

사전 DB를 이용한 웹 메일 문서 생성기

최지연*, 민수홍, 조동섭

*이화여자대학교 과학기술대학원 컴퓨터학과

e-mail : gratel@ewha.ac.kr

Design and Implementation of Web Mail Document Generator using Dictionary DB

Ji-Yeon Choi*, Su-Hong Min, Dong-Sub Cho*

*Dept. of Computer Engineering & Science, Ewha Womans University

요 약

WWW(World Wide Web)은 인터넷 상에서 광범위하게 분산된 다양한 형식의 데이터를 손쉽게 검색할 수 있도록 하며, 모든 응용의 사용자 환경을 손쉽게 포함시킴으로서 새로운 미디어 타입의 통합을 위한 가장 성공적인 기술로 평가받고 있다. E-Mail 서비스는 WWW 시스템의 가장 기본적인 기능으로, 인터넷 기술이 발전하고 사용자가 기하급수적으로 증가함에 따라 E-Mail 사용자도 늘고있다. 그러나 기존의 서비스는 HTML의 텍스트 기반 구조를 통한 비동기적인 형태의 서비스를 계속 유지하고 있어, 최근 인터넷상의 상당수를 차지하는 오디오, 비디오와 같은 실시간 환경의 멀티미디어 데이터들을 통합하는 좀더 동기적인 표현방법이 요구되고 있다. 본 논문에서는 E-Mail 서비스에 대한 동기적 방법론을 제안하고 구현하고자 한다.

1. 서론

WWW이란 거대한 인터넷상에 흩어져있는 정보들을 효율적으로 표현하고 접근할 수 있도록 설계된 분산 하이퍼미디어 시스템으로 새로운 미디어 타입을 통합하기 위한 가장 성공적인 기술로 평가받고 있다 [1]. WWW에서 제공되는 정보는 하이퍼텍스트(hypertext) 형태로 표현된다. 이러한 WWW의 텍스트 기반 구조에 맞추어 HTTP(HyperText Transfer Protocol), HTML(HyperText Markup Language), URL(Uniform Resource Locator)이 제안되어 이를 통해 서버와 클라이언트간의 비동기적인 형태의 검색과 상호작용을 지원하도록 설계되었다. HTML은 1990년 이후 웹 상에서 가장 많이 사용되는 문서 교환의 표준 형식으로 1986년에 제정된(ISO 8879) SGML(Standard Generalized Markup Language)를 바탕으로 정의된 하나의 응용이다. HTML은 구조적 특성을 갖는 이유로 손쉬운 편집이 어려워 웹이 전 세계 이용자들의 각광을 받으면서 HTML을 이용한 문서 편집이 활발해지고 빠르고 간편한 방식의 HTML 편집기도 나오게 되었다.

그러나 HTML은 텍스트 기반의 구조를 통한 비동

기적인 형태의 서비스 유형으로 문서 검색과 단순한 표현 방식을 제공한다. 따라서 다양한 형태의 동기적인 데이터 및 네트워크 자원을 효과적으로 교환 및 검색하기에는 한계를 가지고 있다. 특히 WWW의 대표적인 기능이라 할 수 있는 E-Mail 서비스는 대부분의 인터넷 사용자들이 사용하는 기본적인 기능임에도 불구하고 단순한 텍스트 기반의 표현 방식을 사용하고 있으며, 모든 사용자들이 HTML tag를 알고 있다는 전제 하에 제한된 동기적인 서비스를 제공하고 있다. 그러므로 E-Mail 사용자가 보내고자하는 메일의 본문을 편집할 때, 사용자가 HTML의 tag에 대한 지식이 전혀 없을 경우, 사용자는 단순한 텍스트 형식의 메일만을 전송할 수 있다. 이러한 e-Mail 사용은 메일 수신자에게 효과적인 정보를 제공하는 측면에서는 많은 문제점을 갖고 있다. 이미 알고 있거나 기본적으로 제공될 수 있는 정보들을 사전식으로 등록하여 자동 검색으로 이 문제를 해결 하는 것이 좋다.

그러므로, 본 논문에서는 효과적인 e-Mail 사용을 위해, 사용자가 HTML tag에 대한 지식이 전혀 없어도 표현하고자 하는 개체에 대한 정보가 사전에 미리 등록되어 있다면 e-Mail의 내용을 작성할 때, 표현하고자 하는 텍스트뿐만 아니라 미디어에 대한 정보가 자

이 논문은 2003년도 두뇌한국 21 사업에 의하여 지원되었음.

동으로 삼입되어 효율적으로 문서가 생성되도록 한다. 본 논문은 2 장에서는 관련 연구에 대해서 기술하며, 3 장에서는 제안하는 메일 서비스에 대해 설명한다. 4 장에서는 본 논문의 결론을 맺고, 향후 연구 계획에 대해 언급한다.

2. 관련연구

2.1 하이퍼미디어 / 하이퍼 텍스트 시스템

하이퍼텍스트는 데이터의 내용을 기반으로 하는 데이터 항해(navigation)의 효과적 방법을 지원하는 제어 구조(Control Structure)를 제공한다. 반면 멀티미디어는 정보를 표현하는데 있어서 다양함을 가능하게 하는 풍부한 데이터 타입을 제공한다. 그러나 정적 정보의 결합에 적용되는 개념이 동적 정보의 결합에 완전히 적용되지는 않는다[2]. 하이퍼텍스트 시스템은 정보의 상호연관성을 중심으로 구조화되어 있고 이들을 비선형적으로 접근하도록 되어 있기 때문에, 하이퍼텍스트 링크를 동적으로 관리하는 것이 결코 쉬운 일이 아니다. 종래의 정보검색은 키워드를 얼마나 잘 관리하느냐에 따라 탐색의 성공여부가 결정되었지만, 하이퍼텍스트 시스템에서는 키워드의 변화를 내부적으로 관리해야 할 뿐만 아니라, 시간적으로 보여주어야 하며, 상호 연결된 이웃 자료들에 대한 링크를 수정하여야 하기 때문이다[3].

하이퍼미디어(Hypermedia)는 하이퍼텍스트와 멀티미디어의 합성어이다. 여기서 멀티미디어란 텍스트, 음성, 이미지, 동영상 등의 모노미디어들의 결합을 의미하는데, 이들 모노미디어는 시간적인 특성상 정적 미디어와 동적 미디어로 나눌 수 있다. 정적 미디어는 시간의 변화에 따라 내용의 변화가 없는 모노미디어를 의미하며, 동적 미디어는 시간의 변화가 미디어의 출력에 영향을 미치는 모노미디어들을 의미한다. 정적 미디어의 예로는 텍스트나 이미지 등을 들 수 있고, 동적 미디어의 예로는 동영상이나 음성등을 들 수 있다.

하이퍼미디어 시스템은 비순차적 이라는 특성을 갖는다. 어떤 정해진 순서가 있는 것이 아니라 사용자가 관심을 갖는 부분을 통하여 정보를 검색할 수 있도록 만들어진다[4].

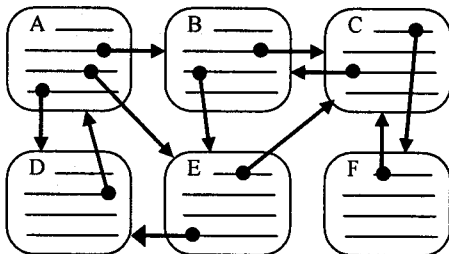


그림 1. 하이퍼미디어 시스템의 비순차적 특성

2.2 메일의 전송과정

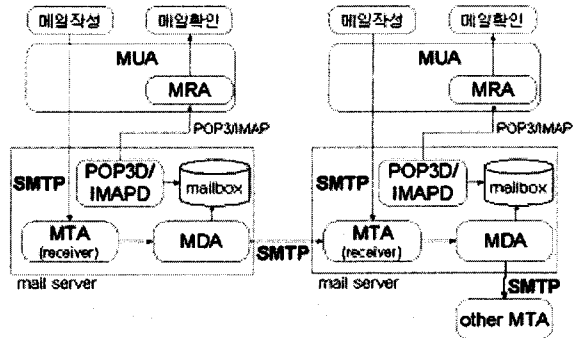


그림 2. 메일의 전송 과정

메일을 전송하기 위해서는 사용자가 일반적으로 MUA (Mail User Agent)에 통합되어 있는 메일 작성 프로그램을 사용하거나, 직접 메일 서버의 MUA 에서 메일을 작성해야 한다[1]. MUA 는 사용자가 메일을 송수신할 때 사용하는 클라이언트 프로그램으로, 마이크로소프트사의 아웃룩 익스프레스나 리눅스, 유닉스에서 사용하는 mail 프로그램을 예로 들 수 있다. MUA 는 사용자가 작성한 메일을 메일서버의 MTA(Mail Transfer Agent)로 전송한다. MTA 는 인터넷상에 있는 하나의 컴퓨터로부터 다른 컴퓨터(수신자측의 메일 서버)로 메일을 전달하는 프로그램이다. MTA 는 수신된 메일의 수신자가 로컬 메일 서버의 사용자인지, 아니면 다른 메일 서버의 사용자인지를 판단한다. 만약 로컬 사용자인 경우에는 해당 MDA(Mail Delivery Agent)를 통해서 로컬 파일시스템에 있는 사용자의 우편함에 메일을 저장한다. MDA 는 메시지를 사용자의 우편함에 쓰기위해 MTA 가 사용하는 프로그램이다. 그리고 로컬 사용자가 아닌 경우에는 별도의 MDA 를 사용해서 다른 MTA 로 메일을 전달한다. 메일을 수신받은 수신자의 MTA 는 다시 로컬 메일서버의 사용자인지를 판단한 후에, MDA 를 통해서 파일 시스템의 사용자 우편함에 메일을 저장한다. MTA 와 MDA 는 sendmail[5]이나 qmail[6]처럼 일반적으로 하나의 메일 서버 어플리케이션에 통합되어 있다. 사용자 MUA 에서 메일 서버로 메일을 전송할 때와, 메일의 수신자가 로컬 사용자가 아니어서 다른 메일 서버에 있는 MTA 로 메일을 전송할 때 SMTP[7]를 사용한다.

사용자가 자신의 메일을 확인하기 위해서는 MUA 에 포함되어 있는 MRA(Mail Retrieval Agent)를 사용해야 한다. MRA 는 원격지 서버에 있는 우편함으로부터 사용자의 MUA 로 메시지를 가져오는 서비스를 제공하는 프로그램이다. MRA 가 메일 서버에 있는 POP3 때문에 접속하면, POP3 때문에 사용자의 우편함에서 메일을 읽어서, MRA 에게 전달한다. 사용자의 우편함으로부터 메일을 읽어올 때는 POP3 를 주로 사용하지만,

요즘에는 많은 네트워크 지향적인 옵션을 제공하는 IMAP(Internet Message Access Protocol)의 사용이 선호되고 있다.

2.3 전자우편 시스템

전자우편 클라이언트 프로그램으로 SMTP 서버에 편지를 보낸다. 전자우편 메시지 수신인의 도메인이 서버가 있는 도메인과 같으면 아스키 텍스트를 서버에 저장하고, 자기 도메인이 아닌 경우 해당 서버에 전달한다. 전자우편 메시지가 수신자의 SMTP 서버에 도착하면 수신자에 할당된 사서함에 저장된다. 수신자는 아웃룩 익스프레스 같은 우편 클라이언트에서 POP3 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. 그림 3 은 전자우편 배달과정을 보여준다.

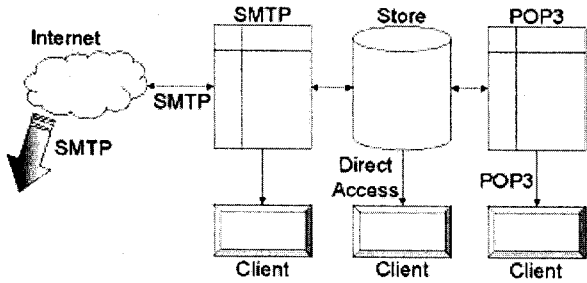


그림 3. 전자우편의 전송

3. 사전 DB 를 이용한 웹 메일 문서 생성기

웹 문서에서 가장 널리 사용되는 HTML 은 HyperText 형식의 문서로서 일반적인 문서에 비해 사용자에게 보다 더 많은 표현의 영역을 제공한다. 이를 통해 검색자가 원하는 내용을 보다 더 효율적이고 알아보기 쉽게 표현할 수 있다. 예를 들어 단어에 색이 들어간 경우 사람들은 그 단어를 다른 단어보다 더 강하게 인식하게 된다. 이러한 기능은 단어를 색으로 표현함으로써 단어의 중요도가 차별화 될 수 있도록 한다.

기존의 메일 서비스는 HTML 의 비동기적인 형태의 서비스이며 사용자가 HTML tag 에 대한 지식을 갖고 있다는 전제 하에서 동적인 서비스를 제공하고 있다. 그러므로 HTML tag 에 대한 지식이 전혀 없는 사용자들은, 인터넷의 상당부분을 차지하는 오디오, 비디오와 같은 연속적이고 동기적인 멀티미디어 데이터들을 메일 본문에 삽입하고자 할 때, 큰 불편함을 느끼게 된다.

따라서 HTML tag 에 대한 지식이 전혀 없는 사용자들도 웹 문서에 원하는 정보를 삽입시킬 수

있도록하며, 일반 사용자들도 보다 편리하게 웹 문서를 다룰 수 있도록 하는 동기적인 형태의 메일 서비스가 필요하다.

웹 메일 문서를 생성하기 위해서는, 먼저 개체들을 각각의 성격이나 형식 등을 기준으로 분류하여야 하며, 이때 분류된 각 개체들은 사용자에게 의해 사전으로 저장된다. 사용자가 메일을 작성하여 보내면, 사용자가 작성한 문서는 전송되기 이전에 사용자가 설정해 놓은 사전을 거친다. 사전을 통해 HTML 문서는 재생산되어 목적지로 전송된다.

3.1 전체 구조

전체적인 구성을 살펴보면, 먼저 각 단어들을 그룹별로 나누고 원래의 문서를 수정하는 전처리 과정과 색을 변경하고 하이퍼링크를 생성하는 등 사용자가 미리 생성해놓은 Dictionary DB 의 내용을 바탕으로 Text 형식의 문서를 HTML 형식의 문서로 변환시키는 과정으로 나뉘어 진다. 사용자가 사전에 등록한 정보는 Dictionary DB 에 저장되며, 메일이 수신되면 이 DB 에 저장된 정보를 바탕으로 하여 문서가 재생산된다.

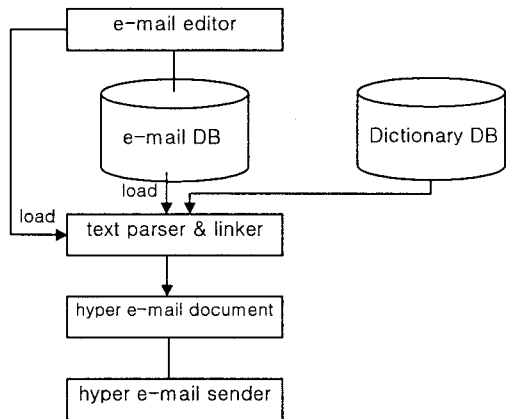


그림 4. 사전기반 hyper e-mail document 응용 구조

전처리 과정에서는 메일 내용을 읽어들이 각 단어가 어떤 특성을 갖는 부류에 속하는지를 파악한다. 이런 과정을 거쳐 분류된 각 개체들을 그룹별로 나누어 원래의 문서를 수정하고 각 그룹에 색을 지정하고 개체마다 링크를 생성하는 등, DB 에 저장된 정보를 자동으로 삽입시킨다. 또한, 사전에 등록된 개체들 위에 마우스가 위치할 때 각 개체마다 어떤 정보를 담고 있는지를 표시하는 Hyper 문서를 생성한다. 다음의 그림은 새로운 이메일이 도착했을 때, 사용자가 설정해놓은 Dictionary DB 를 바탕으로 이메일의 내용이 자동으로 변환되어 새로운 Hyper e-mail 문서를 생성하는 과정을 나타낸 것이다.

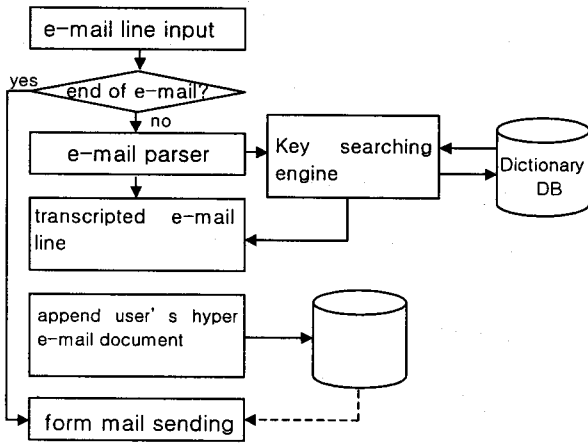


그림 5. Hyper e-Mail 생성과정

3.2 TXT2HTML 전처리의 구성

사용자는 메일을 일반 텍스트 형식으로 작성한다. 사용자가 메일을 전송할 것인지를 결정하면, 메일은 사용자가 미리 만들어 둔 Dictionary DB 를 거치게 된다. 텍스트 문서는 이 Dictionary DB 내의 정보를 바탕으로 텍스트 문서의 내용과 Dictionary DB 안의 목록이 일치하는 것이 있는지 찾는다. 일치하는 정보가 검색되면 DB 에 저장된 목록들의 정보들은 사용자가 작성한 텍스트 메일에 첨부하여 새로운 HTML 문서를 생성한다. 다음의 그림은 이러한 과정을 거쳐 HTML 문서가 생성되는 과정을 도식화 한 것이다.

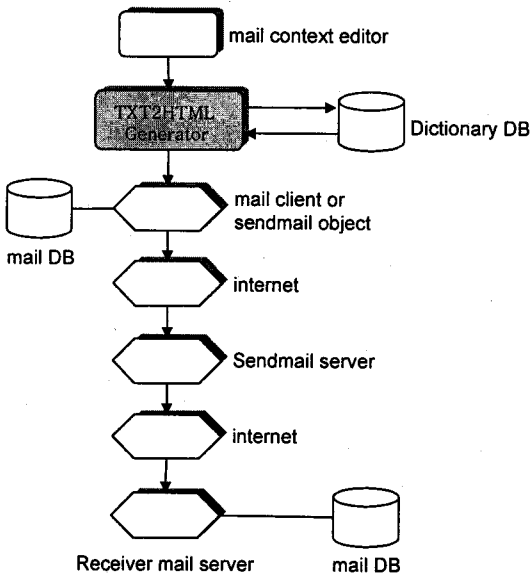


그림 6. TXT2HTML Generator

4. 결론

본 논문에서 제안하는 사전기반 웹 문서 생성기는 텍스트 형식의 문서뿐만 아니라, 모든 형식의 문서를 읽어서 사용자가 생성한 Dictionary DB 에 저장된 정보와 일치하는 정보가 있을 때 기존의 문서에 새로운 내용을 추가하여 HTML 형식의 문서로 바꿔주는 기능을 한다. 이를 이용하여, 사용자는 HTML tag 에 관한 지식이 전혀 없이도 보내고자 하는 메일 내용에 색 기능과 하이퍼링크 기능 등을 추가시킬 수 있으며 더 나아가 미디어나 오디오와 같은 동적인 형태의 멀티미디어 기능들도 삽입시킬 수 있다. 기존 문서상의 키워드에 관련된 다양한 정보들을 자동으로 연결시켜 웹 문서 작성이 편리하도록 하며, 새로운 통합 정보를 구축하는데 용이하도록 한다.

향후 연구과제로는 현재 텍스트 단계에까지만 그친 웹 문서 생성기의 기능을 소리, 동영상 및 가상현실과 같은 정보들과도 연결시켜, 멀티미디어 개체들을 표현할 수 있도록 하이퍼 미디어 기능을 지원하는 메일 시스템을 구축하는 데에 있다.

참고문헌

- [1]w3 team at CERN, The WWW Book, 1995
- [2]이동애, 장덕성, "동적 정보 저장을 위한 자동 하이퍼텍스트 색인 기법의 개발", 한국정보처리학회 논문지 A 1997.09
- [3]은성배, 윤덕호, 윤현주, 윤현수, "하이퍼미디어 시스템에서의 동영상 앵커 처리", 멀티미디어 연구논문 JOM 제 1 권 제 1 호 1999.12
- [4]Athula G., David B.L., and Hohn R, "Hypermedia Authoring" IEEE MultiMedia, Winter 1995
- [5]sendmail 18.11.0, <http://www.sendmail.org>
- [6]Dan Bernstein, qmail-1.03, <http://qmail.org>
- [7]Jonathan B.Postel, "Simple Mail Transfer Protocol" in RFC821, August 1982