

XML 기반의 지능형 무선인터넷 문서 변환기

홍순재*, 조성진*, 김호윤**, 함준호***, 이재원****

*인하대학교 대학원 자동화공학전공

**인하대학교 대학원 첨단정밀공학전공

***인하대학교 자동화공학과 학부

****인하대학교 공과대학 기계항공자동화공학부

e-mail: hsradius@ids.auto.inha.ac.kr

XML-based Intelligent Reformatter for Mobile Access Documents

Soon-Jae Hong*, Seong-Jin Cho*, Ho-Yun Kim**, Jun-Ho Ham***,
Jae-Won Lee****

*Division of Automation Engineering, Post-graduate course of
Inha University

**Division of Advanced Precision Engineering, Post-graduate
course of Inha University

***Department of Automation Engineering, Inha University

****School of Mechanical, Aerospace and Automation Engineering,
College of Engineering, Inha University

요약

본 연구는 서로 다른 형식(format)으로 작성된 무선인터넷 문서들간의 상호 변환(reformatting)을 다루고 있다. 무선인터넷에 접속하는 단말기는 보통 하나의 형식만을 지원하는데 비해서, 무선인터넷 문서를 작성하는데 필요한 마크업 언어(markup language)로는 WML, HDML, mHTML 등 다양한 형식들이 존재하고 있다. 이에 본 연구에서 지식 기반 시스템(knowledge based system)과 XML(eXtensible Markup Language)을 이용하는 문서 형식 변환 기술을 제공하여, 무선인터넷 콘텐츠 구축시 특정 형식에 구애받지 않도록 하였다.

1. 서론

초창기 이동전화(mobile phone) 서비스는 거의 음성통화 위주였지만, 1990년대 말부터 음성 서비스 통화량의 성장 둔화와 더불어 데이터서비스의 통신량이 급격히 증가하고 있다[1]. 인터넷 이용의 저변 확대와 맞물린 이동전화 무선인터넷(mobile internet) 데이터서비스가 차세대 정보통신의 새로운 패러다임(paradigm)을 제시하게 된 것이다. 그러나, 이러한 데이터서비스를 무선인터넷 문서 형태로 작성하는데 필요한 마크업 언어(markup language)는 곧 무선인터넷 접속 프로토콜(protocol)의 문제로서 국내 이동통신사업자간에는 하나로 표준화되어 있지

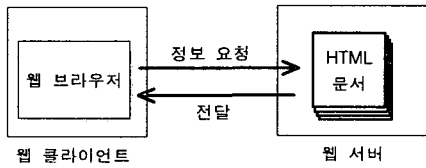
않다. 따라서 동일한 문서라도 이동통신사업자가 다르면 업체 고유의 마크업 언어로 문서를 작성해야 하는 불편이 따른다. 따라서 모든 무선단말기(wireless terminal)에서 접속 가능한 무선 인터넷 콘텐츠를 효율적으로 구축하기 위해서는 문서간의 변환(reformatting)이 필요하다. 이 부분에 대한 학술적 연구는 미미하나, 최근에 와서는 관련 상용 소프트웨어들이 등장하고 있다[2,3,4].

본 연구에서는 이러한 변환을 향상시키기 위한 연구로서 지식 기반 시스템(knowledge based system)과 XML[5]을 이용하고 시스템을 구현한다.

2. 무선인터넷 문서 변환

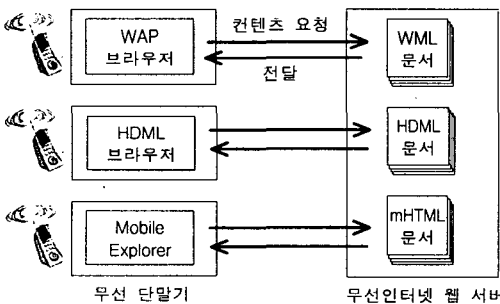
2.1 변환 기술의 필요성

전통적인 인터넷의 웹 서버(World Wide Web Server)는 정보를 요청하는 클라이언트(client)에게 준비된 문서의 내용을 전달하는 역할을 한다. 이때 웹 서버에서 다루는 문서는 (그림 1)과 같이 HTML(Hyper Text Markup Language)이라는 형식을 지닌다.



(그림 1) HTML 문서와 웹브라우저

무선인터넷에서는 이동전화 혹은 PDA(Personal Digital Assistant) 등의 무선 단말기를 클라이언트로 사용하며, 이동통신사의 무선 네트워크를 경유하여 기존의 유선 인터넷과 연결되는 구조이다. 무선인터넷에서도 전통적인 인터넷의 웹 서버와 유사한 무선인터넷 웹 서버가 필요하다. 무선인터넷 웹 서버에서 다루는 문서들은 WML[6], HDML[7], mHTML[8], SHTML[9] 등의 마크업 언어로 작성된다. WML, HDML은 무선 단말기 환경에 맞게 최적화 되어 있으며, mHTML, SHTML은 HTML 4.0의 서브셋(subset)이다. 특정 무선 단말기에 탑재되어 있는 브라우저는 상기의 다양한 언어 중 하나만으로 작성된 문서를 해석할 수 있다. 무선인터넷 마크업 언어로 작성된 문서들의 접근에 필요한 웹브라우저를 (그림 2)에 표현하였다.



(그림 2) 무선인터넷 마크업언어와 브라우저

국내의 5개 이동통신사는 WML, HDML, mHTML중 하나를 자사의 표준으로 채택하고 있다. 따라서 국내 콘텐츠 제공자(Contents Provider)는

무선인터넷 접속이 가능한 모든 이용자를 대상으로 콘텐츠를 제공하기 위해서 세 가지 형식에 맞춘 문서를 모두 갖추어야 한다. 이것은 비효율적이다.

2.2 변환 기술의 분석

무선인터넷 문서간의 상호 변환(reformatting) 기술은 문서 저작자의 의도를 최대한 유지하면서, 또 다른 형식을 준수하는 무선인터넷 문서를 만들어 내는 것이다. 변환 기술은 변환 대상과 결과에 따라서 <표 1>과 같이 세 가지로 구분될 수 있다.

변환 대상	변환 결과	필요 기술
DB, 텍스트	mHTML/WML/HDML	templating
mHTML/WML/HDML	mHTML/WML/HDML	비호환 요소 처리
HTML	mHTML/WML/HDML	필터링, 요약

<표 1> 변환 기술 분류

첫째는, DB 혹은 텍스트 자료로부터 무선인터넷 문서를 생성하는 것이다. 보통 미리 정해진 틀(template)에 따라 문서를 생성하는 비교적 간단한 방법이다. 둘째는, 특정 형식의 무선인터넷 문서를 다른 형식으로 변환하는 것이다. 문서 변환 시에 대상 마크업 언어들이 사용하는 요소(element)들 사이에 서로 1:1 대응하지 않는 요소들에 대한 처리 문제에 어려움이 있다. 셋째는, 기존 유선 인터넷의 HTML 문서를 무선인터넷 문서로 변환하는 것이다. HTML 문서의 복잡한 구조와 많은 양의 텍스트에 대하여 필터링(filtering) 및 요약(summarization)하는 것이 필요한 기술이다.

2.3 제품 분석

무선인터넷 문서 변환 관련 소프트웨어로는 다음과 같은 제품들이 있다.

- 아이비정보시스템의 MoblieNexus는 한가지 언어(형식)로만 콘텐츠를 제작하면, 이를 자동으로 무선 단말기의 언어에 맞게 변환하여 보여주는 게이트웨이(gateway)이다[2].
- 인사이드코리아의 iWap Converter는 HTML 콘텐츠를 WML/HDML 코드로 변환한다[3]. WML/HDML을 mHTML로 변환하는 기능은 없다.
- Oracle의 Portal-to-go는 콘텐츠 관리자가 입력한 DB 자료로부터 이용자가 가지고 있는 무선 단말기에 따라 WML, TinyHTML, VoxML갈

은 마크업 언어를 지원한다[4].

이 외에 기존 유선인터넷의 HTML과의 연계를 통하여 무선인터넷 보급 콘텐츠를 확대하려는 연구도 활발히 진행중이다[10,11,12].

본 연구에서는 국내에서 사용되는 무선인터넷 문서(WML/HDML/mHTML)간의 변환을 다룬다. 상기 언급된 바와 같이 변환 시에 발생할 수 있는 이종 언어의 사용 요소간의 변환 처리는 하나의 요소에 대한 맞대응 요소의 생성을 합성하는 기술이 필요하다. 이의 유연한 처리를 위하여 지식기반 시스템(knowledge-based system)을 이용하여 처리하는 WINTER (Wireless INTERNet) 시스템의 개발 연구를 수행하였다.

3. WINTER의 구현

3.1 시스템 설계

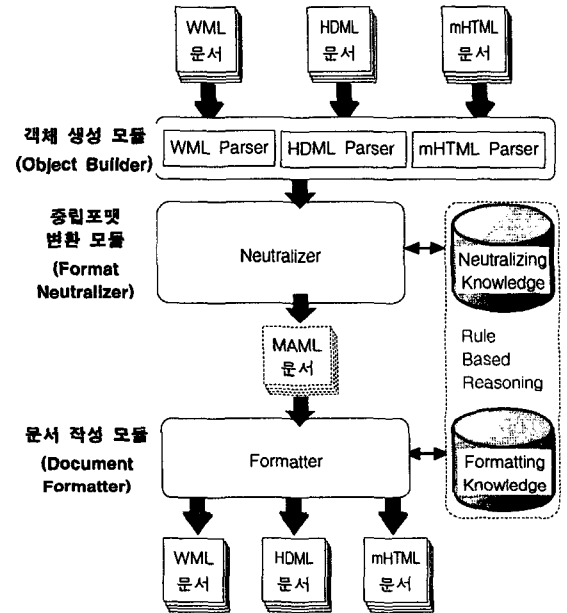
WINTER는 Java 언어의 객체 지향(Object-Oriented) 특성을 활용하여 무선인터넷 문서를 객체화(objectification)한 후, 다시 원하는 형식의 무선인터넷 문서로 포맷(format)하는 방법을 사용하였다. 객체화 단계와 문서포맷 단계 사이에서는 본 연구에서 정의한 MAML(Mobile Access Markup Language) 문서가 중간 산출물로 생성된다. 이 문서는 XML 어플리케이션으로, 모든 무선인터넷 문서의 내용과 동작을 표현할 수 있는 중립형식을 가지도록 정의되었다. MAML을 정의한 또 다른 이유는 형식간 변환이 종종 손실변환이라는 점이다. 문서 변환을 마친 후에 역변환이 필요한 경우 본래의 완전한 문서를 재현하지 못하는 경우가 발생한다. 변환 작업 시 산출되는 MAML 문서를 따로 저장해 두면, 역변환의 정확성을 높일 수 있다.

본 시스템은 단순 파싱(parsing)을 제외한 모든 단계에서 규칙기반추론(Rule Based Reasoning) 방법을 사용하도록 설계되었다. 규칙기반 추론은 서로 다른 무선인터넷 형식간의 단순 변환이 어렵다는 문제점을 해결하기 위해서 채택한 방법이다.

3.2 시스템 구조

시스템은 기능상 (그림 3)과 같이 세 부분으로 구성된다. 객체 생성 모듈(Object Builder)은 무선인터넷 문서(Mobile Access Document)를 읽어들이어 Java 객체로 변환하는 기능을 담당한다. 중립포맷 변환 모듈(Format Neutralizer)은 객체화된 무선인

넷 문서를 MAML 포맷으로 중립화한다. 그리고, 마지막으로 문서 작성 모듈(Document Formatter)이 MAML 문서로부터 원하는 형식의 문서를 생성한다. 중립포맷 변환 모듈과 문서작성 모듈은 모두 지식 베이스(knowledge base)를 기반으로 작동한다.

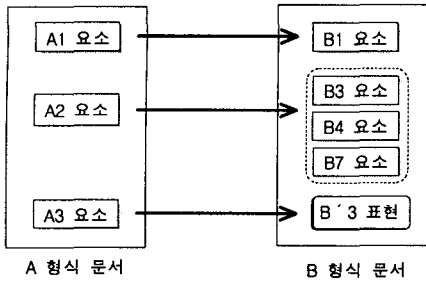


(그림 3) 시스템 구성

3.3 문서 변환 방법

서로 다른 문서의 요소(element)중 개념이 같은 부분은 단순한 1:1 맵핑(mapping)이 가능하다. 이 방법으로 약 50%의 변환이 가능하다. 각 문서의 요소가 1:1로 들어맞지는 않지만, 어느 한쪽의 요소를 몇 가지 결합하면 유사 기능으로 변환할 수 있는 경우도 있다. 이러한 1:N 맵핑 방법이 첫 번째 방법과 같이 사용되면, 약 80%까지의 변환이 가능해진다.

그러나 무선인터넷 문서 형식들간에는 개념상 아예 서로 호환되지 않는 부분들이 있다. 따라서 이 경우는 변환이 불가하므로 문서 변환 시에 기능 상실이 발생된다. 기존의 알고리즘적(algorithmic) 방법으로는 해결하기 어려운 문제이다. 본 연구에서는 이 문제를 해결하기 위해 규칙기반추론(rule-based reasoning) 방법을 사용하였다. 예를 들어, WML의 변수 처리를 mHTML로 표현할 때는 시스템의 기억장소를 변수 값 입출력 용도로 사용한다. 이상의 세 가지 방법을 (그림 4)에 표현하였다.



(그림 4) 변환 방법

본 연구에서는, 세 가지 방법을 모두 사용하여 약 95%까지 변환 율을 높일 수 있다는 결론을 얻었다. 나머지 5%는 S 이동통신사에서 자사에 맞게 특화된 WAP 표준을 채택하고 있기 때문에 생긴 문제이다. 하지만, 이 5%는 심각한 장애를 발생하는 부분이 아니기 때문에 본 논문에서는 논외로 한다.

3.4 시스템의 특징

본 논문에서 제시한 지능형 무선인터넷 문서변환 방법은 기존의 변환 방법에 비해 다음의 특징을 가지고 있다.

- 유연한 확장성 : 규칙기반추론을 기반으로 동작하기 때문에, 규칙이 새롭게 추가되는 것만으로 시스템의 업그레이드가 이루어진다. 따라서 무선인터넷 표준의 새로운 버전(version)으로 업그레이드(upgrade)에 유연하게 대응할 수 있다. 절차적인 언어를 이용하는 기존의 방법은 코드를 수정하고 재컴파일 하는 번거로움이 따른다.
- 모듈성 : Java 기반의 객체지향 방법으로 구현되었기 때문에, 각각의 모듈은 또 다른 무선인터넷 응용 프로그램에서 유용하게 재 사용될 수 있다.

3.5 구현환경

본 시스템은 다음과 같은 환경에서 구현되었다.

- Java 언어 (JDK 1.2.2)
- Jess 5.1

Jess[13]는 Java 기반 전문가시스템 셸로서, 문제의 상황에 준하여 전향추론(forward reasoning) 혹은 후향추론(backward reasoning)이 이용되었다.

4. 결론 및 향후연구

무선인터넷 문서 변환 시스템은 표준화되지 않고

있는 무선인터넷 시장에서 사실상 콘텐츠 형식의 통일이라는 의미를 가질 수 있다. 최근에 와서야 이와 유사한 역할을 하는 변환 소프트웨어가 출시되고 있지만, 대부분은 호환되지 않는 요소를 제거해 버리는 필터링 방식이다. 따라서 본 연구는 지식기반 추론을 기반으로 한 지능형 문서변환 시스템을 개발함으로써 변환의 깊이를 향상시켰다.

본 시스템 보완에 필요한 향후 연구로는 다음과 같은 부분들을 고려하고 있다.

- 변환규칙의 생성에 관한 연구
- HTML 문서로부터 무선인터넷 문서로 변환하는 기능 추가
- 외부 전문가시스템 셸을 사용함으로써 발생하는 시간 지체 최소화화에 관한 연구

참고문헌

- [1] 전자부품연구원 기술기획팀, "무선인터넷 산업 동향", 주간전자정보, V.3, N.7, 2000.03.09.
- [2] <http://www.ibe.co.kr>
- [3] <http://www.insidekorea.net>
- [4] <http://www.oracle.co.kr>
- [5] "Extensible Markup Language", W3C Proposed Recommendation, Feb. 10, 1998. URL : <http://www.w3.org/TR/REC-xml>
- [6] "Wireless Markup Language Specification", WAP Forum, Jun. 16, 1999. URL : <http://www.wapforum.org>
- [7] "Handheld Device Markup Language Specification", May 9, 1997. URL : <http://www.w3.org/TR/NOTE-Submission-HDML-spec.html>
- [8] "Microsoft® Mobile Explorer 1.0 Specification", Microsoft, May 31, 1999.
- [9] "SHTML Specification", 삼성전자, 1999.
- [10] H Ouahid, A. Karmouch, "Converting Web Pages into Well-formed XML Documents", Proceedings of the 1999 IEEE International Conference on Communications, Vol.1 , pp.676-680 , Jun. 6, 1999.
- [11] Marcin Metter, Dr Robert Colomb, "WAP enabling existing HTML applications", Proceedings of the First Australasian User Interface Conference, pp.49-57 , 2000.
- [12] Bickmore T, Girgensohn A, Sullivan JW, "Web page filtering and re-authoring for mobile users", Computer Journal , V.42, N.6, pp.534-546, 1999.
- [13] <http://herzberg.ca.sandia.gov/jess>