

#바글바글

Chat GPT를 이용한 댓글 순화 서비스

Comment purification service using Chat GPT





100%



Agenda

컨텐츠와 댓글

- 매일 생성되는 컨텐츠의 양
 - 댓글의 양면성
- 각 플랫폼별 댓글 검열

기대효과 및 개선방향

- 이용자별 기대효과
- Chat GPT 예시

Meet the Team

Contents
Comments
&
Core

Target Users
&
Approaches

Dev
Schedule

Expected
Outcomes



이용자별 맞춤 접근

- 댓글 쓰는 사람 측면
- 댓글 보는 사람 측면



Meet the Team



조수환
프로젝트 리더
기능구현



안성찬
자료조사
기능구현



김종한
DB
서버



정우성
서버
파이프라인

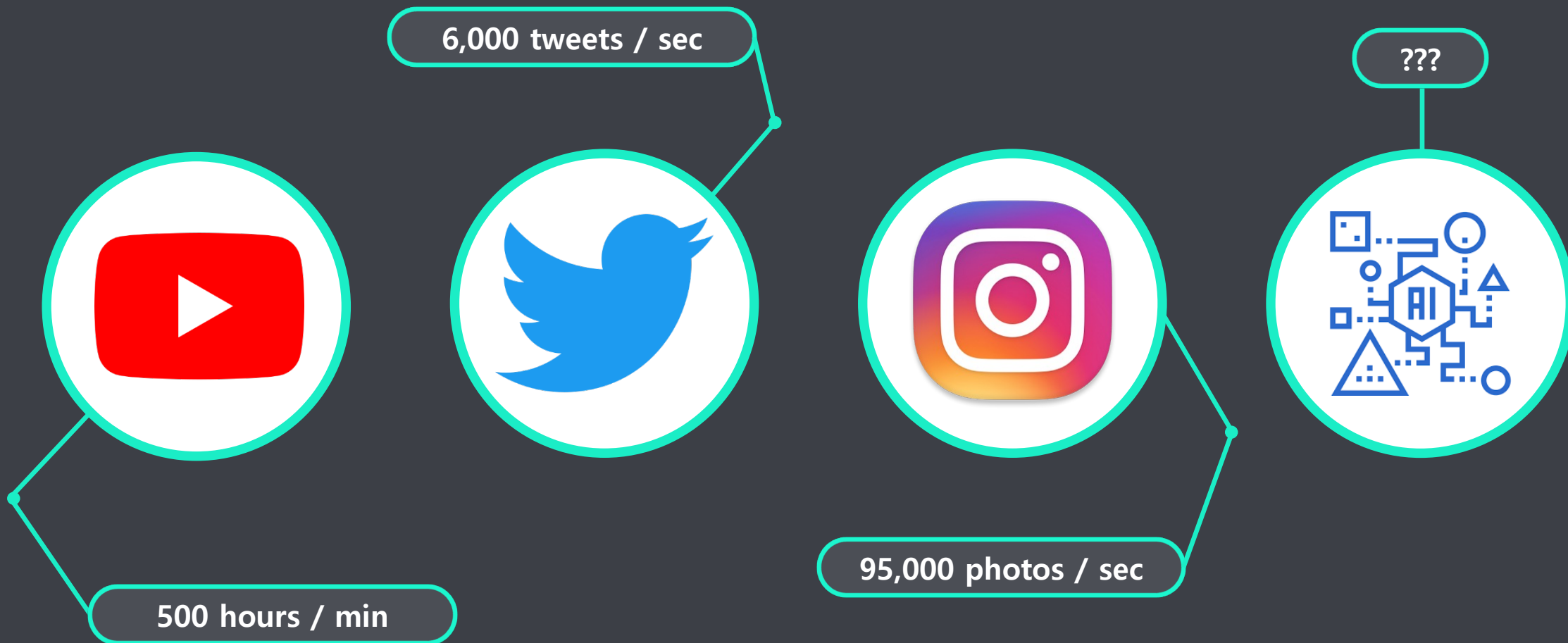


김만서
발표
웹 디자인

Contents



Contents



Comments

명동 한복판 '삼겹살 파티' 들썩

권오용기자 | 사진제공=SK컴즈



입력 : 2009.12.27 18:57 | 수정 : 2009.12.27 19:16



- 댓글** 김성근 **추천** 2660
제가베를이된다면 크리스마스에 명동 한복판에서 혼자 삼겹살을 꾸워 먹겠습니다. (12.04 14:53) 8
댓글의 댓글 453개 ▼
- 댓글** 조영석 **추천** 2205
제가베를이된다면 김성근씨가 삼겹살을 꾸워먹을때 전원에서 노래를 불러주며 흥을 도파워 드리겠습니다. (2.04 14:57) 8
댓글의 댓글 195개 ▼
- 댓글** 김명관 **추천** 2149 **반대** 1
제가베를이된다면 김성근씨가 삼겹살을 꾸워먹고 조영석씨는 노래를 불러주며 흥을 도파워 전 앞에서 린 치겠습니다. (12.04 15:06) 8
댓글의 댓글 204개 ▼



'연예인들 자살 사건에 악플이 영향 미쳤다', 98%

한국언론진흥재단, "댓글 폐지, 실검 폐지에 대한 국민 인식" 발간

"다음 연예뉴스 댓글 폐지...지지한다", 80.8%

실시간 검색어 폐지, '지지한다' 46.7%, '반대한다' 26.8%, '관심 없다' 26.5%

기사입력 : 2019-12-17 17:21:14

좋아요 17개

최근 연예계에는 안타까운 사건이 연이어 발생했다. 젊은 연예인 두 명이 한 달 간격으로 극단적 선택을 했고 그들은 오랜 기간 악성 댓글(이하, 악플)에 시달린 것으로 알려졌다. 가수 겸 배우 설리의 자살 사건 직후, 인터넷포털 다음이 연예뉴스에 대한 댓글 폐지를 전격적으로 단행했다. 이후 다른 인터넷포털들도 동참해야 한다는 목소리가 높았다. 한편, 악플과 악성 댓글이 연예계에 미치는 영향에 대해...

바글바글?

바르게 글쓰고
바르게 글읽기



Comments



Comments

악플이란 무엇인가?

- 악성댓글은 타인을 악의적으로 비하할 목적으로 다는 댓글을 말한다.

모욕죄란 무엇인가?

- 모욕죄: 정보통신망이용촉진및정보보호등에관한법률위반(명예훼손)죄
- 모욕성, 공연성, 특정성을 전부 확보 되는지부터 짚고 넘어가야만 접수가 가능하다.

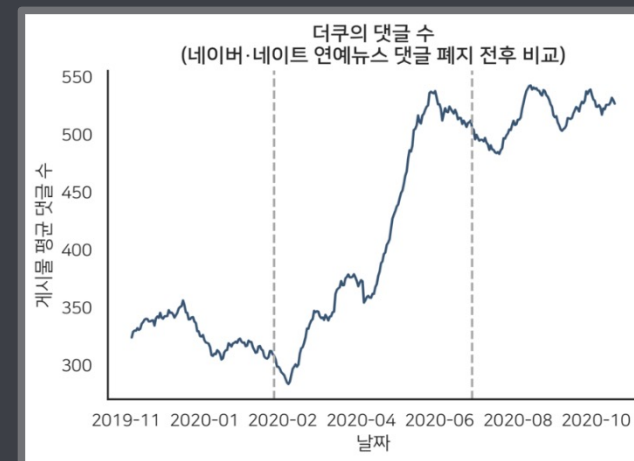
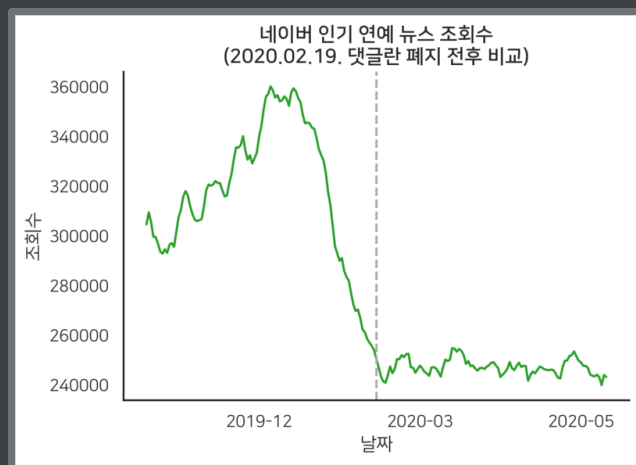
표현의 자유란 무엇인가?

- 표현의 자유는 기본적으로 말 할 권리에 대한 논의

무엇이 악플이고,
어떻게 표현의 자유를 지키고,
얼마나 필터링 해야하는가?

Comments

2020.02 네이버 연예 뉴스 댓글란 폐지 전후 비교



‘무차별적 검열과 삭제는 문제 해결에 도움이 되지 않는다’

Core

“Chat GPT를 이용한 댓글 순화”

- 콘텐츠 생산자와 소비자를 위한 악성 댓글 순화
- 댓글의 여론 평가

Core

왜 'Chat GPT' 로 접근해야 하는가?

- 인간 피드백 기반 강화 학습을 통해 발전했기에 댓글의 의미를 잘 파악한다
- 댓글들을 순화해주는 기술은 존재하나 해당 기능을 사용하는 댓글 순화 서비스는 없다
- 댓글 순화 서비스, 말투 변환, 여론 파악, 콘텐츠 소비자 감정 분석 및 예측 등 '바글바글' 팀만의 독창적 서비스를 만들 것

IBM Research(2018). Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer. // text style transfer

Mst Shapna Akter, Hossain Shahriar, Nova Ahmed, Alfredo Cuzzocrea & Deep Learning(2023). Approach for Classifying the Aggressive Comments on Social Media: Machine Translated Data Vs Real Life Data // gpt-2



Target Users

- 미디어 콘텐츠를 제작 / 소비하는 모든 사람

콘텐츠 제작자

댓글을 쓰는 사람
댓글을 보는 사람



Approaches



댓글을 보는 사람

- 일반 사용자

- 악성댓글을 순화시켜 재작성
- 광고성 댓글 차단

- + 콘텐츠 제작자

- 댓글을 기반으로 콘텐츠에 대한 전체적인 피드백 제공

댓글을 쓰는 사람

- 의도된 악성 댓글

- 댓글 업로드 후 검열

- 의도되지 않은 악성 댓글

- 댓글에 자신의 의도가 잘 반영되었는지 확인





100%



Schedule

1

문제 정의 및 아이디어 도출

많은 아이디어를 도출한다.
브레인 스토밍을 열어 사용자
의 Pain Point를 찾아내고 이
를 위한 해결책을 도출하는
과정을 거친다.

3.10 ~ 3.17

2

아이디어 검증

제시했던 여러 아이디어중 가
장 성공 가능성이 높은 아이
디어를 중점으로 선택할 예정
이다.

기술적으로는 가능한 아이디
어인지, SWOT분석을 통해,
아이디어 검증.

3.17~3.24

3

아이디어 정의

우리 제품이 시장에 있는 다
른 제품과 어떤 점이 다른지
생각해본다.

경쟁 서비스 분석 및 파악

네이버나, 유튜브 등의 기업에
서 현재 악성댓글을 검열을
위한 서비스를 살펴본다.

3.24~3.31



100%



Schedule

4

MID-TERM PRESENTATION

초기 단계에 필요한 최소한의
기능 구현을 완료한다. (MVP)

MVP란 출시에 필요한 기능은
갖췄지만, 제품이 기능하는 데
필요한 이상의 기능은 갖고
있지 않은 제품을 말한다.

3.31~4.21

5

기능 검증 및 개 선

개발한 서비스, 기능의 품질을
보증할 수 있도록 팀원들이나
테스터에게 기능을 테스트 받
는다.

4.21~5.05

6

기능 추가 및 개 선

최소한의 기능뿐만 아니라, 필
요한 기능을 추가하고 개선하
는 과정을 거친다.

5.05~5.19

7

기능 평가

사용자에게 테스트를 하여 피
드백 받는 과정을 거친다.

전문가에게 검토 받는 과정을
거친다.

5.19~6.02

8

FINAL PRESENTATION

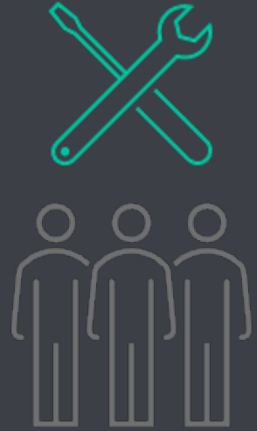
개발 완료한 기능을 모두 포
함한 최종 서비스 출시

6.02~6.16





Expected Outcomes



컨텐츠 제작자 측면

1. 악성 댓글 사용률 감소
2. 컨텐츠 이용률 증가

컨텐츠 이용자 측면

1. 악성 댓글에 의한 피해 최소화
2. 댓글의 순기능 극대화





Expected Outcomes (example)

우성

왜 그렇게 살아? 라는 말을 유하게 바꾸어줘



"왜 그렇게 살아?"라는 말 대신에, "당신이 겪은 어려운 일들은 무엇인가요?" 또는 "어떤 어려움에 직면하고 있나요?"와 같은 관심과 배려를 담은 더 적절한 질문을 사용하는 것이 좋습니다. 이러한 질문은 상대방의 문제나 고통을 이해하고 공감하는 것을 보여줄 뿐만 아니라, 상대방과 함께 대화하며 문제를 해결할 수 있는 가능성을 열어줍니다.  

#바글바글

Thank you

Comment purification service using Chat GPT

