

# 지리정보시스템을 위한 주기억 데이터베이스의 색인 구성에 대한 성능평가

신수미<sup>0</sup>, 편도영<sup>\*</sup>, 김경창<sup>\*</sup>, 김영일<sup>\*\*</sup>

<sup>0</sup>홍익대학교 컴퓨터공학과, <sup>\*</sup>한국과학기술정보연구원 정보서비스실  
{smshin<sup>0</sup>, dypyun, kckim}@cs.hongik.ac.kr, mykim@kisti.re.kr

## Performance Evaluation of Indices based on Main Memory DBMS for GIS

Sumi Shin<sup>0</sup>, Doyoung Pyun<sup>\*</sup>, Kyung Chang Kim<sup>\*</sup>, Myung Il Kim<sup>\*\*</sup>

<sup>\*</sup>Dept. of Computer Engineering, Hongik University

<sup>\*\*</sup>Dept. of Information Marketing, Korea Institute of Science and Technology Information

### 요 약

지리정보시스템에 대한 응용분야가 확대되면서 지리정보시스템의 기반이 되는 데이터베이스의 성능에 대한 정확한 평가가 중요하게 인식된다. 이때 지리정보시스템의 기반이 되는 데이터베이스는 일반적인 데이터베이스와는 달리 공간 질의와 비공간 질의가 동시에 처리될 수 있어야 하므로 이를 위한 효율적인 색인 구성이 요구되며 이에 대한 성능의 평가가 특별히 중요하다. 본 논문에서는 주기억 데이터베이스 기반의 지리정보시스템에 적합하도록 비공간 색인과 공간 색인을 별도로 두는 이중 색인 구성을 제안하고 실제 색인이 지리정보시스템에 적용되었을 때 그에 이 시스템에 대한 성능을 평가하였다. 실험을 통해 색인에 따른 데이터베이스의 성능비교도 함께 측정하였다. 지리정보시스템을 위한 주기억 데이터베이스에 T-tree와 MR-tree가 비공간 및 공간색인을 위해 적용되었을 때 데이터가 증가하여도 질의에 대한 속도가 거의 변화가 없는 우수한 성능을 보여주는 것을 확인할 수 있었다.

### 1. 서 론

지리정보시스템은 지능형 교통시스템, 위치기반 서비스 등과 같은 새로운 공간 응용 분야의 확대에 따라 트랜잭션의 빠른 처리가 필요한 고성능 데이터베이스의 주요 분야로 인식되고 있으며 이러한 실시간 응용분야에서는 데이터베이스의 저장 매체를 디스크가 아닌 주기억 장치(Main Memory)로 대체하는 주기억 데이터베이스의 활용이 확대되고 있다. 이때 주기억 데이터베이스는 그 특성에 적합한 색인이 디스크 기반 색인과 별도로 제시되어 있는데 이들이 실제 지리정보시스템과 같은 응용분야에서 얼마나 효율적인지 응용분야의 활용을 고려한 성능요소의 설계와 평가가 필요하다. 아울러 비공간 질의와 공간질의가 함께 수행되는 지리정보시스템에 이들 색인을 함께 적용하여 어떻게 구성하는 것이 빠른 질의처리에 효율적인지에 대한 평가도 필요하다.

본 논문은 지리정보시스템을 위한 주기억 데이터베이스에 비공간 및 공간 색인을 이중으로 적용한 사례를 소개하고 이 시스템에 몇 가지 색인을 바꾸어 적용하여 공간 및 비공간 질의에 대한 데이터베이스의 성능을 각각 평가한다.

본 논문의 구성은 다음과 같다.

2장에서는 관련연구와 이중 색인 적용방안을 소개하며, 3장에서는 주기억 데이터베이스에 여러 색인이 적용되었을 때의 성능평가 결과와 지리정보시스템에 공간, 비공간 색인이 함께 적용되었을 때의 성능평가 결과를 기술한다. 결론 및 향후 연구는 4장에서 제시한다.

### 2. 비공간 색인과 공간색인

기존의 색인에는 크게 주기억장치 기반의 색인과 디스크 기반 색인이 있다. 널리 알려진 주기억장치 기반의 색인으로 비공간 데이터를 위한 T-tree가 있고 디스크 기반 색인으로는 비공간 데이터를 위한 B-tree와 공간 데이터를 위한 R-tree가 있다.

#### 2.1 비공간데이터 색인 기법

B-tree는 자료에 접근하는 성능을 향상시키기 위해서 대체로 디스크 기반 데이터베이스 시스템에서 사용하는 탐색 트리다. 일반적으로 노드의 크기는 보조 기억장치의 블록 크기와 일치하도록 최적화된다. B-tree의 모든 형태들이 평형을 이루고 대체로 큰 분기 계수를 갖는다.

T-tree는 B-tree와 AVL-tree로부터 발전된 메인 메모리 기반의 색인으로 이진 검색과 높이 균형을 가지는 성질이 있으며 한 노드에 여러 개의 데이터를 갖는다. T-tree는 빠른 처리속도와 메모리 사용최적화라는 메인 메모리의 특성에 적합한 구조로 알려져 있다.

#### 2.2 공간데이터 색인 기법

R-Tree는 객체를 최소 직사각형(MBR : Minimum Bounding Rectangle)으로 표현하며 B 트리에 대한 k 차원 확장 모델에 해당된다.[1] 다차원 데이터에 대하여 높은 성능을 나타내며 디스크에 적합한 색인으로 알려져 있다. 그러나 R-tree는 메인 메모리 데이터베이스에서는 최적의 성능을 보장할 수 없다.

MR-tree는 메인 메모리에 적합하도록 기존 R-tree와

T-tree의 특성을 결합시킨 색인으로서 R-tree와 달리 이진검색이 내부노드에서 완료될 수 있기 때문에 빠른 검색이 가능하다.[3]

### 2.3 지리정보시스템에서의 이중 색인 기법 적용

지리정보시스템은 비공간 데이터 질의와 공간 데이터의 질의가 함께 수행되는데 이때 비공간 데이터를 위한 색인과 공간 데이터를 위한 색인을 별도로 두어 각 질의를 위한 검색에 효율적이게 하는 이중 색인을 둔다. 지리정보시스템 내부에서 이들 색인은 인덱스 관리자에 의해 연동되어 각 질의에 대한 연산을 수행하게 된다.

객체명에 대한 비공간 데이터는 T-트리나 B-트리 등 비공간 색인을 이용하여 검색하고 공간 검색은 레코드 내의 MBR정보를 이용하여 MR-트리나 R-트리 색인을 통해 레코드를 검색하는 방식으로 서로 연동하여 동작한다.

## 3. 성능평가

비공간 색인과 공간색인이 적용된 데이터베이스의 성능평가를 위해서 각각 별도의 평가요소를 적용하였는데 비공간 색인에 대한 평가요소는 Wisconsin 벤치마크를 활용하였으며 공간색인과 지리정보시스템에 대한 평가요소는 각 특성을 고려하여 평가요소를 직접 설계, 적용하였다. 본 논문에서 평가하는 지리정보시스템은 읽기 위주의 응용이 많은 시스템으로서 데이터의 변경이 자주 발생하지 않기 때문에 주기적인 변경작업은 다소의 시간을 허용할 수 있다는 것을 전제로 하고 설계된 평가요소 중 실제 성능의 평가는 selection 질의 대해서만 시행하였다. 실험은 윈도우 2000 서버 운영체제와 256MB의 메모리, CPU Pentium IV 2.0Ghz 의 하드웨어 환경에서 시행하였다.

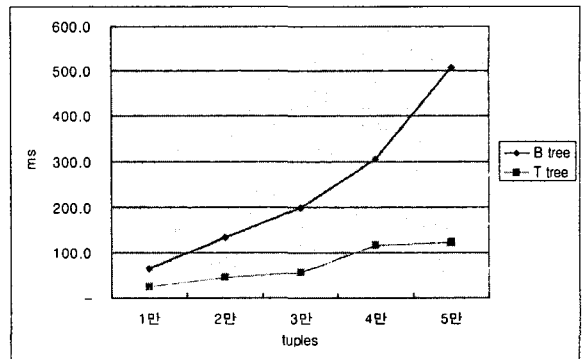
### 3.1 비공간 질의에 대한 성능평가

#### 3.1.1 실험 세부사항

각 테스트는 하나의 Relation에 대한 Tuple수를 10,000개에서 50,000개로 늘이며 시행하였고 색인을 다르게 하였을 때 데이터베이스의 성능 평가를 위하여 B-tree와 T-tree를 각각 적용하여 검색속도를 비교하였다.

#### 3.1.2 실험 결과

[그림1]은 전체 데이터의 10% selection 질의를 T-tree와 B-tree가 적용된 column에 대해 인덱스와 nonclustered 인덱스에 대해 실행한 결과로 T-tree가 적용된 데이터베이스의 질의수행시간이 30ms이하로 보여진다. B-tree가 적용된 데이터베이스의 경우 튜플수가 늘어날 수록 T-tree가 적용된 데이터베이스와 속도차이가 점차 증가하였다.



[그림1] 비공간 질의의 성능 비교(B-tree vs. T-tree)

이를 통해 실제 주기의 데이터베이스에 인덱스로 간단한 코드경로(code path)를 제공하고 메모리 주소를 직접 포인팅 하는 T tree 가 적용된 주기의 데이터베이스의 성능이 우수하다는 것을 알 수 있다.

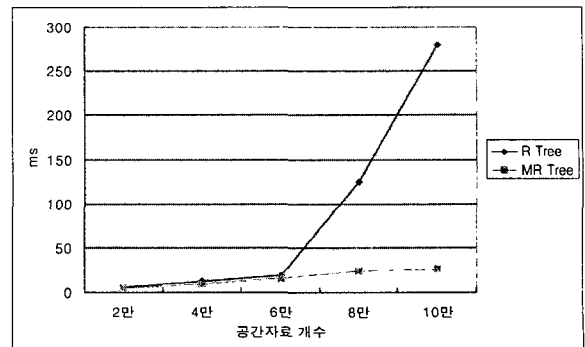
### 3.2 공간 질의에 대한 성능평가

#### 3.2.1 공간 색인의 벤치마크 설계

공간 색인의 성능에 영향을 주는 요인은 공간자료의 형태나 자료의 크기, 분포, 밀도 복잡도 등이 데이터에 대한 부분과 넓이, 둘레, 부피, 최소거리 등에 대한 질의 부분을 함께 고려하여 설계하였다.

#### 3.2.2 실험 세부사항

공간 색인의 성능 평가를 위해서는 한지점에서 전체 영역의 일정 범위에 있는 정사각형의 공간에 있는 데이터를 검색하는 질의를 시행하였다. 실험은 공간데이터의 개수를 2만개에서 10만개로 증가시키면서 시행하였으며 색인을 다르게 하였을 때 데이터베이스의 성능 평가를 위하여 R-tree와 MR-tree를 각각 적용하여 검색속도를 비교하였다.



[그림2] 공간질의의 성능 비교(R-tree vs. MR-tree)

3.2.3 실험 결과

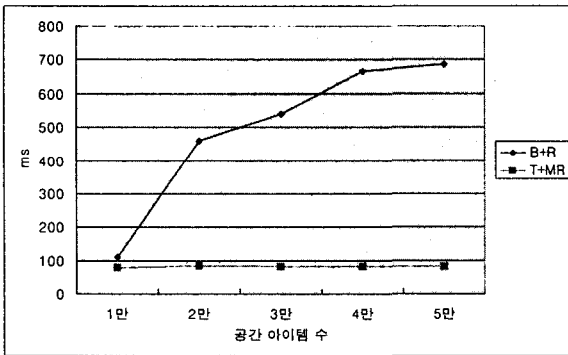
[그림2]은 전체 데이터의 2%에 대한 selection 질의를 실행한 결과로 튜플수가 2만에서 6만까지 R-tree와 MR Tree의 속도가 거의 차이가 없으나 튜플수가 10만으로 증가하면 10배 이상의 속도차이를 보였다.

3.3 지리정보시스템의 성능평가

3.3.1 실험 세부사항

지리정보시스템의 성능평가를 위해서 비공간 질의와 공간질의가 함께 사용된 질의를 설계하였다. 지리정보시스템은 비공간 색인으로 B-tree와 T-tree를, 공간 색인을 위해 R-tree와 MR-tree를 각각 적용하여 성능평가하였다. 테스트는 서울시 데이터 일부를 활용하여 시행하였으며 공간 객체를 10,000개부터 50,000개로 늘이며 각 색인 구성별로 시행하였다.

3.3.2 실험 결과



[그림3] GIS 성능평가 (B+R vs. T+MR)

[그림3]은 전체 데이터 중 한 개의 객체명을 비공간 색인을 이용해 selection한 후 객체 주위 1Km 범위의 공간 데이터를 검색한 결과이다. T-tree와 MR-tree가 색인으로 구성된 경우의 지리정보시스템은 데이터의 수가 증가하여도 검색속도의 변화가 거의 없다. 그에 반해 B tree와 R-tree가 색인으로 적용된 지리정보시스템은 데이터가 데이터가 증가하면서 검색속도가 급격히 증가하였다.

4. 결론

지리정보시스템은 빠른 검색속도가 무엇보다 중요한데 이는 지리정보시스템의 기반 데이터베이스의 성능에 영향을 받는다. 지리정보시스템의 기반 데이터베이스는 일반적인 데이터베이스와는 달리 공간질의와 비공간 질의가 동시에 효율적으로 처리될 수 있어야 하므로 이를 위한 효율적인 색인 구성이 필요하며 특히 주기억 데이터베이스는 전통적인 디스크 기반의 색인과 다르게 제시된

색인이 실제 응용분야에서 얼마나 효율적인지 비교 평가해 볼 수 있는 실험을 시행하였다.

지리정보시스템의 정확한 성능평가를 위해서는 비공간 질의에 대한 성능과 공간질의에 대한 성능 평가가 모두 필요하므로 본 논문은 이 두가지 질의에 대한 개별평가와 연계된 함께 연계된 질의를 모두 평가하였다. 이를 통해 T-tree와 MR-tree가 함께 색인으로 구성된 경우의 지리정보시스템은 데이터의 수가 증가하여도 검색속도의 변화가 거의 없는 것을 볼 수 있었고 데이터의 수가 방대하고 실시간 서비스가 보장되어야 하는 지리정보시스템에서 이와 같은 색인 기반의 주기억 데이터베이스의 성능이 우수할 수 있다는 것을 확인하였다.

본 논문에서 시행한 실험이외에 데이터의 변경 및 삽입에 대한 성능평가는 추후과제로 제안한다.

[참고문헌]

- [1] Guttman, A. "R-tree: A dynamic index structure for spatial searching", In Proceedings of the ACM SIGMOD International conference on Management of data, 1984
- [2] Tobin J. Lehman, Michael J. Carey "A Study of Index Structures for Main Memory Database Management Systems", in Proceedings 12th Int'l. conf. on Very Large Databases, kyoto, Aug. 1986, pp294-303
- [3] 강은호, 김경창 "주기억 데이터베이스에서 공간 데이터에 대한 효율적인 인덱스 구조", 한국정보과학회학술발표논문집, 2003권 2483호, 794-796쪽, 2003
- [4] j. Gray, "The Benchmark Handbook for database and transaction", 3rd, Morgan Kaufmann, 1993