

# XML 문서의 Presentation format 생성기의 설계

박경은\* 이기호  
이화여자대학교 컴퓨터학과  
(starry, khlee}@mm.ewha.ac.kr

## (The design of presentation format generator of XML)

Kyung-Eun Park\* Ki-Ho Lee  
Dept. of Computer Science & Engineering, Ewha Womans University

### 요 약

인터넷에 접속하기 위해 사용하는 단말기가 기존 유선 인터넷의 경우에는 desktop computer 환경이 대부분이었으나 무선 인터넷 접속 단말기는 스마트 폰이나 handheld 기기등 다양하며 이러한 단말기들이 모두 비슷한 환경을 갖지 않기 때문에 웹 콘텐츠를 모든 단말기에서 접근 가능하도록 만들기 어렵다. 그러므로 본 논문에서는 기존의 XML 문서를 기반으로 웹 콘텐츠를 구성하고, 콘텐츠를 표현하기 위한 presentation template를 작성하여 이 두 문서를 분석하여 웹 콘텐츠 페이지를 다양한 단말기에 적합한 포맷과 마크업으로 변환하는 방법을 제시한다.

### 1. 서론

유선 인터넷 뿐 아니라 무선 인터넷의 사용인구가 급속하게 증가하고 있다. 특히 무선 인터넷 사용인구는 2005년 이내에 유선 인터넷 인구를 앞지를 전망이다. (출처 : Anderson Consulting)

또한 기존 유선 인터넷을 접속하기 위한 기기가 desktop computer로 제한적이고 유선 인터넷 콘텐츠를 작성하는 언어도 HTML로 동일했던데 반해 무선 인터넷을 사용자들은 스마트폰, pager, PDA 등 다양한 handheld 기기로 서비스를 요청하며 무선 인터넷 콘텐츠도 WML, HDML, m-HTML, c-HTML 등 다양한 마크업 언어로 작성된다. 또한 무선 인터넷 브라우저도 어느 한 가지 마크업만 이해할 수 있는 등 무선 인터넷을 이용할 수 있는 환경은 매우 다양하다.

그러므로 무선 인터넷으로 서비스를 제공하고자 하는 사업자나 혹은 유선과 무선 인터넷 모두로 서비스를 제공하고자 하는 콘텐츠 제공자들은 다양한 마크업 언어를 사용하여 모든 page를 제작하거나 특정 단말기 사용자에게만 서비스를 제공할 수 밖에 없다. 게다가 각각 다른 단말기 환경을 생각할 때 같은 마크업을 사용하는 단말기라고 할지라도 스크린의 크기 등과 같은 단말기 환경에 적합한 콘텐츠를 제공하지 못할 가능성이 있다. 이렇게 모든 기기를 위해 서비스 콘텐츠를 최적화 시킨다는 것은 많은 노력이 필요할 뿐 아니라 유지 보수에도 큰 비용이 든다. 마찬가지로 특정 기기에만 서비스를 제공한다는 것도 서비스 제공자 입장에서 바람직한 것은 아니다.

한편에서는 데이터를 자유로이 기술할 수 있는 XML(eXtensible Markup Language)을 이용하여 만든 문서들이 증가하고 있으며 XSLT를 이용하면 XML 문서를 다양하게 표현할 수 있다. 그러므로 XML을 이용하면 각각의 마크

업으로 페이지를 만들기 위해서 데이터를 중복해서 작성해야 하는 단점 없이 다양한 단말기에 적합한 페이지를 생성해 낼 수 있다. 그러나 여전히 단말기 특성에 맞는 페이지를 만들기 위해서는 마크업이나 화면 크기 등에 따른 XSL 파일을 여러 개 만들어야 한다는 단점이 존재한다.

그러므로 본 논문에서는 XML 기반으로 웹 콘텐츠를 작성하되 콘텐츠를 어떻게 표현할 것인지를 presentation template 파일에 명세하고 이 XML 기반 웹 콘텐츠와 Presentation template를 분석하여 웹 콘텐츠를 다양한 단말기에 적합한 포맷과 마크업으로 변환하는 방법을 제시한다.

### 2. Presentation Format Generator Input 파일

각 단말기에 맞는 format을 생성하기 위해 우선적으로 콘텐츠 제공자가 제공하고자 하는 콘텐츠를 작성하고 이 콘텐츠가 어떻게 표현되기를 원하는가를 나타내야 한다. 그러므로 사용자가 콘텐츠를 작성할 수 있는 page content 생성 모듈과 콘텐츠 표현 형식을 나타내는 Presentation Template 생성 모듈이 필요하다.

#### ① Page Content 생성 모듈

여러 다른 XML source로부터 다음의 두 가지 방법으로 웹 콘텐츠 페이지를 작성한다.

1. Xpath를 이용해 XML source의 data의 path를 웹 콘텐츠의 element의 애트리뷰트에 명시
  - 원 source data가 바뀌었을 때에도 유지보수가 쉽고 data의 크기가 클 경우 효율적이다
2. 웹 콘텐츠 data를 새로 작성한다.

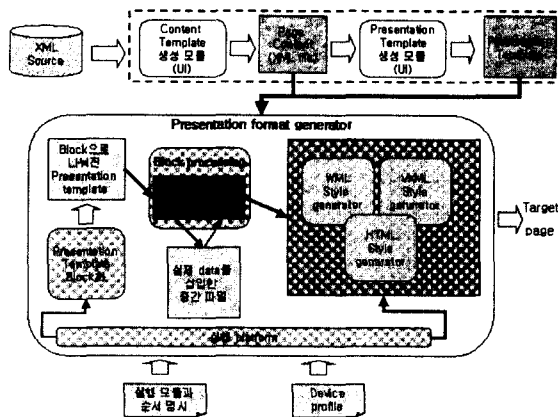


그림 1 Presentation Format generator의 구조

```

<ORDER>
<SOLD-TO>
<PERSON><NAME>홍길동</NAME></PERSON>
</SOLD-TO>
<SOLD-ON>2001년 3월 17일</SOLD-ON>
// 새로 작성된 콘텐츠의 data
<ITEM>
<PRICE>10000</PRICE>
<BOOK>
<TITLE>동화이야기</TITLE>
<AUTHOR>Joseph</AUTHOR>
</BOOK>
</ITEM>
<ITEM>
<PRICE>15000</PRICE>
// 기존 XML source로부터 가져온 data
<BOOK source_xml="computerbook.xml#book[@ID=452]">
<TITLE element=bookname/>
<AUTHOR element=author/>
<ISBN element=isdn/>
</BOOK>
</ITEM>
</ORDER>
    
```

예제 1. Page Content XML 예제 (order.xml)

② Presentation Template 생성 모듈

Presentation Template을 구성하기 위해서 웹 콘텐츠에 나타날 수 있는 요소들을 다음과 같은 일반적인 요소들로 구분하였다. 페이지 작성에 나타날 수 있는 일반적 요소들로 분리하고 자세한 특성은 attribute로 묶어서 target 마크업이 확장되거나 수정되었을 때 바뀐 tag의 특성을 attribute로 나타내어 지원할 수 있다. 이는 target 마크업에 변화가 있을 시 Style generator에서 재작성 되어야 하는 부분을 줄일 수 있다.

Presentation Template 구성 element 정의

- Structure : <template>
- Presentation :
  - content : <text>, <image>, <object>, <form>
  - Blank : <s/>, <n/>, <t/>
  - Layout : <table>, <list>, <frame>
- Link : <link>

Presentation Template 모듈에서는 content template 생성 모듈에서 만들어진 Page content XML file로부터 데이터를 가져와 해당 Presentation element에 매핑한다. 이때

Presentation Template에는 실제 데이터가 들어가는 것이 아니라 실제 데이터의 element path가 들어가게 된다.

Presentation Template 파일에 Page content XML로부터 data를 가지고 오지않고 직접 써야 하는 경우가 있다. 이것은 다음의 두 가지 경우로 제한하며 그 외의 경우는 모두 Page Content XML 파일에 기술되어야 한다.

1. XML data의 제목 표시

XML은 tag에 의미를 부여함으로써 어떤 데이터인지 나타낸다. 그러나 tag 이름은 약자나 합성어 등 웹 콘텐츠의 제목으로 사용하기 어려운 경우가 많다. 그래서 XML data의 제목은 직접 Presentation template에 기술하고 제목이 XML의 어떤 노드에 관한 것인지 매핑한다.

2. XML data의 수식이나 설명

XML data 간에 번호나 머리글 등을 매기거나 XML data를 나열하는 것이 아니라 간단하게 설명하고 싶을 때 데이터에 수식어나 연결어 등을 덧붙이게 된다. 즉, XML 구조에서 암묵적으로 나타나고 있는 설명이나 순서화를 직접적으로 표현할 수 있다. 예제2와 같은 template을 작성하게 되면 '구매자 : 홍길동 날짜 : 2001년 3월 17일' 과 같이 나타나게 되지만 '홍길동님이 2001년 3월 17일 구매한 물품' 이라고 설명식으로 작성할 수도 있다. 두번째의 경우 '님이', '구매한 물품' 과 같은 텍스트를 PD(presentation data)라고 한다.

```

<template source_xml="order.xml">
//직접적인 data를 가져오는 것은 아니지만 주문 관리라는 제목에 order 이하 하위 element들이 매핑됨을 나타냄
<text map=order align="center" size="15">주문 관리</text>
<n size=3/>
<text map=order/sold-to/person/name> 구매자 :</text>
<text element=order/sold-to/person/name /><n/>
<text map=order/sold-on> 날짜 :</text>
<text element=order/sold-on/><n size=2/>

<text map=order/item> 구매 물품 목록 </text>
<table>
<tr>
<td element=order/item/book/title>제목</td>
<td element=order/item/book/author>저자</td>
<td element=order/item/price>가격</td>
</tr>
</table>
</template>
    
```

예제2. Presentation Template 예제 (order-template.xml)

여기서 각각 생성된 Page Content 파일과 Presentation Template 파일은 Presentation format generator의 입력 파일이다.

3. Presentation Format Generator

3.1 모듈 실행 platform

서비스를 요청한 브라우저의 스크린 크기, 색, 지원 마크업 언어와 같은 정보를 HTTP 헤더로부터 알아낸다. 실행 명세 파일을 해석하여 해당 마크업 변환을 위해 명시된 모듈을 차례로 실행한다. 각 모듈에서 요구할 input 정보들을 수집, 관리하고 한 모듈의 output이 다른 모듈에서 input으로 사용될 때 이를 관리한다.

3.1 Presentation Template Block화

page content 파일의 XML 구조 정보와 Presentation element의 줄바꿈 정보(<n/>)에서 나타나는 단락 정보를 이용하여

관계 있는 정보 블록으로 나누고 블록들간의 연결 관계나 중첩 관계를 알아낸다.

<n/>를 중심으로 각 line이 content template 파일의 어떤 element와 매칭되는지를 알아내어 그 element의 EPID를 line의 id로 삼는다. 각 element는 하나의 block으로 취급하고 하나의 line은 line block이 된다.

• EPID 부여 방법

1. root element를 0으로 시작하여 자식 노드는 1로 구분한다.
2. content template에 나타나는 하나의 element에는 하나의 번호가 나타나는 순서의 의해 부여된다.
3. 같은 이름의 element가 반복해서 나타나는 경우 계속해서 번호를 부여하지 않고 (element번호^같은이름의 현재 노드번호)로 EPID를 부여한다.

예를 들면 0/3^2order.xml에서 <sold-on>element의 EPID는 0/2이고 <item> element는 2번 나타나는데 첫 번째 <item>의 0/3^1이고 두 번째 <item>은 0/3^2 이다.

**Presentation Template Block화 알고리즘**

1. Presentation Template의 element를 하나 읽어들인다.
2. element가 <n>이면 다음 <n>이 나타날때까지의 element들의 EPID를 조사하여 가장 최상위 노드의 EPID를 그 line의 id로 한다.
  - A. Presentation element가 layout element이면 내부의 하위 element의 EPID를 계산
  - B. Presentation element가 link element이면 link element의 하위 element들의 EPID 계산
  - C. Presentation element가 content element이면
    - i. Map이나 element attribute를 갖고 있는 element는 EPID를 계산
    - ii. PD는 EPID가 없다.

**Layout Element 블록화 알고리즘**

1. List 이면
  - A. 상위 level이 제목, 하위 level이 내용의 중첩 구조로 블록화 된다.
2. Table 이면
  - A. 제목을 매핑한 cell들과 그 제목에 관련된 내용 cell을 찾아서 제목페이지와 내용 페이지를 작성하고 둘 사이의 링크를 만들어 준다.
3. frame 이면
  - A. 각각의 frame 들을 하나의 블록으로 하고 각각에 대해 Presentation Template Block화 알고리즘을 적용한다.

**3.2 Block Processing**

Style generator의 요청에 따라 한 block씩 Style generator에 넘기고 block 분리 요청이 들어오면 EPID의 관계를 살펴 그 블록을 제목 부분과 내용부분으로 분리하여 Style generator에 넘긴다.

- 중간 파일 만들기
- presentation template element의 path에 명시된 element로부터 실제 데이터를 추출하여 매핑시킨다.
- Block 처리
- 모든 lineBlock들을 읽어서 content template tree에서 같은 가지에 속하는 lineBlock끼리 하나의 block으로 묶어 나간다.
- 이렇게 만들어진 block들 중 처음 block부터 Style generator에게 보낸다. Style generator가 분할 요청을 하

면 block에서 제목이 될만한 element를 분리해낸다. Block의 나머지 element들은 제목에 의해서 링크될 내용이 된다.

**3.3 Style Generator**

Style generator는 block을 실제 페이지로 작성하는 역할을 하며 각 마크업 언어의 navigation 특성에 의해 각각 작성된다. Style generator는 target 마크업의 tag들이 나타내는 navigation logic을 분석하여 페이지를 작성할 수 있어야 한다. 현재 WML, VXML, HTML Style generator를 설계하였다

**1.HTML Style generator**

Presentation template element와 target 마크업 언어간의 mapping table을 구성한다.

Presentation Template Ele.	Presentation Template Attribute	Target markup language
----------------------------	---------------------------------	------------------------

그림 2. mapping table

Presentation template element는 일반화된 element이므로 각 element의 속성값에 따라서 구체적인 target 마크업으로 변환한다. 이 경우 target 마크업의 tag가 수정, 추가 되어도 Presentation template element에 속성값을 추가함으로써 추가된 tag를 지원 할 수 있다.

**2. WML/VXML Style generator**

block processing 모듈로부터 받은 block이 단일 element로 구성된 단일 블록인지 내부 블록이 있는 중첩 블록인지를 확인하여 layout element가 아닌 단일 블록이라면 각 target 마크업의 표현 형식으로 바꾸어주고 중첩 블록이거나 단일 블록 중 layout element이면 제목과 내용으로 분리해 제목을 매뉴화하고 이에 다른 내용을 연결한다.

**4. 결론 및 향후 연구 과제**

본 논문에서는 XML 문서를 다양한 유무선의 단말기 presentation format에 맞도록 변환하기 위해 콘텐츠를 관계 있는 정보 블록으로 나누고 이 블록들을 토대로 마크업 언어의 navigation 특성에 맞는 페이지를 작성할 수 있도록 Style generator를 작성하였다.

일반화된 표현 element들을 구성함으로써 HTML로부터 변환시에 tag 각각의 의미를 분석해야 하는 단점을 줄이고 블록들이 page의 navigation 결정을 도와줌으로써 좀더 효율적으로 target 페이지를 작성할 수 있다. 결과로는 WML, HTML, VXML로 작성된 페이지들이 생성되며 Style generator를 추가함으로써 다른 마크업으로의 변환도 가능하다. 무선 서비스를 위해서는 개인화가 중요하므로 서비스 제공자가 제공하는 페이지 외에 사용자가 선택하도록 하거나 사용자의 성향을 파악하여 자동으로 사용자에게 적합한 페이지를 우선적으로 보여줄 수 있도록 Presentation format generator를 개선하는 것이 필요하다.

**참고문헌**

- [1]W3C, Modularization of XHTML, W3C propped Recommendation, 22 February 2001
- [2]WAP Forum, Wireless Application Protocol Wireless Markup Language Specification Version 1.2, 1999
- [3]WirelessCommunity, <http://www.wirelesscommunity.org>
- [4]VoiceXML forum, VoiceXML(Voice eXtensible Markup Language) ver. 1.0, 2000
- [5]Mosca, <http://www.mosca.co.kr>