

온톨로지-DTD 정합에 의한 질의확장 개선 알고리즘

김명숙, 공용해
순천향대학교 정보기술공학부
e-mail:{krhkms, yhkong}@sch.ac.kr

A Query Expanding Algorithm Improved by Ontology-DTD Matching

Myoung Sook Kim, Yong Hae Kong
Div. of Info. Technology Eng., Soonchunhyang University

요 약

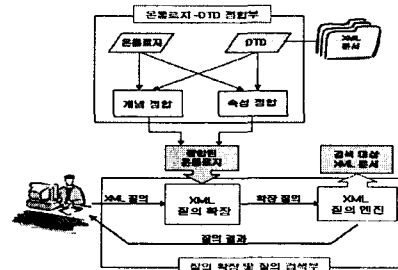
다양한 구조를 가지는 XML 문서에 대하여 온톨로지를 기반으로 확장한 질의를 인가하면 문서의 구조적 차이를 극복할 수 있으므로 검색의 효과를 높일 수 있다. 그러나 온톨로지만을 고려하여 확장되는 질의들은 대상 문서의 구조에 부적합할 수 있으므로 검색의 효율을 저하시키게 된다. 본 논문은 온톨로지와 대상 XML 문서의 구조인 DTD를 고려하여 온톨로지로부터 불필요한 개념과 속성을 제거하는 온톨로지-DTD 정합 알고리즘을 제안하였으며, 이렇게 정합된 온톨로지를 이용하면 대상 XML 문서에 적합하도록 질의들을 생성하므로 질의의 검색 적중률을 높일 수 있다.

1. 서론

XML 문서에 대한 질의검색은 XML 문서들의 구조적 다양성으로 인하여 검색의 효과에 한계가 있다 [1,2]. 온톨로지를 기반으로 확장한 XML 질의를 문서에 인가하면 어느정도 구조적 차이를 극복할 수 있는 검색이 가능해진다[3,4,5]. 온톨로지에 의해 확장되는 XML 질의는 보다 폭넓은 XML 문서의 검색을 가능하게 하지만, 대상 문서의 구조를 고려하지 않기 때문에 존재하지 않는 정보에 대한 불필요한 검색을 초래할 수 있다. 이러한 부적합한 질의들은 검색 적중률(Hit-Ratio)을 저하시키므로, 대상 XML 문서의 DTD와 온톨로지를 정합하여 대상 문서에 종속적인 질의를 생성하도록 함으로써 적중률을 높일 수 있다. 그러나 구조적 정보만을 갖는 DTD와 개념간의 속성상속 및 연관관계까지 내포한 온톨로지를 정합한다는 것은 매우 어려운 문제이다.

본 연구는 대상 XML 문서에 적합한 질의를 생성하기 위하여 온톨로지의 개념과 속성을 대상 문서 구조인 DTD의 엘리먼트와 속성에 일치하도록 정합하는 알고리즘을 제안하였다. 그림 1의 온톨로지-DTD 정합부는 온톨로지와 DTD 간의 개념과 속성을 정합하여 온톨로지를 수정하고, 질의확장 및

질의검색부는 정합된 온톨로지를 사용하여 XML 질의를 확장한 후 질의 엔진에 인가하여 대상 XML 문서의 정보를 검색한다.



(그림 1) 온톨로지-DTD 정합에 의한 XML 질의 확장

한번 정합된 온톨로지는 동일한 DTD를 사용하는 모든 XML 문서에 적용될 수 있으며, 정합된 온톨로지로부터 확장되는 질의는 대상 문서에 대한 검색 적중률을 높일 수 있으므로 XML 문서에 대한 정보 검색의 효율을 증대시킬 수 있다.

2. 온톨로지 기반 질의확장 문제 해결 방안

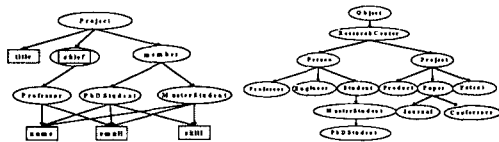
온톨로지를 기반으로 확장된 질의는 XML 문서로부터 폭넓은 정보를 검색할 수 있지만, 대상 문서의 구조와 속성을 고려하지 않기 때문에 문서에 존재하

는 개념이 중간에 위치할 경우, 표 1의 알고리즘에 따라 무조건 그 개념을 제거한다면, 중간개념으로부터 하위로 상속되어야 하는 모든 속성들이 동시에 제거되고 구조적인 개념 확장에 필요한 구조적 연결이 끊기게 된다. 이러한 문제를 해결하기 위해 표 2의 중간개념 정합 알고리즘을 제안하였다.

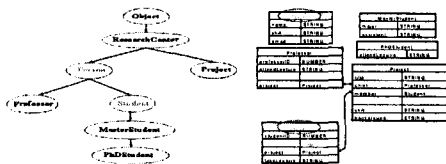
(표 2) 일치하지 않는 중간개념 정합 알고리즘

1. 온톨로지의 중간 개념과 DTD의 엘리먼트를 비교한다.
2. 만약 중간 개념과 일치하는 엘리먼트가 존재하지 않으면, 온톨로지로부터 중간 개념을 비활성 개념으로 설정한다.
3. 온톨로지의 모든 중간 개념에 대하여 과정 2를 반복한다.

그림 5의 (a)와 (b)를 비교해 보면, 온톨로지 중간개념인 Person과 Student는 대상 DTD와 일치하지 않으므로 제거해야 하지만 Person과 Student를 무조건 제거하면 하위개념으로 속성이 상속될 수 없으며 구조적 질의 확장도 수행될 수 없다.



(a) 예제 XML의 DTD 구조 (b) 온톨로지 구조
(그림 5) 중간 개념정합이 필요한 경우의 예



(a) 개념정합 결과 온톨로지 (b) 개념정합 후 속성 재구성
(그림 6) 중간 개념정합 결과 온톨로지 구조 및 속성

따라서 Person이나 Student와 같은 일치하지 않는 중간 개념은 속성 상속과 구조적인 확장에 필요하므로, 질의확장해 적용은 하위 검색에는 사용되지 않는 비활성 개념으로 설정하였다. 그림 6(a)는 표 2의 알고리즘에 의해 정합된 온톨로지이며 Person과 Student는 비활성 개념이 된다. 그림 6(b)는 온톨로지 정합에 따른 속성의 구조를 보인 것이다.

4. 질의 최적화를 위한 속성정합 알고리즘

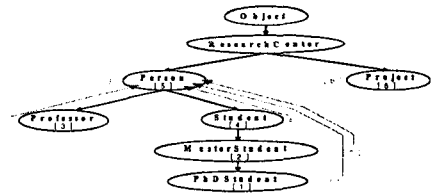
3장에서는 일치하지 않는 개념을 정합하여 불필요한 질의확장을 부분적으로 해결하였다. 다음으로 해결되어야 할 문제는 속성의 불일치 문제이다. 그림 4에서 개념 Student는 PhDStudent, MasterStudent와 상·하 구조를 이루고 있으나, 그림 3의 엘리먼트 PhDStudent, MasterStudent, Student는 서로 같은 레벨에 존재한다. 또한 그림 4의 개념 Person에

존재하는 속성들이 그림 3의 엘리먼트 Person에는 정의되지 않고 최하위에 정의되어 있다. 이 경우 온톨로지의 개념 Person의 속성들을 DTD와 단순정합하면 온톨로지 개념 Person의 모든 속성이 제거된다.

그러나 온톨로지에서 개념 Person의 속성이 모두 삭제될 경우, 모든 하위개념들은 개념 Person의 속성을 상속받을 수 없게 된다. 이러한 문제를 해결하기 위해 속성정합 알고리즘을 제안하였다.

4.1 일치하지 않는 속성정합

속성정합은 하나의 개념 안에 독립적으로 존재하여 상속과 무관한 속성의 정합과, 다른 개념에 종속적으로 존재하여 상속과 관련이 있는 속성의 정합으로 구분된다. 정합 방법은 우선 계층구조가 존재하는 온톨로지에서 상향식 레벨순으로 각 개념에 인덱스를 부여한다. 그림 7의 온톨로지 계층구조에서 표현된 [1], [2],... 등은 각 개념에 할당된 인덱스를 나타내며, 이는 개념에 대한 고유값으로써 탐색 순서 정보로 사용된다.



(그림 7) 온톨로지의 개념 인덱스 설정 및 탐색 경로

또한 온톨로지의 각 개념에 적합한 속성 정보를 제공하기 위해 각 속성에 대하여 초기값이 NULL인 매핑테이블을 구성하였다. 정합 알고리즘은 온톨로지의 각 개념에 정의된 인덱스를 참조하여 인덱스 1인 개념으로부터 시작하여 DTD에서 동일한 엘리먼트를 검색한다. 검색된 엘리먼트의 속성과 개념의 속성들을 비교하여 일치되면, 현재의 개념 인덱스를 해당 속성의 매핑 테이블에 저장한다. 한 개념에 대하여 이러한 과정은 그림 7에 나타난 화살표와 같이 상위개념으로 진행되고, 최상위개념에 도달하였을 때 한 개념의 속성정합이 완료된다. 이러한 과정이 모든 개념에 대하여 수행되면, 온톨로지의 각 속성의 매핑테이블에는 인덱스 정보가 저장된다.

속성의 매핑테이블 값이 NULL이면, 이 속성은 비활성 속성으로 설정되어 온톨로지서 불필요한 속성이 된다. 반면 속성의 매핑테이블에 인덱스값들이 존재하면, 각 인덱스에 해당하는 개념들은 이 속성을 포함하는 개념임을 나타내는 중요 정보가 된다.

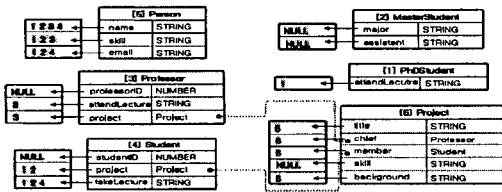
결과적으로 각 속성의 매핑 테이블에 저장된 정보를 이용하면 불필요한 질의확장이 최소화될 수 있다. 속성 정합 알고리즘은 표 5와 같다.

(표 5) 불필요한 속성 정합 알고리즘

1. 온톨로지의 각 개념들에 대하여 상향식 레벨순으로 인덱스를 부여하고, 개념의 각 속성들에 대하여 매핑 테이블을 구성하고 NULL 값으로 초기화한다.
2. 인덱스 초기값 1인 개념을 선택한다.
3. 현재 인덱스에 해당하는 개념에 대하여 일치하는 엘리먼트를 DTD에서 검색한다.
4. 검색된 엘리먼트 속성들과 개념 속성들의 일치여부를 비교한다.
5. 비교하여 속성이 일치하면 현재 개념의 인덱스를 온톨로지 속성의 매핑 테이블에 저장한다.
6. 과정 4, 5를 현재 인덱스에 해당하는 개념의 최상위개념에 이를 때까지 상위개념으로 이동하면서 반복한다.
7. 인덱스값이 최대가 되면 종료하고, 그렇지 않으면 인덱스를 1 증가시킨 후 과정 3부터 수행한다.

5. 온톨로지-DTD 정합 알고리즘 적용 실험

3장과 4장에서 제안한 온톨로지-DTD 정합 알고리즘을 적용하여 그림 3의 DTD를 대상으로 질의확장 효율을 실험하였다. 그림 3의 DTD와 그림 4의 개념정합된 온톨로지에 대하여 표 5의 속성정합 알고리즘을 적용하면 그림 8과 같이 정합된 속성의 매핑 테이블이 구성된다.



(그림 8) 예제1 속성 정합 매핑 테이블

(표 6) 온톨로지와 DTD 정합 전·후 질의확장 결과 비교

질의 확장 예 1	원본질의 //Person[skill="C"]
	온톨로지 정합 전 질의확장 //Person Professor Student MasterStudent PhDStudent[skill="C"]
질의 확장 예 2	원본질의 // Project[member]
	온톨로지 정합 후 질의확장 //Project[Product Paper Patent Journal Conference][member] //Student MasterStudent PhdStudent project 온톨로지 정합 후 질의확장 //Project[member] //PhDStudent MasterStudent project

정합된 개념 및 속성의 매핑 테이블을 적용하여 질의를 확장하고, DTD-온톨로지 정합 전과 후의 질의 확장 결과를 표 6에 비교하였다. 표 6의 질의확장 예 1에서 정합 전 온톨로지를 적용한 경우, Person, Professor 등의 부적합한 엘리먼트에 대해서 질의가 생성되었으며, 질의확장 예 2에서도 역시 Product, Paper, Patent, Journal, Conference, Student 등 대상 문서에 부적합한 질의가 생성되었다.

반면에 질의확장 예제 1에서 정합된 온톨로지를 t 적용한 경우, 개념 Person의 속성인 skill의 매핑 테이블에 저장된 인덱스 정보 1, 2, 3을 참조하여 대상

대상 문서에 적합한 개념으로 질의가 확장되었다. 또한 질의확장 예제 2의 경우도 질의확장시 매핑 테이블의 인덱스 정보를 참조함으로써 불필요한 질의가 모두 제거되었음을 확인할 수 있다.

6. 결론

본 논문은 다양한 구조를 가지는 XML 문서의 질의 검색시 온톨로지 기반의 질의확장에서 발생하는 부적합한 질의 생성을 최소화하기 위한 온톨로지-DTD 정합 알고리즘을 제안하였다. 온톨로지-DTD 정합 알고리즘은 세부적으로 온톨로지와 DTD의 개념-엘리먼트 정합, 온톨로지와 DTD의 속성 정합으로 구성된다. 개념-엘리먼트의 정합은 DTD와 온톨로지의 구조에서 최하위 노드로부터 개념과 엘리먼트를 비교하여, 일치하지 않는 최하위 개념을 제거하고 일치하지 않는 중간 개념은 비활성 개념으로 설정한다. 또한 속성 정합은 온톨로지의 모든 개념에 인덱스를 부여하고 DTD의 모든 엘리먼트의 속성들과 비교한 후, 개념의 각 속성들에 대하여 연관된 개념 인덱스를 저장하는 매핑 테이블을 구성하여 정합된 온톨로지에 포함시킨다.

정합 알고리즘에 의해 정합된 온톨로지는 검색 대상 XML 문서의 구조에 보다 적합한 질의를 생성함으로써 불필요한 정보검색을 제거하고 검색 적중률을 향상시킬 수 있을뿐만 아니라, 동일한 DTD 구조를 갖는 모든 XML 문서들에 대하여 정합된 온톨로지의 재사용이 가능하므로 정보검색의 효율을 높일 수 있다.

참고문헌

[1] M. Altinel, M.Franklin. "Efficient Filtering of XML Document for Selective Dissemination of Informatuon", In VLDB, 2000.
 [2] 김영란, "XML DTD의 효율적인 검색을 위한 구조 정보 및 인덱스 메커니즘", 한국컴퓨터정보학회 논문지, 8권, 3호, pp.80-86, 2003.
 [3] Natalya F. Noy, Mark A. Musen, "The PROMPT suite:interactive tools for ontology merging and mapping", Int. J. Human-Computer Studies 59, pp.983-1024, 2003.
 [4] M. Erdmann, R. Studer, "How to Structure and Access XML Document with Ontologies", Data & Knowledge Engineering, Vol.36, No.3, pp.317-335, 2001.
 [5] Myoung Sook Kim, Kwan Soon Choi and Yong Hae Kong, "Universial DTD Driven XML Access", Int. IC '04, pp.35-41, 2004.