

2024.06.28

'24-1 ybigta 컨퍼런스 최종발표



Personalized YouTube Algorithm Washer

곽민규 · 김예진 · 김현호 · 박경욱 · 이승준 · 임종혁 · 조운영

YourTube 목차

- 01 PROBLEM STATEMENT** _____
- 02 ARCHITECTURE** _____
- 03 MODELING** _____
- 04 CHALLENGING TASKS** _____
- 05 DEMONSTRATION** _____
- 06 CONCLUSION** _____



01

PROBLEM STATEMENT

01 PROBLEM STATEMENT



문제점 (1) : YouTube 알고리즘에 뜨는 쓸 데 없는 콘텐츠만 소비



해결책 (1) : 채널 구독을 통해 내가 보고 싶은 채널의 영상만 모아보기

01 PROBLEM STATEMENT



문제점 (2) : 구독한 채널들의 주제가 다양하고 많아서 필요한 콘텐츠를 찾기 힘들



해결책 (2) : 구독 채널의 영상을 카테고리 별로 필터링해서 볼 수 있도록 하자!

01 PROBLEM STATEMENT

...그리하여 등장한 한 줄기 빛이 있었으니, 그것은 바로



유튜브 알고리즘 세탁기

 github.com/YBIGTA/24-1_YourTube

01 PROBLEM STATEMENT

Q : 그렇다면 **YouTube**는 어떤 프로그램인가요?

A : ... **n**행시 운 띄워주세요 **Drop the Beat**

유튜브

알고리즘

튜닝

브라우저 익스텐션

YouTube

algorithm

Tuning

Browser Extension

01 PROBLEM STATEMENT

다시 말해서,

유튜브 **YouTube**에서 구독한 채널 영상들을
알고리즘 전체 & 맞춤형 카테고리에 따라서
튜닝 필터링함으로써 깔끔하게 재정렬해주는
브라우저 익스텐션 **Chrome** 확장프로그램



02

ARCHITECTURE

02 ARCHITECTURE

YouTube 자체 카테고리 분류 ID를 참고하여 15개의 전체 카테고리 설정



영화/애니메이션



자동차/교통



음악



반려동물/동물



스포츠



여행/이벤트



게임



인물/블로그



코미디



엔터테인먼트



뉴스/정치



노하우/스타일



교육



과학기술



비영리/사회운동

02 ARCHITECTURE

Feature (1)

**YouTube 구독 웹사이트의 구성요소와 UI를 변경하여
기존의 단순한 최근 업로드 순 정렬이 아닌, 여러 카테고리 중 커스텀하여 재정렬**



영화/애니메이션



음악



반려동물/동물



여행/이벤트

조운영의 관심사 카테고리 페르소나



자동차/교통



음악



게임



교육



과학기술

곽민규의 관심사 카테고리 페르소나

02 ARCHITECTURE

Feature (2)

**개별 카테고리에 대해, 사용자가 원하는 키워드와 설명을 세부 카테고리에 추가하면,
그것과 유사도가 높은 순으로 영상을 보여줌 (구독한 채널의 영상 분류 기능)**



영화/애니메이션

귀여운 캐릭터가 나오는 애니메이션, 로맨스 코미디 ...



음악

일할 때 듣기 좋은 노동요, NewJeans, 마라탕후루챌린지 ...



반려동물/동물

매력 넘치는 고양이, 에버랜드에서 뒹굴거리는 푸바오 ...

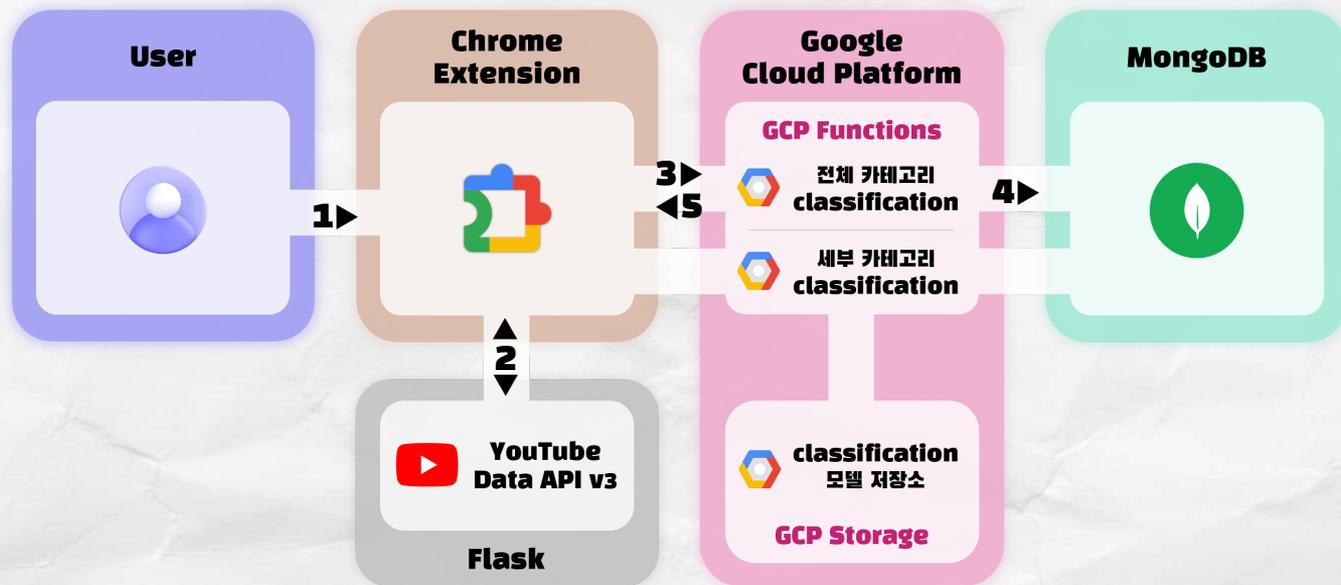


여행/이벤트

경치 좋은 노르웨이 피오르드 트래킹, 빠니보틀 국내여행 ...

02 ARCHITECTURE

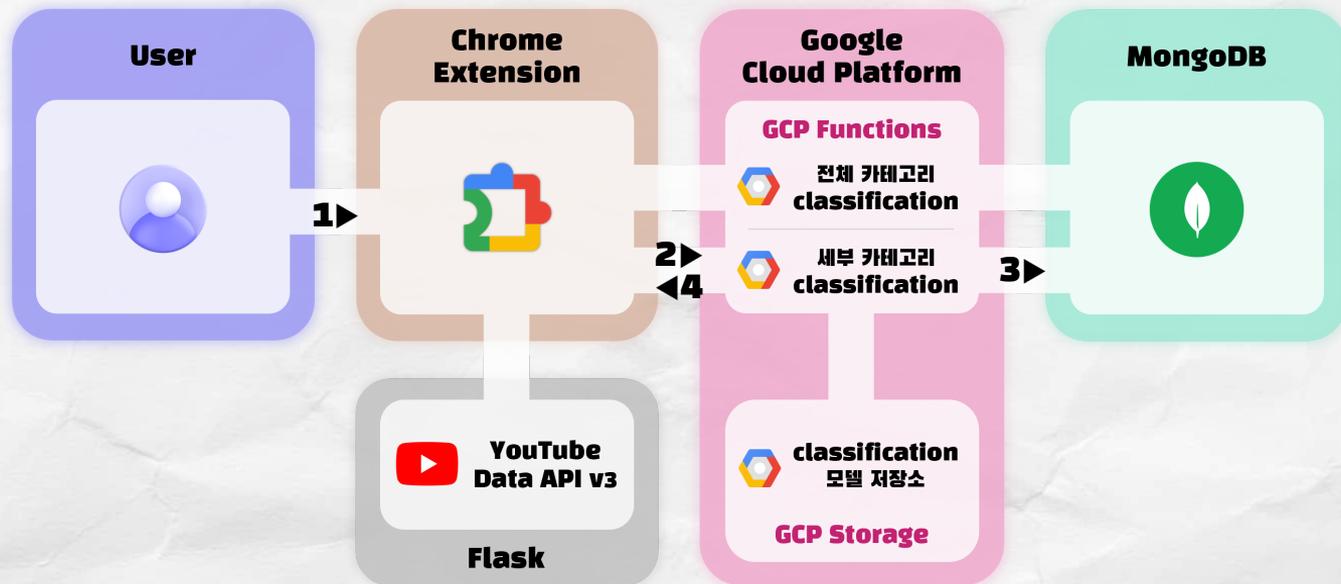
Task (1) : 전체 카테고리 업데이트 및 분류



- (1) 사용자 로그인 (2) 사용자 구독 및 영상 정보 가져오기 (3) 전체 카테고리 분류
(4) 분류 데이터 DB에 저장 (5) 확장 프로그램으로 분류 데이터 전송

02 ARCHITECTURE

Task (2) : 세부 카테고리 업데이트 및 분류



- (1) 사용자가 직접 키워드 및 Description 입력 (2) 세부 카테고리 분류
(3) 분류 데이터 DB에 저장 (4) 확장 프로그램으로 분류 데이터 전송



03

MODELING

03 MODELING

활용 API 및 데이터셋



Description 데이터셋

YouTube 영상 7,000개에
대한 텍스트 데이터 크롤링
(스니펫, 제목, 설명, 태그,
카테고리ID, 캡션)



YouTube API v3

이용자의 구독 채널
목록을 가져오기



OpenAI API

업로더가 카테고리를 설정하지
않았을 경우 Default(22번)
카테고리에 대해 GPT-4o 모델로
카테고리 재분류

카테고리 Classification 모델링



**전체 카테고리는 pre-trained된 BERT for Sequence Classifier를
구축한 데이터에 fine-tuning**

세부 카테고리는 사용자 지정이 가능한 Sentence Transformer 사용



04

CHALLENGING TASKS

04 CHALLENGING TASKS

(1) 느린 inferencing 속도



**summarizer model의 영상
하나 당 inference time이
7초 이상 소요**

전반적인 아키텍처 조정이 필요



**accuracy와 time의 trade-off
관계에서, input 전처리를
발전시켜 raw text를 최대한
의미 있는 문장으로 남긴 뒤,
classifier max length 조절**

04 CHALLENGING TASKS

(2) 세부 카테고리 classification task 문제



처음에는 **BERT classifier**를
이용해 **deterministic**하게
developer가 지정

이는 프로젝트 목적과도 부합하지
않고, 훈련 데이터 직접 라벨링 필요



sentence transformer를
통해 얻는 **latent vector**에
대한 유사도로 사용자가 설정
가능한 알고리즘 고안

실험 결과, 높은 **accuracy** 성능

04 CHALLENGING TASKS

(3) 부족한 YouTube API 일일 할당량



구독 채널 목록의 최근 영상을 불러오는 단계에서 API quota가 100단위로 소모돼 일 할당량 초과

할당량 소모하지 않고 불러오는 방법을 모색해야 함



Oauth2Client 및 구독 목록 불러오는 단계까지는 API 호출

이후 단계는 유튜브 채널 페이지의 RSS 피드를 feedparser로 불러오므로써 해결

04 CHALLENGING TASKS

(4) 처음 시도해보는 확장 프로그램 개발



ybigta 활동에서 한 번도 경험해보지 못한 새로운 형태의 개발 및 산출물

익스텐션 구조를 이해하고 활용하며 설계하는 데 오랜 시간이 소요

지켜야 하는 정책이 많아 까다로움



05

DEMONST- RATIION



06

CONCLUSION



의의 & 시사점

**Windows/MacOS 상관 없이 Chromium 기반 웹 브라우저를 통해
간단하게 설치 및 실행 가능**

serverless 환경에서 모델 배포 경험

**Inference time이 빠른 BERT-base와 sentence-transformer만의
조합으로 사용자가 원하는 유튜브 영상 카테고리 지정**



한계점 & 개선 방안

정책 위반 사유로 Chrome Web Store 등록 거부(빌드 파일 배포는 가능)

**구독 채널마다의 최신 업로드 순 영상으로 카테고리라이징하였으므로
연관도, 조회수, 인기도 등을 반영하지 못함(다양한 요소를 모델에 반영하여 극복 시도)**

시간적 제약으로 인해 현재 세부 카테고리 모델과 확장 프로그램 연결 마무리 절차 중

06 CONCLUSION

24.04.14

Google 계정(페르소나) 생성
가용 API 조사 및 테스트
아키텍처 초안 작성
더미 프로토타입 제작

역할배분, 정기회의 일정 확정
유사 서비스 등 선행연구
문제 재정의 및 구체화

모델 구조 및
데이터 형태 논의

팀 결성

24.05.14

아키텍처 보완
아키텍처 구체화
확장프로그램 프로토타입 개발
유튜브API OAuth2 인증

주제 구체화 및 세분화
개발 내용 정리 및 공유

데이터셋 구축
모델 논의 및 설계
참고 모델 조사

1차 중간발표

24.06.04

로고(BI) 디자인
유튜브API-확장프로그램 연결
유튜브API-모델 연결
확장프로그램-DB 연결

개발 내용 정리 및 공유
종강 직후 매일 마라톤 회의

확장프로그램 개발 및 배포
모델-DB 연결
테스트 및 트러블슈팅

2차 중간발표

24.06.28

컨퍼런스

06 CONCLUSION



조운영(DE)

팀장, 파이프라인 설계&구현



박경욱(DS)

분류 모델 개발, GCP-DB 연결



김현호(DE)

파이프라인 설계&구현, 프론트



곽민규(DS)

데이터셋 구축, 분류 모델 개발



임종혁(DE)

파이프라인 설계&구현, 디자인



이승준(DS)

모델 분류 아키텍처 설계&개발



김예진(DA)

모델 트레이닝, 분류 모델 개발



유튜빙고(마스코트)

귀여운 플라밍고 '튜브' 캐릭터



Thank you!

 github.com/YBIGTA/24-1_YourTube