

# 무선 XML 데이터 방송을 위한 스트림 최적화

박준표\*, 정연돈\*

\*고려대학교 컴퓨터학과

e-mail : {jp\_park, ydchung}@korea.ac.kr

## Stream Optimization for Wireless XML Data Broadcast

Jun Pyo Park\*, Yon Dohn Chung\*

\*Dept. of Computer Science and Engineering, Korea University

### 요 약

최근 사용자가 휴대용 단말기를 사용하여 이동하면서 정보를 주고 받는 무선 모바일 컴퓨팅 환경이 확산되고 있다. 한편, 방송 기법을 통한 데이터의 전달은 대역폭 활용의 이점과 확장성으로 인해 무선 모바일 컴퓨팅에서 효과적인 방법으로 알려져 있다. 본 논문에서는 무선 방송 환경에서 시간에 대한 의존도가 높은 XML 데이터를 전달하기 위한 스트림 최적화 기법을 제안한다. 이를 위해 경로 요약 기법 방법으로 노드들을 그룹화하고, 중복된 속성 이름을 생략하고 속성 값과 듀이 순서만 표시함으로써 스트림의 크기를 최소화하는 속성 요약 방법을 제안한다. 실험을 통해 제안 방법이 기존의 무선 XML 데이터 방송 기법에 비해 접근 시간 및 에너지 효율적임을 보여주고 있다.

### 1. 서론

네트워크 관련 기술이 발전함에 따라, 사용자가 휴대용 단말기를 사용하여 이동하면서 정보를 주고 받는 무선 모바일 컴퓨팅 환경이 확산되고 있다. 한편, 무선 방송 기법은 다수의 이동 사용자에게 데이터를 전달함에 있어, 대역폭 활용의 이점과 확장성으로 인해 무선 모바일 컴퓨팅에서 효과적인 방법으로 알려져 있다. 무선 방송 환경에서 서버는 방송 채널을 통해 데이터를 방송하고, 이동 사용자들은 방송 채널로부터 데이터를 수신한다. 방송 기법을 통한 데이터의 전달은 서버의 부담을 최소화할 수 있으며, 이동 사용자 입장에서는 에너지 소모가 많은 '송신' 작업 없이 원하는 데이터만을 선택적으로 수신할 수 있는 이점이 있다[1].

무선 방송 환경에서 이동 사용자들은 배터리 자원이 한정되어 있으므로, 에너지 보존을 위해 데이터의 선택적인 수신이 요구된다. 또한 이동 사용자에게 빠른 응답을 주기 위해 전체 질의 처리 시간 또한 최소화되어야 한다. 전자를 에너지 효율성, 후자를 접근 시간 효율성이라 일컫는다. 방송 기법을 통한 데이터 전달에 있어, 이동 사용자의 에너지 효율성과 접근 시간 효율성을 측정하기 위해 각각 튜닝 시간(tuning time)과 접근 시간(access time)이 성능 측정의 척도로 활용된다[2]. 튜닝 시간은 이동 사용자가 방송 채널을 수신하며 데이터를 읽은 시간들의 합이고, 접근 시간은 이동 사용자의 질의가 시작한 시점부터 모든 필요한 데이터를 수신 받은 시점까지의 소요 시간이다.

본 논문에서는 시간에 따른 의존도가 높은 XML 데이터를 무선 방송 기법을 통해 전달하고자 할 때,

이동 사용자들의 접근 시간 효율성과 에너지 효율성을 극대화할 수 있는 스트리밍 기법에 대해 다루고 있다. XML(eXtensible Markup Language)[3]은 데이터의 표현 및 전달을 위한 표준으로서 다양한 분야에서 활용되고 있는 언어이다. 그림 1은 간단한 XML 문서의 예를 보여주고 있다. 이 XML 문서는 본 논문 전체에 걸쳐 예제로 활용된다.

```
01:<mondial>
02:  <continent ID="f0_119" NAME="Europe" />
03:  <country ID="f0_136" NAME="Albania" >
04:    <name>Albania</name>
05:    <name>keshilli</name>
06:    <city ID="f0_1461" COUNTRY="f0_136" >
07:      <name>Tirana</name>
08:      <population YEAR="87">172010</population>
09:    </city>
10:    <city ID="f0_36499">
11:      <name>Shkoder</name>
12:      <population YEAR="87">192000</population>
13:    </city>
14:  </border> ... </border>
15:  <city>Shkoder
16:    <name>...</name>
17:  </city>
18:  <langs>English</langs>
19:  <religions>
20:    <population>120130</population>
21:  </religions>
22: </country>
23: ...
24:</mondial>
```

(그림 1) 간단한 XML 문서의 예

지금까지 XML 데이터의 스트리밍 및 XML 데이터 스트림의 방송 기법에 대해 여러 연구들이 이루어져 왔다 [4, 5, 6]. 그러나, 기존의 연구들은 무선 방송 환경에서 이동 사용자의 에너지 효율성 및 접근 시간

효율성 문제를 다루지 않았거나[4], 에너지 효율성에 중점을 두기 때문에 시간에 따른 의존도가 높은 XML 데이터를 방송하기에 취약한 단점이 있다[5, 6].

본 논문에서는 무선 방송 환경에서 시간에 따른 의존도가 높은 XML 데이터 처리에 최적화된 스트림 구조를 제안한다. 접근 시간과 튜닝 시간을 최적화하기 위해 경로 요약 및 속성 요약 방법을 통해 스트림의 크기를 최소로 압축시키고, 압축된 스트림에서의 질의 처리 알고리즘을 제안한다.

이 논문의 구성은 다음과 같다. 관련 연구 및 연구 동기는 2 장에 기술되어 있다. 3 장에서는 XML 데이터 스트림의 구조를 정의하고 스트림의 생성 방법을 설명한다. 4 장에서는 제안 기법에 의해 생성된 XML 데이터 스트림에서의 질의 처리 알고리즘에 대해 설명한다. 제안 기법에 의해 생성된 스트림의 크기 비교 및 성능 측정은 5 장에 기술되어 있다. 마지막으로 6 장에서 논문의 결론을 도출한다.

2. 관련 연구

2.1 관련 연구

무선 XML 데이터 방송 기법과 관련하여, [5]에서는 S-node 라는 스트리밍 단위의 정의를 통해 에너지 효율적인 XML 스트리밍 방법들과 질의 처리 알고리즘들을 제안하고 있다. 이 방법은 XML 문서의 구조적 특징을 이용하여 각각의 S-node 가 같은 태그 이름을 가진 형제 노드(same-tag sibling link)와 다른 태그 이름을 가진 형제 노드(different-tag sibling link)에 대한 주소를 가진 트리 형식의 색인 구조를 구성한다. 이 색인 정보를 통해 이동 사용자는 불필요한 데이터의 수신을 생략하고 원하는 데이터만 선택적으로 수신한다.

[6]는 무선 방송 환경에서 경로 요약(path summary) 기법을 활용하여 에너지 효율적인 XML 스트림의 질의 처리 방법을 제안하였다. 이 방법은 XML 문서의 구조적 정보와 텍스트를 분리하여 색인정보로 활용하며, 중복된 레이블 경로(label path)를 생략함으로써 전체 스트림의 크기를 줄였다. 그림 2 는 이 방법에 의해 생성된 XML 데이터 트리를 보여주고 있다. 모든 중복된 레이블 경로는 요약되었으며, 각 인덱스 노드는 자손 노드들과 텍스트에 대한 위치 정보를 포함하고 있다. 그러나 이 기법에서 노드의 순서는 무시되었고 속성을 생략하였기 때문에 복잡한 XPath 질의 처리를 지원하지 않는다.

2.2 연구 동기

주식정보나 실시간 교통정보와 같이 시간에 따른 의존도가 높은 데이터는 시간에 따라 데이터의 갱신이 빈번하게 일어나거나 일정 시간이 지나면 정보의 가치가 사라지는 특성이 있다. 본 논문에서는 이를 “시간 임계성 데이터”라 부르기로 한다.

**시간 임계성 데이터 (Time Critical Data).** 데이터의 갱신이 빈번하게 일어나거나 시간에 지나면 정보의 가치가 하락 또는 무효화되는 데이터.

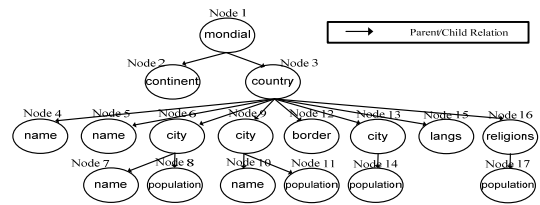
기존의 연구들은 무선 방송 환경에서 이동 사용자의 한정된 배터리 자원을 효율적으로 사용하기 위하여 선택적인 수신을 위한 색인 정보를 구성하고, 이에 따라 증가하는 접근 시간을 최소화하는 방향으로 연구되어 왔다[5, 6]. 그러나 이와 같은 방법들은 속성이 포함된 XPath 경로 질의를 지원하지 않거나[6], 시간에 따라 정보의 가치가 하락 또는 무효화되는 시간 임계성 데이터를 처리하기에 비효율적인 문제가 있다 [5, 6]. 따라서 시간에 따른 의존도가 높은 정보를 무선 방송 기법을 통해 전달하고자 할 때, 접근 시간 효율성에 중점을 두고 에너지 효율성을 높일 수 있는 방향으로 연구되어야 하며, 복잡한 XPath 질의 처리를 위하여 속성을 유지하고 있어야 한다.

3. 제안하는 스트림 구조

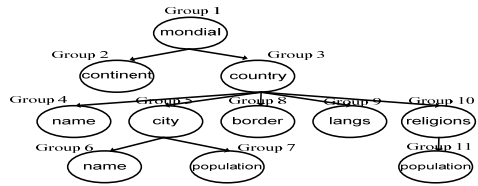
이 장에서는 시간 임계성 데이터를 처리하기 위해 최적화된 XML 데이터 스트림의 구조와 생성 방법, 그리고 색인 방법에 대해 설명한다.

3.1 경로요약

경로 요약 기법은 같은 레이블 경로를 가지는 노드들을 한 노드로 병합함으로써 XML 문서의 크기를 줄일 수 있다. 본 논문에서는 경로 요약 기법을 차용하여 축약된 XML 스트림을 구성한다. SAX 파서를 이용하여 태그 이름, 속성, 그리고 텍스트를 추출하고, 같은 이름을 가진 노드들은 병합하여 “그룹 노드”를 생성한다. 이때, 병합하는 과정에서 같은 레이블 경로를 가진 태그의 이름은 생략되며, 속성과 텍스트는 그룹 노드에 저장된다. 그림 2 는 그림 1 의 원본 XML 문서의 트리 표현(a)과 중복된 노드를 생략하여 생성된 경로 요약 트리(b)를 나타내고 있다.



(a) 원본 XML 문서의 트리 표현



(b) 경로 요약 트리

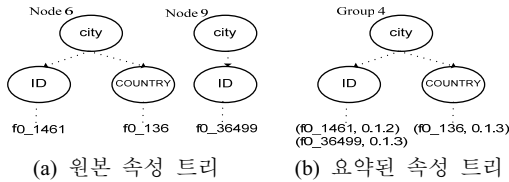
(그림 2) 경로 요약을 이용한 XML 노드들의 그룹화

### 3.2 속성 요약

이 절에서는 스트림의 크기를 최소화하기 위하여 속성을 요약하는 방법과 색인 구조에 대해 설명한다.

XML 에서 속성은 속성 이름과 속성 값의 쌍으로 이루어져 있으며, 하나의 태그가 여러 개의 속성을 가질 수 있다. 이때, 같은 이름을 가진 노드들은 일반적으로 같은 속성 이름을 사용하는 구조적 특성이 있다. 따라서, 같은 속성 이름이 여러 개 반복될 경우, 반복되는 속성 이름을 생략하고 속성 값과 이 속성 값이 해당되는 노드에 대한 색인 정보만을 두어 스트림의 크기를 줄일 수 있다. 이러한 관찰로부터 중복되는 속성 이름은 생략하고, 속성 값과 이 값을 지니는 노드의 듀이 순서(dewey order)를 표시하여 속성이 포함된 XPath 질의를 효과적으로 처리할 수 있는 색인 방법을 제안한다.

**예제 1.** 그림 3 은 원본 XML 문서에서 city 노드의 속성을 나타낸 것(a)와 이를 요약하여 나타낸 것(b)을 보여주고 있다. 중복된 속성 이름을 제거하고, 속성 값과 이 속성 값을 지니는 노드의 듀이 순서로 표시함으로써 스트림의 크기를 줄이고, 속성이 포함된 XPath 질의를 처리할 수 있다. 축약된 속성은 속성 이름과 (속성 값, 듀이순서)의 쌍으로 분리되어 배열형태로 그룹 노드에 저장된다.



(그림 3) XML 노드 속성의 요약

### 3.3 그룹 노드의 구조

경로 요약 및 속성 요약 기법에 의해 생성된 각 그룹 노드는 그림 4 와 같은 구조를 가진다. 이때, Group name 은 이 노드의 태그명이며, Location path 는 이 그룹 노드의 레이블 경로를 나타낸다. Child Addresses 는 이 그룹이 지니는 자식 그룹 노드들의 스트림에서의 주소이다. Attribute name array 는 각 속성의 이름과 이 이름에 해당하는 속성 값들이 지니는 스트림에서의 주소이며, Attribute value array 는 속성 값들과 해당 속성 값을 지니는 노드의 듀이 순서를 의미한다.

Group 4	
city	..... Group name
/mondial/country/city	..... Location path
(name, address), (population, address)	..... Child Addresses
(ID, address), (COUNTRY, address)	..... Attribute name array
{(f0_1461, 0.1.2), (f0_36499, 0.1.3)} {(f0_136, 0.1.3)}	..... Attribute value array
-	..... Text data

(그림 4) 그룹 노드의 구조

### 3.4 스트림 생성 알고리즘

XML 문서의 구조적인 과부하(overhead)를 줄이기 위해 SAX 파서[7]를 이용하여 태그 이름, 애트리뷰트, 그리고 텍스트만 추출하여 그룹 노드로 이루어진 스트림을 생성한다. 스트림의 생성은 2 단계로 이루어진다. 1 단계에서는 노드의 이름과 Location path 를 추출하여 그룹 노드를 생성한다. 2 단계에서는 각 그룹 노드와 Location path 가 같은 노드들의 텍스트 데이터와 속성 값들을 그룹 노드로 병합한다. 이때 속성 값들은 속성 요약된 형태로, 속성 이름은 Attribute name array 에 저장되고, 속성 값과 듀이 순서는 Attribute value array 에 저장된다.

### 4. 스트림에서의 질의 처리 방법

이 장에서는 제안 기법에 의해 생성된 XML 데이터 스트림에서 이동 사용자가 질의 처리하는 방법에 대해 설명한다. 스트림에서의 질의 처리는 질의의 위치 경로(location path)에 따라 이루어진다. 위치 경로란 루트 노드에서부터 마지막 말단 노드까지의 위치 단계(location step)들로 이루어진 일련의 연속된 레이블 명이다.

그림 5 는 이동 사용자가 질의를 처리하는 알고리즘을 보여주고 있다. 질의가 주어지면 이동 사용자는 방송 채널로부터 스트림의 루트 그룹 노드를 수신한다 (5 번째 줄). 루트 그룹 노드에 포함되어 있는 자식 그룹 노드의 주소들로부터 질의의 다음 위치 단계(location step)에 해당하는 자식 그룹 노드의 주소를 가져온다. 이동 사용자는 자식 그룹 노드의 주소를 통해 자신이 수신해야 할 자식 그룹 노드만 선택적으로 수신한다. 이때 만약 질의에 속성 조건(attribute condition)이 포함되어 있으면, 그룹 노드의 Attribute name array 와 Attribute value array 로부터 이 조건을 만족하는 노드의 듀이 순서를 저장한다 (6~14 번째 줄). 질의의 마지막 위치 단계에 해당하는 그룹 노드를 수신하면, 지금까지 저장한 속성을 만족하는 노드의 듀이 순서 정보를 읽어들이어 (15 번째 줄), 이 순서의 자손에 해당하는 노드의 텍스트 데이터만 선별하여 결과를 반환한다 (16~22 번째 줄).

```

01: Algorithm Query_Processing_Over_XML_Stream
02: Input a XML Stream S, Query Q
03: Output result set R satisfying Q
04: begin
05:   Read the root group node from the stream S
06:   while (the last location step of query Q is not appeared)
07:     if (the current location step of Q includes attribute conditions)
08:       Store the matching order satisfying attribute conditions
09:       Get the child group node address satisfying the location step of query Q
10:     else
11:       Get the child group node address satisfying the location step of query Q
12:     endif
13:   Go to the doze mode until the child group node reaches
14: endwhile
15: Load the stored matching orders
16: if (the current location step of Q includes attribute conditions)
17:   Get the text data satisfying both the stored matching orders and the current matching order
18:   Insert the text data into the result set R
19: else
20:   Get the text data satisfying the stored matching orders
21:   Insert the text data into the result set R
22: endif
23: end
    
```

(그림 5) 질의 처리 알고리즘

5. 성능 측정

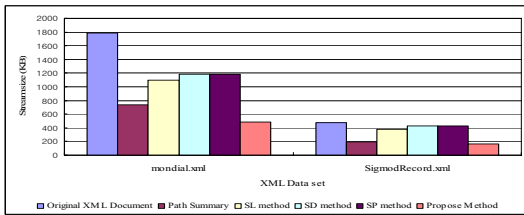
이 장에서는 제안 기법의 성능을 측정하였다. 실험은 mondial 과 SigmodRecord 의 두가지 XML 데이터 집합에 대해 [5]에서 제안한 S-node 방법과 [6]에서 제안한 path summary 방법, 그리고 본 논문에서 제안하는 방법으로 각각 스트림을 생성하여, 스트림의 크기 비교와 XPath 경로 질의를 통한 성능 측정을 하였다. path summary 방법에서는 속성이 모두 생략하였기 때문에 이 실험에서는 정확한 스트림의 크기를 비교하기 위해 각 노드에 포함되어 있는 속성들을 텍스트 형태로 추출하여 path summary 방법에 의해 생성된 스트림에 포함시켰으며, 단순 XPath 경로 질의에 대해서만 접근 시간과 튜닝 시간을 측정하였다. 실험에서 사용한 XPath 경로 질의는 표 1과 같다.

<표 1> 실험에서 사용된 XPaht 경로 질의들

Dataset	Test Query	
mondial (ver. 3.0)	Q1	/mondial/country/city
	Q2	/mondial/country/province[@name="Tyrol"]
	Q3	/mondial/country/province/city/name
Sigmod Record	Q4	/SigmodRecord/issue/articles/article
	Q5	/SigmodRecord/issue/articles/article/title

5.1 스트림 크기 비교

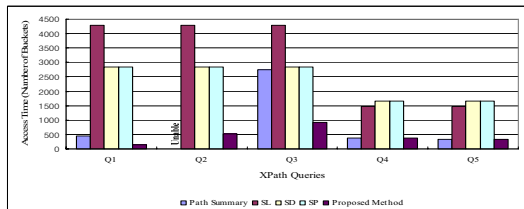
그림 6 은 S-node 와 path summary 와 본 논문에서 제안한 방법에 의해 생성된 스트림의 크기를 보여주고 있다. 실험 결과와 같이 제안 방법은 중복된 태그 이름뿐만 아니라 중복된 속성 이름들도 모두 생략하여 스트림의 크기를 최소화할 수 있었다. 한편 SigmodRecord 문서에서는 속성이 거의 나타나지 않기 때문에 path summary 방법과 본 논문에서 제안하는 방법에 의해 생성된 스트림의 크기가 거의 같다.



(그림 6) 스트림 크기 비교

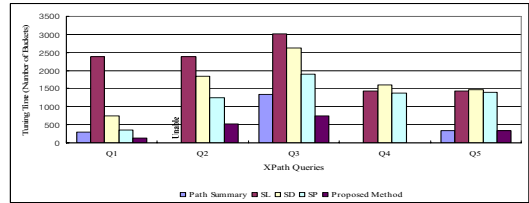
5.2 접근 시간 및 튜닝 시간 측정

그림 7 은 접근 시간을 측정한 결과이다. 제안 기법은 S-node 와 path summary 방법에 비해 가장 작은 크기의 스트림을 생성하기 때문에 이동 사용자의 질의에 대한 접근 시간이 가장 짧았다.



(그림 7) 접근 시간 측정 결과

그림 8 은 튜닝 시간을 측정한 결과이다. 제안 기법은 경로 요약과 속성 요약을 통해 스트림의 크기가 작으며, 자식 그룹 노드에 대한 선택적인 수신으로 인해 튜닝 시간이 향상된 것을 알 수 있다.



(그림 8) 튜닝 시간 측정 결과

6. 결론

본 논문에서는 무선 방송 환경에서 시간에 따른 의존도가 높은 “시간 임계성 데이터”를 처리하기 위해 최적화된 스트리밍 방법을 제안하였다. 일반적으로 XML 문서에서 같은 태그 이름을 가진 노드가 여러 개 나타날 수 있으며 또한 같은 이름을 가진 노드들은 같은 속성 이름을 사용한다는 XML의 구조적 특성을 이용하여 경로 요약 및 속성 요약을 통해 스트림의 크기를 최소화시켰다. 또한 본 논문에서 제안하는 방법은 기존의 경로 요약을 이용한 스트리밍 방법[5]에 비해 속성이 포함된 XPath 경로 질의 처리도 가능한 장점이 있다. 다양한 실험을 통해 본 논문에서 제안하는 방법이 기존의 무선 방송을 위한 XML 스트리밍 방법들에 비해 접근 시간 및 튜닝 시간을 효과적으로 단축시킬 수 있음을 증명하였다. 향후 연구에서는 무선 방송 환경에서 효율적으로 가지 경로 질의 (twig path query)나 분기 질의 (branching query)를 처리할 수 있는 방법에 대한 연구가 고려되어야 할 것이다.

7. 참고 문헌

- [1] S. Acharya, S. Alonso, M. J. Franklin, and S. B. Zdonik. “Broadcast disks: Data management for asymmetric communication” In Proceedings of ACM SIGMOD Conference on Management of Data. pp. 199-210. 1995.
- [2] T. Imielinski, S. Viswanathan, and B. R. Badrinath. “Data on air: Organization and access” IEEE Transactions on Knowledge and Data Engineering. Vol. 9. pp. 353-372. 1997.
- [3] Extensible Markup Language, <http://www.w3.org/XML>.
- [4] E. Y. C. Wong, A. Chan, and H. Leong. “Xstream: A middleware for streaming xml contents over wireless environments” IEEE Transactions on Software Engineering. Vol. 30. pp. 918-935. 2004.
- [5] 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. pp. 223-235. 2005.
- [6] S. H. Park, J. H. Choi, and S. Lee. “An Effective, Efficient XML Data Broadcasting Method in Mobile Wireless Network” In the Proceedings of DEXA Conference. Pp. 358-367. 2006.
- [7] Simple API for XML. <http://www.saxproject.org/>, 2004.