



팀 명 CIDS

팀 원 김정우 유상정 육지영 이태준

지도교수 손태석

멘 토

개발 동기 및 목적

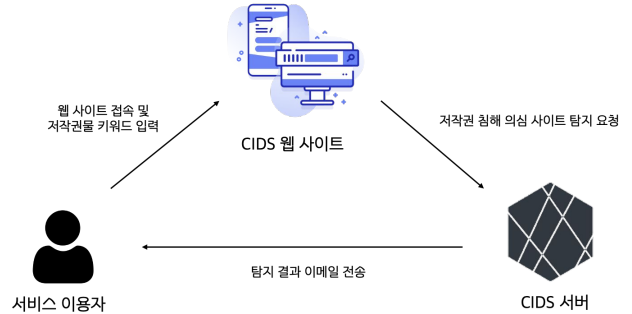


최근 영화, 드라마, 웹툰 등 콘텐츠에 대한 수요 및 사용 빈도가 증가하고 있다. 하지만 수요가 증가함에 따라 콘텐츠를 저작권자의 허가 없이 불법 복제하여 유통하는 저작권 침해 사이트가 늘어나고 있다. 저작권 침해 사이트로 인해 저작권자는 수익 창출에 어려움을 겪고 있으며, 저작권자의 정신적, 금전적 피해가 확대되고 있다. 하지만 저작권 침해 사이트의 탐지가 쉽지 않은 상황이며, 탐지를 위한 서비스가 보편화 되어있지 않다. 이에 따라 저작권 불법 침해 사이트의 빠른 탐지가 가능한 보편화 서비스를 개발하였다.

주요 기술

1. Crawling 기법
 - 요청받은 키워드가 포함된 URL들을 크롤링하기 위한 기술
 - 수집한 URL 내의 불법 광고 배너를 추출하기 위한 기술
2. OCR 기술
 - 불법 광고 배너의 WORD LIST를 추출하기 위한 기술
3. RNN의 LSTM을 활용한 WORD LIST의 학습 및 판단
4. 사용자 접근이 용이한 웹서비스 구축
5. 탐지 결과의 이메일 알림 서비스

개발 내용



저작권 침해 사이트는 수익 창출을 위해 불법 광고 배너를 게시 하며, 불법 광고 배너 안에는 저작권 침해 의심 사이트를 식별할 수 있는 키워드가 존재한다. 따라서 CIDS는 광고 배너 이미지 URL을 추출함으로써 이미지 내부 키워드를 학습하여 저작권 침해 의심 사이트를 탐지하여 결과를 사용자 이메일로 전달한다.

이를 위해 사용자가 서비스를 이용하기 위한 웹 사이트, 저작물 키워드를 바탕으로한 데이터 수집 및 광고 배너 텍스트 추출을 담당할 파이썬 모듈, 추출한 텍스트를 기반으로 저작권 침해 의심 사이트 판단을 진행할 딥러닝 모델이 개발되었다.

결과 및 분석

컨텐츠 저작물에 대한 기존 저작권 침해 탐지 제품들은 사용자 접근이 용이한 시스템과, 키워드만을 가지고 저작권 침해 의심 사이트를 탐지해주는 경우는 존재하지 않는다. 따라서, CIDS는 경쟁사 제품에 대비해 웹툰 혹은 영상만이 아닌 포괄적인 컨텐츠 범위에서 사용할 수 있다. 이러한 이유로 타 제품에 비해서 경쟁성을 충분히 가지고 있다고 생각한다.

본 프로젝트에서 개발하고자 하는 저작권 침해 사이트 탐지 프로그램은 추후 저작권자 본인의 저작물에 대한 금전적 피해 확산을 방지하는데 사용될 수 있다고 예상된다. 또한 저작물 키워드 검색을 통해 탐지한 저작권 침해 사이트 신고를 사용자에게 유도하여 매년 증가하고 있는 저작권 침해 이슈를 줄일 수 있을 것으로 기대된다.