

유해 문자 가중치를 이용한 951 유해 사이트 차단방법

장윤정*, 정규철*, 이태현*, 박기홍*

*군산대학교 컴퓨터정보과학과
e-mail : jyj9510@lycos.co.kr

The Method of Hurtfulness Site Interception Using Poisonous Character Weight

Yun-Jung Jang*, Taehun Lee*, Kyu Cheol Jung*, Kihong Park*
*Dept of Computer Information Science, Kunsan University

요 약

본 논문은 유해 정보 사이트를 차단하는 목적으로 기존의 방법인 제3자가 직접 조사해서 등록하는 방법, 당사자가 직접 등록해서 차단하는 방법, 소프트웨어를 이용해 등록하는 방법, 인터넷서비스 사업자 서비스 제공업체 자체적으로 차단하는 방법 등이 있는데 이런 사이트들의 단점으로는 매일 새로운 사이트가 개설되고 또한 인기 있는 사이트가 포르노 사이트로 둔갑하는 현대사회에는 실정에 맞지 않아 이런 단점을 보완하고자 내용 기반인 테스트에 가중치를 부여하여 유해 정보 사이트를 차단하는 시스템을 제안한 결과 유해 정보 사이트를 접속했을 때 유해 정보 사이트로 판별되는 비율이 78%, 일반 사이트를 접속했을 때 일반사이트로 판별한 비율이 95%로 효과적인 결과가 나타났다.

1. 서론

인터넷은 많은 정보와 지식을 사람들에게 제공해 삶을 풍요롭게 하는 도구이다. 최신 정보의 검색이나 컴퓨터 데이터 처리 및 전송을 빠르게 할 수 있는 장점이 있다. 이러한 장점에도 불구하고 많은 단점을 가지고 있다[1,2,3].

인터넷이 가진 단점들은 인터넷을 통해 음란, 폭력, 마약, 성 매매, 도박, 자살 및 살인 등의 사회적으로 받아들일 수 없는 내용들이다. 그리고 개인 생활의 침해 및 불법행위의 조장 등 유해한 정보들이 이성적인 성인들뿐만 아니라 아직은 감수성이 예민한 청소년들에게까지 무차별적으로 제공되어 진다는 것이다. 인터넷의 단점들을 해결하기 위해 유해 정보를 분류하고 차단하기 위해 여러 가지 방법들이 동원되고 있다. 제3자가 직접 조사해서 등록하는 방법, 당사자가 직접 등록하도록 해서 차단하는 방법, 차단소프트웨어를 이용해 등록하는 방법과 인터넷

서비스사업자(ISP : Internet Service Provider) 자체적으로 차단하는 방법들이 이용되고 있다[4].

인터넷의 단점 해소를 위한 방법에도 문제점이 있다. 하루에도 수십에서 수백의 음란 사이트가 개설되고 있어 음란 사이트를 다 찾아낼 수 없다는 작업의 한계성과 장비가 고장이나 사고가 발생했을 때 데이터가 손실되는 것을 막기 위해 데이터를 하나 이상의 장치에 중복 저장해야하는 문제점이 발생한다. 국내의 유해 정보 차단 소프트웨어들은 대부분이 제3자에 의해서 유해 정보 사이트의 목록을 입력하거나 갱신하게 되는데 그 목록 전체를 다 입력할 수도 없으므로 실효성이 떨어져 청소년들을 유해 정보로부터 보호하는데 한계가 있다[5].

본 연구에서는 기존의 유해 정보 차단의 문제점을 해결하기 위해 내용기반 텍스트를 자동으로 수행하여 유해 정보로의 접근을 차단할 수 있는 시스템을 구현하여 제시하고자 한다. 유해성 여부의 판단 기

준은 정보통신 윤리위원회의 등급 판정 기준을 토대로 각 등급에 해당하는 단어에 가중치를 부여하는 계산법을 제안한다.

다음 2장에서는 기존 유해 정보 차단 방법, 3장에서는 시스템 설계 및 구현, 4장에서는 실험을 통한 시스템의 평가의 유효성을 확인하고 마지막으로 5장에서는 향후 연구 과제 및 결론을 맺는다.

2. 기존 유해 정보의 차단방법

2.1 유해 정보 차단 방법

○ 등급 설정에 의한 차단

제3자가 직접 조사해서 음란 사이트로 등록해서 차단하는 경우와 음란 사이트를 운영하는 당사자가 직접 성인 정보임을 나타내는 등급을 매기는 경우가 있다. 일정 기준의 정의된 등급에 의해 차단되는 인터넷 내용 선별 기술 체계(PICS : Platform for Internet Contents Selection)에 의해 등급 표시 방법을 표준화해서 차단하는 방법이 있다. 또한 국내 정보에 대해서는 정보 제공자가 자율적으로 등급을 표시하고, 보다 쉽고 편리하게 등급표시를 할 수 있도록 인터넷 내용등급 서비스 홈페이지 (<http://www.safenet.ne.kr>)에서 자율 등급표시 웹 서비스를 제공하고 있다. 유럽 경제사회 이사회에서도 유럽의 인터넷 사이트가 아동보호를 위해 의무 등급 표시를 하도록 하고 있다.

○ 인터넷서비스사업자(ISP)를 통한 근본적 차단

기술적으로는 ISP의 인터넷 접속장치에 음란물 차단프로그램을 설치하고 음란사이트의 주소를 차단 프로그램에 입력시켜, 음란 사이트로의 연결을 원천적으로 차단하는 것이며, 음란사이트 목록 데이터베이스는 청소년보호위원회, 정보통신 윤리위원회, 시민감시단체 등의 지속적 정보 수집을 통하여 수시로 갱신한다는 것이다.

○ 차단 소프트웨어를 통한 자율적 규제

유해 정보 차단제품은 지난 95년 7월 미국에서 개발된 사이버패트롤(Cyber Patrol)을 시작으로 서프워치(Surf Watch), 사이버시터(Cyber Sitter), 넷내니(Net Nanny) 등이 개발되어 판매되고 있고 서버용 제품으로는 스마트필터, 리틀브라더 등이 있다. 국내에서는 한국전산원이 개발 중이던 개인용 유해 정보차단 소프트웨어인 NCA패트롤의 핵심 기술을 바탕으로 수호천사, 안티-X, 녹스2, 인터넷지킴이 등 개인용 제품과 수호천사×NET, 넷아거스, 스폰, 보더매니저 등 서버용 제품이 개발되어 있다.

2.2 유해 정보 차단 방법에 있어서의 문제점

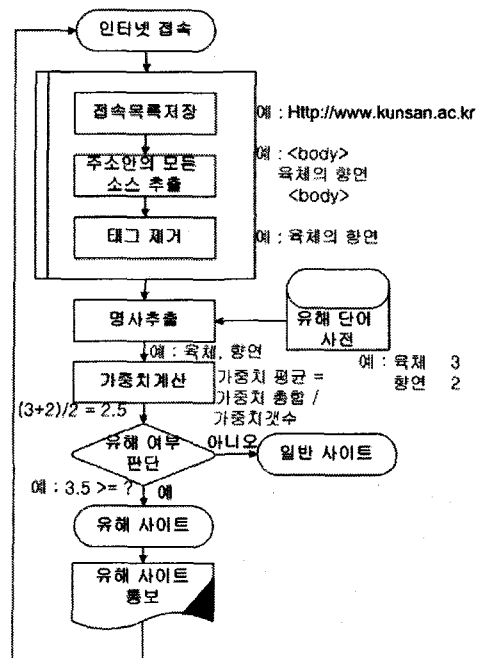
○ 기존 유해 정보 차단의 전반적인 문제

제3자에 의한 등급 설정이나 차단 목록 기반의 차단 소프트웨어는 차단 목록 데이터베이스를 주기적으로 갱신해 주어야한다. 제3자 또는 차단 도구 개발자나 차단 목록 입력자에 의한 사전 검열의 반발이 있을 수 있다. 정보의 제공자에 의한 자율적인 등급 책정은 객관성이 결여되고 주관적인 판단으로 정보의 명확한 등급 책정이 어렵다.

○ '건전 위장' 성인사이트 개설의 문제

인터넷 닷컴 기업들의 경영난이 가중되면서 유명 사이트나 국내 게임 정보를 제공하던 사이트 등 10대 청소년들에게 인기가 있었던 사이트들이 음란 포르노 사이트로 둔갑되어있는 경우가 늘고 있다. 이럴 경우 기존의 차단 방법으로는 게임 정보를 제공하는 사이트로 인식해서 차단이 불가능하다.

본 논문에서는 기존 유해 정보 차단의 문제와 인터넷 불건전 정보 차단의 기술적 문제, '건전 위장' 성인 사이트 개설 문제를 해결하는 방안으로 내용 분석에 의해 청소년 유해 사이트를 차단하는 시스템의 구현에 대해서 다루었다.



[그림 1] 시스템 구성도

3. 시스템의 설계와 구현

3.1 시스템 개요

[그림 1] 시스템 구성도에 대한 시스템의 개요를 설명하면 시스템을 처음 실행해서 인터넷을 접속하면 기본 홈페이지가 접속되고 접속 주소가 파일에 저장된다. 저장된 주소의 소스를 추출해서 소스로부터 태그를 제거하고 텍스트로 저장한다. 텍스트로부터 명사를 추출하기 위해 AC 머신의 알고리즘을 이용한 유해 단어 사전을 사용해서 유해 단어를 추출한다. 추출된 유해 단어의 가중치를 계산해서 유해성 여부를 판단하고 유해 정보 사이트인 경우에 유해 정보 사이트임을 공지하고 유해 정보 사이트의 내용을 읽을 수 없도록 접근을 막았고 확인 버튼을 누르면 일반 사이트로 전환되도록 구성되어 있다.

3.2 주요 시스템 개요 설명

○ 명사 추출

명사를 추출하는 과정에서 태그가 제거된 소스를 읽어 들어서 AC 머신의 알고리즘을 이용하여 단일 패스에서 텍스트 스트링에 포함된 모든 키워드의 위치를 결정할 수 있는 머신[6,7,8]으로 유해 단어 사전을 이용해 내용(텍스트)으로부터 유해 정보를 가진 명사들을 추출한다.

○ 가중치 계산

다음은 유해성 여부를 판단하는 근거자료로 사용한 <정보통신윤리위원회 SafeNet 등급기준>을 나타낸 표이다(<http://www.safenet.net>).

[표 1]의 기준을 적용해서 0등급에 해당하는 단어는 가중치를 1로 부여하고 1등급은 가중치를 2, 2등급은 가중치를 3, 3등급은 가중치를 4, 4등급은 가중치를 5로 부여했다.

○ 유해 여부 판별

[표 1] 정보통신윤리위원회 SafeNet 등급기준

	노출	성행위	폭력	언어
4등급	성기노출	성범죄 또는 노골적인 성행위	잔인한 살해	노골적이고 외설적인 언어
3등급	전신노출	노골적이지 않은 성행위	살해	심한 비속어
2등급	부분노출	착의상태의 성적접촉	상해	거친 비속어
1등급	노출복장	격려한 키스	격투	일상 비속어
0등급	노출없음	성행위없음	폭력없음	비속어 없음

유해 여부를 판별해서 유해 사이트이면 유해 사이트

임을 통보하고 확인버튼을 누르기 전까지는 아무런 동작을 못하도록 한다. 확인버튼을 누르면 처음 시스템을 실행했을 때의 기본 사이트로 이동한다. 유해 정보가 아닌 일반 사이트로 판별되면 아무런 제약조건 없이 접속되도록 했다.

4. 실험 및 평가

4.1 예비 실험

유해 사이트의 유해성 판단 기준 설정을 만들기 위해 유해 단어들의 가중치의 분포를 알아보기 위해서 유해 정보 사이트와 일반 사이트, 성교육 사이트를 대상으로 각각의 가중치 분포를 추출하는 예비 실험을 실시했다.

가중치의 평균을 구하기 위해서는 추출된 명사 가중치의 총합을 추출한 명사의 총 개수로 나눈 값을 평균으로 구했다.

위의 예비 실험을 통해 경험적으로 유해 정보 사이트는 가중치 평균이 3.5 이상인 값으로 나타났고 일반 사이트는 가중치 평균이 2 미만으로 나타나서 유해 정보 사이트임을 판별하는데 가중치 평균이 3.5 이상일 경우에 유해 정보 사이트로 판별하도록 했다.

4.2 실험 및 결과

○ 유해 사이트에 대한 타당성 평가

먼저 5인의 피험자에 의해서 100개의 유해 사이트가 타당한가의 여부에 대해 이하의 4단계로 평가하게 했다.

- A. 유해 사이트로 적절하다.
- B. 유해 사이트로서 위화감이 없다.
- C. 유해 사이트로서 조금 위화감이 있다.
- D. 유해 사이트로서 부적절

그리고 5인 전원이 A평가를 주었던 사이트만이 유해 정보 사이트로 인정한다.

○ 유해 사이트에 대한 실험

[그림 2]에서는 유해 사이트를 일반사이트로 판별 22%가 나왔는데 그 이유는 사이트 내에서 문자를 사용하지 않고 이미지를만 처리한 사이트에 대해서는 유해성이 없는 것으로 판별되었다.

○ 일반 사이트에 대한 실험

[그림 3]에서는 일반 사이트를 유해 사이트로 인식하는 경우 4% 정도 발생했다. 그 이유는 일반 사이트 내에 유해 정보에 속하는 욕설이나 비방 또는 음란 사이트를 홍보하는 내용의 글이 들어있었다.

참고문헌

[1] Taehun Lee, Kazuaki Ando, Masaki Oono and Jun-ichi Aoe, "A Method of Deriving Keywords Using Generation Rules of Compound Word", The Information and Systems Society, The Institute of Electronics, Information and Communication Engineers(Japan) pp.812-821, Apr. 2001.

[2] 이태현, 정규철, 박기홍, "의존관계에 기초한 일본어 키워드 추출방법", 한국정보처리 논문지 제10-B권 제1호. pp.41-46, Feb. 2003.

[3] 이태현, 박기홍, "개념 규칙을 이용한 키워드 도출방법", 2002 한국정보과학회 가을 학술발표 논문집(II), pp.685-687, 2002. 10.

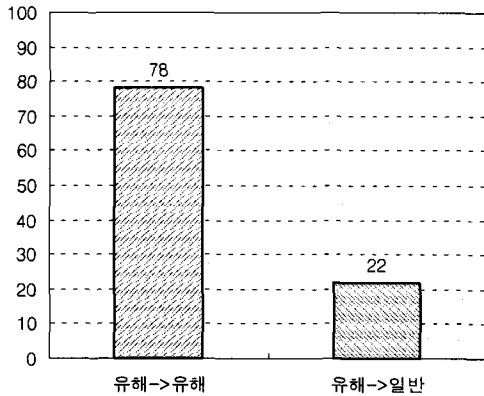
[4] 김현중, "HTML을 지원하는 라이브러리를 이용한 웹 문서 생성 시스템의 설계 및 구현", 이화여자대학교 대학원 전자계산학과, 석사학위 논문, 1996.

[5] 정명숙, "청소년을 위한 유해 웹 영상 차단 시스템의 구현", 경상대학교 교육대학원 전산교육, 2000.

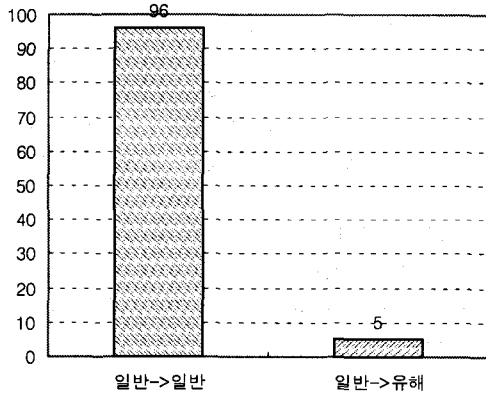
[6] Kazuaki Ando, Toshiharu Kinoshita, Masami Shishibori, Jun-ichi Aoe, "An improvement of the Aho-Corasick machine", International Journal of Information Sciences, Vol 3, pp139-151, 1998.

[7] Kazuaki Ando, Masao Fuketa, Masami Shishibori, Jun-ichi Aoe, "Dictionary Structure for Morphological Analysis of Oriental Languages", Proceedings of 18th ICCPOL, Vol 1, pp533-538, March 1999.

[8] 이진관, "테이블을 이용한 AC 기반의 키워드 검색 기법", 군산대학교 컴퓨터정보과학과, 석사학위 논문, 2002.



[그림 2] 유해 사이트를 접속한 실험 결과



[그림 3] 일반사이트를 접속한 실험 결과

5. 향후 과제 및 결론

본 논문에서는 인터넷상의 청소년의 정신 건강을 해치는 유해 정보들로부터 청소년을 보호하기 위해 유해 사이트를 차단하는 것에 중점을 두고 웹상의 본문 내용(텍스트)을 기반으로 하여 유해 사이트 차단 시스템을 구축하였다. 예비 실험을 통해 단어의 가중치의 평균을 구한 후 유해 사이트를 접속해 본 결과는 유해 정보 사이트로 판명한 비율이 78%를 보였고, 일반 사이트를 접속해 본 결과는 일반사이트로 판명한 비율이 95%의 결과를 얻을 수 있었다.

향후 연구과제로써는 이미지 사이트를 판단하는 시스템과 웹 브라우저 상에서의 한글 이외 영어, 일본어를 판단하는 시스템을 구축 할 예정이다.