

확장된 관계 표현을 이용한 공간 데이터 구조의 설계

최제훈, 김종훈, 배혜영
인하대학교 전자계산공학과

Efficient Spatial Data Structure for Spatial Objects Using Relational Representation

Jae-Hun Choe, Jong-Hoon Kim Hae-Young Bac
Dept. of Computer Science & Engineering, Inha University

요 약

공간 데이터베이스 시스템에 관한 연구는 다양한 분야에서 진행되고 있으며 특히 공간 데이터베이스 시스템의 가장 중요한 응용인 지리정보시스템에 대한 연구가 중점적으로 이루어지고 있다. 지리정보시스템에서 이용되는 공간 객체는 Point, Line, Region 등으로 분류되며 불규칙한 n차원의 가변 길이 속성을 가진다. 이는 또한 공간 객체간의 위상 관계와 위치 정보를 포함하여야 한다. 본 논문에서는 기존의 저장구조를 용이하게 확장 가능할 수 있도록 하는 관계 표현을 이용한 공간 객체 데이터 구조를 제안한다. 제안된 구조는 공간 데이터와 속성 데이터를 효율적으로 연결시키며 공간 위상 관계를 유지하고 관리한다. 이는 또한 레이어 구조에서 관리되는 데이터에 의하여 공간 객체 간의 일관성을 유지시킨다.

1 서론

최근 다양한 분야에서 지도와 관련된 응용 업무에 대한 요구가 증대됨에 따라 공간 데이터베이스 시스템의 대표적인 응용 분야인 지리 정보 시스템에 대한 연구가 날로 증가하고 있다. 지리 정보 시스템(Geographic Information System GIS)이란 지리 및 지형에 관련된 데이터를 디지털화, 스케너, 마우스 또는 키보드 등의 컴퓨터 입력 장치를 이용하여 수집한 후 이를 효율적으로 저장, 처리, 검색하여 사용자가 원하는 형태로 출력 및 분석을 할 수 있도록 설계된 시스템이다[1].

GIS에서 사용되는 지리 데이터는 점, 선, 영역의 타입으로 구분되며 불규칙한 다차원의 가변 길이의 특성을 갖고 위치 정보뿐만 아니라 서로 다른 공간 데이터간의 위상 관련성(topological relationship)을 포함한다[2]. 이러한 지리 데이터는 최단 경로 검색과 같은 네트워크 분석이나 조건에 따른 위치 선정 분석 등과 같은 공간적인 요소에 기반을 둔 분석 처리가 요구됨에 따라 공간 데이터간의 연결성, 인접성, 포함성 등과 같은 공간적 위상관계를 포함하는 효율적인 데이터 구조를 지원하는 것이 필요하다[3]. 따라서 본 논문에서는 공간 데이터와 비공간 데이터의 효과적인 연결을 기하고, 공간적 위상 관계를 유지하고 관리할 수 있으며 공간 연산의 효율성을 높일 수 있는 공간 데이터 구조를 제안한다. 또한 이들을 계층적인 구조로 관리하여 공간 데이터간의 무결성을 지켜줄 수 있도록 설계한다.

본 논문의 구성은 다음과 같다. 2장에서 일반적인 공간 객체를 위한 자료 표현과 구조들을 살펴보고 3장에서 제안하는 공간 데이터 구조의 구성 요소와 구조를 설명하며 또한 이 구조에서의 공간 데이터와 비공간 데이터간의 무결성 유지 방안과 효율적인 연산의 처리 예를 보이고, 4장에서 결론을 맺는다.

2. 관련 연구

공간 데이터베이스는 비공간 데이터와 공간 데이터로 이루어진다. 예를 들어 도로라는 공간 객체는 선 타입의 집합으로 구성된 공간 데이터와 이름, 폭, 준공시기 등과 같은 속성 정보로서의 비공간 데이터로 이루어진다. 이러한 공간 객체를 유지, 관리, 분석하기 위한 데이터베이스 시스템들이 여러 가지 구조로 구축되어 왔다[1,4,7].

2.1 구조

공간 데이터베이스 시스템은 공간 데이터와 비공간 데이터를 어떻게 관리하는가에 따라 이중 구조, 계층 구조 그리고 통합 구조의 형태로 구분된다.

이중 구조는 공간 데이터와 비공간 데이터를 서로 다른 모델을 사용하여 두 데이터를 구분하여 처리한다. ARC/INFO, SICAD, GENASYS 등이 이러한 이중 구조를 갖는 시스템으로 두 개의 저장관리자를 갖는다. 이중 구조의 단점은 동기화, 락킹, 무결성 등을 보장하기 어렵다는 것이다[1,4,7].

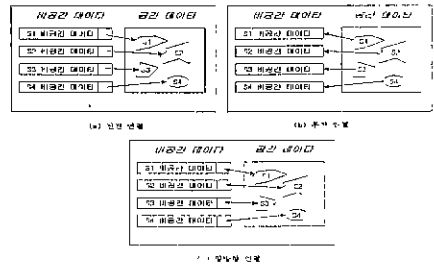
계층 구조는 저장 관리자를 하나만 갖는 구조로 일반적인 RDBMS에 비공간 데이터뿐만 아니라 공간 데이터도 함께 저장한다. 공간 데이터는

반드시 변환되거나 여러 조각으로 분리되어 저장되어야 하며, 여기에는 SIRO-DBMS[6], GEOVIEW 등의 시스템이 있다. 이의 관리를 위해 보통 3계층으로 구성되는데 가장 위에 지리 정보 시스템 층이 있고, 그 다음에 공간 지형 계층이 위치하며, 가장 아래에 RDBMS 층이 위치하게 된다. 이러한 계층 구조는 기존의 DBMS를 이용하므로 이식성이 높으나 공간 연산의 효율성이 떨어진다. 단점이 있다[1,4,7].

통합 구조는 공간 데이터를 지원하기 위해 그들 자신의 ADT(Abstract Data Type)을 가지도록 데이터베이스를 확장한 시스템이다. 주로 확장 가능한 관계형 모델에 의해 구현되며, 실제 구현은 DBMS의 윗부분에 구축되는 경우가 많다. 이러한 경우 공간 연산의 효율성이 떨어지게 되는데, 이를 해결하기 위해 GEOQL에서는 공간 데이터 구조를 따로 갖고 관리하며, 공간 연산에 용이한 데이터 구조를 갖게 된다[1,4,7].

2.2. 공간 및 비공간 데이터의 연결

공간데이터와 비공간 데이터를 통합 관리할 때 발생하는 중요한 문제는 어떻게 객체에서 공간 데이터와 비공간 데이터를 연결(link)시키는 가이다. 보통 다음과 같은 방식을 기본으로 한다. 진전 연결(Forward Link)은 비공간 데이터 릴레이션에 공간 데이터의 식별자로서 튜플 안에 저장된다. 후진 연결(Backward Link)은 공간 데이터 구조 안에 비공간 데이터의 식별자가 저장된다. 식별자는 비공간 데이터의 튜플 식별자로서 구현된다. 양방향 연결(Bi-directional Link)은 진전 연결과 후진 연결을 같이 가진다[4,7].



[그림 1] 공간 및 비공간 데이터의 연결

진전 연결은 객체의 공간 데이터를 얻는데 사용되고 후진 연결은 비공간 데이터를 얻는데 사용된다. SAND와 GEOQL에서는 양방향 연결을 사용하며, 공간 데이터 구조를 따로 관리한다. 그러나, 많은 시스템들이 시스템의 공간이나 비공간 어느 한편으로만 치우쳐있다. 예로 GEOQL에서는 SQL을 확장하여 공간연산을 지원한다. 하지만 관계형 구조에 치우쳐 있기 때문에 비록 공간 자료 구조가 유지되고 있을지라도 공간연산은 중간 계

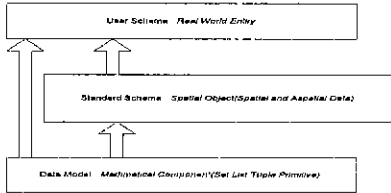
이들의 생성 없이 직접 수행될 수 없다 이것은 공간 질의 처리의 효율성을 떨어뜨린다[4,7]

SAND에서는 공간과 비공간데이터가 양방향으로 연결되어 있으며 공간 자료구조는 스키마에서 공간 속성과 연관되어서 동질의 객체의 집합을 저장하는데 사용되어진다 공간 데이터 구조는 속성의 공간 데이터 타입에 의존하며 공간 연산의 실행 환경으로 사용되어진다[4].

3. 계층적 공간 데이터 구조

3.1 데이터 파라다임

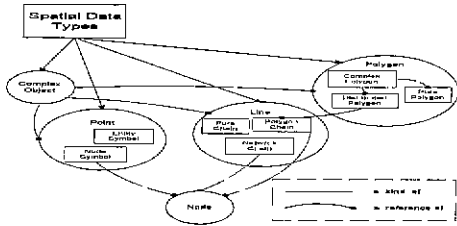
[그림 2]는 본 논문에서 제안하는 고수준 자료 모델의 구조도로 수학적 요소로 구성되는 자료 모델은 집합, 리스트, 튜플, 그리고 프리미티브로 구성되며, 이는 표준 스키마 계층에서 공간 및 비공간 자료를 구성하는 단위로 정리될 수 있다 또한 상위 단계에서는 실제계를 구성하는 요소를 정의하는 것으로 사용자의 관점에서 다루어진다



[그림 2] 데이터 모델 Paradigm

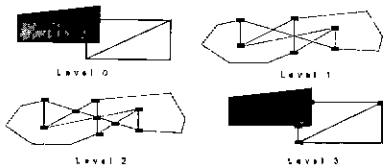
본 연구에서는 공간 데이터 특히 위상 자료 표현에 대해 심도 있게 고려한다 한편, 위상 자료는 공간 자료가 갖는 논리적인 자료 구조로 각 자료를 인스턴스로 간주하며 자신은 복합 객체를 대표로 표현하는 통합 객체 자료로 간주한다 즉, 다양한 표현 자료들의 특성과 그들 간의 관계를 논리적인 스키마로 표현하여 관리한다

이들 공간 데이터 타입은 점, 선, 영역의 기본적인 타입의 지원뿐만 아니라 이들을 계층적으로 결합하여 새로운 사용자 데이터 타입을 제공할 수 있도록 복합 객체를 지원한다 뿐만 아니라 공간 데이터간의 완전 위상을 표현하기 위해 각 기본 데이터 타입은 [그림 3]과 같이 세분화하여 타입을 지원하고, 각 공간 객체는 노드 또는 단순한 표현의 구조를 갖는다



[그림 3] 공간 데이터 타입과 타입간 관계

이러한 공간 데이터 타입에 의해 제공되는 연산들 가운데 가장 중요한 것은 공간 데이터의 관련성을 표현하는 위상 관계이다 따라서 공간 데이터 타입을 표현하기 위해서는 반드시 위상 관계를 고려한 타입을 추가하여야 한다 위상 관계는 한 질의 객체로서 주어진 관계 내에서의 모든 객체에 묻는 것을 가능하게 한다 본 연구에서는 위상 관계를 가장 잘 표현한 DIGEST의 VPF 형식에 주로 사용하는 4 단계의 위상 관계를 이용한다 즉, 위상 관계가 전혀 없는 위상 레벨 0 공간 데이터간에 점과 선을 각각 노드와 아크로 표현하여 서로의 연결성을 표현한 위상 레벨 1, 그리고 각 아크간에 접착된 부분을 노드로 표현하여 교차성을 표현한 위상 레벨 2, 마지막으로 영역간에 관계를 표현하여 인접성 등을 표현한 위상 레벨 3로 나눌 수 있다 [그림 4]는 4 단계의 위상 레벨을 표현한 것이다

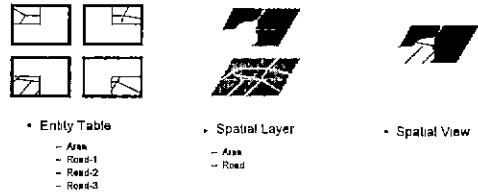


[그림 4] 공간 데이터의 위상 표현

공간 데이터 타입은 위상 관계를 갖는 것과 위상 관계를 갖지 않는 것을 구분하여 지원한다. 이는 위상 관계가 필요 없는 공간 데이터들도 존재할 수 있는데 이러한 공간 데이터를 위상 관계를 가지는 공간 데이터로 표현할 경우 기억 공간의 낭비기 발생하기 때문이다

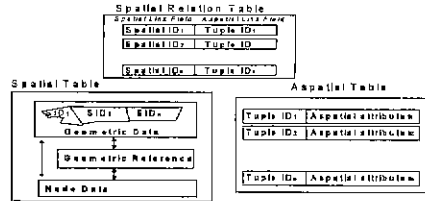
3.3 공간 객체의 논리적 표현

본 구조에서는 공간 객체를 논리적으로 완전히 통합된 공간 뷰 부분과 각 계층별로 분리되었지만 여러 개체를 논리적으로 통합한 공간 레이어 부분, 그리고 각 공간 객체를 물리적인 튜플로 표현한 개체 테이블의 3가지 계층적인 구조로 구분하여 관리한다



[그림 5] 공간 객체의 구성

공간 뷰는 하나이상의 레이어로 이루어지며, 사용자에게 제공되는 공간 객체 그룹의 가장 큰 단위이다 공간 레이어는 실제계의 지형요소를 구분하는 기본 단위의 집합으로, 도시, 도로, 학교, 동경계 등이 이에 해당되며, 하나이상의 엔티티 테이블로 이루어진다 엔티티 테이블은 시스템에 저장되는 물리적인 단위로 공간 데이터와 비공간 데이터가 실제로 저장되고 관리되는 단위이다



[그림 6] 계층적인 공간 객체의 표현

각 엔티티 테이블은 다시 Spatial Relation Table, Spatial Table 그리고 Aspatial Table의 3부분으로 구성된다 공간 데이터와 비공간 데이터의 연결은 양방향 연결을 하고 있으며 이 연결부분은 Spatial Relation Table로 분리하여 관리한다 이는 공간 데이터 처리부와 비공간 데이터 처리부를 명백히 분리하여 시스템 구조의 복잡성을 줄이며, 개층화된 관리를 가능하게 한다 각각의 엔티티 테이블의 관리는 Spatial Relation Part를 통하여 이루어지도록 하여, 공간과 비공간 데이터간의 무결성을 유지하도록 한다 이렇게 계층화함으로써 추후 공간데이터 타입의 확장이 용이해 지고, 비공간 데이터의 외부 DBMS와의 연결이 용이하여 높은 유연성을 제공할 수 있다.

(1) Spatial Relation Table

공간 데이터와 비공간 데이터의 연결을 이루고 있는 부분으로 Spatial Link Field와 Aspatial Link Field로 이루어진다 Spatial Link Field는 Spatial Part의 공간 식별자(Spatial ID)를 갖고, Aspatial Link Field는 Aspatial Part의 튜플 식별자(Tuple ID)를 갖는다

(2) Spatial Table

Spatial Table 부분은 공간 데이터를 저장하며, 이는 다시 위상관계의 표현과 내부 무결성을 유지시키기 위하여 Geometric Data, Geometric Reference Data, 그리고 Node Data의 세 부분으로 나뉘어 저장된다 Geometric Data는 독립적인 공간 객체를 저장하는 부분으로 실제 이 부분만이 Spatial Relation Table에서 연결(Link)될 수 있다 여기에는 공간 객체의 각종 그래픽 속성 값인 칼라, 라인종류 등을 갖는다 Geometric Reference Table은 Geometric Table의 공간 객체에서 참조되는 객체들이 저장된다 이 부분은 단독으로 삭제될 수는 없고 그래픽 속성은 가지지 않는다 마지막으로 Node Table은 각각의 공간 객체에 대한 좌표값이 저장되어 있는 노드들을 저장한다 이러한 분류의 기준은 직접적으로 공간 객체에서 참조되는 부분과 참조되지 않는 부분 그리고 노드 부분을 따로 나누어 이를 계층적으로 관리하고 있어 공간 객체에 대한 삭제나 삽입시의 무결성 유지에 이용이 된다

(3) Aspatial Table

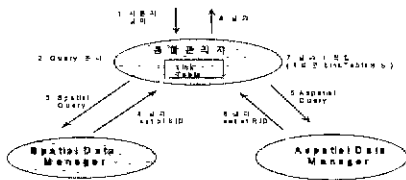
비공간 데이터를 저장하는 부분으로 일반 테이블의 구조로 관리된다

3.4 질의 처리

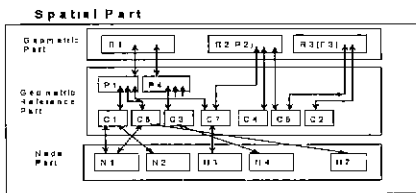
질의 처리는 사용자의 요구 사항을 질의어를 통해 수행하며, 질의 처리는 공간 데이터와 비공간 데이터를 나누어 처리하게 된다 이때 공간 데이터에 대한 질의는 공간 데이터 관리자에 의해 수행되며 비공간 데이터는 비공간 데이터 관리자에 의해 수행된다 이때 사용자의 질의는 통합 관리자를 통해 각각 공간, 데이터 관리자와 비공간 데이터 관리자에게 순차한 공간 질의와 비공간 질의로 나누어 처리할 수 있도록 분할하고 각 데이터 관리자로부터 얻어진 결과를 통합시키는 역할을 한다 이러한 처리과정은 사용자에게 RECORD가 공간 데이터 타입을 포함하는 확장된 RDBMS의 질의 처리 과정 관점으로 보여준다

< 질의 처리 과정 >

- 1 사용자의 질의가 들어오면, Link Table로 접근하게 된다
- 2 여기에서 Parse Tree가 생성되고, Spatial Query와 Aspatial Query로 분할된다
- 3 Spatial Query는 Spatial Table로 보내져 수행된다 결과는 SID(Spatial ID)의 Set으로 다시 Link Table로 보내진다
- 4 Aspatial Query는 Aspatial Data Manager로 보내져 수행된다 결과는 RID(Record ID)의 Set으로 다시 통합관리자로 보내진다
- 5 통합관리자에서는 연산의 결과를 가지고, Aspatial Query와 Spatial Query 사이의 Boolean 연산에 따라 Link Table, 3의 결과(Set of SID), 4의 결과(Set of RID)를 조합하여 새로운 Link Table을 만들어낸다 이때 만들어진 Link Table은 View의 성격과 가지며, 이것이 사용자에게 Query에 대한 결과이며, 필요에 (Create table을 하는 경우) 따라 이것을 가지고 다른 Operation을 수행할 수 있다



[그림 7] 질의 처리 과정



[그림 8] 공간 데이터의 저장 예

[그림 8]은 Geometric Reference Data의 Node Data는 공간 객체들에게 공유되고 객체간의 관계 정보들을 가지고 있다 C6은 공간객체 Region2와 Region3에 공유되고 인접성이란 위상 관계를 표현한다 N2는 C1과 C2에 의해 공유되며 연결성이란 위상 관계를 갖게 된다

이러한 계층적인 구조에서 상위의 공간 데이터들은 참조하는 하위 공간 데이터의 좌표값들을 중복하여 갖도록 한다 데이터의 수정이나 일러서 관련된 상위의 공간 데이터들도 갱신이 필요하므로 비용이 증가하고 저장 공간의 낭비가 발생하지만 Contain이나 Intersect 등의 일반적인 공간 연산들은 전체 데이터를 검색하지 않고 Geometric Part만을 검색하도록 하여 공간 연산의 효율성을 높인다 그리고 Geometric Reference Data와 Node Data는 직접적인 삭제를 허용하지 않는다 이는 공간 관계의 무결성을 지켜주기 위함이다 그리고 기존의 시스템은 그래픽 속성을 따로 저장하거나 공간 객체의 비공간 데이터로 갖도록 하였으나 본 구조에서는 그래픽 속성을 Geometric Part의 공간 데이터에만 갖도록 하여 편리함으로써 비공간 데이터 부분과 공간 데이터 부분을 철저히 구분한다

예를 들어 '집배우명이 제2구역인 공간 객체를 삭제하라'라는 질의가 발생하면, 먼저 검색이 수행되고, 공간 데이터와 비공간 데이터가 삭제되게 된다 이때 공간 데이터의 식체를 위해 몇 가지 규칙을 만족해야 하는

데, 우선 삭제 연산은 항상 상위 계층에서 호출이 되어야 비토스 수행이 된다는 것이다 이렇게 상위 계층에서 호출이 일어나면 효율한 상위 계층의 공간 데이터와의 연결을 먼저 삭제하게 된다 그리고 밑일 상위 계층의 연결된 것이 모두 끊어진 상태가 되면 연결된 하위 계층의 공간 데이터들에게 삭제권 효율하게 되고 자신을 삭제한다 질의의 결과로 공간 데이터 R2의 하위 계층에 연결되는 C6, C7, C11은 R2와의 연결이 끊어지고 C4, R2는 식체된다 계층 관계에서 서로간의 연결 관계를 이용하여 전체 데이터의 무결성을 유지하도록 한다 만일 '시각형 [(10,10),(30,30)]안에 포함되어 있는 모든 모든 객체를 검색하라'는 질의가 오면 검색의 대상은 전체 23계층에서 Spatial Relation Data의 3개의 공간 데이터만으로 충분하여 공간 연산의 효율성이 증대된다

5. 결론

공간 데이터베이스 시스템은 공간 데이터와 그의 속성 데이터인 비공간 데이터를 효율적으로 관리하기 위해 공간 객체에 대한 공간 데이터 타입을 정의해야 하며, 그들간의 위상적인 공간 관련성을 표현할 수 있어야 한다 즉, 공간 데이터가 중간적 위상 관계를 유지하고 관리되어야 하며 이들 사이의 무결성을 지키고 공간 연산의 효율성을 고려한 공간 데이터 모델이 정의되어야 한다 뿐만 아니라 이러한 공간 데이터 모델은 단순한 공간 데이터의 위상 정보뿐만 아니라 실질적인 데이터인 비공간 데이터와 통합되어 연결 처리되어야 한다

기존의 지리 정보 시스템과 같은 공간 데이터베이스 시스템의 응용 분야에서는 공간 데이터를 저장할 경우에 단지 집 선, 영역과 같은 공간 데이터 타입 단위만으로 저장하여 무결성의 유지가 힘들고 일부 공간 연산에서는 효율성이 떨어지며 복잡한 공간 데이터를 지원하기 어렵다는 단점을 갖고 있다

본 논문에서는 공간 데이터를 집 선, 영역의 기본적인 타입으로의 지원 뿐만 아니라 이들을 계층적으로 결합하여 새로운 사용자 데이터 타입을 제공할 수 있도록 복합 객체를 지원하며 공간 데이터간의 완전 위상을 제공하는 공간 데이터 타입을 정의하였다 또한 이러한 공간 객체는 논리적으로 완전히 통합된 공간 뷰 부분과 각 계층별로 분리되었지만 여러 계층을 논리적으로 통합한 공간 데이터 부분, 그리고 각 공간 객체를 물리적인 묶음으로 표현한 계층 데이터의 3가지 계층적인 구조로 구분한 공간 데이터 구조를 제안하였다

제안된 구조는 여러 가지 공간 데이터 타입을 쉽게 지원하며 데이터간의 관계성을 명확히 해주고 특히 위상 관계에 의한 공간 데이터의 무결성을 유지하는데 효과적인 구조를 가지고 있다 또한 향후 연구 방향은 제안된 구조에서 치리할 효율적인 공간 연산자의 처리에 관한 내용이다 또한, 포함이나 교차 등의 일반적인 공간 연산들은 전체 데이터를 검색하지 않고 공간 정보의 일부분만을 참조하므로써 공간 연산의 효율성을 높였다

참고 문헌

- [1] A Belussi, "Designing and Evaluating database schemas for ARC/INFO applications," *EGIS94 Conference Proceeding*, Vol 2, pp 1177-1185, 1994
- [2] B C Ooi, R Sacks-Davis and K J McDoneli, "Extending A DBMS For Geographic Applications," *Proceedings 5th Int Conference Data Engineering*, pp 590-597, 1989
- [3] D J Abel, "SIRO-DBMS A database tool-kit for geographical information systems," *Int Journal of Geographical Information Systems*, Vol 3, No 2, pp 103-116, 1989
- [4] D J Pequet, "Data Models for Very Large Geographic Databases" *Proc. of Int. Workshop on GIS*, 1987
- [5] I Bracken and C Webster, "Towards a typology of geographic information systems," *Int Journal of Geographical Information Systems*, Vol 3, No 2, pp 137-152, 1989
- [6] J Palimaka, et al, "Integration of a Spatial and Relational Database Within a Geographic Information System," *Proc of Int Workshop on GIS*, 1987
- [7] J Star and J Estes, "GEOGRAPHIC INFORMATION SYSTEMS -An Introduction", PRENTICE HALL 1994
- [8] M C Worvoys, "A Generic model for planar geographical objects," *Int Journal of Geographical Information Systems* Vol 6, No 5, pp 353-372, 1992
- [9] Ralf hartmut G'ring, "An Introduction to Spatial Database Systems," *The VLDB Journal*, Vol 3 No 3 pp 357-399, 1994
- [10] Ralf hartmut G'ring and Markus Schneider, "A Foundation for Data Types in Database Systems," *Proc 3rd Symp on Spatial Databases SSD93* pp 14-35, 1993
- [11] Walid G Aref, Hanan Samet, "Extending a DBMS with Spatial Operations," *Lecture Notes in Computer Science 525*, pp 299-318, 1991