

S-XML 미들웨어에서 캐싱을 이용한 효율적인 공간 데이터 변환 기법

*이동욱, *장용일, *박순영, **오영환, *배해영

*인하대학교 컴퓨터·정보공학과

**나사렛대학교 정보통신학과

e-mail : dwlee@dblabinha.ac.kr

An Efficient Spatial Data Transformation Method Using Caching on S-XML Middleware

*Dong-Wook Lee, *Yong-Il Jang, *Soon-Young Park, **Young-Hwan Oh and *Hae-Young Bae

*Dept. of Computer Science & Information Engineering, Inha University

**School of Information Science, Korea Nazarene University

요 약

본 논문에서는 S-XML 미들웨어에서 캐싱을 이용한 효율적인 공간 데이터 변환 기법을 제안한다. 공간 데이터베이스 관리 시스템(SDBMS)을 기반으로 하여 S-XML 을 지원하는 시스템은 SDBMS 에서 사용되는 이진 데이터와 S-XML 간의 상호 변환 과정이 필요하다. 특히 변환 과정에서 공간 데이터는 비공간 데이터에 비해 데이터 크기가 크고, 복잡한 동시에 다양한 데이터 형식을 가지고 있어 비공간 데이터에 비해 변환 비용이 크다. 따라서 공간 데이터의 특성을 고려한 변환 비용 감소 기법이 필요하다.

제안 기법은 SDBMS 의 질의 처리 결과인 이진 데이터를 S-XML 문서로 변환하고, S-XML 미들웨어의 캐싱 관리자에서 공간 데이터에 해당하는 부분만을 캐싱 한다. 캐싱 관리자는 질의를 점 질의와 영역 질의로 구분한다. 점 질의의 결과에 대해서는 공간 데이터의 OID 를 키 값으로 하는 해쉬 색인을 사용하며, 영역 질의의 결과에 대해서는 질의에서 표현된 MBR 에 따라 R-Tree 색인을 사용한다. 변환 과정에서 요구되는 공간 데이터가 캐싱된 데이터로 존재하면 질의 종류에 따른 색인을 사용하여 데이터를 추출하고 이를 S-XML 데이터로 치환한다. 캐싱 데이터를 이용한 공간 데이터 변환 기법에 의해 공간 데이터의 변환 비용을 줄임으로써 S-XML 미들웨어에서의 변환 시간을 단축할 수 있다.

1. 서론

XML 은 W3C 에서 제안한 문서 형식으로, 현재 비중 있는 데이터 유통의 표준으로 사용되고 있으며, XML 전용 DBMS 의 활용도가 높아짐에 따라 그 중요성을 더하고 있다[1]. 한편 OGC(Open GIS Consortium)에서는 XML 에서 공간 데이터 표현을 위한 GML(Geography Markup Language) 을 제안하였다[2]. GML 은 XML 에서 공간 데이터를 위한 스키마가 확장되어 공간 데이터 표현이 가능하다. 최근에는 GML 을 기반으로 하여 공간 데이터의 활용도를 높이기 위한 G-XML, S-XML(Spatial-eXtensible Markup Language) 과 같은 확장된 형태의 문서들이 제안 되고 있다[3, 4, 5].

S-XML 과 같은 공간 데이터 지원이 가능한 XML 의 출현으로 공간 데이터베이스 관리 시스템(SDBMS: Spatial Database Management System)에서도 S-XML 지원이 요구된다. SDBMS 에서 S-XML 을 처리하기 위해서는 SDBMS 에서 처리된 이진 데이터와 S-XML 데이터 간의 상호 변환과정이

요구된다. 하지만 공간 데이터는 일반적으로 대용량이며, 복잡한 스키마를 가지고 있는 동시에 다양한 데이터 형식을 가지고 있어 비공간 데이터에 비해 많은 변환 비용이 요구된다. 따라서, 공간 데이터의 처리 비용을 줄이기 위한 연구가 필요하다.

본 논문은 ¹⁾ SDBMS 의 S-XML 형식의 데이터를 지원하기 위한 변환 과정 중에서 SDBMS 의 질의 처리 결과 데이터를 사용자를 위한 S-XML 데이터로 변환하는 비용을 줄이기 위한 기법을 제안한다. 결과 S-XML 데이터에서 상대적으로 변환 비용이 큰 공간 데이터만을 캐싱하여 이를 메인 메모리 기반으로 저장, 관리한다. 캐싱된 데이터를 치환하여 S-XML 형식의 데이터로 생성이 가능하기 때문에 S-XML 미들웨어의 처리 성능을 향상시킬 수 있다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련연구를 살펴본다. 3 장에서는 SDBMS 기반의 S-XML 미들웨어 구조를 살펴보고, 4 장은 캐싱 관리자를 이용한 S-XML 변환 과정을 기술한다. 5 장에서는 제안 기법의 성능평가를 서술하고, 6 장에서 결론을 맺고 향후 연구과제를 기술한다.

1) 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성·지원사업의 연구결과로 수행되었음.

2. 관련 연구

본 장에서는 공간 데이터를 지원하는 확장된 XML 과 공간 데이터의 처리 비용을 줄이기 위한 관련 연구를 다룬다.

2.1. 공간 데이터 지원을 위한 XML

OGC 는 XML 에서 공간 데이터를 지원 하기 위한 GML 을 3.0 버전까지 제시하고 있다. GML 은 Polygon, Multipoint, Circle 등의 공간 데이터 타입의 표현을 위한 표준 태그를 정의 하고 있다[2]. S-XML 은 GML 을 기반으로 LBS 를 위한 Voice 스키마를 제공하는 확장된 형태의 공간 데이터 지원 문서이다[4, 5].

2.2. XML 지원을 위한 변환 기법

XML 은 웹에서 표준 데이터 유통 형식으로 사용되고 있으며, 이를 기존의 DBMS 에서 지원하기 위한 연구가 진행되었고, XML 질의 처리를 위한 여러 캐싱 기법도 제시 되었다[6, 7]. 기존의 연구에서는 XML 지원을 위한 시스템 비용을 줄이고 부하 분산을 위해 DBMS 와 연동되는 데이터 서버 또는 응용서버에 XML 질의 결과를 캐싱 하거나, 관계형 데이터베이스의 실제화 뷰를 생성하여 이용하였다[8, 9]. 기존의 연구에서는 변환 과정에서 공간 데이터의 특성과 고비용의 변환 연산을 고려하지 않고 있어, 공간 데이터를 포함하는 S-XML 변환 과정에서 시스템 성능 저하의 요소가 있다.

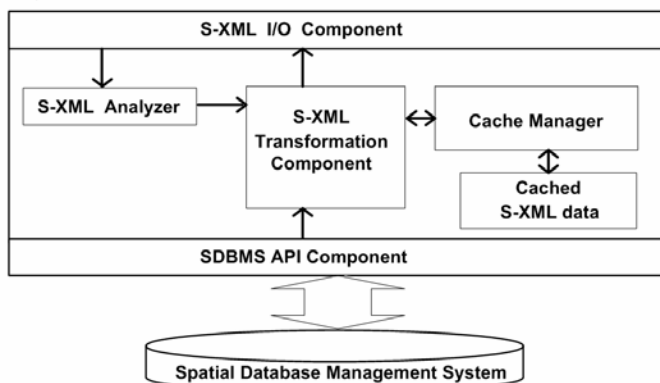
3. 공간 데이터 변환을 위한 S-XML 미들웨어

이 장에서는 SDBMS 에서의 S-XML 지원을 하기 위한 S-XML 미들웨어 구조와 질의 종류에 따른 색인을 설명한다

3.1. S-XML 미들웨어 구조

S-XML 미들웨어는 사용자와 S-XML 형태로 입·출력이 가능하고, SDBMS 에서 질의 처리를 위하여 SQL 형태로의 변환과 SDBMS 로부터 처리된 이진 데이터를 S-XML 로 변환하는 역할을 한다.

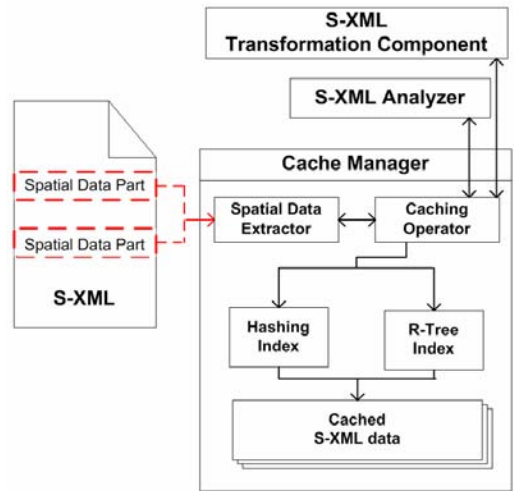
본 논문에서는 S-XML 지원을 위한 변환 구조에서 효율적인 변환을 위하여 S-XML 데이터를 캐싱 관리하여 질의 처리 시 변환 과정에서 사용된다. [그림 1]은 SDBMS 기반의 S-XML 미들웨어 구조이다.



[그림 1] S-XML 미들웨어 구조

Caching Manager 는 변환 컴포넌트의 출력 데이터인 S-XML 의 데이터 중에서 공간 데이터에 해당하는 부분만을 캐싱하여 저장한다. 캐싱 데이터는 S-XML 데이터의 형식을 보존하고 질의 분석기의 질의 종류에 따라 점 기반 질의와

영역 기반 질의로 구분하여 처리한다. 점 질의인 경우 공간 데이터의 OID 값을 키로 하는 해쉬 색인을 사용하여 저장하고, 영역 질의의 경우에는 입력된 질의에 표현된 MBR 을 R-Tree 색인을 이용하여 저장 한다. 캐싱된 데이터는 메인 메모리에 저장되고, 공간데이터의 계층별로 캐싱 데이터를 구축한다. 예를 들어 공간 데이터가 ‘강’, ‘도로’, ‘건물’ 등의 계층 구조로 이루어져 있다면, ‘강’ 공간 데이터를 위해 해싱과 R-Tree 가 사용되고, ‘도로’, ‘건물’ 등도 같은 방법으로 별도의 캐싱 데이터로 저장 된다.. [그림 2]는 캐싱 관리자의 구조를 나타낸다.

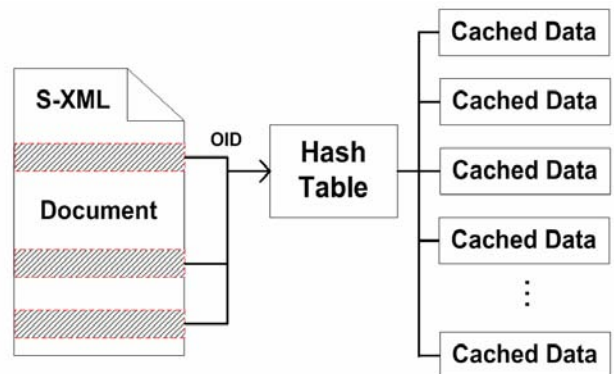


[그림 2] 캐싱 관리자 구조

3.2. 해쉬 색인을 이용한 공간 데이터 캐싱

변환 되는 SDBMS 의 질의 처리 결과가 점 질의에 대한 이진 데이터일 경우 공간 데이터의 OID 를 키 값으로 가지는 해쉬 색인을 이용한다.

점 질의의 결과로 S-XML 문서가 생성되면, S-XML 에서 공간 데이터에 해당하는 모든 부분을 OID 를 기준으로 추출한다. 추출된 데이터는 [그림 3]과 같은 구조를 통하여 저장되며, 변환 컴포넌트의 요청에 따라 해싱으로 검색하여 캐싱된 데이터를 제공한다.



[그림 3] 점 질의를 위한 해쉬 색인 구조

3.3. R-Tree 색인을 이용한 공간 데이터 캐싱

R-tree 색인은 영역 질의에 대한 결과를 갖는 S-XML 데이터를 캐싱 할 때 사용되고, 이를 위해서는 다음의 조건을 만족해야 한다. S-XML 은 공간 데이터를 표현하는 한 계층

