

# MLPQ를 이용한 제약데이터베이스의 공간데이터 연산

우성구<sup>0</sup>, 문경도, 정지문, 류근호  
충북대학교 데이터베이스 연구실

{wsk527, kdmoon67}@chollian.net, jmchung@nsu.ac.kr, khryu@dblab.chungbuk.ac.kr

## Manipulating Spatial Data in Constraint Database using MLPQ

SungKu Woo<sup>0</sup>, KyungDo Moon, JiMoon Chung, KeunHo Ryu  
Dept of Computer Science, Chungbuk National University

### 요 약

이 연구에서는 GIS 등에 적용할 수 있는 제약 데이터베이스의 공간데이터 연산을 구현하였다. 공간데이터를 간결하게 표현할 수 있는 새로운 패러다임인 제약 데이터베이스의 구조와 특징을 검토하고, 제약 데이터 모델용으로 개발된 MLPQ 프로토타입을 이용하여 제약 데이터베이스의 지리정보 모델을 구현하였다. 아울러 구현된 모델을 사용자 질의에 따라 공간질의를 처리하도록 하였다.

### 1. 서 론

지리정보시스템과 같은 대용량 공간데이터 처리에 대한 새로운 패러다임이 바로 제약 데이터 모델이며, 그에 관한 많은 연구가 활발히 전개되고 있다는 사실을 해외 관련 논문 및 저서를 통하여 알 수 있다.

제약 데이터베이스(Constraint Database, CDB)는 1990년에 Kanellakis 등을 통하여 그 이론이 발표되었으며, 관계 데이터베이스(Relational Database, RDB)에서 제약의 논리적인 접속으로 구성된 튜플 표기의 대치가 가능하기 때문에 관계모델을 다양한 방법에 의해서 여러 형태로 그 응용을 확장할 수 있다.

특히 제약데이터 모델에서의 질의처리 표현은 관계 데이터 모델에서와 마찬가지로 관계해석 및 관계대수와 같은 질의언어들의 접근이 가능하며, RDB 모델의 뷰 질의와 마찬가지로 그의 질의처리 결과는 제약 데이터 모델을 출력할 수 있는 닫힘(closure)의 특징이 있음[3,4]을 알 수 있다.

CDB모델은 기존 관계데이터베이스와의 유사한 특징으로 용이하게 그 응용이 가능하며, 각종 대학교 및 연구기관에서는 그에 대한 프로토타입을 발표하고 있다. 현재까지 발표된 CDB모델의 프로토타입(prototype)은 MLPQ, DEDALE, DISCO, SQL/TP, UMB-CDB, C3, COSMOS 등[1,4,7,8]이 있으며, 이 논문에서는 MLPQ를 사용하여, 모의실험용으로 가상의 고속도로와 주요 인터체인지 및 명승지 등을 표현한 지리정보 모델을 구축한다. 그리고 구현된 모델을 사용자 질의에 따라 수행되는 공간질의 처리과정을 기술한다.

### 2. 반 평면 표현

모든 공간객체는 반 평면으로 표현될 수 있다.  $d$ 차원  $R^d$  공간에 있는 반 공간(half-space)  $H$ 는 점(point)들의 집합  $P(x_1, x_2, \dots, x_d)$ 로 정의할 수 있고, 그것은  $a_1x_1 + a_2x_2 + \dots + a_dx_d + a_{d+1} \leq 0$  부등식 형태에 만족한다. 그러므로  $H$ 의 표현은 벡터  $[a_1, a_2, \dots, a_{d+1}]$ 이다. 그래서 볼록  $d$ 차원의 다공간(polytope)은 어떤 닫혀진 반 공간의 무수한 교차로서 정의된다. 어떤  $n$  벡터의 볼록다각형은 직선에 의해 경계가 정해진  $n$ 개 반 평면의 교차로서 정의되고, 영역(region)은 볼록다각형의 합집합(union)으로 표현된다. 또한 라인 세그먼트는 1차원의 다공간(polytope)이며, 그것은 두 개의 반 직선(half-lines) 혹은 선(rays)의 교차에 의해 정의되고, 각각은 세그먼트의 끝점에 의해서 구분된다[7].

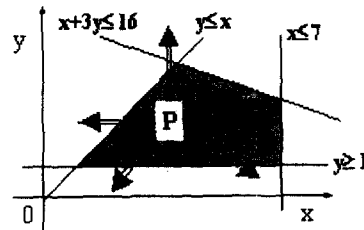


그림 1. 폴리곤의 표현

반 평면 표현에 의한 폴리곤은 그림 1의 예와 같이 네 개의 반 평면 교차에 의해서 정의된다. 즉, 폴리곤은 각 부등식 연산의 교차된 영역 P를 의미한다

3. 공간데이터의 무한 점좌표의 유한한 표현

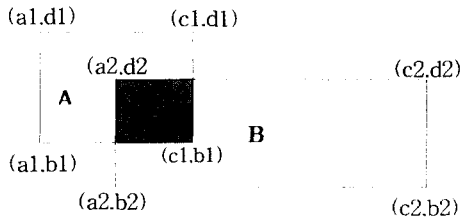


그림 2. 두 개의 직사각형

그림 2와 같이 두 개의 직사각형에 대한 이름과 좌표에 관한 공간데이터를 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)가 존재하는가?”를 해석 질의로서 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,

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

여기서 intersects(a, b, c, d, a', b', c', d')는 n과 n'의 교차된 사각형에 해당되는 사실(fact)의 표현이다. 그런데 이 교차(intersect)라는 프레디카트(predicate)를 다음같이  $\exists x \exists y (a < x < c \wedge b < y < c' \wedge a' < x < c' \wedge b' < y < d')$ 로 표현하면, 이 식은 작동하지 않는다. 그 이유는 RDB 모델의 정량자의 값 x, y가 존재하지 않기 때문이다. 그래서 한 사각형 모서리가 다른 사각형 내에 있다는 필요 충분조건이 되도록 두 사각형의 교차 연산을 다음과 같이 표기해야만 그 해가 가능하다.

$$(a < a' < c \wedge b < d' < d) \vee (a < a' < c \wedge b < b' < d) \vee (a < c' < c \wedge b < b' < d) \vee (a < c' < c \wedge b < d' < d) \vee (a' < a < c' \wedge b' < d < d') \vee (a' < a < c' \wedge b' < b < d') \vee (a' < a < c' \wedge b' < b < d') \vee (a' < c < c' \wedge b' < d < d')$$

질의언어를 이렇게 표현한다는 것은 바람직하다고 볼 수 없다. 좀 더 복잡한 지리객체를 고려할 때 문제를 해결하는 방안으로서는 사각형내의 모든 점의 좌표를 포함하는 개념이다. 즉 R(n, x, y)의 릴레이션으로 표 2의 좌

측 x,y의 무한 표현으로 구성한다면, 앞에서의  $Q_1$ 과  $Q_2$ 의 표현은 다음과 같이 간단하게 표현할 수 있다[4].

$$Q_1 \equiv R(x, y)$$

$$Q_2 \equiv \{(n1, n2) \mid n1 \neq n2 \wedge \exists x \exists y (R(n1, x, y) \wedge R(n2, x, y))\}$$

표 2. 직사각형 릴레이션 - 무한표현과 유한표현

n	x,y의 무한표현		제약식에 의한 유한표현
	x	y	
n1	x x' x'' ...	y y' y'' ...	$x \geq a1 \wedge x \leq c1$ $\wedge y \geq b1 \wedge y \leq d1$
n2	x1 x1' x1'' ...	y1 y1' y1'' ...	$x \geq a2 \wedge x \leq c2$ $\wedge y \geq b2 \wedge y \leq d2$

표 2는 사각형 n에 대해서 왼쪽은 각 개체별 내포된 무한의 점 좌표를 표현한 것이며, 오른쪽은 제약식으로 된 튜플을 표현한 것이다. 즉, CDB는 RDB의 확장된 개념으로 다양하게 사용할 수 있다는 것이다.

4. MLPQ: CDB 프로토타입

CDB모델의 프로토타입에는 현재까지 MLPQ, DEDALE, DISCO, SQL/TP, UMB-CDB, C<sup>3</sup>, COSMOS 등[1,4,7,8]이 발표된 상태이다.

MLPQ는 Nebraska-Lincoln 대학교에서 구현한 것으로서, GUI 환경으로 선형 제약 모델에 기반을 두고, 지리 시공간 데이터의 표현이 가능토록 구현된 시스템이다. MLPQ[1,8]는 Management of Linear Programming Queries의 약어이며, 지리객체와 관련하여 Select, Line, Rectangle, Polygon, Intersection, Union, Difference, Complement, Area, Buffer, Query, Maximum, Minimum, Block, Collide, Similarity등의 기능 아이콘이 있다. 지리객체는 문법에 의해 작성하거나 혹은 윈도우 창에 마우스로 직접 그릴 수가 있다. MLPQ의 프로토타입은 주로 텍스트 파일을 입력 및 저장구조로 사용하고 있다.

그 문법은 begin %moduleName%과 end %moduleName% 사이에 L<sub>1</sub>, L<sub>2</sub>, ..., L<sub>n</sub>을 표시하며, L<sub>n</sub>은 규칙(rule) 혹은 제약튜플을 나타내며, 제약식은 a<sub>1</sub>x<sub>1</sub> + a<sub>2</sub>x<sub>2</sub> + ... + a<sub>n</sub>x<sub>n</sub> θ b이며, a<sub>i</sub>는 상수, x<sub>i</sub>는 변수, θ는 관계연산자 =, <, >, <=, >=를 각각 지칭한다.

5. 제약데이터베이스에서 공간 데이터 연산

MLPQ를 이용한 모의실험용 가상의 관광지도는 3개의 매트릭스로 구성된 고속도로인 폴리라인과 서울(21), 수원(22), 대전(23), 영동(24), 대구(25), 경주(26), 부산(27) 등 7개의 인터체인지인 점좌표를 표시하였다(괄호 안은 Id를 지칭하며, 시각적으로 인식하도록 짧은 라인으로 표시됨). 또한 지도상에는 8곳(31~38번 id)에 분포된 국공립 공원인 폴리곤을 각각 표시하였다.

**시나리오와 질의:** 고속도로 상의 승용차는 지금 부산을 향하여 운행하고 있으며, 대전을 떠난 현재의 위치를 GPS (Global Point System)로 받은 결과  $x = 105.58, y = 222.09$  지점이란 것을 인식했다. 이제 가까운 인터체인지가 있으면 고속도로를 벗어나서 인근 30km주위에 있는 국공립 공원이 있는가를 검색하려고 한다. 가장 가까운 IC에 대한 질의는 다음과 같이 부여한다.

**Nearest\_IC(id, MIN(x), y):-**  $x - x0 \geq 0,$   
**Current\_Pos(id2, x0, y0), IC(id, x, y).**

그 결과는 **Nearest\_IC(id, MIN601, y) :- id = 24,**  
 $x = 151.10, y=212.99.$  이며, 그것은 24번 영동인터체인지가 가장 가까운 인터체인지임을 나타내고 있었다.

그 지점에서 인근 30km Buffer연산 결과는 점으로부터 30km의 8각형의 폴리곤이 만들어진다. 형성된 버퍼 폴리곤 Buf\_Nearest\_IC와 공원 Park 개체와의 동일지점이 있는지를 찾아내는 질의문은 다음과 같다.

**Nearest\_Park(id,x,y) :-** **Park(id, x, y),**  
**Buf\_Nearest\_IC(id1, x, y).**

질의결과는 **id = 36, x >= 152.61, x <= 158.76, y >= 238.76, y <= 243.68** 폴리곤의 인근공원을 찾았다.

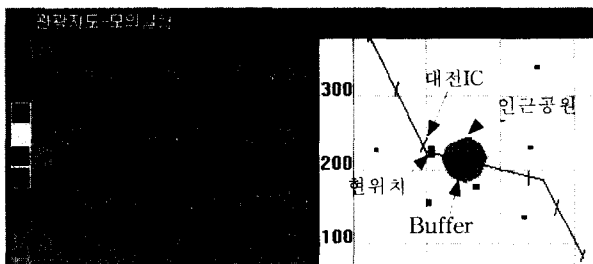


그림 3. 공간질의의 연산결과

최종 질의결과는 그림 3과 같으며, 찾아낸 인근공원 폴리곤에 대한 제약식은 표 3과 같다. 즉 인근공원 Nearest\_Park(id, x, y)는 8각형의 Buf\_Nearest\_IC와 Park의 교집합(intersection)연산을 지칭한 것이므로 2개

의 제약식의 합집합으로 표현된 것임을 알 수 있다.

표 3. 인근공원에 대한 제약식

$\begin{aligned} \text{Nearest\_Park}(id, x, y) :- & \quad id = 36, \\ & \quad x \geq 152.61, \quad x \leq 158.76, \\ & \quad y \geq 238.76, \quad y \leq 243.68, \\ & \quad x + 0.41y \leq 269.32, \quad x + 2.41y \leq 737.73, \\ & \quad -x + 2.41y \leq 435.53, \quad -x + 0.41y \leq -32.88, \\ & \quad -x - 0.41y \leq -209.32, \quad -x - 2.41y \leq -592.88, \\ & \quad x - 2.41y \leq -290.68, \quad x - 0.41y \leq 92.88. \end{aligned}$
--

6. 결론

CDB는 무한 개념의 데이터를 유한한 제약 데이터 표현으로 간결하게 처리할 수 있으며, 관계데이터 모델의 확장된 형태로 연구 개발되었기 때문에 그 응용에 있어서 용이하게 접근 및 확장이 가능하다.

이 논문에서 MLPQ를 이용하여 간단한 모의실험용 지리정보를 구축하고, 시나리오에 따라 공간질의를 테스트 해 본 결과, 구축 및 지리정보의 검색은 간단히 처리할 수는 있었지만, 다양한 연산에 대해서는 프로토타입의 한계도 많이 발견되었다. 향후 연구계획은 현재 발표된 여타 프로토타입의 적용 및 비교분석을 하여 새로운 CDB 모델을 제안하는 것이다.

참고문헌

- [1] Kanjamala, Pradip, Revesz P.Z., and Wang, Younghui, MLPQ/GIS: "A GIS using Linear Constraint Databases"
- [2] Kanellakis, P.C. and Goldin, D.Q., "Constraint Programming and Database Query Languages", vol. 789 of Lecture Notes Science, pp.96-120. Springer-Verlag, 1994.
- [3] Kanellakis, P.C., Kuper, G.M., and Revesz, P.Z., "Constraint Query Languages", Journal of Computer and System Sciences, 51(1):26-52, 1995. A shorter version appeared PODS'90.
- [4] Kuper, G., Libkin, L., and Paredaens, J. (Eds.), *Constraint Databases*, Spinger-Verlag, Berlin, 2000.
- [5] Laurini, Robert and Thompson, Derek, *Fundamentals of Spatial Information Systems*, Academic Press, London, 1995.
- [6] Marriott, Kim and Stuckey, Peter, *Programming with Constraint : An Introduction*, Mit Pr., pp.391-423, 1998.
- [7] Philippe Rigaux, Michel Scholl, and Agnès Voisard, *Spatial Databases With Application to GIS*, Morgan Kaufmann Pub., 2001.
- [8] Revesz, P., Chen, R., Kanjamala, P., Li, Y., Liu, Y., and Wang, Y., "The MLPQ/GIS Constraint Database System", 2000.