

공간 네트워크 데이터베이스를 위한 저장 및 색인 구조의 설계

강홍민^o, 장재우
전북대학교 컴퓨터공학과
{khmsol^o, jwchang}@chonbuk.ac.kr

Design of Storage and Index Structures for Spatial Network Databases

Hong-Min Kang^o, Jae-Woo Chang
Dept of Computer Engineering, Chonbuk National University

요약

이동 객체를 위한 기존 연구는, 제한조건이 없는 이상적인 공간을 가정하고 설계되었으나, 이는 실제 응용에 직접 적용하는데 문제를 지니고 있다. 따라서 본 논문에서는 LBS (location-based service)의 효과적인 지원을 위해, 이상적인 공간대신, 실제 도로나 철도와 같은 공간 네트워크(network)를 고려하며, 아울러, 이러한 공간네트워크 데이터베이스를 위한 효율적인 저장 및 색인 구조를 설계한다.

1. 서론

최근 LBS (location-based service)의 효과적인 지원을 위해, 이상적인 공간 대신, 실제 도로나 철도와 같은 공간 네트워크(network)을 고려한 연구가 활발하게 수행 중에 있다 [1, 2, 3, 4, 5]. 이러한 공간네트워크 데이터베이스 연구는 일반 공간 데이터베이스의 연구와 마찬가지로 크게 3가지 주제로 요약된다. 즉, 공간 네트워크를 위한 데이터 모델, 질의 처리 알고리즘, 마지막으로 공간 네트워크 데이터베이스를 위한 저장 및 색인 구조이다. 이 가운데 저장 및 색인 구조에 관한 연구는 공간 네트워크 데이터베이스의 전체 성능을 좌우하는 중요한 연구이다. 따라서 본 논문에서는 공간네트워크 데이터베이스를 위한 효율적인 저장 및 색인 구조를 설계한다. 이를 위해 저장할 데이터를 크게 3가지, 즉 공간 네트워크 자체의 데이터, 공간 네트워크 상에 존재하는 POI (point of interest) 데이터, 이동 객체의 데이터로 분류한다. 이러한 분류에 근거하여, 첫째 공간 네트워크 자체의 데이터를 효율적으로 저장하기 위해 노드 및 에지 정보를 지니는 공간 네트워크 파일 구조를 설계한다. 둘째, POI 데이터에 대한 효율적인 접근을 위해 POI 색인 구조를 설계한다. 마지막으로, 이동 객체 데이터에 대한 저장 및 효율적인 쿼리 검색을 위해 객체 색인 구조를 설계한다. 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고, 3장에서는 본 논문에서 설계한 공간 저장 및 색인 구조를 제시한다. 4장에서는 공간 저장 및 색인 구조의 구현에 대해 언급하고, 마지막으로 5장에서는 결론을 제시한다.

2. 관련연구

공간 네트워크는 그 자체가 네트워크 구조를 지니고 있으므로, 본 연구는 대학 JT연구센터 육성·지원사업의 연구결과로 수행되었음

공간 네트워크 데이터의 저장이 필수적이다. 일반적으로 이는 그래프 형태로 변환이 용이하므로, 그래프 형태로 변환한 후 저장하는 연구들이 수행되었다 [5, 6]. 공간 네트워크 데이터베이스를 위한 저장 및 색인 구조에 대한 관련 연구는 크게 다음과 같다. 첫째, Univ. of Minesota의 연구에서는, 공간 네트워크를 $G=(N, E)$ 의 그래프 형태로 변환하였다 [6]. 여기서 N 은 노드의 집합을, E 는 에지의 집합을 나타낸다. 아울러 변환된 그래프를 Z -order로 표현하고 연결에 따른 cluster를 형성하여, CCAM의 접근기법에 저장한다. 아울러 이들 노드들에 대해 $B+$ -tree를 구성하여, 노드 ID를 통해 이에 대한 접근을 빠르게 한다.

둘째, HKUST의 연구에서는 도로나 철도와 같은 공간 네트워크를 앞의 연구와 마찬가지로 연결에 따른 cluster를 형성하여 노드들의 adjacent list로 저장한다 [5]. 단, 지역적 clustering 효과를 높이기 위해 Hilbert order를 사용한다. 아울러, 에지들을 polyline으로 저장하며, 이들에 대한 공간 질의를 위해 network R-tree를 사용한다.

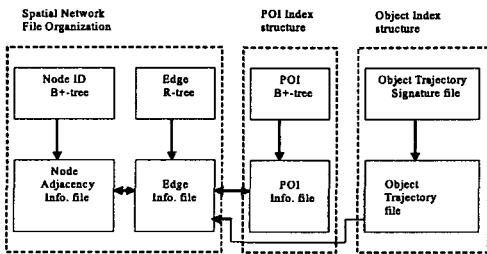
3. 공간 저장/색인 구조의 설계

공간 저장/색인 구조의 설계 목적은 공간 네트워크 데이터 및 공간 네트워크 상의 객체 데이터를 효율적으로 저장하기 위한 파일 구조를 설계하고, 아울러 네트워크 상의 객체 데이터를 효율적으로 검색하기 위한 색인 구조를 설계하는 것이다. 이를 위하여 저장할 데이터를 크게 3가지로 분류한다. 첫째, 공간 네트워크 자체의 데이터와 둘째, 공간 네트워크 상에 존재하는 POI (point of interest) 데이터, 마지막으로 이동객체의 데이터이다.

3.1 공간 네트워크 자체의 데이터

기존의 제한조건이 없는 이상적인 공간과는 달리, 공간 네트워크

크는 그 자체가 네트워크 구조를 지니고 있으므로, 이에 대한 데이터의 저장에 필수적이다. 일반적으로 이는 그래프 형태로 변환이 용이하므로, 그래프 형태로 변환한 후 저장하는 CCAM 및 HKUST의 연구 등이 수행되었다. 따라서 본 논문에서는 CCAM 및 HKUST의 저장 기법의 장점을 결합한 저장/색인 구조를 설계한다. 즉, 노드를 위한 저장 파일에, 노드 자체의 정보뿐만 아니라 노드와 에지(edge)간의 관계성 정보를 포함하며, 아울러 노드에 대한 빠른 접근을 위해 B+-tree를 구성한다. 한편, 에지를 위한 저장 파일에 에지 자체에 대한 정보뿐만 아니라, 에지상에 존재하는 POI들의 list 정보를 지닌다. 또한 에지에 대한 효율적인 공간 질의 처리를 위해, 에지에 대해 R-tree를 구성한다. (그림 1)은 공간 네트워크 데이터베이스를 위한 전체적인 저장/색인 구조를 나타낸다.



(그림 1) 공간 네트워크 데이터베이스를 위한

전체적인 저장/색인 구조

(1) Node Adjacency Information file

노드 자체에 대한 정보 및 노드와 에지간의 관계성을 나타낸다.

$$li = \langle ni, nj, len, Pj \rangle \mid e(ni, nj) \in Ei, i \neq j$$

- li : 노드에 대한 연결 에지와의 리스트,
- e(ni, nj) : 노드 ni, nj 간의 에지,
- Ei : ni와 incident한 edge의 집합

(2) Node ID B+-tree

하나의 노드 정보에 대한 빠른 접근을 위하여 노드 ID에 따른 B+-tree를 구성한다.

(3) Edge Information file

노드 ni, nj 사이의 에지 e(ni, nj)에 대한 에지 정보 Re(ni, nj)는 다음과 같이 구성된다.

$$Re(ni, nj) = \langle Eid, ni, nj, len, MBR(ni, nj), direction, \# \text{ of POIs, SPpoi} \rangle$$

- Eid : 에지 ID, len : 에지 길이
- MBR(ni, nj) : 에지에 대한 2차원 좌표,
- direction : 에지의 방향,
- # of POIs : 에지상에 존재하는 POI의 전체개수,
- SPpoi : POI Info file 내에서의 시작점

(4) Edge R-tree

좌표에 기반한 에지 검색을 빠르게 하기 위해, 에지 MBR에 따른 R-tree를 구성한다.

3.2. POI 데이터

관광지 정보나 주유소, 식당과 같이 사용자의 관심 대상이 되는 장소를 POI (point of interest) 라고 한다. 이것은 자동차나 사

람과 같은 이동객체와 그 특성이 매우 다르기 때문에, POI 데이터는 하나의 독립적인 파일을 구성하여 서비스를 제공하는 것이 필요하다. 즉, 이는 이동 객체와는 달리, 거의 정적인 데이터의 특성을 지닌다.

(1) POI Information file

POI 데이터는 하나의 에지 e(ni, nj)에 놓여있는(중속된) POI들의 리스트 Lij로 나타낸다. 여기서 E는 공간 네트워크 내의 전체 에지 집합을 나타낸다.

$$Lij = \{ \langle Pid, Pname, Ptype, rlen, coord, PGij \rangle \mid e(ni, nj) \in E, i \neq j \}$$

- Pid : POI를 위한 ID, Pname : POI의 이름
- Ptype : POI 분류기호,
- rlen : POI를 포함하고 있는 에지에서, ni로 부터의 길이
- coord : 공간 좌표값,
- PGij : POI의 추가 정보가 저장된 페이지

(2) POI B+-tree

하나의 POI 정보에 대한 빠른 접근을 위하여 POI ID나 이름에 대하여 B+-tree를 구성한다.

3.3. 이동 객체 데이터

이동객체는 계속적으로 그 위치가 변하기 때문에, 일반적으로 데이터의 양이 매우 크다. 아울러 공간 네트워크 상에서는 미래 위치가 현재의 네트워크 세그먼트(에지)와 전체 네트워크에 매우 의존적이므로 상대적으로 미래 위치를 예측하기가 용이하며, 적어도 가까운 미래에 대해서는 전적으로 현재의 네트워크 에지에 의존하므로 상대적으로 정확한 미래 예측이 가능하다.

(1) Object Trajectory

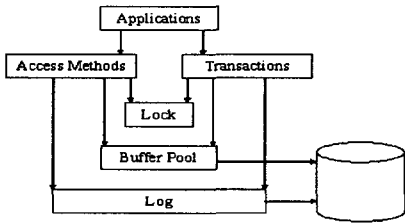
예측된 경로는 과거의 궤적과 마찬가지로 부분적으로는 하나의 에지에 해당되므로, 미래 궤적에 대한 처리도 과거 궤적과 동일하게 처리할 수 있다. 단 시간을 나타내는 애트리뷰트 값을, 예를 들면 마이너스 값으로 표현하여 미래 예측 궤적을 나타낸다.

(2) Object Trajectory signature

객체 궤적에 대한 빠른 검색을 위해서 접근 기법으로 signature 파일 기법을 사용한다. 그 이유는 궤적정보의 전체를 처음부터 끝까지 접근하는 응용보다는, 주어진 에지나 부분 궤적에 해당하는 전체궤적을 검색하는 것이 필요하다. 이러한 요구사항은 TB-tree [7]와 같은 트리 형태의 접근 기법을 사용하여 효과적으로 처리할 수 없다. 따라서 이러한 불연속적인 부분궤적 조건을 포함하는 전체궤적 검색에 대한 효율적인 처리를 위해, signature 파일을 사용하는 것이 효과적이다.

4. 공간 저장/색인 구조의 구현

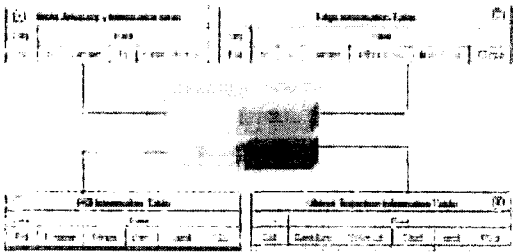
설계된 공간 색인/저장 구조의 구현을 용이하게 하기 위하여, 운영체제 상에서 처음 상태부터 구현하기 보다는, 기존 Berkeley 하부저장시스템 상에서, 이들이 제공하는 파일구조(file organization), 동시성제어(concurrency control) 및 파손 회복(crash recovery) 기능 등을 바탕으로 구현한다. Berkeley 하부저장 시스템의 구조는 (그림 2)과 같다.



(그림 2) Berkeley 하부저장 시스템의 구조

(1) 저장구조의 구현 모듈

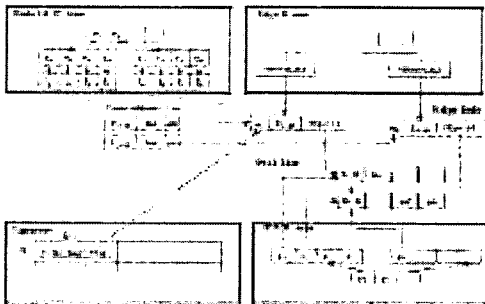
저장구조의 구현모듈의 전체구조는 (그림 3)과 같다. 저장구조는 노드인접 리스트, 에지 리스트, POI 리스트, 이동객체 궤적 리스트의 네 가지의 정보를 저장하는 구조를 가진다. 그리고 Berkely 하부저장시스템의 구조는 Key/Pair(Data)의 형태로 저장을 하기 때문에, 저장 테이블의 구조도 이와 동일한 형식을 따른다.



(그림 3) 저장구조

(2) 색인구조의 구현 모듈

색인구조 모듈의 전체구조는 (그림 4)와 같다. 색인구조는 각각 네 개의 컴포넌트로 구성된다. 먼저 노드를 색인하는 B⁺-tree가 있다. 그리고 에지의 정보를 가지고 MBR을 구성하여 공간질의에 효율적일 수 있도록 한 R-tree가 있다.



(그림 4) 색인구조

세 번째는 POI 정보 중에서 POI의 ID나 타입 등을 가지고 색인되는 B⁺-tree가 있고, 마지막으로 도로 네트워크 상에서 움직이는 이동 객체의 궤적을 색인하는 signature 파일이 존재한다.

5. 결론 및 향후 과제

본 논문에서는 공간 네트워크를 고려한 효율적인 저장구조와 색인구조를 설계하였다. 공간 네트워크 자체의 저장구조에서는 노드와 에지간의 정보를 포함하며, B⁺-tree를 이용해 노드에 대한 빠른 접근을 가능하게 하였으며, R-tree를 통한 에지에 대한 효율적인 공간 질의를 가능하게 하였다. 아울러 네트워크 상에 존재하는 POI들의 색인 구조를 위해, POI list 정보를 지닌다. 마지막으로, 이동객체의 색인 구조를 위해, 객체 궤적 및 signature 파일 구조를 지닌다.

향후 과제로는 설계된 저장 및 색인 구조의 효율성을 보이기 위하여, 저장 및 색인 구조에 대한 성능 비교 및 분석을 수행하는 것이다.

참고문헌

- [1] L. Speicys, C.S. Jensen, and A. Kligys, "Computational Data Modeling for Network-Constrained Moving Objects," Proc. of ACM GIS, pp 118-125, 2003.
- [2] C. Shahabi, M.R. Kolahdouzan, M. Sharifzadeh, "A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases," GeoInformatica, Vol. 7, No. 3., pp 255-273, 2003.
- [3] D. Pfoser and C.S. Jensen, "Indexing of Network Constrained Moving Objects," Proc. of ACM GIS, pp 25-32, 2003.
- [4] T. Brinkhoff, "A Framework for Generating Network-Based Moving Objects," GeoInformatica, Vol. 6, No. 2, pp 153-180, 2002.
- [5] D. Papadias, J. Zhang, N. Mamoulis, and Y. Tao, "Query Processing in Spatial Network Databases" Proc. of VLDB, pp, 802-813, 2003.
- [6] S. Shekhar and D.-R. Liu, "CCAM: A Connectivity-Clustered Access Method for Networks and Network Computations," IEEE Tran. on Knowledge and Data Engineering, Vol. 9, No. 1, pp 102-119, 1997.
- [7] D. Pfoser, C.S. Jensen, and Y. Theodoridis, "Novel Approach to the Indexing of Moving Object Trajectories," Proc. of VLDB, pp 395-406, 2000.