

XML문서를 위한 인덱스기반의 다중 접근 제어*

최남규*, 황정희*, 류근호* 박진수**

*충북대학교 데이터베이스 연구실

**청주 대학교 정보 통신 공학부

e-mail: {cnam9, jhhwang, khryu}@dblab.chungbuk.ac.kr

{jspark}@cheongju.ac.kr

Index-base Multi Access Control for XML Document

Nam Kyu Choi*, Jeong Hee Whang*, Keun Ho Ryu*, Jin Soo Park**

*Database Laboratory, Chungbuk National University

**School of Information & Telecommunication, Cheongju University

요 약

최근, 정보 보호의 중요성이 부각됨에 따라 현재 웹 데이터 교환의 표준인 XML 데이터에 대한 보안과 접근 제어 기법이 주요 연구로 부상하고 있으며, 연구의 초점은 안전함을 보장하면서, 동시에 부가적인 비용 증가를 줄이는데 있다. 그렇지만 이러한 연구의 방식에서는 인덱스를 고려하지 않기 때문에 불필요한 탐색은 물론 데이터 증가에 따른 탐색 비용이 증가한다. 따라서 이 연구에서는 XPath의 원리를 기반으로 인덱스와 접근제어를 동시에 고려하는 인덱스 기반의 접근 제어 맵을 제안하고, 이 기법을 관계형 데이터베이스에 적용하여 구현하였다.

1. 서론¹⁾

XML 문서에 대한 관리를 위하여 순수한 XML 데이터베이스 연구뿐만 아니라 기존의 데이터베이스를 활용하거나 또는 이러한 시스템과 통합하고자 하는 연구가 진행되어 왔다. 현재 일반적으로 사용되고 있는 관계형 데이터베이스와 XML은 서로 다른 구조를 가지고 있기 때문에 이러한 이질적 구조에 대한 맵핑이나 변환에 대한 연구가 되어왔고, XML 문서의 계층적 구조와 순서화 된 특성을 효율적으로 지원하기 위한 인덱싱 기법들이 제안되었다[1,2]. 또한 최근에는 정보 유출의 위험성이 높아지면서 복수의 보안 등급을 갖는 XML문서를 위한 다중 접근 제어 기법에 대해서도 연구가 되고 있다[3,4,5].

그러나 기존의 접근제어 기법은 보안 등급의 표시에 따른 추가적인 비용을 최소화하는데 연구의 초점을 맞추고는 있지만, 탐색 영역 축소나 불필요한 탐색의 제거에 중요한 영향을 미칠 수 있는 XML문서의

인덱스 구조는 고려하지 않고 있다. 또한 XML문서를 위한 인덱스 연구에서는 접근 제어에 대한 연구를 병행하지 않고 있다.

따라서 이 연구에서는 기존의 인덱스 구조를 확장하고, 확장된 인덱스구조를 다중 접근 제어 기법의 기반 구조로 적용한다. 이러한 적용을 구체화 하기 위해서 XML문서의 계층적 구조와 순서화 된 특성을 유지하기 위한 인덱스 정보, 그리고 접근제어의 적용을 위한 보안 등급 정보를 함께 유지하는 인덱스 기반의 접근 제어 맵을 구성한다.

이 논문의 구성은 다음과 같다. 2장에서는 정규 경로 표현을 지원하는 인덱스 기법에 대해 설명하고, 3장에서는 기존의 접근 제어 방식의 문제점과 제안된 인덱스 기반의 다중 접근 제어를 설명한다, 그리고 4장에서는 제안하는 기법을 관계형 데이터베이스에 적용한 결과를 보이고, 5장에서는 결론 및 향후 연구를 기술한다.

2. 인덱스

XML 문서를 관계형 데이터베이스에 저장하고, XML의 정규 경로 표현 또는 질의를 만족하기 위해서

[†] 이 연구는 한국과학재단 지정 청주대 RRC(정보통신연구센터)의 지원으로 수행되었음

는 XML 각 문서에 대한 경로와 위치정보가 유지되어야 하고, 필요한 자료에 대해 보다 빨리 접근하기 위해서는 탐색영역을 최소화해야 한다[6]. 이 연구에서는 이러한 요구 사항을 충족시키기 위해 XPath(XML Path Language) 원리를 기반하는, [6]에서 제안되었던 인덱스 구조를 이용하고, 추가적으로 정규 경로 표현 식을 위한 추가적인 XML문서의 계층적 구조 정보를 함께 유지한다.

2.1 XPath

XPath는 XML Path Language로써 XML문서에 대한 경로와 위치정보의 탐색에 기반이 되는 XML 정규 표현식이다. 이 연구에서는 보다 빠른 문서 접근을 지원하기 위해 XPath 기반의 탐색 영역 축소를 위한 문서 분할과 경로와 위치정보에 대한 질의에 중점을 두어, 문서를 4개의 큰 축(Axis)으로 분할하는 표현식과 경로와 위치 정보 질의를 위한 표현식은 다음과 같이 기술된다.

a에서 j까지의 노드를 갖는 트리를 현재 노드(f)를 기준으로 4개의 축인 Ancestor, preceding, following, descendant로 분할하는 표현식은 다음과 같다[6].

$$\{a..j\} = (\underline{f/descendant} \cup \underline{f/ancestor} \cup \underline{f/following} \cup \underline{f/preceding} \cup \{f\})$$

또한 경로와 위치(position)정보를 동시에 탐색하는 XPath의 표현식은 다음과 같다[7].

$$/ParentPath[position]/ChildPath[position]$$

2.2 탐색 영역 축소를 위한 문서 분할

XML 문서를 그림1과 같이 트리로 표현할 때 preorder과 postorder 번호를 각 노드의 ID로 부여함으로써 XML 문서 전체의 순서와 위치를 유지하면서 동시에 부여된 ID의 간단한 비교를 통하여 문서를 4개의 축으로 분할할 수 있다[6].

2.3 경로와 위치 질의를 위한 인덱스 기법

기존의 인덱스 방식[6]과 같이 ID부여만으로는 각 경로의 관계와 위치를 표현하기 어렵기 때문에 깊이, 태그(문자열), 부모, 위치(position)정보를 추가적으로 유지한다.

이렇게 함으로써 XML문서를 4개의 큰 축으로 분할함으로써 탐색 영역을 축소하고, 추가적인 계층적 구

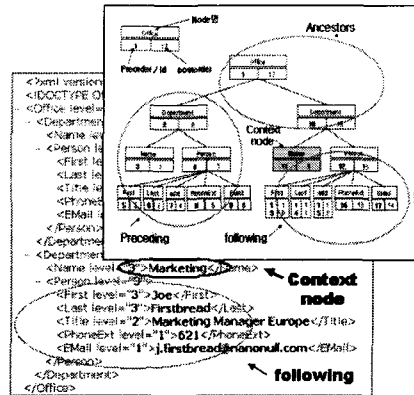


그림 1 축을 기반 한 인덱스 트리 조에 대한 정보를 유지함으로써 경로와 위치(position)질의를 요구하는 XML의 표준 경로식 표현을 지원한다.

3. 인덱스 기반의 다중 접근제어

3.1 접근 제어

XML문서에 대해 별도의 정책 또는 정의를 통해 XML 문서를 세부적으로 접근제어를 하는 방법[10]은 많은 저장 공간은 요구하지는 않지만, 많은 시간과 오버헤드를 유발하고, 반대로 모든 엘리먼트에 대해 속성처럼 보안등급 정보를 유지하는 방법은 시간과 오버헤드를 줄이는 대신, 많은 저장 공간을 요구한다[3]. 그러므로 이러한 문제점들을 해결하기 위해, 최소한의 노드에 대해서만 보안 등급 정보를 유지하고, 그 외의 노드에 대해서는 알고리즘 또는 규칙을 적용함으로써 수행 시간과 저장 공간에 대한 비용을 동시에 줄일 수 있다[3,4].

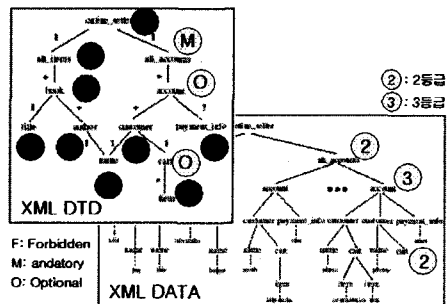


그림2 DTD를 통한 접근 제어 정책

접근제어를 적용하기 위해서 먼저 DTD와 같이 스키마 정의 차원에서 접근제어에 대한 정의를 하고,

그 다음 실제로 XML 문서 작성할 때 직접 각 엘리먼트에 세부적으로 보안 등급을 정의한다[3,5]. 즉, 그림 2와 같이 최소의 노드에만 보안 등급을 표현하고, 나머지 노드는 조상 노드의 보안 등급 상속과 최 근접 노드에 대한 재 정의를 통해 접근 제어를 적용 할 수 있다[3].

3.2 기존[3] 접근 제어 방식의 한계

앞에서는 XML문서를 XPath를 원리를 이용하여 현재 노드를 기준으로 4부분으로 분할하는 원리를 설명하였다.

그러나 기존[3]의 접근제어 방식은 엘리먼트의 탐색 또는 질의 수행시간에 대부분 엘리먼트의 보안 등급을 상속 또는 재 정의를 통해 표시하기 때문에, 분할을 하면 이러한 보안 등급의 전파(propagation)가 올바르게 이루어 질수 없다.

따라서 이 연구에서는 질의 수행 이전에 4개의 큰 축으로 XML문서를 분할 할 뿐 아니라, 사용자의 보안 등급에 따른 해당 등급의 엘리먼트만을 여과할 수 있도록 하는 인덱스 기반의 접근 제어 맵을 제안한다.

3.3 인덱스 기반의 접근 제어 맵

이 연구에서 제안하는 인덱스기반의 다중 접근 제어의 원리는 XML트리를 구성할 때 각 엘리먼트를 위한 인덱스 정보와 보안 등급 정보를 함께 저장하여 인덱스 기반의 접근 제어 맵을 구축하는 것이다.

인덱스 기반의 접근 제어 맵의 구축은 먼저 XML 트리의 표현은 각 엘리먼트 노드에 부여된 preorder와 postorder번호를 좌표로 적용하여 인덱스 맵으로 구성하고[03], 동시에 각 엘리먼트 노드 즉 좌표마다 보안 등급 정보를 속성처럼 저장 및 유지함으로써 가능케 된다. 즉 인덱스 맵에 인덱스 기반의 접근 제어 맵으로 확장 적용한다.

이렇게 구축된 인덱스 기반의 접근 제어 맵을 통해 4개의 기본 축과 각 보안 등급으로 여과 또는 분할을 질의 요청이전에 적용하여 탐색 영역을 축소할 수 있다. 예를 들어 사용자를 3개의 등급으로 나누고, 문서를 4개 부분으로 분할이 가능할 경우 최대 전체 XML문서의 1/12로 문서의 크기를 축소하여, 탐색영역을 줄일 수 있게 된다.

그림3은 현재노드 (preorder=11, post order=9)를 기준으로 3등급이고, preceding 영역을 질의하는 인덱스 기반의 접근 제어 맵의 예이다.

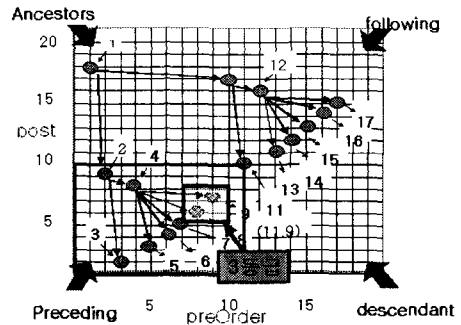


그림3 인덱스 기반의 접근 제어 맵(Map)

3.4 기존 접근 제어 방식[3]과의 비교

[3]의 방식은 사용자가 관심 있는 엘리먼트에 질의할 경우 접근하는 엘리먼트 노드마다 보안 등급 검사뿐만 아니라 보안 등급에 대한 상속 및 재 정의를 절차를 거쳐야 했지만, 이 연구의 인덱스 기반의 접근 제어 맵의 구조를 XML 데이터베이스와 함께 구축하는 경우 XML문서와의 맵핑을 통해 질의 수행 이전에 필요한 문서의 부분만을 추출 또는 분리할 수 있게 된다.

즉 기존의 접근제어[3] 방식은 질의 수행 할 때마다 엘리먼트에 대한 보안 등급의 전파와 검사를 해야 하는 반면 제안하는 인덱스 기반의 접근제어 방식은 최초 문서를 불러올 때만 보안 등급을 확인하고 필요한 문서의 부분만을 사용자에게 제공한다. 그러므로 질의 요청이 빈번 하거나, 특정 부분만을 탐색할 필요가 있는 상황에 대해 더욱 효과적이다.

4. 적용 및 구현

이 구현에서는 사용자가 문서검색 시스템에 보안 등급을 부여받은 사용자가 등록되었다고 가정하고, 특정 정보에 관심 있는 보안 등급이 부여된 사용자가 로그인 할 때 그에 적합한 문서를 제공하는 구현을 하였다.

우선 인덱스기반의 다중 접근 제어 기법을 관계형 데이터베이스에 적용하기 위해서 인덱스 기반의 접근 제어 맵 정보와 content 테이블 구조를 표1, 2와 같이 설계 하였다.

표1 인덱스 기반의 접근 제어 맵 정보

필드 레코드	ID	Pre Order	Post Order	Depth	Level	Parent	position
2	2	07	04	1	2	04	1
3	3	08	05	1	1	04	2

표2 content 테이블

레코드	필드	ID	tag	Content
2		2	Title	Office Manager
3		3	Phone	582-2337

이 구현에서는 인덱스 기반의 접근 제어 맵 정보를 데이터베이스 테이블로 구축한 후 이것을 content 테이블과 조인하는 질의를 통해 해당 문서를 추출도록 하였다. 만일 3.1절과 같이 접근 제어에 대한 정보 저장을 위한 저장 공간을 절약하고, 인덱스 연산 속도를 향상시키기 위해서는 메모리를 활용할 수 있는데, 우선 최소한의 인덱스와 접근 제어 정보를 테이블에 저장하고, 문서를 불러올 때 인덱스 기반의 접근 제어 맵을 메모리에 구축한 후 그 결과 값을 content 테이블과 맵핑함으로써 구현 할 수 있다.

그림 4는 그림1의 XML문서로부터 마케팅 부서의 직원 정보에만 관심 있는 2등급의 사용자를 위해 XML문서의 여과된 부분만을 제공하는 예인데, 이때 “마케팅 부서”의 name에 대한 엘리먼트 노드 ID는 (11, 9)이고, 이 엘리먼트 노드를 기준으로 following 축 부분과 2등급 사용자를 위해 2, 3등급을 갖는 엘리먼트의 내용을 추출한 것이다.

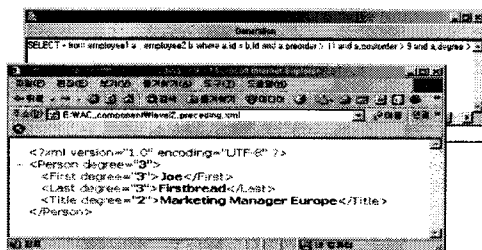


그림4 “마케팅”부서 정보에 관심있는 2등급 사용자를 위한 XML문서의 예

5. 결론 및 향후 연구

이 연구는 기존의 연구되어온 XML문서에 대한 인덱스 방식과 접근제어의 원리를 이용하였으나 기존 연구와 다음과 같은 차이점을 갖는다. 이 연구에서는 [6]의 엘리먼트에 대한 ID 부여 및 문서 분할의 방식을 기반으로 사용하지만, [6]에서 고려되지 않았던 접근 제어와 세부적 위치에 대한 질의에 대한 지원을 확장한다. [3,4]의 접근 제어 방식은 질의 수행 시간에 보안 등급에 대한 전파 및 검사를 하기 때문에 축에 기반을 둔 문서 분할이 불가능하지만, 이 연구에서는 인덱스 기반의 접근 제어 맵을 통해 질의 수행 이전에 축 기반과 보안 등급 기반의 분할

또는 추출을 가능케 한다. [8, 10] 연구는 정책 기반의 접근제어에 중점을 두지만, 이 연구는 [3]의 연구처럼 XML문서의 각 엘리먼트에 보안 등급을 속성처럼 적용한다.

이 연구에서는 위와 같이 인덱스를 고려하지 않은 기존의 접근 제어 기법의 문제점을 분석하여, 이를 개선하고자 인덱스 기반의 접근 제어 기법인 인덱스 기반의 접근 제어 맵을 제안하였고, 이를 어떻게 XML문서를 위한 데이터베이스와 인덱스 설계에 적용할지에 대해 설명하였다, 그리고 관계형 데이터 베이스에 대해 적용한 결과를 보였다.

향후에는 기존의 접근 제어 기법들과의 실험을 통해 비교 평가를 수행하고, 이를 실용에 적용하기 위해 비즈니스 거래를 위한 단일 환경으로 이용되고 있는 ebXML에 적용하고자 한다.

참고문헌

- [1] Igor Tatarinov, Stratis D. Viglas, Kevin Beyer, Jayavel Shanmugasundaram, Eugene Shekita, Chun Zhang. Storing and Querying Ordered XML using a Relational Databases System, In SIGMOD Conference, 2002
- [2] Quanzhong Li, Bongki Moon, Indexing and Querying XML Data for Regular Path Expressions, VLDB2001
- [3] SungRan Cho, Sihem Amer-Yahia, Lakes V.S. Lakshmanan, Divesh Srivastava, Optimizing the Secure Evaluation of Twig Queries, VLDB, 2002
- [4] Ting Yu, Divesh Srivastava, Lakes V.S. Lakshmanan, H.V. Jagadish, compressed Accessibility Map: Efficient Access Control for XML, VLDB2002
- [5] E.Damiani, S.D.C.di Vimercati, S.Paraboschi, and P.Samarati, Design and implementation of an access control processor for XML Documents. Computer Networks, 33(1-6):59-75, 2000. Also in WWW9.
- [6] Torsten Grust. Accelerating Xpath Location Steps, In SIGMOD Conference, 2002
- [7] XML Path Language (XPath) 2.0 <http://www.w3.org/TR/xpath20/>
- [8] M. Kudo and S. Hada, XML document security based on provisional authorization. In ACM Conf. Computer and Communications Security, 2000.
- [9] Chin-Wan Chung, Jun-Ki Min, Kyuseok Shim, APEX: An Adaptive Path Index for XML data, In SIGMOD Conference, 2002
- [10] E.Bertino, S. Castano, and E. Ferrari. Securing XML documents with Author-X. IEEE Internet Computing, 5(3):21-31, 2001