

KISO 욕설 필터링 서비스 출시 1년, 37개사 이용 “댓글창 깨끗해져”

KISO 욕설 방지 시스템 통해 욕설·비속어 33만 건 걸러내
커뮤니티·금융기관·언론사 등 이용업체 37곳으로 확대
지속적 DB 업데이트...‘60만개→80만개’ 신종 욕설 대응

“KSS 도입 이후 커뮤니티 내 불쾌한 표현이 크게 줄었다.”

“뉴스 댓글창이 상당히 깨끗해져... 사업자·이용자 모두 만족도 높다”

한국인터넷자율정책기구(KISO, 의장 이인호)가 건강한 인터넷 문화 조성과 이용자 보호를 위해 지난해 6월 도입한 욕설 필터링 서비스인 이용자보호시스템(KISO Safeguard System, 이하 KSS)이 출시 1년 만에 큰 성과를 거둔 것으로 나타났다. KSS를 이용하는 기업과 기관이 37곳으로 늘어나는 한편, 이를 통한 욕설 탐지 및 보호조치 건수도 1788만 건의 게시글 중 33만 건에 달해 그 효과를 입증했다.

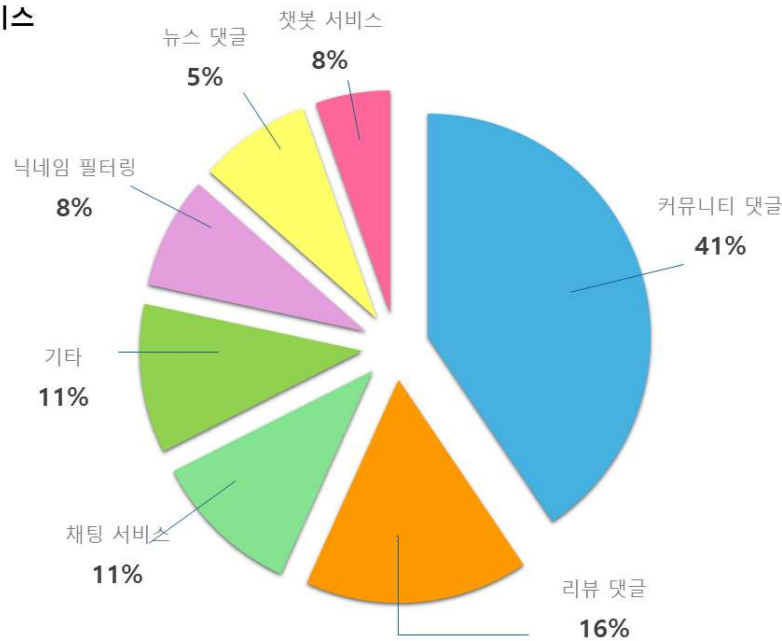
네이버와 카카오가 무상으로 제공한 욕설 데이터베이스(DB) 60만 건을 통합해 KISO가 개발한 KSS는 다양한 플랫폼 환경에서 이용자가 작성한 게시글, 댓글, 아이디 등에 욕설·비속어가 포함된 경우 이를 탐지하고 걸러내 주는 서비스이다.

KISO는 건강한 인터넷문화 조성에 기여한다는 방침에 따라 공공기관과 언론사, KISO 회원사 등에는 KSS를 무료로 제공하고, 일반 기업에도 매우 저렴한 비용으로 이용할 수 있도록 지원하고 있다.

출시 1년 만에 KSS는 온라인 커뮤니티, 인공지능(AI)·메타버스 플랫폼, 금융기관, 교육 서비스 업체, 쇼핑몰, 언론사, 공공기관 등 37곳에서 도입되어, 다양한 산업 분야의 기업들이 이를 적극적으로 활용하고 있다.

대표적으로 EBS에서 만든 교육용 메타버스 ‘위캔버스(WecanVerse)’의 게시글 및 채팅 서비스에서 욕설·비속어를 탐지하고 있으며, ‘TMAP(티맵) 앱 내 장소 리뷰 작성 시 텍스트 필터링, 중고 거래 플랫폼 ‘번개장터’ 내 고객 게시판, 시청자가 생방송 등에 실시간으로 참여하는 KBS ‘티벗(TVut)’ 채팅 서비스에도 욕설·비속어 탐지 및 필터링에 KSS가 사용되고 있다. 교육 서비스 ‘말해보카’와 ‘개념원리’에도 도입되어 있으며 국립중앙도서관, 화성시, 충청북도교육청도 KSS를 도입했다.

KSS 적용 서비스



KSS 이용에 대한 만족도도 높았다. KISO 회원사인 뽀뿌커뮤니케이션 관계자는 “KSS 도입 이후 사용자 간의 소통에서 발생할 수 있는 불쾌한 표현을 상당 부분 줄

일 수 있었다”고 말했다.

동아닷컴도 “KSS 적용 후 욕설·비속어 필터링 기능으로 인해 관련 서비스 댓글창이 상당히 깨끗해졌음을 느낀다”며 “전반적으로 프로그램 사용 만족도가 높다”고 밝혔다.

KSS는 네이버와 카카오가 악성 댓글 등으로부터 이용자를 보호하기 위해 투자 및 개발해 온 자동화된 필터링 기술의 성과물을 공유받아 공개한 것이다. 네이버와 카카오는 이미 자체 시스템을 통해 상당한 성과를 거두고 있다. 네이버는 2019년 4월 인공지능(AI) 기술 기반의 악성댓글 탐지 시스템인 ‘클린봇’을 도입한 이후 악플 노출 비율이 18.5%(2020년)에서 8.9%(2023년)로 절반 이하로 떨어졌으며, 카카오는 2020년 하반기 ‘세이프봇’ 도입 이후 욕설·비속어가 2023년 기준 94.7%나 줄었다.

KISO는 올해 4월 네이버와 카카오로부터 추가적인 DB를 제공받아 기존의 사용빈도가 낮은 단어 및 중복된 내용을 삭제하고 새롭게 추가된 신종 욕설·비속어 20만 건을 추가했다. 이번 업데이트 작업을 통해 현재 약 80만 건의 욕설·비속어 DB를 보유하게 됐다.

KSS는 다양한 특수문자와 숫자가 포함된 변형 형태의 욕설·비속어를 탐지할 수 있었으나, 변형어가 다양해 지면서 모두 걸러내는 데는 한계가 있었다. 그러나 이번 업데이트를 통해 한자와 영어 결합 형태, 단어 중간 삽입형, 완성형 문장 등 새로운 유형의 욕설도 탐지할 수 있게 되어, 기능이 보다 효과적으로 개선되었다. KISO는 앞으로 DB의 분류 기준을 세분화하는 등 다양한 기능을 개발하여, 다양한 회원사의 요구에 맞춰 고도화를 지속할 계획이다.

황용석 KISO 자율규제DB위원회 위원장(건국대 미디어커뮤니케이션학과 교수)은 “인터넷은 시간과 공간적 한계를 뛰어넘어 누구와도 소통할 수 있는 공론장이지만 타인에게 상처를 주거나 불쾌감을 주는 표현으로 인한 부작용도 발생하고 있다”며 “KSS는 자율규제의 중요한 수단인 자동화된 기술을 제공해서 보다 효과적으로 이용자들을 욕설·비속어로부터 보호한다는 점에서 큰 의미가 있다. 이를 통해 즉각적인 이용자 보호조치는 물론 인터넷이 건강한 공론장으로 기능하는데 일조할 것으로 기대된다”고 개발 배경을 설명했다.

이어 그는 “향후 KSS를 지속적으로 고도화해 나가고, 다양한 서비스 내 이용자

를 보호하기 위해 더욱 노력할 것”이라고 덧붙였다.

KISO는 이용자의 표현의 자유 신장과 책임 의식을 고취하고, 포털 사업자의 사회적 책무 이행을 목적으로 국내 포털 사업자들이 2009년 설립한 국내의 대표적인 자율규제 기구로 현재 네이버, 카카오, SK커뮤니케이션즈 등 16개 회원사가 활동하고 있다.

[붙임] KSS 인증로고

