

제약 데이터베이스에서의 효율적인 공간질의 처리

(Efficient Spatial Query Processing in Constraint Databases)

우 성 구* 류 근 호**
 (Sung Koo Woo) (Keun Ho Ryu)

요 약 제약 데이터베이스의 튜플은 제약 논리식으로 구성되었으며, 공간 데이터의 표현 및 질의를 비교적 간결하게 처리할 수 있다. 제약 데이터베이스를 통한 공간 데이터의 실렉션, 합집합, 교집합 등의 질의 연산 처리는 간단하게 관련 튜플 간에 제약식을 포함시켜 주면 된다. 그러나 중복되거나 불필요한 데이터가 증가할 수 있으며, 그에 따른 질의 비용이 증가될 수도 있다. 이 논문에서는 제약 데이터베이스에서의 질의 처리 결과에 대한 문제점을 밝히고, 효율적인 질의처리를 위해서 결과 릴레이션에 대한 튜플 최소화 정리 기법의 제안 및 그 효과를 분석했다. 우리는 튜플 최소화 기법을 이용하여 제약 릴레이션의 불필요한 제약식을 제거함에 따른 질의처리의 효율성이 개선됨을 확인할 수 있었다.

키워드 : 제약 데이터베이스, 공간 데이터베이스, 공간질의, GIS응용

Abstract The tuple of constraint database consists of constraint logical formula and it could process the presentation and query of the constraint database simply. Query operation processing shall include the constraint formula between related tuple such as selection, union, intersection of spatial data through the constraint database. However, this could produce the increasing of duplicated or unnecessary data. Hence, it will drive up the cost as per quality. This paper identified problems regarding query processing result in the constraint database. Also this paper suggested the tuple minimization summary method for result relation and analyzed the effects for efficient query processing. We were able to identify that the effectiveness of the query processing was enhanced by eliminating unnecessary constraint formula of constraint relation using the tuple minimization method.

Keywords : Constraint database, Spatial database, Spatial query, GIS application

1. 서 론

대용량 공간 및 시공간 데이터와 같은 복잡한 데이터를 간결하게 처리할 수 있도록 등장한 패러다임이 제약 데이터베이스(Constraint Database, CDB)이며, Kanellakis 등을 통하여 데이터로그(Datalog)와 제약 논리 언어(Constraint Logic Programing, CLP)의 결합으로 데이터베이스를 표현하고자 그 이론을 정립하여 발표하였다[1,2].

CDB는 대용량의 공간 및 시공간의 복잡한 데이터를 간결하게 처리할 수 있도록 무한의 점 좌표라는 추상적인 개념을 도입했으며, 그 무한한 점들은 구체적인 제약식을 통하여 유한하게 표현하도록 하는 기법을 응용했다.

제약 데이터 모델은 공간 데이터를 제약식을 이용하여

기존 관계 데이터 모델의 일반화 된 튜플(tuple) 형태로 표현할 수 있다[3]. CDB의 튜플은 제약 논리식으로 구성되었으며, 공간 데이터의 표현 및 질의를 비교적 간결하게 처리할 수 있다. 예를 들면, 모서리 좌표가 (0, 0), (0, 3), (5, 3), (5, 0)의 직사각형은 제약식 $0 \leq x \leq 5 \wedge 0 \leq y \leq 3$ 와 같이, 중심점(0, 0)과 반지름 5로 된 양의 반원은 제약식 $X^2+Y^2 \leq 25 \wedge X \geq 0$ 와 같이 표현할 수 있다[1,2,3].

CDB에서의 질의처리 표현은 관계 데이터베이스(RDB)에서와 마찬가지로 SQL은 물론 관계해석(relational calculus) 및 관계대수(relational algebra)와 같은 질의 언어들의 접근이 가능하며, 제약 릴레이션(relation)의 질의처리의 연산결과도 제약 릴레이션이라는 폐쇄 성질(closure property)의 특징이 있음을 알 수 있다[2,4,5,6,7].

[†] 이 논문은 2008년 정부(교육과학기술부)의 지원 (지역거점연구단육성사업/충북BIT연구중심대학육성사업단) 및 2009년도 정부(교육과학기술부)의 지원으로 한국과학재단의 지원을 받아 수행된 연구임(No. R11-2008-014-02002-0)(PTERC).

* 충북대학교 대학원 데이터베이스/바이오인포매틱스연구실, wsk527@naver.com

** 충북대학교 전기전자컴퓨터공학부 교수, khryu@dlab.chungbuk.ac.kr(교신저자)

따라서 CDB는 중간 처리 조정 층을 두어서 상호 운영이 가능하도록 제시되었다[8]. 즉, 그 모델 간에 구조가 달라도 추후 사용함에 있어서 서로 간에 상호운영이 가능하도록 다른 응용프로그램이나 다른 사이트에서 데이터를 공유하여 원활하게 조작되도록 개발되었다.

이 논문에서는 CDB의 개념과 CDB를 이용하여 공간 데이터의 질의처리 응용 및 질의최적화에 대해서 고찰한다. 실험에 적용한 CDB 프로토타입(prototype)은 순수한 제약 데이터 모델로 개발된 MLPQ/PReSTO 도구를 사용한다. CDB의 효율적인 질의처리를 위해서 결과 릴레이션에 대한 튜플 최소화 정리 기법을 적용했다.

이 논문은 튜플 최소화 기법을 통하여 제약 릴레이션의 중복 및 불필요한 제약식을 제거함으로써 그 처리 효율성이 개선됨을 확인했다. 2장과 3장에서는 CDB의 개념과 표현 및 CDB 질의처리에 대한 기존 연구를 고찰했다. 4장에서는 CDB의 질의처리 및 문제점을 제시했고, 5장에서는 CDB에서의 효율적인 질의 처리 방법과 기대 효과를 제시했으며, 마지막 6장에서 결론을 맺었다.

2. CDB의 질의처리

CDB는 지리객체의 무한한 점집합을 모델링하기 위해서 기존 관계모델에 내포된 릴레이션(nested relation)과 같이 확장된 형태로 표현이 가능하며[1,4], 관계대수, 관계해석, 데이터로그와 그 확장형에 이르기까지 모두 적용할 수 있는 넓은 범위의 질의 언어가 유용하다[1,8].

2.1 CDB의 개념과 그 표현

CDB는 RDB가 공간 데이터를 표현함에 있어서 무한 릴레이션을 다루는 능력이 존재하지 않기 때문에 무한 관계를 다루기 위한 것을 주요 목표로 세웠다. 그 결과 기존 공간 데이터에서 그 표현이 가능하고 순수한 관계 언어를 이용하여 공간질의 표현이 가능하도록 했다. 그리고 또 하나의 장점은 통일된 프레임워크에 임의 차원의 데이터를 쉽게 표현하고 다룰 수 있도록 구성되었다[1,3].

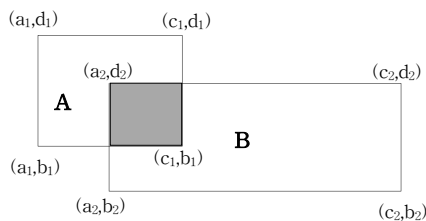


그림 1. 두 개의 직사각형

그림 1과 같이 두 개가 일부 겹쳐진 직사각형에 대한 이름 및 좌표에 대한 공간 데이터를 저장한 릴레이션은 표 1과 같다.

표 1. 두 개 직사각형 릴레이션 R

n	a	b	c	d
A	a1	b1	c1	d1
B	a2	b2	c2	d2

위의 릴레이션에 전형적인 질의 “릴레이션에 점 (x, y)이 존재하는가?”의 질의를 논리식으로 표현하면,

$Q_1 \equiv \exists n \exists a \exists b \exists c \exists d (R(n,a,b,c,d) \wedge a < x < c \wedge b < y < d)$ 이며, 또 다른 질의의 예는 “내부가 꼭 찬 사각형의 교차 연산을 하여서 그 구분되는 사각형의 모든 쌍 (n, n')을 출력하라” 는 질의를 했을 때, 해석 질의는,

$Q_2 \equiv \{(n, n') \exists a \exists b \exists c \exists d \exists a' \exists b' \exists c' \exists d' (n \neq n' \wedge R(n, a, b, c, d) \wedge R(n', a', b', c', d') \wedge \text{intersects}(a, b, c, d, a', b', c', d'))\}$ 으로 표현된다.

여기서 intersects(a, b, c, d, a', b', c', d')는 n과 n'의 교차된 사각형에 해당되는 사실(fact)을 표현하려고 한다. 그런데 이 교차(intersect)연산을 위한 프레디케이트(predicate)를 다음과 같이 $\exists x \exists y (a < x < c \wedge b < y < c \wedge a' < x < c' \wedge b' < y < d')$ 표현하면, 이 식은 작동하지 않는다. 그 이유는 논리식에서 정량자의 값 x, y가 존재하지 않기 때문이다. 따라서 아주 복잡한 논리식이 적용되어야만 그 해를 구할 수 있다[1,9].

그러나 사각형이 무한한 점 좌표 집합이라는 추상적인 개념으로 구성되었다면, 그 해는 아주 용이하게 처리되고, 그것을 구체화 시킨 것이 바로 제약식이다. 그래서 CDB의 튜플 형태는 제약식으로 구성되었다. 그림 1의 두 개의 직사각형은 표 2와 같은 제약식으로 표현할 수 있고, 그 질의는 다음과 같이 아주 간결해진다.

표 2. 제약식에 의한 직사각형의 표현

사각형id	제약식에 의한 유한표현	비고
n1	$x >= a1 \wedge x <= c1 \wedge y >= b1 \wedge y <= d1$	무한 점좌표
n2	$x >= a2 \wedge x <= c2 \wedge y >= b2 \wedge y <= d2$	무한 점좌표

$Q_1 \equiv R(x, y)$ 이며,

$Q_2 \equiv \{(n_1, n_2) \mid n_1 \neq n_2 \wedge \exists x \exists y (R(n_1, x, y) \wedge R(n_2, x, y))\}$ 로서 간단히 해결된다.

제약 데이터(constraint data)는 데이터의 처리방식이나 표현방식에 제한을 가하는 조건으로 그 표현방식은 주로 일차술어논리(first order logic, FOL)이며, 식은 DNF(Disjunctive Normal Form, 선언표준형)형태로 표기했다.

CDB의 튜플의 형태는 $r(x_1, \dots, x_k) :- \emptyset_1 \wedge \dots \wedge \emptyset_n$ 이다. 여기서 r은 릴레이션 심볼이며, Φ 는 원자적 제약 집합, $1 \leq i \leq n$ 범위의 $\emptyset_i \in \Phi$, 그리고 변수 x_1, \dots, x_k 를 칭한다 [6,8,7].

원자적 제약 집합 Φ 는 다음과 같이 표현할 수 있으며,

$$\sum_{i=0}^n a_i x_i \Theta a_0$$

여기서 Θ 는 $=, \leq, \geq$ 와 같은 관계연산자를 의미하며, x_i 는 변수, a_i 는 정수상수, $a_i x_i$ 항목은 $x_i + \dots + x_i$ (a_i times)를 의미한다[4].

다양한 제약식의 형태를 분류해 보면 다음과 같다.

- Equality : $x = 100, y = 200$
- Dense-Order : $x \geq 2, y \leq 3, y \leq x$
- Temporal : $y \leq x + 7$
- Monotone : $y \leq ax + 7, a > 0$
- Linear : $5y + 3x \leq 7, y - 2x \geq 9$
- Polynomial : $x^2 + y^2 = 4, x^2 \leq 3$

위의 제약 형태 분류에서 마지막의 다항식제약은 고차원 곡선의 표현 등이 가능하지만 복잡한 연산비용이 따른다.

그림 2는 4개 제약의 논리곱(conjunction, \vee)으로서 다각형을 표현한 예이며, 그림 3은 그 다각형을 제약식으로 표현한 것이다.

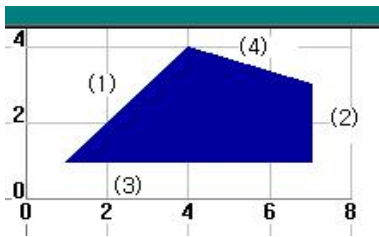


그림 2. 다각형(1) 표현

- (1) $y \leq x$
- (2) $x \leq 7$
- (3) $y \geq 1$
- (4) $x + 3y \leq 16$

그림 3. 다각형(1)의 제약식

각 제약식의 $ax+by+c=0$ 형태는 유클리드 공간에서 반 평면(half-plane)으로 정의한 것이다. 이 때문에 CDB모델에서의 다각형(polygon)은 반드시 볼록다각형 형태로서 정의되어야만 그 표현이 가능하다. 만약 그림 2의 다각형 구역을 x, y 좌표쌍으로 표현한다면, 유리수일 경우에 거의 무한에 가깝지만, 4개 논리곱으로 구성된 제약 튜플로서 간결하게 나타낼 수 있다.

그림 4와 같이 좀 더 복잡한 다각형(2)의 제약식 표현은 볼록다각형으로 되어야 가능하기 때문에 3개의 형태

로 나누어서 제약식 튜플로 표현한 것이 그림 5이다.

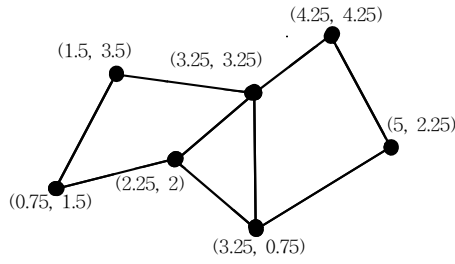


그림 4. 다각형(2) 표현

```
convex(1, x, y) :-
    4x-12y<=-5^16x+6y<=3^2x+14y<=52
    ^20x-16y<=13
convex(2, x, y) :-
    -20x+16y<=-13^20x+16y>=77^4x<=13
convex(3, x, y) :-
    4x>= 13^-x+y<=0^32x+12y<=187
    ^-24x+28y>=-57
```

그림 5. 다각형(2)의 제약식

2.2 CDB에서의 질의표현

CDB에서의 질의처리는 관계해석과 같은 1차 술어논리 뿐만 아니라 SQL 표현 및 관계 대수 질의 표현도 가능하다. 관계대수 질의는 관계해석의 선언적인 방식과는 달리 절차적인 언어로 구성되었으며, 동등한 질의 표현이 가능하다. 두 릴레이션 R_1, R_2 , 그리고 그 각각을 제약 튜플의 기호적 릴레이션을 e_1, e_2 로 놓으면, 다음 공식의 왼쪽은 추상적 수준의 연산을 말하며, 오른쪽은 기호를 이용하여 그 의미를 나타낸다[5].

- $\sigma_F(R_1) = \{t_1 \wedge F, t_1 \in e_1\}$
- $R_1 \times R_2 = \{t_1 \wedge t_2 \mid t_1 \in e_1, t_2 \in e_2\}$
- $R_1 \bowtie R_2 = \{t_1 \wedge t_2 \mid t_1 \in e_1, t_2 \in e_2\}$
- $R_1 \cap R_2 = \{t_1 \wedge t_2 \mid t_1 \in e_1, t_2 \in e_2\}$
- $\pi_{\bar{x}}(R_1) = \{\Pi_{\bar{x}}(t) \mid t_1 \in e_1\}$
- $R_1 \cup R_2 = \{e_1 \cup e_2\}$
- $R_1 - R_2 = \{t_1 \wedge t_2 \mid t_1 \in e_1, t_2 \in (e_2)^c\}$

이러한 관계대수 연산자는 질의최적화와 같은 절차적 연산과정이 필요한 경우 질의의 중간형태 표현으로 사용한다.

CDB의 intersection연산에 관한 관계 해석의 표현은 다음과 같다.

- 질의 1) 강이 통과하는 도시지역은?
- $\{x, y \mid \text{River}(x, y) \wedge \text{City}(x, y)\}$

또한 SQL표현은 기존 RDB에서의 공간 데이터를 처리하는 표현과 거의 흡사하다. 그리고 모든 공간객체가

무한한 점 좌표로 구성된 추상계층이 있기 때문에 질의 처리 사용자가 한층 간결하게 표현할 수 있다.

질의 2) 청주시에 있는 모든 산업지역 검색하라.

```
Select p1
From c in City, l in Land_Use,
     p1 in c.geometry, p2 in l.geometry
where city_name = '청주시'
      and land_use_type = '산업지역'
      and p1.x = p2.x and p1.y = p2.y
```

질의 3) 화면상에 주어진 사각형과 교차하는 도시에 대하여 기술하라.

```
Select city_name, population
From c in City, p in c.geometry
Where p.x = @rect.x and p.y = @rect.y
```

3. CDB 질의 처리에 대한 기존 연구

전통적인 RDB의 접근에 비하여 제약 페러다임은 공간 데이터베이스 및 공간질의 처리와 관련하여 기본적인 문제점을 연구하기 위하여 편리한 수학적 프레임워크를 제공했다. 이러한 문제점 중에 하나는 복잡한 질의평가이며, 또 하나는 질의 언어의 표현력이다. CDB는 일차 술어논리로 구성된 질의이기 때문에 이러한 문제를 좀 더 쉽게 접근하고 표현할 수 있다[10].

CDB에서 제약의 통합을 위한 이론적 프레임워크는 [6]의 논문에서 제시되었다. 그 논문은 제약 질의 언어는 관계해석이나 데이터로그로서 상이한 네 가지 제약 클래스¹⁾를 추가하여 이론적으로 만족함을 제시하였다. 이 논문에서는 각 클래스의 특성에 대한 증명과 질의의 복잡성 평가에 대해서 연구되었다.

기존 데이터베이스 시스템에서 질의최적화의 내용은 CDB에서도 응용하여 연구되어 왔다. 기존 질의 처리의 최적화에서는 SQL 질의를 관계대수 연산으로 변환하여 실행방법을 기술했으며, 특히 선택과 조인 연산에는 최적화를 위한 많은 실행 방법이 제시되었다. 또 다른 연구 내용으로는 의미적 질의 최적화가 제안되었는데 이 기법은 하나의 질의를 더 효율적인 다른 질의로 변화하기 위하여 데이터베이스 스키마에 명시된 제약 조건들을 사용하는 것이다. 즉, 기존 SQL의 WHERE조건에 해당하는 제약 조건이 데이터베이스 스키마에 존재한다고 가정하는 것이다. 만일 제약 조건 검사가 효율적으로 수행될 수 있다면 상당한 시간이 절약되지만, 많은 제약 조건 중에서 주어진 질의에 적용될 수 있고, 최적화할 수 있는 조건을 찾는 데 많은 시간이 소모될 수 있다[11].

그 후 [5]의 논문에서는 관계대수를 통하여 프로젝션, 선택, 자연조인, 합집합 등의 연산을 구체화하여 상세

하게 논의했다. 이 논문에서 제안된 관계대수는 절차적인 언어이기 때문에 질의 평가에 쉽게 응용될 수 있었다.

효율적인 질의처리에 관하여 주요 문제는 사용자가 정의한 연산자는 최적화를 위해서는 블랙박스이다. 그래서 [12]논문에서는 CDB에서 사용자 정의 기능이 포함된 질의 처리 방법을 제안하였다.

[4]는 제약 질의처리에 대한 [5]의 이론제시와 함께 제약 릴레이션의 질의 처리 튜플 내용의 일부를 표로 나타내었으나 그 내용은 신뢰할 수 없는 오류가 포함된 튜플이며, 그 교집합 결과도 역시 오류가 있는 것을 제시하고 있었다.

기존의 GIS 소프트웨어는 공간 데이터의 안정적인 관리와 다양한 서비스를 위해 많은 기능들을 제공하지만, 저장 구조 및 공간 연산의 복잡성으로 인해 대용량 공간 데이터에 대한 검색이 비효율적이다[13]. 현재까지 GIS 질의처리와 관련하여 CDB의 프로토타입을 살펴보면, 선형제약 모델을 기반으로 하여 가장 최초로 구현한 것은 INRIA의 VERSO그룹과 CNAM의 VERTIGO그룹과의 합작 DEDALE이며, O2 DBMS 상에서 구현되었다[1].

그리고 DISCO(Database Logic with Integer Set Constraints)는 미국 Nebraska대학에서 개발했으며, DATALOG(Database Logic Program)로 구현했고, 정량자 제거 기술을 이 시스템에서 사용했다. 또한 기존 SQL에 시간데이터베이스를 추가하여 확장한 SQL/TP가 있으며, 상업용 DBMS로 제안되었으나 복잡하고 분명치 못한 구문 등의 이유로, 그 제안은 실패되었다. 또한 DOAS(Drought Online Analysis System)는 시공간 데이터베이스의 표현과, SQL 및 순환연산이 가능한 시스템이다. 순수한 제약 데이터베이스 모델에는 대수질을 응용한 CDB/CQA와 논리 프로그래밍을 위주로 한 MLPQ가 있다[5,15].

특히 MLPQ/PreSTO (Management of Linear Programming Queries and Parametric Rectangles Spatio-Temporal Objects)시스템은 미국 Nebraska-Lincoln 대학에서 개발되었으며, GIS 및 웹 응용의 시뮬레이션이 가능토록 한 시스템이다[14,16]. 이것은 기존 개발된 여러 프로토타입 중에서 순수 CDB 모델로 개발된 것이다[15]. 현재 유일하게 공개된 MLPQ 도구(ver 3.1)에서는 합집합(\cup) 연산자는 수행되나 교집합(\cap) 연산자는 아주 간단한 것만 처리되는 상태이다.

4. CDB의 질의처리와 그 문제점

그림 6과 그림 7은 폴리라인(poly-line)의 도로와 다각형(polygon)의 농장 형태의 공간객체를 나타낸 것이며, 그림 8은 향후 조성할 공원이란 공간객체를 나타낸 것이다. 그리고 그림 9는 도로, 농장 및 조성할 공원을 모두 나타낸 것이다. 그리고 그림 10은 튜플 제약식의 표현이다.

1) 다항식 부등제약, 선형밀집 부등제약, 무한 도메인상의 등식 제약, 불리안 제약

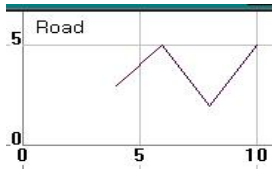


그림 6. 도로(폴리라인)

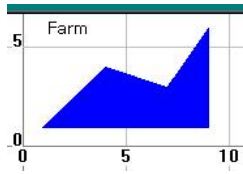


그림 7. 농장(다각형)

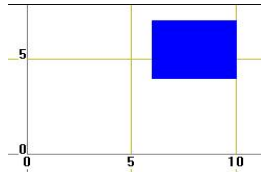


그림 8. 공원(사각형)

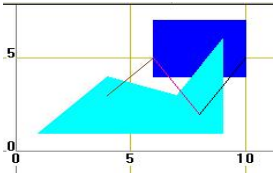


그림 9. 도로, 농장, 공원

<p>Road(id, x, y) :- id = 1, x >= 4, x <= 6, x - y = 1. Road(id, x, y) :- id = 2, x >= 6, x <= 8, 3x + 2y = 28. Road(id, x, y) :- id = 3, x >= 8, x <= 10, 3x - 2y = 20.</p>	<p>Farm(Const_0, x, y) :- Const_0 = 1, x - y >= 0, y >= 1, x + 3y <= 16, 3x - 2y <= 15. Farm(Const_0, x, y) :- Const_0 = 2, x <= 9, y >= 1, 3x - 2y >= 15. Park(id, x, y) :- id=1, y >= 4, y <= 7, x >= 6, x <= 10.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

그림 10. 도로, 농장, 공원에 대한 튜플

위의 공간객체에 대한 선택(selection, δ), 합집합(union, \cup), 교집합(intersection, \cap) 연산이 어떻게 처리되는지를 고찰해 보았다.

농장에 대한 y좌표가 4이하인 지역에 대한 선택(selection) 연산 처리 결과를 나타내 본다. 즉 $\delta_{y \leq 4}(\text{Farm})$ 연산의 결과는 Farm의 각 튜플에 “y <= 4” 라는 제약식이 추가된 튜플을 만들어 주면 된다[4].

합집합은 연합되는 각 객체 간에 논리합(disjunction, \vee) 연산자로 접속하면 DNF 형태를 유지한 새로운 합집합 객체가 생성된다. 결과 릴레이션에는 교집합에 해당하는 부분은 중복해서 선언될 수 있다. 즉, 점과 면, 선과 면, 면과 면의 합집합 연산 결과에서 교집합에 해당되는 부분은 그대로 유지되어 있기 때문에 중복이나 불필요한 요소는 소거해 줄 필요가 있다.

교집합은 면과 면의 경우에는 각 객체 간에 논리곱(conjunction, \wedge) 연산자로 접속하면 해결되지만, 선과 면의 경우에는 논리곱(\wedge) 연산자로 접속하게 되면, 교차

되는 선의 부등식과 면의 부등식이 작용하여 예상과는 전혀 다른 결과를 초래한다. [4]에서 제시된 표의 결과는 단순히 2개의 면과 3개의 선에 해당되는 튜플 제약식들 사이에 카티션 프로덕트(Cartesian product)형식을 취하였기에 왜곡된 결과를 나타내는 것을 확인했다. 또한 x 좌표에 대한 선택선 처리를 위해 y 소거 방식을 취할 경우에는 선분의 기울기가 사라지기 때문에 단순한 수직선 결과만 나타난다.

면과 면의 교집합(\cap) 연산의 경우에 단순하게 두 공간 객체에 해당하는 튜플 제약식을 나열하면, 중복 혹은 불필요한 제약식이 튜플에 존재할 수 있는 문제점이 발생한다. CDB의 교집합 질의처리 결과에 대한 불필요한 튜플을 최소화하는 기존 연구방법은 없었다. 이 논문에서의 교집합(\cap) 연산결과는 휴리스틱한 방법을 통해서 제시하고자 한다. 이 논문에서 질의처리 결과를 그림으로 표현한 것은 MLPQ 도구를 이용한 것이다.

5. 효율적인 질의 처리 방법 제안 및 처리 효과

CDB의 질의처리에서도 기존 공간 데이터베이스와 같이 대량의 기하연산에 관한 질의 처리를 향상시키기 위하여 여과-정제 방법을 채택하고 있다. 여과단계는 MBB (minimal boundary box)를 사용하여 개략적인 후보집합을 걸러내며, 이 단계에서 공간처리방법(SAM)이 효과적으로 지원하게 되며, MBB는 내포된 공간 속성으로 표현될 수가 있다. 정제단계에서는 여과된 객체가 질의구간과 겹치는지를 검사하며, 각 후보객체에 대한 정밀조사를 하게 된다. 이 단계에서 만족하는 객체가 실제 질의를 만족하는 객체로서 질의응답을 한다[1,17]. 정제단계는 여과단계를 거친 질의를 최적화기에 의해서 효율적인 질의평가 계획(QEP)을 산출하는 단계이다. 개략적인 그 실행은 제작성 규칙, 제어 전략, 비용 모델의 단계를 거친다.

제약 데이터 모델의 질의 평가는 다양한 방법으로 구현할 수 있다[1]. 질의 평가를 하지 않고 정의된 데이터를 다루는 기하적 계산 방식, 제약의 단순화된 표현의 해를 찾아내는 산술적 계산 방식, convex hull의 계산과 같은 기하적 문제로 해결하고자 하는 기하적 계산 방식, 위상적인 질의 해결을 위한 위상적 계산 방식 등이 있다.

이 논문에서는 3장에서 제기된 교집합(\cap) 연산의 문제점 중에 선과 면의 왜곡된 결과와 면과 면의 연산에서 중복 및 불필요 제약식 소거에 대해서만 다루기로 한다.

교집합 연산 처리를 위한 튜플 제약식의 최소화 기법을 다음과 같이 정리한다.

- ① 제약 튜플의 스캔 및 계수 정리
 - 교집합을 위한 각 객체 간에 카티션 프로덕트(X) 연산을 수행하여 새로운 튜플 생성함.
 - 제약식 속에 관계연산자를 기준으로 x, y계수는 왼쪽 항에 상수는 오른쪽에 위치함.

2) 텍스트 p.131의 “The intersect of Road and Spat” 잘못된 교집합 연산결과를 제시했음

- x계수를 양수로 변환함(음수일 경우 양변에 -1을 곱하고, 관계연산자를 변경 $\geq \leftrightarrow \leq$).
- 만약 y계수만 단독으로 존재하면서 음수일 경우에도 양수로 변환함.
- ② 선과 면의 교집합 연산의 경우
 - 선의 경우 하나의 버텍스(vertex)는 최소 2개의 제약식이 존재함.
 - 기울기가 있는 경우 3개의 제약식이 존재함.
 - 이들 2~3개의 제약식과 다각형의 각 제약식 간에 부등방정식에 의한 교차점 산출
 - 다각형 내부로 진입하는 선의 진입좌표 산출(부등방정식에 의한 교차점 산출)
 - 선에 대한 제약식은 산출된 교차점을 고려하여 갱신하여 작성함.
- ③ 면과 면의 교집합 연산의 경우
 - 예를 들면, 향후 공원조성을 위해서 농장이 얼마나 수용되는지는 면과 면의 연산임.
 - 선과 면의 경우와 같이 면과 면의 제약식 간에 교차점을 산출함.
 - 새로운 다각형 제약식을 도출하여 적용함.
- ④ 중복 및 불필요 제약식의 소거
 - x 혹은 y의 단항식 관계연산자가 \geq 일 경우 보다 적은 수의 관계식은 소거함.
 - x 혹은 y의 단항식 관계연산자가 \leq 일 경우 보다 큰 수의 관계식은 소거함.

선과 면의 교집합 연산에서 그림 6의 도로와 그림 7의 농장 간에 교집합 연산 결과는 그림 11과 같으며, 해당되는 제약식은 그림 12와 같이 간단해 진다.

그림 8과 같은 직사각형 공원을 새로 조성할 경우에 그림 7의 농장이 어느 정도 흡수되는지 교집합 연산으로 했을 경우에, 단순한 교집합 결과 투플은 그림 13과 같이 표시할 수 있으며, 그림 14는 MLPQ를 이용하여 나타낸 것이다.

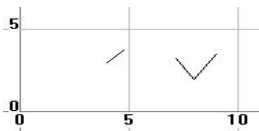


그림 11. Road ∩ Farm

RoadXFarm(i, x, y) :- i = 1, x >= 4, x <= 4.75, x - y = 1.	$3x + 2y = 28.$ RoadXFarm(i, x, y) :- i = 3, x <= 9, x >= 8, $3x - 2y = 20.$
RoadXFarm(i, x, y) :- i = 2, x >= 7.17, x <= 8,	

그림 12. Road ∩ Farm 최소화된 투플

ParkFarm(i, x, y) :- i = 1, y >= 4, y <= 7, x >= 6, x <= 10, x - y >= 0, y >= 1, x + 3y <= 16, 3x - 2y <= 15.	ParkFarm(i, x, y) :- i = 2, y >= 4, ===③ y <= 7, ===④ x >= 6, ===④ x <= 10, ---① x <= 9, y >= 1, ---② $3x - 2y >= 15.$
------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------

그림 13. Park ∩ Farm 투플

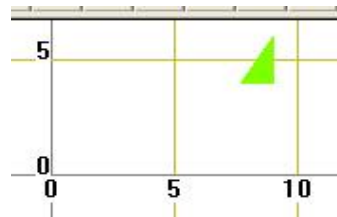


그림 14. Park ∩ Farm

그림 13의 i=1의 투플은 도형이 생성되지 않는 투플이며, i=2의 투플은 그림 14에 해당되는 투플임을 알 수 있다. 그 투플에서 ① $x \leq 10$, ② $y \geq 1$ 은 불필요한 제약식이기 때문에 소거해야 한다.

교집합의 결과는 직각 삼각형(당초 농장의 일부분)이기 때문에 ③ $y \leq 7$, ④ $x \geq 6$ 도 불필요한 투플이 되기에 소거하면, $y \geq 4$, $x \leq 9$, $3x - 2y \geq 15$ 의 세 개의 제약식으로 표현이 가능하다. 그래서 표 3과 같이 모두 15개의 제약식에서 결과 투플에는 3개의 제약식만 남게 된 것이다.

표 3. Park ∩ Farm 질의처리에 대한 투플 최소화 기법의 적용 효과

Park ∩ Farm 공간연산	당초 형태	적용후	비고
투플의 수	2	1	1개의 다각형만 교차됨
제약식의 수	15	3	삼각형 결과물

이와 같은 방법으로 다양한 다각형들 간의 교집합 연산 실험 결과는 표 4이며, 그의 비교그래프는 그림 15와 같다. 꺾은선은 당초 제약식의 수를 나타내었으며, 막대 그래프는 최소화 기법을 적용 후에 최소(공집합은 제외) 및 최대 제약식의 수를 나타낸 것이다.

제안된 투플 최소화기법의 자체 실험결과로 제시된 표 3, 표 4 및 그림 15를 종합해 보면, 다각형 간에 교집합 연산에 있어서 최소 제약식의 형태는 삼각형이 만들어지며, 최대는 당초의 두 다각형의 제약식을 합한 수를 초과하지 않음을 알 수 있었다. 또한 다각형의 각이 많아질수록 최소와 최대의 폭이 증가하며, 최소화 기법에 의한 제약식의 수가 간소화할 확률이 높다는 것을 알 수 있었다.

표 4. 다각형들 간에 교집합 연산과 제약식의 수의 변화

다각형 구분	당초 제약식의 수	적용 후 제약식의 수	
		min	max
삼각형 ∩ 삼각형	6	3	6
삼각형 ∩ 사각형	7	3	7
사각형 ∩ 사각형	8	3	8
사각형 ∩ 오각형	9	3	8
육각형 ∩ 육각형	12	3	12
팔각형 ∩ 팔각형	16	3	16

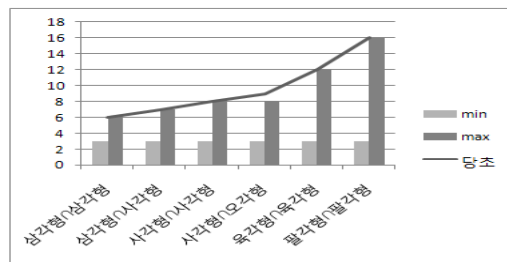


그림 15. 다각형 교집합 연산 결과 제약식의 변화

6. 결론

제약 데이터베이스를 구성하는 튜플들은 제약 논리식으로 되었으며, 그에 따라 무한한 점집합이라는 추상표현이 가능하기 때문에 공간 데이터의 표현 및 질의를 비교적 간결하게 처리할 수 있었다. 제약 데이터베이스를 통한 공간 데이터의 선택, 합집합, 교집합 등의 질의연산 처리는 간단하게 관련 튜플 간에 제약식을 포함시켜 주면 되었다. 그러나 선과 면의 교집합에서는 단순히 제약식을 포함시킬 경우에 왜곡현상이 나타나기 때문에 선 중심의 새로운 해를 도출해야만 한다. 또한 중복되거나 불필요한 제약식의 증가는 질의 비용이 상승하게 된다.

질의 처리 방법을 위하여 교집합 연산 처리를 위한 튜플 제약식의 최소화 기법을 제안했으며, MLPQ 도구를 이용하여 CDB의 질의처리 결과를 표현해 보았다. 질의 처리 과정에서 적용해 본 결과 선과 면의 교집합 연산에서 왜곡현상이 없어짐을 알 수 있었고, 불필요한 제약식을 소거함에 따라 튜플이 간소해 짐을 알 수 있었다. 즉, 공원과 농장의 교집합 연산에서 단순히 제약식을 포함할 경우 2개의 튜플 중에는 총 15개 제약식이 포함되어 있었으나, 간소화 후에는 실제 결과에 해당되는 1개의 튜플 속에 3개의 제약식만 남게 되는 것을 알 수 있었다. 우리는 튜플 최소화 기법을 이용하여 제약 릴레이션의 불필요한 제약식을 제거함에 따른 질의처리의 효율성이 개선됨을 확인할 수 있었다. 다양한 다각형 간에 교집합 연산에 있어서 자체 실험결과에서 얻어진 내용을 종합하

면, 최소 제약식의 형태는 삼각형이 만들어지며, 최대는 당초의 두 다각형의 제약식을 합한 수를 초과하지 않음을 알 수 있었다. 또한 다각형의 각이 많아질수록 최소와 최대의 폭이 증가하며, 최소화 기법에 의한 제약식의 수가 간소화할 확률이 높다는 것을 알 수 있었다.

향후 연구로는 공간 센서 데이터의 효율적인 처리를 위한 공간 DBMS 개발[18], 공간 네트워크 데이터와 관련된 질의처리[19], 시간이력 및 데이터마이닝[20,21] 등의 기존 연구에 제약 데이터베이스 및 질의처리 기술의 접목가능성과 그 효율성을 연구하는 것이다.

참고 문헌

- [1] Kuper, G., Libkin, L., and Paredaens J., Constraint Databases, Springer-Verlag, Berlin Heidelberg, 2000.
- [2] Revesz, P. "Constraint Databases: A Survey," Semantics in Databases, vol. 1358, 1996, pp. 209-246.
- [3] Belussi, A., Bertino, E., and Catania, B., "Manipulation Spatial Data in Constraint Databases," Proc. 5th International Symposium on Spatial Databases, Vol.1262, 1997, pp. 115-141.
- [4] Rigaux, P., Scholl, M., and Voisard, A., Spatial Databases With Application to GIS, Morgan Kaufmann Pub., San Francisco. 2002.
- [5] Kanellakis, P., and Goldin, D., "Constraint Programming and Database Query Languages," Proc. Int'l Symp. Theoretical Aspects of Computer Software, Lecture Notes in Computer Science 789, 1994. pp. 96-120.
- [6] Kanellakis, P., Kuper, G., and Revesz, P., "Constraint Query Languages," Journal of Computer and System Sciences, Vol. 51, no. 1, 1995, pp. 25-52.
- [7] Revesz, P., "Safe Query Languages for Constraint Databases," ACM Transactions on Database Systems, vol. 23, no. 1, 1998. pp. 58-99.
- [8] Chomicki, J., and Revesz, P., "Constraint-Based Interoperability of Spatiotemporal Databases," Geoinformatica, Vol. 3, no. 3, 1999, pp. 211-243.
- [9] SungKu, W., KeunHo, Y., "Manipulation of Geospatial Data based on Constraint Database," Proceedings of International Conference on East-Asian Language Processing and Internet Information Technology 2002 (EALPIIT2002), Hanoi Vietnam, January 8-11, 2002. pp. 71-74.
- [10] Vandeurzen, L., Gyssens, M., and Gucht, D., "On Query Languages for Liner Queries Definable with Polynomial Constraints," In proc. Init. Conf. on Principles and Practice of Constraint

- Programming, LNCS1118, 1996. pp. 468-481.
- [11] Elmasri, R., and Navathe, S., Fundamentals of Database Systems 5th ed., Pearson International Edition, Boston, 2007.
- [12] Gaede, V., and Gunther, O., "Constraint-Based Query Optimization and Processing," Constraint Databases and Applications, ESPRIT WG CONTRSSA Workshop Friedrichshafen, Germany, September 8-9, 1995. pp. 84-101.
- [13] 이기영, 김동오, 신중수, 한기준, "대용량 공간 데이터의 효율적인 검색을 위한 공간 미들웨어의 개발," 한국공간정보시스템학회 논문지, 제10권 제1호, 2008, pp. 1-14.
- [14] Revesz, P., MLPQ/PreSTO User's Manual, University of Nebraska-Lincoln, <http://cse.unl.edu/~revesz/MLPQ/Instruction.pdf>, November 18, 2004.
- [15] Goldin, D., Taking Constraints out of Constraint Databases, - Applications of Constraint Databases - Paris, France(presentation file), June 2004.
- [16] Revesz, P., Chen, R., Kanjamala, P., Li, Y., Liu, Y., and Wang, Y., "The MLPQ/GIS Constraint Database System," In Proc. ACM SIGMOD International Conference on Management of Data, 2000, P. 601.
- [17] Manolopoulos, Y., Papadopoulos, A., Vassilakopoulos, M., Spatial Databases: Technologies, Techniques and Trends, Idea Group Publishing, Hershey, 2005.
- [18] 강홍구, 홍동숙, "공간 센서 데이터의 효율적인 실시간 처리를 위한 공간 DBMS의 개발," 한국공간정보시스템학회 논문지, 제9권 제1호, 2007, pp. 1-13.
- [19] 김용기, 김아름, 장재우, "공간 네트워크 데이터베이스에서 공간 제약을 고려한 경로 내 최근접 질의처리 알고리즘," 한국공간정보시스템학회 논문지, 제10권 제3호, 2008, pp. 19-30
- [20] 지정희, 김승관, 류근호, 김명준, "수치지도를 위한 피쳐 기반 공간자료관리 시스템 설계," 한국공간정보시스템학회 논문지, 제7권 제3호, 2005, pp. 107-118.
- [21] 이양구, 김원태, 정영진, 김광득, 류근호, "날씨 마케팅 적용을 위한 기후 데이터의 군집 분석," 한국공간정보시스템학회 논문지, 제7권 제3호, 2005, pp. 33-44.



우 성 구

1986,1990,1999년 한국방송통신대학교 경영학과(경영학사), 전산학과(이학사), 영문학과(문학사)
 1992년 연세대학교 공학대학원 전산전공(공학석사)
 2001년 충북대학교 대학원 전산학과(박사과정수료)
 1978년~1998년 (주)삼화 전산실, 현대엔진공업(주) 전산실, 해외건설협회 전산실(과장), (주)시에치노 시스템컨설팅(과장), (주)경동보일러 전산실(부장)
 1999년~2008년 (주)ASB SI사업부(이사), (주)지아이에스(수석연구원)
 1996년~2008년 남서울대학교 컴퓨터학과 겸임교수
 1998년~현재 평택대학교 컴퓨터학과 겸임교수
 2001년~현재 한서대학교 컴퓨터정보공학과 겸임교수
 관심분야는 Constraint 데이터베이스, 시간 데이터베이스, 공간 데이터베이스, 시공간 데이터베이스



류 근 호

1976년 숭실대학교 전산학과(이학사)
 1980년 연세대학교 공학대학원 전산전공(공학석사)
 1988년 연세대학교 대학원 전산전공(공학박사)
 1976년~1986년 육군군수 지원사 전산실(ROTC 장교), 한국전자통신 연구원(연구원), 한국방송통신대 전산학과(조교수) 근무
 1989년~1991년 Univ. of Arizona Research Staff (TempIS 연구원, Temporal DB)
 1986년~현재 충북대학교 전기전자 및 컴퓨터공학부 교수
 관심분야는 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS, 지식기반 정보검색 시스템, 유비쿼터스컴퓨팅 및 스트림 데이터 처리, 데이터마ining, 데이터베이스 보안, 바이오인포매틱스 등