

# E-Mail 시스템의 멀티미디어 첨부파일 자동분류 시스템 설계

현영순\*, 정옥란, 조동섭  
이화여자대학교 과학기술대학원 컴퓨터학과

## Classification of the Multimedia Attached File in E-Mail System

Young-Soon Hyun, Ok-Ran Jeong, Dong-Sub Cho  
Dept. of Computer Engineering, Ewha Womans University.

### 요 약

인터넷 사용자의 증가와 함께 e-mail 사용자 또한 증가하게 되면서 대량의 메일을 송수신하는 경우, 메일에 대한 효율적 관리의 중요성이 부각되고 있다. 본 논문에서는 수신된 메일의 내용을 텍스트 형태로 가져온 뒤, **parallel regular expression**을 이용하여 **Keyword**를 검색하고 추출하여 메일에 첨부되어온 멀티미디어 파일들을 자동으로 분류·저장하는 멀티미디어 첨부파일 자동분류 시스템을 제안하였다. 수신된 메일을 일일이 확인하고 분류해야만 했던 기존의 시스템과는 달리 본 논문에서 제안하고자 하는 시스템을 이용했을 경우 노력과 시간을 절감하고 첨부파일들을 효과적으로 관리할 수 있다는 장점이 있다.

### 1. 서론

E-Mail은 현재 가장 활용성이 높은 인터넷 서비스로서 인터넷이 보편화 되어 감에 따라 그 사용자 또한 증가하고 있는 추세이다.

기존의 메일 에이전트 시스템에서는 수신된 메일을 무분별 하게 서버에 적재하는 방식이었다. 그렇기 때문에 대량의 메일을 송수신 해야 하는 경우, 메일의 내용은 물론 메일에 첨부되어온 파일을 보기

위해서는 관리자가 일일이 확인하고 분류해야 하는 번거로움이 있었다.

본 논문에서는 관리자의 업무부담을 줄이고 멀티미디어 첨부파일의 효과적인 분류를 위해, 메일의 내용을 읽어 들어 **Keyword**를 추출하고 폴더를 생성하여 해당 폴더별로 첨부파일을 라우팅 시켜주는 멀티미디어 첨부파일 자동분류 시스템에 대해 제안하고자 한다. 폴더별로 분류된 멀티미디어 파일들은 다시 종류별로 분류, 저장되어지고 첨부파일 처리에 대한 확인 메일은 자동으로 전송자에게 전송되도록 설계하였다.

---

\* 이 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2003-041-D00460).

본 논문의 구성은 다음과 같다. 2 절에서는 메일

시스템에 대한 기존의 연구에 대해 언급한다. 3 절에서는 제안하는 E-Mail 시스템의 멀티미디어 첨부파일 자동분류 에이전트의 전체 구성에 대해 설명하고, 4 절에서는 본 논문의 결론과 추후 연구 계획에 대해 언급한다.

## 2. 관련연구

### 2.1 일반적인 전자우편 시스템

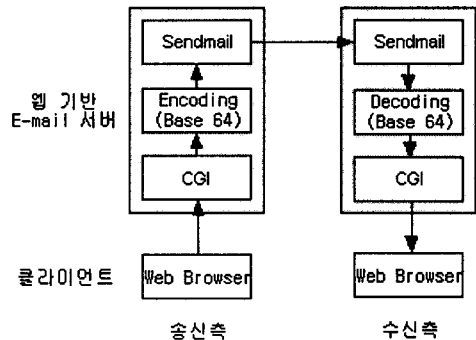
전자우편 시스템은 전자우편 클라이언트 프로그램으로 SMTP(Simple Mail Transfer Protocol)서버에 편지를 보낸다. 전자우편 메시지 수신인의 도메인 서버가 있는 도메인과 같으면 아스키 텍스트를 서버에 저장하고, 자기 도메인이 아닌 경우에는 해당 서버에 전달한다. 전자우편 메시지가 수신자의 SMTP 서버에 도착하면 수신자에게 할당된 사서함에 저장된다. 수신자는 아웃룩 익스프레스 같은 우편 클라이언트에서 POP3(Post Office Protocol) 프로토콜로 서버에 접속해서 자신에게 온 편지를 가져간다. 이 과정에서 포트 25 를 사용하는 SMTP 로 편지를 서버에게 전달하고, 포트 110 을 사용하는 POP3 프로토콜로 편지를 클라이언트까지 배달한다.

일부 시스템(주로 유닉스 플랫폼)에서는 SMTP편지 저장 공간에 직접 접근할 수 있다. 즉, 사용자가 SMTP서버에 접속해서 SMTP의 사서함으로부터 전자우편을 직접 읽을 수 있는 것이다. 그러나, 사용자의 관점에서 보면 아웃룩 익스프레스처럼 일단 사서함에서 클라이언트의 우편함으로 편지를 옮긴 후 읽는 편이 더 편리하다[1].

### 2.2 웹 기반 전자우편 시스템의 동작원리

현재 많은 회사에서 일반 사용자들에게 무료로 서비스하고 있는 웹 기반 전자우편 시스템의 동작원리를 나타내면 [그림 1]과 같다. 사용자는 일반 웹 브라우저를 통해 서버에 접속한 후 보내고자 할 내용을 입력하고 POST 방식으로 서버에 전송하면 서버는 전송 받은 메시지를 CGI 프로그램에서 파싱한다.

파싱된 메시지는 연속적인 8-bit 를 4 개의 ASCII 문자로 변화시키는 Base64 인코딩 작업을 거친 후, sendmail 프로그램을 구동해서 목적지 서버로 전송하게 된다. 목적지 서버에 도착한 메시지는 서버에 저장되어 있다가 수신자가 웹 브라우저를 통해 서버에 접속해서 요청을 하게 되면 반대과정을 거쳐 수신자의 브라우저에 보여지게 된다[2][3][4].



[그림 1] 웹 기반 전자우편 시스템의 동작원리

웹 기반 전자우편 시스템은 사용자가 특정회사의 전자우편 클라이언트 프로그램을 구입하지 않아도 브라우저만 있는 환경이면 전자우편을 송수신할 수 있도록 함으로써 사용자에게 매우 편리한 환경을 제공한다. 이러한 환경은 기업이나 학교 등의 인터넷 환경에 적용될 수 있으며 공동의 사용자 인터페이스를 제공하므로 업무 수행능력이나 소속감 등을 높이는 데에도 기여할 수 있다.

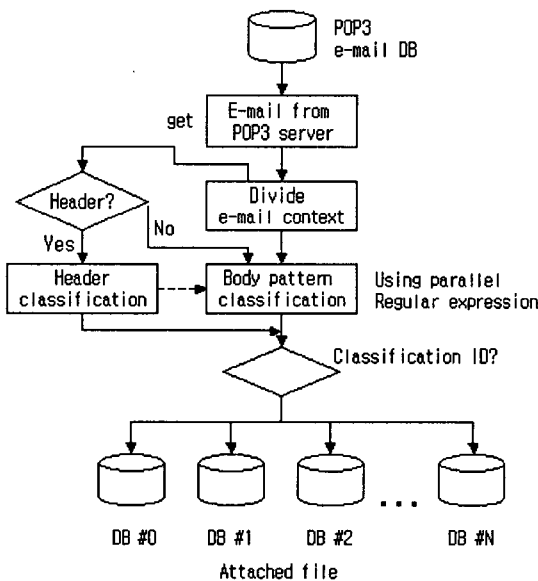
웹 기반 전자우편 시스템을 통하여 수신된 메일들은 다양한 내용을 포함한다. 기업에서 인터넷 환경에 웹 기반 전자우편 시스템을 적용하여 사용할 경우 기업은 매우 많은 양의 메일을 수신할 것이고 그 내용 또한 방대할 것이다. 사용자가 보내는 방대한 양의 메일을 데이터베이스에 저장하기만 한다면 관리자의 입장에서는 메일들을 일일이 읽어보아야만 하는 번거로운 작업이 된다. 따라서 수신된 메일을 적절하게 분류하여 목적에 맞는 부서로 메일을 자동으로 보내주는 에이전트가 필요하다.

## 3. 멀티미디어 첨부파일 자동분류 시스템

### 3.1 시스템의 전체구조

E-Mail 시스템의 멀티미디어 첨부파일 자동분류 시스템에서는 서버에 저장된 메일을 관리자가 POP3 서버를 통해 관리자의 PC로 옮겨온다. Parallel regular expression을 이용하여 옮겨온 메일의 내용을 읽은 뒤 키워드를 추출하고 폴더를 생성, 폴더별로 첨부파일을 분류해서 저장한다. 폴더별로 분류한 멀티미디어 첨부파일은 또 다시 파일의 형식별로 분류되어 저장되어진다. 처리된 첨부파일에 대해서는 송신자에게 처리확인 메일을 자동으로 전송한다.

본 연구에서 제안하고자 하는 멀티미디어 첨부파일 자동분류 시스템의 전체 구성도는 다음 [그림 2] 과 같다.



[그림 2] 멀티미디어 첨부파일 자동분류 시스템

### 3.2 멀티미디어 첨부파일 자동분류 시스템

사용자가 e-mail 을 통해 전송한 메일은 한 곳의 서버에 저장되고 저장된 메일은 관리자가 POP3 서비스를 이용하여 관리자의 PC 로 가져온다. 가져온 메일의 내용은 텍스트 형태로 읽혀지고 parallel regular

expression 을 이용하여 Keyword 의 검색·추출이 이루어진 뒤, 멀티미디어 첨부파일은 이미 구성되어져 있는 데이터베이스 내의 매칭되는 테이블에 저장된다. 이때 일정한 시간이 지나면 새로 수신된 메일이 있는지 확인하고 새로운 메일이 있다면 위의 과정을 거쳐 저장되도록 한다.

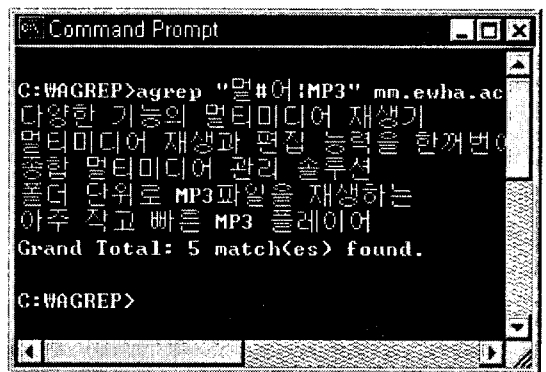
본 연구에서는 데이터베이스 내의 저장을 위해 Keyword 검색 방법을 사용하기 때문에 매칭 테이블 결정을 위한 Keyword 를 미리 지정해 주어야 한다. 어떤 시스템을 설계할 지에 따라서 Keyword 의 내용이 달라지며, Keyword 의 검색은 agrep 이라는 parallel regular expression 을 이용한다.

Parallel regular expression 을 이용하여 내용을 검색한 화면은 다음 [그림 3],[그림 4]와 같다.



[그림 3] 내용 검색 화면 1

‘말’이라는 어두로 시작하여 ‘어’로 끝나는 모든 단어가 포함된 내용을 검색한다.



[그림 4] 내용 검색 화면 2

‘얼’이라는 어두로 시작하여 ‘어’로 끝나는 모든 단어나 또는 ‘MP3’라는 단어가 포함된 모든 내용을 검색한다.

테이블에 저장된 멀티미디어 첨부파일은 시스템에 의해 관련 폴더에 전달·저장되고, 폴더 내에서는 또다시 첨부된 파일의 형식별로 분류·저장되어진다.

처리된 첨부파일에 대해서는 SMTP 서비스를 이용하여 처리확인 자동답장을 보내도록 설계하였다.

#### 4. 결론 및 추후연구

E-Mail 이 현대인의 필수 커뮤니케이션으로 자리잡아가면서 그에 대한 효율적인 관리 또한 중요시되고 있다.

기존의 메일 기반 시스템은 서버에 저장된 메일을 관리자가 하나하나 읽어보아야 하는 비효율적인 시스템이기 때문에 관리자의 업무부담이 크고 시간 낭비가 많았다.

이러한 단점을 보완하기 위해 본 논문에서는 저장된 메일의 내용을 읽어 Keyword 를 검색하고 멀티미디어 첨부파일을 폴더별로 분류·저장, 처리확인 메일까지 자동으로 전송해주는 시스템을 설계하였다. 이는 대량의 메일을 송수신 해야 하는 경우 첨부파일 관리의 필요성이 많은 곳에서 유용하게 사용될 것으로 생각된다.

이 논문의 한계는 첨부파일을 분류하기 위한 방법으로 매칭 테이블을 결정하기 위한 키워드를 미리 지정해 주는 데 있다. 이는 미리 지정해놓은 키워드를 포함하지 않은 문서에 대해서는 분류가 어렵다는 단점을 가지고 있다. 따라서 성능 향상을 위하여 문서들의 특징을 전혀 모르는 상황에서도 문서 내용에서 공통된 패턴을 발견하고 문서를 분류할 수 있는 기계학습 가법을 이용한 자동 문서 분류에 대한 연구가 필요할 것이다[5][6][7].

#### [참고문헌]

[1] 임양원, 권기훈, 임한규, “ 서비스 엔진을 이용

한 웹 기반 메일 에이전트 시스템의 설계 및 구현,” 한국정보처리학회 논문지 제 7 권 제 2 호, 2002.2.

- [2] Stallings, W, Network and Internetwork Security : Principles and Practice. Prentice Hall, 1995.
- [3] 박동욱, 박재희, 김진상, 김일민. “ PGP 방식을 이용한 웹 기반 전자우편 보안 시스템,” 한국정보처리학회 논문지 C, 2001.2.
- [4] Sol, S. and Berznieks, G., CGI/PERL : Web Scripts. M&T Books, 1997.
- [5] C. Buckley, G. Salton and J. Allan “ The Effect of Adding Relevance Information in a Relevance Feedback Environment,” Proc. 17<sup>th</sup> ACM SIGIR International Conference on Research and Development in Information Retrieval, pp.292-298, 1994.
- [6] C.J. van Rijsbergen, Information Retrieval, Butterworths, London. 2<sup>nd</sup> Edition, 1979.
- [7] G. Salton, and M.J. McGill, Introduction to Modern Informaion Retrieval, McGraw-Hill, New York. 1983.