

# 유무선 문서 변환 시스템의 태그변환기 설계 및 구현

이미정<sup>0</sup> 장민수 김재홍 손주찬  
한국전자통신연구원 비즈니스지식처리연구팀  
(1mk63398<sup>0</sup>, minsu, jhkim504, jcsohn)@etri.re.kr

## Design and Implementation of Tag Converter for Contents Adaptation & Transformation System

Mi-Kyoung Lee<sup>0</sup> Min-Su Jang Jae-Hong Kim Ju-Chan Sohn  
Business Knowledge Research Team, ETRI

### 요 약

무선 인터넷의 발달에 따라 이미 만들어진 유선 인터넷용 콘텐츠들을 유무선 환경에 상관없이 이용할 수 있도록 콘텐츠를 변환하는 도구 및 콘텐츠를 저작할 때 도움을 주는 유무선 문서 변환 시스템을 만들고자 한다. 본 논문에서는 유무선 문서 변환 시스템에서 마크업 언어간의 태그 변환 기능을 수행하는 태그 변환기를 설계 및 구현하였다. 본 변환기는 변환 기능의 확장성 및 유연성을 확보하기 위해 태그간 변환 정보를 태그 변환기와 분리하여 관리하며, 미리 지정된 태그 변환 규칙에 따라 태그를 변환한다. 태그 변환 규칙은 모듈 단위로 설계되며 태그 간의 사상정보, 의존관계, 속성 정보에 따라 정의되었으며 모듈별 태그의 유지, 변환, 삭제의 3가지 규칙을 가진다.

### 1. 서론

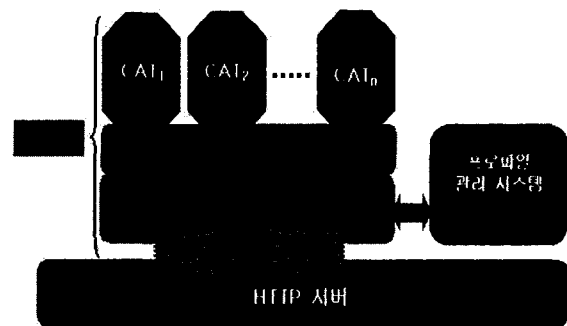
현재 이동통신 사용자의 급증과 유선 인터넷의 급속한 발전으로 모바일 단말기를 이용하여 언제 어디서나 인터넷 서비스에 접근할 수 있는 무선 인터넷(Wireless Internet) 서비스는 빠른 성장을 보이고 있다.

무선 인터넷 서비스 사용이 활성화되기 위해서는 많은 양질의 콘텐츠가 제공되어야 하지만, 기존의 유선 인터넷을 기반으로 제작된 많은 콘텐츠들은 느린 전송 속도, 적은 메모리, 작은 스크린을 갖는 모바일 단말기에서 이용되기에는 많은 제한점을 가지고 있다. 그렇다고 무선 인터넷용 콘텐츠를 새롭게 구축하기 위해서는 추가로 엄청난 시간과 비용이 들게 된다. 화려한 그래픽과 스크립트 언어를 이용하여 작성된 기존의 유선 인터넷용 콘텐츠를 사용하려면 PDA나 모바일 단말기에서 제공되는 성능이나 마크업 언어, 브라우저의 지원 능력 등의 문제로 제대로 된 서비스가 일어나지 않으므로 이러한 문제를 해결하기 위해서 유무선 콘텐츠간의 변환과 여러 종류의 무선 콘텐츠간의 변환을 중간에서 중재할 적절한 콘텐츠 변환기가 필요하다.

따라서, 이미 만들어진 콘텐츠들을 유무선 환경에 상관없이 이용할 수 있도록 콘텐츠를 변환하는 도구 및 콘텐츠를 저작할 때 도움을 주는 유무선 문서 변환 시스템을 만들고자 한다. 본 논문에서는, 유무선 문서변환 시스템에서 문서를 변환할 때, 마크업 언어간의 태그 변환 기능을 수행하는 태그 변환기를 설계하고 구현한다.

### 2. 유무선 문서 변환 시스템의 구조

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



[그림 1] CATS의 기본 구조

#### 2.1 HTTP 서버

CATS는 HTTP 요청을 처리하기 위해 HTTP 서버와

연동해야한다.

### 2.2 CATS 드라이버

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

### 2.3 CAT API

CAT API는 CAT 인터페이스, CAT 요청 및 응답 인터페이스, 컨텍스트 정보 관련 인터페이스로 구성된다.

### 2.4 CAT

CAT은 여러가지 콘텐츠 변환 작업을 구현한 개별 변환 컴포넌트를 말한다. 문서가 변환될 때 이용되는 여러 컴포넌트들로 태그 변환기, 문서 분할기, 이미지 변환기 등의 여러 시스템들을 말한다.

### 2.5 CATBox

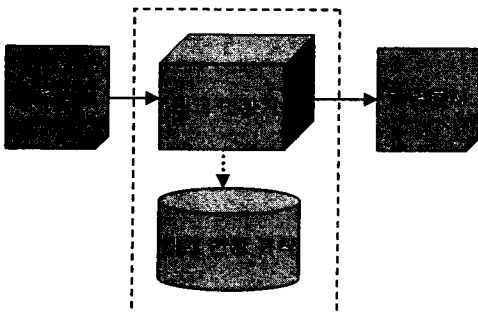
CATBox는 컨테이너인 동시에 전체 변환 작업의 흐름을 관리하는 관리자의 역할을 한다. 이것은 CAT의 설치, 생성, 적재, 실행, 관리, 제거하는 역할을 한다. 그리고 프로파일 정보를 비롯한 각종 컨텍스트 정보를 교환하는 통로 역할을 수행한다.

### 2.6 프로파일 관리 시스템

프로파일 관리 시스템은 변환 작업에 필요한 각종 프로파일 정보를 관리하는 CATS의 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

CATS에서 문서간 태그 변환을 담당하는 태그 변환 CAT 컴포넌트에 대해 설명하겠다. 태그변환기는 CATS에서 마크업 언어 간 태그 변환 기능을 수행하는 부분으로 [그림 2]와 같이 입력 문서를 XHTMLBasic, WML, mHTML등의 출력문서로 변환 해주는 역할을 한다.



[그림 2]태그 변환기의 구조

본 변환기는 변환 기능의 확장성 및 유연성을 확보하기 위해 태그간 변환 정보를 태그 변환기와 분리하여 관리하며 변환 규칙은 각 변환 태그에 따라 여러 규칙으로 미리 저장되어 있다. 태그 변환기에서는 입력 및 출력 태그 집합을 구성하는 XHTML 모듈 정보를 입력 받아 변환에 필요한 규칙을 선택하여 적용한다. 태그 변환 규칙은 기존의 시스템처럼 태그 단위의 변환 규칙이 아니라, 태그의 모듈성을 이용하여 모듈 기반의 변환 규칙을 정의했다. 그래서 무선 인터넷 언어에 대한 모듈 정보를 이용하여 적합한 모듈 단위 변환 규칙을 선택하여 변환이 가능하므로 새로운 언어에 대한 변환의 추가가 용이하다. 변환 규칙에서 새로운 변환 규칙을 추가하거나, 기존의 변환 규칙을 삭제할 수도 있다.

## 4. 태그 변환 규칙

태그 변환 규칙은 태그의 모듈성을 기반으로 하여, 태그 변환 과정에서 태그간의 사상정보, 태그간의 의존 관계, 원본 태그의 속성의 변환, 태그 속성값의 텍스트 콘텐츠 변환 등의 정보들을 참조하여 만들었다. 태그간의 사상 정보와 의존 관계, 속성을 고려하여 태그의 유지, 변환, 삭제의 3가지 규칙으로 크게 분류할 수 있다.

### 4.1 모듈화

출력 태그 집합은 XHTML 모듈로 구성된다. XHTML 모듈은 비슷한 종류의 태그들을 묶어 놓은 태그 집합의 기본 단위로 XHTML family문서들은 모듈의 조합에 의해 하나의 마크업 언어로 구성된다[2]. 기존의 태그 변환 시스템들은 태그 위주로 변환 규칙을 개발한 반면, 차후 개발되는 무선 마크업 언어들은 대부분 XHTML Basic기반의 모듈단위로 구성될 것이므로, 본 시스템에서는 이에 맞춰 변환규칙을 모듈단위로 정의하였다. 따라서 필요한 모듈을 선택적으로 관리할 수 있어서, 새로운 모듈의 추가, 삭제가 용이해 변환의 확장성을 가진다. 본 시스템에서는 HTML 4.01 태그를 기능별로 모듈화시켜 변환 규칙을 정의하였다[3].

### 4.2 모듈 간의 변환 규칙

HTML4.01의 모듈과 XHTMLBasic의 모듈을 이용한 모듈간의 변환 규칙은 [표 1]과 같다[4].

[표 1] 모듈간의 변환 규칙

유지될 모듈	변환, 삭제 모듈	일부선택모듈
structure module	applet module	Forms
text module	presentation module	modules
hypertext module	edit module	Table
list module	bi-directional text	modules
image module	module	
object module	frames module	
metainformation	iframe module	
module	scripting module	
link module	style sheet module	
base module	html specific module	

### 4.3 태그와 속성 변환 규칙

독립적인 모듈 단위로 만들어진 태그들의 변환규칙을 태그 단위에서 살펴본다. HTML4.01은 XHTML basic의 모든 태그들을 수용하고 있어서 변환 과정에서 대부분의 태그들과 속성들이 그대로 유지된다. 하지만 Forms module과 Tables module에서는 좀더 간단한 태그와 속성을 사용하기 때문에 이 모듈의 태그와 속성은 일부 선택적으로 변환하거나 삭제한다. 태그의 변환시에는 유사한 기능을 하는 태그로 변환하거나 기능은 다르지만 모바일 환경에서 콘텐츠를 표현하는데 내용상의 손실이 없다면 그에 적절한 태그로 대체하여 콘텐츠의 손실을 줄였다. 유지, 변환되지 않는 태그들은 삭제 과정을 거치게 되는데, 삭제 시 태그의 내부 콘텐츠 모델을 고려하여 태그만 삭제할 것인지 그 태그에 포함되는 콘텐츠 모델까지 모두 삭제할 것인지를 구분하여 삭제 규칙을 만들었다. 속성의 변환에서도 서로 매핑되지 않는 속성들은 삭제하고, 유사한 기능을 하는 속성이 있다면 변환하거나 태그나 텍스트로 대체하였으며, 매핑되는 속성은 유지시킨다.

### 5. 태그 변환기의 구현

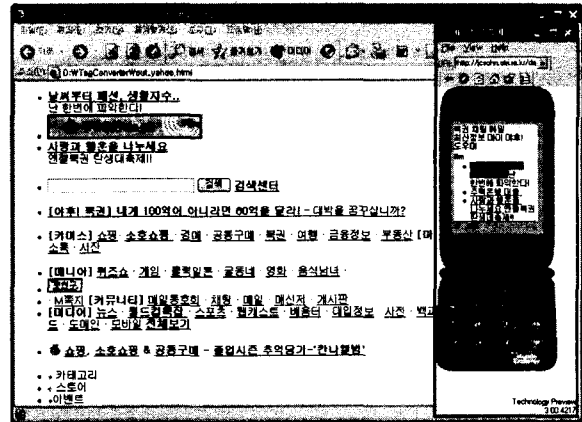
태그 변환기의 태그 변환 규칙은 위에서 설명한 규칙을 XSLT기술을 이용하여 변환 XSL로 구현하였다[5].

```
<?xml version="1.0" encoding="UTF-8" ?>
- <xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns="http://www.w3.org/1999/xhtml">
  <xsl:output method="xml" encoding="EUC-KR" doctype-public="-//W3C//DTD XHTML
    basic 1.0//EN" doctype-system="http://www.w3.org/TR/xhtml1-basic/xhtml1-
    basic1.0.dtd" indent="yes" />
  <!-- ===== input CHANGE ===== -->
  <!-- ===== structure module ===== -->
  <xsl:template match="html">
  <html>
  <xsl:copy-of select="*" />
  <xsl:apply-templates select="head|body|frameset" />
  </html>
  </xsl:template>
  <xsl:template match="body">
  <body>
  <xsl:copy-of
    select="@id|@class|@style|@title|@lang|@dir|@onclick|@ondblclick|@onmouseover" />
  <xsl:apply-templates
    select="p|h1|h2|h3|h4|h5|h6|ul|ol|div|center|noscript|noframes" />
  </body>
  </xsl:template>
```

[그림 3] 태그 변환 XSL 일부

태그 변환 XSL은 [그림 3]과 같이 구성되어 있으며 모듈별로 태그 유지, 변환, 삭제를 처리하고 중첩된 테이블의 경우 하나의 테이블만 남기고 내부 테이블들은 리스트로 나타낸다. PDA나 모바일 단말기의 경우에는 화면이 작아 중첩된 테이블의 경우에는 제대로 보이지 않기 때문이다. 본 XSL은 유지, 삭제, 변환을 모듈별로 구분하여 구현하였으며 각 변환규칙들은 XSLT를 사용하여 구현하여 XML의 장점을 가진다.

[그림 4]는 Yahoo 페이지를 본 논문에서 구현한 태그 변환기로 변환시킨 결과를 보여주고 있다. 원본 문서에서의 불필요한 텍스트들과 이미지는 삭제되었고, 테이블의 경우 리스트 형태로 변환된 것을 볼 수 있다. 결과로 나온 출력 문서는 XHTMLBasic문서이므로, 웹브라우저(IE6.0)와 ME 시뮬레이터(ME3.0)를 실행시켜 테스트하였다.



[그림 4] Yahoo 페이지의 태그 변환 결과

### 6. 결론 및 향후 과제

본 논문에서는 기존의 인터넷 콘텐츠를 유무선 환경에 상관없이 이기종의 클라이언트 장치에서 이용할 수 있도록 중간에서 콘텐츠를 변환하는 도구 및 콘텐츠를 저작할 때 도움을 주는 유무선 문서 변환 시스템의 개발 중에서, 입력 문서의 마크업 언어를 변환하는 태그 변환기를 설계 및 구현하였다. 본 시스템은 CATS에 입력된 문서를 미리 지정된 태그 변환 규칙에 의해 출력 문서로 변환해주는 역할을 한다. 본 변환기는 여러 가지 태그 변환 규칙을 가지는 XSL들로 구성되어 있으며, 적절한 태그 변환 규칙을 가진 XSL을 선택하여 출력문서를 생성해낸다. 본 변환기는 태그 변환기와 변환규칙이 분리되어 있어서 변환 기능의 확장성과 유연성을 가지며, 기존의 변환기처럼 태그 기반의 변환이 아닌 모듈 단위의 변환 규칙을 이용하므로 새로운 언어에 대한 변환규칙의 추가와 삭제가 용이한 장점을 가진다.

현재까지 구현된 시스템은 XHTMLBasic, WML1.1, mHTML에 대한 변환규칙이 제공되어 있으며 향후, WML2.0, XHTML-MP등의 다양한 문서로의 변환도 지원할 예정이다.

### 7. 참고문헌

- [1] J. H. Kim et al., "Internet Content transcoding framework for heterogeneous client devices", CALS/EC KOREA 2001 International Conference, pp.379-391, 2001.
- [2] "Modularization of XHTML", W3C Recommendation, 10, April 2001, <http://www.w3.org/TR/xhtml-modularization>
- [3] "HTML4.01 specification", W3C Recommendation, 24, December 1999, <http://www.w3.org/TR/html4>
- [4] "XHTML Basic", W3C Recommendation, 19 December 2000, <http://www.w3.org/TR/xhtml1-basic>
- [5] "XSL Transformations(XSLT) Version 1.0", W3C Recommendation, 16, November 1999, <http://www.w3.org/TR/XSLT>