

순차적 레코드 저장구조를 갖는 팜 PDA에서 공간데이터 표현 방법

이 재 호, 임 덕 성, 홍 봉 희

부산대학교 컴퓨터공학과,

Representation of Spatial Data in Palm PDA with sequential records

Jai-Ho Lee, Duk-Sung Lim, Bong-Hee Hong

Dept. of Computer Engineering, Pusan National University

요 약

공간데이터는 양이 크고 복잡한 공간 연산을 필요로 한다. 따라서 웹, PC 또는 서버환경에서 사용하는 공간 데이터를 제약물 가진 PDA 환경에서 사용 시 비효율적이다. PDA에서 공간데이터를 효율적으로 처리 및 검색하기 위해서는 데이터 양을 감소시키고 공간 색인 기법을 적용한 저장 구조로 데이터를 변환하는 것이 필요하다.

이 논문에서는 직접 접근이 가능한 순차적인 가변길이 레코드 단위의 저장 구조를 가진 팜 PDA에서 고정 그리드 파일을 색인 구조를 적용한 공간 데이터의 표현 방법을 제시하고 데이터 저장 구조 및 영역 질의, 객체 삽입, 객체 삭제에 대한 처리 방법과 이동성을 고려하여 부가적인 데이터를 수신했을 때 처리 방법을 제시한다.

1. 서론

최근 PDA(Personal Digital Assistance)와 같은 이동단말기를 이용해서 다양한 정보를 취득하고자 하는 요구가 증가되고 있다. 특히 PDA에서 전자지도, 교통정보, GPS를 이용한 위치 추적이나 항법 장치 등의 다양한 벡터 지도 서비스에 대한 활용이 요구되고 있고 관련 연구도 활발하다.

PDA는 현재 많은 종류의 제품이 나와 있지만 이 논문에서는 현재 세계적으로 가장 널리 사용되고 있는 팜 PDA를 대상 환경으로 한다. 작은 화면 크기와 느린 텍스트 입력, 적은 저장 공간, 제한적인 전지 용량, 느린 CPU는 PDA와 같은 이동 단말기의 큰 제약이다. 또한 무선 통신기능이 가능한 기기의 경우 낮은 대역폭으로 속도가 느리고 통신 비용이 크다.

이러한 환경을 가진 PDA에서 현재 웹이나 PC, 서버에서 사용하는 공간데이터를 이용 시 문제점이 있다. 왜냐하면 공간 데이터는 일반적으로 데이터 양이 크고 공간 객체들간의 공간 관련성을 가지며 질의 처리를 위해 복잡한 공간 연산을 수행해야 한다. 따라서 이러한 환경적 제약과 공간데이터의 성질을 고려하여 공간 데이터를 효율적으로 표현하기 위한 방법이 요구된다.

특히 팜 PDA는 직접 접근이 가능한 가변길이의 순차적인 레코드 단위의 저장구조를 가진다. 순차적인 저장구조이기 때문에 레코드 수가 많아질수록 검색속도가 증가하고, 접근 순서가 레코드의 정렬순서와 동일하지 않을 경우 검색속도가 증가한다. 또한 화면 해상도가 낮기 때문에 복잡한 공간 객체를 정확히 표현하기 어렵다. 그리고 한 레코드의 최대크기가 64KB로 제한적이다.

이 논문에서는 위와 같은 문제점을 가진 팜 PDA에서 공간데이터를 저장하는 구조, 기본적인 공간데이터에 대한 처리방법, 무선통신이나 적외선통신 등을 통한 부가적인 데이터 수신 시 처리방법을 제시함으로써 레코드 단위의 저장구조를 가진 PDA를 위한 공간데이터의 표현 방법을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 관련 연구에 대해 언급하고 3장에서는 PDA에서 공간데이터를 표현하기 위한 방법을 기술한다. 4장에서는 실제적으로 고정 그리드 색인을 적용한 공간데이터의 표현 방법에 대해 언급한다. 마지막으로 5장에서는 결론 및 향후 연구 과제에 대하여 기술한다.

2. 관련 연구

현재 PDA도 데이터 전송 방법 중 가장 널리 사용되는 방법은 PC에서 유선 케이블을 통해 데이터를 전송하는 Sync를 통하는 경우이다. 팜 PDA는 공간 데이터를 수송할 수 있는 저장 공간이 적고 저장 구조가 데이터베이스 형태의 직접 접근이 가능한 순차적인 가변길이 레코드 단위의 저장 구조이다[1]. 그리고 느린 CPU때문에 복잡한 공간연산과 같은 공간데이터를 처리하는데 사용자 응답시간이 증가하게 된다. 뿐만 아니라 복잡하고 많은 연산을 함에 따라 에너지 소모가 많아진다[2].

공간데이터에 대한 효율적 처리를 위한 공간 색인 기법으로는 그리드 파일, R-트리, R*-트리, R+-트리 등등 여러 기법들이 있다[3,4]. 그리드 파일은 전체 영역을 분할하고 분할된 셀들이 데이터 버킷을 가리키는 식으로 구성되어 있다[5]. R-트리는 높이 균형 트리로써 B-트리의 다차원

확장이다. R+ 트리는 R-트리와 같이 같은 레벨에 있는 버켓 영역간의 중복을 허용하며 탐색 시간이 증가하고 객체의 MBR을 유지하는 알고리즘이 복잡하고 비용이 크다. R+ 트리는 버켓 영역간의 중복을 허용하지 않고 저장공간 비용이 크고 삭제 시 평균 탐색시간이 증가하고 버켓이 overflow될 가능성이 높아진다.

3. PDA에서 공간데이터를 표현하기 위한 방법

3.1 공간데이터 양 감소 방법

PDA는 저장공간이 제한되어 있기 때문에 공간데이터의 양을 줄이는 것이 필수적이다. 공간데이터를 감소 시키는 몇 가지 방법들이 있다.

첫번째는 모든 공간데이터를 PDA로 전송하는 것이 아니라 사용자의 요구에 맞는 데이터만을 추출하는 방법이 있다. 예를 들면 벡터 지도 데이터 중 사용자가 필요로 하는 레이어만을 선택하는 방법, 실제화 된 공간 뷰의 개념을 적용하는 방법 등이 있다.

두번째는 데이터의 압축 기법이 있다. 그러나 이동단말기에서 압축된 데이터를 처리하기에는 추가적인 압축처리 부분을 필요로 하고 CPU 처리 능력이나 적은 메모리 용량으로는 처리하기 어렵다.

세번째는 지도 일반화 방법이 있다. 즉 복잡한 공간 객체를 단순화 시키는 방법이다. PDA환경에 유용하나 이 작업을 처리하기에는 PDA가 연산 능력이 떨어지기 때문에 서버나 PC에서 처리하는 것이 효과적이다.

네번째는 펌 PDA가 64KB의 레코드 크기 한계를 가지고 있기 때문에 64KB크기 이상의 객체는 PDA로 전송할 데이터에서 제외 시키는 것이다. 실제 64KB이상의 객체는 희소하고 매우 복잡하기 때문에 화면상에 정확한 표현을 할 수 없기 때문이다.

3.2 공간 색인

공간데이터에 대한 효율적인 처리 및 검색을 위해서는 공간 색인의 구축을 필요로 한다. 순차적인 레코드 구조에 색인 구성없이 순서대로 객체를 저장한다면 영역질의나 검색 시 매우 비효율적이다. 여러 가지 공간 색인들이 있지만 각 PDA에 적절한 색인을 선택하고 적용해야 한다.

레코드 단위의 저장 구조를 가진 환경에서는 R-Tree, R+-Tree, R*-Tree 등과 같은 공간 색인 구조는 트리 형태의 구조를 가지고 있으므로 구조 및 연산이 복잡해 지게 된다. 그리고 공간 객체의 삽입이나 삭제 시 트리를 다시 재구성하는 연산을 수행하게 된다.

그리드 파일은 구조 자체가 간단하고 삽입 삭제가 자주 발생하지 않는 정적 상황에 대해서 뿐만 아니라, 삽입과 삭제가 빈번히 발생하는 동적 상황에서도 잘 적용되는 구조이다. 또한 영역 질의에 대해서 성능 및 연산이 용이하다. 그러나 추가적인 영역 추가할 경우 전체 색인을 재구성 하게 되는 단점이 있다.

이동 단말기에서는 지도의 검색은 영역 질의가 일반적이다. 따라서 이 논문에서는 그리드 파일 구조 중 고정 그리드 파일 구조를 응용한 색인을 통해 공간 데이터를 표현하는 방법을 제시한다.

4. 고정 그리드를 적용한 공간데이터 저장 구조 설계

4.1 기본 구조

이 논문에서 제시한 공간 데이터 저장 구조는 기본적으로 그림 1 과 같다. 데이터 공간을 동일한 크기의 셀들로 나누는 구조인 고정 그리드 파일의 구조를 적용한 색인 구조를 가지고 있다. 각 레코드의 구조는 동일한 셀에 속한 다음 객체의 레코드의 인덱스 번호(Next), 공간객체의 MBR영역(MBR), 실제 공간객체 데이터(Obj)로 구성된다. 즉 하나의 레코드에는 하나의 객체를 저장하고 있고 동일 셀에 속한 다른 객체의 레코드 번호를 저장하고 있는 것이다. 동일한 셀 내의 객체들은 레코드 인덱스에 의해 연결 리스트 형태로 저장 된다. 각 셀은 셀 내의 객체가 저장된 연결 리스트의 첫 노드를 포인팅하고 있는 것과 같은 구조이다.

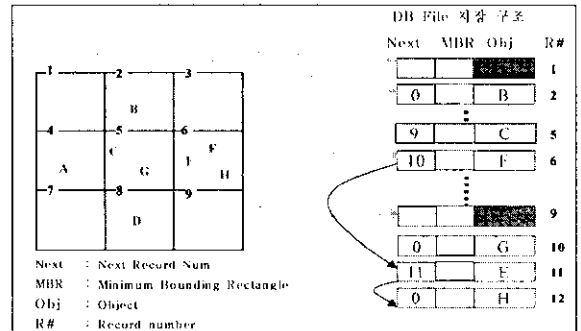


그림 1 기본적인 저장 구조

일반적인 그리드 파일의 저장 구조와는 달리 고정크기의 페이지를 할당하는 것이 아니기 때문에 저장 공간에 대한 낭비가 없다.

4.2 영역 질의

4.2.1 필터링(Filtering) 과정

사용자가 요청한 영역을 포함하는 셀들을 찾는다. 이 경우 요청한 영역의 MBR(Minimum Bounding Rectangle : 최소 경계 사각형) 좌표 값을 이용하여 찾는다. 각 셀이 포인팅하고 있는 레코드로부터 셀에 속한 객체들을 하나씩 순차적으로 탐색하게 된다. 각 객체들을 순차적으로 탐색해 나가면서 객체의 MBR영역과 사용자가 요청한 영역의 MBR을 비교한다.

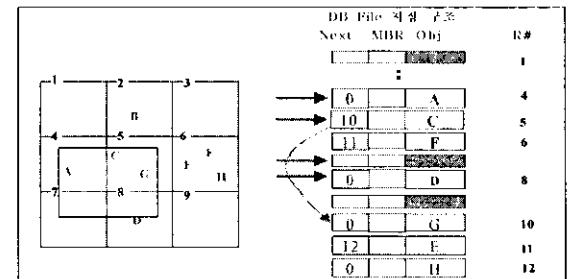


그림 2 영역 질의 시 검색과정

4.2.1 정제(refinement) 과정

필터링 과정 중 객체의 MBR영역과 사용자가 요청한 영역의 MBR이 교차하는 경우 실제 객체의 영역과 사용자가 요청한 영역의 MBR을 비교하는 정제 과정을 수행한다.

4.3 객체 삽입

그림 1에서 새로운 객체를 6번 셀 영역에 "Z" 객체를 삽입한다고 가정 하자. 새로운 레코드를 마지막에 추가하고 "Z" 객체를 새로운 레코드에 저장한다. "Z" 객체가 속한 셀은 6번 셀이다. 6번 셀은 인덱스 6의 레코드를 포인팅하고 있다. 따라서 인덱스 6의 레코드가 가지고 있는 다른 객체의 인덱스를 새로 추가한 "Z" 객체가 저장된 레코드의 인덱스로 대체한다. 그리고 원래 저장하고 있던 인덱스는 새로 추가된 "Z"가 저장된 레코드가 저장하도록 하면 객체 삽입이 이루어 진다. 그림 3은 이 과정을 나타내고 있다.

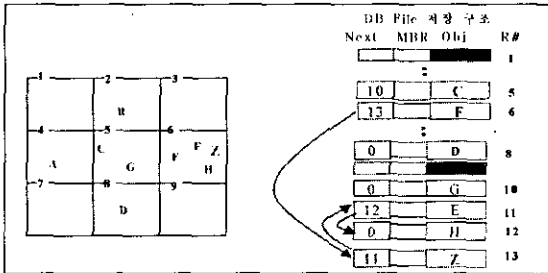


그림 3 객체 삽입 후 저장 구조

연결 리스트에 노드를 삽입하는 것과 동일하며 객체 삽입 비용은 상수 비용($O(1)$)으로 항상 일정하다. 항상 제일 마지막에 새로운 레코드를 만들고 객체를 저장한다. 따라서 기존의 그리드 파일과는 달리 필요한 저장 공간 만큼만을 사용하기 때문에 저장 공간의 낭비가 없고 버퍼의 overflow가 발생하지 않는다

4.4 객체 삭제

그림 1에서 6번 셀 영역의 "E" 객체를 삭제한다고 가정 하면 우선 "E"가 속한 셀이 포인팅하고 있는 레코드로부터 "E" 객체를 검색해 간다. "E" 객체가 발견되면 바로 전 레코드에서 "E" 객체가 저장된 레코드 인덱스를 저장하고 있는데 이것을 "E" 객체가 저장된 레코드에서 저장된 레코드 인덱스로 바꾼다. 삭제 비용은 객체를 찾는 비용 과 삭제 연산의 합이다. 삭제 연산은 항상 상수 비용이므로 삭제 비용은 객체 검색 비용과 동일하다. 객체의 평균 검색 비용은 V 를 셀의 평균 객체 수 라고 하면 $V/2$ 가 된다.

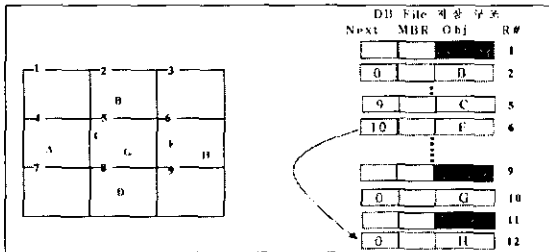


그림 4 기본적인 저장 구조

삭제된 레코드에 대해서는 다음과 같은 방법으로 관리한다. 삭제 된 레코드들을 동일한 셀에 저장된 객체를 저장하는 것과 같이 연결 리스트 형태로 관리할 한다. 그러면 객체 삽입 시에 새로운 레코드를 추가하는 것이 아니라 삭제된 레코드 중에 하나를 대체 한다.

4.5 새로운 영역 추가

무선통신이나 적외선 통신 등으로 필요에 의해 현재 저장된

영역 외에 새로운 영역에 대한 추가적인 요구가 발생할 수 있다. 이러한 경우에 처리하는 방법으로는 기존 색인을 확장하여 데이터를 포함하는 방법, 기존의 영역과 추가된 영역을 전체 영역으로 새로운 색인을 구축하여 데이터를 저장하는 방법, 별개의 색인을 구축하고 각각 관리하는 방법, 이중 색인을 구축하는 등의 방법이 있다.

기존의 고정 그리드 파일 구조는 새로운 영역을 포함하여 색인을 재구성하여야 하지만 여러 제약 조건을 가진 PDA에서 처리하기에는 비용이 크다. 따라서 이 논문에서는 그림 5에서와 같이 기존 색인을 그대로 유지한 채 새롭게 추가되는 영역에 대한 정보와 데이터를 추가한다. 그리고 기존의 영역과 추가된 영역의 비율이 정해진 임계 값 이상이 되면 색인을 재구성한다.

추가되는 셀의 영역정보 및 셀의 객체가 저장된 레코드의 인덱스는 0번 레코드에 저장하여 추가된 영역에 대한 정보를 유지한다. 이 정보를 이용하여 영역질의나 객체의 삽입 삭제에 대한 처리를 한다.

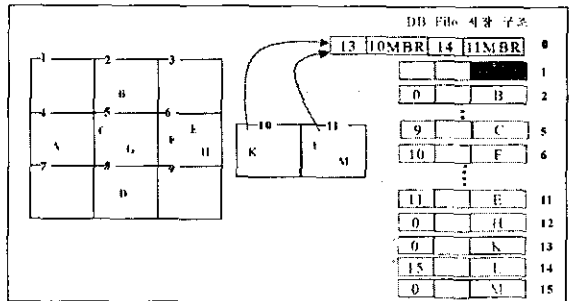


그림 5 새로운 영역이 추가된 저장 구조

확장된 영역이나 새로운 영역이 추가될 때 셀 단위로 추가 될 수 있다. 그러나 기존 영역과 멀리 떨어진 영역이 계속해서 추가될 때는 이 방법으로 처리하기는 어렵다.

5. 결론 및 향후 연구

이 논문에서는 PDA와 같은 이동단말기에서 공간 데이터를 효율적으로 다루기 위한 방법과 직접 접근이 가능한 순차적인 가변길이 레코드 단위의 저장구조를 가진 팜 PDA환경에서 공간데이터를 표현하기 위해 방법으로 고정 그리드 공간 색인을 적용한 공간데이터 표현 방법을 제시하였다.

향후 이 논문에서 제시한 팜 PDA에서의 공간데이터 표현 방법의 세부 설계 및 구현과 공간데이터의 질의 처리에 대한 성능 평가가 필요하다.

6. 참고 문헌

- [1] C. Bey, E. Freeman, and J. Ostrem. Palm OS programmer's companion. <http://www.palmos.com/dev/tech/docs/palmos/>, 2000
- [2] Todd L. Cignetti, Kirill Komarov and Carla Schlatter Ellis, "Energy Estimation Tools for the Palm", *Proceedings of the 3rd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, 2000, Pages 96-103, 2000
- [3] Michael F. Worboys, "GIS A Computing Perspective"
- [4] Volker Gaede, Oliver Günther: Multidimensional Access Methods. *ACM Computing Surveys*, 30(2): 170-231 (1998)
- [5] Nievergelt, j., Hinterberger, H. and Sevcik, K. C(1984) "The grid file : An adaptable, symmetric, multikey file structure". *ACM Transaction on Database Systems*, 9(1), 38-71