

카피킬러라이트 표절 검사 결 과 확 인 서

문서표절률

8%

확 인 서 정 보

검사번호 : 00200088174

발급일자 : 2022.10.14 00:32

아이디 : dorisjb0827@gmail.com

닉네임 : 서진배

검사일자 : 2022.10.14 00:24

검사문서 : 카피킬러 올리기 용.docx

본 확인서는 dorisjb0827@gmail.com 사용자가 카피킬러에서 표절검사를 수행한 표절분석 결과에 대한 문서로 카피킬러 표절검사 시스템이 자동으로 생성한 자료입니다. 문서 작성 기준이 각 학교, 기관마다 다르기 때문에 최종 평가자의 표절평가 결과와는 다를 수 있습니다.

표절 검사 상세 결과

문서표절률	전체문장	동일문장	의심문장	인용/출처	범형/경전
8%	33	0	4	0	0

표절 결과 문서명

카피킬러 올리기 용.docx

1. 주제개인방송인과 시청자들을 위한 욕설 및 비속어 필터링 시스템 제안분반, 팀, 학번, 이름나반 1팀 20221796 서진배

2. 요약방송상 부적절한 언어를 사용하면 그 순간을 자동으로 인식해 필터링 해주는 시스템을 만들고자 한다.이로서 시청자들이 불쾌함을 느끼지 않도록 해주고 방송인들이 더 건전한 방송문화를 **만들 수 있도록 도움을 주고자 한다**. 녹화된 영상을 자동으로 편집해 주는 만큼 실시간 영상을 제공하고, 녹화된 영상 역시 제공해주는 유튜브나 트위치 등과 함께이 시스템을 이용한다면보다 **더 높은 효과를 기대할 수 있을 것이다**. 3. 대표 그림그림 1. 욕설 및 비속어 필터링 시스템

4. 서론위 기사는 유튜브 구독자 216만명의 방송인 감스트가 경기를 앞두고 부상당한 손흥민 선수에게 무의식적으로 욕설을 해 논란이 된 사건을 다룬 기사입니다.유튜브와 SNS등의 매체가 발달하는 만큼 대중들은 보다 더 쉽게 많은 영상을 접할수 있게 되었습니다.그만큼 많은 미디어들이 매일 업로드 되고, 이를 이용해 금전적인 이익을 얻고자 하는 사람들도 늘어나고 있습니다.그렇게 늘어나는 영상 속에서 더 높은 조회수를 얻기 위해 자극적인 내용을 통해 사람들로 하여금 영상을 시청하도록 하는 방송인들이 나타나고 있습니다.하지만 평소 공인으로서 올바르게 방송을 하던 사람들도 실수로 시청자들이 불쾌함을 느낄만한 말을 할 때가 있습니다.단 한번의 실수 임에도 그 사람은 많은 비난과 비판을 받아야 했고 이러한 실수로 인한 사례가 늘어남에 따라 많은 방송인들이정신적인 피해를 받아왔습니다.그래서 저는 많은 방송인들이 한번의 실수로 인해 비난받지 않도록 도와주고자자극적인 말을 듣고 싶지 않은 시청자들을 위한 시스템을 구상하게 되었습니다.욕설을 인식해 자동으로 필터링 해주고 인터넷에 새로 만들어지는 비속어들에 대응할수 있도록 부적절한 단어들을 DB에 저장, 업데이트 후 오디오의 자세한 형태소 분석을 통해 필터링을 진행할 수 있게 만들 예정입니다.

5. 본론본 시스템은 기본적으로 웹을 통해 동작할 예정입니다. 웹을 통해 영상을 업로드하면 웹 서버에서 영상의 음성에서 데이터베이스에 저장된 욕설이나 민감한 발언의 단어와 일치하는지 분석해 영상을 편집해주는 작업을 진행할 예정입니다.욕설을 인식하고 처리하기 위해서는 음성을 텍스트로 바꿔야 합니다. 신경망 모델을 적용하는 Google의 **Speech-to-Text를 사용하면 개발자가 오디오를 텍스트로 변환할 수 있습니다**.120개국 언어 뿐만 아니라 방언까지 인식하여 다양한 사용자층이 모두 사용할 수 있습니다.따라서 **음성명령 및 제어 기능을 구현하고 콜센터의 오디오를 텍스트로 변환하는 등의 작업을 할 수 있고 Google의 머신러닝 기술을 사용하여 실시간 스트리밍 또는 사전 녹음 오디오를 처리할 수 있습니다**. 텍스트로 변환된 욕설을 처리하기 위해서는 문장에서 단어들을 분리해서 처리해야 합니다.KoNLPy 패키지를 사용하면 형태소의 뜻과 문맥을 고려하여 원하는 단어에 마크업을 할 수 있습니다.이를 통해 단어를 품사에 따라 구분하고 확인할수 있고 욕설을 필터링 할 수 있습니다.영상을 편집해 욕설을 필터링 해야 합니다. Moviepy 라이브러리를 사용하면 '프리미어 프로'나 '애프터 이펙터'와 같이 영상편집 프로그램과는 달리 python 코드를 통해 손쉽게 영상 편집이 가능합니다.또한 AudioFileclip이라는 영상에서 오디오를 추출하는 기능까지 제공하기 때문에 이 시스템을 개발하기 적합합니다.

6. 결론우리나라는 2020년 기준가구별 tv보유율이 95.8% 인구별 스마트폰 보유율이 91.1%로 대부분의 사람이 미디어 매체를 사용합니다.이렇게 많은 사람들이 미디어를 사용하기 때문에 방송과 관련한 부분이 많이 발달하였는데, 특히 개인 방송에서 금전적인 이익을 위해 자극적인 내용을 포함하는 경우가 늘어났습니다.개인방송은 보고 싶은 사람만보는 것이기 때문에 상관없지만 이러한 경향이 영향을 끼쳤는지 가끔 공중파 방송에서도 욕이 나오는 등의 장면이나 오기도 합니다.저는 이런 장면들을 없애기 위하여 고의던 실수던 방송에서 부적절한 언어를 사용시에 필터링해 주는 시스템을 구상하였습니다.음성을 텍스트로 변환하여 형태소를 분석해 단어의 명사와 어근을 분리하고 욕설 등이 담겨있는 DB에 있는 단어와 비교해 필터링 하도록 구상했습니다.하지만 초기의 DB만으로는 계속해서 나오는 신조어 등을 모두 검출해내는 것에 한계가 보였고 그래서 새로운 데이터를 DB에 업로드 해주는 AI를 사용해야할 해결하고자 합니다.현재 많은 곳에 자동으로 자막을 달아주는 등의 SW는 많이 나와 있습니다. 하지만 자막만 달아줄 뿐 방송에서 부적절한 단어들까지 모두다 포함해 자막을 띄우기 때문에 시청자들에게 불쾌함을 줄 수도 있습니다.본 시스템을 활용한다면 방송인들은 말실수로 인해 큰 논란이 생기는 일이 없을 것이고 시청자들도 표현에 의해 불쾌할 일도 없을 것입니다.이런 시스템을 제작하기 위해서는 Speech-to-Text와 형태소 분석기술, DB구축이 매우 중요합니다.영상의 말을 잘못 인식해서 필터링이 되지 않는 경우를 방지하기 위해 가장 많은 데이터를 보유하고 있는 구글의 STT API를 사용할 예정인데 Python으로 다양한 기능들을 사용하기 위해 API사용법을 익혀야 합니다.또한 KoNLPy를 사

용해 형태소 분석을 진행할 예정인데 복잡한 한글의 문법 구조 때문에 한번에 형태소 분석이 진행되지 않는 경우가 많이 있습니다. 이를 해결하기 위해 여러 상황들을 테스트해보며 기능을 개선하는 작업을 거쳐야 합니다. 이렇게 분석된 단어들을 필터링하기 위한 다양한 비속어들과 욕설을 포함한 DB를 구축하고 이를 업데이트하는 AI도 개발해야 합니다.

7. 출처