

# 무선 인터넷을 위한 HTML-WML 변환기 설계 및 구현

## Design and Implementation of the HTML-WML Converter

민 영 수\*      강 형 일\*\*      유 재 수\*\*\*  
Young-Soo Min      Hyung-Il Kang      Jae-Soo Yoo

### 요 약

웹 상에 존재하는 HTML로 기술된 방대하고 다양한 정보를 휴대용 무선 장비를 통해 접근할 수 있도록 하려면 HTML로 기술된 문서와 동일한 별도의 WML 문서를 작성해야한다. HTML로 구축된 정보의 양이 방대할 경우, 동일한 정보를 가지는 별도의 WML 사이트를 구축하는 것은 시간적, 공간적으로 많은 비용이 든다. 본 논문에서는 기존의 HTML로 기술된 웹 문서를 휴대용 무선 장비를 이용하여 접근하려고 할 때, 휴대용 무선 장비에 적합한 WML 문서로 변환해주는 HTML-WML 변환기를 설계하고 구현하였다. 구현된 HTML-WML 변환기는 휴대용 무선 장비에 따라 동적으로 WML 문서를 재구성하고 GIF, JPG, BMP 등과 같은 다양한 이미지 형태를 처리할 수 있는 장점을 가졌으며 무선 인터넷 환경에서의 실시간 변환기 뿐만 아니라 WML 에디터와 같은 응용에서도 사용이 가능하다.

### Abstract

To access massive and various HTML documents that are in the web using wireless Internet equipments, another WML document that is equal to the HTML document must be written. In the case Web documents written by HTML are massive, the construction of a WML site with the same information needs much cost of space and time. This paper designs and implements the HTML-WML converter that alleviates such a problem. The HTML-WML converter translates the Web document written by HTML to the WML document for portable wireless equipments. The HTML-WML converter has advantages that it reconstructs WML document dynamically according to portable wireless equipments and processes various image formats such as GIF, JPG, BMP, and so on. The HTML-WML converter can be used as not only a utility of the WML editor but also a real-time converter on wireless Internet.

## 1. 서 론

최근 무선 데이터통신 기술의 발달 및 핸드폰, PDA와 같은 휴대용 무선 장비 수요의 폭발적인 증가에 따라 휴대용 무선 장비를 통해 웹 상의 정보를 이용하고자 하는 요구가 점점 늘어가고 있으며, 휴대용 무선 장비를 통한 웹 상의 정보 접근이 점차 현실화되고 있다[3][7]. 그러나, 웹 문서 기술 언어인 HTML(Hyper Text Markup Language)로 기술된 기존의 문서들은 적은 메모리 양과 작

은 디스플레이 화면, 느린 CPU 속도 등과 같은 제한된 환경의 휴대용 무선 장비를 통해 접근하기에는 많은 무리가 있다. 따라서, 이러한 HTML의 대안으로 휴대용 무선 장비를 통한 효과적인 웹 상의 정보 접근을 위해 HDML(Handheld Markup Language)이 Unwired Planet에 의해 개발되었으며 [1], 현재는 Phone.com, Ericson, Nokia, Motorola 등이 WAP(Wireless Application Protocol) 포럼을 결성하여 무선 응용 프로토콜의 표준을 추진하고 있다[2][3].

WAP에서는 기존의 웹 문서 기술 언어인 HTML을 대신하여 무선망 환경에 맞게 최적화된 HDML 기반의 WML(Wireless Markup Language)를 사용한 대[4][5]. WML은 XML(eXtensible Markup Language)에 기반을 두고 설계된 언어로 HTML과 마찬가지로

\* 비회원 : 충북대학교 대학원 정보통신학과 박사과정  
minys@netdb.chungbuk.ac.kr

\*\* 비회원 : 주성대학 멀티미디어 정보통신공학부 전임강사  
khi69@jsc.ac.kr

\*\*\* 정회원 : 충북대학교 정보통신공학과 부교수  
yjs@chucc.chungbuk.ac.kr

지로 태그(tag)기반으로 이루어진 언어이다. 클라이언트에서는 HTML의 경우와 유사하게 브라우저가 WML을 해석하여 그 내용을 휴대용 무선 장비 화면에 출력하게 된다. 다만, PC에서 사용되는 브라우저와 구별하여 휴대용 무선 장비에서 사용되는 브라우저를 마이크로 브라우저(Micro Browser)라고 한다.

WAP에서는 WML을 사용하기 때문에, HTML로 작성된 기존의 웹 문서들을 사용할 수 없다. 즉, WAP 방식을 사용하여 접속할 수 있는 사이트(site)들은 오직 WML로 작성된 정보를 제공하는 사이트 뿐이다. 만일 HTML을 사용하여 구축된 사이트를 운영하고 있는 업체에서 휴대용 무선 장비를 통해서도 이 사이트를 접속할 수 있도록 지원하려면 HTML을 사용하여 구축된 정보와 동일한 정보를 갖도록 별도의 WML 사이트를 구축해야 한다. HTML로 구축된 정보의 양이 방대할 경우, 동일한 정보를 가지는 별도의 WML 사이트를 구축하는 것은 시간적, 공간적으로 많은 비용이 든다. 더욱이 HTML 사이트의 정보가 변경될 경우, 이를 WML 사이트에 반영하는 문제나 HTML 문서의 내용을 WML로 바꾸기 위한 일관성 있는 규칙의 부재 같은 문제는 WML 사이트를 구축하고 관리하는 일을 더욱 어렵게 만든다 [6,7,8]. 따라서, 이러한 문제를 해결하기 위해 HTML로 기술된 웹 문서를 적절한 형태의 WML 문서로 변환하는 변환기가 필요하다.

본 논문에서는 HTML로 기술된 기존의 웹 문서를 그대로 유지하면서 휴대용 무선 장비를 이용하여 웹 문서에 접근하려고 할 때, HTML로 기

술된 기존의 웹 문서를 휴대용 무선 장비에 적합한 WML 문서로 변환해주는 HTML-WML 변환기를 설계하고 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로써 국내 무선통신회사들이 사용하고 있는 변환기의 동향과 상용화된 HTML-WML 변환기에 대해 살펴보고 기존 웹사이트들의 초기화면 구성 형태와 메뉴 구성 형태를 분석한다. 또한, WAP 환경에서 HTML-WML 변환기가 어디에 적용될 수 있는 위치를 기술한다. 3장에서는 HTML-WML 변환기의 설계와 구현 사항들을 기술하고 4장에서는 간단한 예를 통해 구현한 HTML-WML 변환기의 수행 결과를 나타내고 다른 HTML-WML 변환기와 비교를 통해 구현한 HTML-WML 변환기를 평가한다. 마지막 5장에서는 결론을 맺고 향후 연구 방향을 제시한다.

## 2. 관련연구

최근 휴대용 무선 장비를 통한 인터넷 서비스를 지원하는 몇몇 무선통신회사들은 WML로 작성된 웹사이트를 구축하여 정보를 제공하거나 변환기를 통해 정보를 제공하고 있다. 표 1에서도 볼 수 있는 것처럼 국내 무선통신회사들이 사용하는 변환기는 WML을 지원하는 것이 적고, 실시간 변환을 제공하지만 이미지 변환이 되지 않는 등 초보적인 단계에 머물러 있다.

또한, Spyglass사의 Prism은 HTML로 기술된 웹 문서를 휴대용 무선 장비의 메모리 용량과 디스플레이 화면 크기에 최적화된 WML 문서로 변환

(표 1) 국내에서 개발된 변환기 현황

	SK 텔레콤	한국통신 프리텔	LG 텔레콤	신세기 통신	한솔 엠닷컴
프로토콜	WAP	WIP	WAP	WAP	WIP
언어	WML	M-HTML	HDML	HDML	M-HTML
변환기 특징	· 실시간 변환 제공하지 않음	· 실시간 변환 제공 · 초보적인 단계 · 이미지 변환 안됨	· 실시간 변환 제공하지 않음	· 실시간 변환 제공 · 초보적인 단계 · 이미지 변환 안됨	· 실시간 변환 제공 · 초보적인 단계 · 이미지 변환 안됨

(표 2) 기존 사이트들의 초기화면 구성 형태

문서의 형태	사용율
프레임	67.9 % (87 사이트)
테이블	48.3 % (72 사이트)

하기 위한 변환 루틴을 사용한다. 이 변환기의 특징은 기본적으로 HTML 태그들을 WML 태그들로 대체하고, WML에 존재하지 않는 HTML 태그들은 변환하거나 삭제하는 방법을 사용하였다. 그리고 HTML 문서들이 포함하는 이미지들을 비교적 용량이 작은 텍스트 링크들로 대체했다. 그러나 이미지 맵에 대한 고려는 하지 못하고 있다[9]. 개발된 다른 HTML-WML 변환기도 주로 텍스트 위주의 변환과 필터링 기능을 제공할 뿐이지 이미지에 대한 고려가 되어있지 않다[10].

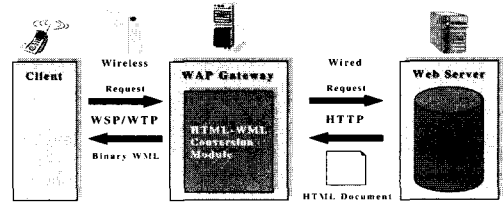
기존 웹사이트들을 분석해보면 유사한 구성 형태를 취하고 있는 것도 많이 있지만, 엄밀히 따지자면 다양한 구성 형태를 취하고 있다. 본 논문에서는 게임, 증권, 유머 사이트 등을 포함한 무선 인터넷 서비스가 가능한 300여 개의 사이트를 분석한 결과 이러한 사이트들의 초기화면 구성 형태는 표 2와 같고, 웹 문서의 메뉴는 표 3과 같은 형태로 구성됨을 알 수 있었다. 단순히 하나의 형태를 취하고 있는 사이트는 드물었으며 복합적인 형태로 구성되어 있는 경우가 많았다.

HTML-WML 변환기는 HTML 문서를 WML 문서로 변환하는 시스템으로서 WAP 게이트웨이, 프락시(proxy) 서버, 웹 서버에서 사용될 수 있다.

WAP 게이트웨이에서 사용되는 경우, WAP 게이트웨이는 사용자의 요구가 들어오면 지정된 URL로부터 HTML 문서를 가져온 후 WAP 게

(표 3) 웹 문서의 메뉴 구성 형태

메뉴의 형태	사용율
이미지 맵	28.9 % (43 사이트)
이미지	65.1 % (97 사이트)
텍스트	18.8 % (28 사이트)
그 외 (애플릿, 스크립트, ...)	39.6 % (59 사이트)

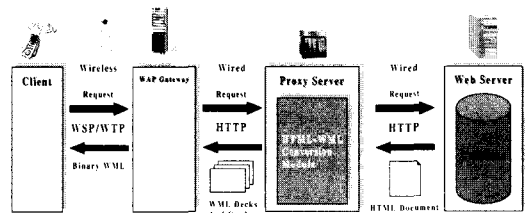


(그림 1) WAP 게이트웨이에서 사용되는 HTML-WML 변환기

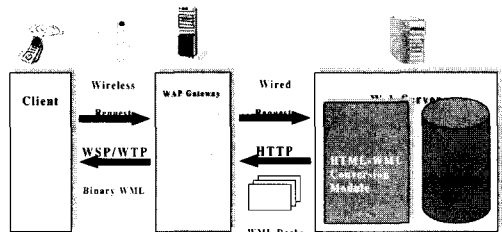
트웨이에서 HTML-WML 변환기를 사용하여 WML로 변환한 뒤 이를 무선망에 적합한 형태로 인코딩해서 무선 단말기로 전송한다. 그 구조는 그림 1과 같다.

프락시 서버에서 사용되는 경우, WAP 게이트웨이는 들어온 사용자의 요구를 프락시 서버로 전달하고, 전달받은 프락시 서버는 지정된 URL로부터 HTML 문서를 가져온 후 HTML-WML 변환기를 사용하여 WML로 변환한 뒤 WAP 게이트웨이로 전달한다. WAP 게이트웨이에서 이를 무선망에 적합한 형태로 인코딩해서 무선 단말기로 전송한다. 그 구조는 그림 2와 같다.

웹 서버에서 사용되는 경우, WAP 게이트웨이



(그림 2) Proxy 서버에서 사용되는 HTML-WML 변환기



(그림 3) Web 서버에서 사용되는 HTML-WML 변환기

는 사용자의 요구가 들어오면 지정된 URL에 따라 웹 서버로 요구를 전달한다. 웹 서버는 HTML-WML 변환기를 사용하여 HTML 문서를 WML로 변환한 뒤 WAP 게이트웨이로 전달한다. WAP 게이트웨이에서 이를 무선망에 적합한 형태로 인코딩해서 휴대용 단말기로 전송한다. 그 구조는 그림 3과 같다.

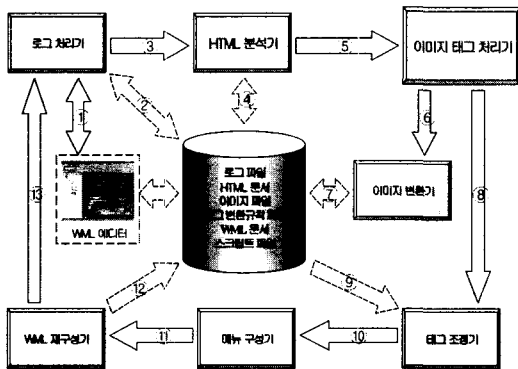
### 3. HTML-WML 변환기의 설계 및 구현

본 논문에서는 기존의 정적 인터넷 환경에서 HTML로 구축해놓은 방대하고 다양한 정보들을 동적 인터넷 환경에서 효과적으로 사용할 수 있도록 WML 문서로 변환하는 HTML-WML 변환기를 그림 4와 같이 설계하고 구현하였다.

HTML-WML 변환기는 로그 처리기, HTML 분석기, 이미지 처리기, 태그 조정기, 메뉴 구성기, WML 재구성기로 구성하였으며 WML 에디터가 결합이 된다면 WML 문서 자체의 수정도 가능하도록 설계하였다. 또한 그림 4에 나타난 WML 에디터 대신에 WAP 게이트웨이와 결합시켜 실시간 변환기로도 사용할 수 있게 설계하였다. 로그 처리기는 로그의 초기화, 로그 분석, 로그 기록의 역할을 수행하며, HTML 분석기는 시작 태그와 종료 태그의 선택적 기록에 대한 것의 일부를 WML에 맞도록 조정하고 HTML 문서를 분석하여 불필요한 요소들을 제거

한 후 HTML 문서로 재생성하는 모듈이다. 이미지 태그 처리기는 이미지에 관련된 태그를 처리하는 모듈로서 이미지 맵의 경우 맵에서의 영역에 따라 이미지 태그로 재구성하고 맵에 대한 정보를 리스트로 구성하며 해당 이미지의 이름을 적절한 WBMP의 이름으로 변환하는 모듈이다. 이미지 변환기는 이미지 태그 처리기로부터 넘겨받은 이미지에 대한 리스트를 토대로 JPG, GIF, BMP와 같은 이미지 포맷을 WBMP의 형태로 변환하고, 부가적으로 영역 추출과 크기 조정의 역할을 수행하는 모듈이다. 태그 조정기는 HTML 태그를 삭제, 유지, 치환, 복합적인 처리의 형태로 정의한 마크업 변환규칙에 따라 WML 태그로 변환하고 시작 태그와 종료 태그가 반드시 존재해야 하는 WML 규칙에 따라 시작 태그와 종료 태그가 없는 부분을 적절하게 조정하는 역할을 수행하는 모듈이다. 메뉴 구성기는 메뉴로서 사용될 수 있는 부분, 내용으로서 사용될 수 있는 부분, FORM에 관련된 부분을 분리하고, 대표가 되는 WML 테크(deck)에 연결하는 역할을 하며, WML 재구성기는 휴대용 무선 장비에서 수용 가능한 적절한 크기의 데크 단위로 카드 단위로 분할하여 최종 WML 문서를 재구성하는 역할을 수행하는 모듈이다.

HTML 문서와 데크 크기, 카드 크기가 입력으로 주어졌을 때, HTML-WML 변환기의 구성도에서 각 모듈 사이에 나타난 정보의 흐름은 다음과 같이 구성하였다.



(그림 4) HTML-WML 변환기 구성도

- ① 외부로부터의 변환요청 또는 변환된 WML 문서의 반환
- ② 로그 파일을 메모리로 적재
- ③ 로그의 분석을 통한 변환의 실행을 지시
- ④ HTML 파일을 적재, 처리 후 스크립트에 해당하는 문서 저장
- ⑤ HTML 분석기를 거친 문서
- ⑥ 변환될 이미지에 관한 이미지 리스트

- ⑦ JPG, GIF, BMP와 같은 이미지 적재, WBMP 로 변환된 이미지 저장
- ⑧ 이미지 태그가 처리된 문서
- ⑨ 태그 변환 규칙 적재
- ⑩ 태그 변환 규칙을 적용하여 적절한 WML 태그로 변환된 문서
- ⑪ 메뉴 구성기를 거쳐 메뉴, 내용, 사용자 입력의 형태로 구성된 문서
- ⑫ WML 재구성기를 거쳐 데크와 카드의 단위로 분리된 문서 저장
- ⑬ 로그 파일에 기록될 정보

### 3.1 로그 처리기

HTML-WML 변환기를 통해 이미 변환된 HTML 문서의 경우 다시 처리하지 않고 곧바로 WML 문서를 찾아 넘겨줄 수 있도록 로그에 변환과 관련된 정보를 기록하도록 설계하였다. 외부로부터의 변환 요청이 있으면 HTML 문서의 최근 수정일자와 로그에 있는 변환일자를 비교하여 변환 여부를 검사하도록 하였다. 로그 정보를 연결 리스트의 형태로 구성하였으며 순서는 파일이름 순서를 원칙으로 하였다. 로그 처리기는 크게 세 부분으로 구성하였다.

#### 3.1.1 초기화

로그 파일의 내용을 읽어서 메모리에 적재하며 로그 레코드의 루트로부터 연결 리스트를 통하여 접근 가능하도록 하였다.

#### 3.1.2 로그 분석

로그의 내용에서 해당 HTML 문서의 이름을 검색하며 해당 HTML 문서의 이름이 존재하지 않을 경우는 새 로그 레코드를 할당하여 로그 연결 리스트에 삽입하고, 존재하는 경우는 해당 문서에 대한 최근 수정일자와 로그에 있는 변환일

자를 비교하여 최근 수정일자와 변환일자가 다를 경우는 로그 정보를 수정하고 변환기를 통해 변환을 하며, 같을 경우는 변환기를 거치지 않고 WML 문서 이름을 반환하도록 하였다.

#### 3.1.3 로그 기록

변환을 마친 후 메모리에 적재되어 있는 로그 테이블의 정보를 로그 파일에 기록하도록 하였다.

### 3.2 HTML 분석기

HTML 문서를 WML 문서로 변환하는데 있어서 우선적으로 고려가 되어야 할 사항은 WML 문서는 XML 문서와 마찬가지로 엠프티(empty) 태그를 제외하고는 시작 태그와 종료 태그가 반드시 존재해야 한다는 점이었다. 하지만 HTML은 시작 태그와 종료 태그가 없어도 되는 경우가 일부 허용이 되므로 WML 문서로 변환하는데 있어서 문제를 발생시킬 수 있었다. 따라서, WML 문서로 변환하기 위해 태그가 없는 것을 적절히 조정할 필요가 있었다. HTML 분석기에서는 모든 태그에 대한 것을 고려한 것이 아니라 HTML, HEAD, BODY, FRAMESET 태그에 대한 것만을 고려하도록 구현하였다. 나머지 태그는 뒤에서 설명할 태그 조정기에서 담당하도록 하였다. FRAMESET 태그의 경우는 BODY 태그가 없는 경우이므로 모든 FRAME에 해당하는 문서를 로드하여 병합할 때 BODY 태그가 없는 문제를 해결하고자 FRAMESET 태그 앞뒤로 BODY 태그를 삽입하였다. 또한 SCRIPT 태그가 HEAD 태그 속에 들어 있다면 BODY 태그 속으로 이동시켰다. 주석의 경우는 SCRIPT 태그 속에서 사용될 때 무시할 수 없는 경우가 있으므로 SCRIPT 태그 속에 있는 주석은 특별히 지우지 않고 SCRIPT 태그와 함께 BODY 태그 속으로 이동시켰다.

HTML 문서는 프레임들로 구성될 수 있으며 각각의 프레임을 모두 병합하여 하나의 문서로

처리해야 할 필요가 있었다. 따라서, HTML 분석기에서는 각 프레임을 모두 병합하여 하나의 문서로 생성하였다. 또한 HTML 문서는 자바스크립트나 애플릿(applet), 주석, 스타일에 관련된 것들과 같이 WML 문서로 변환할 때 처리하기 힘들거나 불필요한 사항들을 포함하고 있는 경우가 많았다. HTML 분석기는 이러한 사항들을 제거함으로써 다음 모듈들의 부하를 줄여주는 역할을 수행하도록 하였다.

### 3.3 이미지 태그 처리기

이미지 태그 처리기는 HTML 문서에서 이미지와 이미지 맵에 관련된 태그들을 재구성하는 모듈이다. 이 모듈에서는 태그만 변환하고 실질적인 이미지 처리는 이미지 변환기에서 수행하도록 하였다. 그림 5는 이미지와 이미지 맵에 관련된 태그들의 변환 예이다.

### 3.4 이미지 변환기

이미지 변환기는 기존의 HTML 문서에 있는 아이콘(icon)이나 기타 그림들을 휴대용 무선 장비에 디스플레이하기 위해 WBMP 포맷으로 변환하는 모듈이다. WBMP 포맷은 보통 2Kbyte 이하의 크기를 가지며 1비트 흑백영상이므로 휴대용

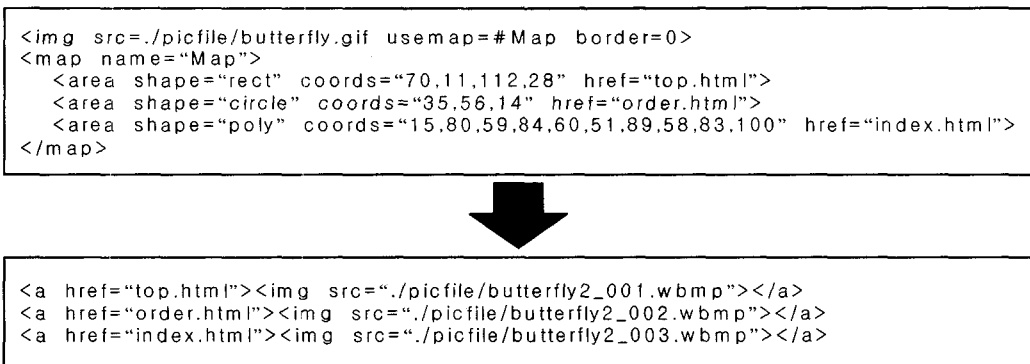
무선 장비에 디스플레이하기 위해선 적절한 크기로 조정되어야 하였다. 이미지 변환기의 경우 다음과 같이 4가지 특징을 갖도록 구현하였다.

- (1) JPEG, TIFF, GIF, BMP 등의 그래픽 이미지 처리
- (2) 휴대용 무선 장비 특성상 이미지를 확대/축소하여 127\*127 크기에 들어가도록 변환
- (3) HTML의 이미지 맵을 처리하기 위해 이미지의 임의의 영역 저장
- (4) 이미지를 1비트 흑백영상으로 변환하기 위해 WBMP에 맞게 이미지 저장

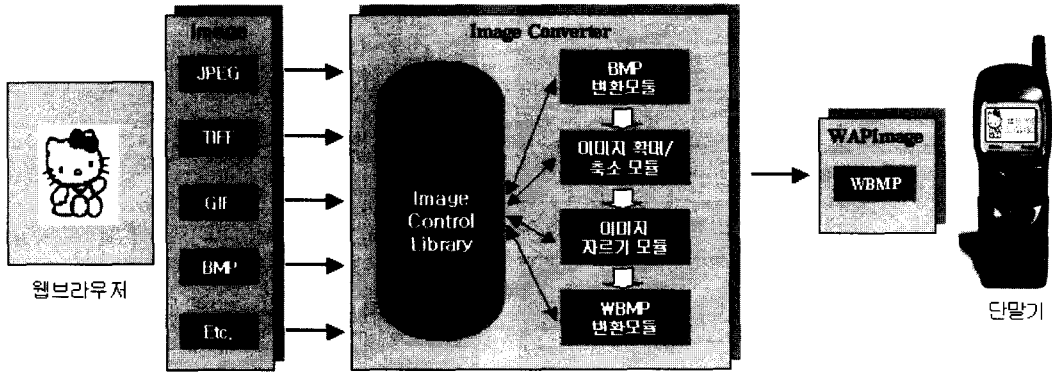
이미지 변환기의 내부적인 처리는 그림 6과 같이 구성하였다.

이미지 변환기는 이미지 태그 처리기로부터 넘겨받은 이미지에 대한 리스트를 토대로 이미지 맵의 경우에는 영역에 해당하는 이미지를 추출하고 WBMP의 형태로 변환한 후 이미지 태그 처리기에서 부여한 이름으로 저장하도록 하였다. 이미지의 경우에는 WBMP의 형태로 변환한 후 이미지 태그 처리기에서 부여한 이름으로 저장하도록 하였다.

HTML 문서에 포함된 이미지의 크기를 휴대용 무선 장비의 제한된 화면 크기에 적합하도록 조정할 필요가 있었다. 이미지 크기 조정은 HTML 문서에 포함된 이미지를 휴대용 무선 장비의 화면에 적절한 크기로 나타내기 위해 원 영상에 대



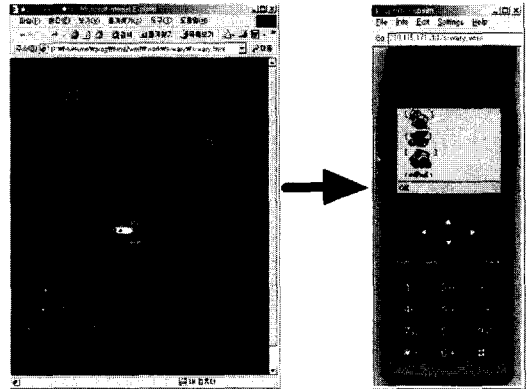
(그림 5) 이미지와 이미지 맵에 관련된 태그 변환 예



(그림 6) 이미지 변환기의 구조

한 x축과 y축의 축소 배율을 계산하여 원 영상을 축소하는 방법을 사용하였다. 축소 배율은 축소하고자 하는 영상의 가로와 세로 크기를 원 영상의 가로와 세로 크기로 나누어 계산하였다. 축소 배율을 계산한 후 x축과 y축의 좌표값으로 표현된 원 영상에 대한 픽셀들에 이 배율을 곱하여 축소 영상에 대한 픽셀들의 x축과 y축의 좌표값을 계산하였다. 예를 들어, 원 영상의 가로와 세로 크기가 4이고, 축소하고자 하는 영상의 가로와 세로 크기를 2로 하고자 한다면 x축과 y축의 축소 배율은 0.5가 된다. x축과 y축의 좌표값으로 표현된 원 영상에 대한 픽셀들의 좌표값 (0,0), (0,1), (0,2), (0,3), ..., (3,0), (3,1), (3,2), (3,3)에 이 축소 배율을 곱하여 계산하면 축소 영상에 대한 픽셀들의 x축과 y축의 좌표값 (0,0), (0,1), (1,0), (1,1)을 계산할 수 있다.

하나의 이미지 맵은 이미지 포맷의 여러 객체들로 이루어져 있다. 이러한 각각의 이미지 객체들은 <AREA> 태그의 'COORDS' 속성값에 따라 사각형(x1, y1, x2, y2), 원(x, y, r), 다각형(x1, y1, x2, y2, x3, y3)의 형태를 갖는다. 이미지 맵에서 이러한 각각의 도형으로 이루어진 이미지 객체들을 사각형으로 매핑하여 추출하는 방법을 사용하였다. 사각형의 이미지 객체는 그대로 추출하고, 좌표값과 반지름이 (5, 5, 2)인 원 형태의 이미지 객체는 좌표값이 (3, 3)에서 (7, 7)까지의 사각형



(그림 7) 이미지 변환기를 통한 이미지 변환 예

으로 매핑하여 추출하고, 좌표값이 (3, 3, 5, 5, 6, 10, 2, 9)인 다각형 형태의 이미지 객체는 최소의 x축과 y축의 좌표값 (2, 3)에서 최대의 x축과 y축의 좌표값 (6, 10)까지 표현된 사각형으로 매핑하여 추출하였다.

HTML 문서에서는 보다 유용하고 효과적인 정보 전달을 위하여 수많은 이미지들을 사용하게 된다. 이러한 이미지들은 GIF, JPEG, TIF 등의 타입을 가지고 있다. 이미지 변환기에서는 이러한 이미지들을 크기가 작은 BMP 형태로 변환한 후 휴대용 무선 장비에 적합하도록 WBMP 형태로 변환하였다. 그림 7은 일반 웹 브라우저에서 보이는 이미지를 휴대용 무선 장비에서 볼 수 있는 이미지로 변환한 예이다.

(표 4) HTML 태그와 WML 태그의 비교

공통적으로 있는 엘리먼트		22개	WML에만 있는 엘리먼트		14개	HTML에만 있는 엘리먼트		69개
<a>	<b>	<big>	<wml>	<card>	<do>	<address>	<applet>	<base>
 	<em>	<fieldset>	<onevent>	<template>	<access>	<body>	<button>	<col>
<head>	<i>	<img>	<go>	<prev>	<refresh>	<div>	<font>	<frame>
<input>	<meta>	<optgroup>	<noop>	<postfield>	<setvar>	<h1>	<html>	<ins>
<option>	<p>	<pre>	<timer>	<anchor>		<label>	<link>	<map>
<select>	<small>	<strong>				<noscript>	<object>	<ol>
<table>	<td>	<tr>				<script>	<style>	<sub>
<u>						<sup>	<title>	...

### 3.5 태그 조정기

WML에서는 HTML에서 사용하는 태그를 모두 지원하지 않는다. 따라서, 적절한 변환 규칙이 필요하였다. 표 4는 이러한 변환 규칙을 결정하기 위해 HTML에서 사용하는 태그와 WML에서 사용하는 태그를 비교, 분석한 결과를 나타낸 것이다.

결과를 살펴보면 WML 태그와 HTML 태그가 일치하는 것이 있기도 하였지만 WML은 상당수의 HTML 태그를 지원하지 않음을 알 수 있었다. 태그 뿐만 아니라 속성을 비교, 분석한 결과도 마찬가지로의 결과를 얻었다. 이러한 결과를 토대로 HTML 태그가 WML에서 적용될 수 있도록 적합한 형태로 변환해 주는 규칙을 정의하였고 삭제, 치환, 유지, 복합적인 처리 형태를 가지도록 하였다. HTML의 태그 및 속성이 WML의 태그 및 속성과 일치한다면 그대로 사용하도록 하였다. HTML 태그와 WML 태그가 일치하지는 않지만, 유사한 기능을 가지는 WML 태그가 있을 경우 HTML 태그를 WML이 지원하는 태그들로 적절하게 변환하였다. 그러나, WML에서 지원하지 않거나 휴대용 무선 장비에서 지원하지 않는 기능을 가진 태그는 삭제하였다.

태그 조정기에서는 최대한 HTML 태그를 삭제하지 않고 사용할 수 있도록 태그 변환 규칙을 정의하였다. HTML 태그의 기능을 최대한 살릴 수 있도록 WML이 지원하는 태그들을 조합하여 변환하도록 하였다. 표 5는 태그 조정기에서 사용하는 변환 규칙을 나타낸 것이다.

삭제할 태그 중에서도 태그만 삭제하는 경우와 태그 사이의 내용까지 삭제하는 경우가 있었다. 즉, 휴대용 무선 장비에서 지원하지 않는 <applet>이나 <object>와 같은 태그는 태그 사이의 내용까지 모두 삭제하여야 하지만 <font>와 같은 태그는 태그만 삭제하고 태그 사이의 내용은 그대로 두어야 하였다. 따라서, 태그들의 특성을 제대로 분석하여 태그들의 기능을 최대한 잃지 않고 삭제하도록 규칙을 정의하였다.

<form>과 관련된 <button>, <textarea>, <isindex> 등과 같은 태그들은 WML에서는 지원하지 않지만 웹 문서에 따라서는 꼭 필요한 요소가 될 수 있으므로 적절한 WML 태그로 변환해 주어야 하였다. 이러한 <form>과 관련된 태그들은 복잡한 변환이 이루어져야 하므로 복합적인 처리의 형태로 처리하도록 하였다. <button> 태그는 <input> 태그로 적절하게 변환하도록 하였고, <textarea>



(표 5) HTML 태그의 WML 태그로의 변환 규칙

삭제			치환		유지		
23개			41개		22개		
<applet>	<base>	<basefont>	<address>	<i>	<a>	<b>	<big>
<col>	<colgroup>	<dd>	<area>	<a><img>	 	<em>	<fieldset >
<dt>	<font>	<hr>	<caption>	<p>	<head>	<i>	<img>
<frame>	<frameset>	<iframe>	<center>	<p align="center">	<input>	<meta>	<optgroup>
<li>	<link>	<map>	<cite>	<i>	<option>	<p>	<pre>
<noframes>	<noscript>	<object>	<code>	<pre>	<select>	<small>	<strong>
<param>	<script>	<style>	<dl>	<p>	<table>	<td>	<tr>
<textarea>	<title>		<html>	<wml>	<uu>		
			<ins>	<u>			
			<label>	<b>			
			<menu>	<fieldset >			
			<sub>	<small>			
			<th>	<td><b>			
			...	...			

역시 주어진 속성값에 따라 <input> 태그로 변환하거나, 일반 내용으로 처리하도록 하였다. <form> 태그는 삭제하여야 하지만, <form> 태그와 관련된 속성이 이후 WML 문서의 네비게이션(navigation) 요소가 되므로 태그 조정기에서는 삭제하지 않고 그대로 유지하도록 하였다. 하지만 이후 WML 재구성기에서 <form> 태그를 삭제하기 때문에 사실상의 태그를 삭제하는 것과 같은 의미를 가지게 된다.

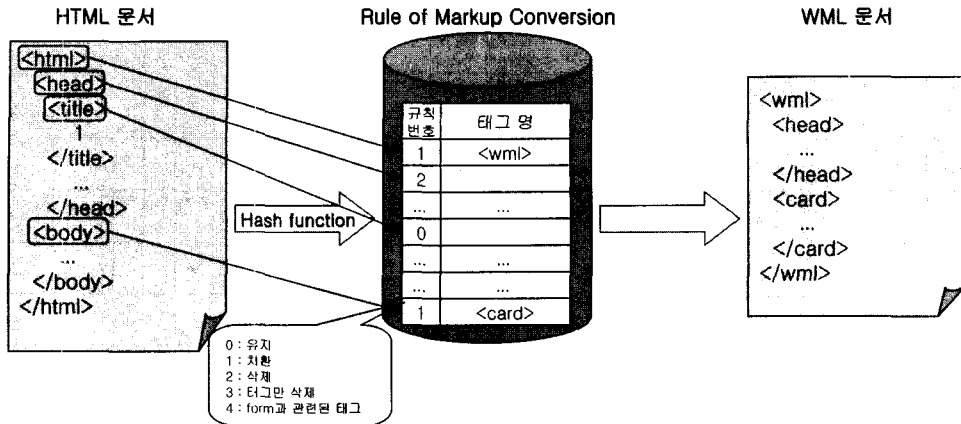
태그 변환 규칙을 위해 변환 규칙이 적용된 변환 테이블을 이용하였으며, 이 변환 테이블로의 접근은 해시 함수를 이용하였다. 해시 함수는 HTML 태그를 키값으로 하여 해시값을 얻어내며, 이를 통해 보다 빠른 변환 테이블로의 접근이 가능하였다. 하지만, 이 해시 함수에서는 HTML 태그가 서로 다른 경우라도 해시값이 서로 같은 경우가 발생할 수 있었다. 따라서, 이러한 충돌 현상을 보완하기 위해, 정해진 해시값이 가지는 태그 변환 정보를 가지는 노드들을 연결 리스트의 형태로 유지하였다. 변환 테이블은 HTML 태그와 태그에 대한 변환 규칙 번호, 그리고 변환

될 WML 태그로 구성하였다.

WML은 XML의 한 응용이므로 앰프티 태그를 제외한 모든 태그들은 시작 태그와 종료 태그가 모두 존재해야 한다. 따라서, 태그 조정기는 시작 태그와 종료 태그가 모두 존재하도록 조정하는 역할도 수행하도록 하였다.

### 3.6 메뉴 구성기

메뉴 구성기에서는 태그 조정기를 거친 문서를 메뉴, 내용 및 폼(form)과 관련된 사용자 입력 부분으로 분리하는 모듈이다. 이는 휴대용 무선 장비로 사용자가 원하는 정보를 보다 빠르게 얻을 수 있도록 하기 위한 방법이다. 즉, 사용자가 어떤 HTML 문서를 요청하였을 경우, 변환되어 가장 처음 전송된 WML 문서에서 메뉴, 내용, 사용자 입력 부분 중 사용자가 원하는 부분을 선택할 수 있도록 하였다. 메뉴로 구성되는 것들은 모두 링크가 설정되어 있다. 이러한 링크가 설정되어 있는 것을 추출하여 메뉴로 구성하고 <form> 태그와 </form> 태그 사이의 것들은 사용자 입력으



(그림 8) HTML-WML 태그 변환

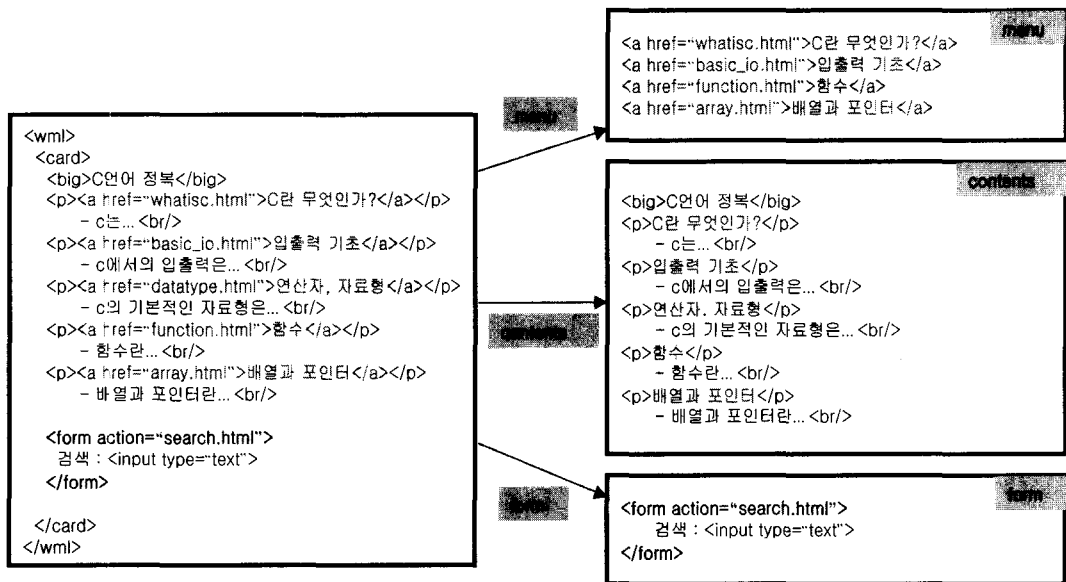
로 구성하였다. 두 개의 경우를 제외한 나머지 것들은 내용으로 구성하였다. 그림 9는 하나의 WML 문서 구조에서 메뉴 부분, contents 부분, form 부분을 따로 분리해 내는 예제를 보인 것이다.

크의 구조이며 다음과 같은 구문의 특징을 가진다.

- (1) 문자셋 : XML 문서의 문자셋을 이용한다. (ISO/IEC-10646 (Unicode 2.0), US-ASCII, ISO-8859-1, or UTF-8)
- (2) 대소문자 구분 : HTML과는 다르게 WML은 대소문자를 구분한다.
- (3) 빈칸 : 새 줄(new line), 스페이스, 탭 등 모

### 3.7 WML 재구성기

XML의 한 응용인 WML은 HDML처럼 카드와 데



(그림 9) 메뉴, 내용, 폼 관련 정보 분리 예

두 다르게 표현한다.

- (4) 문서선언 : XML 문서는 맨 앞에 다음과 같이 선언해야 한다.

```
<?xml version="1.0"?>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//
DTD WML 1.1//EN" http://www.wapforum.
org/DTD/wml_1.1.xml">
```

- (5) Content 형식 : 문서 맨 앞에 정의해야 한다.

```
Content-type: text/vnd.wap.wml
```

- (6) 속성(attributes) : 대부분의 WML 엘리먼트 들은 하나 이상의 속성을 갖는다.

```
<element a1="value1" a2="value2" ...>
content </element>
<element a1="value1" a2="value2" .../>
```

- (7) 변수

유닉스의 셸 1 변수와 비슷한 구조를 가지고 있으며, 변수를 통해서 문자, URL, 아이템 선택, 디폴트 값 등에 이용할 수 있다.  
\$(myvar).

- (8) 특수문자

WML에서 사용하는 특수문자는 표 6과 같다.

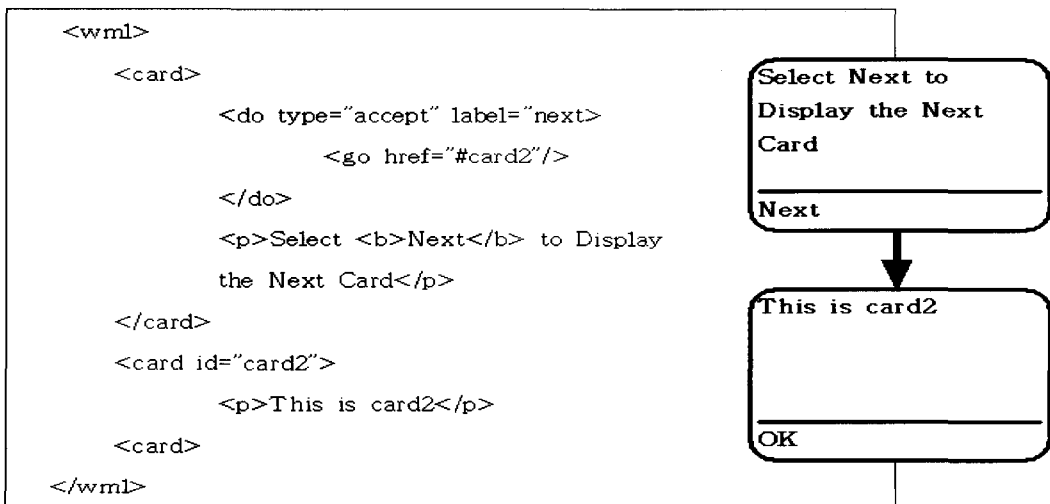
(표 6) 특수문자

special characters	notation
<	&lt; (less than)
>	&gt; (greater than)
'	&apos; (apostrophe)
"	&quot; (quote)
&	&amp; (ampersand)
\$	\$\$ (dollar sign)
Non-breaking space	&nbspbsp;
soft hyphen	&shy;

데크는 하나의 WML 문서를 뜻하며, 카드는 휴대용 무선 장비의 화면에 디스플레이하기 위한 단위이고 하나 이상의 카드가 모여 하나의 데크를 이루게 된다. 데크와 카드 크기는 일정하게 정해질 수가 있으며 해당 내용이 한 데크나 한 카드에 들어가지 않을 경우 다음 데크나 다음 카드와 연결이 되어야 한다. 이때 사용되는 개념이 네비게이션이다. 그림 10은 네비게이션의 간단한 예제이다.

WML 재구성기는 HTML문서를 WML로 변환하기 위한 마지막 단계로써 이전에 모든 처리가

WML은 카드와 데크의 구조로 이루어져 있다.



(그림 10) 네비게이션의 예제 및 단말기의 출력 상태

끝난 수정된 HTML 문서를 카드와 데크 구조의 WML 문서로 만드는 모듈로 휴대용 무선 장비의 화면 크기에 따라 카드 단위로 구성하고 휴대용 무선 장비의 메모리 용량에 따라 데크 단위로 구성하도록 구현하였다. 단위 구성을 한 후에 각 카드와 데크 사이에 네비게이션 정보를 추가적으로 저장해서 WML 문서를 완성하도록 하였다.

#### 4. 수행 결과 및 분석

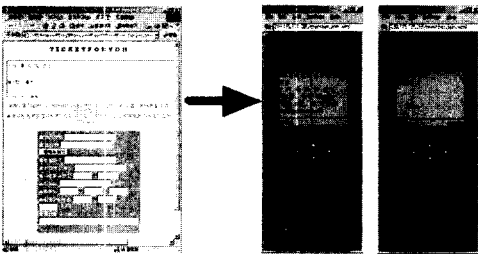
설계한 HTML-WML 변환기의 구현을 통해 HTML로 기술된 웹 문서를 WML 문서로 변환하는 간단한 예제를 그림 11과 그림 12에 나타내었다. 그림 11은 폼과 관련된 정보가 포함되어 있는 웹 문서를 WML 문서로 변환한 예제로써 내용을 나타내는 화면과 사용자의 입력을 받는 화면이 보여지고 있다. 그림 12는 이미지 맵과 이미지로 구성된 메뉴를 포함하는 웹 문서를 WML 문서로 변환한 예제로써 메뉴의 기능을 하는 이미지가 WBMP의 형태로 변환되어 보여지는 메뉴 화면과

나머지 내용 부분을 나타내는 화면이 보여지고 있다. 만일 이미지에 링크가 설정이 되어있지 않으면 그 이미지는 내용으로써 취급되며 텍스트라도 링크가 설정되어 있으면 메뉴로써 취급된다.

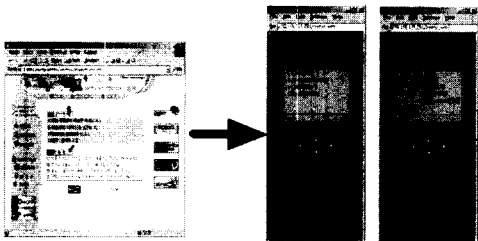
개발한 HTML-WML 변환기는 HTML로 기술된 웹 문서를 휴대용 무선 장비의 메모리 용량과 디스플레이 화면 크기에 최적화된 WML 문서로 변환하는 기능을 제공하였으며 HTML 태그와 속성을 적절한 WML 태그와 속성으로 변환할 수 있도록 변환규칙을 사용하였다. Spyglass사의 Prism은 이미지를 텍스트 링크로 대체할 때 이미지의 요약 정보가 없는 경우가 발생할 수도 있고 확실한 의미 전달이 되지 않는다는 단점을 가지고 있었다. 이러한 단점을 개선하고자 이미지와 이미지 맵을 휴대용 무선 장비에 적합한 WBMP의 형태로 변환해줌으로써 사용자가 쉽게 의미를 파악할 수 있도록 했다. 또한 실시간 변환기로서 사용이 가능할 뿐만 아니라 WML 에디터와 같은 응용에서도 변환기로서 사용할 수 있도록 했다. 그림 11과 그림 12에서 볼 수 있는 바와 같이 다양한 형태의 HTML 문서들을 최대한 수용하고 있으며 HTML 문서가 가지고 있는 정보의 손실을 최소화하고 있다.

#### 5. 결론

본 논문에서는 휴대용 무선 장비를 통해 무선 인터넷 환경을 사용할 수 있도록 기존의 HTML로 기술된 웹 정보를 WML 문서로 변환하는 HTML-WML 변환기를 설계하고 구현하였다. HTML-WML 변환기는 로그 처리기, HTML 분석기, 이미지 태그 처리기, 이미지 변환기, 태그 조정기, WML 재구성기로 구성하였으며 실시간 변환기 뿐만 아니라 WML 에디터와 같은 응용에서도 사용이 가능하도록 하였다. 휴대용 무선 장비의 제한적 환경에 적합하도록 동적으로 WML 문서를 재구성하여 휴대용 무선 장비의 사용자가 무선 인터넷 서비스를 이용할 수 있는 기반을 마련했다. 기존의 다른 변



(그림 11) 폼을 포함하는 HTML 문서의 변환 예



(그림 12) 이미지 맵과 이미지 메뉴를 포함한 HTML 문서의 변환 예

환기의 특징들을 모두 수용하고 있으며 보다 다양한 HTML 문서를 안정적이고 정보의 손실 없이 WML 문서로 변환할 수 있었다. 향후 WMLScript에 대한 처리를 고려하고 보다 다양한 웹 문서에 대한 변환을 수행해 봄으로써 추가적으로 필요한 기능들을 찾아내고 보완해 나갈 계획이다.

## 참고 문헌

- [1] Handheld Device Markup Language Specification, <http://www.w3.org/TR/NOTE-Submission-HDML-spec.html>
- [2] WAP Forum, Ltd., "WAP Architecture Specification," <http://www.wapforum.org>, April 1998.
- [3] WAP Forum, Ltd., "WAP overview," <http://www.wapforum.org>, February 1999.
- [4] WAP Forum, Ltd., "WAP WML Specification Version 1.2," <http://www.wapforum.org>, February 1999.
- [5] WAP Forum, Ltd., "WAP WML Specification Version 1.3," <http://www.wapforum.org>, 2000.
- [6] 박진희, 김명철, 마중수, "HTML(Hyper Text Markup Language) - WML(Wireless Markup Language) Translator 구현", 한국정보과학회 가을 학술발표논문집 Vol. 26. No. 2, 1999.
- [7] 강현관, 윤종선, 강경진, "HTML 문서 내에서의 정보 추출에 관한 연구", 한국정보과학회 가을 학술발표논문집 Vol. 26. No. 2, 1999.
- [8] Lettieri P, Srivastava MB, "Advances in wireless terminals", IEEE Personal Communications, V.6 N.1, 6-19, 1999.
- [9] OPENTV, "Spyglass Prism 3.1.0", <http://www.opentv.com>, 2000.
- [10] 이정은, 최철웅, 장지산, 김성찬, 신동일, 신동규, "WAP기반 무선 콘텐츠 변환기 System의 설계 및 구현", 한국정보처리학회 2001년도 춘계학술발표논문집(하) 제8권 제1호, 2001, 4, pp. 1145-1148
- [11] 김태주, 채경호, 임경식, "무선 인터넷 환경에서 단말에 최적화된 WBMP 이미지 변환기 구현", 한국정보처리학회 2001년도 춘계학술발표논문집(하) 제8권 제1호, 2001, 4, pp. 661-664
- [12] "Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation 6-October-2000, REC-xml-20001006," Tim Bray, et al, October 10, 2000. URL: <http://www.w3c.org/TR/2000/REC-xml-20001006>
- [13] "HTML 4.01 Specification, W3C Recommendation 24-December-1999, REC-html401-19991224," Dave Raggett, et al, December 24, 1999. URL: <http://www.w3c.org/TR/1999/REC-html401-19991224>
- [14] 이상엽, Visual C++ Programming Bible Ver. 6.x, 영진출판사, 1998.
- [15] UP.SDK Release 4.1 Getting Started Guide, URL: <http://www.openwave.com>
- [16] UP.SDK Release 4.1 Developer's Guide, URL: <http://www.openwave.com>

## ◎ 저자 소개 ◎



### 민 영 수

1998년 충북대학교 정보통신공학과(공학사)  
2001년 충북대학교 대학원 정보통신공학과(공학석사)  
2001년 3월~현재 충북대학교 대학원 정보통신공학과 박사과정  
관심분야 : XML, 실시간 시스템, 데이터베이스, etc.  
E-mail : minys@netdb.chungbuk.ac.kr



### 강 형 일

1996년 목포대학교 전산통계학과(이학사)  
1998년 목포대학교 전산통계학과(이학석사)  
2001년 충북대학교 정보통신공학과(공학박사)  
2001년~현재 주성대학 멀티미디어정보통신공학부 전임강사  
관심분야 : 멀티미디어 데이터베이스, XML, 정보검색, etc.  
E-mail : khi69@jsc.ac.kr



### 유 재 수

1989년 전북대학교 컴퓨터공학과(공학사)  
1991년 한국과학기술원 전산학과(공학석사)  
1995년 한국과학기술원 전산학과(공학박사)  
1995년~1996년 8월 목포대학교 전산통계학과 전임강사  
1996년 8월~현재 충북대학교 정보통신공학과 부교수  
관심분야 : 데이터베이스 시스템, XML, 멀티미디어 데이터베이스, 분산채체 컴퓨팅, etc.  
E-mail : yjs@cbucc.chungbuk.ac.kr