

XML 시각화를 통한 자동 질의문 생성

송민영* 이기호
이화여자대학교 컴퓨터학과
(navyblue, khlee}@ewha.ac.kr

Auto generation of a query language through visualization of XML

Minyoung Song* Kiho Lee
Dept. of Computer Science & Engineering, Ewha Womans University

요 약

XML 문서를 효과적으로 관리하고 검색하기 위한 다양한 XML 질의 언어에 대한 연구가 활발히 수행되고 있지만 다양한 XML 질의 언어의 구문은 일반 사용자들이 숙지하는데 어려움이 있고, 또한 텍스트 형태의 XML 문서는 계층적인 구조로 되어있어 이러한 XML 문서의 구조를 알아야만 질의를 할 수 있다는 문제점이 있다. 따라서 본 논문에서는 복잡한 XML 질의어 구문과 질의 대상인 XML 문서의 구조를 동시에 시각화하고, 사용자와의 상호작용을 통해 질의문을 자동으로 생성하고자 한다.

이를 위해 본 논문에서는 다음과 같이 세 가지 컴포넌트를 기반으로 자동 질의어 생성을 위한 visual query interface를 제안하였다. 이 컴포넌트는 계층적 XML 문서 구조를 보여주는 XML structure viewer, 질의에 사용되는 연산자들을 아이콘화한 operator panel, 그리고 이 두 컴포넌트의 상호작용으로 작성된 질의 구문을 문법적으로 검사하기 위한 query statements panel로 구성된다. 이를 통해 다양한 계층적인 XML 문서의 구조와 복잡한 XML 질의 언어의 구문에 대해서 잘 모르는 초보자라 하더라도 쉽게 XML 문서 상에서 질의를 할 수 있다.

1. 서 론

웹 관련 신기술 데이터를 표현하는 기초 언어로 XML을 이용한 문서관리 시스템이 확산되고 있다. IT 분야 뿐만 아니라 인터넷 관련 어플리케이션에서의 XML 분야는 헤아릴 수가 없다. 이처럼 웹에서 관리하는 데이터의 종류가 현저히 증가하고 있다는 것과 다양한 신기술의 등장 때문에 XML 문서상에서 질의를 하고자하는 사용자들 또한 증가하고 있다. 이에 따라 XML 데이터를 대상으로 정보를 추출하고 변환하고 통합하기 위한 질의 언어의 필요성이 대두되었고 다양한 XML 질의 언어(XML-QL, XQL, Lorel, XML-GI, XQuery, 등)가 등장하게 되었다. 그러나 복잡한 질의 언어를 사용한 XML 문서의 검색 방법은 질의 언어의 구조를 이해하고 사용법을 숙지해야하는 어려움이 있어 일반 사용자에게는 적합하지 않다.

또한 텍스트 형태의 XML 문서는 계층적인 구조로 일반 사용자들이 인식하는데 어려움이 있다. 이와 같이 XML 문서 상에서 질의를 할 때, 사용자의 인지적 개념과 질의 언어의 조작 사이에는 큰 차이가 있다. 즉, 사용자는 XML 스키마에 대한 정확한 이해가 필요하며, 텍스트 기반의 질의 언어의 복잡한 문법을 숙지해야한다. 따라서 XML 질의 언어를 시각화함으로써 사용자의 인지적 개념을 그대로 모델링할 수 있도록 하고, 직관적이고 사용하기 쉬운 XML에 기반을 둔 통합 환경을 제공하여 다양한 사용자들이 기본적인 지식만으로도 쉽게 질의어를 작성 및 생성할 수 있는 시스템 개발이 필요하다.

본 연구는 다른 visual query language[1, 2, 3, 4]와는 달

리 XML을 기반으로 하고있으며, XML 문서의 계층화된 구조를 보여주기 위하여 XML 문서 분석을 통한 structure viewer를 제공함으로써 XML 문서에 대한 구조를 잘 모르는 사용자라 하더라도 쉽게 엘리먼트들간의 계층 구조를 파악하여 질의를 할 수 있도록 한다. 또한 XML 질의 언어에 필요한 여러 가지 연산자들을 아이콘화 한 operator panel과 각 expression들을 expression panel로 시각화시킴으로써 XML 질의 언어의 구문에 대해서 잘 모르는 초보자라도 쉽게 질의를 할 수 있도록 한다. 그리고, 이러한 각각의 패널들을 잘 조합해 사용하여 질의 언어로 변환하고, 완성된 질의문이 문법적으로 올바른지를 검사하기 위한 query statements panel을 제공한다. 따라서 사용자는 시각환경에서의 단순 조작만으로 원하는 질의문을 생성할 수 있다.

2. 관련연구

XML(Extensible Markup Language)이 차세대 웹 문서의 표준 마크업 언어로 정해진 이후, XML과 관련하여 XML 문서로부터 사용자가 원하는 정보를 얻기 위한 XML 질의 언어에 대한 연구가 활발히 수행되고 있다.

현재 XML과 관련된 질의 언어는 XML-QL, XQL, Lorel, XML-GI, XQuery 등이 표준화 진행중에 있으며 이들은 XML 문서의 구조적인 특성을 반영한 구조/내용 기반 검색들을 지원하고 있다.

또한, 기존의 다양한 그래픽 질의 언어들에 대한 연구가 있다. 대표적인 것으로는 VISUAL, VODOO, DOODLE, PESTO [1,2,3,4]등이 있으며 이들은 XML을 대상으로 질의

하는 것이 아니라, 관계형 혹은 객체 지향 데이터베이스 모델에 기반을 두고 있다. XML을 대상으로 제안된 그래픽 질의 언어에 대한 연구는 미흡하며 현재 제안된 언어로는 XML-GL[5]이 있다. 이것은 단지 XML-QL을 그래픽한 구조로 보여준 것으로 visual interface가 아니라 구문과 의미를 그래프 구조와 연산들로 정의한 그래픽 기반의 질의 언어이다[4]. XML-GL은 그래픽하게 표현할 수 있다는 점에서 텍스트 기반의 질의 언어에 비해 사용자들에게 친숙한 환경을 제공해주지만, 이것 역시 사용자가 XML-GL의 구문과 문법을 알아야하고 XML 문서의 구조를 알고 있어야만 질의를 할 수 있다는 단점이 있다.

따라서 본 시스템은 이러한 단점을 극복하기 위하여 질의 언어의 구문과 문법을 모르면서도 쉽게 질의할 수 있고 XML 문서의 구조를 자세히 모른다 하여도 XML 문서의 구조를 분석하여 시각화시킨 화면을 통해 질의 할 수 있도록 설계하였다. 또한 우리는 다양한 소스로부터 정보를 검색하고 변환하기 위한 여러 기능들을 내포하고 있는 범용적인 질의어인 XQuery[6]를 이용하여 본 시스템에서 제공하는 시각화 도구들의 기능을 처리하도록 하였다.

3. 시스템의 특성 및 구성도

본 논문에서 제시하는 시스템의 전체 구조는 그림 1과 같다. 본 연구는 다른 visual query language와는 달리 XML을 기반으로 XML 문서의 계층화된 구조를 보여주기 위한 XML 문서 분석을 통한 structure viewer를 제공함으로써 통합된 visual query interface 환경을 제공한다.

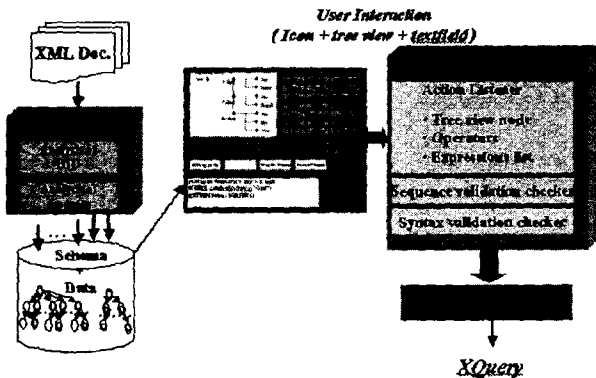


그림 1 전체 시스템 구성도

본 시스템은 다음과 같이 세 부분의 부 시스템으로 이루어져 있다.

① XML structure viewer

입력으로 로딩된 XML 문서가 유효(validation)한지 잘 구성된(well-formedness) 문서인지를 검사하는 과정과정을 거친 후, DOM 트리를 생성하고, Swing을 사용하여 화면에 디스플레이한다. 각 엘리먼트는 노드로써 사용자가 선택을 하게되면 호출한 부분에 해당 경로와 내용이 표시된다.

② Visualized operator panel

데이터베이스와 질의를 위한 visual한 표현 방법으로 Form 기반, Diagram 기반, Icon 기반, Hybrid 기반 방식이 있다. 본 시스템에서는 visual query language의 연산자들을 Icon 기반 방식으로 시각화하였다.

③ Query statements panel

사용자 인터페이스에 입력이 들어오면 각 프레임간 상호작용을 함으로써 입력된 구문을 XQuery문으로 변환하고 문법에 맞는지 검증한다.

3.1 XML structure viewer

XML의 문서를 일컫는 문서 내용부는 문서형 정의부(Document Type Definition, DTD)와 문서 실례부(Document Instance, DI)로 나뉜다. 문서형 정의부는 엘리먼트(Element), 엔티티(Entity), 속성(Attribute) 정보를 구성하여 정의되며, 실례부는 문서형 정의부의 정보로 문서의 내용을 작성한다. XML 문서의 계층적 구조를 표현하기 위해 문서의 구조를 논리적으로 표현한 DTD를 사용한다. DTD는 문서 내에 있는 요소들간의 구조 정보를 XML 문서를 검색할 때에 문서 전체가 아닌 부분 항목들로 처리함으로써 사용자 질의에서 원하는 특정 영역에 바로 접근할 수 있는 구조 기반 정보 검색을 가능하게 한다.

본 시스템은 입력 문서로 DTD 문서, XML 문서를 받아들여 검증 과정을 거치는데, 이때 검증 과정에서 사용되는 XML 구조 검증기는 W3C의 XML 구문 규칙 표준을 준수하고 보다 효율적으로 관리하기 위한 API를 제공하는 객체 모델인 DOM을 표준으로 추상적인 구조 모델을 생성하며 생성된 DOM 구조 모델은 XML 문서를 트리 구조로 표현한다.

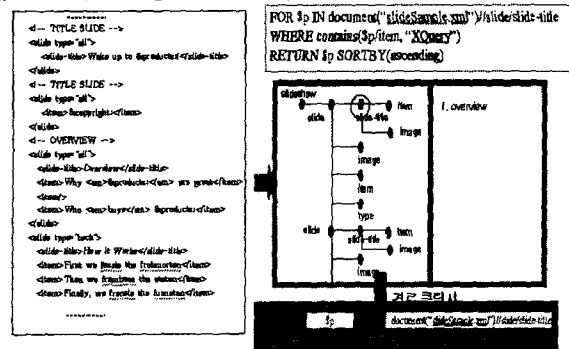


그림 2 XML structure viewer

그림 2는 slide 정보를 표현한 XML 문서를 structure viewer로 읽어 들인 화면이다. 왼쪽 그림과 같은 XML 인스턴스를 DTD 문서에 맞게 작성되었는지 검증한 후 XML 인스턴스의 계층적인 구조를 이루고 있는 엘리먼트들을 트리 구조로 화면에 디스플레이한다. 이 때, 왼쪽 패널의 노드를 클릭하면 해당 경로가 마우스의 포커스가 있는 곳으로 자동 입력된다. 위의 예는 slideSample.xml 문서의 slide-title 노드를 클릭한 경우로, document("slideSample.xml")/slide/slide-title과 같이 XPath의 경로가 텍스트 필드에 표시되며 콘텐츠 정보는 오른쪽 패널에 나타나게 된다.

3.2 Visualized operator panel

본 시스템에서는 XQuery에서 제공하고 있는 operator와 FLWR expression 구문을 표현할 수 있도록 expression panel과 대부분의 질의 언어에서 제공되는 산술, 비교, 논리, 배열, 조건 연산자와 집단(agggregation) 함수들을 제공한다.

산술 연산을 위해 +, -, *, div, mod 연산자를 포함하고 있고, 비교 연산자에는 값 비교, 노드 비교, 순서 비교 연산자로 구분되며 eq, ne, lt, le, gt, ge, ==, !=, <, >, BEFORE, AFTER가 있다. 논리 연산을 위해서 and, or 연산자를 제공하며, 배열 연산을 위한 to, union, |, intersect, except 연산자가 있다. 소트를 위한 sortby 연산과 조건을 명시하기 위한 if-then-else, satisfies 연산자를 제공한다. 이 밖에 집

단 함수로 avg, sum, count, max, min, contains를 제공한다.
 그림 3은 operator 패널의 설계화면이고 그림 4는 expression 패널의 설계화면이다. 각각은 사용자 변수와 경로를 입력받기 위한 텍스트 필드와 버튼들로 구성되어 있다.

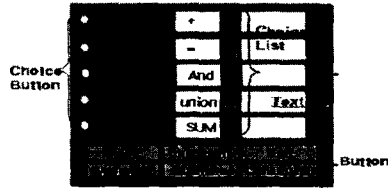


그림 3 operator panel

정렬을 하기 위한 sorting expression 패널에서는 descending, ascending 선택버튼을 사용하여 오름차순 혹은 내림차순 정렬을 가능하게 하고, 조건문을 만들기 위한 conditional expression 패널은 if-then-else 구조를 지원한다. 또한 XQuery에서는 SQL의 SELECT-FROM-WHERE 절과 비슷한 FOR-LET-WHERE-RETURN 절을 사용하는 데 이를 처리할 수 있도록 FLWR expression 패널에서는 네 부분의 패널을 포함하고 있다. quantified expression 패널은 predicate을 만족하는 노드를 검색하기 위한 것으로 조건을 만족하는 모든 노드를 검색하기 위한 EVERY 선택 버튼과 적어도 하나의 노드를 검색하기 위한 SOME 선택 버튼, 그리고 조건을 명시해주기 위한 in-satisfies 구문을 제공한다. typeswitch expression 패널은 표준 데이터타입뿐만 아니라 사용자 정의 데이터타입인 XML 스키마의 모든 데이터 타입을 지원하기 위해 typeswitch-as-default-return 구문을 제공한다.

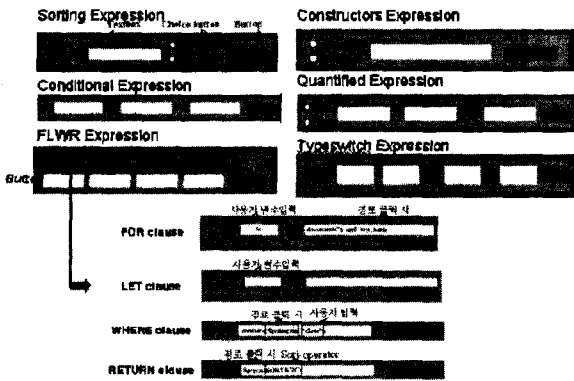


그림 4 expression panel

3.3 Query statements panel

Query statements panel은 각 연산들간의 상호 작용을 통해 입력된 값들을 질의어로 변환하고 전달된 구문이 문법에 맞는지 검증한다. 즉, structure viewer에서 선택한 노드에 의해 path expression이 표현되고 질의 종류는 operator 패널과 FLWR expression에 있는 버튼에 의해 선택되고, 이들을 연결하는 operator는 operator panel에서 선택하여 이들을 조합한 결과가 바로 이 Query statements panel에 텍스트로 표시된다.

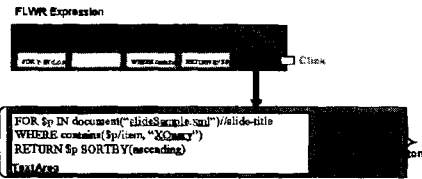


그림 5 Query statements panel

4. 프로토타입 구현 및 실험

전체 화면 구성은 메뉴바와 네 부분의 프레임들로 이루어져 있다. 그림 5는 slideSample.xml 문서를 대상으로 slide-title의 item 엘리먼트에 "XQuery" 텍스트를 포함하고 있는 slide-title을 오름차순으로 출력하라는 XQuery 질의문을 생성하는 예제이다.

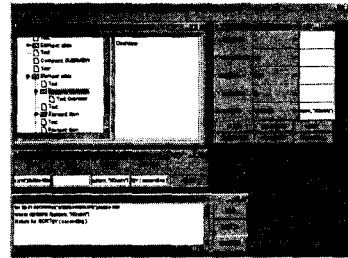


그림 6 실험 화면

5. 결론 및 향후 연구

인터넷 문서의 표준 사양인 XML 문서가 늘어나면서 XML 문서를 효과적으로 관리하고 검색하기 위한 시스템이 개발되고 있지만 지금까지 개발된 XML 문서의 검색 시스템은 질의어의 구조를 이해하지 못한 일반 사용자가 사용하기에는 어려움이 있다. 따라서 사용자가 쉽게 사용할 수 있으면서도 정확한 결과를 가지는 시스템이 요구되었고 이러한 요구사항에 의해 본 논문에서는 사용자들이 보다 쉽게 XML 문서상에서 질의를 할 수 있도록 하고자 자동 질의의 생성을 위한 visual query interface를 제안하였다.

향후 연구 과제로는 본 설계를 바탕으로 시스템을 구현하고, XML 프로세서와의 연동을 통한 결과를 시각화하는 부분에 대한 연구가 필요하다. 또한 이를 기반으로 다른 XML 질의 언어에도 응용이 가능하리라 생각한다.

참고문헌

- [1] N. Balkir, E. Sukan, G. Ozsoyoglu, and Z. Ozsoyoglu. VISUAL: A Graphical Icon-Based Query Language. In Proceedings of the Twelfth International Conference on Data Engineering, 1996.
- [2] L. Fegaras. VOODOO: A Visual Object-Oriented Database Language for ODMG OQL (extended version). <http://lambda.uta.edu/voodoo.ps.gz>, March 1999.
- [3] I. F. Cruz. DOODLE: A Visual Language for Object-Oriented Databases. In Proceedings of the 1992 ACM SIGMOD International Conference on Management of Data, San Diego, California, June 2-5, 1992.
- [4] M. J. Carey, L. M. Haas, V. Maganty, and J. H. Williams. PESTO: An Integrated Query/Browser for Object Databases. In VLDB'96, Proceedings of 22th International Conference on Very Large Data Bases, September 3-6, 1996.
- [5] S. Ceri, S. Comai, E. Damiani, P. Fraternali, S. Paraboschi and L. Tanca. XML-GL: a Graphical Language for Querying and Restructuring WWW Data. May 1999.
- [6] XQuery 1.0. W3C Working Draft, 20 April 2002. <http://www.w3.org/TR/xquery/>