

엘리먼트 정보 기반의 플러그인 XSLT 편집기

곽동규⁰ 조용운 유재우
송실대학교 컴퓨터학과

{coolman⁰, yycho}@ss.ssu.ac.kr, cwyo0@computing.ssu.ac.kr

XSLT PlugIn Editor Based on Element analysis

Donggyu Kawk⁰ Yongyoon Cho Chaewoo Yoo
Dept. of Computing, Soongsil University

요 약

XML 문서를 변환하는 목적은 하나의 어플리케이션에서 사용되는 XML 문서가 다른 어플리케이션에서 재사용될 경우, 사용자에게 동일한 정보를 제공하는데 있다. 문서를 변환하기 위해서는 변환 규칙을 정의해야 한다. 변환 규칙을 생성하기 위해서는 엘리먼트의 정보를 분석하여야 하는데 한 가지 방법으로는 엘리먼트의 정보를 분석하기 어렵다. 그러므로 본 논문은 변환 규칙을 생성하기 위한 엘리먼트 분석으로 다수의 방법을 사용할 수 있는 XSLT 편집기를 제안한다. 본 시스템은 XML의 구조적 특성인 엘리먼트의 포함 관계와 플러그인 방식으로 유사 엘리먼트를 분석하여 사용자에게 제공하는 방법을 보인다. XML의 엘리먼트 구조는 엘리먼트가 표현하고 있는 정보간의 구조와 동일한 구조로 포함되어 있다. 또한, 플러그인 방식 포함할 수 있는 모듈로는 문법적 특성을 이용한 방법을 포함하거나, semantic web의 정보를 이용하는 방법을 추가할 수 있다. 또한 기존의 XSLT 처리기를 내장하고 있어 XSLT 문서를 변환 XML에 적용하여 피 변환 XML 문서를 생성할 수 있다. 제안하는 시스템은 기존의 사용자가 사용 패턴을 분석하여 엘리먼트를 분석하고 변환 규칙을 기술하는 방법과 달리 유사한 엘리먼트에 대한 정보를 사용자에게 제공하는 장점이 있다.

1. 서 론

XML은 현재 어플리케이션에서 데이터를 저장하는 W3C 표준으로 많은 응용 분야에서 사용되고 있다. 기존에 작성된 XML 문서 정보를 다른 XML 어플리케이션에서 재사용하기 위해서 XML 문서의 변환이 필수적이다. 즉, 기존의 XML 문서를 재사용하기 위해서 어떤 어플리케이션을 위해 작성된 XML 문서를 구조와 정보의 손실 없이 다른 어플리케이션의 구조로 변환하여야 한다.

변환 규칙은 변환 XML과 피 변환 XML의 엘리먼트를 분석하여 엘리먼트간의 매칭으로 변환 규칙을 생성하여야 한다. 기존의 변환 방법은 사용자들의 사용패턴을 분석하여 수동으로 변환 규칙을 작성하는 방법[1]과 문법적 규칙이나 의미 정보를 이용하여 변환 규칙을 생성하는 방법[2][3]이 있다. 사용자의 패턴을 분석하는 방법은 수동으로 엘리먼트를 분석하고 변환 규칙을 작성하면 많은 비용을 필요로 하여 재사용에 대한 비용이 크다. 또한, 문법적 규칙을 이용한 방법은 정보에 따른 문법적 특성을 이용하여 문법적 특성이 같은 엘리먼트간의 매칭으로 변환 규칙을 정의한다. 하지만, 엘리먼트의 의미 정보는 문법적인 특성 외에도 엘리먼트의 이름에 정보를 포함하고 있다. 그러나 엘리먼트의 이름을 자동으로 분석하여 정보를 추출하기 어려워 완전히 자동화된 변환 규칙 생성기를 구현하기 어렵다. 그러므로 변환 규칙의 생성하기 위해서는 반자동화된 변환 규칙 생성 방법과 더불어 사용자의 입력이 필요하다.

기존의 변환 규칙 생성기중 가장 널리 이용되고 있는 제품은 XMLSPY[4]가 있다. XMLSPY는 XML 문서의 편집과 더불어 XSLT 편집 기능을 제공하고 있으나 변환 문서로 HTML만을 지원하고 있고 피 변환 문서의 엘리먼트에 대한 정보를 제공하지 않는다. 그러므로 피 변환 엘리먼트에 대한 의미 정보를 사용자가 수동으로 분석하여 적용하여야 하는 단점을 가지고 있다. 본 논문은 플러그인 방식으로 유사 엘리먼트 분석기를 적용할 수 있는 XSLT 편집기를 제안한다. 즉, 제안하는 시스템은 문법적 포함관계와 사용자의 입력을 변환 규칙에 적용하고 플러그인 방식으로 추가할 수 있는 XSLT 편집기이다. 본 시스템은 플러그인 방식으로 문법적 정보나 semantic web[5]를 통해 분석한 엘리먼트의 정보를 사용자에게 제공하여 변환 규칙 작성을 용이하게 한다. 또한, 기존의 XSLT 편집기가 제공하고 있는 직관적인 편집환경을 제공하고 기존의 XSLT 처리기를 내장하여 별도의 처리기를 필요로 하지 않는다.

본 논문은 2장에서 관련 연구를 소개하고 3장에서 본론에서 변환 규칙 생성 방법을 세 가지로 나누어 소개한 후 XSLT 문서 생성 방법을 보인다. 그리고 4장 결론 및 향후 연구 과제를 기술한다.

2. 관련 연구

2.1 패턴을 이용한 변환

특정 어플리케이션에서 사용하는 마크 업 언어를 다른 어플리케이션에서 사용할 수 있도록 하는 문서 변환의 필요성은 여러 차례 강조되었고 변환 방법 또한 연구되

있다. 그 중 W3C의 표준인 XML 문서와 인터넷의 대중화로 가장 널리 사용되고 있는 HTML 문서의 변환은 기존의 작성된 문서가 많으므로 재사용성이 높아 많은 연구가 이루어졌다. 기존의 변환 방법은 세 가지 방법으로 분류할 수 있다. 첫째는 피 변환 마크 업 문서에 나타나는 태그의 의미를 변환 마크 업 문서에 모두 기술하는 방법[1]이고, 두 번째 방법은 유사한 엘리먼트에 1 : 1로 대응시키는 방법, 그리고 세 번째 방법은 어플리케이션의 특성에 따라 논리적인 그룹으로 문서를 분할하여 다수의 문서로 변환하는 방법이다[2][6]. 앞에서 기술한 세 가지 방법은 모두 기존의 문서를 통해 엘리먼트에 따른 사용패턴을 연구하여 문서를 변환하는 변환 규칙을 찾거나 문서를 분할하는 방법이다. 이런 방법들은 모두 패턴을 연구할 수 있는 자료들을 필요로 하고 사용자에게 변환 규칙을 적용할 수 있는 정보를 제공하지 않는다.

2.2 문법적 정보를 이용한 변환[3]

“DTD의 의미구조 분석을 이용한 XML 문서의 변환”은 XML 엘리먼트의 정보 속성에 따라 문법적 속성이 다른 점에 착안하여 엘리먼트의 문법적 특성에 따라 네 가지로 분류하여 의미 구조로 정의하고 유사한 의미 구조의 매칭으로 변환 규칙을 생성하는 방법을 제안하였다. 하지만, 엘리먼트의 의미 정보는 문법 구조이외에도 엘리먼트의 이름을 통해 나타난다. 그러나 엘리먼트의 이름을 자동으로 분석하기 어렵기 때문에 사용자의 입력이 필수적이다.

2.3 XSLT(XML Stylesheet Language Transformations) [7]와 XSLT 편집기[4]

XSLT는 XML의 한 응용으로 XML 문서의 스타일을 정의하는 언어이다. XSLT는 XML 문서인 XSLT 문서를 이용하여 XML 문서를 다른 형태의 문서로 변환할 변환 규칙을 기술한다. XSLT 문서는 어떤 형태의 엘리먼트로 이루어져 있는 데이터를 다른 형태의 문서로 변환할 변환 규칙을 기술하여 문서를 변환한다. 하나의 데이터로 서로 다른 서비스를 제공하기 위해서는 응용프로그램에 맞게 XML 문서를 변환하여야 한다. 그러므로 각각 필요에 따라 XSLT 문서를 만들고 하나의 XML 문서를 다른 형태의 문서로 변환하여 서비스를 제공한다. 또한, XSLT 문서로 구조가 단순하고 XSLT 처리가 많이 개발되어 있어 XSLT 문성만으로 변환이 가능하여 많이 사용되고 있다. 하지만 변환 규칙은 개발자가 직접 정의하여야 하고, 일반적인 XML의 변환 규칙을 정의하기 위한 어떠한 정보도 제공하고 있지 않아 문서 변환 규칙을 정의하기 어렵다. 기존의 XSLT 편집기중 상용화 제품으로 ALTOVA사의 XMLSPY가 있다. XMLSPY는 변환 엘리먼트의 DTD를 분석하여 직관적인 XSLT 문서 작성에 효과적이다. 하지만 피 변환 문서로 HTML만을 지원하고 변환 XML 문서에 대한 정보를 제공하고 있지 않아 변환 XML에 대한 사용자의 분석이 필요하다.

3. 본 론

XML 문서의 변환은 의미적으로 일치하거나 유사한 엘리먼트간의 매칭으로 정의하거나 엘리먼트를 삭제하는 규칙이다. 그러므로 문서 변환을 위해서는 변환 XML 과 피 변환 XML의 엘리먼트를 의미적으로 분석하여 변환 규칙을 생성하여야 한다. 본 논문은 변환 규칙 작성을 생성하는 정보로 세 가지 정보를 적용한다. 첫째는 XML의 문법적 특성인 포함관계이고, 두 번째는 플러그인 방식으로 엘리먼트를 분석하는 모듈이며, 세 번째는 사용자의 입력이다.

첫째, 엘리먼트의 포함구조를 이용한 변환규칙 생성 방법은 정보 속성의 구조에 따라 적용되는 엘리먼트간의 포함 관계를 이용하여 변환 규칙에 적용하는 방법이다. 엘리먼트는 정보가 추상화된 단위로써 정보의 속성이 가지고 있는 포함 관계를 엘리먼트간의 포함구조로 표현한다. 그러므로 엘리먼트 포함구조를 적용하는 방법은 추상화된 정보간의 포함관계를 변환 규칙에 적용하게 한다. 둘째, 플러그인 방식으로 엘리먼트 분석 모듈을 적용하는 방법은 문법적 특성이나 semantic web[5]에 기술된 엘리먼트의 특성을 엘리먼트 분석에 적용하는 방법이다. 의미구조를 이용한 변환[3]에서는 정보에 따른 문법적 특성을 의미 구조로 정의하고 의미 구조를 변환에 적용하였다. 또한, semantic web은 엘리먼트간의 의미 관계를 표현하는 기술로 현재 활발하게 연구가 이루어지고 있다. 그러므로 플러그인 방식으로 엘리먼트 분석 모듈을 내장하는 시스템은 사용자에게 문서 변환을 위한 엘리먼트 정보를 제공한다. 하지만 앞에서 소개한 방법만으로는 엘리먼트의 의미 정보를 완전히 분석하기 어렵다. 왜냐하면 엘리먼트의 의미 정보는 문법적 특성 그리고 다른 문서, 엘리먼트의 이름, XML 문서를 사용하는 응용프로그램의 기능, 도메인에 대한 정보 등을 통해 표현된다. 그런데 엘리먼트의 이름이나 XML 문서를 사용하는 응용프로그램의 기능, 도메인에 대한 정보 등은 자동으로 분석하기 어렵다. 이러한 엘리먼트의 정보는 사용자가 직접 분석하여 적용하여야만 엘리먼트 정보에 합당한 변환 규칙을 생성할 수 있다. 그러므로 본 시스템은 사용자에게 엘리먼트에 대한 정보를 제공하고 사용자의 입력을 통해 문서를 생성하는 방법을 제안한다.

본 논문은 앞에서 소개한 세 가지 엘리먼트 분석 방법을 변환 규칙 기술하기 위해서 XSLT 문서를 사용한다. XSLT는 XML 문서의 변환을 위한 XML의 한 응용으로 W3C에서 지정한 표준으로써 변환 규칙을 XSLT 문서로 기술하면 기존의 XSLT 처리기를 이용하여 문서를 변환할 수 있다. 표 1은 변환 규칙을 XSLT 문서로 기술하는 방법이다.

· 표 1 변환 규칙에 따른 XSLT 문서

변환 규칙	XSLT 문서
XML element A → XML element B	<xsl:template match="A"> <xsl:apply-templates /> </xsl:template>
XML element A → ε	<xsl:template match="A"> </xsl:template>

표 1은 앞에서 소개한 방법에 따라 생성한 변환 규칙을 XSLT 문서로 기술하는 규칙이다. 본 시스템의 변환 규칙 생성과정과 변환 규칙의 적용은 다음과 같은 순서로 진행된다. XSLT 편집기는 변환 XML 문서의 DTD와 피 변환 문서의 DTD를 입력으로 받아 포함구조에 맞는 엘리먼트를 찾는다. 하지만 일반적인 XML 문서에서는 동일한 포함구조를 갖는 엘리먼트가 다수 존재한다. XSLT 편집기는 DTD와 엘리먼트의 정보를 가지고 있는 다른 문서를 바탕으로 유사한 엘리먼트를 분석하여 사용자에게 제공한다. 사용자는 분석된 결과를 바탕으로 변환 규칙을 생성하고 생성된 변환 규칙은 XSLT 문서로 기술된다. 그림 1은 XSLT 문서를 생성하고 XSLT 처리기를 내장하고 있는 XSLT 생성기이다.

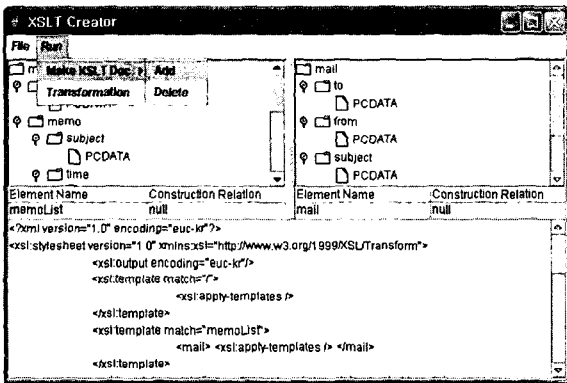


그림 1 XSLT 문서 생성기

그림 2는 그림 1과 같이 생성한 XSLT 문서를 정보 제 사용을 위해 변환을 적용하는 화면이다.

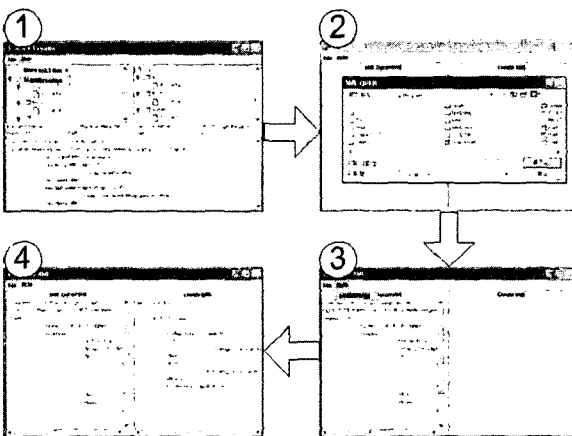


그림 2 XML 문서 변환기

그림 2에서 1은 XSLT 생성기를 통해 생성된 XSLT 문서가 적용된 XSLT 처리기를 실행 시킨다. XSLT 처리기의 엔진은 JAVA에서 제공하는 JAXP[8]를 사용한다. 2에서

피 변환 XML은 DTD에 유효한 XML 파일을 연다. 3은 변환 XML 문서를 변환 규칙에 의해 변환을 실행하고 4와 같이 피 변환 문서를 생성한다.

4. 결론 및 향후 연구 과제

XML은 데이터를 표현하는 마크 업 언어의 W3C의 표준으로 많은 어플리케이션에서 자료를 기술하는 방법으로 사용되고 있다. XML 문서의 변환은 하나의 어플리케이션을 위한 XML 문서를 다른 어플리케이션에서 사용할 수 있는 문서로 변환하여 한번 작성한 문서를 재사용할 수 있는 효과가 있다. XML 문서 변환을 위해서는 변환 XML과 피 변환 XML의 엘리먼트를 의미적으로 분석하여 엘리먼트간의 매칭으로 변환 규칙을 정의해야 한다.

본 논문은 XML 변환 규칙을 작성하기 위해서 엘리먼트의 정보를 분석하여 사용자에게 제공하는 XSLT 편집기를 제안한다. 변환 규칙을 생성하기 위해서는 XML 문서가 가지고 있는 엘리먼트의 포함구조와 문법적 분석, 다른 문서의 엘리먼트 정보 등을 적용하여야 한다. 제안하는 시스템은 엘리먼트의 포함구조를 분석하여 변환 규칙을 적용하고, 플러그인 방식으로 엘리먼트 분석 모듈을 내장하여 엘리먼트의 정보를 사용자에게 제공하는 XSLT 편집기이다. 플러그인 방식의 모듈은 문법적인 엘리먼트의 속성을 분석하거나 다른 문서에 기술되어 있는 엘리먼트의 정보[5]를 분석하는 모듈이다.

본 시스템은 엘리먼트의 포함구조와 플러그인 방식의 엘리먼트 분석 방법으로 변환 규칙 작성을 위한 정보를 사용자에게 제공한다. 하지만, 엘리먼트의 이름이나 응용 프로그램의 기능, 도메인에 관한 지식과 같은 정보를 분석하는 방법을 제시하지 못하였다. 앞에서 기술한 엘리먼트의 정보를 분석하는 방법은 향후 연구 과제로 남는다.

참고 문헌

- [1] Narayan Annamalai, Gopal Gupta, B. Prabhakaran, "An Extensible Transcoder for HTML to VoiceXML Conversion", ICCHP 339-346, 2004.
- [2] 광동규, 최종영, 유재우, "마크 업 문서의 대수적 분석에 의한 분할과 변환", 한국정보과학회 프로그래밍언어 논문지 제17권 제3호, pp.57~66, 2003.
- [3] 광동규, 최종영, 조용운, 유재우, "DTD의 의미구조 분석을 이용한 XML 문서의 변환", 한국정보과학회 '2004 춘계학술발표대회 논문집, pp.859-861, 2004.4.
- [4] XMLSPY, <http://www.altova.com>.
- [5] Semantic Web, <http://www.w3.org/2001/sw>.
- [6] 최훈일, 장영건, "HTMLtoVoiceXML 변환기의 설계 및 구현", 한국정보과학회 논문지C, 7권 6호, pp.559~569, 2001.
- [7] XSL Transformations (XSLT) Version 1.0 <http://www.w3.org/TR/1999/REC-xslt-19991116>.
- [8] Java API for XML Processing (JAXP), <http://java.sun.com/xml/jaxp>.