

# 무선 방송 환경에서 복합 XPath 표현식의 질의를 지원하기 위한 XML 스트리밍 기법\*

박상현, 하중우, 류병걸, 이상근  
고려대학교 정보통신대학 컴퓨터통신공학부  
e-mail:{condols, okcomputer, smart123, yalphy}@korea.ac.kr

## XML Streaming Method to Support Complex XPath Expressions in Wireless Broadcasting Environments

Sang-Hyun Park, Byung-Gul Ryu, Myong-Soo Lee, SangKeun Lee  
Division of Computer and Communication Engineering, Korea University

### 요 약

XML 데이터 사용의 증가와 그 응용분야가 다양해짐에 따라 무선 방송 환경에서 효율적으로 XML 데이터를 전송하기 위한 기법이 요구된다. 본 논문에서는 무선으로 방송되는 XML 스트림에 대하여 복합 XPath 표현식으로 구성되는 질의를 처리하기 위한 기법을 제안한다. XML 스트림은 기존의 반구조적인 데이터의 경로요약 기법을 활용하여 생성한다. 성능평가를 통하여 제안하는 기법이 복합 XPath 표현식이 포함된 질의에 대해서도 에너지 효율적임을 보인다.

### 1. 서 론

각종 모바일 장비와 무선 기술의 발달에 따라 다양한 서비스들이 무선 방송 환경을 기반으로 제공되고 있다. 대표적인 서비스로서 교통정보나 날씨정보 그리고 DMB 서비스[7]등을 예로 들 수 있다. eXtensible Markup Language (XML)[2]는 데이터의 표현 및 교환의 표준으로서 그 연구 분야가 무선 환경으로 확대되고 있다[4][9][12]. 또한, 무선 방송 기법을 통한 데이터 전송은 많은 사용자들에게 서비스를 제공하는 것을 가능하게 한다[3]. 이때 무선으로 방송되는 데이터 중 사용자가 원하는 데이터를 접근하기 위해 색인정보를 사용함으로써 사용자는 방송되는 모든 데이터를 수신할 필요가 없다[6]. 즉, 색인정보를 통해 자신이 원하는 데이터가 방송되는 시점을 예측하고 원하는 데이터만을 수신함으로써 사용자 장비의 에너지를 절약하는 것이 가능하다.

XML 내에 저장된 데이터는 계층적으로 구조화(hierarchically structured) 되어 있으며 자기 기술적(self-describing)이다. 이때 XML 문서의 구조정보는 방송 스트림의 색인으로서 활용할 수 있으며 이를 이용하여 방송채널을 통해 XML을 전송하는 것이 가능하다.

본 논문에서는 무선 방송의 장점을 이용하여 XML 데이터를 무선으로 방송하는 기법을 제안한다. 이때 무선 방

송 환경에서는 데이터가 전송되는 한 주기 안에 사용자가 원하는 데이터를 모두 수신할 수 있어야 한다. 또한 방송되는 XML 데이터로부터 원하는 데이터만을 수신하기 위해서는 XPath 와 같은 질의를 통해 방송되는 XML 데이터로부터 질의에 대한 결과를 얻어야만 한다. 무선 방송 환경의 경우 방송의 한 주기 내에서 원하는 데이터가 방송되는 시점에 채널에 튜닝하지 못할 경우 다음 주기까지 대기해야 하며 이로 인해 접근시간의 지연이 발생한다. 이러한 문제를 해결하기 위해 XML의 구조정보를 색인으로 활용하여 무선으로 방송되는 XML 데이터에 대하여 단순 XPath 표현식(Simple XPath Expressions)으로 구성되는 질의를 처리하기 위한 XML 스트리밍 기법을 선행연구로 진행하였다[10]. 그러나 단순 XPath 표현식으로 구성되는 질의의 경우 그 질의가 얻을 수 있는 XML 문서 내에서의 데이터의 범위가 작다. 즉, 질의의 단순함으로 인해 얻을 수 있는 데이터의 다양성이 제약 받게 되는 것이다. 이러한 문제점을 보완하기 위해 본 논문에서는 단순 XPath 표현식의 질의뿐만 아니라 복잡 XPath 표현식(Complex XPath Expressions)의 질의를 지원하기 위한 XML 스트리밍 기법을 제안한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 살펴보고 3장에서는 제안하는 XML 스트리밍 기법 및 질의 처리 기법을 설명한다. 그리고 4장에서는 성능평가 결과를 살펴보고 마지막으로 5장에서는 결론을 내리고 향후 연구방향을 제시한다.

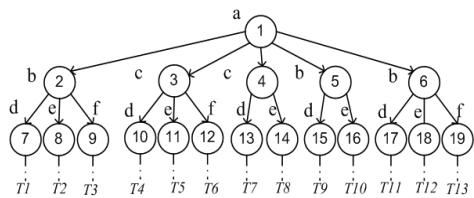
\* 이 연구에 참여한 연구자는 '2단계 BK21'의 지원비를 받았음.

2. 관련 연구

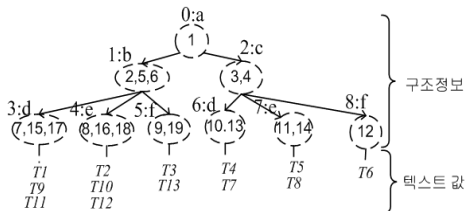
무선 환경에서 XML 데이터를 전송하기 위한 대표적인 기법으로 Xstream 을 예로 들 수 있다[12]. Xstream 은 XML Fragmentation 과 Packetizing 기법을 통해 무선 환경으로 XML 데이터를 전송하는 기법을 제안하였다. 그러나 Xstream은 XML 데이터를 무선으로 전송하는데 있어 에너지 효율성을 고려하지 않았다. 본 논문의 연구와 유사한 연구로서 S-Node[9]가 제안되었다. 그러나 S-Node 는 XML 스트림에 label path 의 중복이 발생하기 때문에 접근시간의 지연이 발생한다. 또한 Chung and Lee[4]는 무선 XML 스트림을 위한 두 개의 계층적 구조와 색인기법을 제안하였다. 그러나 이 또한 단순 XPath 표현식으로 구성되는 질의만을 고려하였다. 본 논문에서는 단순 XPath 표현식의 질의뿐만 아니라 복합 XPath 표현식의 질의를 지원하기 위한 XML 스트리밍 기법을 제안한다. 이때 XML 스트림 데이터는 기존의 반구조적 데이터의 경로요약(Path Summary) 기법[5]을 활용하여 생성되며 이에 대한 복합 XPath 표현식 질의의 처리 기법 또한 제시한다.

3. 제안하는 XML 스트리밍 기법

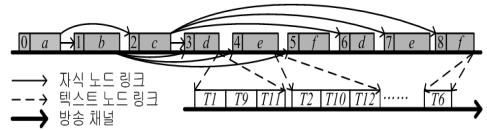
그림 1은 XML 데이터 트리의 예를 보여준다. 예를 들어 Q1:a/b/d/text() 와 같은 단순 XPath 표현식으로 구성된 질의에 대한 결과값은 노드 7, 15, 17의 텍스트 값인 T1, T9, T11 이다. 이때 Q1의 경우 원본 XML 데이터 트리에서 3개의 중복된 경로를 가지게 된다. 따라서 기존의 반구조적 데이터의 경로 요약기법으로 원본 XML 데이터 트리를 축약하면 그림 2와 같이 나타낼 수 있다.



(그림 1) 원본 XML 데이터 트리



(그림 2) 원본 XML 데이터 트리의 경로 요약



(그림 3) XML 스트림 구조

그림 1의 원본 XML 데이터 트리를 그림 2와 같이 표현하게 되면 Q1에 해당하는 3개의 중복된 경로가 하나로 축약되어진다. 즉, Q1에 대한 질의의 결과값은 그림 2에서 3번 노드의 텍스트 값과 같으므로 한 경로만을 읽음으로써 원하는 질의의 결과를 모두 얻을 수 있다. 그림 2의 경로요약을 기반으로 무선 방송을 위한 XML 스트림을 구성한다. 연속된 일련의 스트림 데이터를 생성하기 위해 그림 2의 경로요약본의 넓이 우선 탐색을 통해 그림 3과 같은 XML 스트림을 생성한다. 이때 각 노드간의 구조정보를 유지하기 위해 각각의 노드는 자신의 자식노드에 대한 링크 정보를 가지게 되는데 이는 다음 자식노드가 방송되는 시점을 나타내게 된다.

이때 경로 요약 기법으로 인해 원본 XML 문서의 구조 정보를 잃게 되는 일이 발생하며 스트림을 수신할 시 원본 데이터로의 동일한 복원이 불가능하다는 문제가 발생한다. 이러한 문제는 기존의 다양한 XML 문서의 색인기법을 활용함으로써 구조정보를 복원하는데 응용될 수 있다. 즉, 비록 원본 XML 문서를 경로 요약기법으로 축약하였으나 축약되기 전의 노드간의 계층정보를 부가적인 색인정보로 스트림 정보에 추가할 수 있다. 예를 들어 그림 1의 2번노드는 9번 노드(f노드)를 자식노드로 가지고 있으나 5번 노드(b 노드)의 경우에는 자식노드중 f가 존재하지 않는다. 이때, 그림 2와 같이 경로요약이 이루어지게 되면 2, 5, 6번 노드가 한 노드(b 노드)로 축약이 되게 되고 자식노드로서 d, e, f를 모두 가지게 되기 때문에 축약된 b노드 중 어떤 노드가 f를 자식 노드로 가지고 있지 않은지 식별하기 어렵게 된다. 이때 이를 구별하기 위해서는 b노드에 10011 과 같은 bit sequence로 이루어진 색인정보를 추가한다. 이것은 b노드가 1번째 depth에서 넓이 우선 탐색으로 볼 때 1번째, 4번째 그리고 5번째 노드라는 것을 나타낸다. 그리고 그림 2에서 b 노드의 자식노드인 d와 e노드는 모두 10011 의 부가적인 색인정보를 가지고 f 노드는 10001의 부가적인 색인정보를 가진다. b노드의 색인정보인 10011과 f 노드색인정보인 10001을 비교하게 되면 f 노드는 첫 번째 b노드와 세 번째 b 노드에만 존재함을 알 수 있는 것이다. 위의 예와 같이 경로요약을 통해 스트림의 크기를 줄이고 원본 XML 문서의 색인정보를 추가하는 것이 단순히 경로요약 없이 원본 XML 문서를 스트리밍 하는 것보다 접근시간 및 튜닝시간에서 더

나은 성능을 보여줄 수 있다.

본 연구의 선행연구[10]에서는 위와 같이 구성된 XML 스트림 데이터에 대한 단순 XPath 표현식의 질의에 대한 성능 평가 결과만을 제시하였다. 그러나 Q2://c/d/text()와 같은 복합 XPath 표현식이 포함된 질의를 처리하고자 할 경우 단순 XPath 표현식만으로 구성된 질의의 경우와는 달리 방송되는 스트림에서 질의에 상응하는 최초 노드 c를 확인할 수 있어야 한다. 또한 Q2와 같은 질의의 패턴이 XML 스트림내에 중복되어 나타날 수 있다. 따라서 질의에 “//” 혹은 “\*”와 같은 복합 XPath 표현식이 포함된 경우에는 해당 표현식 이후에 질의와 대응되는 자식 노드를 찾기 위해 링크를 연속적으로 탐색한다. 예를 들어 Q2의 질의의 경우 첫 번째 노드인 c가 루트 노드로부터 몇 번째 깊이(depth)에 존재하는지 알 수 없다. 왜냐하면 “//”의 표현식으로 기술되어 있기 때문에 임의의 깊이에서 중복적으로 나타날 수 있는 것이다. 따라서 이러한 문제점을 해결하기 위하여 스트림의 각 노드는 자신의 깊이 정보를 가진다. XML 스트림은 경로 요약 기법을 기반으로 축약되어 있기 때문에 특정 깊이에서 c노드를 발견하게 되었을 때 그 이후의 자식노드들의 링크만 확인하게 되면 임의의 깊이에 존재하는 질의 패턴을 모두 찾는 것이 가능하다. 만약 이미 상위의 깊이에서 c를 찾았다고 하더라도 깊이가 한 단계 깊어질 때마다 해당 깊이에 또 다른 c노드가 존재하는지 검사한다. 만약 동일한 c노드가 존재하면 그 자식노드의 링크도 탐색하게 된다. 만약 스트림 검색 중에 질의 패턴과 맞지 않는 하위 노드를 확인한 경우에는 바로 탐색을 중지하고 지금까지 일치했던 질의 패턴만을 계속적으로 탐색한다. 이때 c노드를 발견할 때까지는 방송 채널을 모두 튜닝하게 된다. 이것은 질의에 복합 XPath 표현식이 포함되어 있기 때문에 단순 XPath 표현식의 질의보다 더 많은 튜닝시간이 불가피하게 요구된다. 그러나, 경로요약 기법을 사용하지 않고 XML 스트림을 생성할 경우 Q2와 같은 질의를 처리하기 위해서는 경로요약을 사용했을 때보다 더 많은 접근시간의 지연을 유발하게 된다. 따라서 비록 단순 XPath 표현식의 질의보다는 성능에서 비효율적이거나 복합 XPath 표현식의 질의 성능은 경로요약을 사용하지 않은 경우보다는 더 나은 성능을 보여주게 된다.

#### 4. 성능평가

제안하는 XML 스트리밍 기법을 평가하기 위한 실험 데이터로서 XMark [11], DBLP [8], SwissProt 을 사용하였으며, 실험은 펜티엄 III 2.8Ghz CPU 와 1GB 메모리의 Windows XP 운영체제에서 이루어졌다. XML 스트림을 생성하고 질의를 처리하는 알고리즘은 Java 언어로 개발하였으며 XML 파싱을 위해 Apache Xerces 파서[1]를 사

용하였다. XML 스트림은 64KB의 버킷단위로 전송되도록 시뮬레이션을 수행하였다.

일반적으로 무선 방송 기법의 성능을 측정하는 평가 척도로서 접근시간과 튜닝시간[6]이 사용되며 본 논문에서도 동일한 척도를 사용하였다. 이때 접근시간과 튜닝시간의 상대적인 성능향상을 비교하기 위하여 원본 XML 데이터를 전송할시 소요되는 버킷수에 대하여 제안하는 스트리밍 기법으로 XML 데이터를 전송할때의 버킷수의 비율로 나타내며 다음과 같이 정의된다.

$$\text{시간비율(\%)} = \frac{\text{XML 스트림의 버킷수}}{\text{원본 XML 데이터의 버킷수}} \times 100$$

실험에 사용된 XML 데이터는 표1 과 같다. 각각의 데이터에 대하여 서로 다른 특성을 지닌 두가지 XPath 질의를 수행하였다. 질의는 복합 XPath 표현식과 단순 XPath 표현식이 모두 포함되도록 작성하였다.

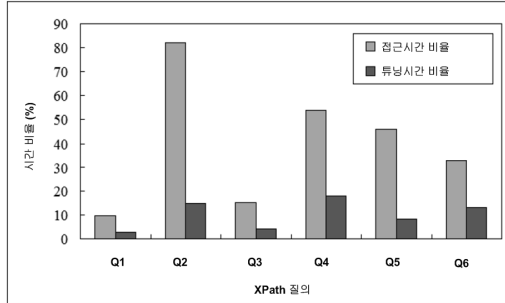
기존의 무선 방송 환경에서 XML 스트리밍을 다룬 연구에서는 본 논문에서와 같은 복합 XPath 표현식의 질의를 다루고 있지 않으며 단순 XPath 표현식의 질의에 대한 성능만을 제시하였다. 단순 XPath 표현식의 질의에 대한 비교평가는 선행연구[10]에서 수행하였으며 단순 XPath 표현식의 질의에서도 본 논문에서 제안하는 기법이 접근시간 및 튜닝시간에서 나은 성능을 보여주었다. 표1의 데이터와 질의에 대한 성능평가 결과는 그림 4에서 보여주고 있다.

(표 1) XML 데이터와 XPath 질의

XML 데이터	크기(MB)	질의
XMark	4.54	Q1://regions/*/item/name/text() Q2://description/parlist/listitem/text/text()
DBLP	2.95	Q3://dblp/booktitle/text() Q4://dblp/*/title/
SwissProt	9.93	Q5://Entry/Features/*/Descr/text() Q6://Entry/*/author/text()

평균 접근시간과 튜닝시간 비율은 각각 40% 와 10%이다. XPath 질의에 조상-후손 관계의 경로표현(“//”)과 와일드 카드 경로표현(“\*”)이 존재하는 경우 모바일 클라이언트는 부모-자식 관계의 경로표현(“/”)이 포함된 경우보다 더 많은 버킷을 튜닝하게 된다. 또한 이러한 경로 패턴이 여러번 나타나는 경우에는 더 많은 버킷을 튜닝하게 되므로 평균 접근시간은 원본 XML 데이터에 비해 향상되지만 많은 차이를 보이지는 않는다. 그러나 튜닝시간의 경우에는 매우 향상된 성능을 보여주고 있다. 이는 비록

XPath 질의가 복합 XPath 표현식을 포함하고 있다고 하더라도 경로요약에 의해 생성된 XML 스트림에서 질의에 해당하는 각각의 노드는 한번만 튜닝하면 되기 때문에 튜닝시간에서는 월등한 성능 향상을 나타내게 된다.



(그림 4) 제안하는 기법의 성능평가 결과

## 5. 결론 및 향후연구

본 논문에서는 무선 방송 환경에서 복합 XPath 표현식의 질의를 지원하는 XML 스트리밍 기법을 제안하였다. 경로 요약 방법을 활용하여 XML 스트림을 생성하였으며 모바일 클라이언트는 방송되는 XML 스트림 중 원하는 데이터를 XPath 질의를 통해 얻을 수 있었다. 복합 XPath 표현식이 포함된 질의는 단순 XPath 표현식이 포함된 질의보다 접근시간은 더 소요되나 튜닝시간에서는 동등한 성능향상을 보인다. 즉, 복합 XPath 표현식의 질의에 대하여도 모바일 장비의 에너지를 절약할 수 있는 것이다. 본 논문에서는 한 경로만을 탐색하는 질의의 확장 에 관한 연구결과를 제시하였다. 향후 연구로서 Twig 패턴 질의와 같이 보다 다양한 XPath 질의를 지원할 수 있는 스트리밍 기법을 개발할 것이다.

## 참고문헌

- [1] Apache XML Project. <http://xml.apache.org/>.
- [2] Extensible Markup Language. <http://www.w3.org/XML>.
- [3] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik. Broadcast disks: Data Management for Asymmetric Communication. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 199-210, 1995.
- [4] Y. D. Chung and J. Y. Lee. An Indexing Method for Wireless Broadcast XML data. *Information Sciences*, 177(9):1931-1953, 2007.
- [5] R. Goldman and J. Widom. Dataguides: Enabling Query Formulation and Optimization in Semistructured Databases. In *Proceedings of International Conference on Very Large Data Bases*, pages 436-445, 1997.
- [6] T. Imielinski, S. Viswanathan, and B. Badrinath. Data on Air: Organization and Access. *IEEE Transactions on Knowledge and Data Engineering*, 9(3):353-372, 1997.
- [7] Y. Jeong and W. Kim. A Novel TPEG Application for Location Based Service Using Terrestrial-DMB. *IEEE Transactions on Consumer Electronics*,

52(1):281-286, 2006.

- [8] M. Ley. DBLP XML records. <http://www.informatik.uni-trier.de/~ley/db/>.
- [9] C.-S. Park, C. S. Kim, and Y. D. Chung. Efficient Stream Organization for Wireless Broadcasting of XML Data. In *Proceedings of Asian Computing Science Conference*, pages 223-235, 2005.
- [10] S.-H. Park, J.-H. Choi, and S. Lee. An Effective, Efficient XML Data Broadcasting Method in a Mobile Wireless Network. In *Proceedings of the International Conference on Database and Expert Systems Applications*, pages 358-367, 2006.
- [11] A. Schmidt, F. Waas, M. L. Kersten, M. J. Carey, I. Manolescu, and R. Busse. XMark: A Benchmark for XML Data Management. In *Proceedings of the International Conference on Very Large Data Bases*, pages 974-985, 2002.
- [12] E. Y. Wong, A. T. Chan, and H. V. Leong. Xstream: A Middleware for Streaming XML Contents over Wireless Environments. *IEEE Transactions on Software Engineering*, 30(12):918-935, 2004.