



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

청소년으로부터
SNS 내 불법 스포츠 도박 광고를
차단하기 위한 방안

2018년 2월

조선대학교 교육대학원

정보컴퓨터교육

김 지 아

청소년으로부터
SNS 내 불법 스포츠 도박 광고를
차단하기 위한 방안

A Plan to Block Illegal Gambling Advertisements
in Social Network Services of Teenagers

2018년 2월

조선대학교 교육대학원

정보컴퓨터교육

김 지 아

청소년으로부터
SNS 내 불법 스포츠 도박 광고를
차단하기 위한 방안

지도교수 조 범 준

이 논문을 교육학석사(정보컴퓨터교육)학위
청구논문으로 제출함

2017년 10월

조선대학교 교육대학원

정보컴퓨터교육

김 지 아

김지아의 교육학 석사학위 논문을 인준함.

심사위원장 조선대학교 교수 배 용 근 인

심사위원 조선대학교 교수 강 문 수 인

심사위원 조선대학교 교수 조 범 준 인

2017년 12월

조선대학교 교육대학원

목 차

ABSTRACT	i
I. 서 론	1
1. 연구의 필요성	1
2. 연구내용과 범위	4
II. 관련연구	7
1. 불법 스포츠 도박 실태 조사	7
2. 유해 콘텐츠 차단을 위한 노력	10
3. 가중치 기법에 관한 연구	13
III. 빈도 별 가중치 공식을 적용한 광고 차단 방법	16
1. 이미지화 된 단어 분석	18
1) A 그룹	21
2) B 그룹	21
3) C 그룹	22
2. 사용자정의 가중치 및 빈도 별 가중치 공식	23
1) 사용자정의 가중치	23
2) 빈도 별 가중치	25
3. 차단 여부 판정	29

IV. 실험 및 결과	32
V. 결론	38
참고문헌	41

〈표〉 목 차

〈표 1〉 기존 가중치 기법의 종류	14
〈표 2〉 그룹 별 단어 분류	22
〈표 3〉 단어 빈도수와 그룹 별 빈도수 평균	26
〈표 4〉 <i>TF</i> 데이터베이스	28
〈표 5〉 i의 결과 값	35
〈표 6〉 ii의 결과 값	35
〈표 7〉 iii의 결과 값	35

〈그림〉 목 차

〈그림 1〉 청소년 불법 스포츠 도박 금지 홍보 동영상.....	13
〈그림 2〉 청소년 불법 스포츠 도박 금지 캠페인	13
〈그림 3〉 불법 스포츠 도박 차단 모형	18
〈그림 4〉 실험 대상에 대한 빈도 별 가중치 공식 적용을 나타낸 분포도	31
〈그림 5〉 실험 이미지 i	33,36
〈그림 6〉 실험 이미지 ii	33,36
〈그림 7〉 실험 이미지 iii	33,36
〈그림 8〉 실험 이미지 iv	33
〈그림 9〉 실험 이미지 v	33
〈그림 10〉 실험 이미지 vi	33

ABSTRACT

A plan to block illegal gambling advertisements in Social Network Services of teenagers

Ji A Kim

Advisor : Prof. Beom Joon Cho

Major in Information and Computer Science Education

Graduate School of Education, Chosun University

In the last decade, the penetration rate of smartphones is growing rapidly. As smartphones become common, usage rate of internet and social networking has increased. Young people in teenager and 20s have the advantage of enjoying the amount of informations they can get fast in social networking, but they also have the disadvantage of getting easily informations are harmful to youths.

Among them, illegal sports gambling advertisement is the most accessible. Most of the illegal sports gambling vendors post pictures are inserted a saying they want to promote, so they can avoid text-based filtering methods. By monitoring advertisements for illegal sports gambling, I judged that social networking programs need stronger prevention methods than passive prevention,

which tipped off about illegal ads and blocked that.

Therefore, this dissertation proposes the following methods to aggressively block illegal sports gambling ads. Firstly, get samples of illegal sports gambling advertising images are used for advertising, analyze words are used frequently and classify by the group according to their importance. Secondly, calculate the value by applying the weighted equation for each word. Based on the results of the survey, it is judged whether or not an illegal sports gambling advertisement is classified..

The result of the experiment in the proposed method, Of the 200 images of the 200 images, 193 images were appropriately identified. In other words, 96.5 % of the success rate was shown in the experiment. Research is likely to be needed for failure rates that have been disallowed for later years.

I. 서 론

1. 연구의 필요성

스마트폰은 2000년대에 인류의 삶에 들어섰으며 지금까지 많은 이들의 사랑을 받고 있다. 우리나라에도 역시 2009년부터 보급되기 시작했고, 현재는 우리나라의 85%가 넘는 인구가 스마트폰을 사용하고 있다. 이는 한국인 5명 중 4명이 스마트폰을 사용하고 있다는 것을 의미한다. 이처럼 스마트폰은 우리 삶에 없어서는 안 되는 필수적인 요소로 자리매김하였다.

스마트폰 보급률이 높아지면서 동시에 인터넷 이용률도 크게 증가하였다. 국내에 스마트폰이 도입되기 전인 2008년에는 남성 81.6%, 여성 71.5%가 인터넷을 사용했다면 2016년에는 남성 91.0%, 여성 85.6%이나 되는 인구가 인터넷을 이용하고 있다. PC로 이용하던 쇼핑, 게임, 은행업무, 문서작성, 멀티미디어 콘텐츠 이용 등을 스마트폰의 보급 이후로 장소 제약 없이 몇 번의 조작만으로 편리하게 이용할 수 있게 되었다.[1]

통계에 따르면 사람들은 스마트폰을 주로 메신저(97.6%), SNS(96.2%), 게임(90.3%) 등의 용도로 많이 이용한다고 답했다. 그 중에서도 청소년은 스마트폰으로 메신저를 99.5%, SNS를, 97.5%, 게임을 94.9%만큼 이용한다고 답했다. 청소년 10명 중 9명 이상이 이와 같은 용도로 스마트폰을 이용하고 있는 셈이다. 또한 이를 통해서 전체 연령이 메신저, SNS, 게임을 이용하는 것보다 청소년이 이용하는 빈도가 더 많다는 것을 알 수 있다. 그 중 청소년들은 스마트폰을 하루 평균 4.8시간 이용한다고 답했는데, 이는 아동 1.7시간, 성인 4.6시간에 비해 높은 것을 알 수 있다. 그 중에서도 청소년들이 스마트폰에 의존하는 경향이 높다는 것을 확인하였다.

청소년의 스마트폰 과의존 위험군은 무려 전체 스마트폰 이용 청소년 중 31.6%, 고 위험군은 4%, 잠재적위험군은 27.6%로 나타났다. 이는 성인의 스마트폰 과의존 위험군 13.5, 고위험군 2.1%, 잠재적위험군 11.4%에 비해 2~3배는 높은 비율로 대처가 필요한 상황임을 알 수 있다.[2] 또 다른 연구에 따르면 청소년들이 스마트폰으로 가장 많이 하는 일로 대부분 SNS를 꼽았고, 그 다음으로는 게임을 꼽았다.[3] 즉, 청소년들이 스마트폰을 하는 데에 대부분의 시간이 SNS 혹은 게임으로 소요되고 있다는 사실을 알 수 있다.

또, SNS를 이용하는 주된 목적으로는 ‘재미있어서, 혹은 스트레스 해소를 위해서’가 61.8%로 1위를 차지했고, 그 밖에 ‘사람들과 교류하기 위해’(45.7%), ‘새로운 정보를 획득하고 공유하기 위해’(28.8%) SNS를 이용한다고 답했다. 여기서 SNS란 Social Network Services의 준말로 각각의 사용자가 글이나 영상 등의 정보를 다수의 사용자에게 공유하면서 타인과 네트워크를 형성하는 서비스를 의미한다. 즉, SNS를 이용함으로써 다양한 정보를 쉽게 접할 수 있다. 이것은 SNS의 장점이자 단점으로도 작용하고 있는데, 그 이유는 사용자가 얻는 정보가 원하는 정보일 수도 있고 원치 않는 정보일 수도 있기 때문이다.[2]

원하지 않는 정보에는 청소년에게 해가 될 수 있는 매체가 다수 포함되어 있다. 청소년보호위원회는 이처럼 청소년에게 유해한 것이라고 판단되는 정보들을 「청소년보호법」 규정에 따라 청소년유해매체물이라고 정의하고 있다. 청소년보호법에 따른 매체물에는 영화 및 비디오물, 게임물, 음반, 음악파일, 음악영상, 공연, 전기통신을 통한 부호·문언·음향 또는 영상정보, 방송프로그램, 특수일간신문, 일반주간신문, 특수주간신문, 일반주간신문, 인터넷 신문 및 인터넷 뉴스 서비스, 잡지 등 정기간행물, 출판간행물, 전자출판물, 외국간행물, 옥외광고물, 상업적 광고 선전물 등이 있다.

즉, 청소년들은 SNS를 통해서 원치 않는 정보인 청소년유해매체물에 노출되어 있다.

문제는 SNS 이용목적과 청소년유해매체물의 상관관계에 있다. 위에서 언급한 바와 같이 SNS를 이용하는 가장 큰 목적이 재미있어서, 혹은 스트레스 해소를 위해서인데, SNS에서 찾는 즐거움은 지인들과의 교류, 재미있는 정보, 자극적인 정보도 있지만 충동과 사행심을 조장하는 불법 광고물도 존재한다. 선정적인 사이트를 홍보하는 광고물과 주로 불법 스포츠 도박 사이트를 홍보하는 광고물이 그 예이다. 불법 광고업자가 접근이 쉽다는 SNS의 특성을 이용해서 광고를 하면 그 광고에 청소년들도 쉽게 노출이 된다.

자극적이고 충동적인 것에 쉽게 반응하는 청소년들은 평소에도 돈내기 게임에 많은 관심을 보이고 있었는데, 조사에 따르면 청소년 중 한 번이라도 돈내기 게임에 참여한 적이 있다고 답한 청소년들은 온라인용 내기 게임을 하루 평균 87.3분, 불법 인터넷 도박을 78.7분, 한게임이나 넷마블 등에 있는 카드나 화투 게임을 75.5분 정도 이용한다고 말했다. 또 그 중에서도 자주 이용한 돈내기 게임에서 이용한 금액이 불법 인터넷 도박에서는 234.476원, 온라인용 내기 게임에서는 152.947원 등 적지 않은 금액을 투자한다고도 이야기했다. 도박을 해본 적 있는 청소년은 전체 청소년의 24.2%이고, 그 중에서도 위험군이 4.0%, 문제군이 1.1%이다. 일반 성인의 도박 위험군, 문제군 비율보다도 높다.[1]

이러한 문제 상황을 인식하고 불법 스포츠 도박 광고를 제재하려는 시도가 SNS 내에서도 이루어지고 있다. 현존하는 SNS 중에서 접근성이 높은 편이며 불법 스포츠 도박 광고를 가장 빈번하게 접할 수 있는 페이스북에서도 다양한 노력을 하고 있다. 불법 광고를 발견할 시 신고를 부탁하는 문구나, 커뮤니티 규정을 개정하여 페이스북 이용자들이 불법 스포츠 도박 광고에 대처할 수 있도록 하고 있다. 그러나 위와 같은 규정은 주로 이용자들이 쉽게 찾기 어려운 곳에 안내되어 있다. 가정이나 학교, 개인에게 당부하는 안내문도 마찬가지이다. 물론 사용자의 마음가짐이나 적극성이 요구되는 사항이긴 하나 관리자의 노력 또한 필요하다. 심지어 고객센터 페이지조차 불법 스포츠 도박 광고물로 도배가 되어 있는 모습을 확인하였다.

가장 적극적이라고 볼 수 있는 방법은 회원들에게 신고를 해달라고 안내를 하고 접수된 신고에 한해 문제를 해결하는 방법인데 이조차도 접수된 신고에 제한된다는 점에 있어서 아주 적극적이라고 볼 수는 없다. 또 자체적으로 모니터링을 한다고 안내하였으나 방대한 양의 정보를 모두 검열하기에는 어려운 부분이 있다.

이에 본 논문에서는 불법 스포츠 도박 광고가 청소년에게 심각하게 노출되어 있음을 인지하고, 단어를 기반으로 한 가중치 알고리즘으로 청소년으로부터 불법 스포츠 도박 광고를 적극적으로 차단할 수 있는 방법을 제시하고자 한다.

2. 연구내용과 범위

본 논문에서는 불법 스포츠 도박 광고가 가장 많이 노출되어 있는 SNS인 페이스북을 중심으로 하여 연구했음을 밝힌다. 페이스북은 접근성이 높은 SNS이다. 가입을 할 때에는 이름과 휴대폰 번호 또는 이메일, 그리고 사용할 비밀번호만 있다면 누구나 쉽게 자신의 정보를 등록할 수 있다. 글이나 영상 등의 정보를 공유할 때에도 간편하게 올릴 수 있으며 친구가 올린 정보는 타임라인에 최신 순 혹은 인기 순으로 나열이 되어서 확인할 수 있다. 이때 사용자가 올린 정보에 타인이 ‘좋아요’라는 버튼을 클릭하게 되면 타인의 친구에게도 해당 정보가 공개될 수 있도록 설정할 수 있다. 이러한 방법으로 서로를 알지 못하는 관계 속에서도 정보를 주고 받을 수 있다.

그렇다면 불법 광고업자들이 불법 스포츠 도박 광고를 어떤 식으로 행하고 있는지에 대해 알아볼 필요가 있다. 먼저, 많은 양의 광고를 적은 인력으로 하는 데에는 한계가 있기 때문에 일명 ‘댓글 아르바이트생’을 모집해서 광고 인력을 대체하고 있다. 광고업자들은 쉬운 노동으로 많은 돈을 벌 수 있게 해준다고 유혹하며 학생들의 개인정보가 담겨있는 페이스북 계정을 광고에 이용하도록 유도한다. 특히

이런 종류의 아르바이트는 중개업을 거치거나 허가받은 아르바이트가 아니므로 유의해야 한다.

페이스북 내에서 불법 스포츠 도박을 홍보하는 가장 기초적인 방법으로 정보를 공유하는 글을 게재할 때 하단에 주석처럼 불법 스포츠 도박 광고 문구를 적는 방법이 있다. 또, 각 정보마다 타인이 코멘트를 할 수 있도록 설정이 되어 있는데, 대부분의 업주들이 이 점을 이용해 불법 스포츠 도박을 광고하고 있다. 코멘트를 단순한 문자로 표현할 수도 있지만 영상을 이용해 나타내기도 한다. 업주들은 글로 표현하는 방식보다는 영상으로 교묘히 홍보하는 방식을 주로 이용한다. 대부분 사진 속에 홍보하고 싶은 문구를 삽입해서 이미지화한 뒤에 그 영상을 코멘트에 올린다. 이는 영상물을 이용함으로써 페이스북 사용자의 눈에 띄기 쉽게 하고, 게시물 하단에 주석처럼 홍보문구를 삽입하거나 코멘트에 문자로 홍보를 했을 때 특정 단어에 의해 제재될 수 있기 때문에 번거롭지만 위와 같은 방식을 이용하는 것으로 보인다.

단순히 모니터링을 통해 제재가 힘든 부분은 또 있다. 불법 광고업자들은 불법이라는 위험성과 메신저의 편의성 때문에 대부분 따로 홈페이지를 제작하지 않고 메신저를 이용해 불법 스포츠 도박을 행하고 있다. 그렇기 때문에 해당 불법 스포츠 도박 광고가 유해 사이트로 차단되기에 어려운 부분이 있다.

이러한 특수한 상황으로 보았을 때 단순히 모니터링을 하는 기존의 방식보다는 새롭고 더 강력하게 불법 스포츠 도박을 차단할 수 있는 방법이 필요하다는 것을 제안한다.

본 논문에서 제안하는 방법은 불법 스포츠 도박 광고마다 사용되는 특정한 단어들을 수집하고 특성에 맞게 분류하여 분류된 그룹마다 가중치를 다르게 부여한 뒤 빈도별 가중치 공식에 적용하고 그 결과 값을 바탕으로 차단 여부를 판별하는 방

범이다.

2장에서는 선행된 연구 중 본 논문과 관련된 연구 방법에 대해 기술한다.

3장에서는 제안하는 방법에 대해 상세히 기술한다. 불법 스포츠 도박 광고에서 자주 쓰이는 단어를 수집하고 분류하는 기준에 대해 설명하고 사용자정의 가중치, 빈도 별 가중치 공식에 대해 상세히 제시한다.

4장에서는 본 논문에서 제안한 방법을 이용해 실험을 한 내용에 대해 기술한다. 또한 실험을 토대로 한 결과에 대해서도 설명한다.

5장에서는 방법과 방법을 바탕으로 한 실험 내용을 종합해 결론을 맺는다.

II. 관련연구

1. 불법 스포츠 도박 실태 조사

사행산업통합감독위원회법에 따르면 사행사업과 관련하여 개별 법률에서 금지 또는 제한하는 행위, 사행성게임물을 이용하여 사람들이 사행행위를 할 수 있게 서비스를 제공하는 행위, 정보통신망을 통하여 사람들이 사행행위를 할 수 있게 서비스를 제공하는 행위를 불법사행산업으로 정의하고 있다. 여기서 법률에서 제한하지 않는 사행사업은 카지노업, 경마, 경륜, 경정, 복권, 체육진흥 투표권, 소싸움경기를 의미한다.

현재 합법 스포츠평토토를 관리하고 있는 문화체육관광부와 국민체육진흥공단, 수탁사업자인 케이토토(주)는 국민체육진흥법 제24조를 근거로 하여 체육진흥투표권 발행 사업을 진행하고 있다. 나아가 국민체육진흥공단과 수탁사업자가 아닌 자는 체육진흥투표권이나 이와 유사한 것을 발행하여 결과를 적중시킨 자에게 재물이나 재산상의 이익을 제공하는 행위를 국민체육진흥법을 통해 금지하고 있다.[3]

그러나 합법적인 스포츠 도박을 이용하지 않고 불법 스포츠 도박을 이용하는 사람들은 베팅 금액에 제한이 없고, 보다 다양한 운동경기에 대해 베팅이 가능하다는 점, 그리고 합법 스포츠평토토와는 다르게 공익사업으로 사용되는 돈이 없기 때문에 정부 허가 사업에서의 배당액보다 높다는 점을 들어서 높은 기대치를 가지고 불법 스포츠 도박을 더 많이 이용하고 있다.[4]

합법 스포츠평토토는 10만원의 제한이 있는 반면에 불법 스포츠 도박은 최소 1만원부터 많게는 1000만원 이상의 베팅 한도를 가지고 있기 때문에 실상 그 한도가 무의미하다고 보아야 한다. 또, 다양한 결과에 대한 베팅으로 참가자들의 흥미를

유발하여 그 유혹에서 빠져나올 수 없게 한다. 스포츠토토 신고센터의 자료에 따르면, 불법 스포츠 도박 신고는 2009년 5395건에서 2011년 1만3755건으로 2년 새에 약 2.5배 증가했다.[5]

불법 스포츠 도박은 해외 서버 운영 등의 이유를 들어 국가의 통제가 어려운 환경에서 불특정 다수가 높은 사행성과 중독성으로 금전적 피해와 심리적 제반 문제에 노출될 수 있다는 점에서 문제가 되고 있다. 또한 소비자 권익보호가 어렵고 검은 돈의 유통창구로 이용되고 있어 사회적으로도 악영향을 미치고 있다.[6]

최근에는 불법 스포츠 도박을 홍보하기 위해 다양한 유인책을 사용하는 것으로 확인되었다. 블로그, 카페, 인터넷 방송, SNS 등 인터넷을 활용해 참여자를 모집하는 방법이 있는가 하면 문자 메시지를 이용해 불특정 다수의 참여자를 모집하기도 하고 기존에 불법 스포츠 도박을 하던 이용자가 지인에게 홍보를 하는 방법도 존재했다.[7] 특히 컴퓨터만 있으면 누구나 쉽게 접근이 가능한데다 가입절차가 간단해서 참여가 지속적으로 늘고 있는 불법 스포츠 도박은 수사기관의 단속망을 피하기 위해 해외 서버를 운영하기도 하는데 지속적인 홍보를 막기가 어려워 최근에는 10대 청소년들 사이에서까지 유행처럼 번지고 있다.[8]

불법 온라인 도박 사이트를 포털사이트에 검색하도록 하였을 때 네이버를 통해 114개, 구글 112개, 다음 10개, 네이트 5개의 결과가 검색되었고, 그밖에도 SNS를 통해 사이트에 대한 홍보가 이루어지는 경우가 있었다. 특히, 지인과의 관계성에 기초한 홍보가 이루어지는 페이스북에서는 거의 대부분 불법 스포츠 도박 광고였다.[9]

최근에는 불법 스포츠 도박을 이용하는 청소년들이 대부분 고등학교 2학년 때 처음으로 불법 스포츠 도박을 접했고, 주로 인터넷, 휴대폰 어플을 이용해서 불법 스포츠 도박을 이용하며 대부분이 본인 확인 인증 절차 없이 불법 스포츠 도박을 이용하고 있다는 연구가 발표됐다.[10] 사람들이 많이 하는 불법 스포츠 도박으로는 ‘축구’, ‘카지노베팅’ 순으로 인기가 가장 많았고, 불법 스포츠 도박을 끊을 수 없

는 이유는 대부분 재미 때문이거나, 돈을 딸 수 있을 것 같은 기대감 때문이었다.[11] 최근 연구에 따르면 사람들이 최초로 도박을 경험한 시기는 대부분이 대학교 이상이었으나 고등학교 때 처음 접한 경우나 중학교 때 처음 접한 경우도 적지 않았다. 그리고 도박을 경험한 대학생들 중에서도 대학교 때 도박을 경험한 사람보다 중, 고등학교 때 도박을 경험한 사람, 즉 더 어린 나이에 도박을 경험할수록 도박중독 비율이 높게 나타났다고 보고했다.[12]

따라서 불법 스포츠 도박은 많은 위험 요소를 포함하고 있음을 인지해야 한다. 특히 청소년들의 직·간접적인 경험 가능성을 높이는 불법 스포츠 도박은 다른 유형의 불법 도박에 비해 더 큰 잠재적 위험성을 가지고 있다. 연구에 따르면 23.6%의 청소년이 불법 스포츠 도박 사이트에 방문해 본 경험이 있다고 답했고, 전체의 10%에 해당하는 청소년들이 지금도 정기적으로 인터넷 도박을 하고 있다고 한다.[13]

2017년 2월에 열린 청소년 불법 스포츠 도박 확산 방지를 위한 간담회에서 언급된 사례도 있다. 고등학교 3학년 학생이 PC방 화장실에 붙어있는 광고지를 보고 불법 스포츠 도박을 시작했다가 결국 1500만원의 빚을 지게 됐고, 부모가 해결해주었다. 그러나 다시 도박에 빠져서 700만원의 빚을 지게 되었고, 도박 외에는 갚을 방법이 없다고 생각한 학생은 또 다시 도박에 빠지고, 결석, 가출이라는 결과까지 낳게 되었다.

불법 스포츠 도박이 성행한 이래로 수많은 사건, 사고들이 주변에서 발생하는 것도 문제가 되고 있다. 대표적인 예가 2015년에 발생한 30대, 40대 남성 동반 자살 사건이다. 경찰에 따르면 평창의 한 펜션에서 두 남성이 동반 자살을 하는 사건이 발생했는데, 이들은 각각 불법 스포츠 도박으로 인해 1억 원, 5억 원의 빚을 지고 있는 것으로 밝혀졌다. 이와 같이 불법 스포츠 도박은 결국 가정불화, 각종 범죄, 심하면 살인이나 자살에까지 이르는 심각한 문제이므로 특히 충동적인 행동에 약한 청소년들을 불법 스포츠 도박으로부터 보호해야 한다.

2. 유해 콘텐츠 차단을 위한 노력

위와 같은 실태가 원인이 되어 현재도 불법 스포츠 도박 광고를 포함한 유해 콘텐츠를 차단하기 위한 노력들이 이어지고 있다.

2009년에는 내용을 기반으로 한 유해 멀티미디어 차단 기술 동향에 대한 연구가 보고되었다. 먼저, 메타 정보 및 텍스트 기반 차단 기술은 멀티미디어의 제목, 제작 정보, IP 주소, 또는 멀티미디어의 부가 설명에 포함된 텍스트 정보를 기반으로 하여 자주 등장하는 특정 키워드를 이용하여 유해 멀티미디어를 찾아내어 차단하는 방법이다. 해쉬 및 DB 기반 차단 기술은 유해 멀티미디어에 대한 해쉬값을 계산, 그 값을 DB화 하여 유해 멀티미디어 차단에 활용하는 방법이다. 전처리 기술은 멀티미디어의 정보를 최소화, 혹은 극대화 시켜 내용을 분석하는 방법이다. 특징 추출 기술은 노출이 심한 유해 멀티미디어의 특징을 이용해 영상으로부터 피부색을 검출하여 판별하는 방법이다.[14][15]

2000년에는 웹 문서 속에 포함된 영상을 19*19의 크기로 분할하여 분석하였고, 특징에 따라 비유해 영상과 유해 영상으로 구분하도록 한 연구가 발표되었으며,[16] 2004년에는 정보통신윤리 위원회에서 제시한 유해단어의 가중치를 이용, 가중치 평균을 더해 유해도를 판단하는 방식을 이용해 유해 사이트를 차단하는 방법이 제안되기도 했다.[17]

또 2015년에는 동영상에서 프레임을 추출, N개의 유해 프레임 분류기가 유해한 콘텐츠인지 유해성 여부를 판단하고 최종 결과를 반환하는 실험을 진행한 연구가 발표됐다. 이때 사용된 융합 방법은 Precision 위주의 융합 기법, Recall 위주의 융합 기법, 그리고 가중합 기반의 다중 분류기 융합 기법이다.[18]

유해한 문자에 대한 차단 방법도 제안되었다. 2009년에는 어떤 어휘로부터 쓰여진 문장에 대해 확률을 측정할 수 있는 확률인 언어 모델을 이용해서 유해와 무해 토픽으로 분류하는 시스템을 구현한 논문이 발표됐고, 해당 프로그램을 통하여 자질어 추출 알고리즘 사용에 대한 정확도를 평가하고자 하였다.[19]

2004년에는 음란 비디오 영상의 소리를 파악하고 이를 기반으로 해서 음란 유해 사이트를 차단하고자 하였다. 음란 유해사이트를 차단하는 기존의 방법인 목록기반과 단어기반 방법이 가지는 문제를 해결하기 위해 소리를 분석하여 차단하는 새로운 방법을 제시하였다.[20]

2009년에는 기존에 있는 유해정보 차단 소프트웨어 11개를 분석하여 주요기능을 정리한 연구가 발표되었고, 유해정보 차단 소프트웨어의 부족한 부분에 유해사이트, 유해파일, 프로그램 차단 기능, 사용시간 제한설정 기능, 관리자 알림 기능, 감사 기록 생성 기능 등이 필요하다고 주장하였다..[21]

또, 안드로이드의 유해 애플리케이션인 일명 “블랙마켓”을 예로 들면서 청소년 유해 콘텐츠의 쉬운 접근을 지적한 연구도 있었다. 안드로이드의 유해 콘텐츠 차단 애플리케이션의 일부는 유해 사이트는 차단이 가능하지만 게시글에 기재된 음란물 링크는 차단되지 않는다고 말하며 안드로이드의 로컬 프록시를 이용해 웹 사이트의 접속과 차단되지 않은 웹 페이지의 유해 콘텐츠까지 차단할 수 있음을 증명했다.[22] 안드로이드의 유해 콘텐츠를 차단하기 위한 또 다른 방법도 제시되었는데 유해 콘텐츠를 차단하기 위해 안드로이드 플랫폼에 있는 인텐트를 모니터링하는 방법을 제안하기도 하였다.[23]

또, 웹 사이트들이 하이퍼링크로 연결 관계를 형성한다는 특성을 이용하여 유해 사이트를 판별하는 실험에 대한 연구도 있었다. 기존 유해사이트에 관한 웹 페이지를 수집하고 URL을 추출한다. 추출된 정보를 이용하여 유방향 그래프를 생성, 그래프에 존재하는 순환을 찾아낸 뒤에 이 순환에 포함되는 웹 사이트는 유해사이트로 판별하도록 하였다.[24]

유해사이트 DB를 구축하기 위해 검색 요원들이 유해사이트를 돌아다니며 데이터를 수집하여 분류하거나, 웹 검색 로봇을 이용해 패턴 매칭 방법으로 분류하는 기존의 방식보다 정확도를 높일 수 있는 링크 정보를 기반으로 1-Depth, 2-Depth 사이트의 유해지수를 계산해 유해성 여부를 검증하기도 하였다.[25]

유해한 사이트를 차단하는 방법에 대해서도 제시되었는데, 해당 논문에 제시된 방법은 길이가 긴 주소를 Short URL 서비스 서버로 보내고 서버에서 Short URL을 반환해주는 서비스인 Short URL 서비스를 분석하여 문제점을 분석하고 Short URL 서비스를 기반으로 한 유해 사이트 접속 차단 기법이다.[26]

또, 구글 검색을 통해 개인정보 판매 게시글을 차단하는 기법을 제안한 연구도 있었는데, 먼저 판매 게시글에 대한 단어 형태를 분석하고, Google Custom Search Api를 이용하여 자주 사용하는 단어가 DB에 저장되도록 설계한 뒤, 검색한 결과를 바탕으로 차단하는 방법이다.[27]

기존의 유해 콘텐츠를 차단하는 국내외 프로그램에 대해 분석한 연구 결과도 있다. 유해 콘텐츠를 차단하는 기존 국내외 프로그램들을 분석하고, 인터넷 내용 등급 서비스를 기반으로 하는 선별 차단 기술을 이용하는 인터넷 내용 선별 플랫폼 기반 선별 차단 프로그램을 제안한 논문이다.[28]

그밖에도 스포츠토토 수탁사업자인 케이토토에서는 청소년에게 기승을 부리고 있는 불법 스포츠 도박의 위험성을 알리고 근절하기 위한 캠페인인 ‘꿈지꿈이’를 전개하고 있으며, 2013년에도 불법 스포츠 도박을 근절하기 위해 게임 형식으로 이벤트를 진행한 바 있다. 또 한국도박문제관리센터에서도 불법 스포츠 도박 등으로 인해 도박 문제의 폐해가 증가함에 따라 스포츠 경기와 연계하여 도박 문제 예방에 대한 홍보를 진행하였다.

위와 같이 유해 콘텐츠를 차단하기 위한 많은 연구들이 있었고 지금도 계속되고 있다. 하지만 그동안 진행된 유해 콘텐츠를 차단하기 위한 연구의 대부분은 유해 콘텐츠 중에서도 선정적인 콘텐츠에 집중하고 있었고 불법 스포츠 도박에 관련된

컨텐츠를 차단하기 위한 연구 논문은 쉽게 찾아볼 수 없었다. 또 차단을 해야하는 대상이나 차단 방법으로는 애플리케이션이나 웹 사이트를 차단하는 방법에 초점을 둔 연구가 많음을 확인할 수 있었다. 그리고 텍스트 추출을 기반으로 한 연구는 많았지만 이미지화 된 텍스트를 분석하는 방법은 아직 제시되지 않고 있음을 알 수 있다.



<그림 1> 청소년 불법 인터넷 도박 금지 홍보 동영상



<그림 2> 청소년 불법 인터넷 도박 금지 캠페인

3. 가중치 기법에 관한 연구

기존의 연구에 따르면 가중치를 부여하는 기법은 다양하게 존재한다. 특히 용어에 대한 가중치 기법에 대한 선행연구를 중점적으로 조사하였다.

이름	공식	특징
이진 TF	$TF = 1(\text{if } tf > 0), 0$	단어가 출현한 경우 모두 1
단순 TF	$TF = tf$	단어빈도 tf 를 그대로 사용
로그 TF	$TF = 1 + \log(tf)$	단어의 영향력 차이를 낮추기 위해 TREC-1의 SMART팀이 제안
더블로그 TF	$TF = 1 + \log(1 + \log(tf))$	로그를 두 번 취하는 방식
루트 TF	$TF = \sqrt{tf}$	로그 TF보다 덜 축소하는 공식
보정 TF	$TF = (1 - w) + w \times \frac{tf}{max_tf}$	가중치를 일정 범위로 한정시키는 방식
Okapi TF	$TF = \frac{tf}{2 + tf}$	2-포아송 모형을 적용
더블로그2 TF	$TF = 1 + \log_2(1 + \log_2(tf))$	tf 가 3 이상일 때부터 가중치가 로그 곡선을 띄는 방식
루트직선 TF	$TF = \frac{tf + 3}{4}$	루트 TF의 공식과 가중치 기울기가 유사하도록 고안된 방식

<표 1> 기존 가중치 기법의 종류

기존 연구에 따르면 단어빈도 가중치 공식을 다음과 같이 정리할 수 있다. 단어가 출현한 경우를 모두 1로 지정하는 이진 TF, 단어빈도 tf 를 그대로 사용한 단순 TF, 단어의 영향력 차이를 낮추기 위해 TREC-1의 SMART팀이 제안한 로그 TF, TREC-7에서 제안된 로그를 두 번 취하는 방식인 더블로그 TF, 로그 TF보다 덜 축소하는 공식인 루트 TF, 가중치를 일정 범위로 한정 시키는 방식의 보정 TF, 2-포아송 모형을 적용한 공식인 Okapi TF, tf 가 3 이상일 때부터 가중치가 로그 곡선을 띄는 방식인 더블로그2 TF, 루트 TF의 공식과 가중치 선의 기울기가 유사하도록 고안된 방식인 루트직선 TF 등이 있다.[29]

2001년에는 한글 문서에서 추출된 용어의 가중치를 결정하기 위해 가중치를 계산하는 방법을 제안한 연구가 있었는데, 해당 연구에서는 용어의 유형과 조사 정보에 의해서 경험적인 방법으로 유형에 따라 가중치를 부여하는 방식을 택하였다.[30]

또 스팸 메일을 필터링하기 위한 메일 주소 유효성과 제목-내용의 중요도에 따른 가중치 기법을 제안한 연구도 있었다. 메일 헤더에 기록된 주소와 송/수신자 메일 주소가 일치하지 않는 점과 스팸 메일의 제목의 특징을 이용해 스팸 메일 필터링 시스템에 메일 주소 유효성 검사와 제목 및 내용을 구분하여 스팸 확률을 계산하는 기법을 제안하였다. [31]

그 밖에 자료를 정규분포로 만드는 Box-Cox 변환기법을 토대로 가중치를 부여하는 계산식을 만들고, 용어 빈도를 정규화 할 수 있는 방식을 제안 논문도 있었다.[32]

Ⅲ. 빈도 별 가중치 공식을 적용한 광고 차단 방법

청소년에게 유해한 청소년유해매체물과 관련되어 있거나 유해한 콘텐츠를 차단하기 위한 방법을 제시하는 일은 이전부터 현재까지 끊임없이 연구되어지고 있는 문제이다. 그러나 기존의 연구는 음란 유해사이트와 관련된 연구가 대부분이었고, 주로 웹 사이트를 차단하는 방식과 애플리케이션을 차단하는 방식을 이용했다.

이는 과거에는 불법 스포츠 도박 광고보다는 음란 유해사이트를 차단하는 문제가 우선이었기 때문이다. 한때 ‘바다이야기’ 사건으로 오프라인 도박장이 이슈가 되기 전에는 온라인에서 성행하는 불법 스포츠 도박은 크게 활성화되지 않았고, 오프라인 도박장이 이슈가 된 이후에 불법 도박 업주들이 오프라인보다는 온라인에 눈길을 돌렸다. 바다이야기가 이슈가 된 해가 2006년인 점을 감안하면 약 10여 년 만에 많은 변화를 이룬 셈이다. 게시글을 이용해 텍스트로 불법 스포츠 도박을 광고하던 시기부터 꾸준히 발전해 종내에는 영상을 이용해 불법 스포츠 도박을 광고하는 방법까지 고안한 것이다. 또, 음란 유해사이트 역시 못지않게 큰 문제였기 때문에 대부분의 연구자들이 겉으로 잘 드러나지 않는 불법 스포츠 도박보다는 음란 유해사이트에 초점을 맞추어 연구를 한 것으로 보인다.

이뿐 아니라 차단하는 방식에도 한계가 있었다. 과거에는 불법 스포츠 도박을 홍보하기 위해 웹 사이트를 이용하거나 포털사이트의 게시글을 이용하였다. 그래서 기존의 연구는 웹 사이트를 차단하는 방법이나 단어빈도를 중심으로 한 가중치 기법을 이용해 차단하는 방법을 이용하였다. 그러나 불법 스포츠 도박을 홍보하는 방식도 발전하여 영상 속에 문구를 삽입해 홍보를 하는 방식이나 움직이는 동영상 속에 문구를 삽입해 홍보를 하는 방식에 대해서는 대안이 미비한 것이 사실이다.

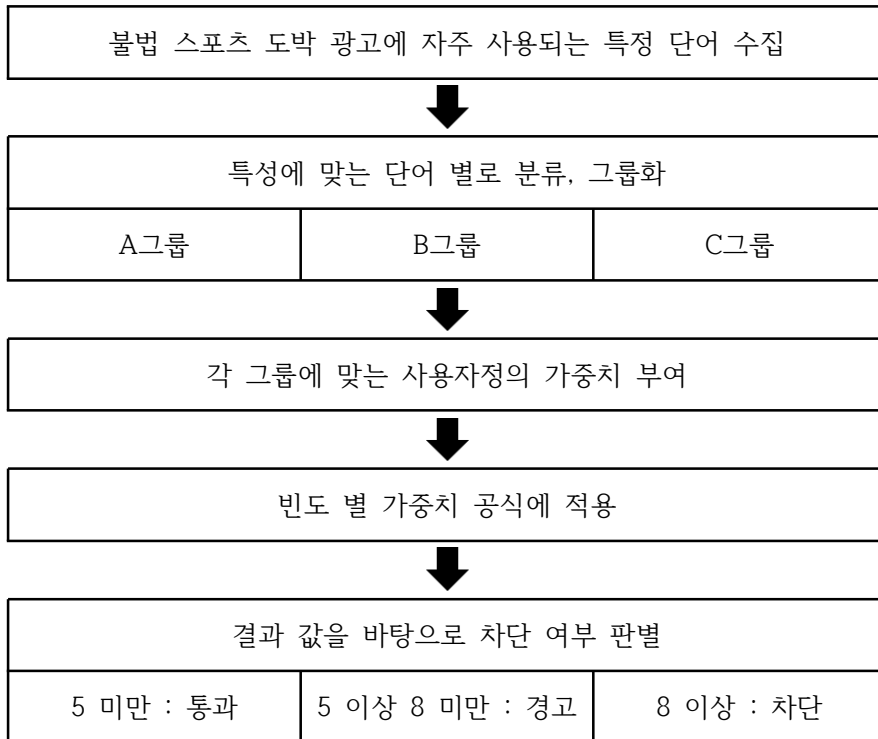
또한 웹 사이트에서 애플리케이션으로, 게시글에서 SNS로 매체가 이동된 것도 어려움을 겪는 데에 한 몫 하고 있는 것으로 보인다.

즉 최근에는 SNS를 통해서 불법 스포츠 도박을 광고하는 횡수가 빈번해지고 있다는 것이다. 특히 SNS에서 자주 이용되는 홍보 방법은 이미지 속에 홍보하고자 하는 문구를 삽입하여 이미지화하여 홍보를 하는 것이다. 게다가 모든 정보를 한 눈에 확인할 수 있는 타임라인에서 바로 확인할 수 있도록 게시글에 홍보 이미지를 게재하는 것이 아니라 인기 있는 게시글의 코멘트에 홍보 이미지를 게재하는 교묘한 방법을 주로 사용하고 있다. 따라서 쉽게 눈에 띄지 않아 감시가 어려워지고, 특정 단어를 이용해 필터링을 하는 것도 힘들어진다. 그러나 앞서 문제 삼은 불법 스포츠 도박을 홍보하는 이미지를 자체적으로 차단하는 방법은 아직 제시되지 않고 있다. 따라서 본 논문에서는 불법 스포츠 도박에 쉽게 접근하게 하는 문구가 삽입된 이미지를 사전에 처리하는 빈도 별 가중치 공식을 이용한 불법 스포츠 도박 광고 차단 방법을 제안하고자 한다.

본 논문에서 제시하고자 하는 방법은 빈도 별 가중치 공식을 적용한 SNS 속 불법 스포츠 도박 광고 차단 방법이다. 페이스북에 게재된 불법 스포츠 도박 광고 이미지를 중심으로 하여 불법 스포츠 도박 관계자들이 자주 이용하는 단어를 파악, 수집하고, 수집된 단어들을 각 단어의 특성과 유형에 맞게 A그룹, B그룹, C그룹으로 분류한다. 그리고 해당 그룹에 맞게 경험적인 방법으로 산출해 낸 사용자정의 가중치를 각각 부여한 뒤, 빈도 별 가중치 공식은 어떤 방식을 취하는지에 대해서도 기술한다. 마지막으로 어떤 기준으로 차단 여부를 판별할 수 있는지에 대해 근거를 제시한다.

본 장에서는 불법 스포츠 도박 광고에 자주 사용되는 특정 단어의 종류와 그 기준에 대해 나누어보고 해당 단어의 특성에 맞는 단어 별로 분류를 하는 과정에 대

해 상세하게 다룰 예정이다. 그리고 A등급, B등급, C등급에 맞추어 경험으로 획득한 경험적 사용자정의 가중치에 부여한 뒤 TF 데이터베이스를 생성, 결과 값을 바탕으로 차단 여부를 판별하는 기준과 과정에 대해 살펴본다.



<그림 3> 불법 스포츠 도박 차단 모형

1. 이미지화 된 단어 분석

단어를 분석하기에 앞서 불법 스포츠 도박 광고 방법에 대해 파악할 필요가 있

다. 불법 스포츠 도박 광고 관계자들은 지금껏 꾸준히 광고 방식을 발전시켜 왔다. 이때 유의할 것은 광고 방법이 바뀌고 홍보하는 도박의 종류가 바뀌면 또 다른 광고가 탄생한다는 것이다. 그러므로 예전에 홍보했던 방식으로 지금은 홍보하지 않고, 지금 홍보하는 방식으로 예전에는 홍보하지 않았던 것이다.

현재 SNS에서 이루어지는 대부분의 불법 스포츠 도박 광고는 영상물로 이루어져 있다. 불법 스포츠 도박 광고 관계자들은 특정 단어가 필터링 되는 것과 같은 감시망을 피하기 위해 이미지 속에 전하고 싶은 말을 담아서 페이스북 사용자들에게 홍보를 하고 있었다. 주로 자극적인 글이나 재미있는 사진 등 사용자들의 관심을 끌 수 있는 글을 찾고, 해당 글에 불법 스포츠 도박 광고 영상을 코멘트로 올리는 방식을 채택했다. 즉, 사용자들은 자신들의 타임라인을 통해 친구나 가족, 지인이 반응한 글을 확인할 수 있는데, 이때 코멘트에 있는 불법 스포츠 도박 광고를 보게 될 수 있다.

또, 코멘트는 직접 확인을 하지 않는 이상 매 순간마다 노출이 되어있는 것이 아니므로 위와 같은 방법으로 자연스럽게 광고를 하면서도 감시자들의 눈을 피할 수 있다. 이는 불법 스포츠 도박 광고업자들이 페이스북의 장점이자 단점인 쉬운 접근성을 파악하고 이를 악용하는 것으로 판단된다.

불법 스포츠 도박 광고 이미지를 만들 때 대부분은 쉽게 인식하기 어려운 사진을 사용한다. 가끔 단조로운 배경에 문구를 삽입해 깔끔한 이미지를 심어주는 경우도 있지만 주로 연예인들의 화보 사진에 문구를 입히는 경우가 많으며 자극적인 여성의 나체 사진을 이용하는 경우도 적지 않다. 그들이 이미지 속에 삽입하는 문구는 가족방 홍보, 진행되고 있는 게임의 종류, 진행 중인 이벤트 등이었다.

불법 스포츠 도박을 운영하거나 이용하는 관계자들은 비관계자는 이해하기 어려운 은어를 사용하기도 한다. 이는 해당 이미지가 불법 스포츠 도박 광고인지 아닌지를 구분하여 차단 여부를 결정하는 데에 중요한 근거가 될 수 있다. 또, 그들이 사용하는 단어가 일상생활에서 쓰는 표현과 또 다른 중의적인 의미를 지니는 경우

나 일상생활과 같은 의미로 사용하는 단어들도 존재한다.

그래서 본 논문에서는 불법 스포츠 도박 광고를 판가름하는 틀을 마련하기 위해 불법 스포츠 도박 광고 시 가장 빈번히 사용되는 몇 가지 단어를 선정했다.

불법 스포츠 도박 광고물에서 자주 언급되는 문구는 ‘카톡’, ‘문의’, ‘이벤트’, ‘사다리’, ‘충전’, ‘가족방’, ‘com’, ‘스포츠’, ‘루틴’, ‘마틴’, ‘첫충’, ‘가입’, ‘단폴’, ‘달팽이’, ‘다리다리’, ‘매충’, ‘과워볼’, ‘배팅’, ‘놀이터’, ‘포인트’, ‘돌발’, ‘메이저’, ‘꽂머니’, ‘슈어매니’, ‘공원’, ‘다음드’, ‘해외안전’, ‘상담’, ‘적립’ 등이 있었다.

자주 언급되는 단어들을 분석한 결과 ‘카톡’처럼 불법 스포츠 도박과 직접적인 연관이 없을뿐더러 일상생활에서 자주 쓰이는 단어들이 있는가 하면 ‘이벤트’처럼 사행성을 조장하는 데 쓰일 수 있는, 그리고 일상생활에서 사용하는 단어들, 혹은 일상생활에서 쓰이지만 불법 스포츠 도박에서 사용하는 단어의 의미와는 전혀 다른 의미를 가진 ‘사다리’ 같은 단어들. 마지막으로 불법 스포츠 도박 관계자와 이용자들만 알 수 있는 단어로 일상생활에서 전혀 쓰이지 않는 ‘가족방’ 같은 단어들로 구분할 수 있었다.

단어를 수집한 뒤에는 단어를 각 특성에 맞게 분류하는 작업이 필요하다. 단어마다 사용되는 빈도 및 중요도가 상이하기 때문에 모두 같은 기준을 적용할 수 없다고 판단하였다. 예를 들어 ‘가족방’ 같은 특정 단어는 불법 스포츠 도박 관계자들만 사용하는 단어이기 때문에 이와 같은 단어가 포함되면 불법 스포츠 도박 판정을 해야 한다. 또, ‘com’ 같은 단어는 불법 스포츠 도박 관계자 외의 사람들도 같은 의미로 사용하는 단어이기 때문에 이 단어를 가지고 불법 스포츠 도박 광고라고 판단할 수 없다. 따라서 본 논문에서는 단어의 중요도와 특성에 맞게 A, B, C 총 세 그룹으로 분류하였다.

(a) A그룹

A 그룹은 중요도가 가장 높은 단어를 분류한 그룹으로 지정하였다. 즉, 불법 스포츠 도박 광고에서 자주 쓰이는 단어 중 관련도가 가장 높은 단어인 관계자 간에 이해할 수 있는 은어들을 해당 그룹으로 분류하였다. A 그룹으로 분류된 단어는 ‘가족방’, ‘루틴’, ‘마틴’, ‘첫충’, ‘단폴’, ‘다리다리’, ‘매충’, ‘파워볼’, ‘퐁머니’, ‘슈어맨’, ‘다음드’ 등 총 11개의 단어이다.

해당 단어들은 불법 스포츠 도박 광고에서만 쓰이는 용어들로 일상생활에서는 사용하지 않는 단어들이다. 대부분의 불법 스포츠 도박 관계자들이 A 그룹의 용어들을 이용하고 있었고, 그 중에서도 자주 사용되는 단어들을 분류하였다. 따라서 불법 스포츠 도박 광고로 추정되는 이미지에 해당 단어가 등장하면 불법 스포츠 도박 광고로 간주한다.

(b) B 그룹

B 그룹은 A 그룹으로 분류되지 않은 나머지 단어들 중에 일상생활에서 사용하는 단어를 중의적인 표현으로 사용하는 단어나 의미는 같지만 불법 스포츠 도박과 밀접하게 관련이 있는 단어들을 중심으로 분류하였다. B 그룹으로 분류된 단어는 ‘이벤트’, ‘사다리’, ‘충전’, ‘스포츠’, ‘달팽이’, ‘배팅’, ‘놀이터’, ‘포인트’, ‘돌발’, ‘메이저’, ‘공원’, ‘지급’ 등 총 12개의 단어이다.

해당 그룹으로 분류된 단어들은 각 특징에 따라 두 가지로 나뉜다. 첫 번째는 일상생활에서도 사용되는 용어들을 가져다가 다른 의미로 사용하고 있는 경우이고,

두 번째는 혹은 일상생활에서 사용되는 용어와 뜻이 같지만 밀접한 관련이 있는 경우이다.

전자의 경우에는 ‘사다리’, ‘스포츠’, ‘달팽이’, ‘놀이터’, ‘돌발’, ‘공원’ 등이 있다. 후자의 경우는 ‘충전’, ‘배팅’, ‘포인트’, ‘메이저’, ‘지급’ 등이 있다. 특히 후자의 경우 A 그룹과의 다른 점은 A 그룹은 밀접한 관련이 있으며 일상생활에서는 사용하지 않는 단어들을 분류한 그룹이고, B 그룹은 밀접한 관련이 있는 단어인 것은 같지만 일상생활에서도 사용하는 단어들을 분류한 그룹이라는 데에 차이가 있다.

(c) C 그룹

C 그룹은 A 그룹과 B 그룹처럼 불법 스포츠 도박과 밀접한 관련이 있는 단어는 아니지만 자주 쓰이는 단어들을 중심으로 분류한 그룹이다. C 그룹으로 분류된 단어는 ‘카톡’, ‘문의’, ‘com’, ‘가입’, ‘상담’ 등 총 5개의 단어이다.

해당 그룹으로 분류된 단어들은 특히 일상생활에서 같은 뜻으로 사용되는 단어들이지만 불법 스포츠 도박 광고에 자주 등장하는 단어들이다. C 그룹의 단어들은 일상생활에서 사용할 때의 의미와 불법 스포츠 도박에서 사용할 때의 의미가 같다. 즉, B 그룹과의 차이점은 B 그룹은 일상생활에서 사용되는 용어지만 불법 스포츠 도박에서 사용할 때는 다른 의미를 사용하는 단어들이고, C 그룹은 일상생활과 불법 스포츠 도박에서 사용할 때의 의미가 모두 같은 단어들을 분류한 것이다.

그룹	단어					
A 그룹	가족방	루틴	마틴	첫충	단폴	다리다리
	매충	파워볼	꽁머니	슈어맨	다음드	X
B 그룹	이벤트	사다리	스포츠	충전	달팽이	배팅
	지급	포인트	놀이터	돌발	메이저	공원
C 그룹	카톡	문의	가입	com	상담	X

<표 2> 그룹 별 단어 분류

2. 사용자정의 가중치 및 빈도 별 가중치

(1) 사용자정의 가중치

앞서 분류한 A, B, C 세 종류의 그룹은 각각의 특성이 모두 다르다. A 그룹은 불법 스포츠 도박을 할 때 자주 사용하는 단어들을 분류한 그룹이고, B 그룹은 일상생활에서 쓰이는 뜻과 다르게 사용하고 있는 불법 스포츠 도박과 관련이 있는 단어들을 분류한 그룹이다. 여기에는 일상생활에서도 사행성과 관련이 짙은 단어들 이 포함되어 있다. 그리고 C 그룹은 불법 스포츠 도박보다는 일상생활에서 더 자주 쓰이는 단어지만 불법 스포츠 도박 광고에 자주 등장하는 단어들을 분류해 놓은 그룹이다.

그렇기 때문에 각 그룹마다 연관성과 중요도가 다를 수밖에 없다. A그룹은 불법 스포츠 도박과 가장 밀접한 관련이 있는 은어들을 분류한 그룹이기 때문에 중요도가 가장 높다. B 그룹은 중요도가 가장 높은 A 그룹을 제외한 나머지 단어 중 불

법 스포츠 도박과 관련되어 있으나 일상생활에서도 사용 가능한 단어를 분류한 그룹이기 때문에 중요도는 A 그룹보다 낮다. 그리고 C 그룹은 불법 스포츠 도박과 관련되지 않지만 자주 등장하는 단어를 분류한 그룹이기 때문에 그 중요도는 A 그룹이나 B 그룹보다 낮다. 그렇기 때문에 모든 그룹에 동일한 가중치를 부여할 수 없다.

또, 중요도도 중요하지만 그 중요도에 비해 단어가 사용된 빈도가 그룹마다 다르기 때문에 사용자정의 가중치가 필요하다. 가장 많이 등장하는 단어는 중요도가 가장 높은 A 그룹이 아닌 중요도가 가장 낮은 C 그룹이고, 반대로 그룹 중 가장 적게 등장하는 단어는 중요도가 가장 높은 A 그룹이기 때문이다. 따라서 그룹마다 다른 중요도와 빈도수의 격차를 보완하기 위해 사용자정의 가중치를 다르게 부여하였다.

본 실험에 앞서 사전 실험을 진행하였다. 이 실험은 불법 스포츠 도박 광고를 차단하기 위해 필요한 것들을 추출하기 위해 진행된 실험이다.

100개의 이미지를 추출하여 기존에 진행하고자 하는 실험과 동일한 방법으로 진행하였다. 불법 스포츠 도박 광고 이미지를 찾아내어 빈번히 등장하는 단어들을 빈도수에 따라 중요도에 따라 정리하였다. 그리고 빈도 별 가중치 공식에 임의로 정한 사용자정의 가중치를 대입하여 가장 이상적인 사용자정의 가중치를 찾아내었다. 여러 번의 시행착오 끝에 가장 적절한 사용자정의 가중치를 경험적인 방법으로 산출해냈다.

다시 말하여 본 논문에서는 중요도를 중심으로 하여 경험적인 방법을 통해 사용자정의 가중치 α 를 부여하였다. 가장 중요도가 높은 그룹인 A 그룹은 $\alpha=50$, 다음으로 중요도가 높은 그룹인 B 그룹은 $\alpha=40$, 마지막으로 가장 중요도가 낮은 그룹인 C 그룹은 $\alpha=10$ 이다.

이렇게 중요도에 따라 그룹 별로 부여한 α 는 빈도 별 가중치 공식에 적용된다.

(2) 빈도 별 가중치

그룹별로 사용자정의 가중치를 경험적인 방법으로 구해 지정한 다음에는 빈도 별 가중치 공식을 이용해 각 단어마다 고유의 TF 값을 산출한다. 단어마다 지정된 이 값은 후에 불법 스포츠 도박인지 아닌지 판단할 수 있는 중요한 단서로 쓰인다. 사용된 단어를 모두 종합해서 해당 단어의 TF 값을 도합하고 모두 합한 결과 값을 바탕으로 불법 스포츠 도박인지 아닌지를 구분할 수 있게 한다. 본 논문에서 제안하는 빈도 별 가중치 공식은 다음과 같다.

$$TF = \alpha \times \frac{tf}{\max tf}$$

공식에서 α 는 사용자정의 가중치, tf 는 단어 빈도수, $\max tf$ 는 그룹 내 단어의 전체 빈도수이다. 이 빈도 별 가중치 공식은 기존의 가중치 기법 중 하나인 보정 TF 방법을 일부 변형한 것이다. 빈도수의 편차가 큰 점을 들어서 그룹 별 사용자정의 가중치에 맞게 수정한 공식이다. 기존의 보정 TF 공식은 다음과 같다.

$$TF = (1 - w) + w \times \frac{tf}{\max tf}$$

불법 스포츠 도박 광고인지 아닌지를 판별하기 위해 빈도 별 가중치 공식을 이용해 TF 값을 계산한다.

중요한 것은 TF 값을 구할 때 각각의 단어에 그룹마다 다른 사용자정의 가중치를 적용해서 그 중요도와 빈도에 따라 값의 차이가 있도록 설정되어 있다는 점이

다. 각 단어가 같은 사용빈도를 가진다고 해도 해당 그룹의 사용자정의 가중치에 따라 값이 달라질 수 있다.

예를 들어서 A 그룹의 ‘가족방’과 C 그룹의 ‘카톡’을 비교하겠다. ‘가족방’의 TF 값을 TF_1 이라 하고 ‘카톡’의 TF 값을 TF_2 라 할 때, ‘가족방’과 ‘카톡’이라는 단어가 모두 10의 tf 를 가지고 100의 $\max tf$ 를 갖는다고 하자. 일상생활에서 쓰이지 않고 불법 스포츠 도박 내에서만 사용하는 단어인 ‘가족방’은 A 그룹이므로 ‘가족방’의 α 는 50이다. 또, 일상생활에서 자주 쓰이고 불법 스포츠 도박과는 관련도가 낮지만 광고 이미지에 자주 쓰이는 단어인 ‘카톡’은 C 그룹이므로 ‘카톡’의 α 는 10이다. 이 때 각각의 TF 값은 $TF_1 = 5$, $TF_2 = 1$ 이 되는 것이다. 즉, 같은 빈도수를 가지고 있다고 하더라도 해당 단어와 불법 스포츠 도박의 관련도에 따라 TF 값은 다르게 산출될 수 있다.

또 다른 예를 들어보겠다. 이번에는 B 그룹의 ‘이벤트’와 C 그룹의 ‘문의’를 비교하겠다. ‘이벤트’의 TF 값을 TF_3 이라 하고 ‘문의’의 TF 값을 TF_4 라 할 때, ‘이벤트’의 tf 는 10, ‘문의’의 tf 는 40, 또 각각의 $\max tf$ 는 모두 100이라고 정의하였다. 이 때도 마찬가지로 각각의 중요도에 따라 다른 α 가 지정된다. 일상생활에서 사용하는 단어의 뜻과는 같지만 사행성이 짙은 불법 스포츠 도박에 자주 이용될 수 있는 단어인 ‘이벤트’는 B 그룹이므로 ‘이벤트’의 α 는 40이다. 또, 불법 스포츠 도박과 밀접한 관련은 없지만 홍보 이미지에 자주 쓰이는 단어인 ‘문의’는 C 그룹이므로 ‘문의’의 α 는 10이다. 이제 두 단어의 TF 를 구하면 $TF_3 = 4$, $TF_4 = 4$ 이다. 즉, 서로 다른 빈도수를 가지고 있음에도 불구하고 그 단어와 불법 스포츠 도박의 연관성에 따라서 같은 TF 를 갖거나 더 자주 등장하는 단어의 TF 가 낮을 수도 있다는 의미가 된다.

이 두 가지의 예를 통해 알 수 있는 점은 본 논문에서 제안하는 빈도 별 가중치 공식을 이용하여 값을 구할 때 단어의 빈도수보다 해당 그룹의 중요도가 더 크게 작용한다는 점이다. 가중치 기법으로 불법 스포츠 도박을 차단할 수 있는 방법에

대해 연구하면서 가중치를 적용할 수 있는 부분 중 단어의 빈도보다는 단어의 특성에 맞게 값을 다르게 해야 한다는 사실을 알게 되었다. 또, 실제로 100개의 이미지를 가지고 사전 실험을 할 때 빈도수를 조사한 바에 따르면 관련성이 더 높은 그룹 단어들의 빈도수 평균이 관련성이 낮은 C 그룹 단어들의 빈도수 평균보다 낮았다.

그룹	단어 빈도수						평균
A 그룹	가족방	루틴	마틴	첫충	단폴	다리다리	31.45
	48	45	43	40	37	33	
B 그룹	매충	파워볼	콩머니	슈어맨	다음드		32
	31	31	16	14	8		
C 그룹	이벤트	사다리	스포츠	총전	달팽이	배팅	45.8
	58	57	48	47	35	26	
C 그룹	지급	포인트	놀이터	돌발	메이저	공원	45.8
	25	21	20	20	17	10	
C 그룹	카톡	문의	가입	com	상담		45.8
	64	57	52	51	5		

<표 3> 단어 빈도수와 그룹 별 빈도수 평균

즉 단순히 단어의 빈도에 초점을 맞춘다기보다 해당 단어가 불법 스포츠 도박과 얼마나 관련이 있는지, 또 얼마나 중요한지에 대해 초점을 맞추고 관련성이나 중요도가 높은 단어가 등장하면 불법 스포츠 도박 광고일 확률이 월등히 높아진다는 점을 이용하였다. 그러므로 본 논문에서는 가중치 기법을 적용할 때 단어의 빈도보다는 사용자정의 가중치인 α 가 더 중요하다고 주장하는 바이다.

빈도 별 가중치 공식에 적용하는 방법은 다음과 같다. 먼저 해당 단어의 빈도수

를 그룹 내 단어의 전체 빈도수로 나눈다. 그룹마다 빈도수의 평균이 다르기 때문에 그룹별로 다르게 설정해주어야 한다. 다음으로 설정한 사용자정의 가중치를 곱해 주는데 그렇게 하면 한 단어에 대한 TF 값이 나온다. 본 논문에서 제시하고 있는 단어들을 모두 빈도 별 가중치 공식에 대입하면 하나의 TF 데이터베이스가 완성되는 것이다.

이렇게 생성된 TF 데이터베이스는 해당 이미지가 불법 스포츠 도박인지 아닌지를 판가름할 수 있는 중요한 정보로 사용된다. 본 논문을 제안하며 저장된 TF 데이터베이스에는 A 그룹 11개의 단어, B 그룹 12개의 단어, C 그룹 5개의 단어를 포함해 총 22개의 단어가 존재한다.

그룹	TF 데이터베이스					
A 그룹	가족방	루틴	마틴	첫충	단폴	다리다리
	6.94	6.50	6.21	5.78	5.35	4.77
	매충	파워볼	콩머니	슈어맨	다음드	/
	4.48	4.48	2.31	2.02	1.16	
B 그룹	이벤트	사다리	스포츠	총전	달팽이	배팅
	6.04	5.94	5.00	4.90	3.65	2.71
	지급	포인트	놀이터	돌발	메이저	공원
	2.60	2.19	2.08	2.08	1.77	1.04
C 그룹	카톡	문의	가입	com	상담	/
	2.79	2.49	2.27	2.23	0.22	

<표 4> TF 데이터베이스

A 그룹으로 분류된 단어들의 TF 값은 각각 ‘가족방’ : 6.94, ‘루틴’ : 6.50, ‘마틴’

: 6.21, ‘첫충’ : 5.78, ‘단폴’ : 5.35, ‘다리다리’ : 4.77, ‘매충’ : 4.48, ‘과워볼’ : 4.48, ‘꽂머니’ : 2.31, ‘슈어맨’ : 2.02, ‘다음드’ : 1.16이며 가장 높은 *TF* 값을 가진 단어는 ‘가족방’이다.

B 그룹으로 분류된 단어들의 *TF* 값은 각각 ‘이벤트’ : 6.04, ‘사다리’ : 5.94, ‘스포츠’ : 5.00, ‘충전’ : 4.90, ‘달팽이’ : 3.65, ‘배팅’ : 2.71, ‘지급’ : 2.60, ‘포인트’ : 2.19, ‘놀이터’ : 2.08, ‘돌발’ : 2.08, ‘메이저’ : 1.77, ‘공원’ : 1.04이며 가장 높은 *TF* 값을 가진 단어는 ‘이벤트’이다.

C 그룹으로 분류된 단어들의 *TF* 값은 각각 ‘카톡’ : 2.79, ‘문의’ : 2.49, ‘가입’ : 2.27, ‘com’ : 2.23, ‘상담’ : 0.22이며 가장 높은 *TF* 값을 가진 단어는 ‘카톡’이다.

3. 차단 여부 판정

각각의 단어마다 *TF* 값을 산출하고 나면 생성된 *TF* 데이터베이스를 바탕으로 해당 이미지가 불법 스포츠 도박 광고인지 아닌지를 판정하게 된다. 이미지화 된 문구를 분석하여 *TF* 값이 존재하는 단어를 골라낸 다음 그 값들을 모두 합산해서 이미지의 결과 값이 얼마인지를 알아낸다. 이때 중복되는 단어는 중복해서 더하지 않고 하나로 계산한다.

TF 값의 합계를 구하고 나면 판정 기준을 바탕으로 불법 스포츠 도박 광고인지 아닌지를 결정한다. 판정 기준은 세 가지로 분류된다. 중요도가 높은 단어가 다수 존재해서 불법 스포츠 도박 광고라고 확정이 되었을 때는 차단 판정을 한다. 특정 문구가 많이 나타났지만 자주 사용되는 단어가 부족해서 불법 스포츠 도박 광고라고 확정 짓기 어려운 경우에는 경고 판정을 한다. 그리고 *TF* 값이 존재하는 단어가 아예 등장하지 않았거나 중요도가 낮고 빈도수가 적은 단어가 한 개 혹은 두 개가 등장했을 때는 통과 판정을 하도록 하였다.

이 때 판정을 하는 기준은 사전 실험의 결과를 바탕으로 책정하였다. 99개의 불법 스포츠 도박 광고 이미지와 1개의 일반 광고 이미지, 즉 100개의 광고 이미지를 가지고 본 실험의 과정과 같은 방법으로 사전 실험을 한 결과, 판정 범위를 5 미만, 5 이상 8 미만, 8 이상으로 했을 때 가장 이상적으로 차단할 수 있다는 것을 파악했다. 가장 관련성이 높은 A 그룹의 단어가 분석되었다고 했을 때, 아주 높은 확률로 불법 스포츠 도박과 관련이 있다는 점을 들어 범위를 지정하였다.

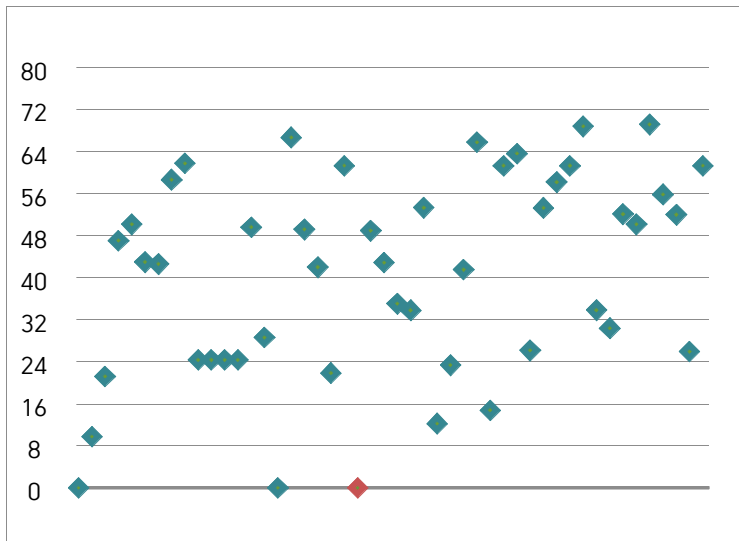
A 그룹의 평균 tf 가 31.45라고 했을 때 평균값을 바탕으로 한 TF 는 4.54이다. 또, B 그룹의 평균 tf 가 32라고 했을 때 평균값을 바탕으로 한 TF 는 3.33이며 C 그룹의 평균 tf 가 45.8이라고 했을 때 평균값을 바탕으로 한 TF 는 2이다. 이때 A 그룹의 단어들은 모두 일상생활에서 쓰이지 않고 불법 스포츠 도박 관계자들 사이에서만 사용되는 은어이기 때문에 하나의 단어라도 사용이 된다면 높은 확률로 불법 스포츠 도박 광고이다. 따라서 평균값의 TF 인 4.54를 최하 범위로 지정할 수도 있다. 그러나 다른 그룹 단어들 TF 값의 합계가 4.54를 충분히 넘을 수 있으므로 그를 올림 한 5를 통과 최댓값으로 기준을 설정하였다. 즉 결과 값이 5 미만일 시에는 불법 스포츠 도박 광고가 아니라고 판정하고 통과를 시키도록 하였다.

판단 여부의 조건이 되는 수치를 측정하기 위해 진행한 사전 실험에 사용한 이미지는 총 100개이고 그 중 99개는 불법 스포츠 도박 광고 이미지, 1개는 일반 광고 이미지이다. 99개의 불법 스포츠 도박 광고 이미지 중 3개의 이미지를 제외한 나머지 광고 이미지에 쓰인 2개 이상의 단어가 TF 데이터베이스에 포함되어 있는 단어였고, 99개 중 96개의 불법 스포츠 도박 광고 이미지가 결과 값이 8을 넘고 있었으므로 차단 최솟값으로 기준을 설정하였다. 즉 결과 값이 8 이상일 때에는 불법 스포츠 도박 광고가 맞다고 판정하고 이미지를 게재할 수 없도록 차단하도록 하였다.

또 결과 값이 5 이상 8 미만인 이미지들이 존재했는데 통과와 차단 사이의 값이므로 해당 이미지는 불법 스포츠 도박 광고인지 아닌지 사람의 눈으로만 판단할

수 있다고 생각, 이미지에 대해 경고 판정을 하고 해당 이미지를 게재할 때 경고 메시지를 출력하도록 하였다.

사전 실험 결과 99개의 불법 스포츠 도박 광고 이미지 중 96개의 불법 스포츠 도박 광고 이미지가 차단 판정을 받았고 1개의 일반 광고 이미지 중 1개의 일반 광고 이미지가 통과 판정을 받았다. 아래 차트는 본 논문에서 제안하고자 하는 빈도 별 가중치 공식으로 불법 스포츠 도박을 차단하는 방법으로 알아본 100개의 이미지의 결과 값이다. 빨간 점은 일반 광고 이미지이다.

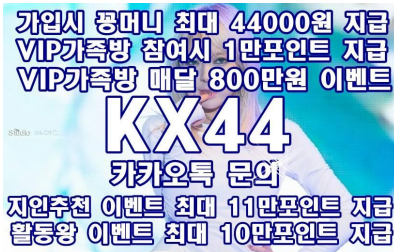


IV. 실험 및 결과

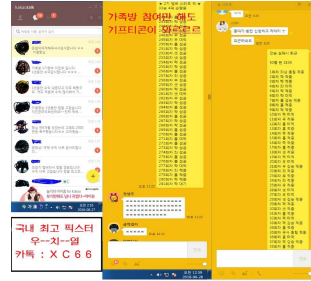
본 실험은 불법 스포츠 도박과 연관된 단어를 선정, 중요도 별로 분류하고 가중치 공식으로 TF 데이터베이스를 생성한 뒤 관련 이미지 200개를 이용하여 TF 합계를 연산하는 방법으로 진행하였다. 이때 불법 스포츠 도박 관련 단어는 사전 실험 때에 얻은 TF 데이터베이스를 바탕으로 하였으며 TF 값 역시 같은 값으로 진행하였다. 가장 중요도가 높은 그룹인 A 그룹부터 그 다음으로 중요도가 높았던 B 그룹, 마지막으로 중요도가 가장 낮은 그룹인 C 그룹의 단어들을 그대로 사용하였다. 또한 본 실험에 사용된 이미지는 200개 모두가 불법 스포츠 도박 광고가 아니라 불법 스포츠 도박 관련 광고 161개와 관련 없는 일반 이미지 39개를 이용하여 실험을 진행하였다. 이 이미지들은 기존의 불법 스포츠 도박 광고가 가장 흔한 방법으로 널리 홍보되고 있는 코멘트에서 얻은 이미지들이다.

이미지들을 분석하여 본 논문에서 제안한 사용자정의 가중치를 그룹 별로 적용하였다. 여기서 그룹 별로 적용한 사용자정의 가중치는 A 그룹은 $\alpha = 50$, B 그룹은 $\alpha = 40$, C 그룹은 $\alpha = 10$ 이다. 그리고 빈도 별 가중치 공식에 적용하여 이용된 단어마다 TF 값을 구하고, 이 데이터들을 모아 TF 데이터베이스를 마련하였다.

또, 제안한 방법으로 생성된 TF 데이터베이스를 참고로 하여 불법 스포츠 도박 광고 이미지에서 사용된 단어들을 추출한 뒤 사용된 단어 TF 값의 합을 구했다. 이를 결과 값이라 하고, 결과 값이 8 이상이면 차단 판정을, 5 이상 8 미만이면 경고 판정을 했고, 5 미만일 때는 통과 판정을 하도록 했다. 이때 차단 판정을 받으면 해당 이미지를 게재할 수 없도록 차단을 하고, 경고 판정을 받으면 이미지를 게재할 때에 메시지를 출력하도록 한다. 그리고 통과 판정을 받으면 이미지를 게재할 수 있도록 한다.



<그림 5> 실험 이미지 i



<그림 6> 실험 이미지 ii



<그림 7> 실험 이미지 iii



<그림 8> 실험 이미지 iv



<그림 9> 실험 이미지 v



<그림 10> 실험 이미지 vi

위는 실험에 사용된 이미지들의 예시를 일부 나열한 것이다. $i \sim iv$ 는 불법 스포츠 도박 광고 이미지이고 v 는 일반 광고 이미지, vi 은 광고가 아닌 일반 이미지이다.

i 는 불법 스포츠 도박 광고 이미지이다. 이미지를 분석하면 TF 데이터베이스에 포함된 단어들을 추출할 수 있다. 이미지와 TF 데이터베이스를 비교분석한 결과 ‘꽂머니’, ‘지급’, ‘가족방’, ‘포인트’, ‘이벤트’, ‘문의’라는 단어를 사용한 것을 확인했다. 본 논문에서 제안한 빈도 별 가중치 공식을 바탕으로 연산한 각 단어의 TF 값과 합산한 결과 값은 다음과 같다.

‘꽂머니’, ‘지급’, ‘가족방’, ‘포인트’, ‘이벤트’, ‘문의’의 TF 를 각각 TF_1 , TF_2 , TF_3 , TF_4 , TF_5 , TF_6 이라고 할 때 각각의 값은 $TF_1 = 2.31$, $TF_2 = 2.60$, $TF_3 = 6.94$, $TF_4 = 2.19$, $TF_5 = 6.04$, $TF_6 = 2.49$ 이다. 그러므로 모두 합산한 결과 값은 22.67이다. 결과 값을 바탕으로 차단 여부를 판정하면 차단 최소 값인 8을 넘었으므로 해당 이미지는 차단 판정을 받고, 이미지는 게재를 할 수 없도록 차단한다.

ii 역시 불법 스포츠 도박 광고 이미지이다. 해당 이미지에 사용된 단어를 분석한 결과 ‘가족방’, ‘카톡’의 두 단어가 TF 데이터베이스에 있는 단어와 일치하는 것을 확인했다. 마찬가지로 빈도 별 가중치 공식을 바탕으로 연산한 각 단어의 TF 값과 합산한 결과 값은 다음과 같다.

‘가족방’, ‘카톡’의 TF 를 TF_1 , TF_2 라고 할 때 각각의 값은 $TF_1 = 6.94$, $TF_2 = 2.79$ 이다. 그러므로 해당 이미지의 결과 값은 9.73이다. 즉, 결과 값을 바탕으로 하여 차단 여부를 판정하면 차단 최소값인 8을 넘었으므로 해당 이미지 역시 차단 판정을 받고, 이미지는 게재를 할 수 없도록 차단하게 된다.

iii은 불법 스포츠 도박 광고 이미지이다. 이미지와 TF 데이터베이스를 비교분석

하게 되면 ‘카톡’과 ‘com’을 사용하고 있는 것을 확인할 수 있다. 빈도 별 가중치 공식을 바탕으로 연산한 각 단어의 TF 값과 합산한 결과 값은 다음과 같다.

‘카톡’과 ‘com’의 TF 를 각각 TF_1 , TF_2 라고 할 때 값은 $TF_1 = 2.79$, $TF_2 = 2.23$ 이다. 그러므로 해당 이미지의 결과 값은 5.02이다. 이때 차단 최솟값인 8보다는 작으며 통과 최댓값인 5보다는 크다는 것을 알 수 있다. 즉 해당 이미지의 결과 값은 5 이상 8 미만이므로 경고 판정을 받았다. 경고 판정을 받은 이미지는 게재할 때 경고 메시지가 출력된다. 육안으로 보기에 해당 이미지는 불법 스포츠 도박 광고가 맞으나 사용된 단어가 부족해서 결과 값이 비교적 낮은 것으로 보인다. 그러나 불법 스포츠 도박 광고가 아니라고 통과 판정을 한 것이 아니고, 의심되는 이미지라고 경고 메시지를 출력하므로 실험을 성공한 것이 맞다.

iv는 역시 불법 스포츠 도박 광고 이미지이다. 광고에 사용된 단어와 TF 데이터 베이스를 비교분석한 결과 해당 이미지에는 ‘배팅’과 ‘com’이 사용되었다. ‘배팅’과 ‘com’의 TF 를 각각 TF_1 , TF_2 라고 하였을 때, 이들의 값은 $TF_1 = 2.71$, $TF_2 = 2.23$ 이다. 이들을 합한 결과 값은 4.94이다. 즉, 해당 이미지의 결과 값은 4.94이다. 이미지의 차단 여부를 판정하게 되면 통과 최댓값이 5 미만이므로 해당 이미지는 통과 판정을 받는다.

불법 스포츠 도박 광고 이미지인 것을 육안으로 확인했으나 사용된 단어의 부족과 사용빈도가 낮은 단어의 사용으로 해당 이미지는 차단되지 못하고 통과 판정을 받았다.

v는 불법 스포츠 도박 광고 이미지가 아닌 일반 광고 이미지이다. 마찬가지로 vi는 불법 스포츠 도박 광고 이미지가 아닌 일반 이미지이다. 해당 이미지들에 사용된 단어와 TF 데이터베이스를 비교분석했을 때 일치하는 단어가 없으므로 해당 이미지들의 결과 값은 0이다. 즉, 해당 이미지는 불법 스포츠 도박 광고 이미지가

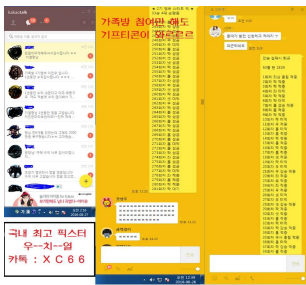
아니라고 판단하여 통과 판정을 받았다.



<그림 5> 실험 이미지 i

단어	콩머니	지금	가족방	합계
<i>TF</i>	2.31	2.60	6.94	
단어	포인트	이벤트	문의	22.67
<i>TF</i>	2.19	6.04	2.49	

<표 5> i의 결과 값



<그림 6> 실험 이미지 ii

단어	가족방	카톡	합계
<i>TF</i>	6.94	2.79	9.73

<표 6> ii의 결과 값



<그림 7> 실험 이미지 iii

단어	카톡	com	합계
<i>TF</i>	2.79	2.23	5.02

<표 7> iii의 결과 값

본 논문에서 제안한 방법을 토대로 200개의 관련 이미지를 통해 실험한 결과 불법 스포츠 도박 광고가 아닌 이미지 39개는 모두 통과 판정을 받았고 불법 스포츠 도박 광고 161개 중 153개의 이미지가 차단 판정을, 1개의 이미지가 경고 판정, 7개의 이미지가 통과 판정을 받았다. 불법 스포츠 도박 광고임에도 불구하고 관련된 어휘가 아예 등장하지 않았거나 적게 등장해서 차단 판정이나 경고 판정을 받지 못한 7개의 이미지를 제외한 나머지 193개의 이미지가 옳은 판정을 받았다. 즉 빈도 별 가중치 공식을 이용해 불법 스포츠 도박 광고를 차단한 실험은 96.5%의 성공률을 보였다.

V. 결론

인터넷의 발달로 사회는 수많은 정보를 접하며 살아간다. 그러나 그 역기능으로 인해서 또한 우리는 수많은 문제를 마주하고 살게 되었다. 이것은 고스란히 우리에게도 또 청소년들에게도 나아가서 미래의 청소년이 될 아이들에게도 옮겨간다. 최근에 문제가 되고 있는 것은 청소년들의 불법 스포츠 도박 관련 문제이다. 불법 스포츠 도박 광고는 SNS에서 많이 이루어지고 있었는데, 특히 페이스북 게시글의 코멘트에서 쉽게 발견할 수 있었다. 대부분이 감시망을 피하기 위해 이미지 속에 불법 스포츠 도박 광고임을 나타내는 문구들을 포함시켜서 교묘히 광고를 하고 있었고, 그로 인해 청소년들도 하루 몇 십 개 이상의 불법 스포츠 도박 광고 이미지를 접할 수 있게 되었다.

위와 같은 문제에 대해서 청소년의 불법 스포츠 도박 노출 정도에 대한 심각성과 기존의 소극적인 방식보다 더 강력하게 불법 스포츠 도박 광고를 차단할 필요성을 느꼈고, 다양한 방법을 생각하다가 빈도 별 가중치 공식을 이용해 더 적극적으로 차단할 수 있는 방법을 본 논문에서 제안하였다.

빈도 별 가중치 공식을 이용하기 전에 이미지화 되어 불법 스포츠 도박에 자주 사용되는 단어들을 선정하고 각각의 특성에 맞게 A, B, C 세 그룹으로 나누었다. 불법 스포츠 도박에만 쓰이는 은어를 A 그룹으로, 일상생활에서와 다른 의미로 쓰이는 불법 스포츠 도박 용어나 연관성이 깊은 단어를 B 그룹으로, 그리고 불법 스포츠 도박과 깊은 관련은 없지만 자주 쓰이는 단어를 C 그룹으로 분류하였다. 각 그룹마다 중요도가 다르기 때문에 그룹 별로 가중치 α 를 상이하게 설정하였다. 가장 중요도가 높은 A 그룹은 $\alpha=50$, 다음으로 중요도가 높은 B 그룹은 $\alpha=40$, 마지막으로 중요도가 비교적 낮은 C 그룹은 $\alpha=10$ 으로 정했다. 그리고 본 논문에서 제시

한 빈도 별 가중치 공식을 이용해 각 단어마다 TF 값을 산출해냈다. 그리고 이 TF 값을 모아놓은 데이터베이스를 바탕으로 불법 스포츠 도박 광고인지 아닌지를 판단하도록 하였다.

본 논문에서 제안한 방법으로 실험을 한 결과 200개의 실험 이미지 중 193개가 이미지의 상황에 맞게 차단, 경고, 통과 판정을 받았고 나머지 7개의 이미지가 불법 스포츠 도박 광고임에도 불구하고 통과 판정을 받았다. 즉, 본 실험을 통해 빈도 별 가중치 공식을 통해 불법 스포츠 도박 광고를 사전에 차단하는 방법은 96.5%의 성공률을 보인다는 것을 확인했다.

즉, 불법 스포츠 도박을 차단하기 위한 방법으로 불법 스포츠 도박 광고에 이용되는 이미지들을 분석하여 자주 사용되는 언어들을 분류하고 해당 단어들의 빈도와 각 단어의 중요도에 따라 빈도 별 가중치 공식을 이용해 TF 값을 계산한 뒤, TF 데이터베이스를 생성하여 데이터베이스에 저장된 단어와 수치를 바탕으로 불법 스포츠 도박 광고인지 아닌지에 분석하는 실험을 진행한 결과, 96.5%의 높은 성공률을 보였고 이는 빈도 별 가중치 공식을 이용해 불법 스포츠 도박을 차단하는 방법이 유의미함을 보였다.

특히, 불법 스포츠 도박을 직접적인 모니터링과 사용자들의 제보를 통해 차단하는 기존의 방식과는 다르게 본 논문에서 제안하는 방법을 이용하면 불법 스포츠 도박 광고를 사전에 차단할 수 있다. 단순히 텍스트를 필터링 하여 차단하는 기존의 방법과도 차이가 있다. 최근에 자주 행해지고 있는 광고 방법에 맞추어 이미지화 된 단어를 분석하는 방법을 제안하였기 때문에 그 활용도가 더 높을 것으로 예상된다.

단, 향후에 차단되지 못한 3.5%의 불법 스포츠 도박 광고를 차단시키기 위한 연구가 필요하다고 여겨진다. 또한 이미지를 분석할 수 있는 구체적인 방법이 필요하다고 생각한다. 그리고 아직도 불법 스포츠 도박 광고는 끊임없이 발전하고 있다. 최근에는 동영상을 통해 불법 스포츠 도박 광고를 하는 모습을 확인하였기에 그에

대한 빠른 대처 방안도 제시해야 한다. 마지막으로 최근에는 ‘바다이야기’라는 불법 스포츠 도박 광고가 떠오르고 있는데, 그에 맞추어서 TF 데이터베이스를 늘려가는 것도 필요할 것이다.

본 논문에서 제안하는 빈도 별 가중치 공식을 이용한 불법 스포츠 도박 광고 차단 방법을 통해서 불법 스포츠 도박을 광고하는 것을 사전에 차단할 수 있을 것으로 예상된다. 기존의 이미 홍보되고 있는 불법 스포츠 도박 광고를 사후에 처리하는 방법보다 사전에 차단하기 때문에 더 강력하게 차단할 수 있을 것으로 예상된다. 만약 본 논문에서 제안하는 방법과 기존의 방법을 병행하여 사용한다면 불법 스포츠 도박 광고를 차단하는 더 큰 효과를 가져 올 것으로 예상된다.

참고문헌

- [1] 고진주.김민재.신창석 (2015)청소년 도박문제 실태조사, 한국도박문제관리센터, 2015
- [2] 권미수.남길우.서보경. 2015년 인터넷 과의존 실태조사, 한국정보화진흥원, 2015
- [3] 윤태현. 특성화 고등학교 학생의 스마트폰에 대한 인식:성별, 학업성취별 분석, 인하대학교 교육대학원, 2015
- [4] 김인수. 체육진흥투표권(sportstoto) 사업의 구조개선과 불법스포츠도박 및 승부조작 근절에 관한 연구 : IPA Matrix분석으로, 2017
- [5] 박준휘.최성락.오시영. 스포츠 토토 불법 행위에 대한 처벌 강화 정책의 효과 분석, 한국사회와 행정연구, 2014, pp.35-57
- [6] 김혁. 불법 스포츠평토토 실태와 개선방안탐색, 한국체육과학회지, 2015, pp.81-98
- [7] 김나혜.박성언.전정우. 대학생들의 불법 스포츠도박 지속행동에 대한 탐색, 한국체육학회지, 2015, pp95-105
- [8] 김서연. 남자 중·고등학생의 불법 스포츠도박 실태 조사, 단국대학교 교육대학원, 2015
- [9] 김도우.박경래.이창무. 불법 사행산업의 실태 및 대응전략에 관한 연구, 한국공안행정학회, 2012, pp10-42
- [10] 윤우석. 불법온라인 도박의 운영방식과 현황파악에 관한 연구, 한국공안행정학회보, 2016 pp.139-165
- [11] 서동우. 대학생의 불법스포츠도박 경험과 도덕적 의사결정태도, 강원대학교 교육대학원, 2016
- [12] 권복순.김영호. 한국 대학생의 도박참여 실태와 도박중독 유병률 조사, 정신보건과 사회사업, 2011, pp.5-28
- [13] 권선중.김교현.성한기.이민규.강성균. 성인과 남자 청소년의 불법 인터넷 도박 문제와 위험 요인 그리고 예방 전략, 2007, pp1-19

- [14] 한승완.김세민.정병호.노용만. 내용 기반 유해 멀티미디어 차단 기술 동향, 정보보호 학회지, 2009, pp.38-44
- [15] 한승완.최병철.임재덕. 유해 멀티미디어 분석 및 차단 기술 동향, 정보보호학회지, 2012, pp.19-24
- [16] 이은애.정명숙.김재건.하석운. 청소년을 위한 유해 웹영상 차단 시스템의 구현, 한국 멀티미디어학회 춘계학술발표논문집, 2000, pp.319-323
- [17] 정규철.이진관.이태헌.박기홍. 컴퓨터과학 : 문자 기반 유해사이트 판별 기법, 컴퓨터 교육학회논문지, 2004, pp.83-91
- [18] 최진우.한승완. 다중 유해 프레임 분류기를 이용한 유해 콘텐츠 차단 방법에 관한 연구, 한국통신학회 학술대회논문집, 2015, pp.1165-1166
- [19] 임석향.이원휘.안동연.정성종. 언어 모델을 이용한 유해 문서 분류, 한국정보과학회 한국컴퓨터종합학술대회 논문집, 2009, pp.301-304
- [20] 조동욱.김지영. 음란 유해사이트 차단을 위한 음향 신호 처리 및 분석, 한국콘텐츠 학회논문지, 2004, pp.1-6
- [21] 전용렬.이현승.허순행.김경신.원동호.김승주. 유해정보 차단 S/W 개선방안 연구, 정보처리학회논문지, 2009, pp.543-554
- [22] 양민수. 안드로이드에서 로컬 프록시를 이용한 유해 콘텐츠 차단에 관한 연구, 2012
- [23] 여세환.이진.김정선. 안드로이드 플랫폼에서 인텐트 모니터링을 통한 유해 어플리케이션 차단 방법, 정보과학회논문지, 2003, pp273-277
- [24] 신정훈. 웹 사이트간 연결 관계를 이용한 유해사이트 판별 방법, 송실대학교 일반대학원, 2014
- [25] 장영현. 하이퍼링크 연관성을 이용한 유해사이트의 자동 분류, 고려대학교 컴퓨터과 학기술대학원, 2005
- [26] 김학현.이광우.김승주.원동호. Short URL을 이용한 유해사이트의 접속차단기법 연구, 한국통신학회 종합 학술 발표회 논문, 2011, pp.938-940
- [27] 배영효.이금녀. 구글 검색을 통한 불법정보 유통 차단 기법, 디지털 포렌식 연구,

2013, pp.77-92

- [28] 이재선.전용희.장정숙. 유해 인터넷 정보 차단을 위한 내용 등급 서비스 기반 사용자 인터페이스 설계 및 구현, 한국통신학회논문지, 2003, pp.957-964
- [29] 이재윤.최보영.정영미. 문헌 자동분류에서 용어가중치 기법에 대한 연구, 한국정보관리학회 학술대회 논문집, 2000, pp.41-44
- [30] 강승식.이하규.손소현.홍기채.문병주. 조사 유형 및 복합명사 인식에 의한 용어가중치 부여 기법, 한국정보과학회 2001년도 가을 학술발표논문집, 2001, pp.196-198
- [31] 강승식. 메일 주소 유효성과 제목-내용 가중치 기법에 의한 스팸 메일 필터링, 멀티미디어학회지, 2006, pp.255-263
- [32] 김수진.김민수.백장선.박혁로. 정규화 용어빈도가중치에 의한 자동문서분류, 한국정보과학회 2003년도 봄 학술발표논문집, 2003, pp.510-512