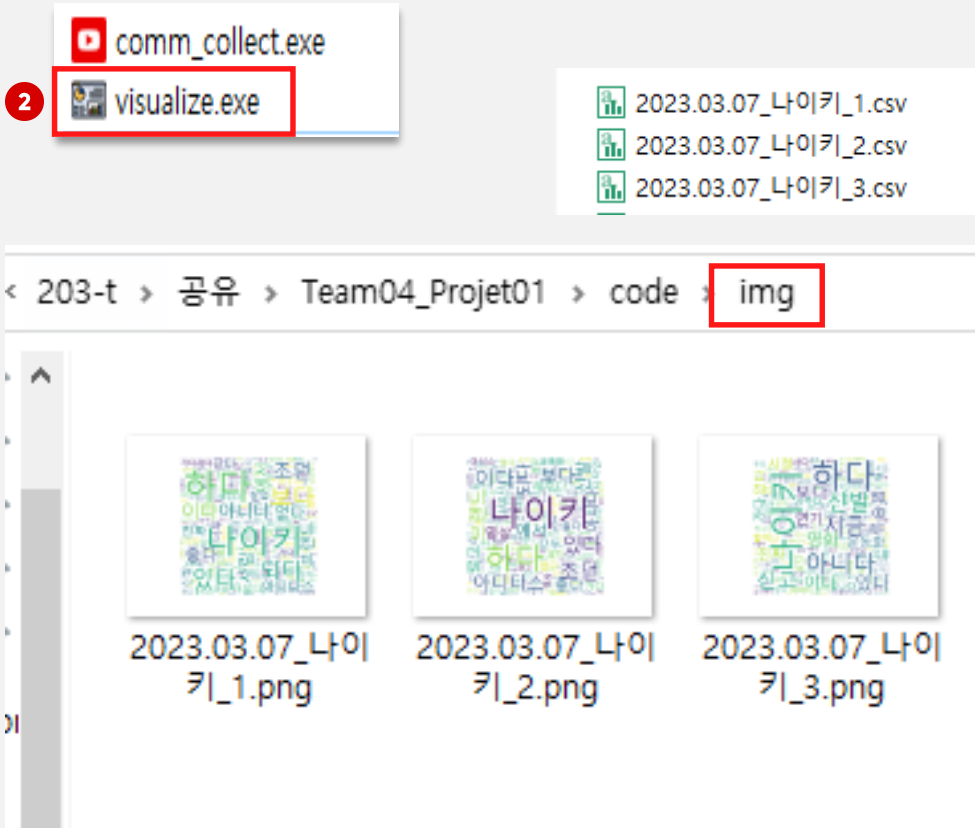
visualize.exe



1. comm_collect.exe 실행 시

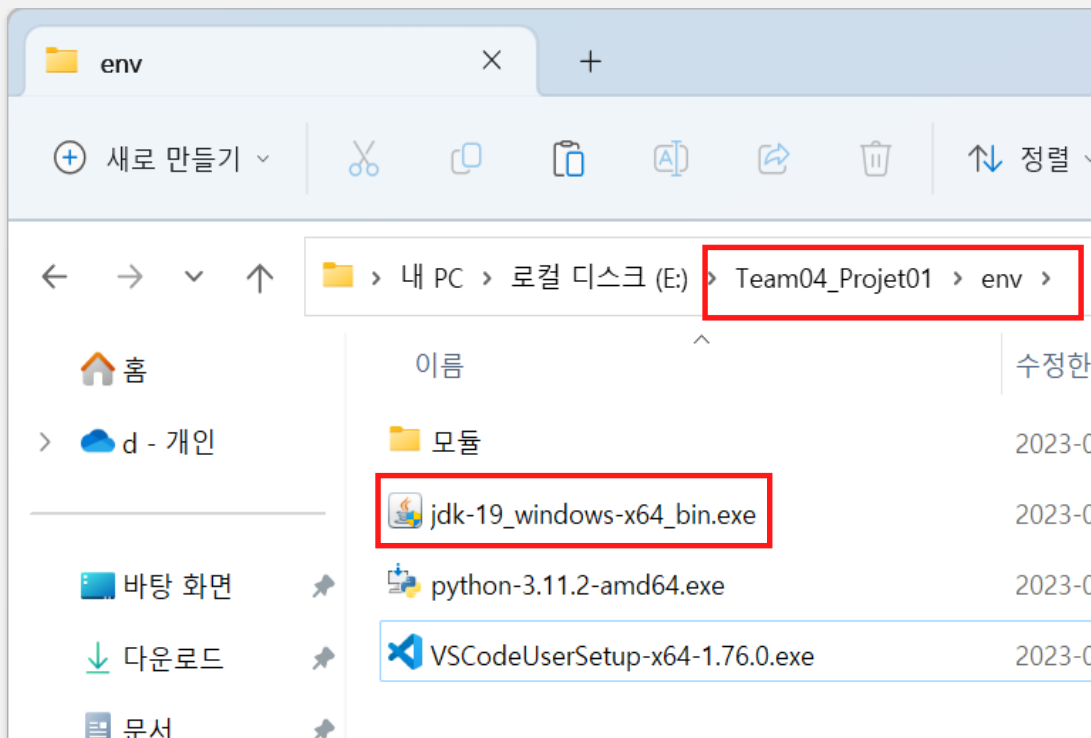
- comment 폴더 안에 3개의 csv파일
- nocomment 폴더 안에 7개의 csv파일

이 생성되면 실행완료

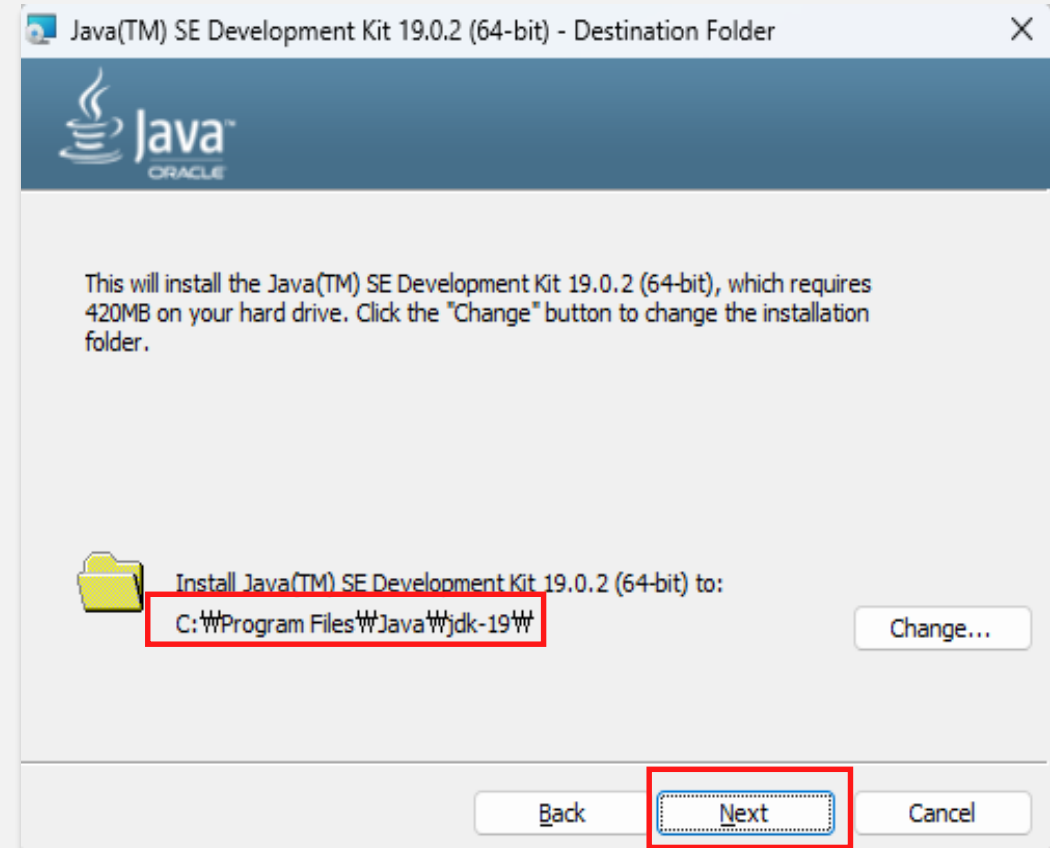


2. visualize.exe 실행 시

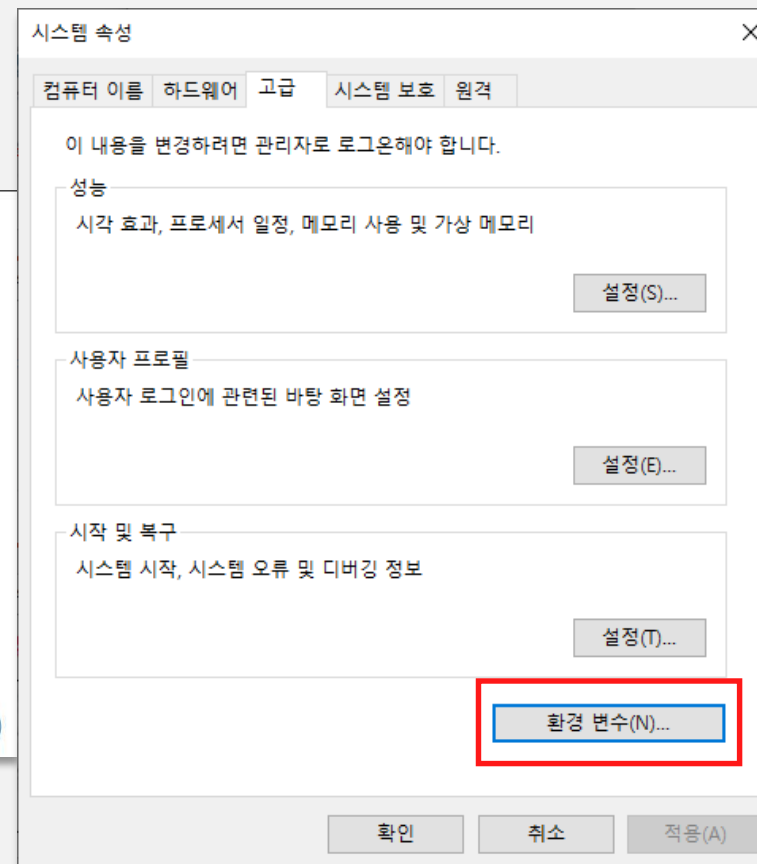
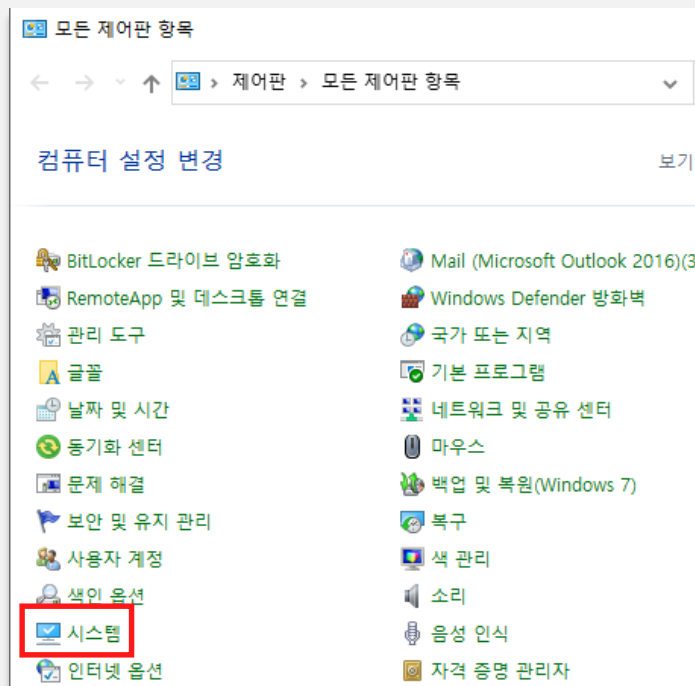
comment 폴더의 csv파일이
img폴더에 워드클라우드의 형태로 생성
되어 있으면 실행완료



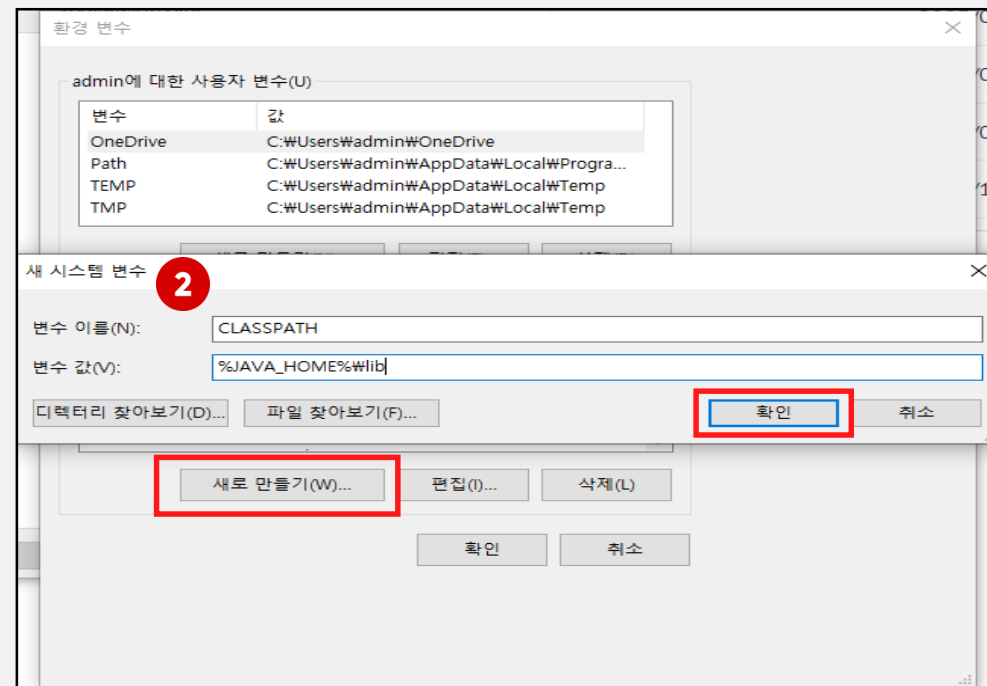
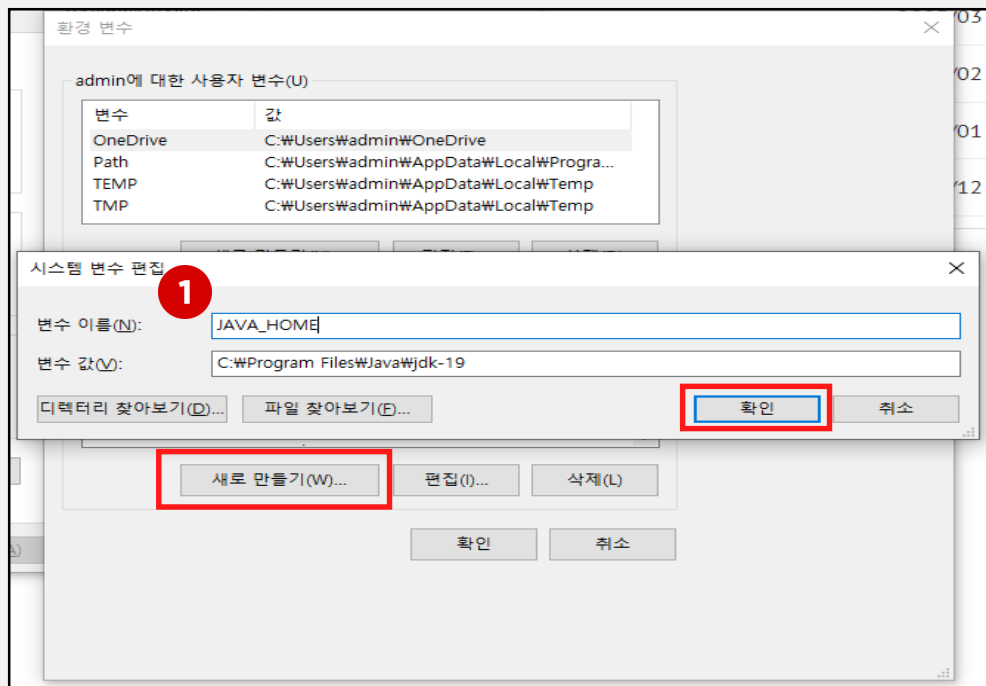
env 폴더에 있는 jdk-19 설치 파일 실행



- 설치 경로 확인 후 설치 진행



- 제어판 > 시스템 > 고급시스템 설정 > 환경변수



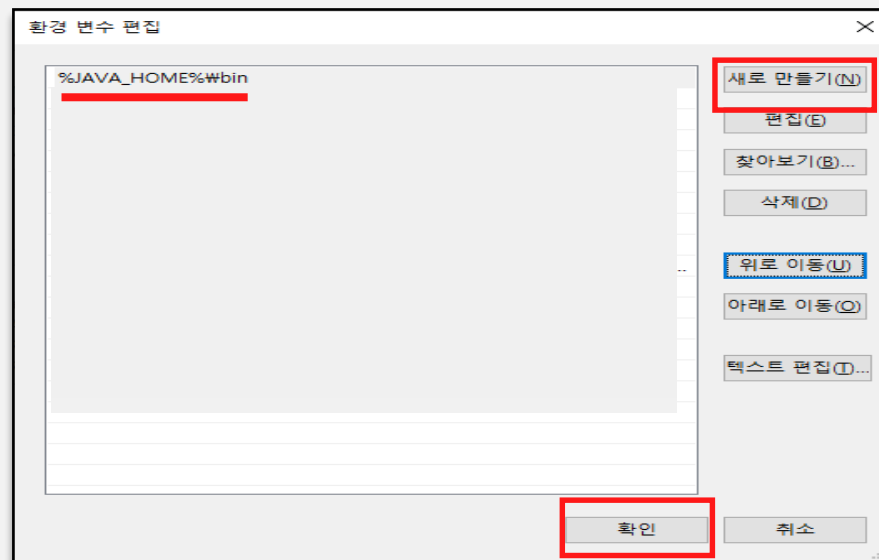
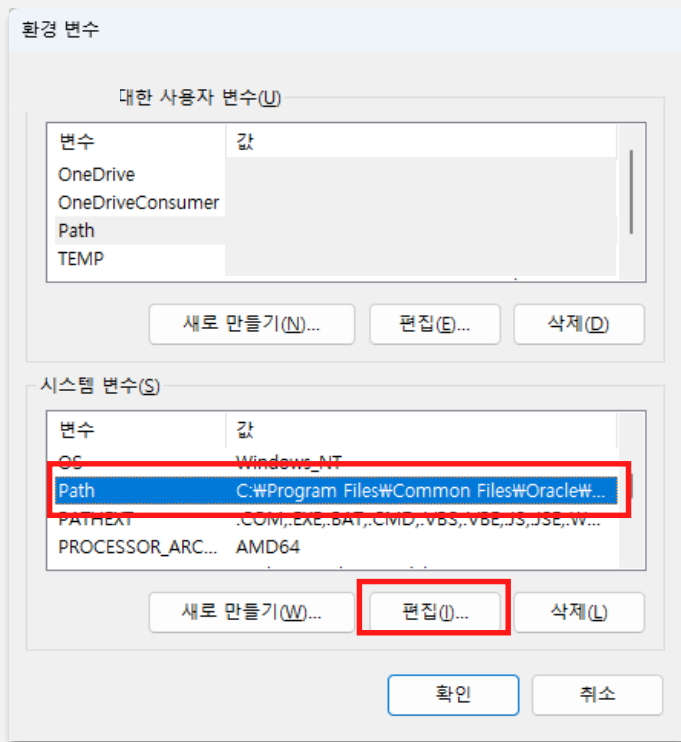
새로 만들기 선택

1

변수 명 : JAVA_HOME
변수 값 : C:\Program Files\Java\jdk-19

2

변수 명 : CLASSPATH
변수 값 : %JAVA_HOME%\lib

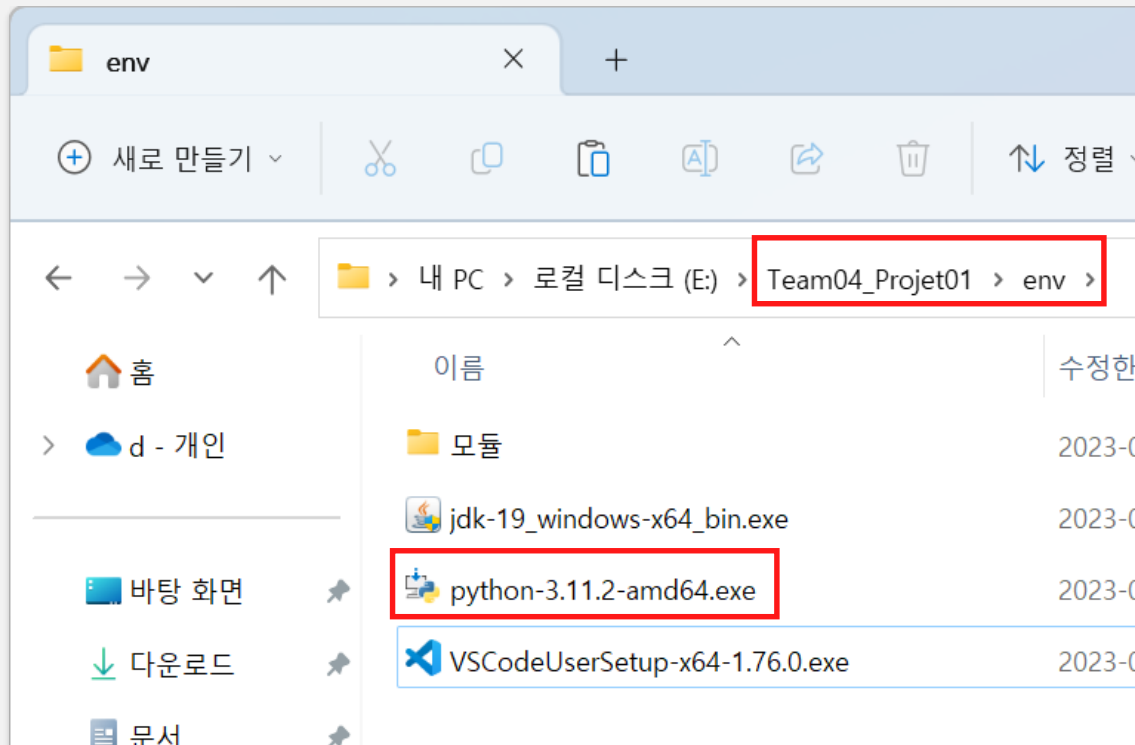


```
C:\Users\admin>java -version
java version "19.0.1" 2022-10-18
Java(TM) SE Runtime Environment
Java HotSpot(TM) 64-Bit Server VM

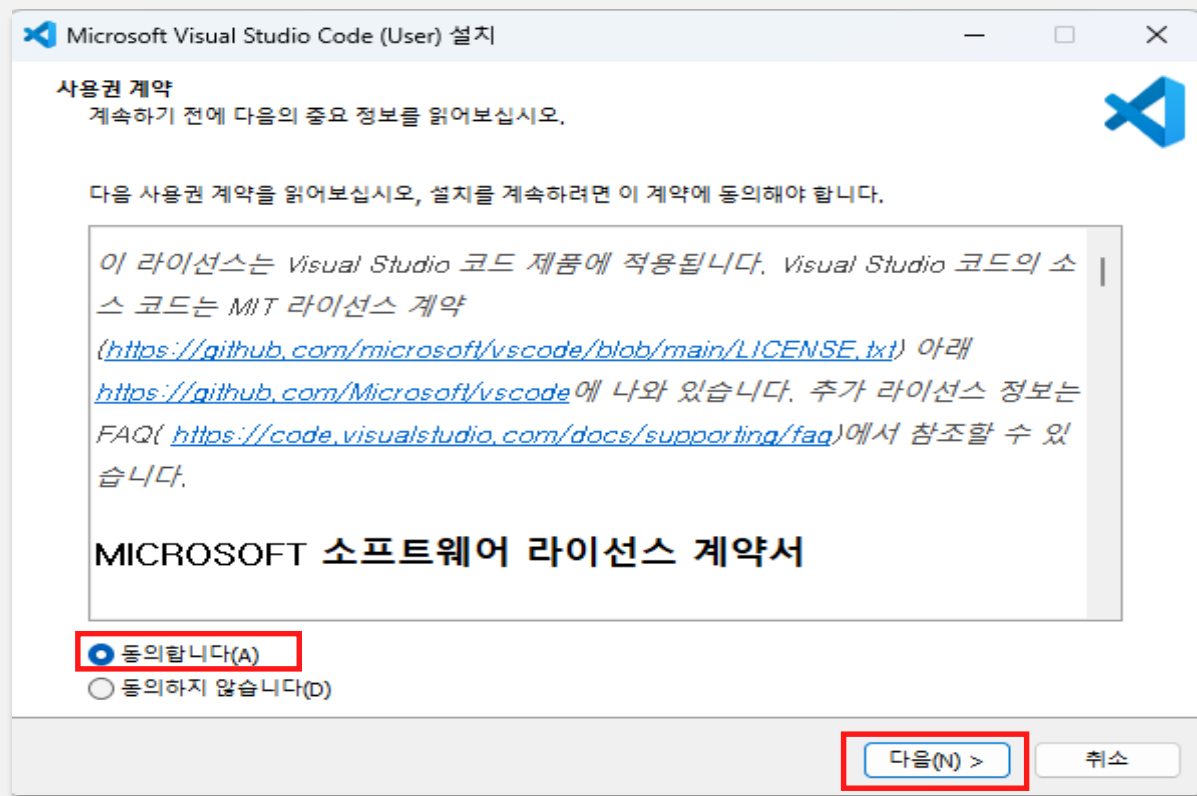
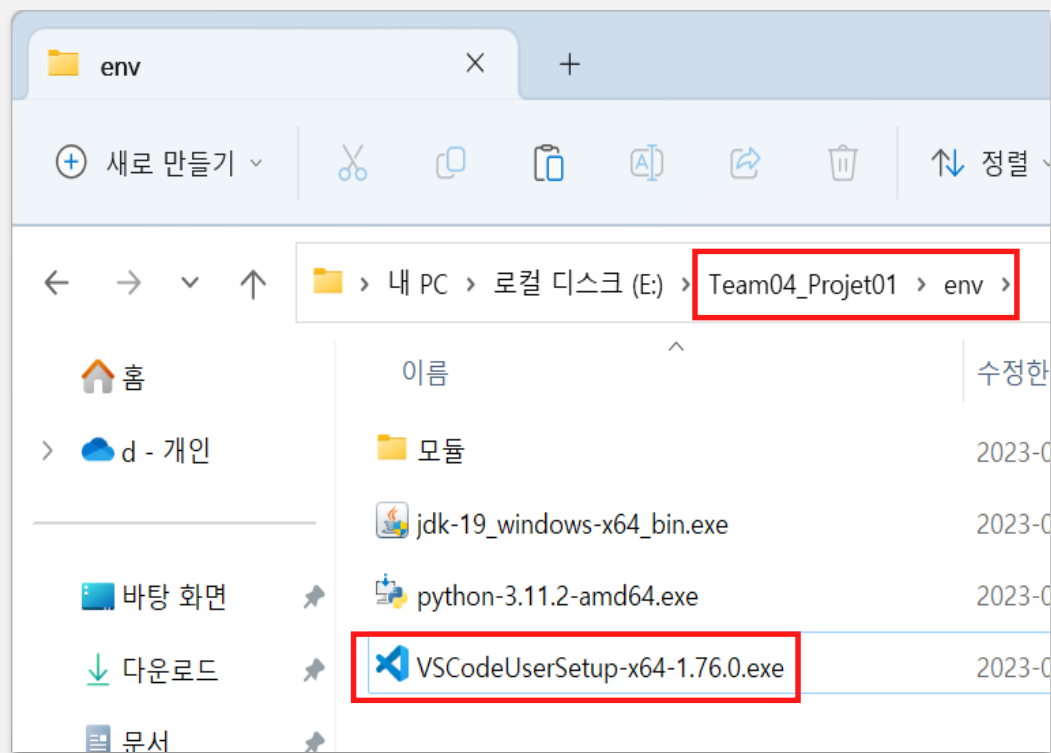
C:\Users\admin>javac -version
javac 19.0.1

C:\Users\admin>javap -version
19.0.1
```

- Path 선택 > 편집 > 새로 만들기 > %JAVA_HOME%\bin > 확인



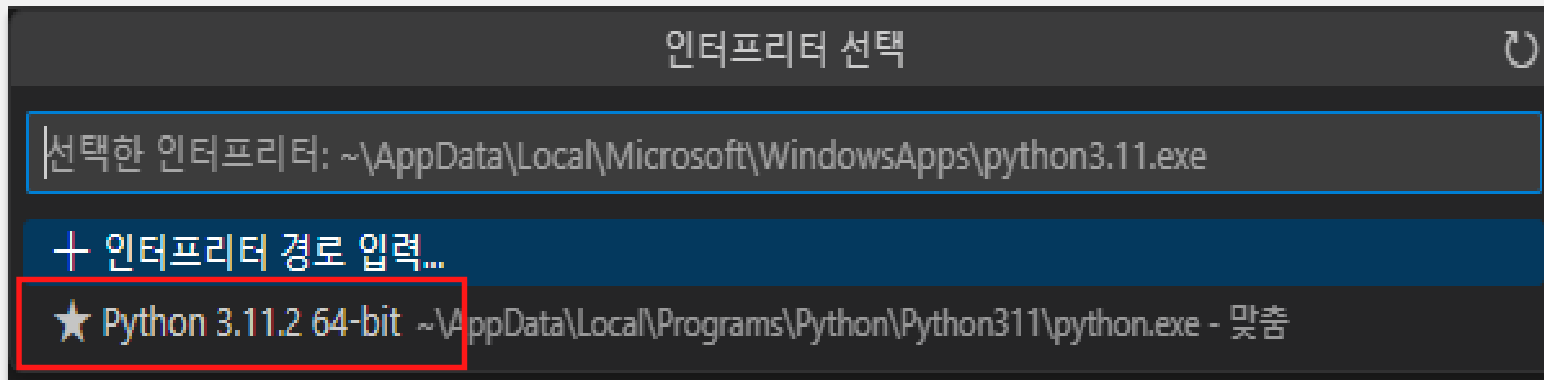
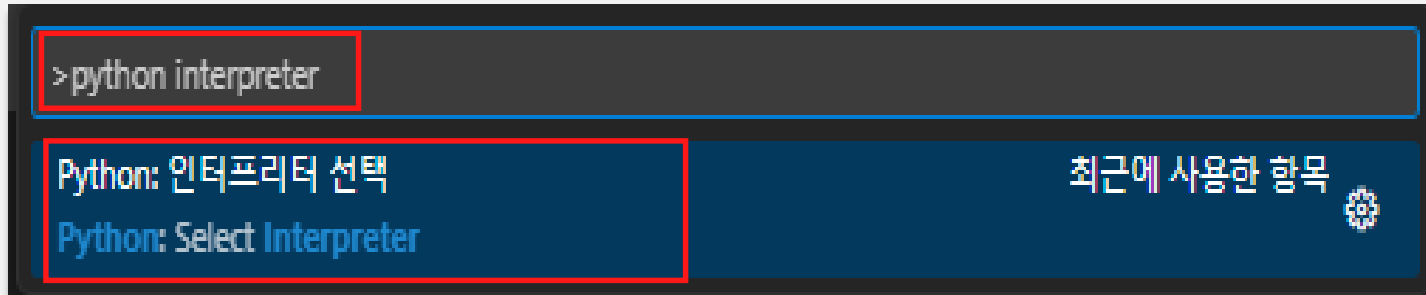
- env 폴더에 python 설치 파일 실행 > Add python.exe to PATH 체크 > 설치 진행



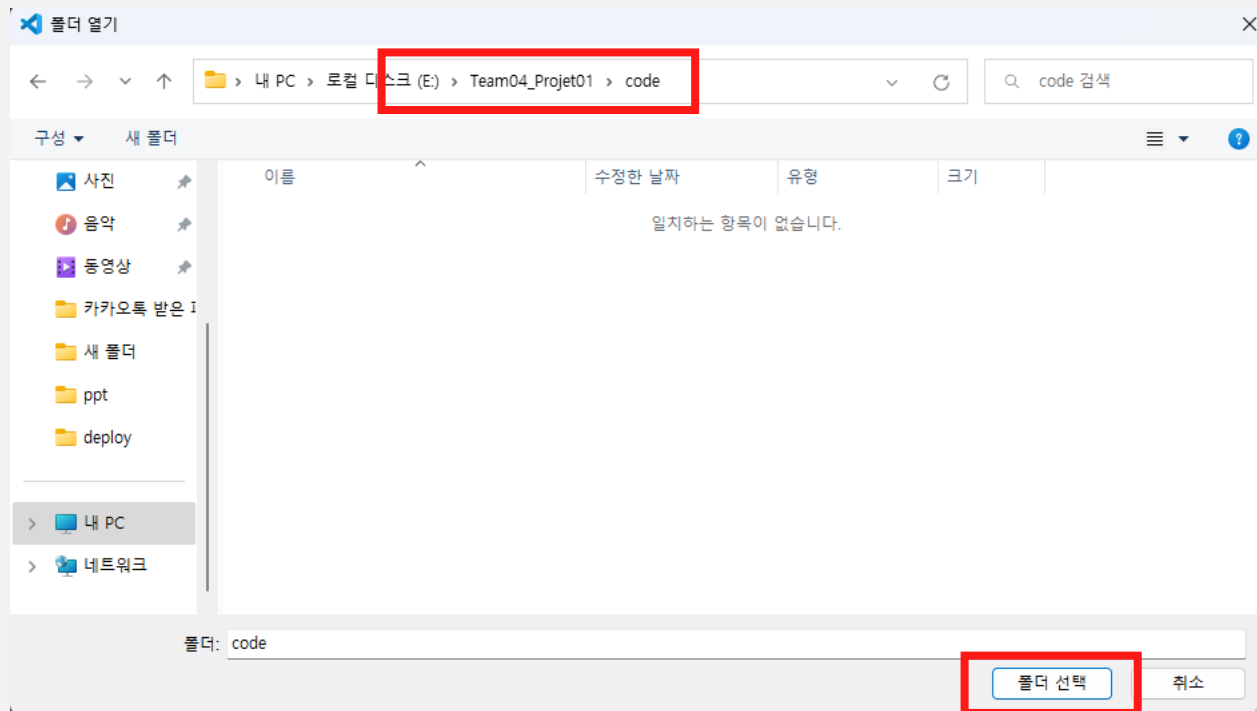
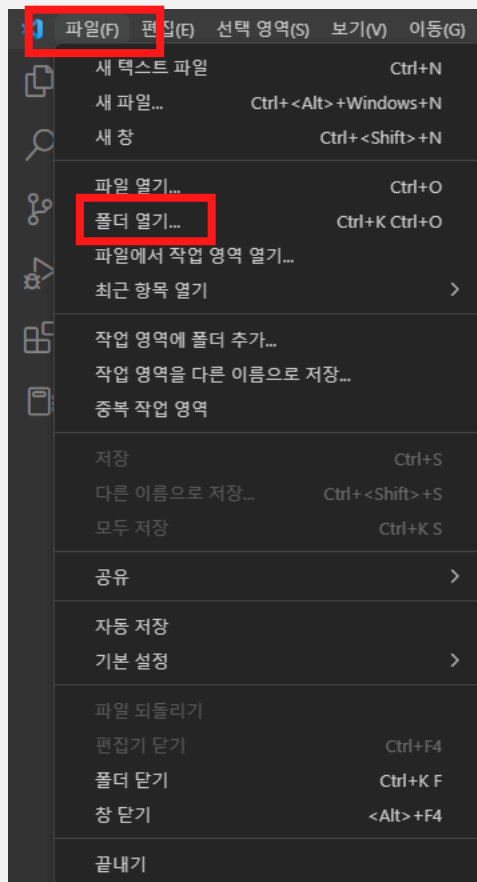
- env 폴더에 VScode 설치 파일 실행 > 동의합니다 체크 > 설치 진행



- 확장 탭 선택 > python 검색 > 설치



- Ctrl + shift + p 단축키 입력 > python interpreter 검색 > 파이썬 버전 선택



- 파일(F) -> 폴더열기... -> code 폴더 선택



1 Run in Interactive Window

2 대화형 창에서 현재 파일 실행

대화형 창의 줄에서 실행

대화형 창에서 선택/행 실행 <Shift>+Enter

대화형 창에서 라인까지 실행

커서까지 실행

터미널에서 Python 파일 실행

Python 터미널에서 선택/줄 실행 <Shift>+Enter

가져오기 정렬

명령 팔레트... Ctrl+<Shift>+P

except:

print("... 모듈을 선택합니다.")

1 comm_collect.py 선택 -> 마우스 우클릭

-> 1 실행 -> 2 실행



```

댓글정보 수집을 시작합니다.

수집 날짜 : 2023.03.10
영상 이름 : [Track 49] 정애를 찾아서 / Park Seo Jun in Japan
채널 이름 : 박서준 Record PARK's
조회수 : 468,766회
게시일 : 2023. 2. 6.
원래어 : 박서준
스크립 시작 : 17:18:56
대댓글 읽기 시작 : 17:21:04
댓글 수집 시작 : 17:21:04
유틸리티 분석 결과
{"you": 195, "to": 139, "and": 92, "the": 81, "so": 66, "for": 63, "love": 68, "your": 53, "my": 58, "in": 46, "of": 43, "is": 43, "me": 43, "과다": 41,

파일이 저장되었습니다.
경로 : c:\Users\admin\Desktop\deploy\comment\2023.03.10_박서준_1.csv

댓글정보 수집을 시작합니다.

수집 날짜 : 2023.03.10
영상 이름 : 박서준이 어머무들에게 딸을 절대 안하는 충격적인 이유 --
채널 이름 : 이슈북북
조회수 : 34,702회
게시일 : 2023. 2. 22.
원래어 : 박서준
스크립 시작 : 17:21:31
대댓글 읽기 시작 : 17:21:34
댓글 수집 시작 : 17:21:34
유틸리티 분석 결과
{"박서준": 4, "과다": 4, "보다": 3, "연애인": 3, "그렇다": 3, "이름": 2, "연애": 2, "전과": 2, "원다": 2, "남자": 2, "드다": 2, "과기": 2, "관리": 2, "마루

파일이 저장되었습니다.
경로 : c:\Users\admin\Desktop\deploy\comment\2023.03.10_박서준_2.csv

댓글정보 수집을 시작합니다.

수집 날짜 : 2023.03.10
영상 이름 : [#어름발라] BTS 뷔x박서준x최우식 == 친절한 이유ㅋㅋㅋ 전과 달자마자 아메추부터 물어봄 이 조만 절대 정설래 | #나중에드름발라
채널 이름 : 디글 :Digggle
조회수 : 550,007회
게시일 : 2022. 7. 15.
원래어 : 박서준
스크립 시작 : 17:21:44
대댓글 읽기 시작 : 17:22:00
댓글 수집 시작 : 17:22:00
유틸리티 분석 결과
{"하다": 10, "종다": 7, "보다": 7, "너무": 6, "타다": 6, "남자": 6, "이탈": 5, "박서준": 5, "전과": 5, "is": 5, "인어송": 5, "subtitles": 5, "English": 5,

파일이 저장되었습니다.
경로 : c:\Users\admin\Desktop\deploy\comment\2023.03.10_박서준_3.csv

댓글정보 수집을 시작합니다.

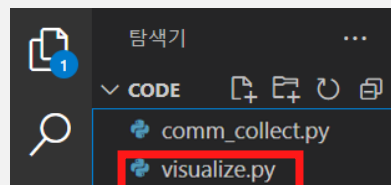
수집 날짜 : 2023.03.10
영상 이름 : 음력달과 제니 사별 편성소에서 달항아리 이유 + 지드래곤이 전다본 이유 + 박서준
채널 이름 : 우드홀라
조회수 : 94,827회
게시일 : 2023. 3. 8.
원래어 : 박서준
파일이 저장되었습니다.
경로 : c:\Users\admin\Desktop\deploy\nocomment\2023.03.10_박서준_4.csv

댓글정보 수집을 시작합니다.

수집 날짜 : 2023.03.10
영상 이름 : 어름이있다... 5월은 달력 부활지 많은 조식부터 보기만 해도 시청률 바다 수만까지 박서준x최우식과 달래라면 이게 정설라 뭐까지 | #어름발라 #살일
채널 이름 : 살일 #ing
조회수 : 29,590회
게시일 : 2022. 8. 14.
원래어 : 박서준
파일이 저장되었습니다.
경로 : c:\Users\admin\Desktop\deploy\nocomment\2023.03.10_박서준_5.csv
  
```

- 1 : comment 폴더에 저장
- 2 : nocomment 폴더에 저장

네트워크 > 203-t > 공유 > Team04_Projet01 > code				
이름	수정한 날짜	유형		
comm_collect.exe	2023-03-10 오후 1:42	응용 프로그램		
visualize.exe	2023-03-10 오후 1:43	응용 프로그램		
comment	2023-03-08 오후 12:32	파일 폴더		
deploy	2023-03-10 오후 1:50	파일 폴더		
img	2023-03-07 오후 4:12	파일 폴더		
nocomment	2023-03-08 오후 12:19	파일 폴더		



- visualize.py 실행 -> 결과를 img 폴더에 저장

```

Team 임대(Rental)

프로그램명 : visualize

수집된 데이터 저장 위치
- 시각화된 이미지는 img 폴더에 저장됩니다.

시각화 프로그램을 시작합니다.

-----

pandas 라이브러리 설치 여부를 체크합니다.
pandas 라이브러리가 존재합니다.

wordcloud 라이브러리 설치 여부를 체크합니다.
wordcloud 라이브러리가 존재합니다.

matplotlib 라이브러리 설치 여부를 체크합니다.
matplotlib 라이브러리가 존재합니다.

-----

파일 리스트 목록
2023.03.10_박서준_1.csv 2023.03.10_박서준_2.csv 2023.03.10_박서준_3.csv

2023.03.10_박서준_1.csv 파일 wordcloud 시작
2023.03.10_박서준_1.png로 img 폴더에 저장

2023.03.10_박서준_2.csv 파일 wordcloud 시작
2023.03.10_박서준_2.png로 img 폴더에 저장

2023.03.10_박서준_3.csv 파일 wordcloud 시작
2023.03.10_박서준_3.png로 img 폴더에 저장

-----

Team 임대(Rental)

프로그램명 : visualize

수집된 데이터 저장 위치
- 시각화된 이미지는 img 폴더에 저장되었습니다.

프로그램을 종료합니다.
  
```

네트워크 > 203-t > 공유 > Team04_Projet01 > code			
이름	수정한 날짜	유형	
comm_collect.exe	2023-03-10 오후 1:42	응용 프로그램	
visualize.exe	2023-03-10 오후 1:43	응용 프로그램	
comment	2023-03-08 오후 12:32	파일 폴더	
deploy	2023-03-10 오후 1:50	파일 폴더	
img	2023-03-07 오후 4:12	파일 폴더	
nocomment	2023-03-08 오후 12:19	파일 폴더	



```

Team 임대(Rental)

프로그램명 : visualize

수집된 데이터 저장 위치
- 시각화된 이미지는 img 폴더에 저장됩니다.

시각화 프로그램을 시작합니다.

-----

pandas 라이브러리 설치 여부를 체크합니다.
pandas 라이브러리가 존재합니다.

wordcloud 라이브러리 설치 여부를 체크합니다.
wordcloud 라이브러리가 존재합니다.

matplotlib 라이브러리 설치 여부를 체크합니다.
matplotlib 라이브러리가 존재합니다.

-----

파일 리스트 목록
2023.03.10_박서준_1.csv 2023.03.10_박서준_2.csv 2023.03.10_박서준_3.csv

2023.03.10_박서준_1.csv 파일 wordcloud 시작
2023.03.10_박서준_1.png로 img 폴더에 저장

2023.03.10_박서준_2.csv 파일 wordcloud 시작
2023.03.10_박서준_2.png로 img 폴더에 저장

2023.03.10_박서준_3.csv 파일 wordcloud 시작
2023.03.10_박서준_3.png로 img 폴더에 저장

-----

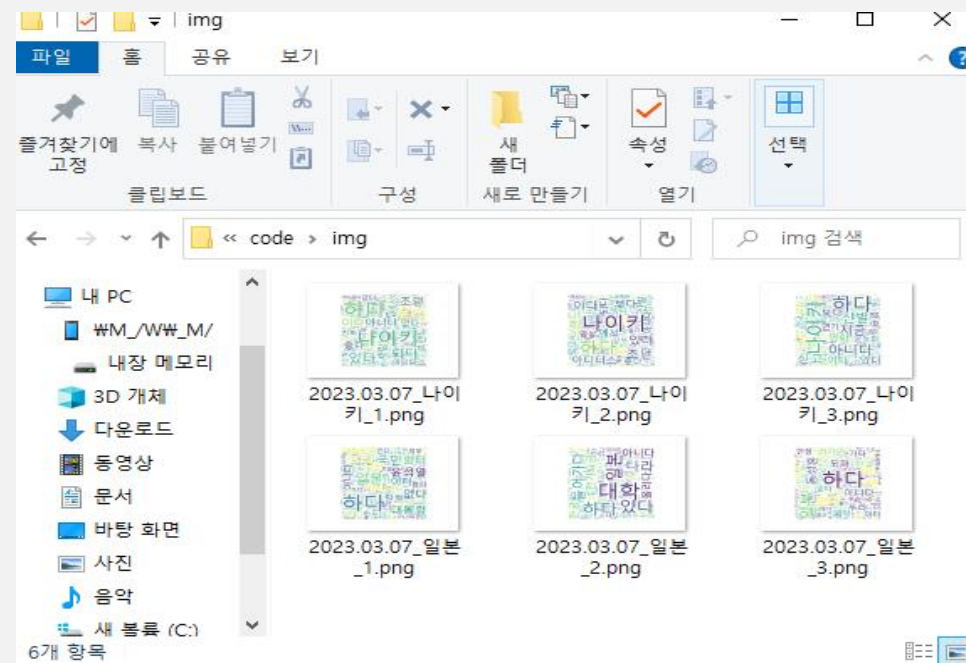
Team 임대(Rental)

프로그램명 : visualize

수집된 데이터 저장 위치
- 시각화된 이미지는 img 폴더에 저장되어 있습니다.

프로그램을 종료합니다.
  
```

저장 결과



감사합니다