**CYBERAFFITI 사이버래피티**

**유해도 기반의 필터링 시스템**

**빅데이터 청년인재 고려대학교 3조**

**남궁찬 박재은 장민교**

**정우태 조민수 채수영**

목차

**Ⅰ 서론1**

1. 개발 배경2
2. 시장 조사 및 필요성3

**Ⅱ 본론4**

1. 데이터 수집5

1.1) 수집 대상6

1.2) 수집 방법6

1. 전처리5

2.1) 투표 기반 레이블링6

2.2) 토크나이즈6

2.2.1) 형태소 분석 기반6

2.2.2) Bype Pair Encoding6

2.2.3) 직접 개발한 방식6

3. 모델5

3.1) 1dcn6

3.2) Self-Attention Bidirectional-LSTM6

4. 확장프로그램5

1.1) CHROME 확장 프로그램6

1.2) 확장프로그램 서버 연동6

**Ⅲ 결론4**

1. 개발 결과5
2. 기대 효과 6

**Ⅰ 서론**

1. 개발 배경

정보화 시대의 흐름 속에서 현대인들은 여러 매체를 이용하여 다양한 정보를 얻을 수 있게 되었습니다. 인터넷과 모바일을 이용하면 사용자는 언제든 어디서든 원하는 정보를 쉽게 구할 수 있지만, 이러한 장점에는 그에 버금가는 부작용도 존재하고 있습니다. 선정적이고 폭력적인 미디어들에 대한 접근 또한 용용 해진 것은 물론이고, 심의 규정에 따라 제재를 하는 경우에도 검색을 통해 우회하여 접속하는 방법을 쉽게 알아낼 수 있게 되었기 때문입니다.

1. 시장 조사 및 필요성

이러한 제재의 경우에도 인터넷 방송 플랫폼의 경우에는 아주 최소한으로만 적용되고 있습니다. 인터넷 방송의 경우에는 범용 인터넷 망을 사용하지 않고 있는데, 이는 방송 구역이라는 개념이 적용되지 않아서 플랫폼의 자체적인 제재에만 기댈 수밖에 없기 때문입니다. 따라서 현재 대표적인 인터넷 방송 플랫폼으로 손 꼽히는 유투브, 아프리카티비, 트위치 등에서는 선정적이고 폭력적인 멘트 및 영상들이 인터넷 망을 통해 유포되고 있습니다. 문제가 되는 부분은 이 유해 영상들에 대한 필터링이 제대로 수행되고 있지 않는다는 점입니다. 트위치 및 아프리카 티비에서는 채팅방에 필터링을 수행하는 기능이 존재하지만 정확한 비속어 및 욕설을 사용해야만 필터링이 되기에 변형된 욕설들까지는 잡아내지 못하고 있습니다. 또한 유투브는 신고 기반으로 필터링이 되고 있기에 유해 영상에 대한 신고가 들어오지 않는다면 제재가 되지 않는 채로 공개되는 문제가 발생하고 있습니다.

2018년 방송 통신위원회에서 조사한 결과에 따르면 10대 청소년 10명 중 1명이 인터넷 방송을 시청하며 평균적으로 하루에 2시간씩 시청을 한다고 합니다. 그리고 어린이 및 청소년들이 주로 시청하는 영상 중 27개의 콘텐츠에 대한 모니터링을 수행한 경우, 총 1034개의 유해 내용들이 발견된 점을 밝히며 아직 가치관이 확립되지 않은 어린이 및 청소년들에게 끼칠 수 있는 부정적인 영향력의 상관 관계를 보여주었습니다. 따라서 저희는 사이버 공간상에서 존재하는 낙서를 필터링하고 지워주는 목적을 통해 사이버래피티[[1]](#footnote-1) 프로젝트를 수행하게 되었습니다.

1. 사이버래피티: 사이버와 그래피티(낙서)의 합성어. [↑](#footnote-ref-1)