

# 무선 콘텐츠 서비스를 위한 HTML Filter System의 설계

김용빈\*, 김민수, 신동규, 신동일

세종대학교 컴퓨터공학과,

## HTML Filter System for Wirelss Contents Service

Kim, Yong Bin, Kim, Min-Su, Shin, Dong-Kyu, Shin, Dong-il

Sejong University

E-mail : kimybin@gce.sejong.ac.kr, mskim@gce.sejong.ac.kr,

shindk@sejong.ac.kr, dshin@sejong.ac.kr

### 요약

WAP을 기반으로 한 무선 인터넷 콘텐츠 서비스는 무선 인터넷 마크업 언어인 WML(Wireless Markup Language)을 사용하여 실현하게 된다. 휴대폰 사용자들은 무선 인터넷 서비스를 통하여 WML로 작성된 콘텐츠를 휴대폰을 통해 볼 수 있는 것이다. 이에 본 논문에서는 무선인터넷의 효율적인 콘텐츠 서비스를 위하여 기존 웹 페이지를 표현하기 위한 HTML(Hypertext Markup Language)을 WML로 변환하기 위해 개발된 기존 제품들에 대한 비교분석을 통하여 효과적인 필터링 기능을 수행하는 시스템을 제안하였다.

### 1. 서론

전 세계적인 인터넷의 폭발적인 확산은 기존의 사회기반을 급격히 변화시키는 촉매가 되고 있으며, 인터넷과 통합된 멀티미디어 기술의 발전 및 초고속 통신망의 보편적인 사용으로 인하여 기존 서비스 방식을 탈피한 다양한 멀티미디어 응용 및 정보 서비스가 현재 웹 상에서 증가하고 있다. 반면, 통신이용자들은 시간과 장소, 가격에 얽매이지 않고 인터넷에 접속하여 자신이 원하는 정보를 취득할 수 있는 환경에 대한 요구가 증가하는 추세에 따라 무선 인터넷 환경과 각종 서비스에 대한 출현이 가속화되고 있다. 그 중에서 인터넷에 산재되어 있는 정보를 이동성이 강한 장비, 예를 들어 휴대폰이나 PDA를 통해 정보를 취득하는 방법이 관심의 대상으로 부각되고 있다. 또한, 현재 무선 인터넷에서 서비스되고 있는 분야는 특정분야 즉,

뉴스, 증권, 게임 등의 일부 분야에 제한되어 있는 실정이다.

지금까지 다양하게 출현된 휴대폰을 통한 인터넷 정보로의 접근 기술 중, 대표적인 것이 WAP 포럼(Forum)에서 제정한 WAP(Wireless Application Protocol)[1]이 있다. WAP은 현재 무선통신 환경에 산재해 있는 휴대 단말기의 크기, 컴퓨팅 능력, 무선망의 낮은 대역폭, 불안정한 접속, 데이터 전송의 지연 등의 산재해 있는 제약들을 완화하면서 안정된 무선 인터넷 서비스를 제공하기 위해 설계되었다. 또한, WAP을 기반으로 한 무선 인터넷 콘텐츠 서비스는 무선 인터넷 마크업 언어인 WML(Wireless Markup Language)[2]을 사용하여 실현하게 된다. 휴대폰 사용자들은 무선 인터넷 서비스를 통하여 WML로 작성된 콘텐츠를 휴대폰을 통해 볼 수 있는 것이다. 이러한 WML 콘텐츠를 서비스하는 콘텐츠 제공자(Contents

Provider)들에게는 WML로 작성된 문서와 관련된 애플리케이션을 효과적으로 무선 인터넷 사용자에게 서비스하는 것이 사용자의 서비스에 대한 만족도를 높이는 길일 것이다.

그러나, 현재 유선인터넷 상에서의 양질의 멀티미디어 콘텐츠 및 다양한 정보들이 무선 인터넷으로 원활하게 서비스되고 있지 않은 상황에서 일본에서 성공을 거둔 NTT Docomo의 IMode는 양질의 콘텐츠의 제공으로 인해 수많은 사용자들을 무선 인터넷으로 흡수하면서 성공을 거둔 사례라 할 수 있다.

이에 본 논문에서는 무선인터넷의 효율적인 콘텐츠 서비스를 위하여 기존 웹 페이지를 표현하기 위한 HTML(Hypertext Markup Language)을 WML로 변환하기 위해 개발된 기존 제품들에 대한 비교분석을 통하여 효과적인 필터링 기능을 수행하는 시스템을 제안하였다.

## 2. 본론

### 2. 관련연구

#### 2.1 무선인터넷 동향

현재 유선인터넷에는 HTML을 사용하여 웹 페이지를 표현하는 반면, 무선인터넷에서는 무선 인터넷의 제약을 고려하여 만든 WML을 사용하여 콘텐츠를 표현하고 있다. 물론, WML이 HTML보다 표현 할 수 태그가 제한되어 있고, 콘텐츠의 다양한 표현에 있어서는 비교가 되지 않을 수 있으나, 무선통신 환경의 제약사항을 고려해 볼 때 WML의 사용은 필수불가결한 것이다.

현재 무선 인터넷은 초기 유선인터넷 도입기와 비슷한 상황에 있다고 볼 수 있다. 초기 유선 인터넷은 기술적 진입 장벽과 통신 인프라의 제약으로 인해 다소간 폐쇄적인 상황이었으나 인터넷 접근이 대중화되고 활성화됨에 따라 자유로운 공간으로 자리매김 하였다. 양질의 무선 인터넷 콘텐츠의 신속하고 효율적인 개발이라는 필요에 의해 기존의 수많은 인터넷 콘텐츠들중 무선 인터넷에 적합한 콘텐츠에 대한 연구와 이러한 콘텐츠들의 무선 단말기 상에서의 접근에 대한 요구 또한 증대되고 있다.[9]

하지만, 유선인터넷의 개발자들만큼 무선 인터넷의 전문 개발자들의 업계의 수요에 비해 상당히 부족한 현실이다.

따라서, 본 논문의 연구 방향은 기존의 양질의

콘텐츠를 WML로 변환시킴으로써 기존 콘텐츠의 다양함을 유지하는 것이다. 이런 목적으로 유, 무선 인터넷을 통합하는 이상적인 방안은 XML(eXtensible Markup Language)을 이용해서 사이트를 새롭게 실제 데이터와 프리젠테이션을 분리하는 방안이다. XML에 의해 사이트의 기반 데이터들을 DB에 저장하고 사이트에 접근하고자 하는 무선 단말기의 종류에 따라 적절한 XSL(eXtensible Stylesheet Language)에 의해서 적절히 변형되어 단말기에 사이트 콘텐츠들을 제공하게 된다. XML기반의 신규 사이트 구축은 장기적으로 바라볼 때 유, 무선 인터넷의 통합을 위해 추진되어야 될 바른 방안으로 보이지만 신규 구축에 필요한 비용과 노력을 고려할 때 결정하기 어려운 문제이다.

이러한 모든 상황을 고려하여 현재 상용 무선 인터넷 변환기(Filter or Converter)들이 다수 출시되어 있다.

다음에서 기존 상용제품들에 대해 알아보고 제안한 Filter System에 대하여 알아보도록 하겠다..

#### 2.2 기존 제품의 고찰

현재 Filter System은 크게 두가지로 나뉘어 진다.

##### ■ Full Automated

- Phone.com WAP Gateway[4]

변환의 목표가 되는 HTML 파일의 제목, 텍스트, 링크들을 자동으로 추출해서 WML로 변환하기 때문에, 실제로 변환된 결과의 QoS(Quality of Service)보장은 어렵다.

- Argo Actigate[5]

마크업 언어의 변환은 HTML에서 WML, cHTML, HDML, XML등의 변환을 지원한다. 현재 까지 시장에서 접할 수 있는 자동화된 제품들 중 최상의 기능을 제공한다.

##### ■ Configurable

- Oracle 9i Application Server Wireless Edition[6]

Target HTML 파일을 분석해서 얻어지는 HTML 구성 요소인 제목, 텍스트 구문, 이미지, 입력 폼, 테이블등에 대하여 사이트 운영자나 개발자가 불필요한 항목에 대한 삭제, 일부항목에 대한 수정, 신규 항목의 추가등의 설정이 가능하므로 보다 완성도 높은 무선 사이트 구성이 가능하다.

- IBM Websphere Everyplace Suite[7]

신규 마크업 언어에 대한 지원과 변환 규칙의 설정을 지원한다.

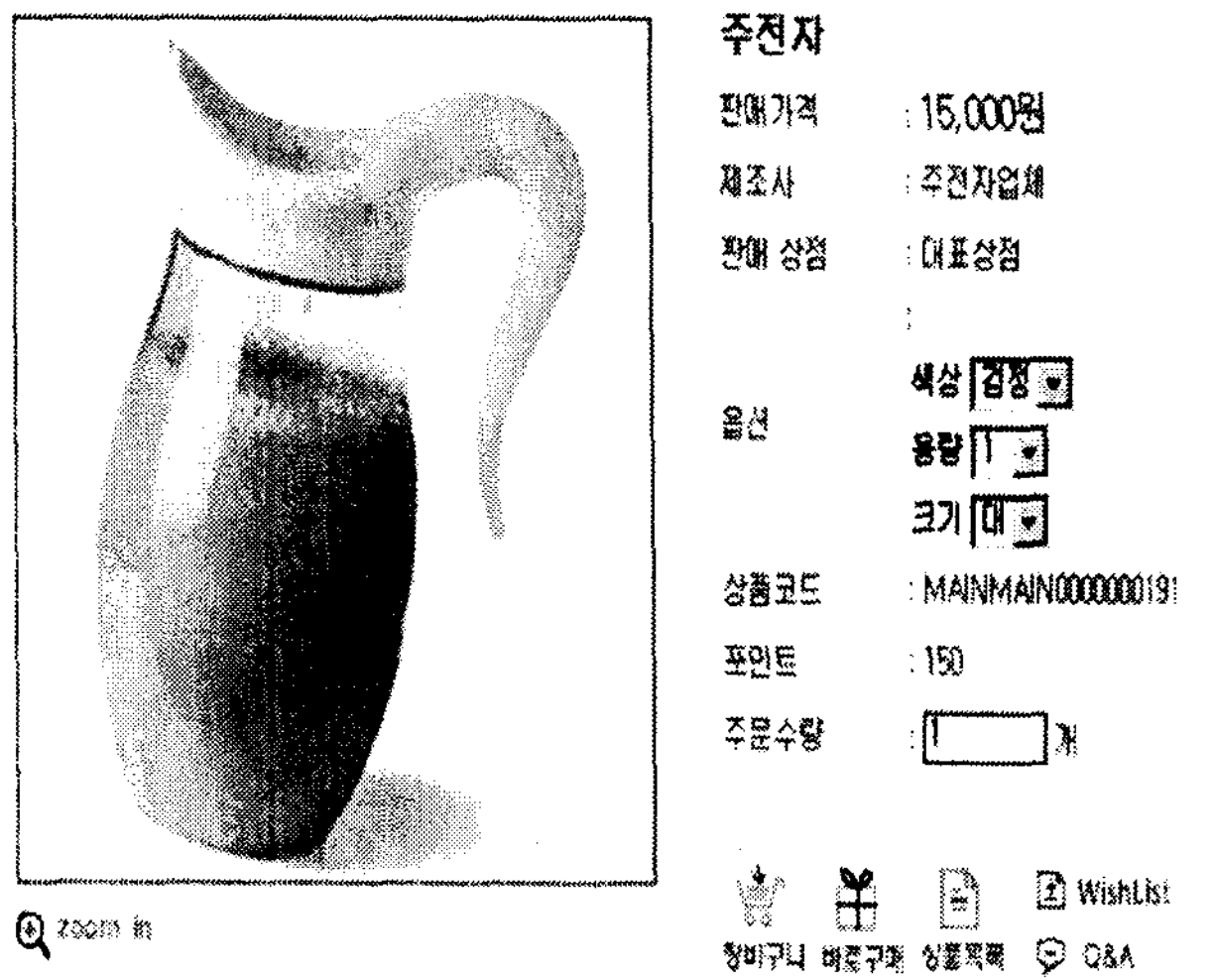
- OpenTV SpyGlass Prism[8]

컨텐츠들 중 변환의 대상이 되는 부분을 지정하고 지정된 부분만을 변환할 수 있는 설정 기능을 제공한다.

또한, CORBA를 이용한 분산 시스템으로 이루어져 있으므로 유연한 확장과 관리가 가능하다.

### 2.3. Filter System의 설계

현재 상용으로 나온 제품들은 모두 일반적인 Filter기능을 수행하고 있다. 하지만 일반적인 모든 문서들에 대한 Filter를 적용한다고 생각하면 출력된 WML문서의 품질에 대한 보장은 기대하기 힘들다. 또한 Configurable Filter의 경우엔 문서의 양이 방대할 경우 그 작업또한 번거롭지 않을 수 없다. 이에 본 논문에서 제안한 Filter System은 전자상거래 솔루션(現, Epion, 제품명: Equss)에서 기존의 상품구매를 위해 상품에 대한 HTML로 제작된 문서에 대하여만 적용 및 설계하였다. 현 전자 상거래 시스템의 상품구매를 위한 Html문서의 형식은 [그림1]과 같다.



[그림 1] 전자상거래 시스템에서 기본적인 상품구매화면

[그림1]의 HTML문서의 형식의 경우 가장 일반적

인 상품홍보 및 상품구매 화면이라 할 수 있다. 화면의 좌측 상단에 제품에 대한 이미지가 출력되고 우측 하단에는 제품을 구매하기 위한 가격정보, 수량정보, 색상정보등을 표시하고, 하단 부분에 제품에 대한 상품정보등이 일반적으로 표시되게 된다. 이러한 일반적인 문서의 경우에만 Filter System을 적용함으로써 출력물의 품질을 높이도록 한다.

제안된 필터 시스템의 구현을 위해서는, 기존의 HTML 문서를 분석하기 위한 파서가 필요한데 이는 IBM의 XML4J, 또는 SUN의 JAXP을 이용 SAX Parser 혹은 DOM Parser를 사용하도록 한다. 파서가 분석한 변환가능한 태그들을 규칙을 적용할 템플릿의 기준에 맞추어 변환하도록 한다.

적절한 태그로의 변환할 시 템플릿파일로부터 변환 규칙들을 읽어 들어 변환 하도록 설계되었으며 이는 특정 변환 규칙을 추가 또는 기존 규칙을 제거하게 함으로써 보다 확장가능 형식의, 유연한 System을 만들 수 있도록 한다.

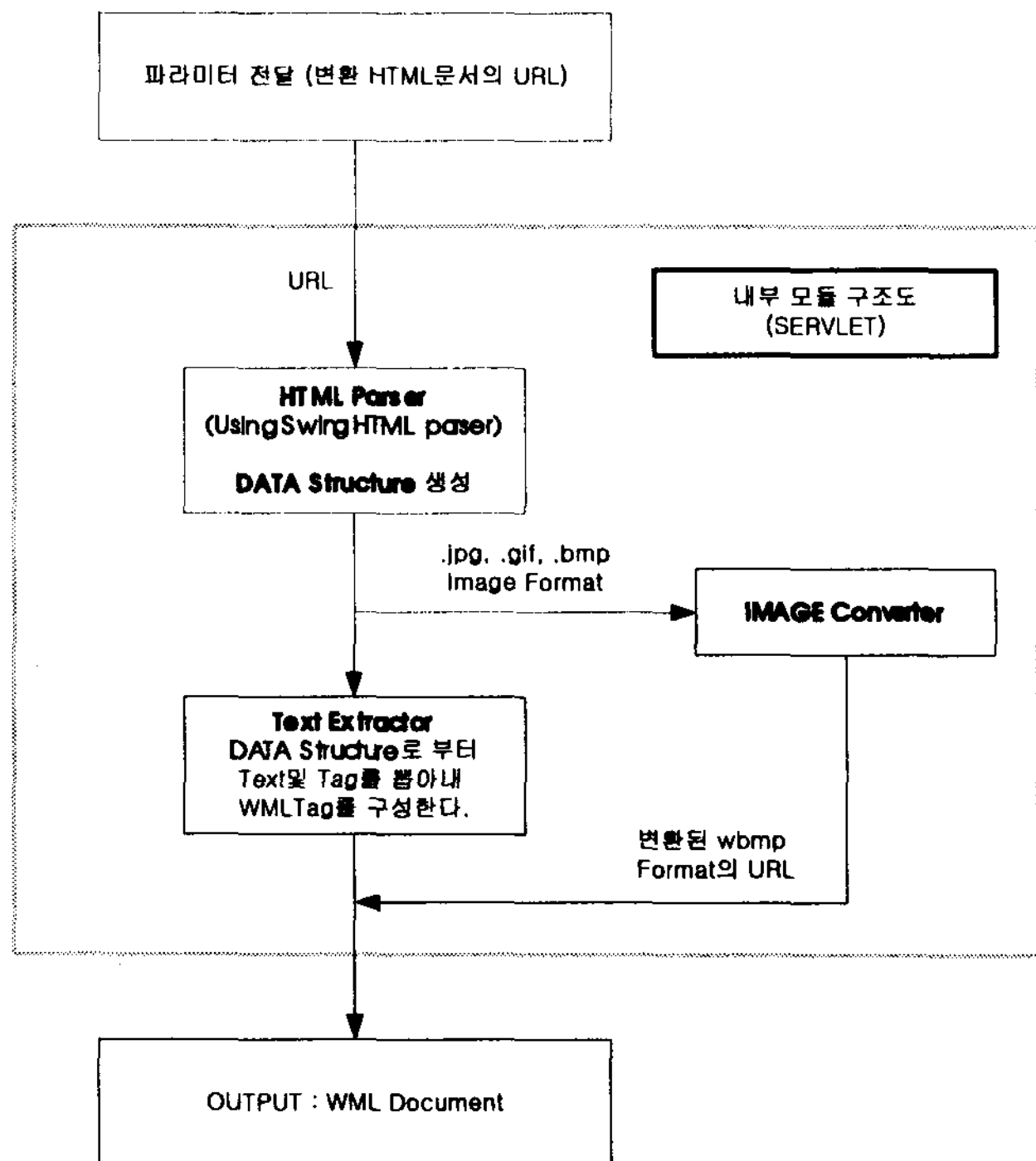
다음으로 고려해야 할 부분이 이미지 변환이다.

현재 무선인터넷을 위한 단말기들은

16bit(65000color)를 지원하는 단말기가 출시되고 있는 반면에, 범용적인 단말기들은 WBMP 형식의 파일들만 지원하고 있다. 따라서, HTML문서에서 추출한 제품 이미지(.gif, .jpg, .bmp 등)를 단말기에서 지원할 수 있는 포맷으로 변환할 수 있는 이미지변환 모듈이 필요하다. 이미지 변환 모듈은 확장을 고려하여 256 Color 변환 및 MONO Color로 변환할 수 있도록 설계하였다.[10] 다만, 이미지 변환 모듈에서 특히 고려해야할 사항은 단말기의 종류에 따른 변환 이미지의 크기 조절이다. 만약, Filter System이 Wap Gateway와 연동하여 동작한다면 연결되는 단말기의 정보를

CC/PP(Composite Capability/Prefrence Profiles)를 사용하거나 WSP(Wireless Session Protocol)헤더를 분석함으로써 알아 낼수 있다. CC/PP는 단말기에 적절한 컨텐츠를 제공하기 위해 클라이언트와 서버 사이에 주고 받는 정보를 정의 하기 위한 프레임워크로서 이를 이용하여 단말기의 특성 및 자원, 네트워크에 대한 정보등을 명시할 수 있다. 또한, WSP헤더에는 단말기의 모델정보 및 기타 정보들이 있으므로 이를 분석하여 단말기의 스크린 크기등을 알아 낼 수 있으나, Filter System은 단말기의 정보를 알아 낼 수 없으므로 단말기에 적절한 이미지로 변환을 고려하지 않는다. 위의 고려

사항들을 고려한 시스템 전체 구조는 [그림 2]와 같다.



[그림 2] 시스템 전체구조

HTML 문서의 URL 또는 File Path를 받아 Parsing Module에서 문서를 분석하게 된다. 전체 모듈은 서블릿으로 동작하며 전달될 파라미터(변환될 HTML문서의 URL)를 받아 해당 HTML문서를 로드하여 HTML Parser를 이용하여 HTML문서를 파싱한 후 최종 WML 문서를 출력하게 된다.

#### 2.4 세부 모듈별 구조 및 설명

전체 프로그램은 크게 다음과 같이 네가지 Module들로 구분되어진다.

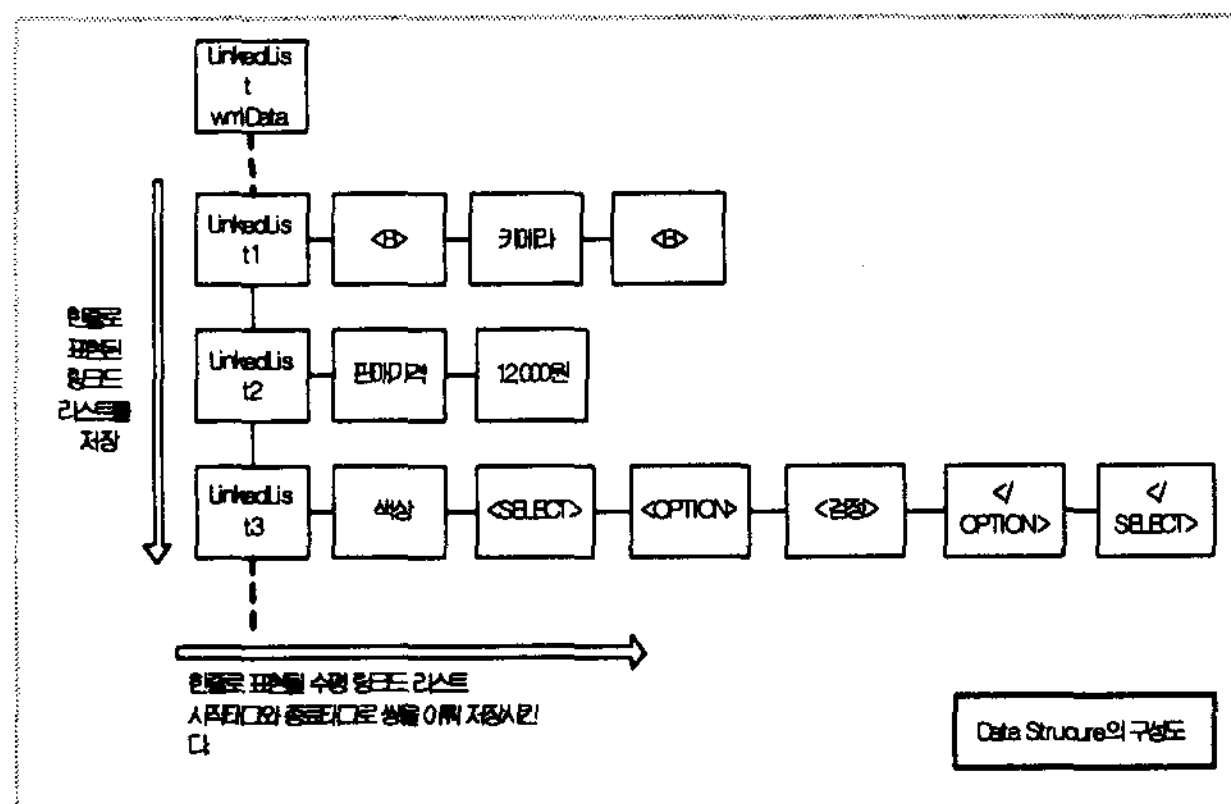
##### ● HTML Parsing Module

전체 프로그램에서 가장 중요한 모듈로 변환을 위해 입력된 HTML문서의 URL로 해당 HTML문서를 변환을 위해 메모리로 읽어 들인다. HTML파서로는 JDK내의 Swing에 포함된 HTML파서(javax.swing.text.html.parser)를 사용하여 HTML 문서로부터 변환에 필요한 태그 및 태그의 속성값, Text를 읽어와 데이터구조에 저장하게 된다.

##### ● Data Structure

HTML Parsing Module에서 읽어낸 태그 및 속성값, Text를 저장한다. 2개의 LinkedList를 사용하여 Data를 저장하는데, 하나의 LinkedList는 Sub LinkedList를 저장하고 저장된 LinkedList에는 한 줄로 변환될 HTML Tag 및 Text들이 입력되게 된다. [그림 3] HTML Tag일 경우 해당 속성값 또한 저장된다.

링크드 리스트에 저장되는 클래스, Tag나 Text별로 각각 저장되는데, 다음과 같다.



[그림 3] Data Structure의 자료 형태

##### ● Text Extractor Module

저장된 Data Structure에서 Data를 추출하여 WML Tag를 만드는 부분이다. 이미 Data를 입력하는 HTML Parsing Module에서 필요한 데이터만 추출하여 저장되었기 때문에 Data Structure에 저장된 모든 데이터를 출력하도록 한다. 단, Data Structure에 Text가 포함되지 않은 Data가 포함될 수 있기 때문에 이 자료는 Data Structure내에서 삭제하도록 한다. 이 부분에서 모든 WML태그를 구성하여 실제 폰 화면에 보여질 WML 태그를 출력하게 된다.

##### ● Image Converter Module

Data Structure내에 저장된 자료형 중에서 화면에 보여질 상품정보를 가르키는 URL을 찾아 내어 해당 URL을 이 모듈로 입력하면 해당 URL이 지시하는 이미지(.jpg, .gif, .bmp format)를 메모리로 읽어 들어 해당 이미지를 Width 120 Pixel이내의 WBMP Format으로 변환 시킨다.

Image 변환을 하기 위한 Class로 변환될 Image의



URL을 파라미터로 입력받아 해당 이미지를 메모리로 로드한후 이미지 변환 과정을 거치게 된다. Image변환에는 GINGCO New Media[12]사의 pic2wbmp에서 사용되는 API를 사용하여 변환 하였다. 이미지의 크기는 단말기의 좁은 액정화면에 나타나야 하기 때문에 크기는 가로 120 pixel의 크기로 제한을 하였다.

## 2.5 설계된 WMLTag Class 구조 및 설명

### ● WMLTag class

링크드 리스트에 저장되는 클래스. Tag나 Text별로 저장하는 클래스

```
class WMLTag {
    int type; // Tag 종류
    int checked; // option일 경우 checked
    boolean start; // 시작 태그인지 아닌지
    시작태그=TRUE 종료태그=FALSE
    TAGATTR attr; // Attribute
    String str; // Text
} // eof WMLTag class
```

### ● WMLTagDefine Class

Tag에 대한 데이터를 저장할 경우 저장되는 데이터(WMLTag)의 Type를 결정하기 위해 사용되는 태그들을 정의한 클래스

```
예) class WMLTagDefine {
    int WML = 00;
    int CARD = WML + 1;
    int FILDSET = WML + 2;
    ..
    int STR = WML + 30;
}
```

### ● TAGATTR Class

WMLTag Class에서 Text가 아닌 Tag에 대한 정보를 저장할 경우, Tag가 속성값을 가지고 있을 때 Tag의 속성(name)과 속성값(value)을 저장하기 위한 String 배열을 포함하고 있는 클래스

### ● TagForm Class

파라미터로 입력된 Tag와 Text, 속성 및 속성값이 있으면 입력받아, String값을 반환하는 메소드를

포함하고 있는 Class

## 3. 향후전망 및 개발방향

### 3.1 향후전망

유선인터넷과는 다른 여러 가지 제약속에서 무선인터넷의 발전 전망은 밝다고 할수 있다. 정보통신분야의 눈부신 기술 발전은 모든 컴퓨터간의 연결과 그에 의한 정보의 흐름을 보장하는 과정을 거쳐왔다. 이러한 발전 과정에서 불과 100년 사이에 PSTN, 인터넷, 이동통신 등 다양한 새로운 통신 기술 발전을 잇는 새로운 분야가 될것이며, 향후 정보통신 발전의 지향점이 될 것이다. 먼저 무선인터넷은 향후 유,무선 통합 환경에서 기본적인 통신 수단이 될 것이다. 현재까지 통신 네트워크는 유선 네트워크와 무선 네트워크, 인터넷이 별도의 네트워크로 발전하여 왔으며 개별 네트워크간의 상호 연동이 중심이 되었다. 현재 무선인터넷에서 주로 제공되는 서비스는 콘텐츠 응용형 서비스인 게임, 연예, 스포츠 등에 대한 단문 정보와 통신 응용형 서비스인 이메일 서비스 등이다. 이것은 앞으로 금융 응용 서비스인 인터넷 뱅킹과 보험이나 상거래 응용 서비스인 m-Commerce등 새로운 서비스로 발전하고 수직 시장에서는 Telemetry, Mobile Intranet등이 새로운 시장으로 소비자를 확보해갈 것이다.

### 3.2 개발 방향

본 논문에서 제안한 Filter System은 전자상거래 솔루션에서 만들어진 상품 구매를 위한 페이지를 무선 단말기에서 상품 검색 및 상품 구매를 할 수 있도록 하였고, 일반적인 상품 구매 화면을 Filter System을 통하여 일괄적으로 WML문서를 만드는 System을 제안하였다. 이를 이용하여 기존 전자상거래 솔루션을 유선인터넷만이 아닌 무선인터넷으로 확장시킬수 있다.

Filter System을 확장하여 일반적인 HTML문서의 필터링 기능을 실현 함으로써 가능하게 하고 함으로써 범용적인 Filter 시스템에 대한 구현에 대하여 연구할 예정이다.

### [참고문헌]

[1] Wireless Application Protocol Architecture,

- WAP Forum, Apr. 30. 1998.  
<http://www.wapforum.org>
- [2] Wireless Markup Language, WAP Forum,  
Apr.30. 1998. URL:<http://www.wapforum.org>
- [3] Imode, NTT Docomo,  
<http://www.nttdocomo.com/top.shtml>
- [4] Phone.com,  
WapGateway, <http://www.phone.com>
- [5] Argo,Argo  
Actigate, <http://www.argogroup.com>
- [6] Oracle, Oracle 9I Application Server  
WirelessEdition,  
<http://technet.oracle.com/products/iaswe/>
- [7] IBM, IBM Websphere Everyplace Suite,  
<http://www-3.ibm.com/pvc/products/wes/index.shtml>
- [8] SpyGlass Prism, OpenTV SpyGlass Prism,  
<http://www.opentv.com/products/technologies/prism/>
- [9] 권 오성, Converter : Integration between  
Wired Internet and Wireless Internet,  
[http://matilda.snu.ac.kr/doc/analysis/wap\\_converter.htm](http://matilda.snu.ac.kr/doc/analysis/wap_converter.htm)
- [10] Milan Sonka, Vaclav Hlavac and Roger Boyle, Image Processing, Analysis and Machine Vision, Chapman & Hall, 1993.
- [11] 무선인터넷 백서
- [12] GINGCO New Media  
<http://www.gingco-newmedia.de>