



선플수호자

인터넷 커뮤니티 악플 수집 및
악플 내려받기 서비스



소리없는 흥기,
악플

악플이란?

인터넷상에서 다른 사람이 올린 글
또는 콘텐츠에 대해 악의적인 댓글을
달거나 험담을 하는 것

협오표현 + 악플

= 소수자 집단 공격

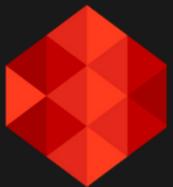
문제 악화

인터넷 댓글 :
효과적 감시 無
영향력 高



악플로 고생하는
사람들을 위한
해결책은 없을까?

월간 방문자 수 천만 이상 커뮤니티

NATE 

dcinside

PPOMPPU



에펨코리아
DIGITAL GALLERY



국내 1위 중고차쇼핑몰

보배드림

the qoo theqoo.net



Humoruniv.com
웃기다대학


MLBPARK

INVEN



'댓글' 데이터 분석

1. 구어체의 텍스트 데이터 상당수

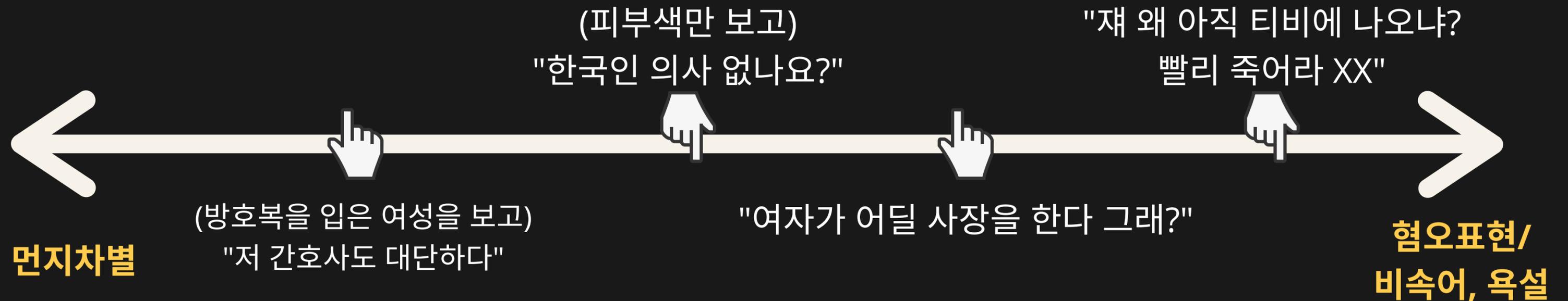
예시: "우왁ㅋㅋㅋㅋㅋㅋ천재아니냐?"

+ 단어 변형 맞춤

예시: 유대종 -> 유머벨

'댓글' 데이터 분석

2. '악플'의 기준은 필연적으로 주관적



악플 분석,
사람이 직접?



**딤러닝,
감정 소모 NO!
시간 소요 NO!**

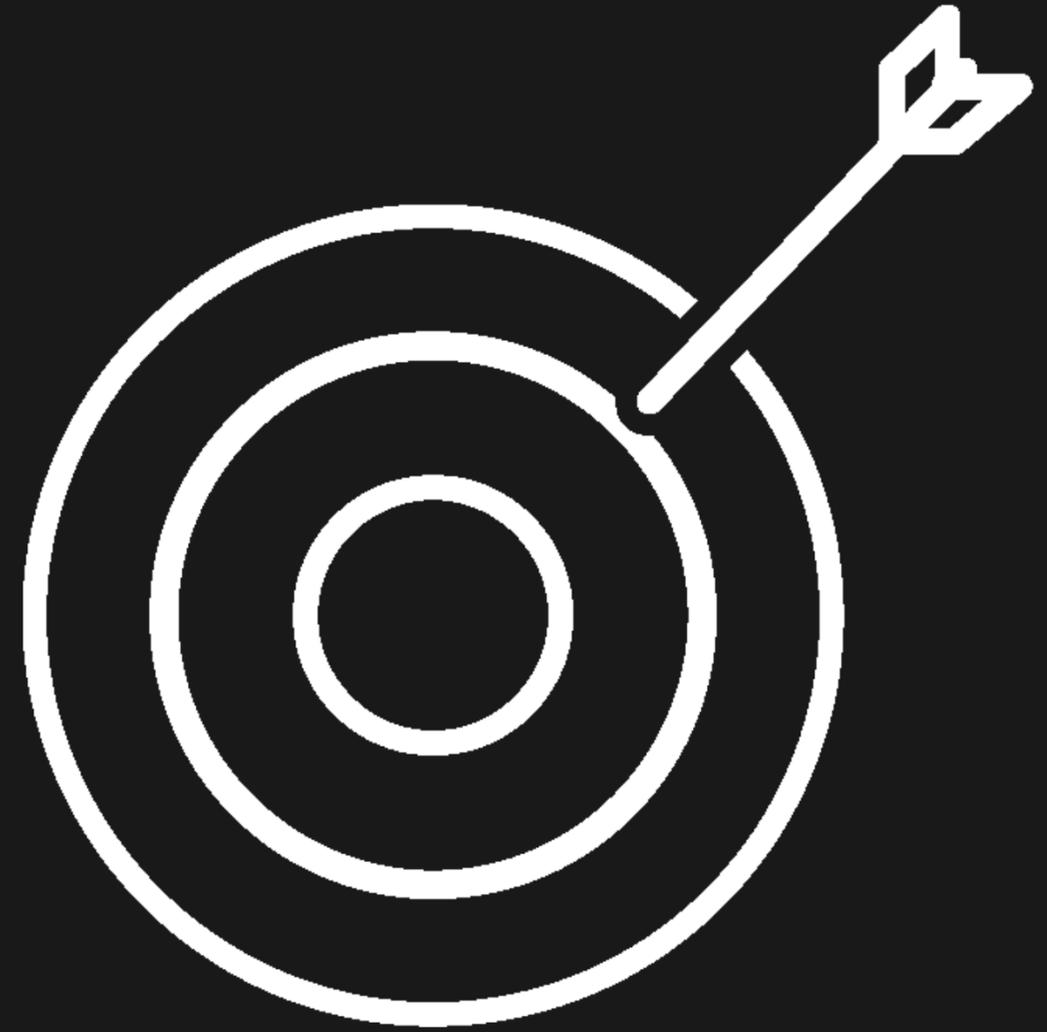
딥러닝,

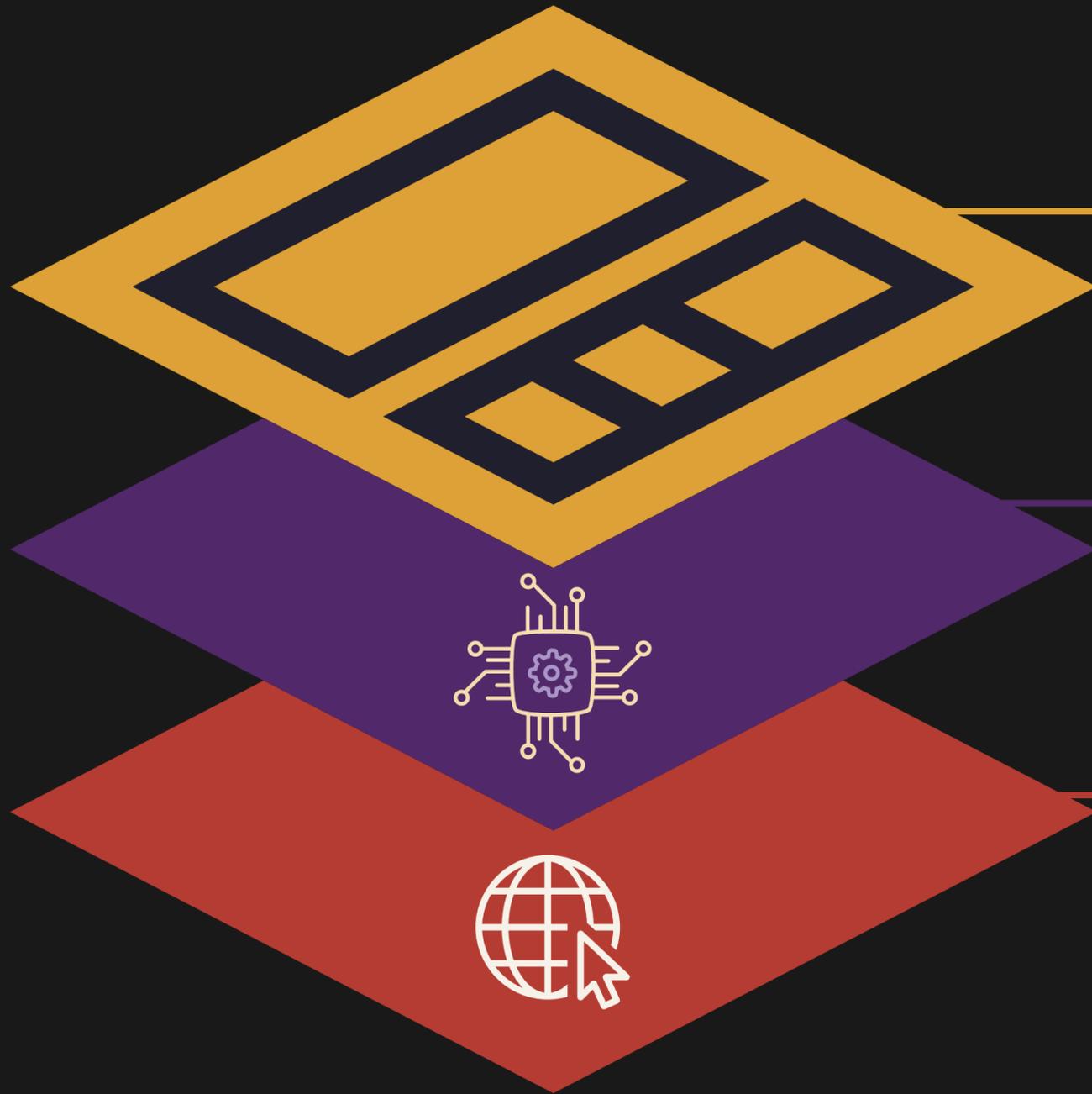
1. **문맥** 반영 가능

2. 유사도에 대해

확률적 접근 가능

악플
자동
수집기





프론트엔드

Progressive Web App 의 형태로, 누구나 PC/모바일 등에서 쉽게 접속해 악플을 검색하고 다운로드 받을 수 있게 제작

악플 예측 모델

Korean Hate Speech Dataset 을 기반으로 훈련된 DistilKoBERT 모델이 악플일 확률을 예측

크롤링 백엔드

국내 인터넷 커뮤니티(일베, 인벤)의 인기 게시글을 대상으로 댓글을 수집해 내부 DB에 저장

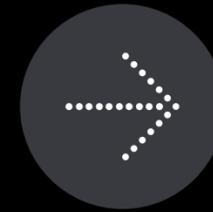
1. 크롤링 백엔드



인터넷 커뮤니티
인기 게시물 자동 수집



수집한 게시물에 올라온
댓글 정보를 수집



게시글과 댓글 수집 결과를
DB에 저장

분석 대상 :
인터넷 커뮤니티 댓글

500,000

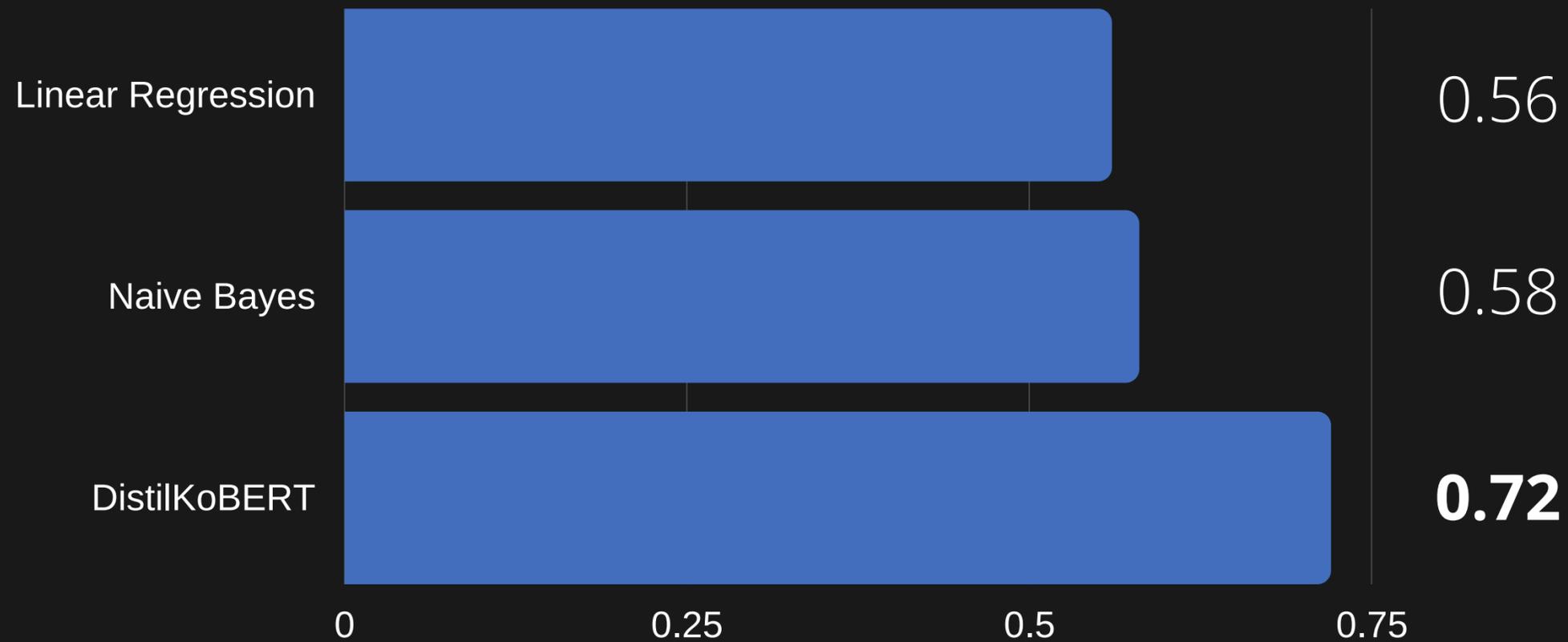
오십만!

2. 학습데이터: Korean HateSpeech Dataset

```
comments          contain_gender_bias    bias    hate
송중기 시대극은 믿고본다. 첫회 신선하고 좋았다.    False  none    none
지현우 나쁜놈          False    none    offensive
알바쓰고많이만들면되지 돈욕심없으면골목식당왜나온겨 기댁기게나하고 산에가서팔어라    False  none    hate
설마 ㄹ 현정 작가 아니지??    True    gender  hate
```

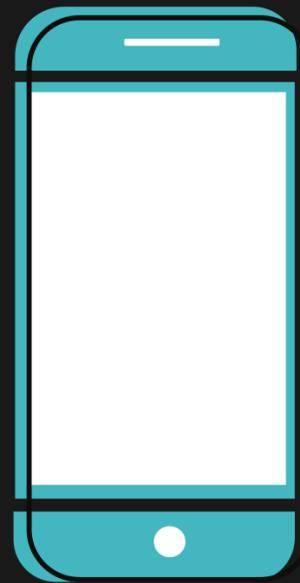
2. 학습 예측 모델

DistilKoBERT

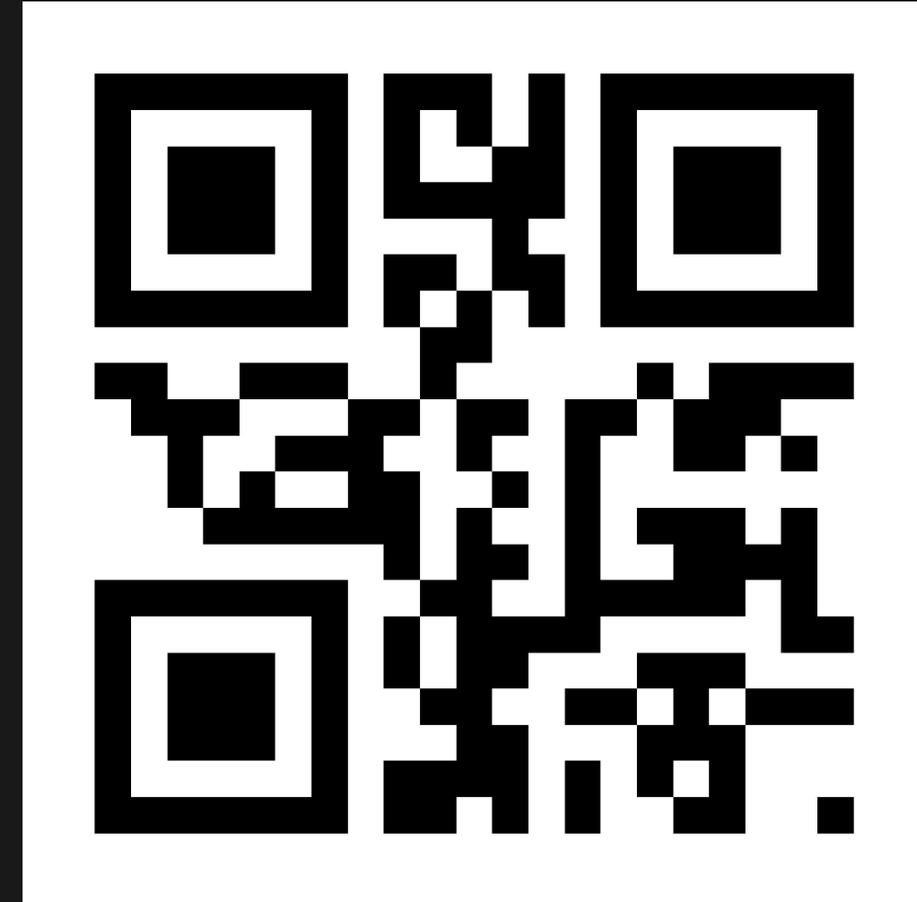


Korean Hate Speech Dataset 검증 셋에 대한 정확도 비교

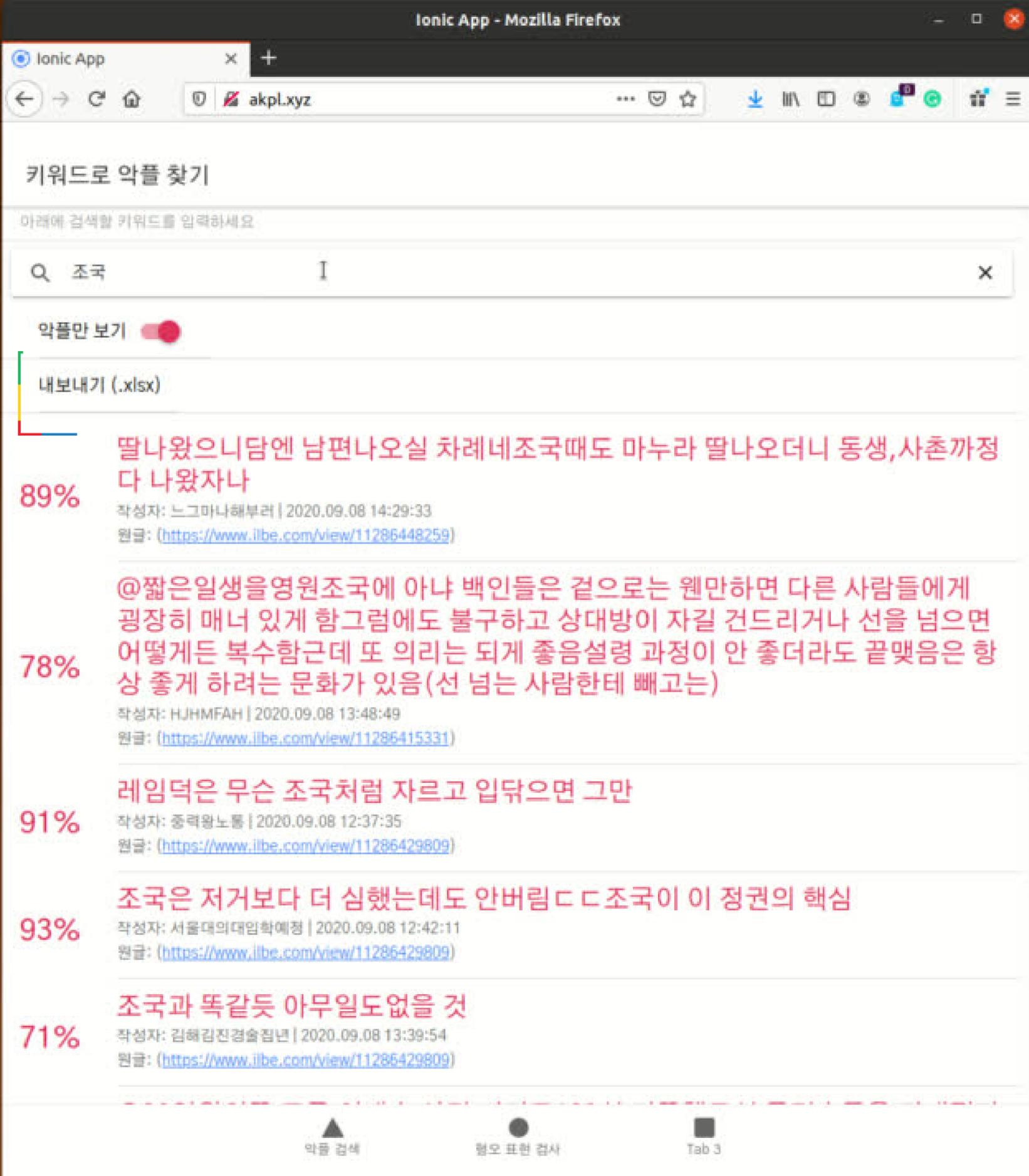
3. 프론트엔드 : Progressive Web App



스마트폰, PC, 태블릿
어디에서든,
akpl.xyz
접속하실 수 있습니다 :)



**악플자동수집기,
지금 보러갑시다~**



**악플, 이제
클릭 한번으로
다운로드!**

사회적 활용

1. 악플 피해자의 든든한 지원군

2. 각 커뮤니티 별

전체 댓글 중 악플 비율 수치화 가능

학술적 활용

1. 학술연구의 도구

예시) COVID-19 전후로

특정 이민자 집단에 대한 인식 변화 조사

2. 커뮤니티 추가, 예측모델 개선 :

무궁무진한 활용 가능!

악플자동수집기

BY

선플수호자