



유튜브 알고리즘 세탁기

24기 DE 조운영



목차

- 왜?
- 무엇을?
- 어떻게? (어떻게든...)
- 누구랑?

왜 ?

1. 유튜브 알고리즘이 추천해주는 콘텐츠들로 내 시간을 채우는 거 별로...



2. 구독을 난사하기 시작

4

왜 ?

3. 구독한 채널의 컨텐츠들로도 제대로된 필터링 불가

- 구독 채널 수 너무 많음 (한 계정에만 133개...?!)
- 다양한 관심사를 가지고 있거나, 시시때때로 관심사가 변함

왜 ?

4. 구독한 채널들의 콘텐츠를 카테고리에 따라 자동으로 필터링해주는 서비스가 있으면 좋겠다

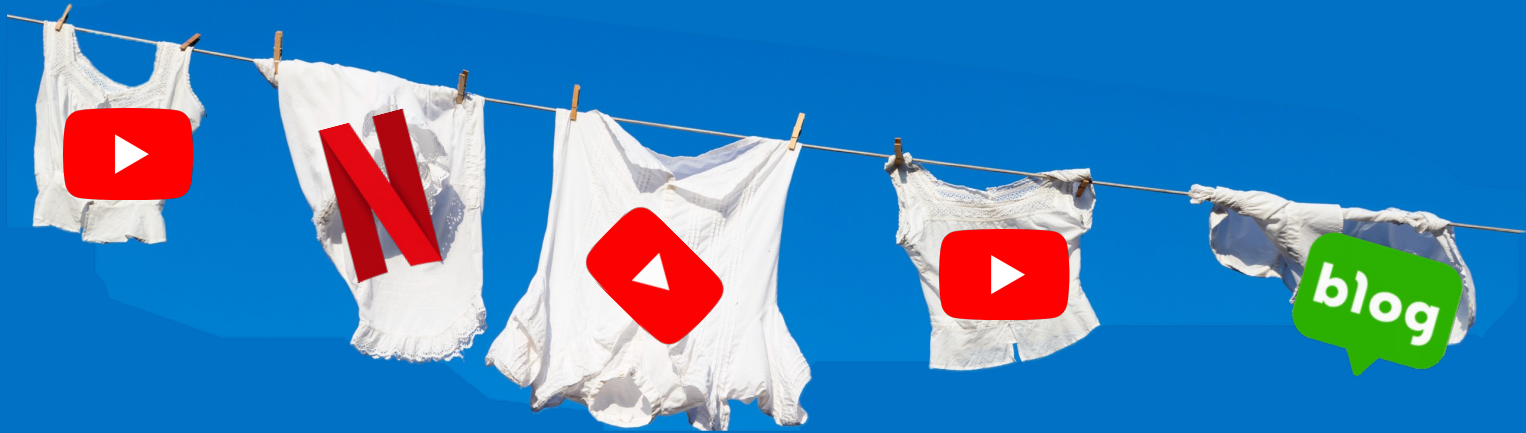
- 내 알고리즘은 내가 설계한다!
(내 인생은 내가 설계한다!)



무엇을 ?

. 구현하고자 하는 서비스 :

내가 구독한 채널의 컨텐츠들을
카테고리(노래/여행/영화/춤/수학/옷/요리/브이로그 등)에 따라
필터링해서 볼 수 있는 웹페이지



무엇을 ?



Premium ^{KR}

검색



홈



Shorts



구독



YouTube Mu...



나



오프라인 저장 ...

홈

동영상

Shorts

재생목록

커뮤니티



플레이리스트

영화

요리

브이로그

춤

옷

여행

운동

수학

코딩

스포츠



37:55

[playlist] 나른한 오후, 그 따스한 햇살과 함께:)

조회수 681만회 • 4년 전



3:25

keshi - 2 soon [가사해석/번역/한글자막]

조회수 605만회 • 4년 전



32:44

[playlist] 그냥.. 오늘 니가 꼭 들었음 하는 노래

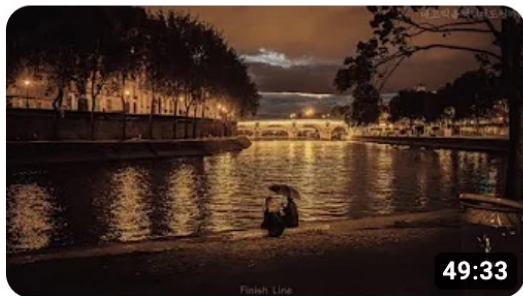
조회수 604만회 • 4년 전



1:22:29

[playlist] 초여름의 피크닉

조회수 593만회 • 1년 전



49:33



41:11



1:28:33



46:46

어떻게 ?

. 필요한 데이터셋 :

유튜브 API  YouTube > Data API

<https://developers.google.com/youtube/v3/docs/subscriptions?hl=ko>

<code>snippet.title</code>	string 동영상 제목입니다. 속성 값의 최대 길이는 100자이며 < 및 > 유효한 UTF-8 문자를 모두 포함할 수 있습니다. videos.update 메서드를 호출하고 video 리소스의 snippet 부분을 업데이트할 때 이 속성 값을 설정해야 합니다.
<code>snippet.description</code>	string 동영상 설명. 속성 값의 최대 길이는 5,000바이트이며 < 및 > 고 유효한 모든 UTF-8 문자를 포함할 수 있습니다.

`snippet.tags[]`

list

동영상과 연결된 키워드 태그의 목록입니다. 태그에는 공백이 수 있습니다. 속성 값의 최대 길이는 500자(영문 기준)입니다 제한 계산 방법과 관련된 다음 규칙을 참조하세요.

- 속성 값은 목록이며 목록 항목 사이의 쉼표가 한도에 포함
- 태그에 공백이 포함된 경우 API 서버는 태그 값을 따옴표로 처리하며, 따옴표는 글자 수 제한에 합산됩니다. 따옴표 수 제한을 위해 **Foo-Baz** 태그에는 7자가 포함되지만 **Foo**에는 9자가 포함됩니다.

`snippet.categoryId`

string

동영상과 연결된 YouTube 동영상 카테고리입니다. **videos.update** 메서드를 호출하고 **video** 리소스의 **snippet** 부분을 업데이트하는 경우 이 속성의 값을 설정해야 합니다.

어떻게 ?

- 필요한 데이터셋(확장) :

(그 외 OTT들의 API 또는 OTT 콘텐츠들에 대한 정보를 담은 데이터

<https://www.justwatch.com/us/new>)



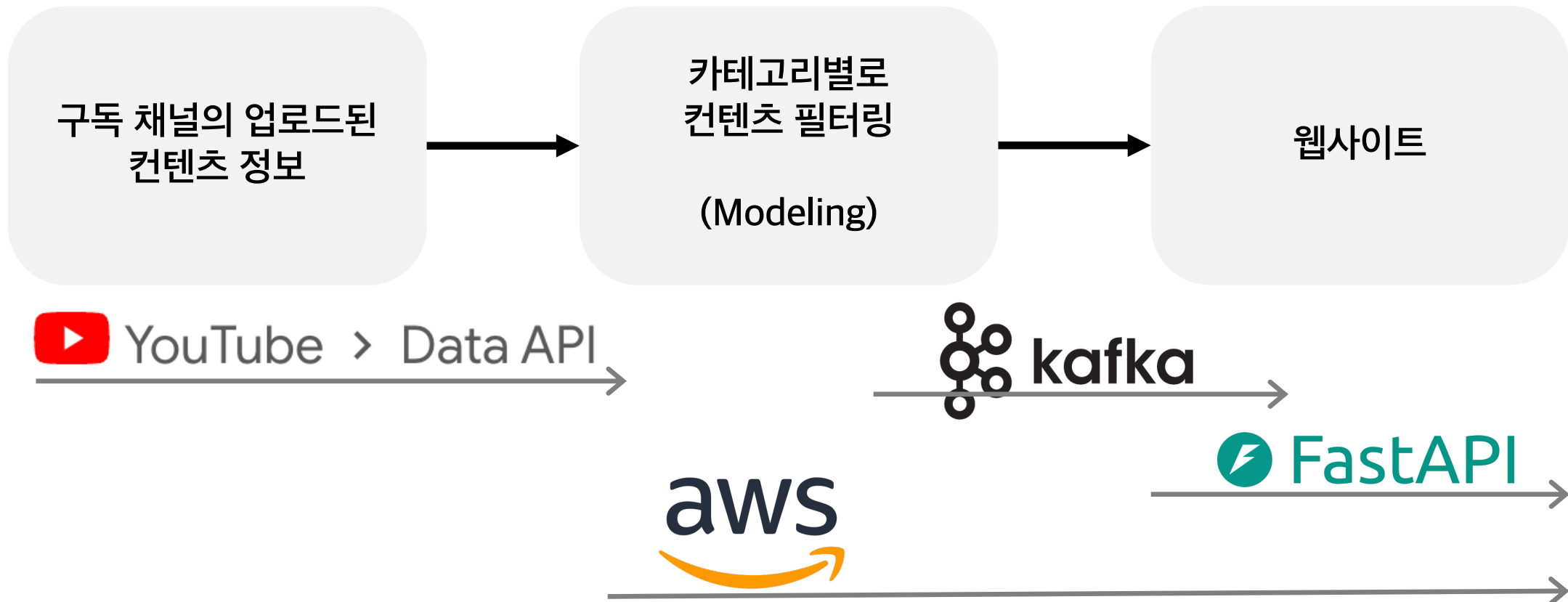
어떻게 ?

. 데이터 활용 및 기술스택 (...?) : (+ AWS, Docker)

- 데이터 처리: 필터링한 콘텐츠 정보를 데이터베이스로 담고 실시간으로 업데이트 - Kafka? Kinesis?
- 모델링: API를 통해 불러온 동영상의 카테고리 정보, 영상 제목 및 설명 등을 분석하여 카테고리 별로 분류 - NLP, Clustering 등
- 백엔드, 프론트엔드: 웹사이트 구현 - FastAPI? Spring Boot?, React?

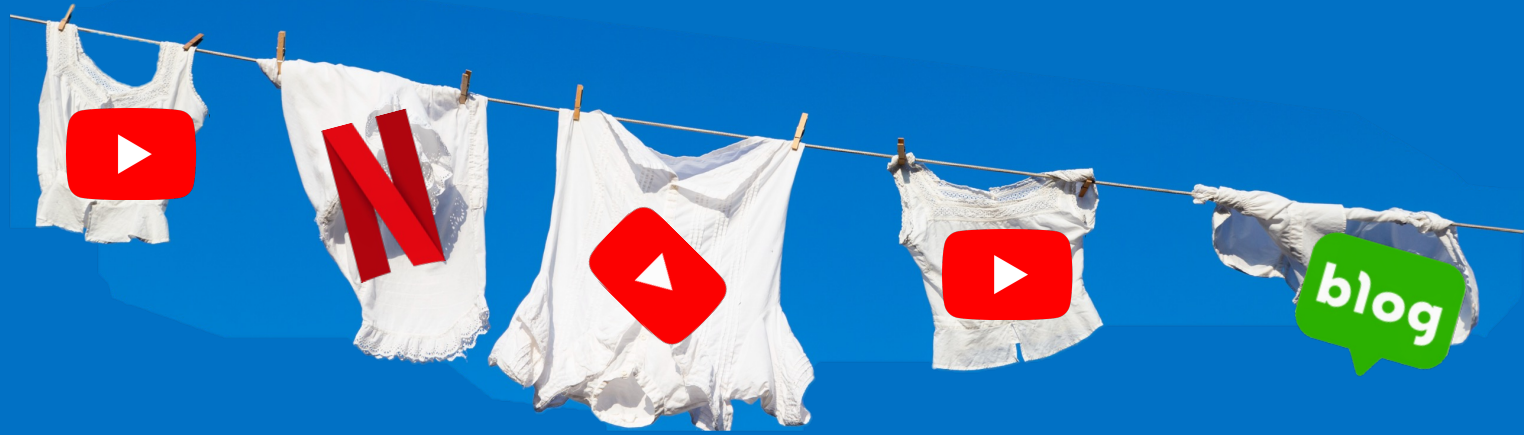
어떻게 ?

. 결과물의 전반적인 구조 (...?) :



누구랑 ?

- 어떻게든 끝까지 해보고 싶은 사람
- 적극적으로 팀원들과 소통하면서 공부할 사람
(노선을 통해 실시간으로 진행 상황 공유, 주 1회 대면 회의 예정)
- 주제가 재밌을 거 같은 사람, 유튜브 좋아하는 사람
- 유튜브 API, 스트리밍 데이터 처리, NLP와 클러스터링,
백엔드 및 프론트엔드 구현 등에 관심 있는 사람



THANK YOU