

# 웹 콘텐츠 변환을 위한 태그 변환기의 설계 및 구현

이미경, 장민수, 김재홍, 손주찬\*

\* 한국 전자 통신 연구원 인터넷 컴퓨팅 연구부  
e-mail : lmk63398@etri.re.kr

## Design and Implementation of Tag Converter for Transformation of Web Contents

Mi-Kyoung Lee Min-Su Jang Jae-Hong Kim Ju-Chan Sohn\*  
\*Dept. of Internet Computing, ETRI

### 요 약

웹상의 콘텐츠에 인터넷 브라우저가 아닌 다양한 기기종의 클라이언트 장치가 접근하더라도, 그 장치에 필요한 형식의 콘텐츠를 실시간으로 변환해주는 문서 변환 시스템(Contents Adaptation & Transformation System)을 구현하였다. 본 논문에서 구현한 태그 변환기는 CATS 문서 변환 시스템에서 마크업 언어간의 태그 변환 기능을 수행한다. 태그 변환기는 변환 기능의 확장성 및 유연성을 확보하기 위해 태그간 변환 정보를 태그 변환기와 분리하여 관리하고 있다. 원본 문서가 컨텍스트 정보와 함께 태그 변환기로 입력되면 태그 변환기는 설정파일을 참고하여 태그 변환용 XSL 문서들 중에서 변환에 사용할 XSL 문서를 결정하고 프로세서는 원본문서와 XSL 문서를 이용하여 새로운 문서를 생성하게 된다. 현재 태그 변환기는 HTML로 작성된 문서를 XHTMLBasic, WML, mHTML로 변환해주고 무선 콘텐츠간의 변환(mHTML $\leftrightarrow$  WML)도 지원해준다. 그리고 본 태그 변환기의 변환 규칙을 구현한 XSL 문서는 유효하지 않은 형태의 HTML 문서의 변환을 지원해주고 있다.

### 1. 서론

모바일 단말기를 이용하여 언제 어디서나 인터넷 서비스에 접근하여 서비스를 사용할 수 있는 무선 인터넷 서비스는 현재 빠른 성장을 보이면서 발전하고 있으며 많은 이용자를 가지고 있다.

무선 인터넷 서비스 사용이 활성화되면서 많은 양의 콘텐츠가 제공되고 있지만, 현재의 무선 콘텐츠들은 서로 다른 무선인터넷 방식에 따른 무선 마크업 언어들로 새롭게 구축되고 있다. 현재 우리나라에서 무선 인터넷을 제공하는 이동통신 업체들이 채택하고 있는 표준들을 살펴보면 SKTelecom과 LGTelecom의 WAP 방식과 KTF에서 채택한 ME 방식으로 크게 나눌 수 있다. 그리고 SKTelecom은 WAPForum에서 제안한 WML1.1을 기준으로 제작된 콘텐츠로 서비스를 제공하고 있으며, LGTelecom은 OPENWAVE사(PHONE.COM)에서 제안한 확장된 형식의 WML1.1을 이용하여 제작된 콘텐츠로 서비스를 제공하고 있다.

그리고 KTF에서는 HTML의 서브셋으로 볼 수 있는 mHTML을 이용하여 제작된 콘텐츠를 제공하고 있다 [1].

기존의 유선 인터넷을 기반으로 제작된 많은 양질의 콘텐츠들은 느린 전송 속도, 적은 메모리, 작은 스크린을 갖는 모바일 단말기에서 이용되기에는 많은 제한점을 가지고 있다. 그리고 지금까지 무선 인터넷 콘텐츠를 새롭게 구축하기 위해서는 추가로 엄청난 시간과 비용이 들게 된다. 이러한 문제점을 해결하기 위해서 서로 다른 콘텐츠간의 변환을 중간에서 중재할 적절한 콘텐츠 변환기가 필요하다.

따라서, 이미 만들어진 콘텐츠를 유무선 환경에 상관없이 다양한 기기종의 클라이언트 장치에서 이용할 수 있도록 실시간으로 변환해주는 웹 서버용 유무선 문서 변환 시스템을 만들었다[2]. 본 논문에서는, 콘텐츠 변환 시스템에서 문서를 변환할 때 마크업 언어간의 태그 변환 기능을 수행하는 태그 변환기를 설계하

고 구현한다.

## 2. 웹 콘텐츠 변환을 위한 시스템의 구조

CATS(Contents Adaptation & Transformation System)는 웹 콘텐츠 변환 시스템으로 단일 웹 콘텐츠를 다양한 사용자 장치에 맞는 형태로 변형 및 재단하는 기능을 수행하는 시스템이다. 본 시스템은 기존의 웹 콘텐츠에 여러 가지 다른 기종의 클라이언트 장치가 접근하더라도 그에 적절한 콘텐츠를 자동으로 생성하여 제공해주는 역할을 한다. 이 시스템에서는 HTTP 로 전송되는 장치 프로파일, 사용자 프로파일, 네트워크 프로파일을 이용하여 디바이스의 정보를 파악한 후, 웹 서버에 존재하는 콘텐츠를 그에 적절한 포맷의 출력 문서로 변환하여 이를 클라이언트 장치에 전송하는 작업을 한다. CATS 는 그림 1 과 같은 기본 구조를 가지고 있다[2].

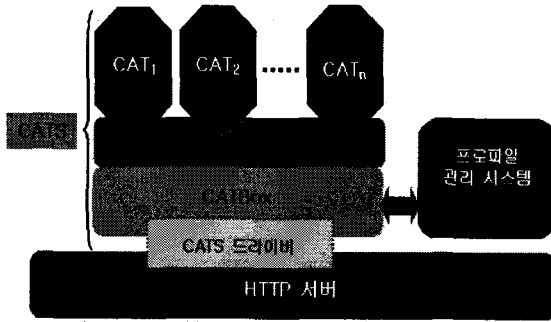


그림 1. CATS 의 기본 구조

### 2.1 HTTP 서버

CATS 는 HTTP 요청을 처리하기 위해 HTTP 서버와 연동하고 있다. 본 시스템은 Tomcat4.1 을 HTTP 서버로 이용하고 있다.

### 2.2 CATS 드라이버

CATS 드라이버는 CATS 와 HTTP 서버간 연동을 담당한다. 이 드라이버의 역할은 HTTP 요청과 응답을 가로채어 CATS 로 전달하고 변환된 결과를 HTTP 서버로 되돌려주는 역할을 한다.

### 2.3 CAT API

CAT API 는 CAT 인터페이스, CAT 요청 및 응답 인터페이스, 콘텐츠 정보 관련 인터페이스로 구성된다.

### 2.4 CAT

CAT 은 여러가지 콘텐츠 변환 작업을 구현한 개별 변환 컴포넌트를 말한다. 문서가 변환될 때 이용되는 여러 컴포넌트들로, 태그를 변환하는 기능을 수행하는 태그 변환기, 문서를 단말기에 보여지기 적절한 크기로 자르는 문서 분할기, 각 클라이언트 단말기가 지원하는 이미지 형식으로 변환하여 이미지를 전송해주는 이미지 변환기 등의 여러 시스템들로 구성되어 있다.

## 2.5 CATBox

CATBox 는 컨테이너인 동시에 전체 변환 작업의 흐름을 관리하는 관리자의 역할을 한다. 이것은 CAT 의 설치, 생성, 적재, 실행, 관리, 제거하는 역할을 한다. 그리고 프로파일 정보를 비롯한 각종 컨텍스트 정보를 교환하는 통로 역할을 수행한다.

## 2.6 프로파일 관리 시스템

프로파일 관리 시스템은 변환 작업에 필요한 각종 프로파일 정보를 관리하는 CATS 의 구성 시스템이다. 이 시스템은 프로파일 정보들을 클라이언트 장치, 웹 서버 및 기타 소스로부터 입수한 다음 해석하여 변환 작업에 참조할 수 있도록 준비하며 프로파일 정보 기술 및 해석, 프로파일 정보 정합의 기능을 가진다.

## 3. 태그 변환기의 구조

콘텐츠의 태그간 변환을 담당하는 태그 변환기에 대해서 설명하겠다. 태그 변환기는 입력 문서와 그 문서를 변환하는데 필요한 정보를 받아서, 변환 규칙 중에서 알맞은 XSL 문서를 찾고, XSL 문서의 규칙에 따라서 출력 문서로 변환해 주는 역할을 한다. 태그 변환기는 CATBox 로부터 태그 변환에 관련된 정보를 받아서 변환을 실행한다. 태그 변환기의 구조는 아래 그림 2 와 같다.

원본 문서가 CatResponse 객체로 컨텍스트 정보와 함께 태그 변환기(TagConverterCat)로 입력되면 태그 변환기는 설정파일을 참고하여 변환에 사용될 XSL 을 선택하고 선택한 XSL 문서의 위치(URI)를 파악하게 된다. 변환에 사용할 XSL 문서를 결정하게 되면 태그 변환기는 원본문서와 태그 변환 XSL 문서를 이용하여 Apache Xalan XSLT processor를 실행시켜 새로운 마크업 문서를 생성하게 된다.

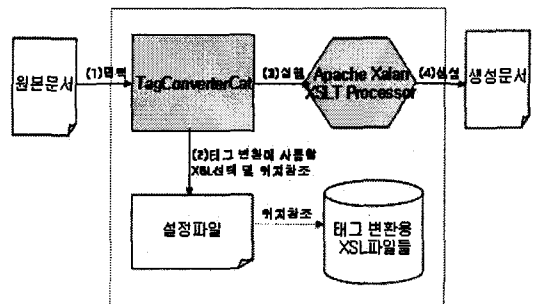


그림 2. 태그 변환기의 구조

태그 변환기에서는 변환 기능의 확장성 및 유연성을 확보하기 위해 태그간 변환 정보를 태그 변환기와 분리하여 관리하며, 변환 규칙은 변환을 지원할 마크업 언어에 따라 미리 저장되어 있다. 따라서 새로운 마크업 언어에 대한 변환 기능이 추가 되거나 수정되더라도 쉽게 적용이 가능해진다.

#### 4. 태그 변환용 XSL

태그 변환용 XSL 문서들은 CATS 변환 시스템에서 변환이 지원되는 언어들에 대한 변환 XSL 문서로 각 태그들의 변환 규칙에 의해 미리 저장되어 있다.

현재 시스템에서 지원되는 변환으로는 HTML 문서를 XHTMLBasic, SKWML (SKTelecom 전용 WML1.1), UPWML(LGTelecom 전용 WML1.1), WML2.0 으로 변환이 가능하며, WML 과 mHTML 의 무선 콘텐츠간의 변환도 지원하고 있다[1, 3, 4, 5, 6, 7].

##### 4.1 태그 변환 규칙

태그 변환 규칙은 태그들 간의 유지, 변환, 삭제의 3 가지 규칙으로 나눌 수 있다. 그리고 태그간의 사상 정보, 태그간의 의존 관계, 원본 태그의 속성의 변환, 태그 속성값의 텍스트 콘텐츠 변환 등의 정보들을 참조하여 만들었다.

변환 규칙을 만들 때 가장 중요한 것은 콘텐츠의 내용의 손실을 없애고 최대한 유사하게 표현하는 것에 중점을 두었기 때문에, 반드시 필요한 태그이지만, 유사한 태그가 없을 경우 조금 다른 기능을 가지더라도 그 역할을 할 수 있는 태그들로 대체시키고, 기능이 비슷한 태그는 기능을 유지하는 태그로 변환시키고, 기능면에서 불필요한 태그들은 삭제 시켰다. 예를 들면, HTML 의 체크박스과 라디오버튼의 경우는 WML 로 변환시 지원되지 않지만, WML 에서 선택에 이용되는 비슷한 기능을 하는 <select>,<option>태그로 대체 할 수 있다.

대부분의 유선 환경에서만 필요한 태그들은 삭제시키나 대체하였는데, 예를 들면 디스플레이에 관련된 모든 태그들은 태그를 삭제하고, 프레임 태그<frame>는 소스코드의 링크를 <a>태그로 대체하여 링크를 뽑아내고 테이블의 경우는 작은 모바일 단말기 브라우저에서는 보여주기가 어려우므로 리스트 태그<ol><li>로 변경해 주었다.

변환 규칙은 HTML4.01, WML1.1, mHTML1.1.3, XHTMLBasic 의 스펙을 기준으로 마크업 언어의 전체 태그에 대한 변환 기능을 구현해 주었다[4, 5, 6, 7].

##### 4.2 태그 변환 XSL

태그 변환 XSL 들은 HTML 문서를 XHTMLBasic, SKWML, UPWML, mHTML 로의 변환을 지원하는 유무선간 변환 XSL 문서들, 그리고 무선간의 변환을 지원하는 mHTML ~ SKWML, mHTML ~ UPWML, WML ~ mHTML 로 구성되어 있다. 유무선간의 변환에서는 HTML 문서를 이용하여 변환하기 때문에, HTML 문서의 Well-formed 를 맞춰준다.

그리고 현재 만들어진 XSL 문서들은 변환될 마크업 언어의 DTD 를 기준으로 만들어졌기 때문에, DTD 에 유효하지 않은 문서들은 정확한 변환을 나타내지 못한다. 하지만 웹 콘텐츠를 나타내는 거의 대부분의 HTML 문서들은 유효하지 않은 형태이므로, 본 XSL 문서들은 유효하지 않은 문서에 대한 변환 기능을 모두 처리해 주며, 유효하지 않은 형태의 문서를 변환하

여 유효한 형태의 XML 문서로 생성해주는 기능을 가지고 있다. 아래의 그림 3 은 변환 XSL 문서들 중, HTML2WML XSL 문서의 일부분을 보여주고 있다.

```
<?xml version="1.0" encoding="UTF-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
<xsl:output method="xml" encoding="EUC-KR" coctype="public" ../WAP-FORUM/DTD/WML1.1//EN"
doctype-system="http://www.wapforum.org/DTD/wml_1.1.xml" indent="yes" />
<xsl:key name="radiobutton" match="input" use="@type" />
<xsl:key name="checkboxbox" match="input" use="@type" />
<!-- =====>
<!-->
<!-->
<!-->
<xsl:template match="html">
<wml>
<xsl:if test="boolean(@lang)">
<xsl:attribute name="xml:lang">
<xsl:value-of select="@lang" />
</xsl:if>
<xsl:apply-templates select="body|frameset" />
</wml>
</xsl:template>
<!--
PDA에서는 card 부분이 2 레코드를 띄어준다
단말기의 화면폭은, screen...
-->
<xsl:template match="body">
<card>
<xsl:if test="boolean(@lang)">
<xsl:attribute name="xml:lang">
<xsl:value-of select="@lang" />
</xsl:if>
</xsl:template>
```

그림 3. HTML2SKWML XSL 의 일부

#### 5. 태그 변환기의 구현 실험 결과

태그 변환기는 Java(JDK1.3)를 이용하여 구현되었으며, Apache Xalan XSLT Processor 를 이용한다. CATS 변환 시스템은 Tomcat4.1 환경에서 실행되며, 본 시스템을 테스트하기 위해서 유선 환경에서의 접속은 Internet Explorer6.0 을 이용하였고, 모바일 클라이언트 단말기로는 ME1.3 브라우저를 가지는 KTF 전용 SPH-5900 단말기를, WAP 브라우저를 가지는 SKTelecom 전용 SHP-V300, iPAQ PDA 를 이용하였고, 시뮬레이터로는 ME 용 브라우저인 MME3.0 과 WAP 용 브라우저인 UP.SDK 4.0 을 이용하여 테스트 해 보았다.

그림 4 는 Internet Browser 를 이용하여 변환 시스템이 실행되는 웹 서버의 콘텐츠에 접속한 화면이다. 웹 서버에 설치된 콘텐츠는 보험 업무를 수행하는 포털 사이트로 그림 4 는 사용자의 정보를 입력하는 페이지이다.

## Information Gathering

Items	Details
Sex	Male <input checked="" type="checkbox"/>
Driver	<input type="radio"/> Insurant <input type="radio"/> None
Accident No.	<input type="text"/>

Memo

Finish

그림 4. HTML 문서

그림 5 에 보여지는 화면은 UP.SDK4.0 과 MME3.0 에플래이터를 이용하여 그림 4 와 같은 페이지를 요청했을 때 보여지는 화면이다. 각 에플래이터에서는 디스플레이창의 크기가 작기 때문에, 테이블은 리스트 형태로 바뀐 것을 볼 수 있다. 그리고 테이블의 제목을 표시하기 위해서 WML 의 경우는 제목을 진하게, mHTML 의 경우에는 폰트를 진하게 하는 기능이 없기 때문에 [ ]를 사용하여 제목임을 나타내었다.

컨텐츠에 접근하는 단말기 장치의 특성에 따라서 변환 시스템에서는 실시간으로 장치에서 필요로 하는 페이지를 생성해서 전송해준다.

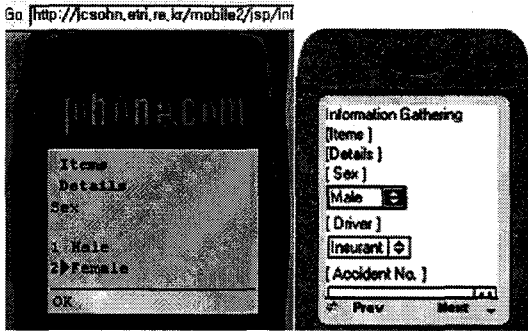


그림 5. WAP Browser 와 ME Browser 의 결과

## 6. 결론 및 향후 과제

본 논문에서는 웹 컨텐츠를 유무선 환경에 상관없이 기기종의 클라이언트 장치에서 이용할 수 있도록 중간에서 컨텐츠를 변환하는 도구 및 컨텐츠를 저작할 때 도움을 주는 유무선 문서 변환 시스템에서 입력 문서의 마크업 언어를 클라이언트 장치에서 필요한 마크업 언어로 변환하는 기능을 하는 태그 변환기를 설계 및 구현하였다. 본 태그 변환기는 태그 변환기와 태그 변환 규칙으로 나뉘어 있으며, CATS 에 입력된 문서를 태그 변환 규칙에 의해 구현된 태그 변환용 XSL 을 이용하여 장치가 요청하는 형식의 컨텐츠 문서로 변환해주는 역할을 한다. 태그 변환용 XSL 문서들은 미리 정해진 태그 변환 규칙에 따라 구성되어 있으며, CATBox 로부터 태그 변환 관련 정보를 받아서 변환에 적합한 태그 변환 규칙을 가진 XSL 을 선택하여 태그 변환을 실행한다. 본 변환기는 태그 변환 기능과 변환 규칙이 분리되어 있어서 변환 기능의 확장성과 유연성을 가지고 있으며, 새로운 언어에 대한 변환규칙의 추가와 삭제가 용이한 장점을 가진다. 그리고 XSL 을 이용하여 만들어졌기 때문에 변환 규칙에 독립성을 가진다. 현재 태그 변환기의 변환 규칙으로는 HTML to XHTMLBasic, HTML to SKWML, HTML to UPWML, HTML to mHTML, mHTML to SKWML, mHTML to UPWML, WML to mHTML, WML2.0 을 지원하고 있다.

그리고 본 논문의 태그 변환기에서 이용하는 XSL 문서는 DTD 에 유효하지 않은 HTML 문서의 변환도 100% 지원해주는 장점을 가진다.

향후, 변환 규칙 XSL 문서는 마크업 언어 단위의 변환 규칙이 아닌, XHTML 의 모듈성을 이용한 모듈 단위의 변환 규칙을 지원할 수 있도록 보완할 계획이다.

## 참고문헌

- [1] “무선 인터넷 홈페이지 만들기”, (주) 애니빌 무선 인터넷 연구소, 영진출판사, 2002.
- [2] J.H.Kim “Internet Content transcoding framework for heterogeneous client devices”, CALS/EC KOREA 2001 International Conference, pp.379-391, 2001.
- [3] “XSL Transformation(XSLT) Version 1.0”, W3C Recommendation, 16, November 1999. <http://www.w3.org/TR/XSLT>
- [4] “HTML4.01 specification” W3C Recommendation, 24, December 1999. <http://www.w3.org/TR/html4>
- [5] “XHTML Basic”, W3C Recommendation, 19 December 2000, <http://www.w3.org/TR/xhtml-basic>
- [6] “Wireless Application Protocol Wireless Markup Language Specification Version 1.1”, Version16-Jun-1999, WAPForum, <http://www1.wapforum.org/tech/documents/SPEC-WML-19990616.pdf>
- [7] “KTF 무선 Contents 개발 권고안 ME1.1.1 용”, 2000.