

Java를 이용한 E-Mail 수집 및 발송기의 설계 및 구현

이규명*, 서정민**, 이성종*, 이상문*,
* 충주대학교 전자계산학과
** 한경대학교 정보보안학과

A Design and Implementation of E-Mail Collecting and Sending System using by Java

Gyu Myung Lee*, Jeong Min Seo**, Seong Jong Lee*, Sang Moon, Lee*
* Dept. of Computer Science, Chungju Nat'l University
** Dept. of Information Security, Hankyong Nat'l University

요 약

인터넷의 보편화로 인하여 이메일 사용자는 계속 증가하고 있으며 이로 인해 이메일 마케팅은 주목받는 마케팅 방법의 하나로 제시되고 있다. 따라서, 이메일 수집 및 발송기의 개발은 이메일마케팅에 있어서 매우 의미있는 일이라 할 수 있다. 이에 본 연구에서는 Java를 이용한 이 메일 추출 및 분석기를 설계 구현하였다

1. 서론

인터넷 사용의 폭발적인 증가로 이메일 사용자는 계속 증가하고 있으며, 이로 인해 이메일마케팅은 최근들어 가장 주목받는 마케팅 방법의 하나로 자리를 잡아가고 있다.

이메일은 인터넷상에서 소비자와 생산자(혹은 판매자)를 연결하는 마케팅 효과가 뛰어난 수단으로, 비용이 매우 저렴하고 고객과 관련된 다양한 데이터를 실시간 수집·분석할 수 있어 인터넷 마케팅에서 가장 효과적인 기술로 각광받고 있으며 확대 구매를 이끌 수 있도록 잠재된 고객에게 관심을 일으키는 마케팅 방법이다.

인터넷 마케팅이란, 잠재고객이 원하는 바를 발견하고 그들이 원하는 제품을 생산, 공급하여 고객만족을 창출함으로써 조직의 목표를 달성하려는 접근방법[5]으로서, 마케터는 경쟁자에 비해 소비자의 욕구를 보다 더 효율적으로 충족시켜 줄 수단을 개발할 뿐만 아니라 그러한 수단에 관한 정보를 적극적으로 유포하고 바람직한 반응을 보이도록 소비자를 설득하여 장기적인 이윤을 극대화할 수 있도록 해야할 것이다.

최근에는 이러한 마케팅의 원리가 기업뿐 아니라 정치분야, 병원, 사회단체 등 비기업 조직에까지 급속히

확산되고 있다. 더욱이 마케팅의 대상도 다양한 현대 산업사회의 특성을 반영하여 단순한 제품에서부터 서비스, 정치 후보, 휴양지나 쇼펍센터, 조직, 아이디어, 활동 등을 포괄하도록 범위가 확대되고 있다. 이메일 수집 및 발송기의 구현은 마케터가 직접 이메일주소를 수집하여 추적하고 마케팅도구로 활용함으로써 스스로 이메일마케팅의 주체가 될 수 있음을 의미한다.

본 논문의 구성은 다음과 같다. 우선 2장에서 이 연구를 수행하기 위한 관련 연구를 조사·분석하고, 3장에서 시스템 및 데이터베이스를 설계하며, 4장에서 이메일 수집 및 발송기의 구현결과에 대해 기술하고, 5장에서 결론 및 향후 과제를 기술하였다.

2. 관련연구

2.1 HTTP 및 Java

◆ HTTP

HTTP(HyperText Transfer Protocol)은 월드 와이드 웹(world wide web: WWW)[1]에서 데이터를 액세스하는 데 사용되는 주된 프로토콜이다. 텍스트, 하이퍼텍스트, 오디오, 영상 등의 다양한 형식의 정보를 액세스할 수 있는데도 HyperText로 명명된 이유는, 규격이 만들어질 당시에는 Hyper Link를 통한 주로

텍스트 문서 간의 빠른 이동이 주요 이슈였기 때문이다. HTTP는 1990년부터 사용되어 왔으나 1993년에 이르러서야 HTTP 1.0이 문서로 공표되었고, 1995년에는 HTTP 1.1이 RFC 2068로 문서화되었다[6]. HTTP의 동작은 매우 간단하다. 서버 프로그램인 웹 서버는 실행되어 있는 상태로 규정된 포트에서 대기하고 있다가, 브라우저로 불리는 웹 클라이언트가 웹 서버와 TCP연결을 설정하여 데이터를 주고받는 것이 HTTP인 것이다.

웹클라이언트의 요청을 받은 웹 서버는 웹 클라이언트에게 응답을 보낸 후 TCP연결을 해제한다. 요청과 응답메시지 모두 MIME(multipurpose internet mail extension)와 비슷한 전자우편 형태를 띤다. 브라우저는 응답메시지의 몸체에 실린 HTML(hypertext markup language)[2]문서로 표현된 정보를 정의된 방식으로 처리하여 사용자에게 보여 준다. 요청종류(request type)는 대표적으로 GET, HEAD, POST 등의 방식이 있다.

◆ Java

Java[3]는 썬마이크로시스템즈(SUN Microsystems)에서 개발한 객체 지향 프로그래밍 언어로서, 1995년 5월에 발표, 1996년 1월부터 정식 버전의 배포를 개시하였다. C++를 바탕으로 언어 규격을 규정하였다. 버그(bug)의 원인이 되기 쉬운 기능, 예를 들면 포인터 연산을 생략하였다. 또한 C++에서는 사용이 끝난 객체(object)를 명시적으로 폐기하도록 프로그램에 기술하였으나 그 대신 자동 쓰레기 수집(garbage collection) 기능을 추가하였다. 자바는 객체 클래스의 계승(inheritance) 관계를 실행할 때에 확정하는 동적 모델을 채용하였다. 자바에서는 클래스의 계승 관계를 실행할 때 확정하기 때문에, 하위의 클래스를 재컴파일할 필요가 없다. 자바로 작성한 프로그램의 원시 코드는 자바 컴파일러로 컴파일한다. 자바 컴파일러는 바이트 코드라는 중간 코드를 생성한다. 이 중간 코드를 자바 가상 머신이라는 소프트웨어로 해석해서 실행한다. 이 자바 가상머신을 설치한 컴퓨터라면 컴퓨터의 명령 집합 아키텍처나 운영 체제에 관계없이 실행할 수 있다. 자바의 원시 코드를 고쳐 쓰거나 재컴파일할 필요가 없다. 그렇기 때문에 자바는 기종이나 운영 체제와 무관한 응용 프로그램의 개발 도구로 각광받고 있다.

2.2 파서(Parser)

파서(parser)는 컴파일러에게 들어온 문서나 소스코드가 규정된 문법에 맞게 씌어졌는지를 검사하는 부

분이다. 일반적인 파싱(parsing)은 크게 어휘 분석(lexical analysis)과정과 구문 분석(syntactic analysis)과정으로 나누어진다[7].

어휘 분석은 입력된 일련의 문자들을 토큰(token)이란 논리적 단위로 구분하는 과정이다. 어휘 분석기(lexical analyzer)에서 일련의 문자들이 토큰으로 인식되면, 인식된 토큰과 관련 정보를 다음 단계의 구문 분석기에 전달한다. 토큰은 의미 정보(semantic information)를 갖는 최소의 구문적 단위이다. 구문 분석기(syntactic analyzer)는 어휘 분석기가 공급하는 토큰들을 검사하고, 중간 형태(intermediate form)로 변환하여 인지된 문장을 효율적으로 처리한다[8].

Java는 키보드로 입력받아서 화면에 출력하는 작업, 파일로부터 데이터를 읽어서 다시 파일에 쓰는 작업, 네트워크를 통해 데이터를 송수신하는 작업 등의 모든 입출력을 스트림(stream)이라는 추상적인 개념과 java.io 패키지에 있는 클래스를 이용해 쉽게 처리할 수 있다. Java에서의 스트림(stream)은 물리적인 디스크상의 파일과 각종 장치(키보드, 화면, 통신포트, 프린터 등)를 하나의 통일된 방식으로 다루기 위한 추상화된 논리적 장치를 의미하는데, 이렇게 함으로써 데이터 입출력에 상당한 융통성과 효율성을 준다[9].

2.3 SMTP

인터넷 서비스 중 가장 큰 비중을 차지하고 있는 것 중의 하나가 바로 이메일 서비스이다. 인터넷에서는 메시지를 송신하기 위해 SMTP 프로토콜을 사용한다[4]. SMTP란 Simple mail Transfer protocol의 약자로서, 메시지를 교환하기 위한 프로토콜 규약이다. 그림 1은 SMTP를 이용한 메일의 전송과정을 나타낸다.

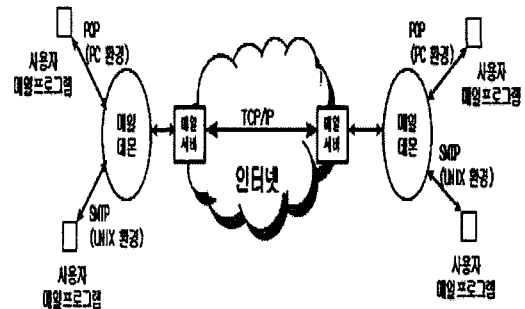


그림 1 SMTP 서버의 구조 및 사용

사용자는 사용자 에이전트(user agent: UA)를 통해서 메일을 작성한 후, SMTP를 사용하여 Mail Daemon으로 메시지를 전송한다. Mail Daemon은 종단간 클라이언트의 주소를 분석하고 가장 가까운 메일 서버로 메시

지와 정보를 보낸다. 송신자가 보낸 편지가 일단 수신자 측의 전자우편을 관리하는 메일 서버에 전달되면, 메일 서버는 수신자의 전자우편 주소를 분석해서 최단 경로를 찾아 근접한 메일 서버에 편지를 전달하게 된다. 최종 수신자 측의 메일 서버에 도착하기까지 연속적으로 전달하는 중계작업이 계속된다. 사용자가 UA를 통해 수신된 메일을 확인하기 위한 프로토콜로 POP(post office protocol) 또는 IMAP(internet mail access protocol)가 사용된다[10].

3. 시스템 설계

3.1 프로그램 동작원리

본 이메일 주소 수집기의 동작 원리는 우선, 주 URL 주소를 입력하면, 그곳의 HTML 소스를 읽어 들여 링크되어 있는 URL 주소를 따라 계속 html 소스를 읽어 들이면서 그 페이지에 나타난 이메일 주소를 추출한다.

3.2 메일주소 수집기

- (1) 분류 생성, 수정 및 삭제 : 잠재고객의 성격에 맞는 마케팅을 위해 관심 분야별로 분류를 생성하여, 분류에 연관된 URL을 관리한다.
- (2) 파일 생성 : 초기 주 URL을 통해 수집한 이메일주소를 주어진 파일로 생성 한다.
- (3) URL 추가 : 생성된 분류에 연관된 URL을 저장할 수 있다.

3.3 이메일 발송기

- (1) 제목 및 발송자 정보 입력 : 발송할 이메일의 제목과 발송자의 이름 및 회신주소를 입력한다.
- (2) 메일 내용의 저장, 수정 및 삭제 : 발송할 이메일의 내용을 작성하여 DB에 저장하고, 저장된 mail 내용을 불러와서 편집할 수 있다.
- (3) SMTP 서버 설정 : SMTP 메일 서버를 선택할 수 있다.
- (4) 수집된 이메일주소 선택 : 이메일주소 수집기에서 수집된 Mail 주소가 저장된 파일 이름을 입력하여 저장되어 있는 모든 Mail 주소에 이메일을 발송한다.

3.4 DB Schema

데이터베이스관리 서버(DBMS)로는 마이크로소프트사의 SQL_Server 2000을 사용하였다. 테이블의 구조는 표 1과 같다.

테이블명	내 용
menu	분류이름 저장
menu2	분류이름에 관련된 사이트 저장
mailto	전송할 메일 관련 사항 저장

표 1 기본 테이블 구조

각각의 테이블은 표 2와 같이 정의하였다.

테이블명	컬럼명	자료형	내 용
menu	no	int(기본키)	분류명의 인덱스번호
	name	varchar	분류명
menu2	no	int(기본키)	site의 인덱스번호
	thread	int(참조키)	분류명 인덱스번호
mailto	site	varchar	분류명
	no	int(기본키)	Mail내용 인덱스번호
	name	varchar	발송자 이름
	addr	varchar	발송자 Mail 주소
	content	varchar	발송 Mail의 제목
	memo	varchar	발송 Mail의 내용

표 2 테이블별 속성

4. 구현 및 결과

4.1 메일주소 수집기

그림 2는 이메일주소 수집기의 메인화면으로, 특정 분류를 생성, 수정 및 삭제가 가능하며 초기 주 URL을 특정분류에 연관시켜 저장 및 불러올 수 있어, 분류이름과 연결되어 있는 URL주소가 분류리스트에 나타난다.

메일주소 수집 버튼을 클릭하면 웹브라우저와 같은 방법으로 웹서버에 자료를 요청해 html 문서를 전달 받게 되는데, 'mailto'나 '@' 등 이메일주소와 관련된 부분만 별도로 구분하여 수집한다.

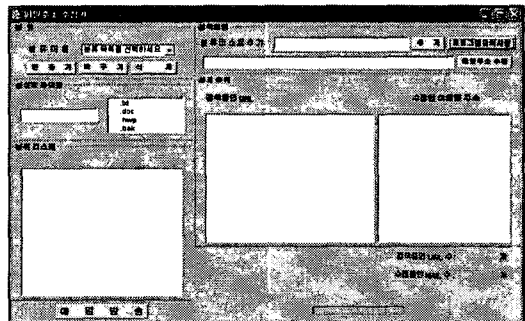


그림 2 메인화면

그림 3은 프로그램 실행이 완료된 화면으로, 수집한 이메일주소를 파일로 생성할 수 있다. 파일 형식은 txt, doc, hwp, bak로 제한되어 있다.

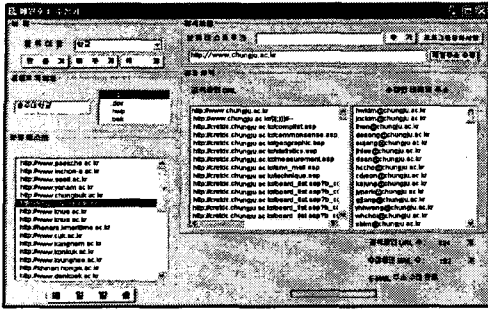


그림 3 프로그램 실행화면

4.2 이메일 발송기

SMTP 서버로 메일을 발송하기 위해서는 메일 제목, 발송자의 메일 주소, 수신자의 메일 주소, 내용이 포함되어 있어야 한다. 그림 4는 메일 발송기의 실행 준비화면이다. SMTP 서버를 선택할 수 있으며(현재는 충주대학교 SMTP 서버를 사용), 수집된 메일주소가 저장되어 있는 파일이름을 입력하여, 파일에 있는 모든 메일 주소에 송신할 수 있다.

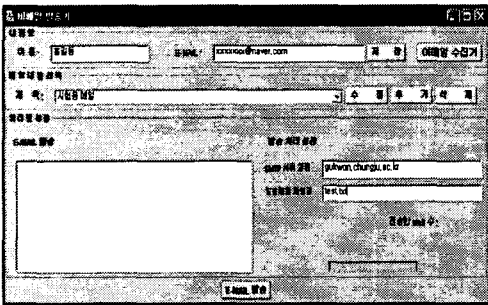


그림 4 메일 발송기 실행 준비화면

그림 5는 메일 발송기 실행화면으로, 발송한 메일의 전송상태를 보여준다.

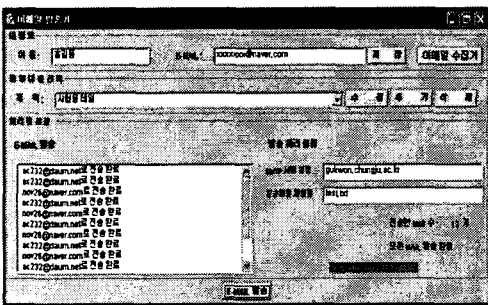


그림 5 메일 발송기 실행화면

5. 결론 및 향후 과제

지금까지 자바를 이용한 이메일주소 수집 및 발송

기의 개발과정에 대해 살펴보았다. 본 논문에서 자바를 이용한 이유는 2장에서 기술한 바와 같이 인터넷 상에서 시스템을 구현하는데 있어 가장 적합하며, 또한 하드웨어 독립적으로 실행이 가능해 어떤 시스템에서도 그대로 사용할 수 있다는 유용성이 뛰어나기 때문이다. 그러나 현재는 웹페이지 요청(request type)을 GET 방식을 사용할 경우 게시판 및 공지사항에 연결되어 있는 이메일주소 수집이 가능하게 되는데, POST 방식을 사용하거나, 자바스크립트를 이용하여 페이지를 넘긴 경우 URL 주소를 읽어올 수 없어 이메일 주소를 수집할 수 없다. 또한, 정보통신 관련법상 무단으로 이메일 주소를 수집하는 행위는 불법이기 때문에 이메일 주소 수집 거부 의사가 표명된 사이트를 선별해야 할 것이다.

향후에는 정보검색 기법과 접목하여 검색어 입력으로 분류를 자동 생성하도록 하여 편리성을 높여야 할 것이다.

[참고문헌]

- [1] Berners-Lee, T., R. Cailliau, A. Loutonen, H. F. Nielsen and A. Secret, "The World-Wide Web," Communications of the ACM, v.37, n.8, pp. 76-82, August 1994.
- [2] Connolly, DW., Hypertext Markup Language, World Wide Web Consortium, Boston, 1995; <http://www.w3.org/hypertext/WWW/MarkUp/MarkUp.html>
- [3] Ken Arnold, James Gosling, 'The Java Programming Language', Addison Wesley, 1996.
- [4] Gary Govanus, 'TCP/IP 24seven', SYBEX, 1999.
- [5] 유동근, "통합마켓팅", 미래원, 1997.
- [6] 차동완, 정용주, 윤문길, "인터넷 기술세계", 교보문고, 2002.
- [7] 이기철, 표창우, 변영태, "시스템 소프트웨어", 도서출판 생능, 1994
- [8] 김병철, 박찬영, 심영철, 이재광, 이재훈, 홍충선, "TCP/IP 프로토콜", 도서출판 MRC 미래컴, 2000.