



YOUTUBE 댓글 수집 시각화 통합테스트 시나리오

Selenium, Konlpy, WebDriver_Manager, Pandas, BeautifulSoup4



대우직업능력개발원 4팀 TEAM 임대 임서인 황대명

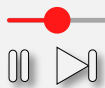


대우직업능력개발원
Daeu Development Institute for Vocational ability

START



목차



1. 테스트 개요
2. 테스트 목적
3. 테스트 전제조건
4. 댓글 수집 프로그램
 - 테스트 개요
 - 용어 설명
 - 테스트 시나리오
5. 시각화 프로그램
 - 테스트 개요
 - 테스트 시나리오



통합테스트란?

각각의 모듈이 독립적으로 작용하며,
시스템 전체가 예상대로 작동하는지를 검증하는 테스트입니다.
각 모듈을 개별적으로 테스트하고, 이를 통합하여 전체 시스템을
대상으로 테스트합니다. 이 과정에서 발생한 문제를 검출하고
수정하여 시스템의 안정성을 보장할 수 있습니다.

따라서,

프로그램의 흐름도를 통해 시나리오를 전개하여,
각 모듈 간의 호환성 문제나 시스템 전체의 문제점을 발견하기 위하여
통합테스트에 대한 내용을 다루고자 합니다.



이 문서는 유튜브 댓글 수집 및 시각화 프로젝트의 실행과정을 증명하기 위해 제작되었습니다.

이 테스트는 다음의 능력을 검증하기 위해 수행됩니다:

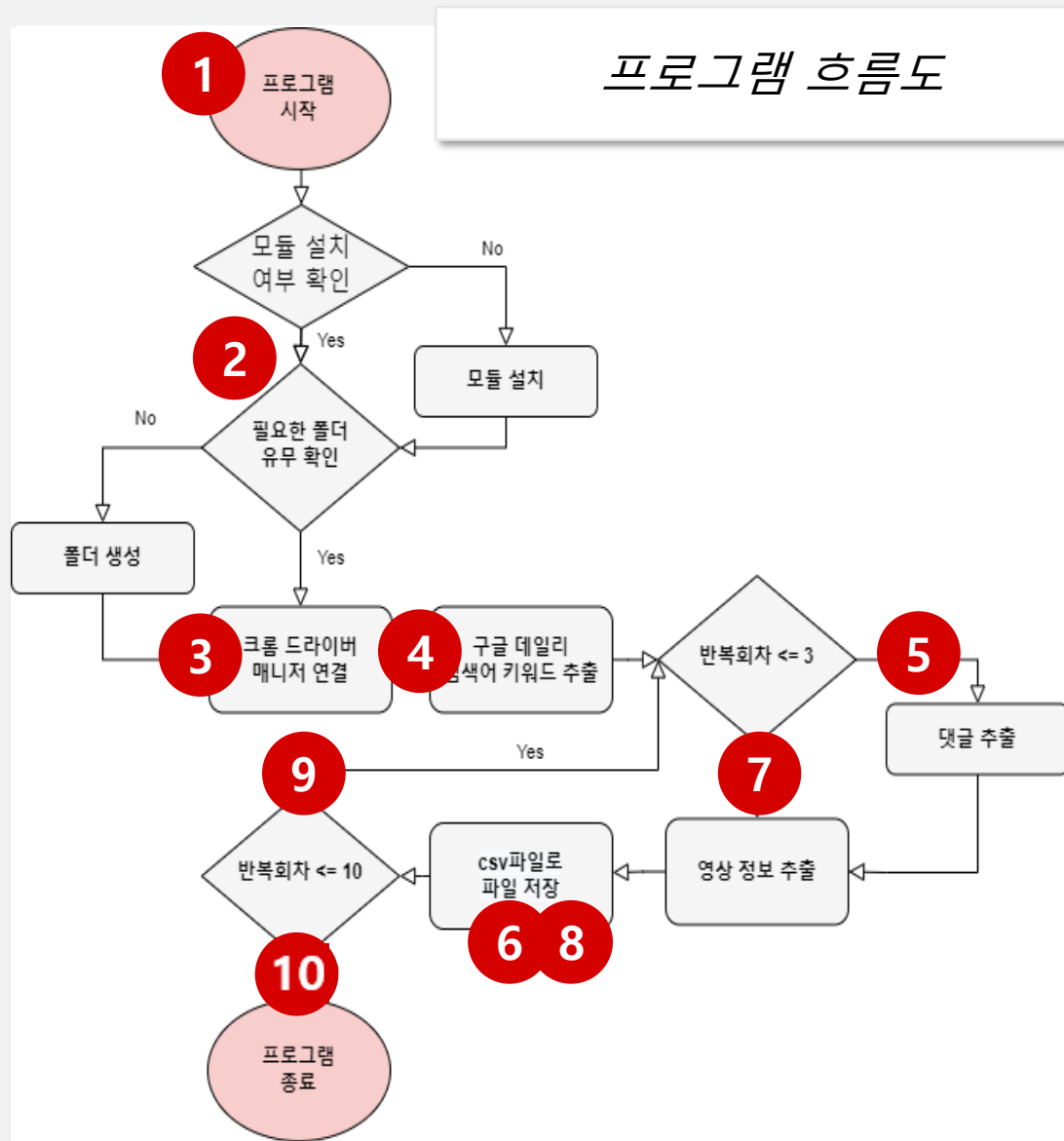
1. 필요한 라이브러리 자동 다운로드 및 전제 조건 충족
2. 구글 트렌드의 일간 인기 키워드 추출
3. 정확하고 안정적으로 유튜브 의견 수집 및 정제
- 4 분석된 데이터를 쉽게 이해할 수 있는 형식으로 시각화



해당 시나리오를 수행하기 위해서는 다음과 같은 전제조건을 충족해야 합니다.
이러한 전제조건이 모두 충족되었는지 확인하신 후에 테스트를 진행하시기 바랍니다.

1. exe 파일을 실행할 수 있는 컴퓨터 운영체제
2. 구글 크롬 웹 브라우저 설치
3. 구글 트렌드 및 유튜브에서의 검색을 위한 인터넷 연결

테스트 예상 시간은 유튜브 댓글과 대댓글 개수에 따라 다를 수 있습니다.
댓글 1000개의 수행 예상 시간은 약 2~30분 정도 소요됩니다.



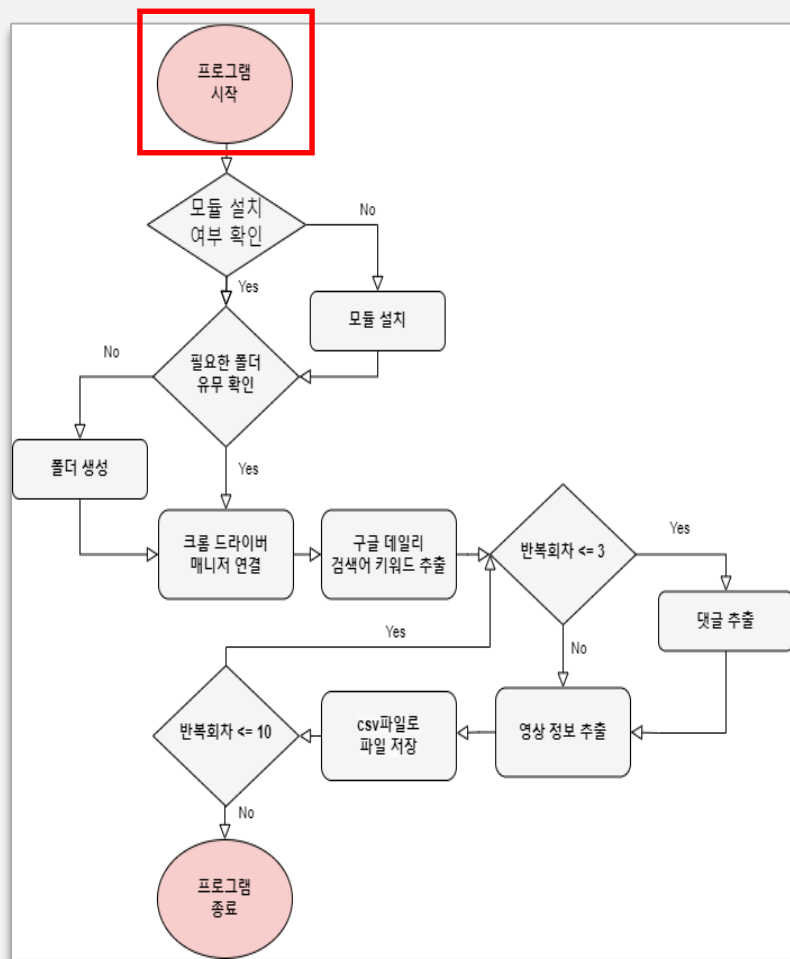
테스트 ID	테스트 시나리오
TS_CC01	댓글 수집 프로그램 시작
TS_CC02	라이브러리 설치, 필수 폴더 유무 확인
TS_CC03	크롬 드라이버 매니저 연결
TS_CC04	구글 데일리 검색어 키워드 추출
TS_CC05	영상 정보 및 댓글 수집, 정제
TS_CC06	추출된 영상정보와 댓글 저장
TS_CC07	영상 정보 수집
TS_CC08	수집된 영상정보 저장
TS_CC09	댓글 수집 프로그램 회차 반복
TS_CC10	댓글 수집 프로그램 종료

- TS : Test Scenario

- CC : Collecting Comments

TS_CC01

- 댓글 수집 프로그램 실행



실행 결과



WW203-tW공유WTeam04_Projet01WcodeWcomm_collect.exe

Team 임대(Rental)

프로그램명 : comm_collect

수집된 데이터 저장 위치

- 댓글 및 영상 수집 데이터는 comment 폴더에 저장됩니다.
- 영상정보 수집 데이터는 nocomment 폴더에 저장됩니다.

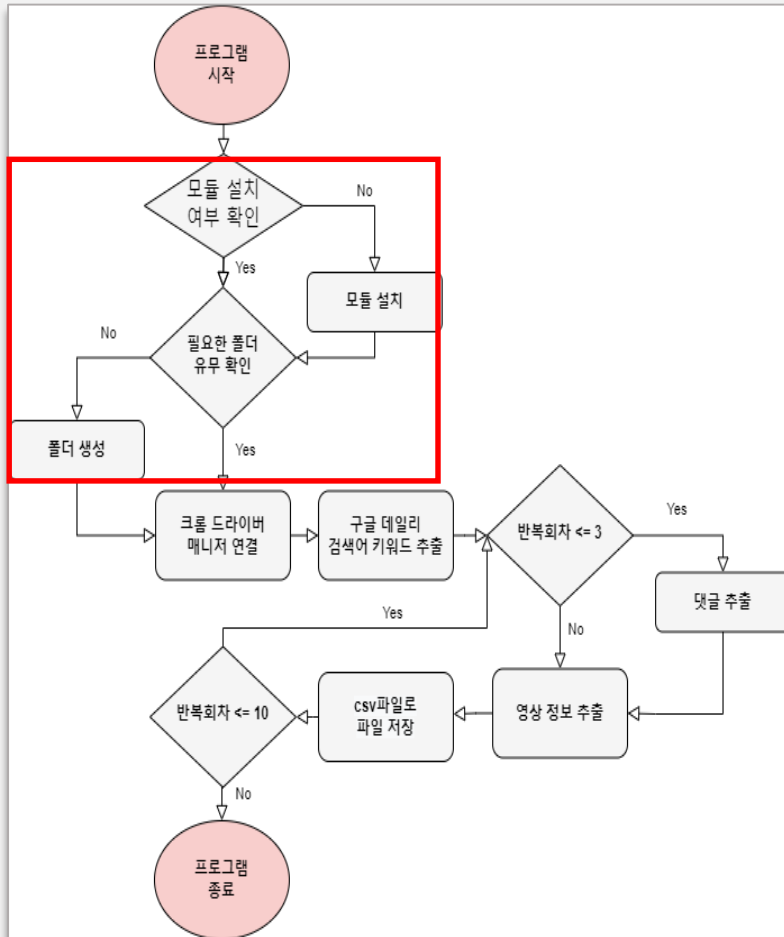
시각화를 위해 visualize.exe 를 실행시켜 주세요.

유튜브 댓글 수집 프로그램을 시작합니다.

- comm_collect.exe 실행

TS_CC02

- 라이브러리 설치 확인
- 필수 폴더 생성 확인



실행 결과

1

selenium 라이브러리 설치 여부를 체크합니다.
selenium 라이브러리가 존재합니다.

webdriver_manager 라이브러리 설치 여부를 체크합니다.
webdriver_manager 라이브러리가 존재합니다.

pandas 라이브러리 설치 여부를 체크합니다.
pandas 라이브러리가 존재합니다.

bs4 라이브러리 설치 여부를 체크합니다.
bs4 라이브러리가 존재합니다.

konlpy 라이브러리 설치 여부를 체크합니다.
konlpy 라이브러리가 존재합니다.

폴더 생성 여부를 체크합니다.

2

생성된 필수 폴더 목록

 nocomment

 comment

1

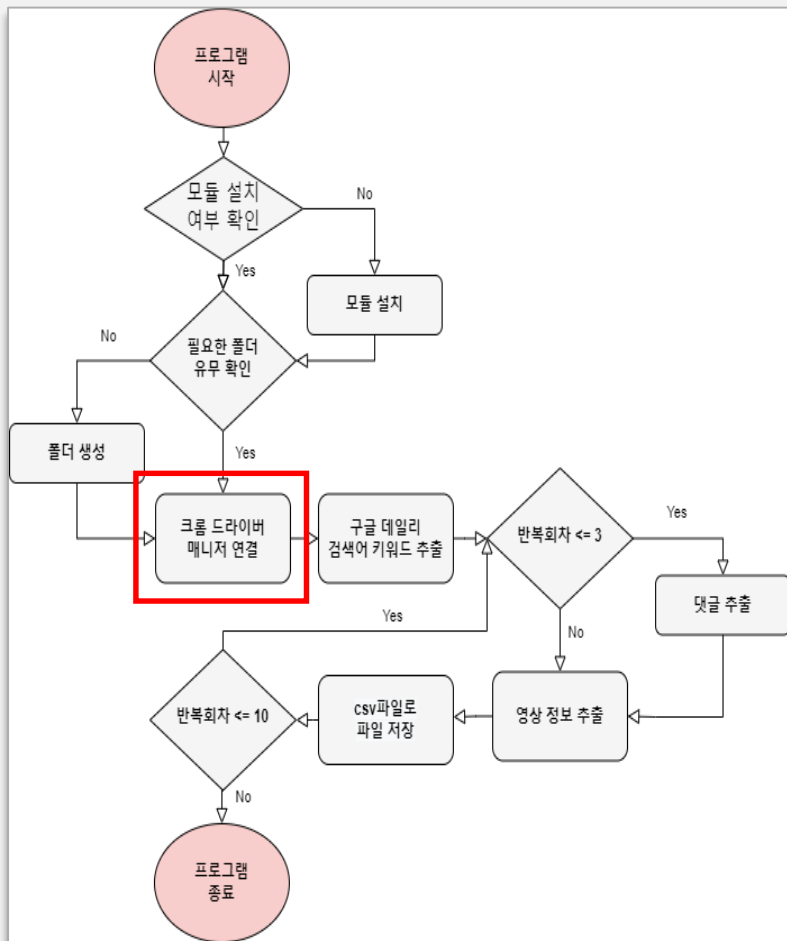
모듈 설치 여부 확인 후 해당 모듈 설치

2

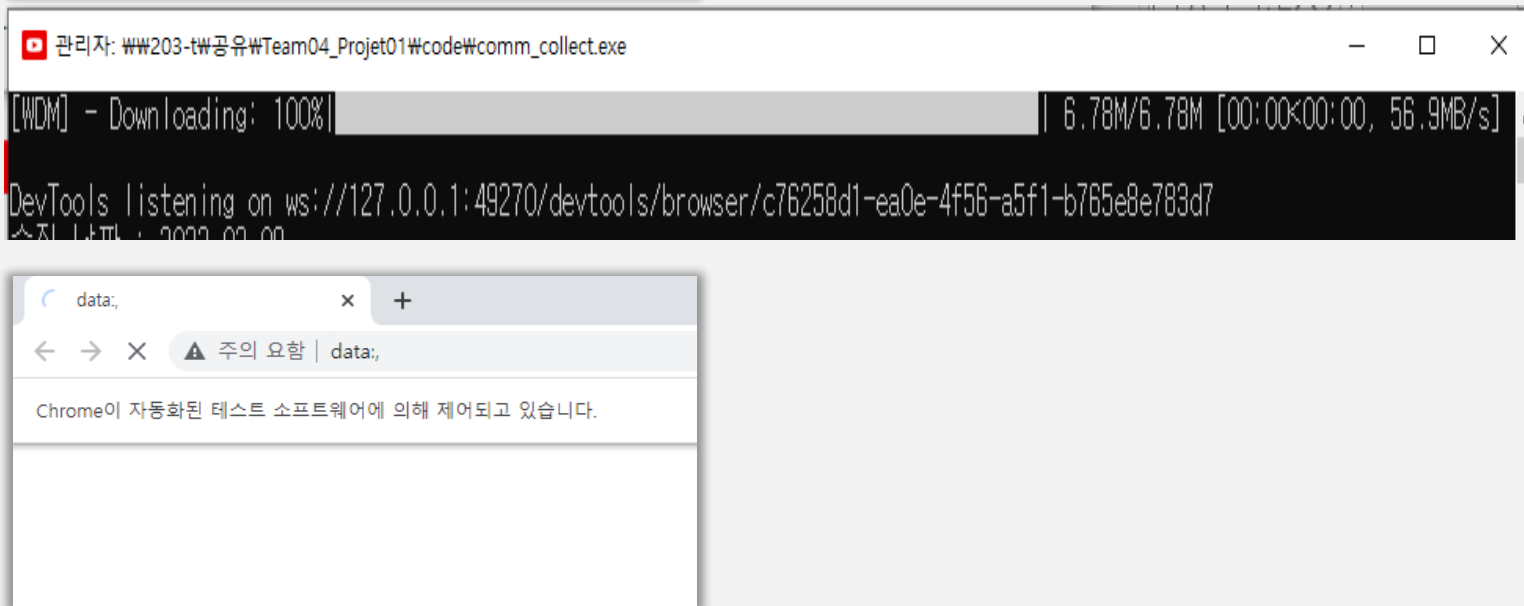
필수 폴더 유무 확인 후 폴더 생성

TS_CC03

- 크롬 드라이버 매니저 연결



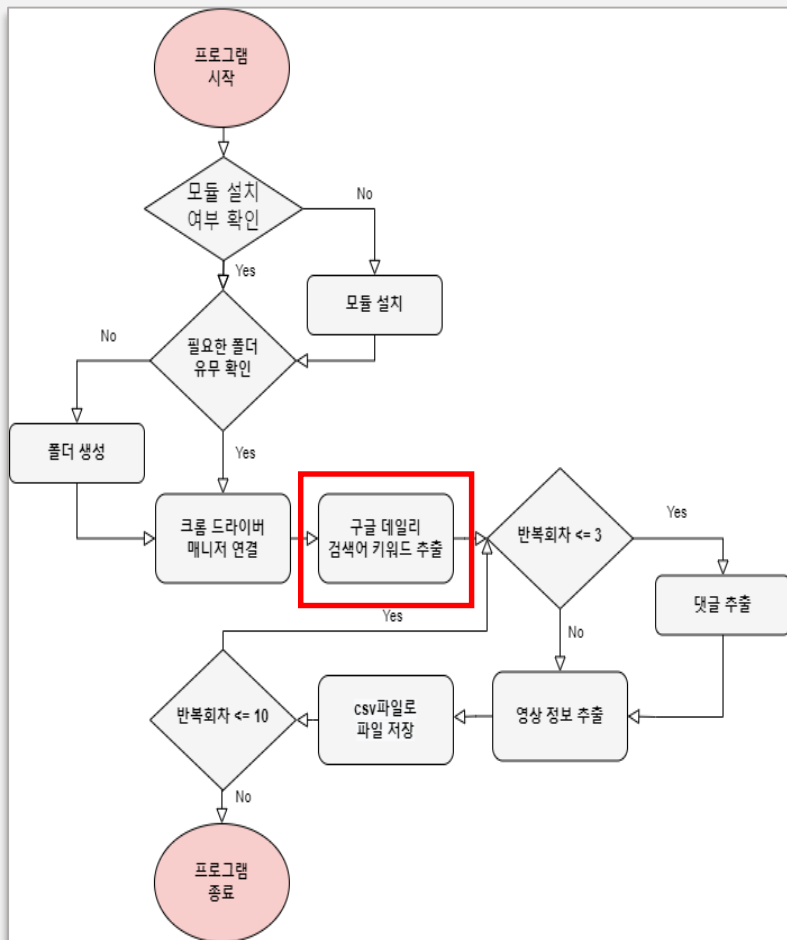
실행 결과



- 버전에 맞는 크롬 드라이버를 다운로드 하여 크롬 실행

TS_CC04

- 구글 데일리 검색어 키워드 추출



실행 결과

1

주제 탐색

일반 인기 급상승 검색어

2023년 3월 7일 화요일

1 나이키

나이키 티파니 에어포스1 출시, 한국발매일·가격은? 국제·

2 더불어민주당

아침논단...연속 개원 이어가는데 일하지 않는 국회 대구문

2

나이키 -Shorts

관련 필터

광고·추천 제품

검색 필터 열기



나이키 스우시 남성
트레이닝 상하의 ...
수량 W79,800



나이키코리아 덩크 로
우 맨즈 RETRO 빈고...
수량 W275,000



나이키 DRY 라이트웨
이트 글러브 다용도...
수량 W26,800



나이키 부니 버킷 햇
Nike
수량 W22,700



나이키 NIKE 나이키
세티 클래식 부츠...
수량 W83,000

Google은 상품 판매의 당사자가 아닙니다.



예측이 틀려도 성공할 수 있다는 증거, 나이키
조회수 114만회 · 2년 전
14F 일사애프
세계 최고의 스포츠 패션 브랜드 나이키. 하지만 나이키도 처음부터

1

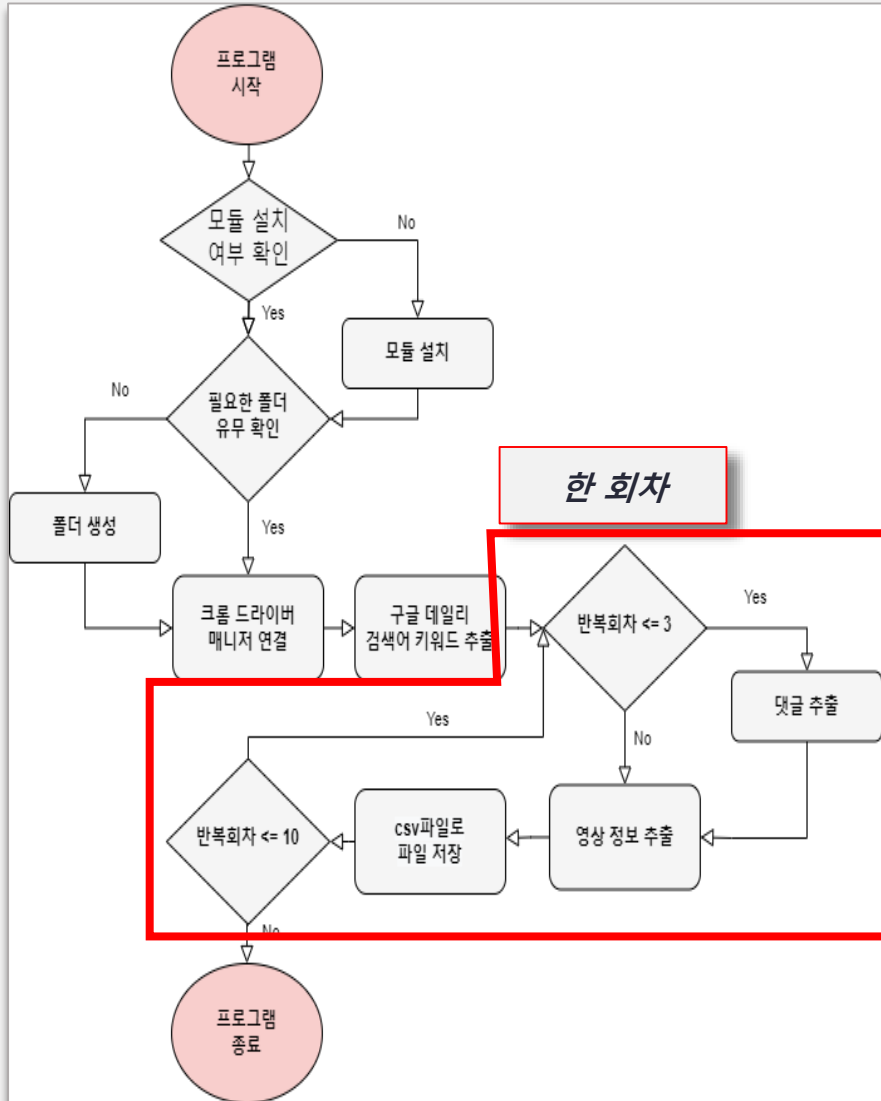
구글 트렌드에 접속 후 1위 키워드 추출

2

유튜브 검색 창에 전송 후 검색

개선 사항

예외 처리를 통해 Shorts 영상과 실시간 영상을 제외

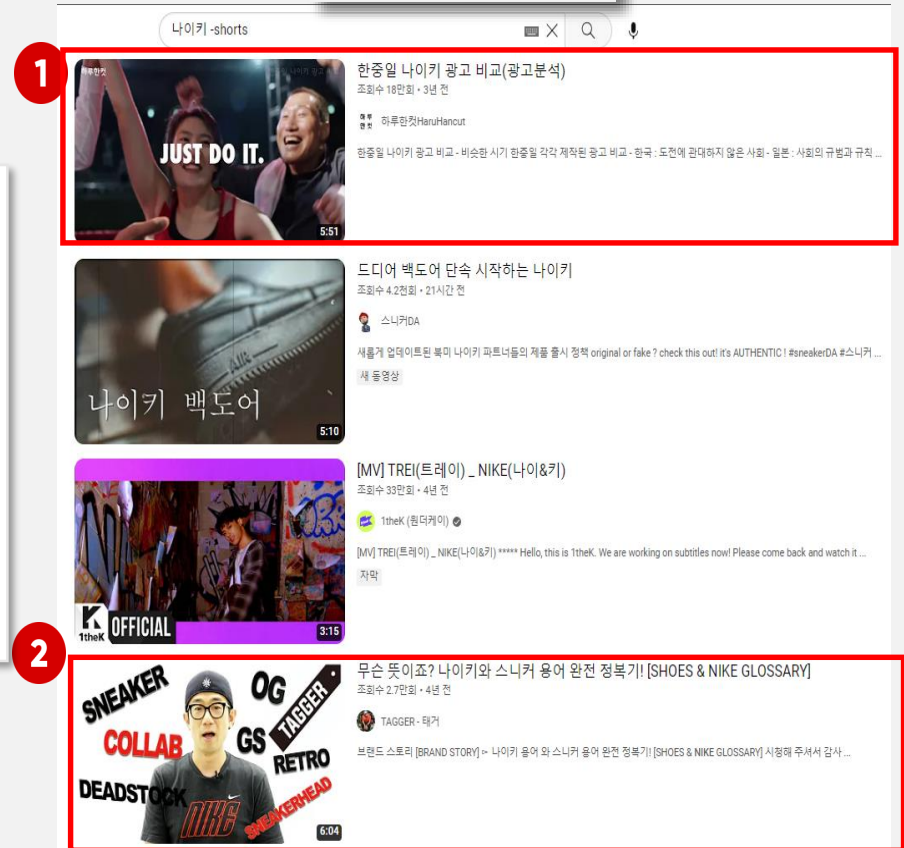


반복회차란?

유튜브에 키워드 검색 후 뜨는
첫번째 동영상을 index = 1,
반복회차=1이라 칭한다.

한 회차 후에 다시 키워드 검색
화면으로 돌아와 두번째 영상을
클릭 시 반복회차 = 2가 된다.

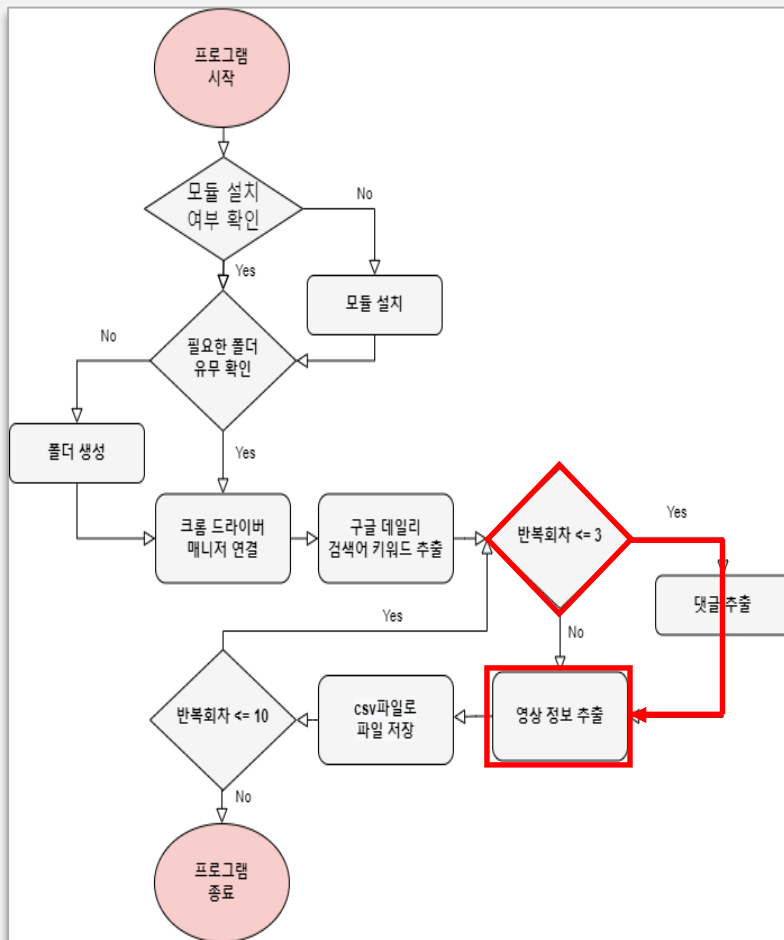
키워드 검색 화면



- 1 반복회차 ≤ 3 인 경우, 영상정보와 댓글을 수집
- 2 $3 > \text{반복회차} \geq 10$ 인 경우, 영상정보만 수집

TS_CC05

- 영상 정보 및 댓글 수집
- 댓글 및 대댓글 정제



실행 결과

1

herry 2년 전
조민... 인간 역사상 최고의 스포츠스타
4 0 댓글

현민
이현민 1년 전
나이키의 just do it이라는 말도 한 사형수의 말에서 영감을 얻어 사용하게됐다고 하
5 0 댓글

경선
양경선 2년 전
조현용님 영상 능한 줄 알고 지난 2주간 계속 들춰볼라 했습니다. 다시 영상 보게되
2 0 댓글

스님
스님 4개월 전
진짜 이런 비하인드 이야기들은거 너무 재미있당...
0 0 댓글

산스
산스 2년 전
동족이지만 인색이 좋았군요... 또한 남의 의견을 잘 수용하는 인정이 있었기에 가능
잘 보고 가요!

2

날짜 : 2023.03.07
이름 : 예측이 틀려도 성공할 수 있다는 증거, 나이키 스토리 (그러
채널 이름 : 14F 일사에프
조회수 : 1,146,350회
게시일 : 2021. 1. 17.
검색어 : 나이키
스크롤 시작 :14:48:55
대댓글 열기 시작 :14:51:12
댓글 추출 시작 :15:15:12
{*하다*: 261, *나이키*: 237, *보다*: 98, *있다*: 91, *되다*: 84, *

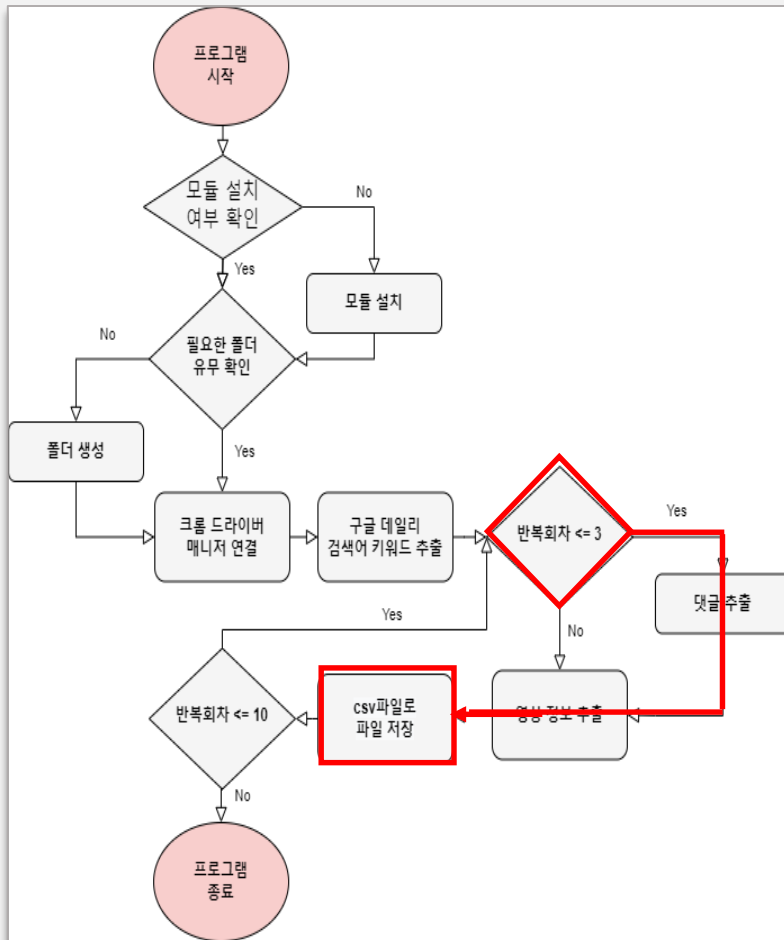
- 반복회차가 3회 이하인 경우, 영상정보를 수집한 뒤 스크롤 및 대댓글 토글 선택으로 댓글과 대댓글 수집
- 수집된 댓글을 자연어 처리하여 정제한다.

개선 사항

예외 처리를 통해 충분하지 않은 양의 댓글을 가진 영상 제외

TS_CC06

- 수집된 영상정보와 댓글 저장



실행 결과

2023.03.07_나이키_1
2023.03.07_나이키_2
2023.03.07_나이키_3

[comment 디렉토리 내부]

1	수집 날짜	2023.03.07			
	제목	나이키 운동화를 빼앗긴 동생의 원수를 갚			
	채널	영리한곰			
	조회수	190,839회			
	게시일	2020. 3. 30.			
2	나이키	26			
	하다	23			
	아니다	11			
	신발	9			
	지금	9			
	신고	8			
	있다	8			
	이다	7			
	영화	7			
	보다	7			
	류승범	6			
	진짜	5			
	연기	5			

[csv 파일 내부]

1 영상정보

2 댓글

- 반복회차 <= 3 인 경우,
수집한 영상정보와 댓글을 csv파일 형태로 #comment 폴더에 저장

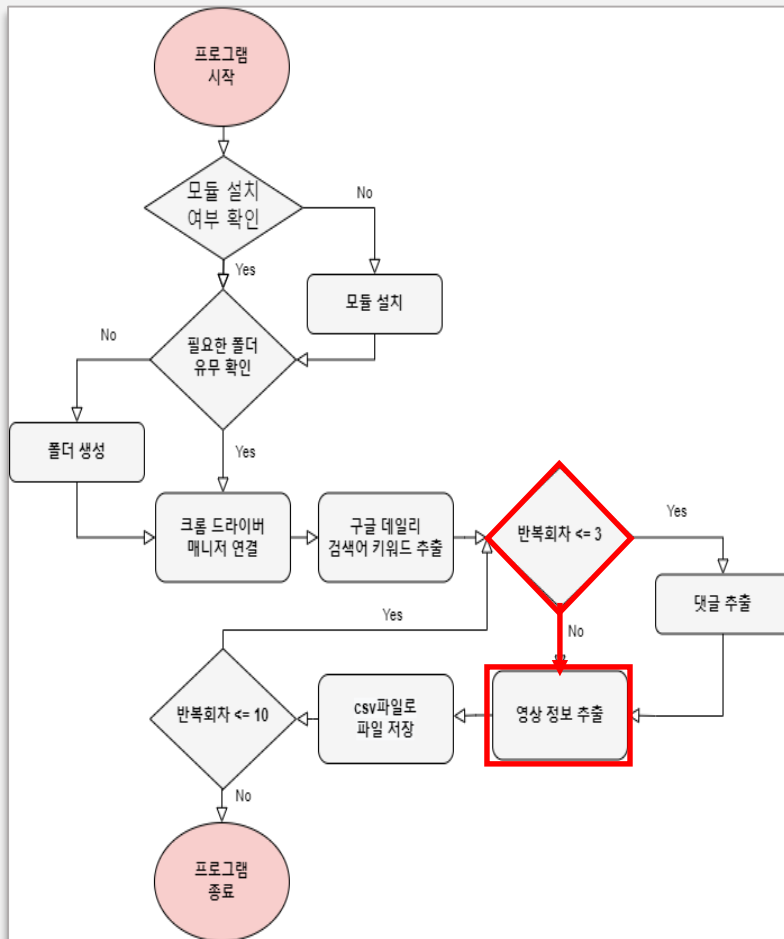
TS_CC07

- 영상 정보 수집

실행 결과

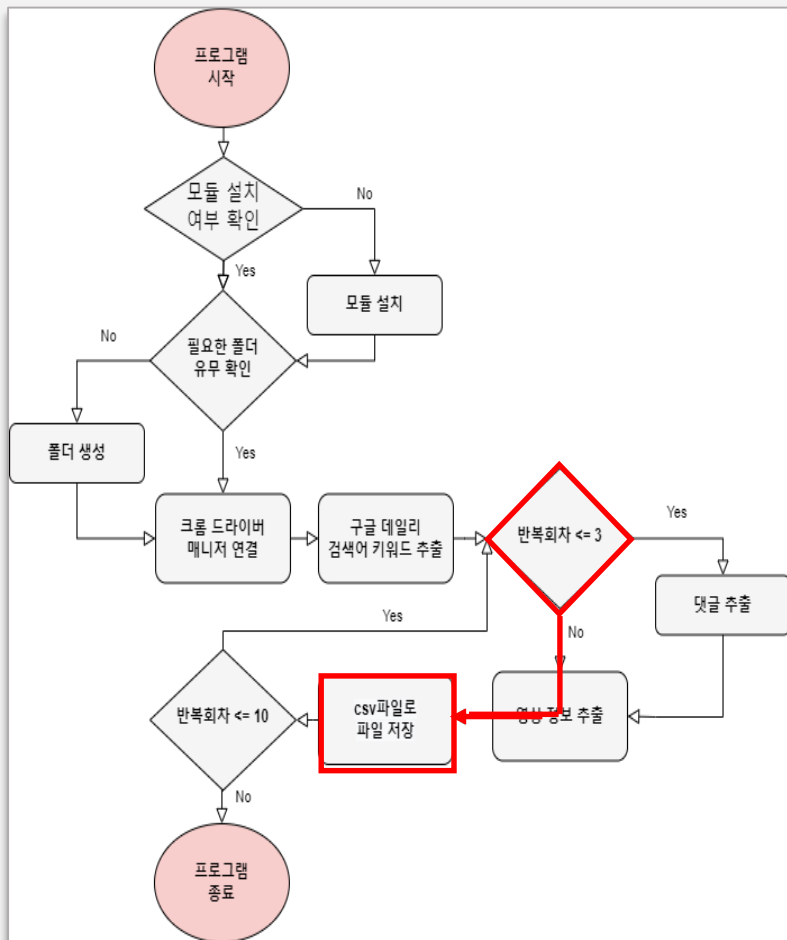
수집 날짜 : 2023.03.07
 영상 이름 : 예측이 틀려도 성공할 수 있다는 증거, 나이키 스토리 (그러니까 좌절금지!
 채널 이름 : 14F 일사에프
 조회수 : 1,146,350회
 게시일 : 2021. 1. 17.
 검색어 : 나이키

- 반복회차가 3회 초과 10회 이하인 경우, 영상정보만을 수집한다.



TS_CC08

- 수집된 영상정보 저장



실행 결과

2023.03.07_나이키_4
 2023.03.07_나이키_5
 2023.03.07_나이키_6
 2023.03.07_나이키_7
 2023.03.07_나이키_8
 2023.03.07_나이키_9
 2023.03.07_

[nocomment 디렉토리 내부]

수집 날짜	2023.03.07			
제목	[아울렛 탐방기] 이천 롯데 프리미엄 아울렛 - 나이키,			
채널	포모스트			
조회수	2,482회			
게시일	2023. 3. 5.			

수집 날짜	2023.03.07			
제목	나이키 공홈 선착구매 줌 보메로 5 포톤 더스트 앤 메탈			
채널	신발왕			
조회수	1,311회			
게시일	2023. 3. 3.			

[csv 파일 내부]

수집 날짜	2023.03.07			
제목	Nike Football : The Last Game - 한국 자막			
채널	Nightcus			
조회수	227,364회			
게시일	2014. 6. 11.			

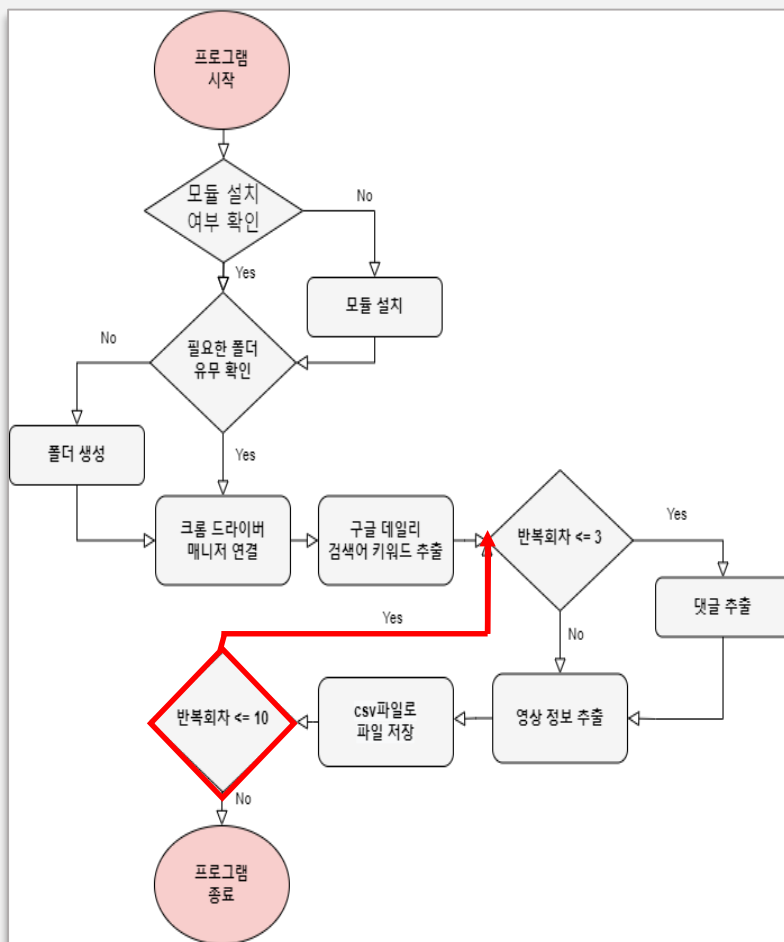
- 3 < 반복회차 <= 10인 경우,
 영상의 정보를 csv파일 형태로 #nocomment 폴더에 저장

TS_CC09

- 댓글 수집 프로그램 회차 반복

실행 결과

- 반복회차가 10회 이하일 경우, 한 회차를 다시 반복

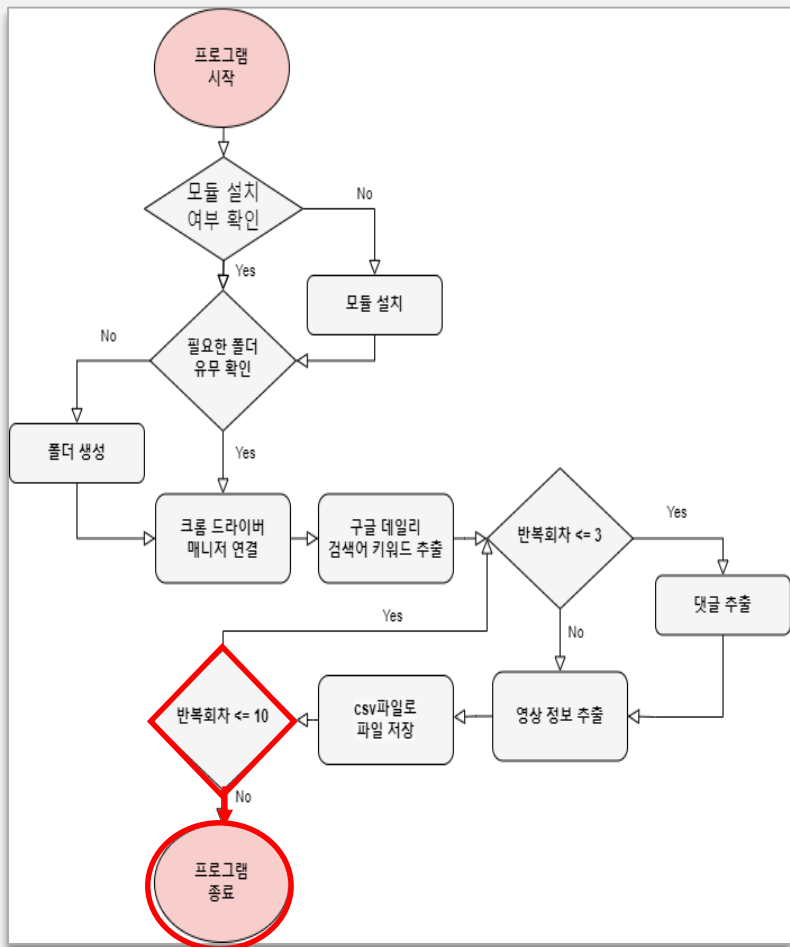


TS_CC10

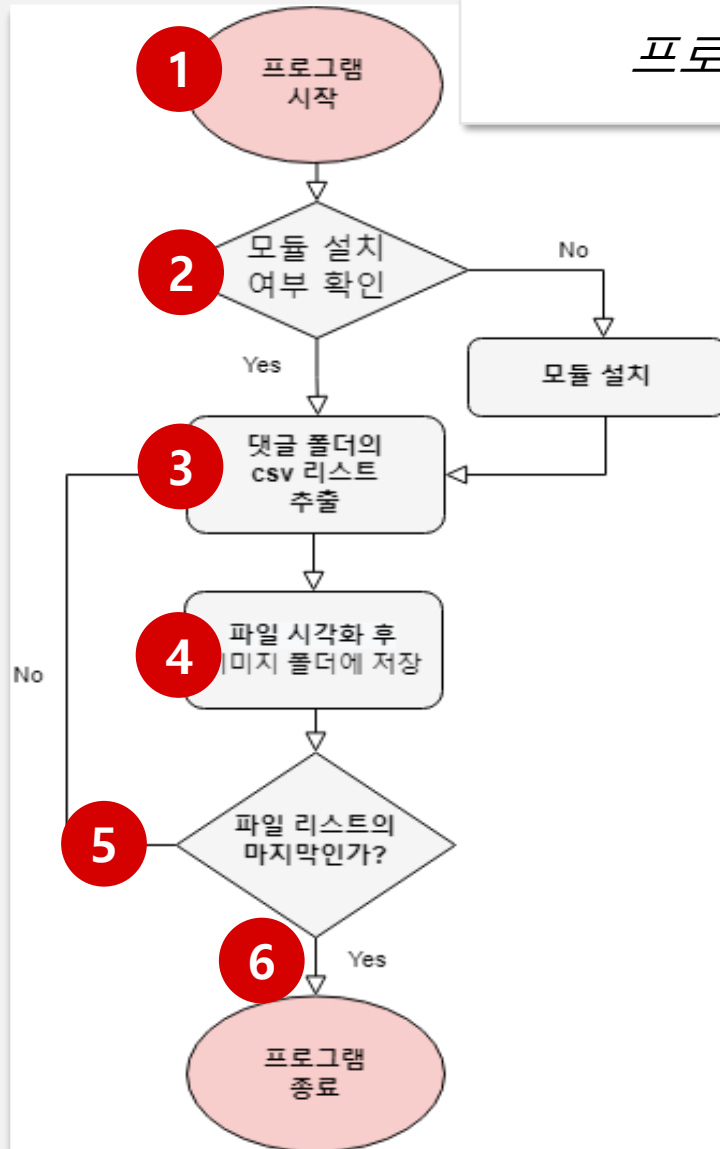
- 댓글 수집 프로그램 종료

실행 결과

- 반복회차가 10회 초과일 경우, 프로그램을 종료



프로그램 흐름도



테스트 ID

테스트 시나리오

TS_VSL01

시각화 프로그램 시작

TS_VSL02

라이브러리 설치, 필수 폴더 유무 확인

TS_VSL03

댓글 폴더의 csv리스트 추출

TS_VSL04

파일 시각화 후 이미지 폴더에 저장

TS_VSL05

시각화 프로그램 반복

TS_VSL06

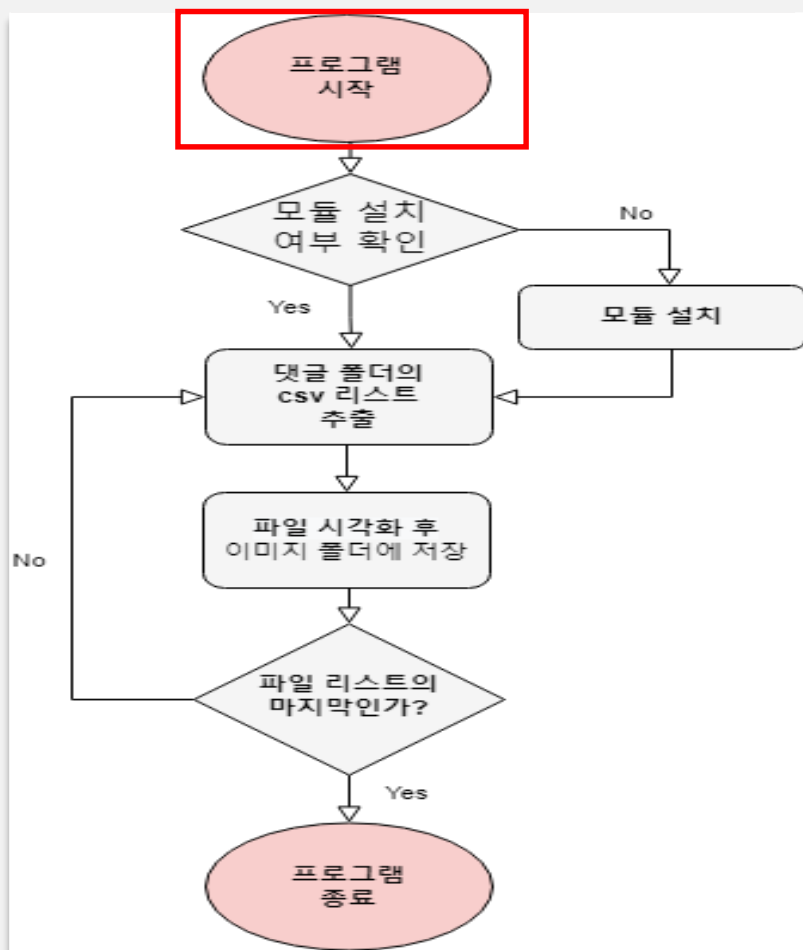
시각화 프로그램 종료

- TS : Test Scenario

- VSL : Visualize

TS_VSL01

- 시각화 프로그램 시작



실행 결과



visualize.exe

```

WWW203-t\공유\Team04_Projet01\code\visualize.exe

Team 임대(Rental)

프로그램명 : visualize

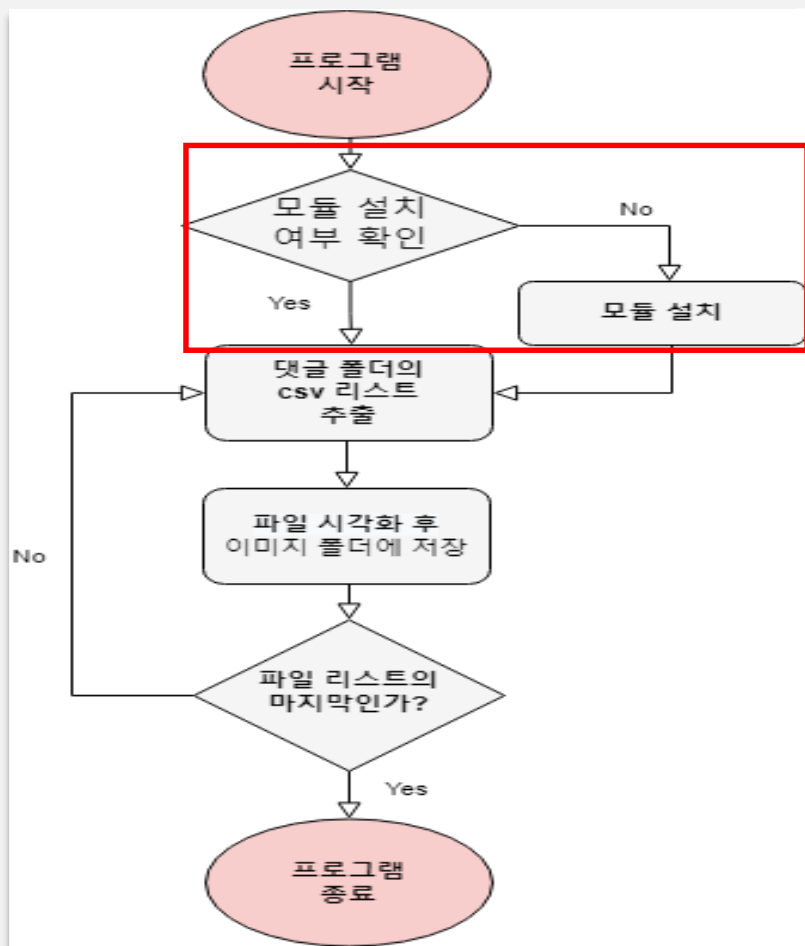
수집된 데이터 저장 위치
- 시각화된 이미지는 img 폴더에 저장됩니다.

시각화 프로그램을 시작합니다.
  
```

- visualize.exe 실행

TS_VSL02

- 라이브러리 설치 확인
- 필수 폴더 생성 확인



실행 결과

1

```

pandas 라이브러리 설치 여부를 체크합니다.
pandas 라이브러리가 존재합니다.

wordcloud 라이브러리 설치 여부를 체크합니다.
wordcloud 라이브러리가 존재합니다.

matplotlib 라이브러리 설치 여부를 체크합니다.
matplotlib 라이브러리가 존재합니다.
  
```

2

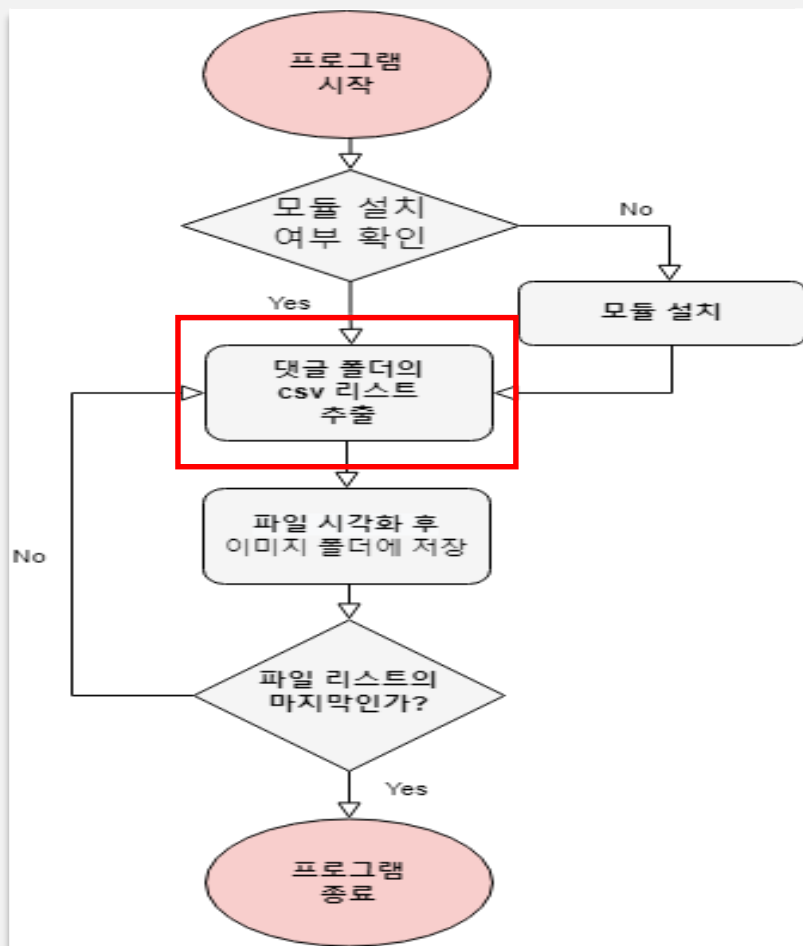


생성된 필수 폴더

1. 모듈 설치 여부 확인 후 해당 모듈 설치
2. 필수 폴더 유무 확인 후 폴더 생성

TS_VSL03

- 댓글 폴더의 csv리스트 추출



실행 결과

```

a_ 2023.03.07_나이키_1
a_ 2023.03.07_나이키_2
a_ 2023.03.07_나이키_3 [ comment 디렉토리 내부 ]
  
```

파일 리스트 목록

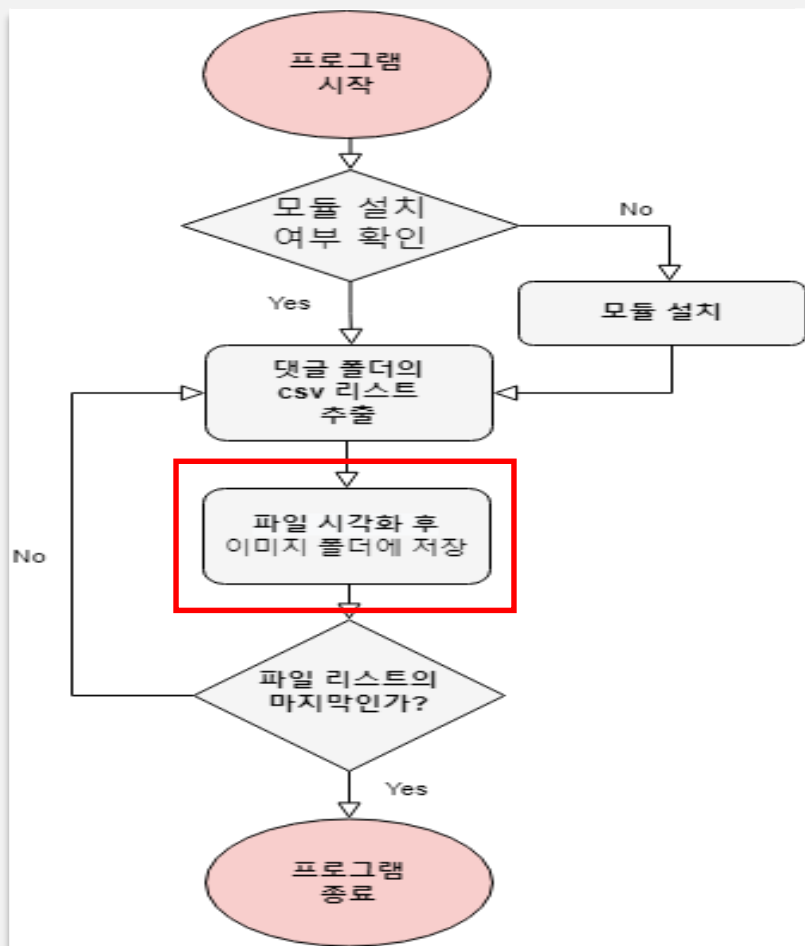
```

2023.03.07_나이키_1.csv 2023.03.07_나이키_2.csv 2023.03.07_나이키_3
  
```

- 댓글 폴더 내 파일 리스트를 불러와 내용을 추출

TS_VSL04

- 파일 시각화 후 이미지 폴더에 저장



실행 결과

1

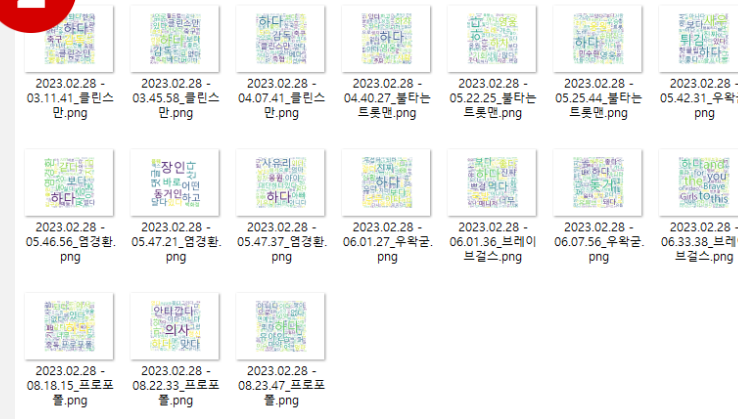
2023.03.07_나이키_1.csv 파일 wordcloud 시작
 2023.03.07_나이키_1.png로 img 폴더에 저장

2023.03.07_나이키_2.csv 파일 wordcloud 시작
 2023.03.07_나이키_2.png로 img 폴더에 저장

2023.03.07_나이키_3.csv 파일 wordcloud 시작
 2023.03.07_나이키_3.png로 img 폴더에 저장

2023.03.07_나이키_1.png
 2023.03.07_나이키_2.png
 2023.03.07_나이키_3.png

2

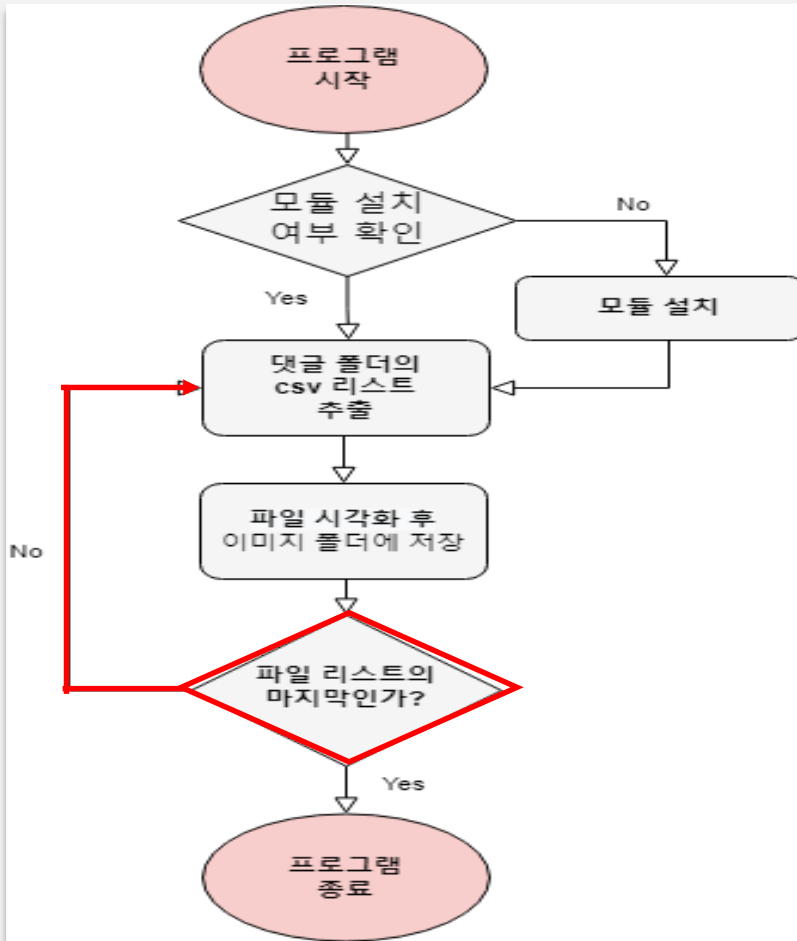


- 1 csv파일 목록을 시각화 하여 이미지 파일로 저장
- 2 comment문서 내의 모든 csv파일을 한번에 시각화

TS_VSL05

- 시각화 프로그램 반복

실행 결과

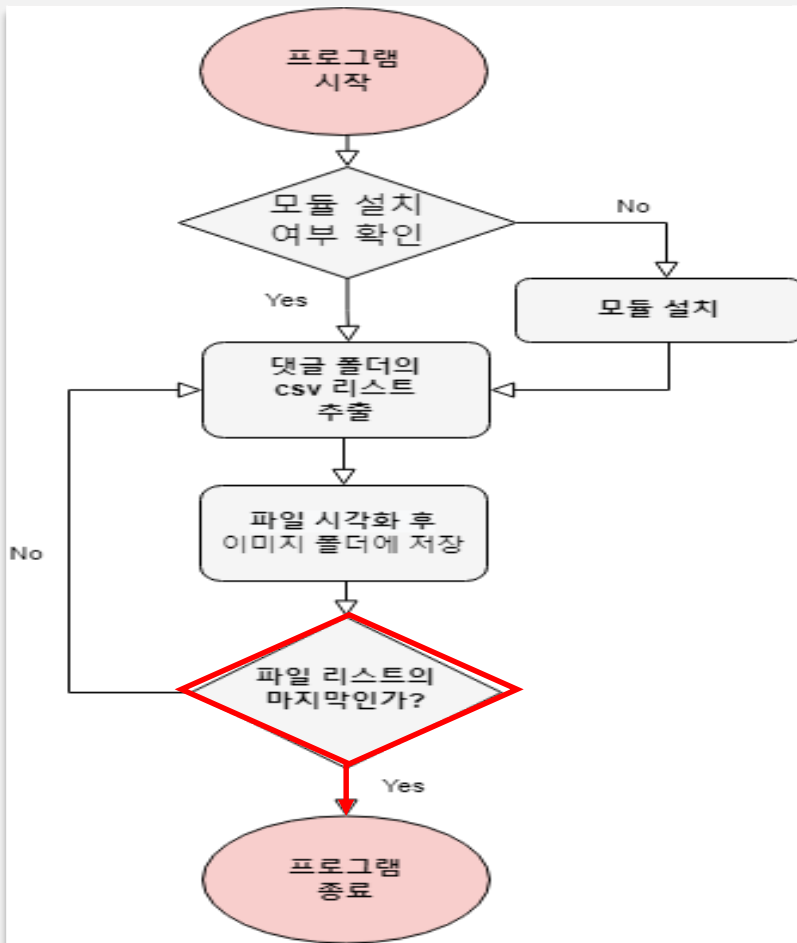


- 시각화한 파일이 comment폴더의 마지막 요소가 아닐 때, 다시 csv리스트를 추출 후 시각화

TS_VSL06

- 시각화 프로그램 종료

실행 결과



- comment 폴더의 마지막 요소까지 시각화 했을 경우,
프로그램 종료

감사합니다