

# 인터넷 게시판에서 정보통신윤리 교육을 위한 유해단어 필터링 시스템의 설계와 구현

김치민\*, 김응곤\*

\*순천대학교 컴퓨터과학과

\*e-mail:sinkimch@chol.com

## Design and Implementation of Harmful Word Filtering System for Education of Information Communication Ethics on Web Board

Chi-Min Kim\*, Eung-kon Kim\*

\*Dept. of Computer Science, Sunchon National University

### 요 약

사이버 공간은 개방성, 익명성, 탈중심성의 특성을 갖는다. 이러한 특성으로 인하여 긍정적인 측면과 부정적 측면이 동시에 나타나고 있다. 청소년들의 사이버 공간에서의 활동은 제공되는 정보를 단순히 탐색하는 수동적 역할에 그치지 않는다. 청소년들은 자신의 의견을 제시하거나 정보를 제공하는 역할도 동시에 하고 있다. 이 과정에서 스스로 유해한 정보를 생산하거나 게시판에서 타인에게 해를 끼치는 행위가 발생하고 있다. 본 논문은 청소년들의 적극적인 정보제공 활동 시점에 발생하고 있는 정보화 역기능 현상을 해결하는 방안으로 인터넷 게시판에서 유해단어 필터링에 의한 정보윤리 교육 기법을 제안한다.

인터넷 게시판에서 유해단어 필터링 기법은 초·중등학교 홈페이지 게시판에서 나타나는 부적절한 행동과 상대방에 대한 성적 모욕, 욕설의 사용, 상대방 비하 등에 관련된 유해단어를 추출하여 유해단어 사진을 구축하고 필터링하는 방법이다. 필터링 된 결과에 따라 글 쓰는 시점에서 정보윤리 콘텐츠를 제공한다. 이 기법을 학교 홈페이지 게시판에 적용한 결과 그렇지 않은 경우에 비하여 학생들의 글쓰기 자세와 글 내용이 바른 것으로 나타났다.

### 1. 서론

인터넷의 등장은 사회의 변화를 가속시키면서 사이버 공간이라는 가상 사회를 형성시켰다. 개방형으로 구성된 그물형의 네트워크에는 어떠한 중심도 없는 특성을 지니고 있다[1]. 즉, 사이버 공간은 개방적이고 탈중심적인 공간인 동시에 익명성에 의해 지배되는 공간이다. 또한 현실 사회에서 사람들끼리의 만남이 직접적인 대면인 반면, 사이버 공간에서의 만남은 탈맥락성을 갖는다. 서로 얼굴을 대면하지 못한 가운데, 상대방이 제공하는 단편적 정보에 의존하여 만남이 이루어진다.

최근들어 급속하게 추진된 학교 정보화 사업은 학생들에게 사이버 공간으로 접근 할 수 있는 다양한

방법을 제공하고 있다. 학생들은 사이버 공간에서 개방적이고 평등한 개인의 자유를 만끽하면서 사회의 지적 자원을 확장해 가는 자율 문화, 현실 개혁을 지향하는 참여·저항 문화를 형성하면서 긍정적인 활동을 하고 있다.

그러나 사이버 공간의 특성은 이러한 긍정적인 측면 외에 음란물 유통, 인터넷 중독, 사이버 폭력, 언어의 오용, 개인정보의 오·남용, 유해 사이트의 만연 등 심각한 사회적 문제를 발생시키기도 한다[2]. 각급학교에서 공식적인 웹사이트를 운영하게 되면서 기존 웹사이트에서 제공되는 정보의 유해성과 더불어, 학교 웹사이트 게시판을 이용한 청소년들의 유해한 행위들이 폭발적으로 늘어나고 있다.

본 연구에서는 초·중등학교 웹사이트 게시판에서 학생들이 올리는 글의 내용을 유해단어 사전으로 필터링하여, 유해한 내용에 대하여 정보윤리 의식을 고취하고 바른 글쓰기를 유도하는 콘텐츠를 제공하는 시스템을 설계하고 구현하였다.

본 논문에서 개발하고자하는 유해단어 필터링 시스템의 내용은 다음과 같다.

- ▷ 유해단어 목록의 수집과 DB구축
- ▷ 글쓰기 행동 누가 기록 시스템 구현
- ▷ 부적절 글쓰기 행동 판별 모듈 구현
- ▷ 정보윤리 교육 내용 제작

## 2. 관련연구

### 2.1. 정보통신 윤리 교육

2000년 한국학술정보원의 조사 자료에서 학교에서의 정보화 역기능 현상에 대하여 응답자의 45.9%가 심각하다고 응답하였다[3]. 정보화 역기능 현상으로 언어폭력, 음란물 유통, 개인정보의 오·남용, 바이러스 유포 등의 순으로 심각하게 인식하고 있다.

인터넷 상에서의 언어폭력은 감수성이 예민한 청소년기의 학생들에게 심각한 피해를 끼칠 수 있다. 이러한 피해를 예방하기 위하여 불건전 정보 차단 시스템을 설치하거나 유해단어 필터링 기법이 사용되고 있다.

불건전 정보 차단 시스템은 차단사이트의 목록을 이용하여 사용자 컴퓨터 또는 서버, ISP 제공자 단위에서 특정 인터넷 사이트의 접속을 차단하는 방법이다[4]. 그러나 차단을 저조와 차단장치 설치율이 낮아 문제가 있으며, 유해사이트 목록의 지속적인 갱신이 이루어지지 못하고 있다. 또한 학생들의 컴퓨터 조작 능력이 보호자보다 우월하여 보호자들이 효율적으로 통제하지 못하는 문제도 발생하고 있다.

유해단어 필터링 기법은 게시판 글 작성 시 작성된 글에 대하여 유해단어를 검색하여 경고한 후 등록을 차단하는 필터링 기법이다. 이 방법은 사전에 수집된 유해단어로 게시판에 올리는 글의 내용을 필터링하여 글 등록을 차단하게 된다. 필터링 기법은 필터링 되는 단어가 한정되어 있어 사용자가 단어를 변용하게 되면 유해단어를 필터링 할 수 없게 된다. 또한 필터링 결과에 대한 처리가 단순하여, 청소년들에게 자신의 행위를 반성하고 행동을 수정하도록 지도하지 못하고 있다. 이는 부적절한 내용을 등록하지 못하게 하는 효과는 있으나 청소년들의 윤리의

식을 각성시키지 못하고 부적절한 내용을 더욱 교묘하게 작성하게 하는 동기를 제공할 수 있다.

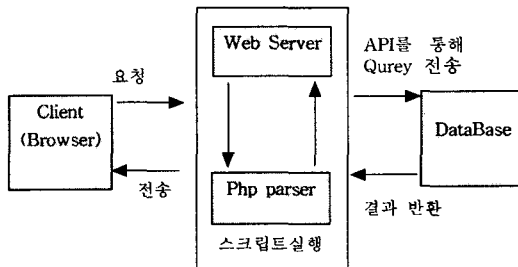
### 2.1.1. 웹 데이터베이스 연동

MYSQL은 웹상에서 서버 스크립트를 통하여 관리할 수 있는 WEB\_DB이다. 공개된 관계형 데이터베이스로 리눅스를 기반으로 한 웹서버에서 많이 사용된다. MYSQL은 속도가 빠르고 표준 SQL문을 충실히 지원하며 Apache 웹서버, PHP Script 언어와 함께 연동할 수 있다[5].

DB와 웹을 연동하는 방법은 하드웨어 환경과 운영체제에 따라 다르다. 서버 확장을 통한 데이터베이스 접근 방식 중 현재 가장 많이 사용하고 있는 서버 자체의 확장 API를 이용하여 서버가 직접 데이터베이스에 접근하는 방식을 기본으로 하였을 경우 (그림1)과 같이 나타낼 수 있다[6].

PHP(Professional Hypertext Preprocessor)는 HTML 문서 내부에 포함되어 웹 서버에서 실행될 수 있는 서버 스크립트 언어이다. 다양한 데이터베이스를 지원하여 데이터베이스와 연동시키기가 쉬우며 실행 속도가 빠르다.

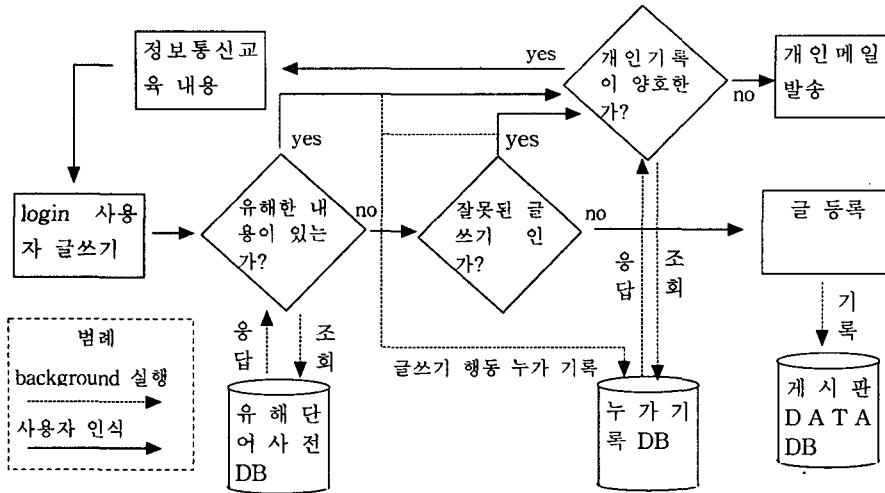
PHP는 CGI방식, 웹서버 모듈 방식, 프러그인 방식 등으로 실행시킨다[7].



(그림 1) PHP를 이용한 데이터베이스 연동원리

## 3. 설계 및 구현

유해단어 필터링 시스템은 유해단어 사전 DB와 개인행동 누가기록 DB, 그리고 필터링 모듈, 부적절 글쓰기 행위 판별 모듈, 정보윤리 콘텐츠로 구성된다. 각각의 모듈을 설계하고 구현하기 위하여 자료의 검색과 갱신의 효율이 높으며 GNU규약에 의거하여 경제적 부담이 적은 MySQL과 PHP Script 언어를 사용하였다.



(그림 2) 유해단어 필터링에 의한 정보통신윤리교육 실행모형

### 3.1. 유해단어 필터링 시스템 실행 모형

(그림2)의 유해단어 필터링 시스템 실행 모형은 학생들이 인터넷 게시판에서 작성한 글을 등록하는 시점에 작동한다. 사용자가 작성한 글은 유해단어 사전으로 필터링 되어 작성된 내용에서 유해한 단어를 검색하고 분류한다.

유해단어를 검색한 후, 무의미한 문자의 나열 등 부적절한 글쓰기 행위 여부를 판별하고 적절한 글쓰기로 판별되었을 경우 글이 등록된다. 유해단어가 검색되거나 부적절한 글쓰기 행위로 판별될 경우 검색된 유해단어와 해당하는 정보윤리 교육 내용을 제시하여 사용자가 잘못된 표현을 수정하도록 유도한다. 부적절한 글쓰기 행위일 경우 해당하는 정보윤리 교육내용을 제시하도록 하였다.

사용자가 정보윤리 교육내용을 학습하는 동안 시스템은 사용자의 ip, id, 유해단어 코드, 부적절한 글쓰기 유형 등을 기록하게 된다. 기록된 결과는 누적되어 학생들의 행동 유형 판단의 자료가 되며 잘못된 글쓰기가 반복되면 해당 사용자에게 메일을 발송하여 지도하고, Off-Line 교육 대상을 선별할 수 있도록 설계하였다.

### 3.2. 유해단어 사전 분류

분류코드는 <표 1>과 같이 세부적 분류를 위하여 3단위로 하였으며 정보윤리 교육내용과 연계되는 코드이다. 분류코드가 400 미만일 경우 글 등록을 차

단하게 된다.

<표 1> 유해단어 사전 분류코드

분류코드	유해단어군
100	부적절한 성적표현
200	타인비하 수치심 유발표현
300	폭력적 표현
400	언어오용
500	무의미 문자 사용

### 3.3. 부적절한 글쓰기 판별

부적절한 글쓰기 여부를 판별하는 기준을 다음과 같이 정하고 각각에 해당하는 경우 사용자에게 내용의 적절성 여부를 확인하여 등록하도록 유도하도록 하였다.

- ▷ 동일한 단어를 3회 이상 연속 입력한 경우
- ▷ 동일한 문자를 10회 이상 연속 입력한 경우(줄긋기 문자 제외)
- ▷ 내용의 전체 길이가 30자 이내인 경우
- ▷ 20초 이내에 연속하여 게시판에 글을 올리는 경우
- ▷ 동일인이 올린 글의 내용이 90%이상 일치하는 경우

### 3.4. 정보통신윤리 교육내용의 구성

정보통신윤리 교육내용은 유해단어 사전 분류에 대응하여 해당되는 코드에 적절한 내용으로 구성하였다.

### 3.5. 개발 환경

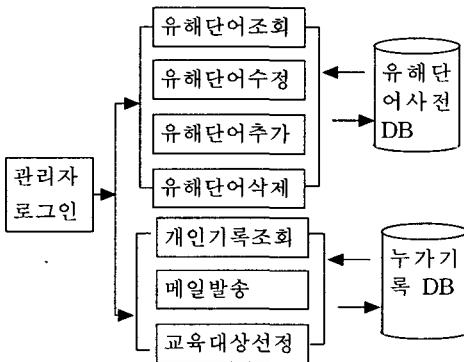
본 모듈은 Linux를 기본 운영체제로 설치하고, Apache 웹서버 환경에서 [표 2]와 같이 DB는 MySQL을 사용하고 PHP에서 제공하는 API 모듈을 이용하여 연동시켰다. 사용자 인터페이스의 부분 처리를 위해 JavaScript를 도입하였다. 웹브라우저는 많은 사용자들을 확보하고 있는 MS 익스플로러를 기준으로 작성하였다.

<표 2> 개발환경 구성요소

구분	구분	사양
하드웨어	CPU	IntelPentiumIV 1Ghz
	HDD	30G
	RAM	1.25G
소프트웨어	OS	Linux
	Webserver	Apache
	Database	MySQL
	Script language	PHP

3.6. 관리 모듈

인터넷 게시판에서 유해단어 필터링이 사용될 경우 사용자들은 여러 가지 음절의 조합으로 새로운 유해단어를 만들어 사용하는 경우가 많다. 따라서 유해단어 사전의 계속적 보완이 필요하며 특히 유해단어와 관련단어를 연계시키는 작업이 필요하다. 이를 위하여 (그림3)과 같이 유해단어 관리자 모듈을 통하여 지속적인 보완이 가능하도록 하였다.



(그림 3) 관리자 모듈 실행 모형

4. 결론

사이버 공간은 청소년들에게 학교와 함께 제2의 생활공간으로 자리하고 있다. 청소년들의 주된 생활공간으로 등장한 사이버 공간은 새로운 생활 환경과 문화를 형성하고 있다. 교사들은 학습지도 활동에 적극적으로 인터넷을 활용하고 있으며, 학생들도 학습활동에 인터넷을 활용하는 빈도가 늘고 있

다. 현재 각급 학교에서는 학교 홈페이지를 통하여 학생들의 학습활동을 지원하며 학교생활과 학생활동에 필요한 공간을 제공하고 있다.

그러나 사이버공간의 특성인 익명성과 개방성, 탈중심성으로 인하여 학생들은 게시판에 등록하는 글에서 부적절한 내용을 담거나, 인신공격, 상대방 비하, 근거 없는 비방, 무의미한 내용의 반복 등록 등의 부정적인 행동을 하고 있다. 이러한 행동은 아직 가치관이 완성되지 못한 학생들에게 도덕적 가치의 혼란을 초래할 수 있으며 바른 글쓰기 자세와 자신의 의견을 설득력 있게 전달하는 능력의 발달에 지장을 초래할 수 있다.

본 연구에서는 초·중학교에서 홈페이지에서 나타나는 부적절한 글쓰기 행동을 교정할 수 있는 인터넷 게시판에서 유해단어 필터링을 통한 정보윤리교육 시스템을 설계하고 구현하였다. 학생들이 인터넷 게시판에서 글을 작성하고 등록하는 시점에서 유해단어를 필터링하고, 부적절한 글쓰기 행동을 판별하여 실시간으로 정보윤리 의식을 각성시키는 교육내용을 제시하였다.

유해단어 필터링에 의한 정보통신윤리 교육 시스템을 적용한 학교 홈페이지의 게시판에 등록된 글에 나타난 유해단어와 부적절 글쓰기 행동 빈도는 그렇지 않은 경우에 비하여 80% 감소하여 의미 있는 결과를 얻을 수 있었다.

참고문헌

- [1] 한국학술정보원, 교육기관 정보화 역기능 방지에 관한 연구, 방문사, 2000.
- [2] 추병완, "사이버윤리의 정립방안", 청소년보호 정책토론 자료집, 2001.
- [3] 한국교육학술정보원, 교육기관 정보화 역기능 방지에 관한 연구, 방문사, 2000.
- [4] 김성조, "기술적 접근 전략", 청소년보호 정책토론 자료집, 2001.
- [5] 허정수, MySQL Advanced Class, 베스트북, 2001
- [6] 허정수, MySQL Web DB연동, 베스트북, 2000.
- [7] 이승혁, PHP 프로그래밍 가이드, 마이트Press, 2000.