

유사성을 이용한 효율적인 메일 그룹핑 전략

정옥란*, 조동섭
이화여자대학교 컴퓨터학과
e-mail:{orchung*,dscho}@ewha.ac.kr

Effective Mail Grouping Strategy Using vertical & horizontal similarity

Ok-Ran Jeong*, Dong-Sub Cho
*Dept of Computer Science and Engineering,
Ewha Womans University

요약

이메일의 사용자가 인터넷 빠른 증가에 따라 기하급수적으로 늘어나고 있다. 또한 전자상거래에서 무수한 정보 서비스를 메일 시스템을 이용하여 제공할 때, 빠른 처리를 가능하게 하는 방법이 현안이 되고 있다. 이는 메일 관리 시스템의 SMTP 성능 향상이 이 현안의 핵심이라고 할 수 있는데, 본 연구에서는 전자상거래의 메일 내용의 특성을 고려하여 각각의 클라이언트에게 보내는 내용이 같은 항목 하에 일부 내용만이 바뀌는 수평적 유사성, 보내지는 품의 전체적인 디자인의 변화나 회사의 정보가 변경되는 수직적 유사성의 특징을 이용한 메일 그룹핑(grouping) 방법을 제안한다.

1. 서론

인터넷 이메일은 사용자들이 가장 자주 애용하는 프로그램이며, 앞으로 오랫동안 이메일의 응용분야는 늘어날 것이다. 이를 위해 인터넷 이메일은 사용자들의 다양한 기능적 요구에 부응해야 하며, 프로그래머와 메일 관리자들은 더 많은 기능을 추가하기 위해 노력해야 한다[1]. 본 연구는 이러한 현실에 당면한 문제를 해결하기 위하여 특히 대량의 메일을 항상 송신해야 하는 전자상거래의 메일 내용의 특성에 맞추어 SMTP 성능 향상을 위해 매크로를 이용한 그룹핑(grouping) 하는 방법을 제안해 보고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 메일 시스템의 일반적인 구성 요소 및 기능, 전송방법에 대해서 설명하고, 3장에서는 상업용 메일의 수직적 유사성과 수직적 유사성에 대한 정의를 내리고

이를 이용한 메일 서버와 메일 클라이언트의 그룹핑 전략을 제안한다. 마지막으로 4장에서는 결론을 맺는다.

2. 관련연구

2.1 기존 메일 전송 시스템

기존의 인터넷 메일 시스템은 동격 서버들을 가진 분산 클라이언트/서버 시스템이다. 다시 말해 클라이언트는 서버와 통신하여 메일을 송수신하고, 그리고 서버들이 서로 통신하는 형태이다. 클라이언트가 나가는 메시지를 직접 서버(MTA)로 보내면, 그 서버는 메시지를 수신자의 우편함으로 배달하거나 혹은 그것을 전달(forwarding)할 다른 MTA로 보낸다. 이러한 시스템은 서버들을 계층적으로 배열함으로써 높은 확장성을 갖도록 하기 위한 것이다.[2]

인터넷 메일 메시지는 인터넷 사용자가 상대방에게 보내고자 하는 정보이다.

이 논문은 2002년도 두뇌한국21사업에 의하여 지원되었음.

메시지에는 이진 파일(binary file)을 비롯해서 메시지를 첨부할 수 있는 여러 부분이 포함될 수 있다. 메시지는 MUA로부터 보내지고 또한 다른 사람이 보낸 메시지를 읽는 데 사용되기도 한다.

인터넷 메일 시스템의 전송 요소들은 다음과 같다 -MTA(Mail Transfer Agent)

인터넷상에 있는 하나의 컴퓨터로부터 다른 컴퓨터(메일 서버)로 전자 메일을 전송하는 서버 프로그램이다.

- MUA(Mail User Agent)

사용자가 전자 메일을 송수신할 때 사용하는 클라이언트 프로그램으로 MUA는 하나의 프로그램일 수도 있고, 또는 전자 메일을 송수신하는 전형적인 MUA의 행위를 에뮬레이트(emulate)하는 스크립트일 수도 있다.

-MDA(Mail Delivery Agent)

메시지를 사용자의 우편함에 쓰기 위해 MTA가 사용하는 프로그램이다.

-MRA(Mail Retrieval Agent)

원격지 서버에 있는 우편함으로부터 사용자의 MUA로 메시지를 가져오는 서비스이다.

2.2 메일 헤더

메일 헤더의 각 구성요소는 다음과 같다

“Return-Path”는 수신자에게 메시지를 배달하는 마지막 메일 서버에서 덧붙이는 필드로 송신자로 메일을 다시 반환할 주소와 경로 등의 명확한 정보를 포함하고 있다.

“Received”는 편지가 배달되는 경로를 나타내며 “Received”가 한 줄 이상 나타나는 경우에는 메일 보내고 받는 서버 이외의 다른 서버들을 통과해온 것을 의미한다.

“From”은 보내는 사람의 주소를 의미한다.

“Reply-To”필드는 메시지를 처음 받은 메일 서버에서 추가하는 정보로 수신자가 메시지에 대한 답변을 회신할 주소가 된다. 만약에 송신자가 이 필드를 비워둔다면 “From”필드의 주소로 회신하게 된다.

“To”는 수신자의 주소를 나타내며 “Subject”는 이메일의 제목, “X-mailer”은 송신자가 사용한 메일 클라이언트 프로그램, “Data”는 이메일이 보내진 날짜를 의미한다.

“Message-Id”는 해당 이메일에 지정된 식별 번호로 메일 서버가 메시지를 외부로 보내며 붙이는 일련번호로 해당 메시지가 어떤 컴퓨터에서 보내졌는지 알

수 있다.

“Content-Type”은 메일 본문이 어떤 형태인지 알려 주는데 text/plain은 일반 문자열을 사용한 본문이고, 일반 문자열과 여러 인코딩 방식이 섞여 있을 경우에, multipart/mixed는 일반 문자열과 파일을 첨부하였을 때, multipart/alternative는 같은 내용이 일반 문자열과 HTML로 반복하여 선택하여 읽을 수 있는 경우, multipart/related는 HTML 형식의 메시지를 보내며 배경그림을 첨부했을 때 사용된다.

“Content-Transfer-Encoding”은 본문이 인코딩된 방식을 표시하는데, 한글 메시지에 8비트라고 표시되어 있으면 인코딩 없이 본문을 그대로 보낸 것이고 BASE64라든가 Quoted printable라고 적혀 있으면 그런 방식으로 인코딩 했다는 의미이다[2,3]

2.3 메일 전송 과정

다음 그림 1은 메일이 전송되는 일반적인 과정을 보여준다.

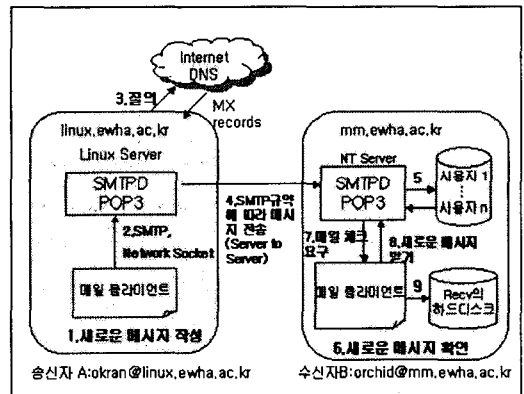


그림 1 메일 전송 과정

A가 B에게 메일을 보낸다고 가정할 때 이메일을 보내는 과정은 다음과 같다.

먼저 보내는 사람 A(okran@linux.ewha.ac.kr)가 자신의 메일 클라이언트(MUA)에서 새로운 메시지를 작성한 후, 받는 사람 필드에 수신자 B의 메일 주소인 orchid@mm.ewha.ac.kr를 입력한 후 메시지를 보낸다. 메일 클라이언트 프로그램은 메일 서버 linux.ewha.ac.kr의 25번 기본 포트로 TCP 네트워크 연결을 시도하고 메일 서버는 사용자 ID와 비밀번호를 체크한 후 접속을 인증하고 해당 메시지를 받는다. 메시지를 받은 메일 서버는 받는 사람의 이메일 주소를 보고 자신의 도메인인지 다른 도메인인지 확인을 하고 다른 도메인의 경우, 인터넷 DNS로 B의

메일 주소 중 mm.ewha.ac.kr의 IP주소가 무엇인지 질의를 던진다. 그래서 mm.ewha.ac.kr에 대한 MX 레코드를 받는다. 만약 최근에 linux.ewha.ac.kr에서 mm.ewha.ac.kr에 대한 MX레코드를 읽어 우선순위 (MX preference) 숫자가 가장 낮은 MX 레코드를 받는다. 만약 보낸 적이 없다면 A의 DNS는 다른 DNS서버에게 질의를 던질 것이다. mm.ewha.ac.kr에 대한 수신 메일 서버에 대한 이름과 노드를 찾고 A가 사용하는 SMTP 서버는 mm.ewha.ac.kr에 TCP 네트워크 연결을 한다. SMTP 규약에 의해 메시지를 전송한다. B의 SMTP 메일서버는 받는 사람의 이메일 주소를 확인하고 사용자 계정이 있는지를 확인하고 B의 메일박스에 A가 보낸 메시지를 저장한다. B의 SMTP 서버가 메시지를 완벽하게 받으면 A의 메일 서버는 네트워크 연결을 끊는다. B가 메시지를 받는 과정은 다음과 같다. B가 자신의 메일 클라이언트를 통해 새로운 메시지를 확인하고자 하면 110번 포트를 이용해 POP3 서버로 연결을 한다. POP3서버는 B의 사용자 ID와 비밀번호를 체크하고 인증을 해준다. B의 메일 클라이언트는 메일 박스에 있는 메시지의 리스트를 요구하고 전체의 메시지를 요구하고, 더 이상 새로운 메시지가 없으면 B의 메일 클라이언트는 연결을 끊는다.[4]

3. Effective Mail Grouping Strategy

3.1 수직적 유사성과 수평적 유사성

전자상거래에서 고객에게 전송되는 상업용 메일은 대부분 공통된 정보를 가지고 있으며, 고객에 따라 특정 항목의 정보만 바뀌는 특징이 있다. 예를 들어, 사용자에게 이동전화 요금 고지서를 보낸다고 할 때 고지서의 양식, 회사의 정보, 광고 등은 공통된 정보이고 사용자의 이름, 요금, 주소 등은 사용자에게 따라 달라지는 정보이다. 본 논문에서는 진자의 정보를 프로토타입(Prototype)이라 정의하며 후자의 정보를 사용자 정보라 정의한다.

프로토타입은 항상 일정한 정보를 갖는 것은 아니나 전체 내용이 전부 변하는 것은 매우 드문 일이다. 즉, 회사의 정보의 일부가 변하거나 광고 내용이 변하는 등 이전 프로토타입의 일부 내용만 변경되는 경우가 많다. 따라서 변경된 프로토타입은 이전 프로토타입과 유사성을 갖으며, 프로토타입은 시간의 흐름에 따라 변하는 정보이므로 시간축에 대해 수직적 유사성(vertical similarity)을 갖는다고 정의

할 수 있다.

사용자 정보는 프로토타입 정보에 비해 메일 전체의 일부에 지나지 않으므로 특정 회사에서 고객에게 전송하는 상업용 메일들은 서로 유사성을 갖는다고 말할 수 있다. 이러한 유사성은 메일을 전송하는 특정 시간에 대해 모든 고객 메일이 갖는 유사성이므로 본 논문은 이를 수평적 유사성(horizontal similarity)이라 정의한다.

3.2 시스템 구조

메일 서버는 보통 메일과 같이 메일 클라이언트에게 메일을 전송한다. 메일 클라이언트에서 동작하는 메일 에이전트(Mail Agent)는 프로토타입을 저장하고 메일 서버에서 전송되는 메일을 토대로 프로토타입을 변경하거나 본래의 메일을 복원한다. 따라서, 메일 서버가 전송하는 메일은 사용자 정보만 포함되므로 전체 메일 전송량은 크게 줄어들게 된다. 전체적인 시스템 구조는 다음 그림 2 와 같다.

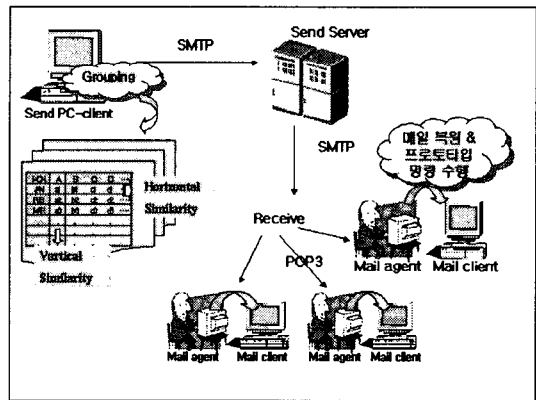


그림 2 메일 그룹핑 시스템 구조

3.3 그룹핑 전략

본 논문에서 제시한 메일 그룹핑 전략의 핵심은 프로토타입에 있다. 즉, 회사마다 고객에게 보내는 메일들은 수직적 유사성과 수평적 유사성을 가지므로 이를 하나의 프로토타입 형태로 그룹핑 할 수 있고 메일 에이전트는 클라이언트 컴퓨터에서 이 프로토타입을 관리하고 이를 이용해 메일을 복원함으로써 전체 메일 전송량을 현저하게 줄일 수 있다.

메일의 복원은 사용자 정보를 프로토타입의 항목에 설정함으로써 이루어질 수 있다. 이러한 복원을 위해 프로토타입의 항목들은 매크로 함수 형태로 저장되며 메일 서버는 매크로 함수값을 메일로 전송하게 된다. 현재 상업용 메일은 html 문서 형태로 전송되므로 이러한 매크로는 html 문서를 매크로 함수

를 사용하여 압축한 형태[5]와 비슷한 양식을 형태를 갖는다. 다음은 매크로 함수를 이용한 메일 복원의 한 예이다.

* 프로토타입 내 매크로 형태

```

요금(x, y, z) {
<table> <tr> <td> 사용액: <td> $x
<tr> <td> 세금 <td> $y
<tr> <td> 총액 <td> $z </table> }
    
```

* 메일 전송 형태

```

요금(10000, 1000, 11000)
    
```

* 복원 메일

```

<table> <tr> <td> 사용액: <td> 10000
<tr> <td> 세금 <td> 1000
<tr> <td> 총액 <td> 11000 </table>
    
```

프로토타입은 매크로의 집합으로 구성되며 프로토타입의 변경은 결국 매크로 함수의 추가, 삭제, 혹은 바디의 변경으로 볼 수 있다. 일례로, 요금 청구서에 할인액이 포함된다면 메일 서버는 다음과 같은 새로운 매크로 함수를 보냄으로써 이를 반영할 수 있다.

* 프로토타입 내 매크로 형태 (변경시)

```

요금(x, y, z, p) {
<table> <tr> <td> 사용액: <td> $x
<tr> <td> 세금 <td> $y
<tr> <td> 할인액 <td> $p
<tr> <td> 총액 <td> $z </table> }
    
```

4. 결론

인터넷이 세계적으로 확산된 이후, 많은 네트워크 서비스들 중에 웹 서비스와 더불어 가장 많이 사용되고 있는 네트워크 서비스는 전자상거래, 이메일 서비스라 할 수 있다. 하지만 기존의 이메일 시스템이 점차적으로 늘어나고 있는 전자 메일 트래픽 처리 시 부하 문제점을 보이고 있어 많은 사용자들의 요구를 만족시키지 못하고 있다. 즉, 각 사용자들이 좀 더 많은 용량의 메일을 더 많은 사람에게 송신할 수 있는 메일 처리 시스템이 개발된다면 사용자들의 요구를 만족시킬 수 있을 것이다. 본 연구는 이러한

노력의 일환으로 메일 내용의 특성을 살려 다수의 수신자를 위한 메일 그룹핑 전략(Mail Grouping Strategy)을 제안하였다. 이 연구가 전자상거래 시스템에서 활용된다면 무수한 고객의 정보 서비스를 제공할 때 효과적인 성능으로 빠른 처리를 가능하게 할 것이다.

참 고 문 헌

[1]David Wood. 채규혁 역. “인터넷 이메일 프로그래밍”, 한빛미디어, 2000년 4월.
 [2]”E-메일과워가이드“<http://dbfa.ce.pusan.ac.kr/dblab/member/yminpark/reading/Email/>
 [3]David H. Crocker, “STANDARD FOR THE FORMAT OF ARPA INTERNET TEXT MESSAGES,”<http://tearoom.chongju.ac.kr/CIE/RFC/822/index.htm>, August 1982.
 [4]”알아두면 유용한 e-mail 지식“, <http://www.amail.co.kr/email2000/concept>
 [5] 정옥란, 조동섭 “웹문서의 효율적인 전송을 위한 시스템 설계”, 한국정보과학회 춘계학술대회 논문지, 2001년 4월.