

# 온톨로지-DTD 정합에 의한 XML 질의 확장

김 명 숙<sup>+</sup> · 공 용 해<sup>\*\*</sup>

## 요 약

XML 질의를 온톨로지 기반으로 확장하면 보다 폭넓은 정보검색이 가능해지는 반면에, 대상 문서의 구조에 부적합하게 확장된 질의들은 검색의 효율을 저하시킬 수 있다. 본 연구는 온톨로지와 대상 문서의 DTD를 정합한 결과인 축소된 온톨로지를 기반으로 질의를 확장함으로써 질의의 적합도를 향상시키는 방법을 제안한다. 온톨로지 개념과 DTD 엘리먼트 정합 및 온톨로지와 DTD 속성 정합에 의해 한번 축소된 온톨로지는 질의의 적중률을 높일 수 있을 뿐만 아니라 동일한 구조를 가지는 문서 집단에 재사용될 수 있으므로 검색의 효율을 향상시킬 수 있다.

키워드 : 온톨로지, XML, 질의 확장

## XML Query-Expansion by Ontology-DTD Match

Myung Sook Kim<sup>+</sup> · Yong Hae Kong<sup>\*\*</sup>

### ABSTRACT

If XML queries are expanded based on ontology, broader search may be possible. On the other hand, queries that are not adequate to target documents may also degrade the search efficiency. We try to improve query adequacy by expanding query based on a reduced ontology, which is the result of ontology and target DTD match. The match considers ontology concepts and DTD elements as well as ontology and DTD attributes. Since the reduced ontology can improve the hit ratio of queries and also be successively reusable for XML documents of a kind, the proposed method can improve XML search efficiency.

Key Words : Ontology, XML, Query-Expansion

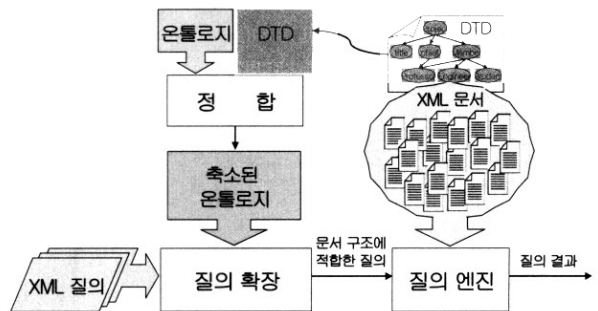
### 1. 서 론

XML 문서는 동일한 내용의 정보가 다양한 구조로 표현되므로 XQL, XPath, XQuery 등의 질의 언어에 의한 단순 구조적 접근으로는 의미적으로 중요한 정보를 검색하는데 한계가 있다[1]. 현재까지 연구되어진 링크질의, BRS/Search, XSearch, 포함질의 등의 검색 방법들도 단편적인 정보검색의 한계를 벗어나지 못하고 있다.

이러한 한계를 극복하기 위해 온톨로지의 개념 구조 및 상속 특성에 근거해 질의를 구조적으로 확장하는 연구가 수행된 바 있다[2, 3]. 이 방법들은 다수의 질의를 생성함으로써 단순 질의 검색의 한계를 어느 정도 보완할 수 있었으나, 대상 문서의 구조를 고려하지 않으므로 검색에 부적합한 질의를 과도하게 생성하는 부작용을 초래할 수 있다. 따라서 XML 문서의 구조적 다양성을 적극적으로 극복하고 검색에 유효한 질의만을 체계적으로 생성하기 위해서는 온톨로지와

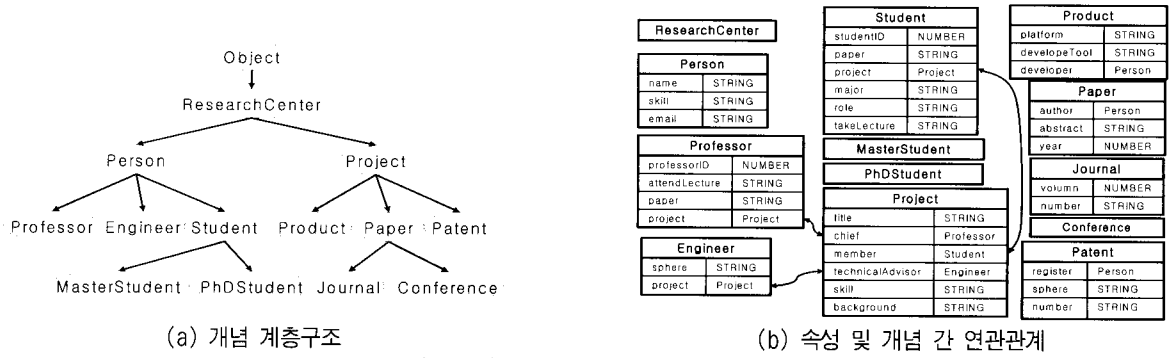
문서의 구조를 모두 고려하는 질의 확장 방법이 필요하다.

본 연구는 단순히 온톨로지의 사용만으로는 검색에 유효한 질의를 생성하는 것이 어려우므로, 대상 문서의 구조가 질의 확장 시 반영될 수 있도록 온톨로지와 DTD를 정합하는 알고리즘을 제안하였다. 세부적으로는 온톨로지의 개념 정합 알고리즘과 온톨로지의 속성 정합 알고리즘을 개발함으로써, 각 단계마다 불필요한 개념과 속성이 차례로 제거되어 (그림 1)과 같이 온톨로지가 축소되도록 하였다. 이렇게 축소된 최종 온톨로지는 대상 문서의 DTD 구조와 부합



(그림 1) 온톨로지-DTD 정합에 의한 XML 질의 검색 과정

\* 본 논문은 정보통신부 정보통신연구진흥원에서 지원하고 있는 정보통신기초기술연구지원사업의 연구결과입니다.  
<sup>+</sup> 준 회원 : 순천향대학교 전산학과  
<sup>\*\*</sup> 정 회원 : 순천향대학교 정보기술공학부 교수  
 논문접수 : 2005년 4월 14일, 심사완료 : 2005년 8월 10일



(그림 2) "대학연구센터" 온톨로지

되는 정보만을 가지므로 유효한 질의의 생성을 가능하게 한다. 또한, 검색이 처음 시도될 때 대상 사이트에 정의된 DTD와 온톨로지를 한번만 정합하면, 이후의 검색부터는 축소된 온톨로지를 재사용함으로써 질의를 신속하게 확장할 수 있다.

본 논문은 2장에서 관련 연구에 대하여 살펴보고, 3장과 4장에서는 온톨로지-DTD 정합 알고리즘과 정합된 온톨로지를 이용한 질의 확장 알고리즘을 설명한다. 5장에서 예제 XML 문서를 사용하여 온톨로지-DTD 정합에 의한 질의 확장의 효과를 실험하였다.

2. 관련 연구

본 장에서는 XML 정보검색을 위해 연구된 기존의 질의 검색 방법을 살펴보고, 이를 바탕으로 본 연구의 접근 방법을 기술하고자 한다.

현재 XML 질의 검색을 위해 제시된 XQL, XML-QL, XPath, XQuery 등의 질의 언어는 XML 문서의 구조와 내용을 기반으로 한 검색을 가능하게 하며, 이러한 질의 언어를 근간으로 링크질의, BRS/Search, 포함질의, XSearch 등의 다양한 XML 정보검색 방법이 연구되었다[4-12]. 이러한 방법들은 정보 추출의 성능적인 향상을 위해 필요한 연구이나, 여전히 인가된 질의에 대한 XML 문서의 구조나 문맥에 국한된 단순 질의 검색 방법에 불과하다.

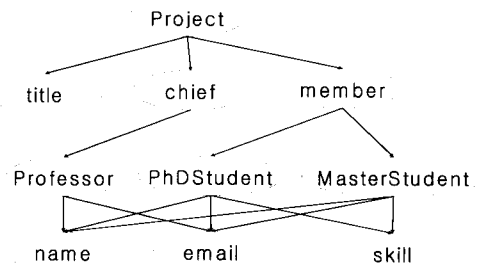
이러한 질의 검색의 한계를 극복하기 위해 Erdmann에 의해 수행된 온톨로지 기반의 구조적 질의 확장 방법은 온톨로지에 명시적으로 표현된 개념 계층구조를 이용하여 질의를 확장한다[2, 3]. 예를 들어, (그림 2)의 "대학연구센터" 온톨로지에서 "C"언어 skill을 보유한 사람을 검색하는 //Person[skill="C"]와 같은 질의는 온톨로지의 개념 Person의 하위 개념인 Professor, Engineer, Student, MasterStudent, PhDStudent로 확장된다. 그러나 이러한 질의 확장 방법은 대상 문서의 구조와는 상관없이 오직 온톨로지의 구조에만 의존하므로 부적합한 질의를 다수 생성할 수 있으므로 검색 효율을 저하시킬 수 있다.

본 논문은 대상 문서의 구조를 질의 확장 시 반영함으로써, 생성되는 질의가 대상문서에 보다 적합하게 하는 방법을 추구하였다.

3. 온톨로지-DTD 정합 알고리즘

온톨로지의 개념만을 고려하면 DTD와 같은 단순한 트리라 생각할 수 있으나, 개념에 포함된 속성들의 연관관계나 상속까지 고려하게 되면 온톨로지는 매우 복잡한 그래프가 되므로 정합 시에 세심한 주의가 필요하다. 본 연구는 불필요한 질의 생성을 최소화하기 위해 대상 XML 문서들의 구조를 정의한 DTD를 고려하여 온톨로지를 정합하는 알고리즘을 제안하였다. 온톨로지-DTD 정합 알고리즘은 세부적으로 두 단계로 구성된다.

온톨로지에서 개념의 속성은 하위 개념으로 상속되나, DTD의 속성은 상속되지 않으므로 일치하지 않는 개념을 제거할 때는 이러한 상속 특성이 반드시 고려되어야 한다. 첫 번째 단계에서 온톨로지의 개념을 DTD와 정합하여 개념적으로 축소된 온톨로지를 생성한다. 두 번째 단계인 속성 정합은 온톨로지의 개념에 대응하는 DTD의 엘리먼트에서 하위 요소들을 검색하여 공통이 되는 요소를 추출하는 것이다. 이 경우 상위 개념에 포함된 속성은 하위 개념에도 포함되므로 개념 간의 상속 특성이 고려되어야 하며, 온톨로지와 DTD의 공통 요소 추출에 소요되는 탐색의 횟수를 최소화하는 방법이 고려되어야 한다. 결과적으로 온톨로지는 재차 축소됨으로서 검색에 반드시 필요한 개념과 속성만을 포함하게 된다. 각 단계별 세부 알고리즘은 (그림 2)의 "대학연구센터" 온톨로지와 (그림 3)의 예제 DTD를 사용하여 설명하였다.



(그림 3) 예제 DTD 구조

3.1 개념 정합 알고리즘

(그림 2)의 "대학연구센터" 온톨로지를 사용하여 질의

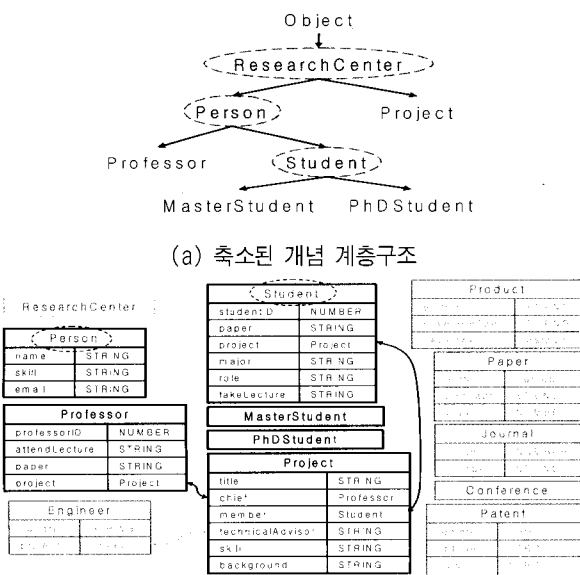
//Project[title="XML"]을 확장하면 //(Project|Product|Paper|Patent|Journal|Conference)[title="XML"]이 된다. 확장된 개념들을 (그림 3)의 DTD를 갖는 XML 문서에 적용하면, Project를 제외한 나머지 5개의 개념은 DTD에 존재하지 않으므로 불필요한 질의일 뿐이다. 이러한 불일치는 온톨로지가 대상 문서의 구조와 일치하는 최소한의 개념만을 가지도록 축소함으로써 가능해진다. 이를 위하여, 온톨로지와 대상 문서의 DTD를 비교하여 일치하지 않는 개념을 찾아 모두 제거하면 되지만, 제거되는 개념으로 인하여 하위 개념으로 상속되어야 할 속성까지 동시에 제거되는 문제가 발생한다.

개념 정합은 상속에 영향을 미치는 중간 개념과 상속과 무관한 최하위 개념을 구분하여 이루어진다. 만약 일치하지 않는 개념이 최하위 개념일 경우, 온톨로지로부터 해당 개념은 물론 그 속성까지 모두 제거한다. 반면에 일치하지 않는 개념이 중간 개념일 경우, 해당 개념은 질의 확장에 적용은 하되 검색에는 직접적으로 사용되지 않는 비활성 개념으로 만든다. <표 1>에 이러한 알고리즘을 나타내었다.

(그림 2)의 온톨로지와 (그림 3)의 DTD의 정합에 의해 일치하지 않는 최하위 개념인 "Engineer, Conference, Products, Paper, Patent"는 <표 1>의 알고리즘에 의해 (그림 4) (a)와 같이 모두 제거되었으며, 중간 개념 "Person"과 "Student"는 비활성 개념 변환되어 점선으로 표시되었다. 제

<표 1> 개념 정합 알고리즘

- For all the ontology concepts,
1. Select a concept.
  2. Compare the concept with elements of DTD.
- If the concept is not part of the DTD,
- If the concept is a leaf concept,
  - Remove the concept and its attributes from ontology.
  - Otherwise, the concept is made inactive.



(b) 축소된 속성

(그림 4) "대학연구센터" 온톨로지의 개념 정합 결과

거된 개념의 속성들도 (그림 4) (b)와 같이 제거되어 흐리게 표시되었으나, 비활성 개념의 속성들은 상속을 위해 그대로 유지되었음을 보였다.

3.2 속성 정합 알고리즘

속성 정합은 온톨로지의 개념에 대응하는 DTD의 엘리먼트에서 하위 요소들을 검색하여 공통이 되는 요소를 추출하는 것이다. 이러한 정합은 단순히 개념의 속성들과 엘리먼트의 하위 요소를 비교하여 공통되는 속성을 추출함으로써 가능하다, 온톨로지의 개념 사이에 존재하는 상속이 고려되어야 한다. 또한 온톨로지와 DTD의 단순 비교는 중복 비교 및 수많은 탐색 횟수에 의한 오버헤드를 초래할 수 있으므로, 이러한 비효율을 최소화하는 방법으로 속성이 정합되어야 한다.

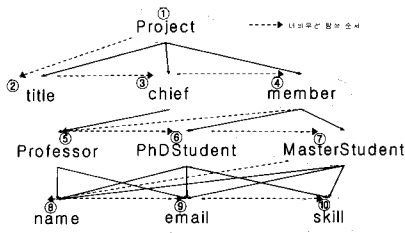
예를 들어, (그림 4)의 온톨로지에서 개념 Professor의 속성은 상위 개념으로부터 상속 받은 name, email, skill 과 자기 자신의 고유한 속성인 professorID, attendLecture, paper, project 등으로 이루어져 있으며, (그림 3)의 DTD에서 엘리먼트 Professor의 속성은 name과 email이다. 개념 Professor의 속성들을 엘리먼트 Professor의 속성들과 단순 비교하면, name과 email을 제외한 나머지 속성들이 서로 일치하지 않는다. 이 경우 일치하지 않는 속성을 모두 제거하게 되면, 다른 개념 PhDStudent, MasterStudent에 포함되어야 하는 속성 skill이 제거되므로 다른 개념에 필요한 속성까지 모두 제거되는 부작용이 발생할 수 있다. 즉, (그림 4)의 각 개념에 존재하는 속성들은 (그림 3)의 엘리먼트에 정의되어 있지 않거나 또한 정의되어 있다 하더라도 그 위치가 서로 다르기 때문에, 온톨로지의 속성 상속을 고려하지 않은 채 온톨로지의 개념에 대하여 일치하지 않는 속성이 모두 제거될 경우, 다른 하위 개념들은 제거된 속성을 상속받을 수 없게 된다.

<표 2>에 제안한 속성 정합 알고리즘은 탐색 오버헤드를 최소화하기 위하여, 3.1절의 알고리즘에 의해서 1단계 축소된 온톨로지에서 DTD의 요소가 정합된 개념에 포함되지 않으면 그 요소는 온톨로지 상에서 반드시 속성으로 정의된다는 사실을 적극적으로 이용하였다.

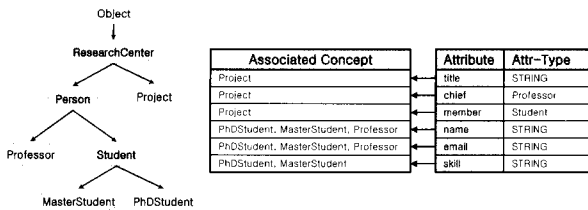
<표 2>의 알고리즘은 DTD의 노드를 차례로 방문하면서, 현재 노드가 이미 정합된 개념 정보에 포함되어 있는지의 여부를 확인하면, 현재 노드가 온톨로지상에서 개념인지 속성인지를 간단하게 식별할 수 있다. 만약 현재 노드가 개념이면 속성 정합을 수행하지 않고 다음 노드로 이동하고, 만약 속성이면 현재 노드의 모든 부모 노드를 속성과 연관된 개념 정보로 저장한다. 이는 DTD의 구조에서 속성으로 존

<표 2> 속성 정합 알고리즘

- For all the nodes in DTD, starting from the root node,
1. Select the next node in DTD by BFS.
  2. If the selected node is included in section 3.1 reduced ontology, go to Step 1.
  3. Otherwise(when the selected node is an attribute),
    - 3.1 Add all the parent nodes of the selected node to its attribute list.
    - 3.2 Go to Step 1.



(그림 5) DTD 탐색에 의한 속성 정합 과정



(a) 개념 구조 (b) 개념들의 속성  
(그림 6) "대학연구센터" 온톨로지의 속성 정합 결과

재하는 노드의 상위 노드는 온톨로지의 구조에서 반드시 개념으로 대응되기 때문에, DTD의 현재 노드가 속성임이 확인 되면 현재 노드의 모든 부모 노드가 연관된 개념 정보가 된다. (그림 5)는 <표 2>의 알고리즘에 의한 속성 정합 과정을 나타낸 것이며, DTD의 최상위 노드로부터 너비우선탐색(BFS: Breadth First Search)을 수행한다. (그림 5)의 점선과 번호는 DTD의 노드 탐색 순서를 나타낸 것이다. 제시한 속성 정합 방법은 온톨로지의 모든 속성과 DTD의 모든 속성을 전부 비교하지 않고, 단 한 번의 DTD 트리 탐색에 의해 (그림 6)과 같이 속성 정합을 효율적으로 완료할 수 있다.

#### 4. 정합된 온톨로지에 의한 질의 확장

XML 질의는 3장의 온톨로지-DTD 정합 알고리즘에 의해 축소된 온톨로지를 기반으로 확장된다. XML 문서의 구조적 다양성은 개념의 상속을 고려한 <표 3>의 개념 계층구조에 의한 질의 확장 알고리즘에 의해 극복할 수 있다. 예를 들어, (그림 6)에서 개념 Person을 개념 계층구조에 의해 확장하면 비활성 개념인 Person, Student를 제외한 나머지 하위 개념인 Professor, MasterStudent, PhDStudent로 질의가 확장된

<표 3> 개념 계층구조에 의한 질의 확장

- If a query contains no attribute at all,  
For all the concepts of a query,
  - Select a concept.
  - Search for all its descendant concepts.
  - If the selected concept and its descendant concepts is active,  
Add the descendant concepts to a query set.
- If a query contains any attribute,  
For all the concepts associated with the selected attribute in a reduced ontology,  
If a concept is a descendant of the concepts in the query,  
Add the concept to the query set.

다. 이러한 질의 확장은 단순 질의에 의한 정보검색에 비해 보다 폭넓은 정보의 검색을 가능하게 하며 동시에 과도한 질의의 생성을 억제한다.

질의에 속성이 포함되어 있을 경우의 질의 확장은 (그림 6) (b)의 결과를 적용하여 질의 확장을 수행한다. 예를 들어, 질의 //Person[skill="C"]는 (그림 6) (b)의 축소된 속성 테이블에서 개념 Person의 속성 skill과 연관된 개념을 검색하여 그 하위 개념으로 질의를 확장한다. 따라서 //Person [skill="C"]는 //(PhDStudent|MasterStudent)[skill="C"]로 확장된다.

온톨로지에는 개념 및 속성 간의 상호 연관성이 존재한다. (그림 4) (b)에 표현된 연결선분은 이러한 관계를 나타내며, 상호연관 규칙은 <표 4>의 연관관계 생성 알고리즘에 의해 추론될 수 있다. 추론된 규칙은 질의 확장에 사용되어 보다 의미 있는 정보의 검색을 가능하게 한다.

<표 4> 상호연관 규칙 생성 알고리즘

- 
- For all the ontology concepts,
- Select a concept A.
  - For all the attributes of concept A,
    - Select an attribute B.
    - Search concept C that matches attribute B.
    - Set association between concepts A and C
- 

(그림 4) (b)의 경우, 두 개념 Student와 Project 그리고 Professor와 Project 사이에 상호연관 규칙이 존재한다. 이러한 상호연관 관계는 다음과 같이 규칙으로 생성된다.

- Rule 1 : FORALL Pers1, Proj1  
 Proj1 : Project[member ->> Pers1]  
 <-> Pers1:Student[project ->> Proj1]
- Rule 2 : FORALL Prof1, Proj1  
 Proj1 : Project[chief ->> Prof1]  
 <-> Prof1 : Professor[project ->> Proj1]

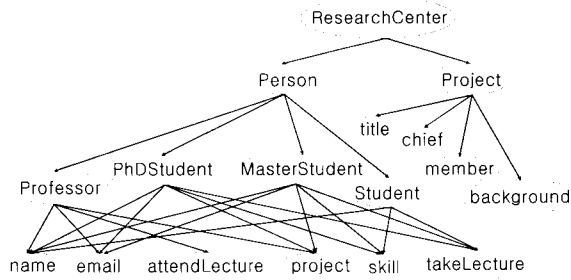
인가된 질의의 개념 중에 연관된 규칙이 존재하면 규칙에 포함된 개념을 질의에 추가시키고, <표 3>의 질의 확장 알고리즘을 사용하여 질의에 포함된 모든 개념에 대하여 질의를 확장한다. 이러한 방법은 인가된 질의와 의미적으로 연관된 정보를 검색하게 함으로써 구조적 질의 확장과 함께 보다 깊이 있는 정보검색을 가능하게 한다.

#### 5. 실험

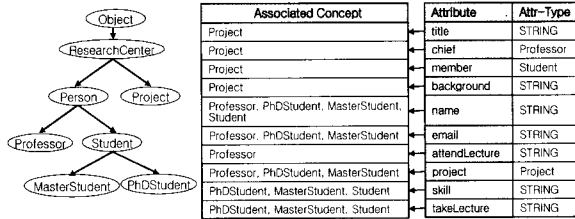
3장과 4장에서 제안한 알고리즘이 실제 XML 문서와 일치하지 않는 요소들을 온톨로지로부터 효과적으로 제거하고, 정합되어 축소된 온톨로지의 사용이 불필요한 질의를 최소화하는지를 다양한 예제 DTD를 대상으로 실험하였다.

##### 5.1 온톨로지-DTD 정합 실험

3장에서 제안한 온톨로지-DTD 정합 알고리즘을 다양한 예제 DTD에 적용하였다. (그림 7)(a), (그림 8)(a), (그림 9)(a), (그림 10)(a), (그림 11)(a) DTD들은 서로 다른 구조로 정의되어 있는 것처럼 보이지만 (그림 2)의 "대학연구센터

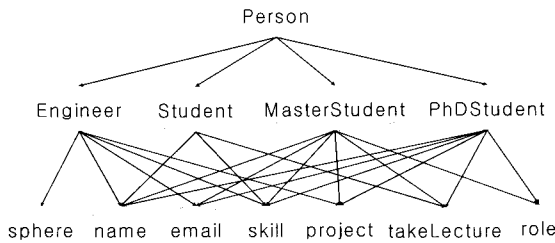


(a) 예제 DTD1

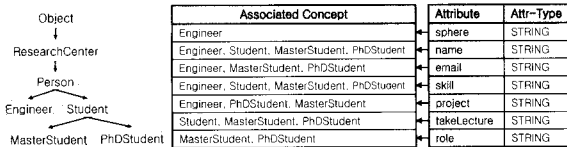


(b) 축소된 온톨로지

(그림 7) 온톨로지-DTD1 정합

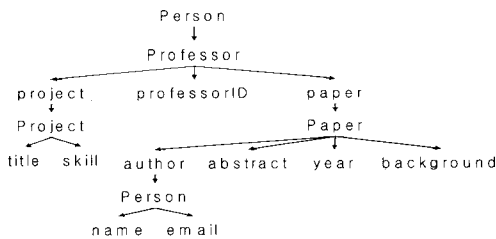


(a) 예제 DTD2

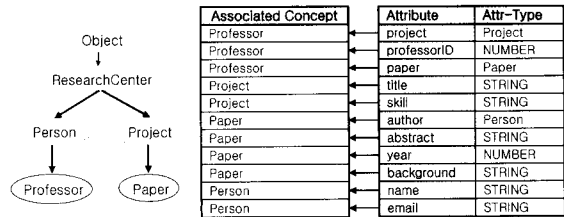


(b) 축소된 온톨로지

(그림 8) 온톨로지-DTD2 정합

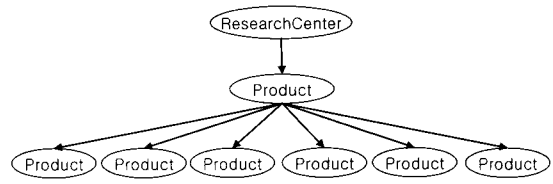


(a) 예제 DTD3

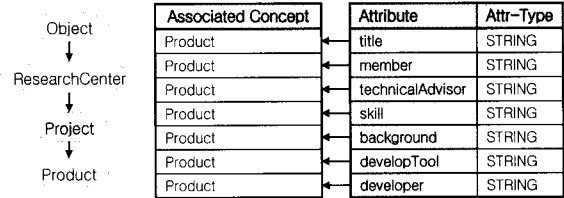


(b) 축소된 온톨로지

(그림 9) 온톨로지-DTD3 정합

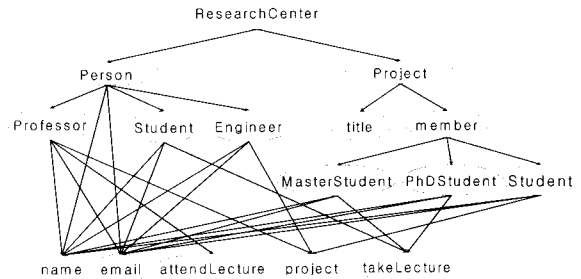


(a) 예제 DTD4

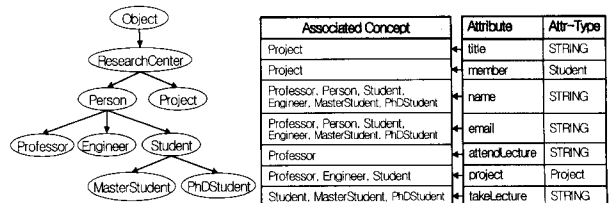


(b) 축소된 온톨로지

(그림 10) 온톨로지-DTD4 정합



(a) 예제 DTD5

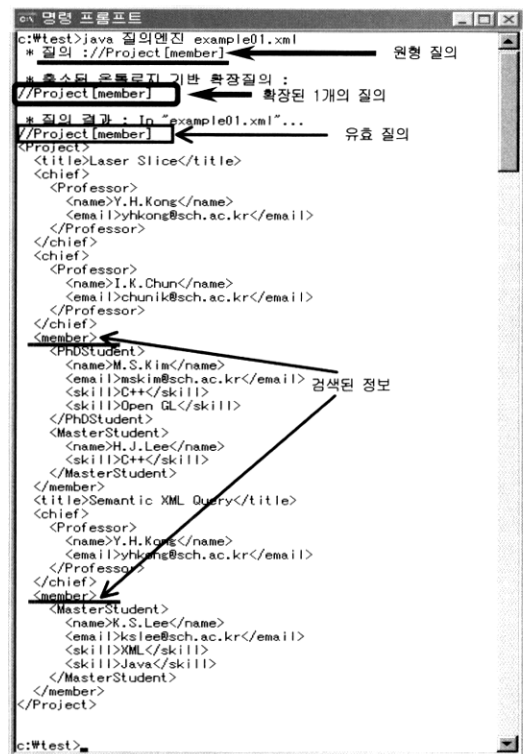
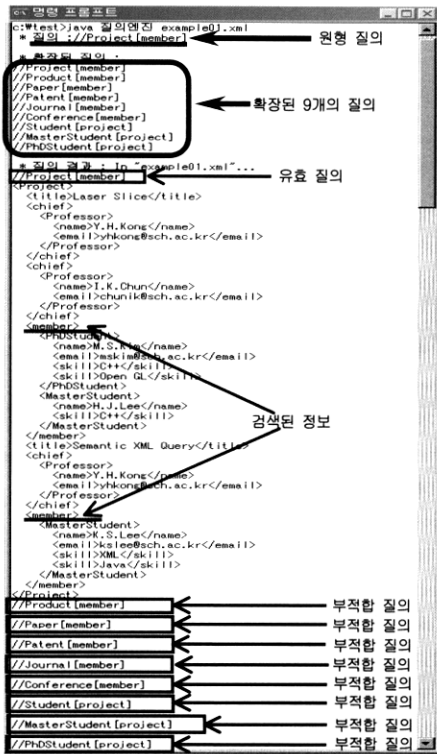


(b) 축소된 온톨로지

(그림 11) 온톨로지-DTD5 정합

터” 온톨로지와 유사한 정보를 포함하고 있다. (그림 2)의 온톨로지는 각각의 예제 DTD와 정합되어 (그림 7)(b), (그림 8)(b), (그림 9)(b), (그림 10)(b), (그림 11)(b)와 같이 적합하게 축소되었다.

이렇게 정합되어 축소된 온톨로지를 사용한 질의 확장의 효과를 다음과 같이 보였다. 먼저 //Project [member]라는 질의가 (그림 7)(a)의 구조를 가지는 특정 XML 문서에 인가되었을 때, (그림 2)의 원형 온톨로지로부터 확장되어 인가된 결과를 (그림 12)(a)에 보였으며, (그림 7)(b)의 축소된 온톨로지에 의해 확장된 질의가 인가된 결과를 (그림 12) (b)에 보였다. (그림 12)(a)에서 보듯이, //Project[member]는 6개의 질의 //(Project|Product|Paper|Patent|Journal|Conference)[member]로 확장될 뿐만 아니라, Project와 연관된 개념 Student에 대해서도 3개의 질의 //(Student|MasterStudent|PhDStudent)[project]를 추가하여 확장하였다. 그러나 원형 온톨로지서 확장된 9개의 질의 중에서 단지 1개의 질의 //Project[member]에



(a) 원형 온톨로지에 의한 질의 확장

(b) 축소된 온톨로지에 의한 질의 확장

(그림 12) DTD1 구조의 XML 문서에 대한 //Project[member] 질의 결과 비교

서만 정보가 검색되고 나머지 8개의 질의는 검색에 실패하였다. 반면에 (그림 12)(b)는 축소된 온톨로지가 Project[member]라는 하나의 질의만을 생성하여 검색에 성공하였음을 보였다.

이러한 결과는 원형 온톨로지가 동일 영역의 모든 문서 구조를 포함하도록 설계되어 있지만, 검색에 적합하지 않은 불필요한 요소들도 다수 포함하고 있음을 나타낸다. 반면에 정합된 온톨로지에는 이러한 불필요한 요소가 DTD와 정합되는 과정에서 모두 제거되어, 대상 문서에 부적합한 질의의 생성을 최소화하였음을 의미한다.

5.2 질의 확장 실험

본 절에서는 (그림 2)의 원형 온톨로지와 5.1절의 정합된 온톨로지를 사용하여 질의를 확장하고, 확장된 질의를 예제 XML 문서에 인가함으로써 검색의 효율을 실험하였다.

우선, 질의 //Project[member]를 (그림 2)의 원형 온톨로지가 확장한 결과와 (그림 7)(b), (그림 8)(b), (그림 9)(b), (그림 10)(b), (그림 11)(b)의 축소된 온톨로지가 확장한 결과를 <표 5>에 비교하였다. 구조가 서로 다른 5개의 DTD를 대상으로, 원형 온톨로지는 총 45개의 질의를 생성하였다. 반면에, 축소된 온톨로지는 7개의 질의만을 생성함으로써 원형 온톨로지에 비해 매우 적은 수의 질의만을 생성하였다. <표 6>에서 질의 //Person[skill]은 각각 30개와 7개의 질의를 생성함으로써 축소된 온톨로지가 원형 온톨로지에 비해 1/4만의 질의를 생성함을 보였다. 또한, <표 7>은 질의 //Engineer [project]가 35개와 3개의 질의를 생성하여 질

의의 수를 약 1/11로 감소시켰다. 이러한 일련의 실험 결과로부터 축소된 온톨로지가 원형 온톨로지에 비해 매우 적은 수의 질의만을 생성시킴을 확인할 수 있었다.

<표 5> 질의 확장 예 : //Project[member]

| DTD     | 질의  | //Project[member] <=> (적용된 연관규칙)//Student[project]                    |          |
|---------|---|---|----------|
|         |   | 원형 온톨로지   | 축소된 온톨로지 |
| DTD1    | //Project[member] //Product[member]<br>//Paper[member] //Patent[member]<br>//Journal[member] //Conference[member]<br>//Student[project] //MasterStudent[project]<br>//PhDStudent[project] | //Project[member]<br>//MasterStudent [member]<br>/PhDStudent[project] |          |
| DTD2    | //Project[member] //Product[member]<br>//Paper[member] //Patent[member]<br>//Journal[member] //Conference[member]<br>//Student[project] //MasterStudent[project]<br>//PhDStudent[project] | //MasterStudent [project]<br>//PhDStudent[project]                    |          |
| DTD3    | //Project[member] //Product[member]<br>//Paper[member] //Patent[member]<br>//Journal[member] //Conference[member]<br>//Student[project] //MasterStudent[project]<br>//PhDStudent[project] | No query  |          |
| DTD4    | //Project[member] //Product[member]<br>//Paper[member] //Patent[member]<br>//Journal[member] //Conference[member]<br>//Student[project] //MasterStudent[project]<br>//PhDStudent[project] | //Product[member]   |          |
| DTD5    | //Project[member] //Product[member]<br>//Paper[member] //Patent[member]<br>//Journal[member] //Conference[member]<br>//Student[project] //MasterStudent[project]<br>//PhDStudent[project] | //Project[member]   |          |
| 확장된 총질의 |   | 45개   | 7개       |

〈표 6〉 질의 확장 예 : //Person[skill]

| 질의<br>DTD  | //Person[skill] <=> (적용된 연관규칙 없음)  |  |
|------------|--|--|
|            | 원형 온톨로지  | 축소된 온톨로지   |
| DTD1       | //Person[skill] //Professor[skill]<br>//Engineer[skill] //Student[skill]<br>//MasterStudent[skill] //PhDStudent[skill] | //Student[skill="C"]<br>//PhDStudent[skill="C"]<br>//MasterStudent[skill="C"]                          |
| DTD2       | //Person[skill] //Professor[skill]<br>//Engineer[skill] //Student[skill]<br>//MasterStudent[skill] //PhDStudent[skill] | //Engineer[skill="C"]<br>//Student[skill="C"]<br>//MasterStudent[skill="C"]<br>//PhDStudent[skill="C"] |
| DTD3       | //Person[skill] //Professor[skill]<br>//Engineer[skill] //Student[skill]<br>//MasterStudent[skill] //PhDStudent[skill] | No query   |
| DTD4       | //Person[skill] //Professor[skill]<br>//Engineer[skill] //Student[skill]<br>//MasterStudent[skill] //PhDStudent[skill] | No query   |
| DTD5       | //Person[skill] //Professor[skill]<br>//Engineer[skill] //Student[skill]<br>//MasterStudent[skill] //PhDStudent[skill] | No query   |
| 확장된<br>총질의 | 30개  | 7개   |

〈표 7〉 질의 확장 예 : //Engineer[project]

| 질의<br>DTD  | //Engineer[project] <=> (적용된 연관규칙) //Project[technicalAdvisor]   |                                 |
|------------|--|---------------------------------|
|            | 원형 온톨로지  | 축소된 온톨로지                        |
| DTD1       | //Engineer[project] //Project[technicalAdvisor]<br>//Product[technicalAdvisor]/Paper[technicalAdvisor]<br>//Patent[technicalAdvisor] //Journal[technicalAdvisor]<br>//Conference[technicalAdvisor] | No query                        |
| DTD2       | //Engineer[project] //Project[technicalAdvisor]<br>//Product[technicalAdvisor]/Paper[technicalAdvisor]<br>//Patent[technicalAdvisor] //Journal[technicalAdvisor]<br>//Conference[technicalAdvisor] | //Engineer<br>[project]         |
| DTD3       | //Engineer[project] //Project[technicalAdvisor]<br>//Product[technicalAdvisor]/Paper[technicalAdvisor]<br>//Patent[technicalAdvisor] //Journal[technicalAdvisor]<br>//Conference[technicalAdvisor] | No query                        |
| DTD4       | //Engineer[project] //Project[technicalAdvisor]<br>//Product[technicalAdvisor]/Paper[technicalAdvisor]<br>//Patent[technicalAdvisor] //Journal[technicalAdvisor]<br>//Conference[technicalAdvisor] | //Product<br>[technicalAdvisor] |
| DTD5       | //Engineer[project] //Project[technicalAdvisor]<br>//Product[technicalAdvisor]/Paper[technicalAdvisor]<br>//Patent[technicalAdvisor] //Journal[technicalAdvisor]<br>//Conference[technicalAdvisor] | //Engineer[project]             |
| 확장된<br>총질의 | 35개  | 3개                              |

마지막으로 (그림 7)(a), (그림 8)(a), (그림 9)(a), (그림 10)(a), (그림 11)(a)의 DTD 구조를 가지는 예제 XML 문서

에 대하여 //Project[member], //Person[skill], //Engineer[project]의 세 가지 질의를 인가하는 실험을 하였다. 질의 //Project[member]의 경우, 원형 온톨로지로부터 확장된 질의 총 45개 중 6개만이 검색에 성공한 반면, 축소된 온톨로지로부터 확장된 질의는 7개 중 6개가 검색에 성공하였다. 질의 //Person[skill]에서는 원형 온톨로지는 30개의 질의 중 3개만이 검색에 성공하였으나, 축소된 온톨로지는 7개 중 6개가 성공하였다. //Engineer[project]의 경우에서도 원형 온톨로지는 35개의 질의 중 극히 일부인 3개만이 검색에 성공하였으나, 축소된 온톨로지로는 3개의 질의 모두가 검색에 성공하였다.

결과적으로 원형 온톨로지를 사용한 경우, 생성된 총 110개의 질의 중 15개 질의만이 검색에 성공하여 14%의 적중도를 나타내었다. 반면에 정합된 온톨로지는 총 17개의 질의를 생성하였으며, 생성된 질의의 대부분이 검색에 성공하여 약 88%의 적중도를 보였다. 즉, 원형 온톨로지에 의해 생성된 110개의 질의 중에 17개를 제외한 93개의 불필요한 질의가 축소된 온톨로지에 의해 효과적으로 제거됨을 확인할 수 있었다.

## 6. 결론

다양한 구조의 XML 문서를 대상으로 하는 정보검색에서 온톨로지 기반의 질의 확장은 단순 질의에 비해 보다 폭넓은 검색을 가능하게 하나, 이러한 방법은 대상 문서의 구조를 고려하지 않기 때문에 불필요한 질의를 과도하게 생성하여 검색의 효율을 저하시킬 수 있다.

본 연구는 대상 XML 문서의 DTD를 고려하여 질의가 최소한으로 확장될 수 있도록 온톨로지와 DTD를 효율적으로 정합하는 방법을 제안하였다. 제안한 방법은 온톨로지 개념 정합 알고리즘과 온톨로지 속성 정합 알고리즘의 두 단계로 구성되며, 각 단계는 불필요한 개념과 속성이 차례로 제거되어 온톨로지가 대상 문서에 적합하게 축소되도록 하였다. 정합 결과인 축소된 온톨로지를 기반으로 질의를 확장하게 되면, 대상 문서에 대한 질의의 적합도를 향상시킬 수 있다.

제안한 온톨로지-DTD 정합 방법에 의한 질의 확장의 효

〈표 8〉 질의 검색 적중도 비교

| 질의 분류<br>DTD 분류 | 적중도 (검색에 성공한 질의 수 / 총 질의 수) |              |                 |              |                     |               |
|-----------------|-----------------------------|--------------|-----------------|--------------|---------------------|---------------|
|                 | //Project[member]           |              | //Person[skill] |              | //Engineer[project] |               |
|                 | 원형 온톨로지                     | 축소된 온톨로지     | 원형 온톨로지         | 축소된 온톨로지     | 원형 온톨로지             | 축소된 온톨로지      |
| DTD1            | 3/9                         | 3/3          | 3/6             | 3/3          | 0/7                 | 0/0           |
| DTD2            | 1/9                         | 1/2          | 3/6             | 3/4          | 1/7                 | 1/1           |
| DTD3            | 0/9                         | 0/0          | 0/6             | 0/0          | 0/7                 | 0/0           |
| DTD4            | 1/9                         | 1/1          | 0/6             | 0/0          | 1/7                 | 1/1           |
| DTD5            | 1/9                         | 1/1          | 0/6             | 0/0          | 1/7                 | 1/1           |
| PTO합            | 6/45<br>(13%)               | 6/7<br>(88%) | 6/30<br>(20%)   | 6/7<br>(86%) | 3/35<br>(13%)       | 3/3<br>(100%) |
| 원형 온톨로지 합       | 15/110(14%)                 |              |                 |              |                     |               |
| 축소된 온톨로지 합      | 15/17(88%)                  |              |                 |              |                     |               |

과를 다양한 구조의 예제 XML 문서를 대상으로 실험하였다. 먼저, 제안한 정합 방법은 온톨로지를 대상 문서에 적합하도록 효과적으로 축소하였다. 또한, 원형 온톨로지에 의해 확장된 질의들은 검색 적중도가 14%에 불과하였으나, 정합된 온톨로지로부터 확장된 질의들은 88%의 높은 검색 적중도를 나타내었다.

결과적으로 온톨로지-DTD 정합에 의해 한번 축소된 온톨로지는 질의의 적중률을 높일 수 있을 뿐만 아니라 동일한 구조를 가지는 문서 집단에 재사용될 수 있으므로 검색의 효율을 향상시킬 수 있다.

### 참 고 문 헌

[1] J. Kamps, et al., "Best-Match Querying from Document-Centric XML", WebDB, 2004.

[2] M. Erdmann and S. Decker, "Ontology-award XML queries," WebDB, 2000.

[3] M. Erdmann and R. Studer, "How to Structure and Access XML Document with Ontologies," Data & Knowledge Engineering, Vol.36, No.3, pp.317-335, 2001.

[4] J. Robie, et al., XML Query Language (XQL), <http://www.w3.org/TandS/QL/QL98/pp/xql.html>, 1998.

[5] A. Deutsch, et al., XML-QL: A Query Language for XML, <http://www.w3.org/TR/NOTE-xml-ql/>, World Wide Web Consortium, August-1998.

[6] J. Clark, et al., XML Path Language (XPath) 1.0, <http://www.w3.org/TR/xpath>, W3C Recommendation, November 1999.

[7] S. Boag, et al., XQuery 1.0: An XML Query Language, <http://www.w3.org/TR/xquery/>, W3C Working Draft, February, 2005.

[8] 문찬호, 강현철, "링크 질의를 통한 XML 문서의 검색 기법", 정보처리논문지, 제8권, 제4호, pp.313-326, 2001.

[9] 손충범, 이병엽, 유재수, "BRS/Search 시스템을 이용한 XML 문서 검색시스템 설계 및 구현", 인터넷정보학회, 제2권, 제2호, 2001.

[10] J. Schanmugasundaram, "Relational database for querying XML documents : limitation and opportunities", VLDB, 1999.

[11] 이상원, "XML 포함질의를 위한 확장형 인덱스", 정보처리학회, 제11권, 제2호, pp.317-324, 2004.

[12] S. Cohen, et al., "XSEarch: A Semantic Search Engine for XML", VLDB, 2003.

### 김 명 속



e-mail : krhkms@sch.ac.kr  
 1995년 순천향대학교 전산학과(공학사)  
 1999년 순천향대학교 전산학과(공학석사)  
 2005년 순천향대학교 전산학과  
 (공학박사)  
 관심분야 : 웹응용, 영상처리, 알고리즘,  
 인공지능 등

### 공 용 해



e-mail : yhkong@sch.ac.kr  
 1982년 연세대학교 전자공학과(공학사)  
 1986년 미국 Polytechnic Univ. 전산학과  
 (공학석사)  
 1991년 미국 Polytechnic Univ. 전산학과  
 (공학박사)

1982년 한진중공업 연구원  
 1983년 삼성전자 연구원  
 1991년~현재 순천향대학교 정보기술공학부 교수  
 관심분야 : 시멘틱웹, 신경회로망, 멀티미디어 응용 등