

# 다차원 인덱스를 지원하는 4D 데이터 제공자†

이선준\* 김상호\* 류근호\* 이성호\*\*

충북대학교 데이터베이스 연구실\*

(sunjun\*, shkim\*, khryu\*)@dblab.chungbuk.ac.kr, shlee@etri.re.kr\*\*

## 4D Data Provider Supporting a Multidimensional Index

Sun Jun Lee\*, Sang Ho Kim\*, Keun Ho Ryu\*, Seong Ho Lee\*\*

Database Laboratory, Chungbuk National University\*

Electronics and Telecommunications Research Institute\*\*

### 요약

사용자에게 제공되고 있는 대부분의 시간 지리 정보 시스템은 소프트웨어 형태로 패키징되어 제공되므로 과도한 구축 비용과 재사용성의 문제점을 가지고 있다. 그리고 시간 지리 정보 시스템은 복잡하며 대용량인 특성을 가진 시공간 데이터를 처리해야만 한다. 따라서 시공간 데이터의 효율적인 관리 및 재사용성을 위해서 소프트웨어 부품을 제공하는 컴포넌트 개발 방법론을 사용해야하며 시공간 데이터에 빠르게 접근하기 위한 인덱스 구조가 필요하다.

이 논문에서는 시공간 데이터를 효율적으로 관리하기 위하여 4차원 시공간 객체에 대한 인덱스를 컴포넌트 개발 방법론을 적용하여 설계하였다. 그리고 다차원 인덱스를 이용하는 4D 데이터 제공자를 COM 기반의 OLE DB 인터페이스를 이용하여 구현하고 성능을 평가하였다.

## 1. 서론

시공간데이터베이스는 최근 활발히 진행되고 있는 연구분야 중 하나로서 실세계에 존재하는 여러 가지 공간 객체에 관한 데이터 입력, 저장 및 분석, 처리할 뿐만 아니라 시간에 따른 공간 객체들의 변화와 이력을 효율적으로 관리한다[1]. 시공간 데이터베이스 시스템의 대표적인 응용 분야로서 시간 지리 정보 시스템(Temporal Geographic Information System: TGIS)이 있다. 현재 사용자에게 제공되고 있는 대부분의 시간 지리 정보 시스템은 소프트웨어 형태로 패키징되어 제공되므로 과도한 구축 비용과 재사용성의 문제점, 폐쇄적인 시스템 구조를 가지고 있다. 그리고 시간지리정보시스템은 시공간 데이터를 효율적으로 저장하고 검색해야한다. 이를 위해 적절한 시공간 데이터 모델링, 질의 처리, 시공간 인덱스 구조에 대한 연구가 필요하다.

이 논문에서는 현 시간지리정보시스템들의 문제점을 해결하기 위해 개방형 GIS 서비스 아키텍처를 시간을 지원하도록 확장한 객체모델을 이용하였으며 컴포넌트 개발 방법론을 적용하여 각 서비스들을 컴포넌트 형태로 설계한다. 각 서비스를 컴포넌트 형태로 설계 및 구현함으로써 구축비용 및 다양한 사용자 요구사항을 효율적으로 충족시키며 컴포넌트 소프트웨어의 특징인 재사용성을 충족하므로 향후 유지보수나 기능확장에 많은 이점을 갖는다. 그리고 복잡하고 대용량인 시공간 데이터를 빠르고 효율적으로 다루기 위해서 적절한 인덱스를 구성한다. 효율적인 질의 처리를 위하여 공간데이터는 R/R\*-트리를 이용하고 시공간데이터는 R\*-트리에 시간차원을 추가한 인덱스를 이용한다. 따라서 이 논문은 데이터 속성에 따른 적절한 인덱스를 지원하는 4D 데이터 제공자 시스템을 제안한다.

제안하는 다차원 인덱스를 지원하는 4D데이터 제공자는 개방형 GIS의 2차원 공간모델을 확장한 4차원 시공간 객체 모델을 수용한다. 이 논문을 전개해 나가기 위하여 차원은 3차원공간(x,y,z)과 1차원의 시간(유효)를 다루고 시간단위(간격)은 동일하며 이산적이라고 가정한다.

이 논문의 전체적인 구성은 다음과 같다. 먼저 2장은 관련연구로서 시공간데이터베이스의 개념과 기존의 공간, 시공간 인

덱스에 관하여 소개한다. 3장에서는 이 논문에서 제안한 다차원 인덱스를 지원하는 4D데이터 제공자에 필요한 컴포넌트들을 설계한다. 4장에서는 4D데이터 제공자를 구현하여 실험 및 평가한다. 마지막으로 5장에서는 결론을 맺는다.

## 2. 관련연구

### 2.1 시공간 데이터베이스

시공간데이터베이스는 공간상에 존재하는 객체에 대하여 비공간적인 요소뿐만 아니라 공간적인 요소, 시공간적 요소 그리고 이들에 대한 효율적인 관리와 서비스를 가능하게 해주는 데이터베이스의 연구분야이다. 공간객체에 대한 이력을 지원하는 시공간 데이터베이스의 대표적인 응용 분야 중의 하나는 시간 지리 정보 시스템이다. 시간 지리 정보 시스템은 시공간객체의 시공간 정보 및 시공간객체에 포함된 비공간 데이터를 표현하고 시공간에 따른 자료검색 및 갱신 등을 지원해야하며 시간에 따른 위치변화와 공간객체 상호간의 위상관계, 변화에 대한 이력을 지원해야한다. 많은 이력 자료를 가지고 있는 시간 지리 정보 시스템에서 사용자가 원하는 자료를 질의시 마다 기본 데이터베이스에 대한 질의를 하여 다시 얻는다는 많은 비용을 요구한다. 그러므로 사용자가 요구하는 자료를 찾을 때 비용을 줄일 수 있는 방법이 필요하다. 첫번째로 공간, 시간, 시공간 객체에 따른 적절한 데이터 모델 설계이다. 데이터베이스에서 중요하고 또한 기본적인 기능은 데이터의 저장과 저장된 데이터의 조회이다. 따라서 데이터 모델 설계시 대상 데이터의 표현 방법과 데이터를 효율적으로 조회할 수 있는 연산자가 고려되어야한다. 두 번째로 객체에 빠른 접근을 위한 인덱스가 필요하다. 시공간 데이터는 시간의 흐름에 따라 변화하는 데이터로서 매우 복잡하고 대용량이라는 특성을 가진다. 그러므로 다차원이며 시간속성을 효율적으로 다룰 수 있는 다차원 인덱스나 시공간 인덱스가 필요하다. 이 논문에서는 시공간 데이터 모델링과 질의 처리의 속도를 높이기 위한 접근구조의 최적화에 초점을 맞춘다.

### 2.2 다차원 인덱스

R-트리[2], R\*-트리[3]는 B-트리와 K-D-B트리를 합쳐놓은 개념과 유사하다. B-트리와 유사하게 인덱스 레코드들로 구성된 높이 균형트리로서 단말노드들에 데이터 객체에 대한 포

†이 연구는 2002년 ETRI의 "다차원 공간 Indexer와 Simple API 개발"에 관한 연구과제로 수행된 결과임

인터를 포함하고 있다.

2.2.1 R-트리

공간에 존재하는 객체를 최소경계사각형(MBR)으로 표현하여 사각형들간의 포함과 겹침관계로서 인덱스를 구성하여 공간적 위치에 따라 신속하게 데이터를 검색하는 인덱스이다. R-트리의 노드 형태는 다음과 같다. 비단말노드는 (I, child-pointer) 형태의 개체들을 포함한다. 여기서 child-pointer는 자식 노드의 주소이며 I는 자식 노드의 개체들의 사각형들을 모두 포함하는 최소경계사각형이다. 데이터베이스의 실제 객체들을 나타내는 단말노드는 (I, tuple-identifier) 형태의 개체들을 포함한다. 여기서 tuple-identifier는 데이터베이스에서 실제 튜플들을 나타내며 I는 공간객체의 인덱스된 최소경계사각형이다.

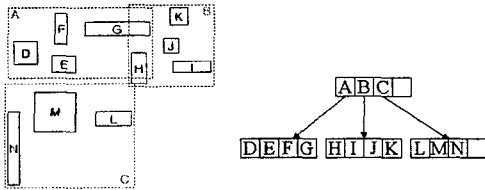


그림 1. R-트리

균형트리를 유지하고 공간 근접성 위주로 트리를 형성한다. 하지만 중간노드의 MBR들 사이의 겹침문제로 탐색성능이 저하된다.

2.2.2 R\*-트리

R-트리는 경험적인 최적화에 기초를 두고 있으며 최적화 기준으로 사각형의 영역 최소화만을 고려하였다. R\*-트리는 영역, 겹침, 둘레의 합등의 여러 매개변수를 이용하여 R-트리를 최적화하였다. R\*-트리는 R-트리와 동일한 인덱스 구조를 가진다. 겹침은 MBR이 여전히 존재하지만 최소화 시켰다. 노드의 분할시 강제적인 재삽입을 이용한 최적의 분할 방법을 이용하여 겹침을 최소화하고 저장 공간도 효율적으로 사용하였다.

2.3 데이터 제공자

다양한 GIS 데이터 소스에 대한 단일한 접근 방법을 제공하기 위해서는 데이터 제공자 컴포넌트가 필요하다.

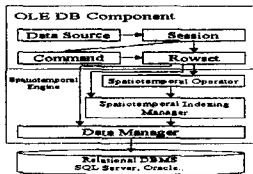


그림 2. 데이터 제공자 컴포넌트 프로토타입

데이터 제공자의 역할은 공간 데이터 소스에 관한 메타정보를 제공하고 공간 데이터에 대한 공간 검색 방법을 지원한다. 메타정보로서 공간 정보를 가지고 있는 피쳐데이터의 정보, 피쳐 데이터들의 최대 맵 영역에 관한 정보, 공간 좌표계에 관한 정보를 나타낸다. 시공간 검색은 시공간 색인 방법을 이용하여 후보객체들을 검색하는 필터링단계와 후보 객체들에 대한 실제 검색연산인 정제단계로 나눈다.

3. 다차원 인덱스를 지원하는 4D 데이터 제공자

개방형 GIS는 2차원 공간 객체만을 다루고 있다. 그래서 3차원 객체 및 시공간 객체 모델을 표현하기 위해 개방형 GIS 명세서의 모델을 확장하는 방식을 선택하였다[4]. 그리고 복잡하며 대용량인 시공간 데이터를 빠르게 접근하기 위한 인덱스 구성이 필요하다. 인덱스 구성은 R\*-트리를 이용하였다.

4차원 객체 모델은 표준 모델을 기반으로 사용함으로써 각 시스템간 데이터의 확장성과 이식성을 보장한다. 그리고 컴포넌트 형태로 구현함으로써 시스템의 유지 보수 및 구축 비용 절감의 이점을 얻을 수 있다.

공간데이터의 저장 형태는 WKB(Well Known Binary)구조를 사용한다. WKB의 구조는 개방형 모델에서 제안한 WKB를 4차원 객체 모델의 인스턴스 객체에 맞도록 확장하여 설계하였다. WKB는 일반적인 데이터 타입의 리스트로 구성되므로 시스템간의 데이터 이동시 타입 호환성 및 이식성을 보장한다[5].

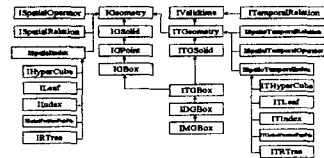


그림 3. 시공간 객체 컴포넌트 모델 구조

시공간 데이터를 컴포넌트 모델로 제공하는 데이터 제공자는 그림 3에서와 같은 구조로 되어 있다.

ISpatioTemporalRelation 인터페이스는 시공간 데이터들간의 위상 관계를 식별할 수 있는 멤버 함수들을 제공한다. ISpatioTemporalOperator 컴포넌트는 시공간 객체의 기하 관계 연산들을 제공한다. 시공간에 대한 객체는 IGeometry 컴포넌트에서 관리한다. IGeometry 컴포넌트에서 공간에 대한 관리를 담당한다. ISpatialIndex 컴포넌트와 ISpatioTemporalIndex 컴포넌트는 각각 공간과 시공간 객체에 대한 인덱스 구성을 담당한다.

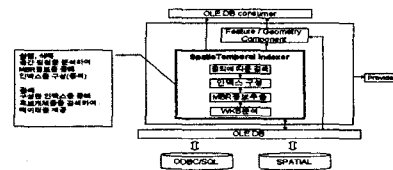


그림 4. 인덱스 컴포넌트 모델

그림 4에서 4D 데이터제공자의 인덱스 컴포넌트 부분은 다음과 같이 실행된다. 공간 정보를 나타내는 WKB를 분석하여 객체의 MBR정보를 추출한 후 MBR정보를 통해 R-트리를 구성한다. 인덱스 구성은 질의시 동적으로 구성되며 검색시에는 구성된 인덱스를 통해 후보객체를 검색하여 데이터를 제공한다.

4. 전체적인 시스템 구조

4.1 R\*-트리의 구현

공간 데이터에 대한 일반적인 질의로는 "어느 특정 지역에 속한 모든 객체를 탐색하라", "특정 지점을 포함하는 객체를 찾아라"등이 될 수 있다. 이러한 공간데이터를 효율적으로 다루기 위해서는 자료 항목들 빠르게 검색하기 위한 인덱스 구조가 필요하다. 이러한 공간 검색에 적합한 접근구조로 R-트리, R\*-트리가 있다. R-트리와 R\*-트리를 구현한 클래스에 대한 설명을 보인다.

클래스는 크게 다음과 같이 9개로 나누었다. 즉, HyperCube, Node, AbstractNode, Leaf, Index, PageFile, PersistentPageFile, CachedPersistentPageFile, RTree 등으로 구성하였다.

HyperCube 클래스는 차원에 따른 MBR에 대한 정보와 MBR간의 기하, 위상연산에 대한 함수를 포함하고 있다. 그림

5는 차원에 따른 사각형간의 포함관계를 연산하는 contains 함수를 나타낸다. AbstractNode 클래스는 Index클래스와 Leaf클래스의 부모클래스이다. 노드의 정보인 레벨, 페이지 번호, MBR정보, 포인터(실제 객체 ID)정보 등을 가지고 있다. Leaf 클래스는 실제 객체에 대한 정보를 포함하고 있는 단말노드 클래스이다. Index 클래스는 인덱스 역할을 하는 비단말노드 클래스이다.

```
bool HyperCube::contains(const HyperCube& h)
{
    assert(dimension == h.dimension);
    for (int i = 0; i < dimension; i++)
    {
        if (low[i] > h.low[i] || high[i] < h.high[i])
        {
            return false;
        }
    }
    return true;
}
```

그림 5. MBR간의 포함 연산

PageFile 클래스는 트리의 구성을 파일로 유지하기 위한 최상위 클래스이다. PersistentPageFile 클래스는 PageFile 클래스를 상속 받는다. 파일포인터와 페이지에 대한 정보를 유지한다. CachedPesisenPageFile 클래스는 파일정보를 유지하기 위한 최하위 클래스이다. RTree 클래스는 전체적인 트리 구성 정보와 삽입, 삭제, 검색 함수를 유지하는 클래스이다.

4.2 데이터 제공자의 구현

OLE DB는 다양한 데이터 자원을 위한 Microsoft의 시스템 레벨 프로그래밍 인터페이스이다. OLE DB는 다양한 데이터베이스 관리 시스템 서비스를 포괄하는 COM 인터페이스 단위를 지정한다.

데이터 제공자는 SQL 데이터베이스, 문서, 공간데이터와 같은 다양한 데이터를 대표하는 컨포넌트이다. 제공자는 로우셋이라고 불리는 객체를 사용하여 공통적으로 데이터를 제공한다.

데이터소스 객체는 특정한 데이터 제공자를 확인하는 기능을 포함하고 사용자가 제공자에 연결할 수 있도록 적절한 허가를 받았는지를 확인하며 특정한 데이터 자원에 대한 연결을 초기화한다. 세션 객체는 연결 내부에서 이루어지는 트랜잭션의 범위를 정의한다. 커맨드 객체는 소비자로부터 하여금 데이터 정의 또는 질의와 같은 데이터 조작 명령을 실행할 수 있도록 한다.

로우셋 객체는 데이터의 공통 표현 방법을 제공한다. 이 객체들은 대개 세션으로부터 직접 생성되거나 명령 실행의 결과로 생성되지만, 데이터 또는 메타데이터를 회산하는 기타 메소드의 결과로 생성될 수도 있다.

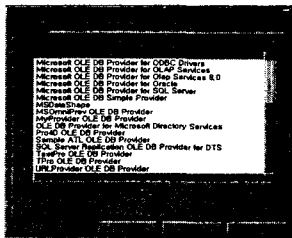


그림 6. 4D 데이터 제공자 컴포넌트 연결

그림 6은 4차원 데이터 제공자 컴포넌트에 연결하는 화면이다. OLE DB 데이터 제공자는 컴포넌트 형태로 존재하며 클라이언트에 등록된다. WorldProv4DV는 OLE DB인터페이스를 지원하는 4D 데이터 제공자 컴포넌트이다.

4D 데이터 제공자는 명령어를 수행하기 전에 매개변수(SPATIOTEMPORAL\_FILTER, SPATIOTEMPORAL\_OPERATOR)를 설정함으로써 시공간 필터링을 수행한다. 질의에 임의의 매개변수가 설정되면 시공간 매개변수는 임의의 매개변수 후에 처리된다.

시공간 매개변수의 순서는 SPATIOTEMPORAL\_FILTER, SPATIOTEMPORAL\_OPERATOR의 순이다.

그림 7은 시공간 질의를 처리한 후 4D 데이터 제공자에서 제공되는 4차원의 공간 인덱스를 통해 질의에 해당하는 객체를 찾은 후 이를 사용자에게 데이터를 제공하는 화면이다.

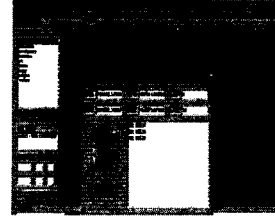


그림 7. 인덱스를 이용한 3차원 객체의 검색화면

4.3 평가

기존의 시간 지리 정보 시스템은 소프트웨어 형태로 패키지가 되어 제공되므로 과도한 구축비용과 낮은 재사용성을 가지고 있었다. 그리고 인덱스 구성없이 기하알고리즘을 통해 모든 실제객체들을 비교하여 질의에 맞는 객체들을 검색하여 비효율적이었다. 하지만 이 논문의 4D 데이터 제공자는 컴포넌트 형태로 설계 및 구현되어 확장성과 재사용성을 높였고 4차원 데이터에 적합한 R\*-트리 인덱스를 데이터 제공자에 통합하였다. 인덱스를 지원하는 데이터 제공자를 통해 클라이언트들은 시공간데이터를 빠르게 접근할 수 있었다. 또한 개방형 GIS 표준 사양을 반영하여 상호 운용성을 보장하였다. 향후에 데이터양이 많은 시공간 데이터 집합에 대한 성능평가를 수행할 것이다.

5. 결론

시간 지리 정보 시스템은 시공간 데이터를 다룬다. 시공간 데이터는 복잡하며 대용량인 특성이 있다. 그러므로 시공간 데이터에 빠르게 접근하기 위해서는 인덱스 구성이 필요하다. 또한 재사용성과 확장성을 위하여 컴포넌트 개발 방법을 사용하여야 한다.

이 논문에서 제안한 4D 데이터 제공자는 효율적인 검색연산을 위하여 공간 속성에는 R/R\*-트리를 이용하였으며 시공간 속성에는 R\*-트리에 시간 차원을 추가한 인덱스를 이용하였다. 이러한 속성에 따른 적절한 인덱스 구성으로 4D 데이터 제공자의 성능을 향상시켰고 각 서비스를 COM 기반으로 구현하여 재사용성, 확장성을 높였다.

향후 연구로서 많은 변화를 가진 시공간 객체에 적합한 시공간 인덱스의 설계 및 컴포넌트 구현에 대한 연구가 필요하다.

6. 참고문헌

[1] 강병국, 시공간 데이터베이스에서의 시간 버전 모델, 박사학위논문, 2002년2월  
 [2] Guttman, A., R-trees: A dynamic index structure for spatial searching, In Proc. ACM SIGMOD Int. Conf. on Management of Data, 1984, pp.47-54.  
 [3] Beckmann, N., H.-P. Kriegel, R. Schneider, and B. Seeger, The R\*-Tree : An efficient and robust access method for points and rectangles, In Proc. ACM SIGMOD Int. Conf. on Management of Data, 1990, pp.322-331.  
 [4] 강규, 컴포넌트 기반의 4차원 시공간 위상 관계 연산자의 설계 및 구현, 석사학위논문, 2002년2월  
 [5] Open GIS Consortium, Inc. OpenGIS, "Simple Features Specification For OLE/COM Revision 1.1," OpenGIS Project Document, 99-050, 1999.  
 [6] Y.Theodoridis, M.Vazirgiannis, T.Sellis, Spatio-Temporal Indexing for Large Multimedia Applications, Submitted to the 3rd IEEE International Conference on Multimedia Computing and Systems(ICMCS 1996), pp.1-23.