
사용자 패턴을 이용한 지능형 e-메일 시스템의 연구

A Research on the Intelligent E-mail System Using User Patterns

임양원, 임한규
안동대학교 멀티미디어공학과

Yang-Won Lim(limyw@andong.ac.kr), Han-Kyu Lim(hklim@andong.ac.kr)

요약

전자우편은 인터넷을 이용하는 사용자들에게 중요한 커뮤니케이션의 역할을 담당하고 있다. 하지만, 원하지 않는 광고가 많은 스팸 메일, 악의를 가진 폭탄메일 등 대부분이 불필요한 자료들로 인해 전자우편이 가지는 본연의 의미와는 무색하게 사용되어지고 있다. 본 논문에서는 이러한 불필요한 정보와 자료들을 최대한 방지하고 보다 깨끗한 환경에서 이용할 수 있는 전자우편을 만들기 위해 사용자 패턴을 이용한 지능형 전자우편 조정 관리 시스템에 대한 연구이다. 사용자가 전자우편을 이용하는 형태, 즉 수신된 전자우편에 대해 사용자의 행동 패턴에 대한 집중적인 분석으로 불필요한 정보와 필요한 정보를 자동으로 분류하여 스팸 메일을 빠르게 처리할 수 있도록 하였다.

■ 중심어 : | 스팸메일 | E-mail | 필터링 | 사용자행동패턴 |

Abstract

Electronic mail (E-mail) is an integral part of communication for the recent Internet users. However, e-mail has also come to serve as a means to support flood of unwanted spam mails and junk mails having bad purposes. This paper was conducted in order to develop an intelligent e-mail system using user behavior pattern that can prevent these unnecessary information and enable the user to enjoy communication via e-mail in a cleaner environment. The concentrated analysis of the user behavior in terms of using e-mail functions has resulted in better classification between unnecessary and necessary information, thereby facilitating faster disposal of spam mails.

■ keyword : | Spam | E-mail | Filtering | Behavior Pattern |

1. 서 론

인터넷의 급속한 성장과 더불어 인터넷 사용자가 늘어나면서 전자우편(E-mail) 시스템은 현재 정보 및

의사 교환의 필수적인 매체로 개인적인 커뮤니케이션의 중요한 역할을 담당하고 있으며, 그 가용성 또한 빠르게 발전해가고 있다. 하지만 개인에게 소중한 고 필요한 메일뿐만 아니라 대부분이 개인에게는 불

* 이 논문은 2003년도 안동대학교 특별학술연구지원사업에 의하여 연구되었습니다.

접수번호 : #051223-001

접수일자 : 2005년 12월 23일

심사완료일 : 2006년 01월 11일

교신저자 : 임한규, e-mail : hklim@andong.ac.kr

필요한 광고성 스팸 메일, 악의를 가지고 보내어지는 폭탄 메일 그리고 메일을 통해 전달되어지는 바이러스성 메일 등(이후 “스팸 메일”)으로 인해 개인의 시간과 비용을 부담할 수밖에 없어 메일의 역기능을 불러오고 있는 것이 현실이다[1]. 시만텍을 비롯한 세계 유수의 스팸 차단 솔루션 업체들은 2004년 5월 발송된 전체 이메일 가운데 대략 75%를 스팸으로 분류했으며, 스팸으로 인한 우리나라의 사회, 경제적인 손실 비용은 연간 2조 6천억 원에 이른다[2]. 2005년 하반기에는 일반사용자 1인당 일평균 스팸 메일 수신량은 6.9통으로 2004년 12월의 조사결과(13.8통)에 비해 50.2% 감소하였지만 수신자가 원하지 않는 데도 계속 뿌려지는 스팸 메일은 더 이상 광고로서의 가치가 없는 사회적 공해에 불과하다[3].

이에 따라 스팸 메일을 차단하기 위해 여러 가지 방법들이 연구되고 있으나 스팸 메일을 차단하기 위해서는 기술적인 어려움도 있지만 개인성향이 다른 것으로 인한 어려움도 있다. 예를 들어 어떤 사람에게서는 스팸으로 분류되는 대출광고가 다른 사람에게겐 원하는 정보일 수 있기 때문에 어떤 것을 스팸 메일로 구분해야 되는가에 대한 판단이 어려울 수 있다.

대표적인 스팸 메일 방지 기술로는 메일의 발신자, 제목 등의 특정 조건을 필터링하여 차단하는 방법과 각각의 메일시스템에서 제공하는 사용자의 스팸 신고 기능을 이용한 방법 등이 있다[4]. 그러나 이와 같은 단순한 특정 조건을 기반한 필터링 기법으로는 스팸 메일을 걸러내는 것이 어려울 뿐 아니라, 스팸 메일이 아닌 경우에도 조건으로 검색된 메일은 스팸으로 처리되는 경우가 발생할 수 있다.

이러한 문제의 보완을 위해 본 논문에서는 사용자와 관리자가 보다 편리한 형태로 스팸 메일을 구분할 수 있도록 스팸 메일임을 확인할 수 있는 단어뿐 아니라 사용자의 행동 패턴을 데이터베이스화하여 불필요한 정보와 필요한 정보를 자동으로 분류하고 처리하여 사용자가 스팸 메일 처리를 좀 더 쉽게 접근하고 빠르게 처리할 수 있도록 제안한다.

II. 관련연구

스팸 메일을 걸러내기 위한 방법으로 기존에 연구되어지고 있는 문서분류, 필터링 등의 기법들이 사용되어지고 있다[5][6]. 그 밖에 확률적인 방법을 적용한 메일 필터링 시스템들이 개발되어지고 있다[7][8].

1. RBL(Real-time Black-hole List)에 의한 차단방법

RBL은 사용자들의 신고에 의해 스팸 메일 발송서버의 IP를 차단하게 된다[9]. 전 세계적으로 RBL을 사용하는 곳이 많아지고 있다. RBL에서 차단하는 IP들은 각 포털이나 기관으로부터 스팸성 메일로 신고된 것이며, 각 호스팅이나 포털에서 자신들이 인정하는 white list를 통해 특정 IP에 대해서는 black list에 들어가지 않도록 먼저 등록을 해야만 한다. 우리나라의 경우 RBL은 KISA에서 진행하는 프로젝트이며, 차단리스트도 KISA에서 작성된다[10]. RBL은 SMTP(Simple Mail Transfer Protocol)[11]상에서 스팸 메일을 차단하기 때문에 메일서버의 부담도 적을 뿐 아니라, 스팸 메일 검사시간도 짧다. 그러나 메일 서버 사용자가 보낸 스팸 메일로 인해서 동일한 메일서버를 사용하는 전체 사용자들이 피해를 볼 수 있다.

2. 메시지 규칙을 이용한 차단 방법

메시지 규칙을 이용한 필터링 방법은 스팸의 특징으로 대표될 수 있는 단어들을 찾아내어 스팸 메일의 여부를 판단하는 방법으로서 확률적인 방법들에 비해 단순하며 재현율과 정확도에 있어서도 비교적 좋은 성능을 얻을 수 있다[12]. 반면에 사용자가 직접 메시지 규칙을 입력해야 하며, 스팸 메일의 형태가 변화함에 따라 메시지 규칙도 지속적으로 갱신시켜야 하는 문제가 발생한다. 뿐만 아니라 참과 거짓 두 가지의 경우만을 결과로 갖기 때문에 보다 정확한 필터링을 하는데 한계가 있다[4].

3. Web-based Mail System

웹메일 시스템 계정과 서비스를 제공하는 서버와 사용자와의 인터페이스 역할을 하는 클라이언트로 구성되며 웹 브라우저가 클라이언트의 역할을 담당하도록 구성되어있다. 즉, 일반 메일서버에 계정을 만들 수 없는 일반 사용자들에게 웹 환경을 통하여 메일서비스를 제공받을 수 있도록 하는 것이다[13]. 본 논문에서 실험을 위해 웹메일 시스템을 이용하였다.

III. 제안하는 스팸 메일 필터링 시스템

1. 시스템의 구조

본 논문에서 제시한 시스템의 테스트 및 설계를 위해 메일 시스템은 일반적인 웹기반 메일 시스템 구조를 바탕으로 하였다. 아래 [그림 1]은 수신된 메일이 메일서버의 데이터베이스로 저장되어 사용자가 읽기 전 패턴을 저장하는 데이터베이스에서 패턴을 읽어와서 선처리하는 과정을 보여주고 있으며, 메일서버로 수신된 메일은 다시 패턴을 처리하여 저장하도록 하였다. 본 논문에서 제시한 사용자 패턴을 이용한 지능형 전자우편 조정 관리 시스템의 동작원리는 다음과 같다.

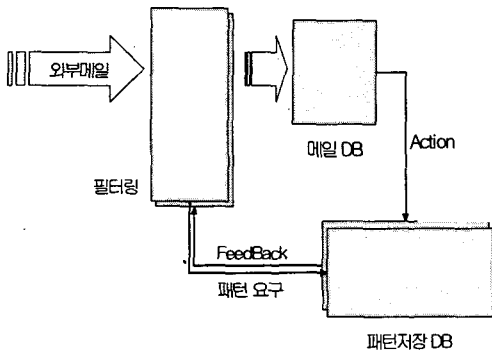


그림 1. 시스템의 구조

• 시스템의 처리과정

- ① 새로운 메일 수신 및 저장
수신된 메일은 메일 데이터베이스에 저장된다.

② Action

사용자가 메일시스템에 접속 후 메일 데이터베이스에 저장된 새로운 메일을 읽거나, 읽지 않거나, 삭제하거나, 또는 저장하는 사용자의 행동을 의미하며, 이러한 사용자의 행동을 패턴저장 데이터베이스에 저장한다.

③ Feedback 및 패턴요구

수신된 메일에 대한 스팸 여부의 패턴 확인을 위해 패턴이 저장되어 있는 데이터베이스에 저장된 내용과 수신된 메일을 분석한다. 불필요한 정보일 경우 수신된 해당 메일을 삭제한 후, 패턴저장 데이터베이스는 특정한 패턴에 의해 삭제되었다는 정보를 Feedback 과정을 통해 저장하게 된다. 필터링에서 삭제되지 않는 메일(안전하다고 판단되는 메일)은 메일서버의 데이터베이스에 저장된다.

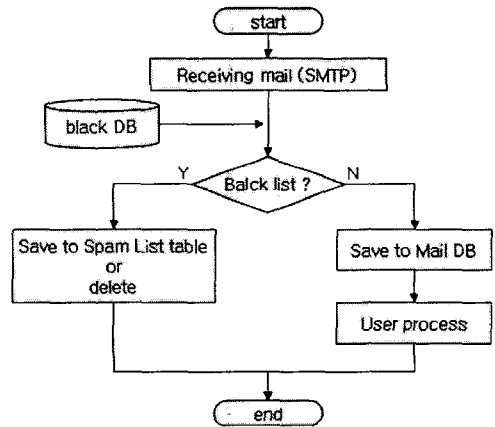


그림 2. 수신된 메일의 처리과정

• 수신된 메일의 처리과정

[그림 2]는 수신되었을 때의 메일서버의 동작과 필터링 과정을 나타낸다.

- ① 수신된 메일은 패턴 데이터베이스에 필터링된 메일들이 저장되어 있는 블랙 데이터베이스에서 블랙리스트인지 아닌지를 검사하여
- ② 블랙리스트에 저장된 정보이면 삭제되거나 스팸 리스트 테이블로 저장하고
- ③ 블랙리스트가 아니라면 메일 데이터베이스로 저

장된 후 사용자 처리과정을 거쳐게 된다.

• 사용자 처리과정(action)

[그림 3]은 수신된 메일의 확인하는 사용자의 행동 처리과정을 보여준다.

- ① 사용자가 메일 데이터베이스로부터 새로운 메일을 읽지 않고 삭제했다면 해당 메일은 스팸 메일일 가능성이 있기 때문에 패턴 데이터베이스로 메일의 정보(발신자, 발신자IP, 메일제목 등)와 함께 사용자의 행동(삭제)을 함께 저장한다.
- ② 메일을 읽고 삭제한 경우도 사용자 행동과 함께 메일정보를 패턴 데이터베이스로 저장한다.

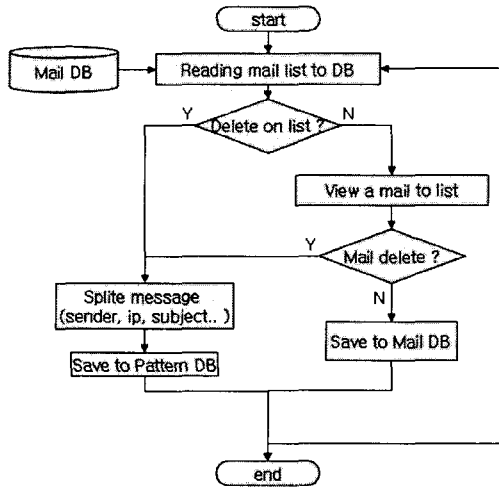


그림 3. 사용자 처리과정(action)

2. 사용자의 행동 패턴에 대한 분류

본 논문에서 사용되어지는 사용자 행동 패턴의 형태는 다음과 같이 크게 개별 사용자에게 대한 행동 패턴과 사용자 각각의 패턴을 전체 통계로 분석하여 패턴을 분류하는 방법으로 연구하였다.

먼저, 사용자가 수신되는 메일에 대한 행동 패턴의 종류는 다음과 같이 분류하였다.

- 사용자가 메일의 내용을 확인하지 않고 삭제한 경우에
 - ① 메일서버 접속 후 바로 메일을 읽은 경우

- ② 메일서버 접속 후 바로 메일을 읽지 않고, n 번의 접속 후 읽은 경우

- 사용자가 메일의 내용을 확인하고 삭제한 경우에
 - ① 메일을 읽으면서 바로(n초 이하) 삭제한 경우
 - ② 메일 확인 후 n초(또는 n분)의 시간이 경과된 후 삭제한 경우

- 사용자가 메일의 내용을 확인하지 않고 삭제하지 않은 경우

사용자의 행동 패턴에 대한 정보와 사용자의 메일 사용에 대한 기초 자료를 바탕으로 각각의 사용자와 전체 사용자에게 적용되는 패턴을 분석하여 적용하였다.

다음은 사용자의 수신된 메일사용에 대한 기초 자료이다.

- ① 발신 IP
- ② 발신 도메인
- ③ 발신자 메일 주소
- ④ 발신자 이름
- ⑤ 메일제목 (음절단위로 저장)
- ⑥ 메일용량

3. 패턴저장 데이터베이스 설계

본 논문에서 제안하는 시스템의 패턴저장 데이터베이스는 [그림 4]와 같은 구조로 설계하였다.

tblUser는 메일서버의 사용자 테이블이며, 패턴분석시에 사용자의 고유코드 구분을 위해 사용되어졌다. tblTempPattern는 사용자의 행동 패턴에 따라 수신된 메일의 정보와 패턴정보를 저장하도록 하였으며, 패턴분석을 통한 필터링에서 사용되어지는 테이블은 아니며, 단지 동일한 행동패턴의 분석과 통계를 위해 저장되어지는 임시 테이블로 사용되어졌다. 테이블이름이 "Black"으로 명명된 테이블들은 임시 테이블에 저장된 행동 패턴과 정보를 분석하여 "불필요한 정보"로 확인된 정보들로 저장된 것이며, "White"로 명

명된 테이블들은 “필요한 정보”로 확인된 정보들로 저장되었다. tblBlackDomainAll, tblBlackWordAll, tblBlackIPAll, tblWhiteDomainAll, tblWhiteWordAll, tblWhiteIPAll 등과 같은 테이블은 행동 패턴과 메일 정보를 최종적으로 분석하여 기본적인 해당 정보를 저장하였으며, 메일서버가 새로운 메일 수신시 필터링 단계에서 접근 가능하도록 설계하였다.

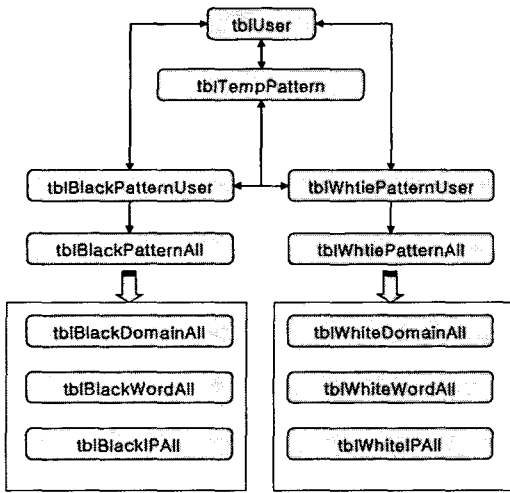


그림 4. 제안된 시스템의 테이블 관계도

4. 주요 행동 패턴 처리

본 논문에서의 행동 패턴의 처리는 먼저 사용자가 수신된 메일에 대한 행동을 데이터베이스에 저장하여야 하며, 저장되는 정보는 수신된 메일에 대한 IP, 도메인, 발신자의 메일주소, 메일 제목의 음절단위의 단어 등과 같은 메일에 대한 일반적인 정보와 수신된 메일을 사용자가 읽었는지 읽지 않았는지, 수신된 메일을 읽고 삭제했는지 삭제하지 않았는지 등의 행동 패턴에 대한 정보를 함께 저장하였다. 본 논문에서는 주요한 행동 패턴 처리에 있어서 위와 같이 수행하고, 각각의 처리는 세분화되어 데이터베이스에 저장되고 패턴분석이 이루어지도록 하였다.

5. 필터링 처리과정

메일서버로 수신되는 메일은 먼저 필터링을 하도록

설계하였으며, 이것은 기존의 메일시스템의 스팸 메일 처리 흐름과 동일한 순서이다. 하지만, 본 논문의 필터링 처리과정은 사용자의 행동 패턴으로 분석된 정보가 저장되어 있는 패턴 데이터베이스의 정보를 이용해서 수신된 메일이 불필요한 메일인지 필요한 메일인지를 처리하도록 하였다. 패턴을 이용한 처리 과정은 다음과 같다.

- ① 새로운 메일이 메일서버로 도착하면
- ② 수신된 메일의 기본적인 정보, 즉 발신 도메인, 발신 IP, 발신자 메일주소 등을 분석한다.
- ③ 먼저, 수신된 메일의 정보가 메일서버 전체 블랙 리스트(불필요한 정보로 분석된 패턴정보)의 정보와 일치하면 삭제한다.
- ④ 메일서버 전체 블랙 리스트와 일치하지 않는다면 수신된 메일의 해당 수신자의 블랙 리스트와 비교한다.
- ⑤ 사용자의 화이트 리스트(필요한 정보로 분석된 패턴정보) 정보와 수신된 메일의 기본정보를 비교하여 일치하면 사용자 메일 정보 테이블에 저장한다.
- ⑥ 사용자의 화이트 리스트 정보와 일치하지 않으면, 해당 사용자의 블랙리스트 정보와 수신된 메일의 기본정보를 다시 비교하여 일치하면 삭제하고, 일치하지 않으면 패턴정보를 저장할 수 있도록 메일 정보 테이블에 저장한다.

6. 실험 및 고찰

본 논문에서 제안 하는 사용자 패턴을 이용한 지능형 전자우편 조정관리 시스템의 테스트를 위해 MS windows2000의 IIS5.0에서 서버 설계를 하였으며, 웹 메일(web-based mail) 시스템은 IIS5.0의 SMTP 서비스를 이용하여 개발하고, 데이터베이스는 간단히 구현할 수 있는 MS의 ACCESS 2000으로 설계하였다. 전체적인 코딩작업을 위해 MS의 ASP(Active Server Page)를 이용하여 실험하였다. 실험을 위한 스팸 메일 자료는, 본 연구를 위해 수집된 500여 개의 메일 원본 내용 중 수신자의 주소만을 동일하게 변경한 후 실험하였다.

IIS 5.0의 SMTP 서비스는 "drop" 폴더에 메시지와 일을(확장자가 *.msg) 넣으면 서버에 메일이 수신된 것으로 인식하여 "drop" 폴더를 읽게 되는데, 이 과정에 ASP로 해당 폴더에 있는 메일을 데이터베이스로 저장할 수 있게 된다. 즉, 수신되는 메시지의 빠른 테스트를 위해 실험용 메시지를 IIS의 메일 수신 폴더인 "drop" 폴더에 복사해 넣는 방식을 사용하였다.

"drop" 폴더에 있는 메시지는 웹메일 시스템이 수신된 메시지로 인식하여 해당 사용자의 데이터베이스로 저장하게 되며, 최초 이와 같은 메일은 "읽지 않은 메일"로 분류가 되어 저장되어 있다.

[그림 5]는 일반적인 메시지 규칙을 이용한 필터링을 하기 위해 특정 단어들을 등록시켜두었다. [그림 5]와 같이 일반적인 메시지 규칙을 통한 필터링을 기본으로 한 이후 제안된 방식과 같이 사용자 행동 패턴에 대해 실험하였다.

그림 5. 단어중심의 필터링

순위	분류 조건	삭제
1	제목에 <성인광고> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
2	제목에 Yang_won (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
3	제목에 <광고> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
4	제목에 <광고> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
5	제목에 [광고] (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
6	제목에 {광고} (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
7	제목에 (광고) (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
8	제목에 [광고] (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
9	제목에 <카드값> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
10	제목에 <카드값> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
11	제목에 <카드값> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
12	제목에 <카드값> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
13	제목에 <카드값> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>
14	제목에 <카드값> (01)가 포함된 메일을 바로삭제	<input checked="" type="checkbox"/>

그림 5. 단어중심의 필터링

[그림 6]은 사용자 행동 패턴을 통해 저장된 패턴정보를 바탕으로 만들어진 스팸차단 리스트 결과이다.

스팸메일 차단 > E-Mail 차단		[E-Mail 차단 >]	
E-Mail	설명	등록일	삭제
1004ix@korea.com	성인광고 (미유니)	2005-11-18	<input checked="" type="checkbox"/>
asunjean12911@netian.com	성인광고 (유니)	2005-11-18	<input checked="" type="checkbox"/>
aca3krkr@yahoo.co.kr	성인광고 .. (사용자 lm에게 오는)	2005-11-19	<input checked="" type="checkbox"/>
ad@ksize.com	성인광고 (르사이즈 업체)	2005-11-18	<input checked="" type="checkbox"/>
	바이러스성 메일을 주소만 바꿔서 발송	2005-11-19	<input checked="" type="checkbox"/>
angel00msh@lycos.co.kr	성인광고 (민선)	2005-11-19	<input checked="" type="checkbox"/>
big@boss.com	악성 바이러스	2005-11-22	<input checked="" type="checkbox"/>
bluko@hanmail.net	성인광고 (홍고)	2005-11-22	<input checked="" type="checkbox"/>
058ulkrway7da@yahoo.co.kr	성인광고 (미성년)	2005-11-22	<input checked="" type="checkbox"/>
dialmdejr2000@hanml.com	성인광고 (규현미)	2005-11-22	<input checked="" type="checkbox"/>
hamjules@yahoo.com	성인 광고	2005-11-22	<input checked="" type="checkbox"/>
hnh7910@lycos.co.kr	성인광고 (현미환)	2005-11-22	<input checked="" type="checkbox"/>
hongek3@lycos.co.kr	성인광고 (김미혜)	2005-11-22	<input checked="" type="checkbox"/>

그림 6. 스팸 메일 차단 결과화면

수신자의 반복적인 행동에 따라 리스트에 등록이 된다. 예를 들어 수신자가 동일한 패턴을 가지는 메일들에 대해서 "읽지 않고 삭제"를 3번 이상 반복하였다면 해당 메일은 더 이상 사용자가 원하는 메일이 아닐 것이다. 우선적으로 [그림 6]과 같이 스팸차단 리스트에 자동으로 저장된다.

[그림 7]은 스팸 리스트로 자동 처리되어 패턴 DB에 저장된 사용자 패턴정보의 화면이다.

tblTempPattern : HUS						
UserCode	Code	Target	Keyword	Action	ActionPara	
-2024586240	1	0	해당누군가	2	12	
-2024586240	1	4	광고	0	0	
-2024586240	2	4	성인	0	0	
-2024586240	3	4	무료	0	0	
-2024586240	4	4	정보	0	0	
-2024586240	5	4	포로노	0	0	
-2024586240	6	4	력	0	0	
-2024586240	8	4	대생	0	0	
-2024586240	9	4	음	0	0	
-2024586240	10	4	대박	0	0	
-2024586240	11	4	부자	0	0	
-2024586240	12	4	info	2	-2	
-2024586240	13	4	성	0	0	
-2024586240	14	4	영교생	0	0	
-2024586240	15	4	학원	0	0	
-2024586240	16	4	채취연습	0	0	
-2024586240	17	4	만식재	0	0	
-2024586240	18	4	{광고}	0	0	
-2024586240	19	0	test	0	0	
-2007906977	3	4	싸고	0	0	

그림 7. 사용자 행동 패턴 정보

실험은 동일한 웹메일 시스템에서 일반적인 메시지 규칙만을 위한 방법과 본 논문에서 제안하는 사용자 패턴을 이용한 방법으로 진행하였다. [표 1]은 두 가지 실험에 대한 스팸 메일 처리율을 보여준다.

표 1. 스팸 메일 처리율

메일 수(개)	메시지 규칙을 이용한 방법	사용자 패턴을 이용한 방법
100	80.5%	80.5%
200	81.9%	85.1%
300	88.4%	90.2%
400	70.5%	85.9%
500	83.1%	89.1%
평균	80.9%	86.2%

실험을 위해 메일을 100개씩 나눠서 "drop" 폴더에 복사해 넣고, 사용자 계정으로 웹메일 시스템으로 접속 후 새로 온 메일을 읽어 들이고, 본 논문의 3장 2절에서 언급한 사용자의 여러 가지 행동 패턴을 이용해서 수신된 메일을 읽거나 삭제하였다. 이와 같이 반복해서 그 다음 100개의 메일도 동일한 방법으로 계속해서 실험하였다. 실험결과, 메일의 수가 적을 때는 비슷한 스팸 메일 처리율을 나타내지만 메일의 양이 많아지게 되면 본 논문에서 제안한 사용자 패턴을 이용한 방법이 스팸 메일 처리율이 증가하는 것을 볼 수 있다.(스팸 메일 처리율이 높은 이유는 실험을 위해 수집된 메일이 대부분이 광고성 메일이기 때문이다)

이와 같이 제안된 방법은 일반적인 메시지 규칙을 사용하는 방법과는 다르게 장기적으로 사용할수록 스팸 처리가 원활히 이루어지 것을 볼 수 있다.

IV. 결 론

현재 나와 있는 여러 가지 스팸 메일 처리는 다양한 기술과 특징들을 제공하고 있으나, 본 논문에서 제시하는 사용자 패턴을 이용한 스팸 메일 처리 방법은 확실적인 스팸 처리 방식에서 벗어나 더욱 지능화 되고 다양한 환경에서 적용할 수 있도록 제공한다. 특히, 사용자의 행동 패턴의 분석과 개인의 행동 패턴의 정보를 다시 전체 메일서버의 패턴으로 분석함으로써 시간이 경과함에 따라 더욱 지능화되고 관리자가 별도로 개입하지 않아도 새로운 스팸 메일의 유형을 판독하여 변화하는 스팸 메일의 다양성에 적

응할 수 있도록 하였다.

본 논문에서 제시한 시스템의 설계는 사용자의 행동 패턴 분석에 의한 스팸 처리를 위한 메일서버의 지능화에 목표를 두고 연구하였다. 향후, 패턴분석으로 인한 패턴 데이터베이스의 크기를 최적화하여 최소한의 데이터베이스의 크기로 필요한 패턴정보를 저장할 수 있는 연구를 수행할 계획이다. 또한, 이를 통해 패턴분석을 통한 스팸 메일처리를 원활히 할 수 있는 메일서버 시스템과 패턴 분석에 들어가는 비용과 시간을 최소로 줄일 수 있는 데이터베이스의 가용성을 높일 수 있는 연구가 필요하다.

참 고 문 헌

- [1] 김민중, "스팸 메일-정보인가 법적 규율의 대상인가", 한국 인터넷 법학회, 2002.
- [2] <http://www.kisa.or.kr>
- [3] <http://www.spamcop.or.kr>
- [4] 한길성, "스팸 메일에 대한 현황 분석 및 방지 기술에 대한 고찰", 한국콘텐츠학회논문지, 제1권, 제2호, pp.92-97, 2003.
- [5] 고주영, 심재창, 김현기, "비밀단어의 회신을 이용한 스팸 메일 차단 시스템의 구현", 멀티미디어 학회논문지, 제8권, 제1호, pp.111-118, 2005.
- [6] 백기영, 김승해, 최장원, 류재철, "수집과 빈도분석을 통한 스팸 메일 차단 방법", 정보처리학회 논문지, 제12-C권, 제1호, pp.137-146, 2005.
- [7] <http://email.about.com/od/bayesianfilteringsoftware/>
- [8] 김현준, 정재은, 조근식, "가중치가 부여된 베이저안 분류자를 이용한 스팸 메일 필터링 시스템", 정보과학회논문지, 제31권, 제8호, pp.1092-1100, 2004.
- [9] <http://www.mailabuses.com>
- [10] <http://www.kisarbl.or.kr/web/main.jsp>
- [11] RFC821, "Simple Mail Transfer Protocol"
- [12] K. Johnson, *Internet Email Protocols A*

Developer's Guide, Addison Wesley, 1999.

- [13] 임양원, 권기훈, 임한규 “서비스엔진을 이용한 웹 기반 메일 에이전트 시스템의 설계 및 구현”, 정보처리학회논문지, 제7권, 제25호, pp.608-620, 2000.

저자 소개

임 양 원(Yang-Won Lim)

정회원



- 1992년 : 충주대학교 컴퓨터공학과 (공학사)
- 2001년 : 안동대학교 컴퓨터공학과 (공학석사)
- 2004년 : 안동대학교 멀티미디어공학전공 박사과정

<관심분야> : 멀티미디어, 에이전트, 웹프로그래밍 등

임 한 규(Han-Kyu Lim)

중신회원



- 1981년 : 경북대학교 전자계산기공학전공(공학사)
- 1984년 : 연세대학교 전산전공(공학석사)
- 1997년 : 성균관대학교 컴퓨터공학전공(공학 박사)

• 1998년 3월~현재 : 안동대학교 멀티미디어공학전공 부교수

<관심분야> 멀티미디어, 자연언어처리 등