김일성종합대학학보 주체107(2018)년
정보과학 제64권 제2호

조선어질문응답을 위한 XML문서검색에서 XQL표준의 적용에 대한 연구

최 명 옥

경애하는 최고령도자 **김정은**동지께서는 다음과 같이 말씀하시였다.

《과학기술강국은 사회주의강국건설에서 오늘 우리가 선차적으로 점령하여야 할 중요한 목표입니다.》(《조선로동당 제7차대회에서 한 중앙위원회사업총화보고》 단행본 38페지)

선행연구들에서는 XML문서검색체계[1, 2]에 대하여 연구되였지만 XQL표준을 리용한 검색체계실현에 대하여 론의된것이 없다.

XQL은 XSL의 패턴언어를 확장한 형태이며 마디를 찾기 위해 제공하는 XSL의 기능들에 추가적으로 불론리, 려파기, 마디의 집합에 대한 색인화 등의 기능들을 확장하였다.

조선어질문응답을 위해서는 자원으로 존재하는 조선어문장들을 XML로 표기하고 XML문서검색방법으로 질문에 대한 응답을 생성할수 있다.

론문에서는 조선어질문응답을 위한 XML문서검색에서 XQL표준을 적용하기 위한 구문분석 및 검색방법을 제안하였다.

1. 조선어질문응답에서 나서는 문제

질문응답체계에서 자연언어에 대한 연구는 지난 시기부터 많이 진행되여왔다. 그러나 그러한 연구들은 대부분이 문장의 문법적인 구조해석을 위한 측면에 많이 집중되여왔으며 의미구조해석과 그 표현을 위한 연구는 크게 진행되지 않았다.

문장에 대한 문법적해석 즉 구문분석에 대한 연구가 일정하게 론의되고 문장구조의 의미적인 고찰에 대한 요구가 제기되여 의미구조에 대한 연구가 진행되기 시작하였다. 특히 조선어와 일본어를 비롯한 교착언어들에서는 문장을 이루는 문장성분들의 의미적인 역할이 그러한 성분들에 부여된 교착물(토 및 조사)들에 의하여 일정하게 주어지는것으로 하여 이 언어들에 대한 해석에서는 격문법을 리용하여 문장의 의미구조를 해석하기 위한 연구가 진행되였다.

영어를 비롯한 일부 언어들에 대해서는 의존성리론에 대한 연구가 심화되여 중간적인 의미구조로써 의존구조를 해석하기 위한 연구가 진행되였다.[3]

문장의 의미론적해석에 대한 연구는 자연언어로 표현된 문장을 XML로 표기하기 시작하면서 더욱 심화되였다.

2. XQL에 관한 고찰

W3C에서 공식적인 XML질문어에 대한 표준이 현재 론의중에 있으며 XML질문어인 XQL을 XML문서검색에 리용하기 위한 연구가 진행되고있다.

XQL표현식은 쉽게 해석되고 리용하기 편리하며 다양한 쏘프트웨어환경에서 사용할수 있다. 따라서 XQL은 특별히 XML문서를 위해 설계되였으며 하나의 구문을 사용하여 질문, 주소화, 패턴작성 등을 수행할수 있는 다중질문어라고 할수 있다.

XQL표준에서는 출력에 대한 형식을 특별히 지정하지 않으며 질문에 대한 결과로 1개의 마디, 마디목록, XML문서, 배렬 또는 다른 구조가 될수 있다. 일부 구현에서는 질문에 대한 결과가 XML문서 또는 나무구조가 될수도 있다.

SQL과 XQL을 비교하여 표 1에 보여주었다.

표 1. SQL과 XQL의 비교

SQL	XQL
자료기지는 표들의 집합이다.	자료기지는 XML문서들의 집합이다.
SQL질문어의 기본모형으로서 표를 리용한 언어이다.	XQL질문어의 기본모형으로서 XML구조를 리용한 언어이다.
FROM절은 질문어에 의하여 검색되는 표를 결정한다.	질문어는 여러 문서들의 입력마디들에 적용되며 이 마디와 자식마디들을 검색한다.
질문어에 대한 결과는 여러 행들의 집합으로 구성되는 표이다.	질문어의 결과는 XML문서마디들의 집합이며 적합한 XML문서를 생성한다.

3. XQL구문분석방법

XQL질문어를 구성한 후 사용자가 이 질문을 입력하면 XQL질문어에 대한 구문분석이 진행되여야 한다.

XQL구문분석기는 XQL질문어를 입력받고 해석나무를 생성하여 매개 명령과 의미에 대한 처리를 진행하여야 한다. 이 처리에서는 XML자료화일의 색인, 구조정보 및 내용물정보를 리용한다.

XML질문문에 서술될수 있는 기호들은 다음과 같다.

//, /, 요소, *, @속성이름, 려파기, 첨수

질문문에 //가 오면 그것은 자손마디들의 집합을 의미하며 /가 오면 그것은 하위자식마디들의 집합을 의미한다. 질문문에서 요소가 직접 얻어지는 경우 현재 위치에서 해당 요소를 검색하며 *은 현재위치에서 존재하는 모든 요소들의 집합을 의미한다. 또한 /의 다음에 @속성이름이 오는 경우 @속성이름을 검색하라는 의미이며 요소의 다음에 려파기가 오는 경우 이 려파기에는 @속성이름 또는 요소가 올수 있다. 그리고 첨수는 검색된 요소의 순서와 범위를 지정할수 있다.

질문어가 입력되였을 때 질문어구문분석을 진행하는 과정을 론의하자.

질문문을 구성하는 가능한 첫 기호는 //, /, 요소, *이다. //의 다음에는 요소가 올수 있다. /의 다음에는 @속성이름 혹은 요소가 올수 있다.

요소의 다음에는 //, /, 려파기, *, 첨수가 올수 있다. *의 다음에는 /가 올수 있다. @속성이름의 다음에는 반드시 =가 오며 =의 다음에는 문자렬이 온다. 그리고 려파기의 다음에는 요소, 첨수, @속성이름이 올수 있다.

우와 같은 론의로부터의 XQL질문어구문분석방법을 그림 1에 보여주었다.

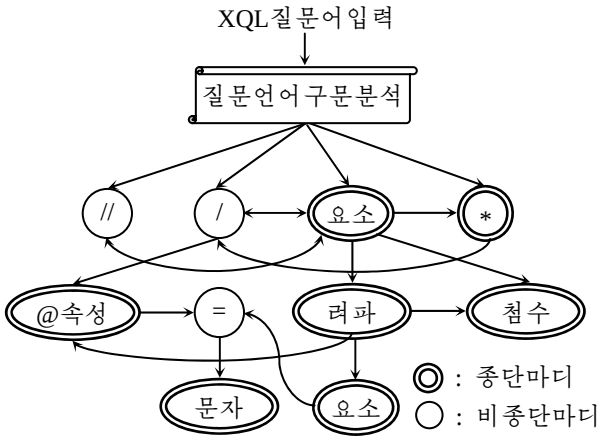

그림 1. XQL질문어구문분석방법

XQL질문어구문분석결과 최종검색하려고 하는 목적이 요소인가, 속성인가, 문자렬인가를 결정하며 입력한 XQL질문어가 문법에 맞는가를 검사할수 있다. 최종결과로 마디정보나무가 얻어진다. 이 마디정보나무와 색인문서구조나무를 리용하여 검색을 수행한다.

XML문서검색방법을 그림 2에 보여주었다.

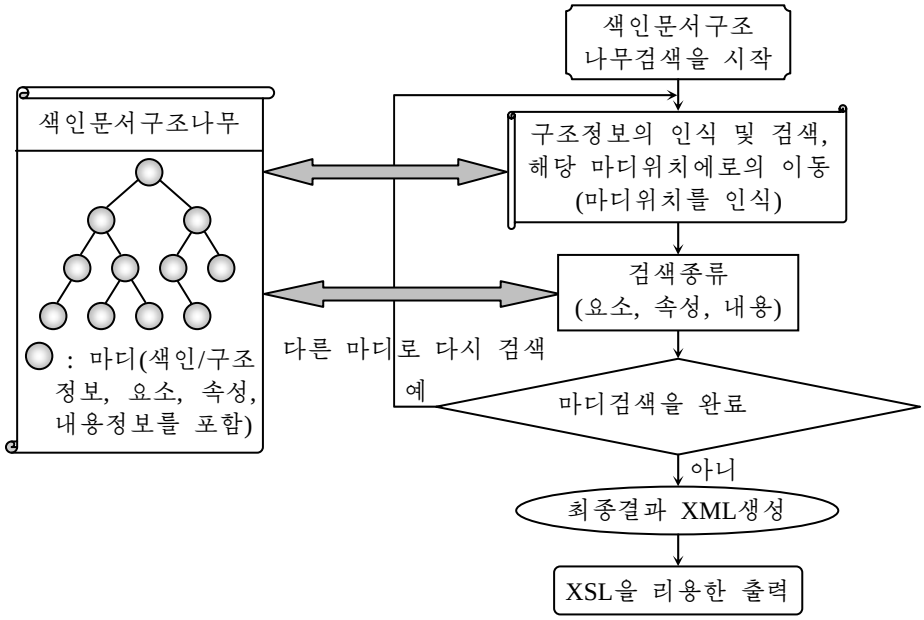

그림 2. XML문서검색방법

4. 제안한 방법의 효과성평가

XQL표준에 의한 문서검색방법을 리용하여 조선어질문응답체계를 구성하고 그 효과성에 대하여 학과목《객체지향언어》를 대상으로 평가하였다.

실험에서는 1 200개 문장의 질문을 입력하였을 때 체계의 응답정형을 3가지 지표로써

평가를 진행하였다.

체계의 효과성평가를 표 2에 보여주었다.

표 2. 체계의 효과성평가

No.	질문형(QT)	질문개수(QN)	대답없음(NF)	대답틀림(IC)	정확한 대답(CA)
1	사실형	300	32	19	249
2	개요형	300	35	21	244
3	정의형	300	41	50	209
4	기타	300	72	59	169
결과		1 200	180	149	871

이전의 자연언어처리체계들은 일반적으로 많은 량의 언어지식을 사전형태로 가지고있는것으로 하여 리용하기 어렵고 사용하는 단어가 사전에 제한되여있기때문에 입력문의 자유도가 높지 못한 결함을 가지고있었다.

론문에서는 조선어질문문의 구조적특징과 질문의 형태식별에 기초하여 XQL검색을 진행할수 있는 방법을 제안함으로써 조선어질문응답의 정확도를 높였다.

맺 는 말

XQL은 내용, 구조, 속성에 기초한 검색을 지원하는 XML문서의 질문어이다. 론문에서는 조선어질문응답을 위한 XML문서검색에서 XQL표준을 적용하기 위한 구문분석 및 검색방법을 제안하였다.

참 고 문 헌

[1] 리일남 등; XML응용기술, 공업출판사, 20, 주체97(2008).

[2] Valentin Tablan et al.; A Natural Language Query Interface to Structured Information, Department of Computer Science, 53, 2011.

[3] Majei Pavla; Automatic Question Answering System, Masarykova Univerzita, 30, 2014.

주체107(2018)년 2월 5일 원고접수

A Study for Application of XQL Standard in XML Document Retrieval for Korean Question Answering

Choe Myong Ok

XQL is a query language for XML document that supports retrieval based on content, structure and property.

This paper suggests phrase parsing and retrieval method for application of XQL standard in XML document retrieval for Korean question answering.

Key words: XML Document Retrieval, XQL, Question Answering