

XSL를 이용한 XML 문서 검색에 관한 연구

김충성(cskim@cs.chonbuk.ac.kr)^{○†}, 김용성(yskim@moak.chonbuk.ac.kr)[†]

[†] 전북대학교 컴퓨터과학과

Study for XML document retrieval to use XSL

Chung-Sung Kim^{○†}, Yong-Sung Kim[†]

[†] Dept. of Computer Science, Chonbuk National University

요약

최근 들어 이기종 간의 문서 교환을 위해 SGML(Standard General Markup Language) 문서보다 XML(eXtend Markup Language) 문서가 인터넷 기반에서 표준으로 자리잡고 있다. 앞으로 인터넷상의 수많은 정보들은 XML을 기반으로 할 것이고 이를 위해 문서 정보 검색 시스템이 필요하게 된다. 문서의 논리 구조를 표현하는 DTD(Document Type Definition) 기반으로 구조 검색을 할 수 있지만 본 논문에서는 XSL(XML Style Sheet Language) 문서에서 DTD의 Element를 지정하는 Pattern을 이용하여 문서 구조와 속성을 새로운 Tree로 표현하며 검색에 필요한 질의어 또한 XSL의 Pattern 자체를 이용하고 있다. 사용자에게 편하고 효율적인 검색 환경을 위해서 검색 인터페이스의 모형을 제안하였다.

1. 서론

대량의 정보 속에서 살아가는 현대인들은 정보를 효율적이며 체계적으로 관리하고 이를 의사결정의 수단으로 사용하고자 한다. 그리고, 인터넷의 발달과 정보 통신의 발달로 정보가 수많은 곳에 산재되어있다. 이들 정보의 효율적인 이용을 위해서는 상호작용을 위한 형식과 절차를 다루는 표준화 규약이 필요하며, 산재된 정보의 검색을 통해서 사용자가 원하는 정보를 추출해야한다.

이러하듯 다양한 정보를 저장하고 관리하기 위한 표준화 방법과 이기종 시스템간의 정보 교환과 다양한 검색환경을 제공하기 위해 국제표준(ISO 8879)인 SGML이 제정되었다. SGML은 마크업 언어로써 문서의 논리 구조와 내용을 기술할 수 있지만 문법이 복잡하며 인쇄를 위한 표준이다. 최근 들어 W3C에서 제안되고 있는 XML은 SGML의 복잡한 문법이 생략되며 인터넷을 기반으로 하고 있으며 인터넷 문서 표현 방법의 표준으로

자리잡을 것이다. XML문서를 쉽게 검색한다면 사용자는 정보를 손쉽게 얻을 수 있을 것이다.

2. 관련연구

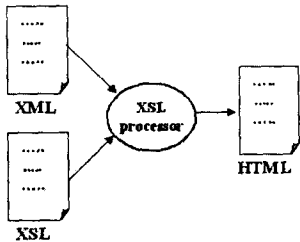
본 논문에서는 XML 문서를 검색하는 시스템을 설계할 것이며 XML 문서의 효과적인 검색을 위해서 XSL의 사용을 제시한다.

지금까지 대부분의 자료 검색 시스템의 경우 텍스트 위주의 검색 방법이었다. 하지만 SGML과 XML은 DTD로 문서의 논리구조를 표현할 수 있다. 즉 DTD를 통해 데이터베이스를 구성함으로써 구조 검색이 가능하게 되었다. TheoSEARCH, XSearcher, STEER 등이 이에 해당되는 시스템들이며 DTD에서 문서구조를 나타내는 엘리먼트로 구조검색을 지원하고 있다. 또한 각 검색 시스템들은 구조검색과 내용검색을 모두 지원하고 있다. 본 논문은 XSL의 Transformation 기능을 이용한 XML 문서의 구조검색과 내용검색을 하고자 한다.[2]

3. XSL

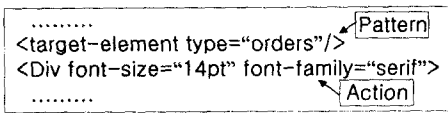
본 장에서는 XSL의 특징을 살펴봄으로써 효과적인 XML 검색 방법에 대해 알아보려고 한다.

XSL은 DTD와 DI(Document Instance)로 구성된 XML 문서를 브라우저에 표현하는 Style Sheet 언어이다. 다음 [그림 1]은 XML 문서를 XSL processor를 통해 HTML 문서로 변경하여 브라우징 하는 절차를 나타낸 것으로 XSL이 사용되고 있다.[1]



[그림 1] XML 브라우징

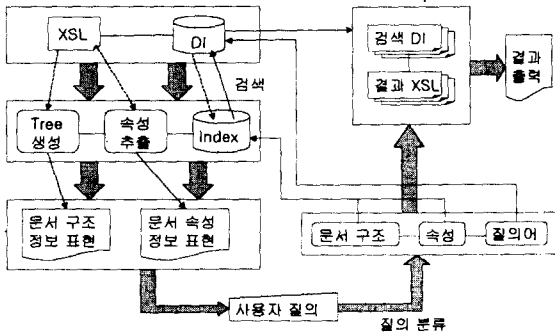
XSL은 문서의 구조를 포함하는 transformation기능 (pattern)과 스타일을 지정하는 formatting기능 (action)이 있다.[4] 즉, pattern은 DTD의 엘리먼트를 지정하고 action은 문서의 내용에 스타일을 지정하며 다음 [그림 2] 는 이를 나타내고 있다.



[그림 2] XSL의 기본 구조

4. XML 문서 검색 시스템

본 장에서는 XSL을 기반으로 하는 XML 문서 검색 시스템에 대해서 살펴보도록 한다. 다음 [그림 3]은 XML 문서 검색을 위한 시스템 구성도이다.



[그림 3] XML 문서 검색 시스템 구성도

4.1 DB 관리 시스템

본 논문에서 설계한 전체적인 시스템에서 XML의 문서를 저장하고 검색하기 위해서 DB를 사용한다. 기존의 DTD로 스키마를 구성하지 않고 XSL의 패턴으로 필드를 구성하여 검색을 효과적으로 하고자한다. 다음 [그림 4]는 XSL 문서의 한 예를 나타내고 있다.

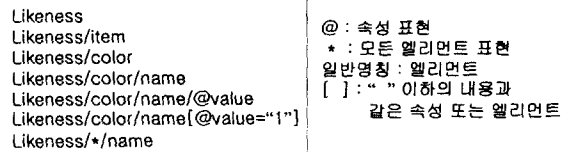
```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WO-xsl">
<xsl:template match="/">
<xsl:value-of select="/likeness/item"/>
<select>
<xsl:attribute name="NAME">
<xsl:value-of select="/likeness/item"/>
</xsl:attribute>
<xsl:for-each select="/likeness/color">
<option>
<xsl:attribute name="VALUE">
<xsl:value-of for="name/@value"/>
</xsl:attribute>
<xsl:value-of for="name"/>
</option>
.....
</xsl:stylesheet>
```

[그림 4] XSL 문서

[그림 4]의 문서에서 밑줄친 부분이 패턴이며 이는 DTD의 문서 구조 tree를 나타내고 있다. 즉, DB의 스키마를 구성할 수 있다.

4.2 질의어

4.1절에서 설계된 DB를 검색하기 위한 질의어로 마이크로소프트사에서 W3C(World Wide Web Consortium)에 제안한 XQL(XML Query Language) 제안서를 이용한다.[3] 이는 XML 문서의 구조와 속성 검색을 위해 XSL의 패턴 자체를 XQL로 사용하였다. [그림 4]를 통해 간단한 쿼리를 생성해 보면 [그림 5]와 같이 나타낼 수 있다.



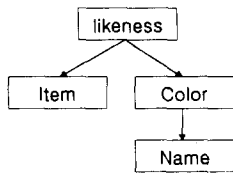
[그림 5] 간단한 쿼리와 특징

4.3 문서 구조 표현

본 절에서는 XSL의 패턴을 이용해서 문서 구조를 표현할 수 있는 트리를 생성하며 사용자에게 편한 인터페이스를 제공함으로써 질의를 쉽게하도록 한다.

4.3.1 Tree 생성

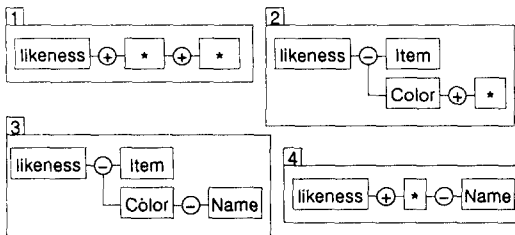
XSL의 패턴에서 표현하는 엘리먼트는 문서의 논리적 구조를 표현하는 DTD의 엘리먼트이며 계층관계를 표현하고 있다. 또한 특정 엘리먼트를 지정하고 이를 formatting하기 때문에 DTD의 순서, 반복을 고려하지 않아도 된다. 즉, DTD의 Tree보다 간단한 형태를 갖춘다. 다음 [그림 6]은 [그림 4]의 XSL 문서에서 패턴으로 만든 트리를 보여주고 있다.



[그림 6] XSL 패턴 트리

4.3.2 문서 구조 인터페이스

생성된 트리로 문서의 구조를 표현하려면 윈도우의 파일관리자처럼 계층관계를 표현한 Tree 모양이어야 한다. 또한 모든 엘리먼트를 나타내는 질의어 '*'를 위해 다음 [그림 7]과 같은 형태로 구조를 표현한다.



[그림 7] 패턴을 이용한 트리 구조

4.4 문서 속성 표현

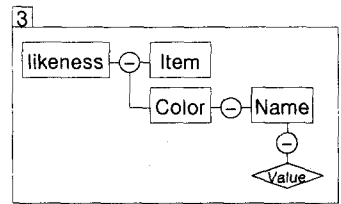
본 절에서는 DTD의 Attribute를 표현함으로써 다양한 검색 조건과 정확한 검색을 제공한다.

4.4.1 속성 추출

XSL의 패턴에서 Attribute를 표현할 때는 [그림 4]와 같이 '@'를 사용한다. 즉, '@' 다음의 명칭은 Attribute 이름에 해당하며, 각 Attribute의 값은 '=' 다음의 값으로 지정된다. XML 문서 검색을 위해 필요한 필드는 Attribute의 명칭이며 Attribute의 값은 검색어를 통해 입력받는다.

4.4.2 문서 속성 인터페이스

Attribute를 사용자가 쉽게 검색하고 정확한 구조 표현을 위해서 [그림 7]의 트리 구조에 이를 추가한다. 엘리먼트와 Attribute를 구분하기 위해서 모양을 달리하여 표현한다. 다음 [그림 8]은 [그림 7]의 3번째 그림에 Name 엘리먼트의 Attribute를 나타내고 있다.



[그림 8] 속성을 표현한 트리 구조

5. 결론 및 향후 연구과제

본 논문에서는 XSL의 Transformation 기능에서 DTD의 엘리먼트를 나타내는 Pattern을 기반으로 XML 문서 검색 시스템을 구성하였다. Pattern의 특성상 문서의 구조 검색이 가능하게되어 기존의 검색보다 정확하고 다양한 조건으로 검색할 수 있다.[5] DTD가 없는 Well-formed XML 문서일 경우 DTD를 새로 만들어 DB에 저장하는 것보다 XSL의 패턴을 이용해 DB를 구성하는 것이 더 적합하다. 본 논문의 검색 시스템에 일반적인 내용 검색 방법도 포함하여 다양한 검색 기능을 가지게 한다.

향후 연구로는 본 논문에서 다루고 있는 XSL의 패턴을 이용한 DB의 정확한 구성이 필요하며 문서 구조가 서로 다른 XML 문서일 경우 구조 검색할 수 있는 연구가 필요하다.

[참고 문헌]

[1] WWW 문서 작성을 위한 차세대 언어 XML 가이드, 1998. 8, 도서출판 그린, 정희경
http://myhome.shinbiro.com/~fancy2/xml/what_xsl.html

[2] TheoSEARCH, K4M
<http://www.k4m.com/product/theo-search03.html>

[3] XQL Proposal, Microsoft Corporation
<http://www.w3.org/TandS/QL/QL98/pp/xql.html>

[4] Chapter 14 of the XML Bible : XSL Transformations, Elliotte Rusty Harold
<http://metalab.unc.edu/xml/books/bible/updates/14.html>

[5] 문서구조화와 정보검색, 정보과학회지 특집 정보검색, 1998. 8
 맹성현, 주종철