

계층적 행정구역 기반의 효율적인 위치정보 표현 방식

김우철, 이상윤, 박상현, 원정임
연세대학교 컴퓨터과학과

e-mail : {twelvepp, sylee, sanghyun, jiwon}@cs.yonsei.ac.kr

An Efficient Location Encoding Method Based on Hierarchical Administrative District

Woo-Cheol Kim, SangYoon Lee, Sanghyun Park, JungIm Won
Dept. of Computer Science, Yonsei University

요 약

최근 이동 통신 기술의 급속한 발달로 인해 휴대폰, PDA 등과 같은 휴대용 단말기의 사용이 보편화 되고 있다. 따라서 이동 객체의 시간에 따른 공간적인 위치정보를 활용하여 다양한 서비스를 제공하는 위치 기반 서비스(Location-Based Service)에 관한 많은 연구가 진행되고 있다. 위치 기반 서비스의 효율적 제공을 위하여 시간에 따라 지속적으로 변하는 이동 객체의 대용량 시공간 정보를 신속하게 저장, 관리, 검색할 수 있는 인덱싱 및 질의 처리 기술이 수반되어야 한다. 본 논문에서는 대용량 이동 객체 데이터베이스를 대상으로 한 효율적인 인덱스 구축을 위한 데이터 압축 표현 방식에 대하여 논한다. 이를 위해 본 논문에서는 기존의 주요 연구에서 2 차원 공간 좌표 (X, Y)로 표현되던 이동 객체의 공간 정보를 계층적 구조를 갖는 행정구역에 기반하여 1 차원의 공간 정보로 압축 표현하는 방식을 제안한다. 이를 이용하여 대용량의 공간 정보를 저장하고 있는 이동 객체 데이터베이스의 인덱스 크기의 감소 및 질의 처리 시간의 향상 효과를 얻을 수 있다. 또한, 제안된 방식은 2 차원 공간 좌표를 1 차원의 행정구역 기반의 위치정보로 표현하기 때문에 이로 인한 데이터 손실이 발생할 수 있으나, 일정 공간 내의 객체분포를 필요로 하는 교통 상황 파악, 근사적(approximate) 공간 정보를 필요로 하는 사람·차량 위치 추적 서비스와 같은 위치 기반 서비스를 신속하게 처리하는데 유용하게 사용될 수 있다.

1. 서론

이동 통신 기술의 발달로 인해 무선 이동 기기의 사용이 보편화 되면서 위치 기반 서비스(LBS: Location-Based Service)의 필요성이 증가하고 있다. 위치 기반 서비스는 이동 객체에 의해서 일정 시간마다 생성되는 공간적인 위치정보를 활용하여 다양한 응용 서비스를 제공하는 것을 말하며, 예를 들어 교통 정보 서비스, 물류 운송 정보 서비스, 버스·지하철 노선 안내 서비스, 사람·차량 위치 추적 서비스 등을 들 수 있다.

이러한 위치 기반 서비스에서의 이동 객체는 다음과 같은 특징을 갖는다. 첫째, 시간에 따라 지속적으로 공간적인 위치정보를 갱신하기 때문에 높은 갱신 비용이 든다. 둘째, 이동 객체는 공간 정보, 시간 정보, 객체 정보 등으로 이루어진 다차원의 데이터로 표현되므로 이동 객체의 저장 시 공간적인 오버헤드가 발생한다. 셋째, 이동 객체의 궤적 검색을 위하여

다수의 이동 객체로부터 획득된 과거 시점의 시공간 정보를 그대로 보존하기 때문에 시간이 증가함에 따라 정보의 양이 급속히 증가하여 대용량의 정보가 된다. 넷째, 검색 시 대용량 정보를 대상으로 하기 때문에 높은 검색 비용을 필요로 한다. 따라서, 이러한 이동 객체의 대용량 시공간 정보를 신속하게 저장, 관리, 검색할 수 있는 인덱싱 및 질의 처리 기술이 수반되어야 한다.

본 논문에서는 대용량 이동 객체 데이터베이스를 대상으로 한 효율적인 인덱스 구축을 위한 데이터 압축 표현 방식에 대하여 논한다. 이를 위해, 본 연구에서는 기존의 연구에서 2 차원의 공간 좌표 (X, Y)로 표현되던 이동 객체의 위치정보를 실세계 모델에 부합하는 계층적 행정구역에 기반하여 1 차원의 공간 정보로 표현하는 방식을 제안한다. 제안된 방식은 2 차원 공간 좌표를 1 차원의 행정구역 기반의 공간 정보로 표현하기 때문에 이동 객체의 정확한 공간적인 위

치정보에 대한 손실이 발생할 수 있다. 그러나, 위치 기반 서비스를 이용하는 사용자 측면에서 살펴보면 이동 객체의 정확한 위치를 나타내는 (X, Y)의 공간 좌표 보다는 행정구역 단위의 근사적인 공간 정보가 더욱 직관적일 수 있다. 예를 들어 차량 위치 추적 서비스에서 검색 대상의 자동차가 울림픽 대로를 따라 이동하고 있다면 서비스를 이용하고 있는 일반 사용자에게 자동차의 정확한 공간 위치인 (X, Y) 보다는 '반포 부근의 울림픽 대로' 라는 위치정보가 보다 직관적일 수 있다.

본 연구에서 제안된 방식을 이용하여 대용량의 이동 객체 데이터베이스를 대상으로 구축된 인덱스의 크기를 감소시킬 수 있으며, 그에 따른 질의 처리 시간의 향상 효과를 기대할 수 있다.

2. 관련연구

R-tree [1], 3DR-tree[2], HR-tree [3], TIR-tree [4], TB-tree [5] 같은 기존의 인덱스들은 이동 객체의 위치 정보를 (X, Y)의 2 차원 공간 좌표로 표현하고, 이를 3 차원 이상의 R-Tree 구조를 사용하여 인덱싱한다. 또한, 참고 문헌 [1], [2], [3], [4]는 영역 질의에 효과적이며, 참고 문헌 [5]는 계적 질의에서 높은 성능을 보였다. 그러나, 이들 인덱스들은 시간이 증가함에 따라 인덱스의 크기가 급속히 커지기 때문에 높은 검색 비용을 필요로 한다. 따라서 최근 이러한 문제점을 개선하기 위한 방안으로 이동 객체를 표현하는 다차원 데이터를 저차원으로 변환하여 저장, 관리하는 연구들이 진행되고 있다.

그 대표적인 예로 이동 객체의 위치정보를 2 차원의 공간 좌표가 아닌 실세계의 도로 네트워크(Road Network) 정보를 이용하여 표현하는 연구들이 진행되고 있다. 도로 네트워크는 이동 객체의 이동 가능한 범위를 제한하고; 이동 객체의 위치정보를 1 차원의 도로 정보에 기반하여 표현한다. 참고 문헌 [6]에서는 도로 네트워크를 기반으로 하는 위치 정보 표현의 유용성을 보인다. 예를 들어, "현재 위치에서 가장 가까운 호텔을 찾아라" 라는 질의에 대해 공간 좌표 상에서 가장 가까운 위치에 있는 호텔을 검색 결과로 제시하는 대신에 도로망을 따라 이동 거리가 가장 짧은 호텔을 검색 결과로 제시한다. 그러나, 참고 문헌 [6]은 공간 질의만 주요 대상으로 하기 때문에 시간 개념이 포함된 질의는 처리하지 못한다는 단점이 있다.

이와 달리, 이동 객체를 표현하는 데이터의 차원이 아닌 인덱스의 차원을 줄이는 방법을 제시한 참고 문헌 [7]은 3 차원의 인덱스를 영역 정보를 표현하는 2 차원 인덱스와 계적 정보를 표현하는 2 차원 인덱스로 각각 분리하여 저장하는 방식을 사용한다. 주어진 질의에 따라 적합한 2 차원 인덱스를 선택하여 처리하기 때문에 기존의 3 차원의 인덱스 기법보다 검색 성능이 우수하다. 그러나 이동 객체의 위치정보가 여전히 (X, Y)의 2 차원 공간 좌표로 표현되기 때문에 도로망 혹은 행정 구역 단위의 위치 체계를 갖는 실세계 모델

에 적합하다고 할 수 없다.

3. 이동 객체 데이터의 특징

대부분의 위치 기반 서비스에서의 이동 객체는 위치를 나타내는 (X, Y)의 2 차원 공간 좌표, 시간을 나타내는 T, 객체를 표현하는 Obj 로 이루어진 (X, Y, T, Obj)의 4 차원 데이터로 표현된다. 또한, 이와 같은 4 차원 데이터로 표현되는 이동 객체를 위한 인덱스 구조로 R-Tree 를 기본 구조로 하는 3 차원 이상의 인덱스 구조를 사용하는 것이 일반적이다. 그러나, 이들 인덱스들의 대부분은 실세계에서의 대부분의 이동 객체가 일정 도로를 따라 이동한다는 특성을 고려하지 않고 이동 객체의 위치를 (X, Y)의 2 차원 공간 좌표로 표현하기 때문에 실세계에 객체가 존재할 수 없는 영역까지도 인덱스 공간으로 표현하게 된다. 따라서 다수의 이동 객체를 인덱싱하는 경우 인덱스 크기가 크고, 중첩 영역의 증가로 인한 성능 저하를 발생시킬 수 있으므로 효율적이지 못하다.

본 논문에서는 이러한 문제점을 해결하기 위한 방안으로 실세계의 위치정보 표현 모델을 반영하여 이동 객체의 위치정보를 1 차원으로 압축 표현하는 방식을 제안한다. 실세계에서 객체의 위치정보는 행정구역 단위로 표현된다. 예를 들어 "서대문구청"의 위치는 "서울특별시 서대문구 연희로 165"와 같이 계층 구조를 갖는 행정구역 단위로 표현된다. 즉, 실세계에서 위치정보는 (동경 120 도 55 분, 북위 30 도 18 분)과 같이 2 차원의 공간 좌표 형태로 표현되는 대신에 영역을 세분화시켜 행정구역 기반의 계층적 구조 형태로 표현된다.

4. 새로운 위치정보의 표현 방법

본 논문에서는 3 장에서 언급한 실세계의 위치정보 표현 방식의 특징을 이용한 계층적 행정구역 기반의 위치 표현 방법을 제안한다. 즉, 기존의 좌표 형태의 2 차원 데이터를 계층적 행정 구역 기반의 위치정보를 표현하는 이진 스트링으로 변환한 후, 이를 다시 하나의 숫자로 변환하여 1 차원 데이터로 압축 표현하는 방법을 사용한다.

4.1. 위치정보를 위한 영역 분할 방법

계층 구조를 갖는 행정구역 정보를 이용하여 위치 정보를 표현할 수 있다. 그러나 "서울시 서대문구 연희로 165"의 정보를 단순히 "도/시" 혹은 "시/군/구"과 같이 각각의 행정구역을 각각의 차원으로 표시하게 되면 (서울시, 서대문구, 연희로, 165)와 같이 4 차원으로 표현되어 기존의 2 차원 공간 좌표보다 오히려 차원이 커지게 된다. 따라서 본 논문에서는 위치정보를 각각의 행정구역을 해당 이진 스트링으로 변환한 후, 변환된 각각의 이진 스트링을 하나로 묶어서 그림 1 과 같이 하나의 숫자로 표현하는 방식을 사용한다.

이때, 각각의 행정구역에 대한 이진 스트링은 단순히 사전식 배열 방식에 의하여 생성하는 것이 아니라 각각의 행정구역간의 위치적 상관관계를 파악할 수

있도록 방향정보를 사용하여 생성한다. 예를 들어 “서울특별시”라는 행정구역 안에서 “구” 단위의 행정구역을 표현하는 이진 스트링을 “강남구(00000)”, “강동구(00001)”, “강북구(00010)”과 같이 임의의 일련 번호를 사용하여 생성하게 되면 강남구와 강동구사이의 위치적 상관관계에 대한 정보를 알 수 없게 된다.

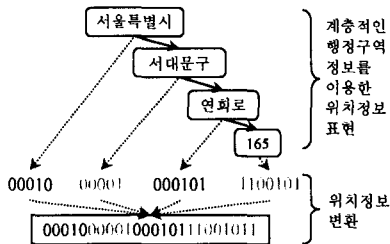


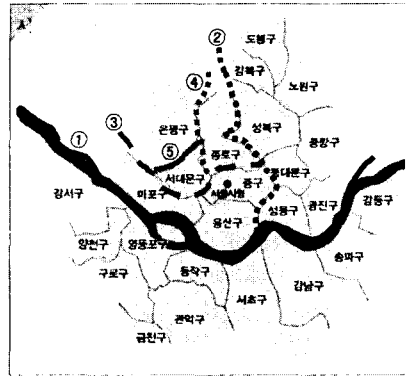
그림 1. 계층구조를 갖는 행정구역의 위치 정보 표현 예

따라서 본 논문에서는 전체 영역을 방향을 고려하여 행정구역 단위로 분할한 후, 이를 이진 스트링으로 표현하는 방식을 제안한다. 이진 스트링 생성 방법은 다음과 같다.

- (1) 행정구역들을 남/북으로 분할한 후, 행정구역이 남쪽에 있으면 “1”을 북쪽에 있으면 “0”을 이진 스트링에 추가한다.
- (2) (1)에서 분할된 행정구역들을 다시 동/서로 분할한 후, 행정구역이 동쪽에 있으면 “1”을 서쪽에 있으면 “0”을 이진 스트링에 추가한다.
- (3) (1)과(2)를 모든 행정구역에 대하여 반복하여 이진 스트링을 생성한다.
- (4) (3)에 의해 생성된 각각의 행정구역에 대하여 생성된 이진 스트링을 하나로 묶어서 일정 길이의 이진 스트링을 생성 한 후, 이를 하나의 숫자로 표현한다.

단, 행정구역 분할 과정에서 다음의 사항을 고려한다. 첫째, 분할 후의 행정구역의 개수가 가능한 같게 분할 한다. 둘째, 최근에 주소 체계 개편을 통해서 위치 표현 방식이 “동” 단위의 행정구역에서 “로” 단위의 길 중심으로 변경이 되었지만, 아직까지도 “동” 단위의 위치 체계 역시 많이 사용되고 있다. 따라서 최하위 행정구역인 “로”는 일반적으로 그 상위 행정구역인 하나의 “동”에 포함 되지 않고 여러 “동”에 걸쳐서 나타난다. 본 논문에서는 “동”의 경계선에 따라서 “로”를 분할하고 각각 다른 “로”로 표시함으로써 “로” 단위도 위와 동일한 방법으로 분할하여 표현한다.

그림 2에 “서울특별시”의 행정구역 중에 하나인 “서대문구”의 위치를 표현하기 위한 이진 스트링의 생성 예를 보인다. 동일한 방법으로 “서대문구”의 하위 행정구역인 “연희동”과 “연희로”도 표현할 수 있다.



“서울특별시” 안에 있는 “서대문구”를 표현하는 과정	이진스트링
① 전체 행정구역을 대상으로 남/북으로 분할 -> “서대문구”가 북쪽에 있기 때문에 0을 추가	0__
② “서대문구”가 속해 있는 영역을 동/서로 분할 -> “서대문구”가 서쪽에 있기 때문에 0을 추가	00__
③ “서대문구”가 속해 있는 영역을 남/북으로 분할 -> “서대문구”가 북쪽에 있기 때문에 0을 추가	000__
④ “서대문구”가 속해 있는 영역을 동/서로 분할 -> “서대문구”가 서쪽에 있기 때문에 0을 추가	0000__
⑤ “서대문구”가 속해있는 영역을 남/북으로 분할 -> “서대문구”가 남쪽에 있기 때문에 1을 추가	00001
⑥ 더 이상 분할 가능한 행정구역이 없기 때문에 종료	00001

그림 2. “서대문구”의 이진 스트링 생성 예

4.2. 행정구역에 기반한 위치정보의 특징

4.1 절처럼 변환된 위치정보 이진 스트링은 다음과 같은 4 가지의 특징을 가지고 있다. (1) 먼저 각 이진 스트링은 동일한 길이를 갖는다. (2) 두 개의 이진 스트링의 일정 길이의 접두어가 같다면 접두어가 표현하는 의미만큼 같은 행정구역 안에 속해 있다. 예를 들어 “서울시 서대문구 신촌동 백양로”와 “서울시 서대문구 연희동 연희로”는 “서울시 서대문구”를 표시하는 이진 스트링의 길이만큼 같은 접두어를 가지고 있고 그 접두어가 표현하는 “서울시 서대문구”에 같이 속해있는 행정구역이다. (3) “서울시 서대문구”에 포함된 모든 행정구역은 “서울시 서대문구”를 표현하는 이진 스트링 뒤에 “0”을 일정 길이만큼 붙여 생성한 최소값과 “1”을 일정 길이만큼 붙여 생성한 최대값 사이의 범위에 포함된다. (4) 마지막으로 각 위치정보간에는 포함 관계만이 생기며, 중첩 관계는 생기지 않는다.

4.3. 이동 객체의 위치 및 궤적 정보의 표현 방법

4.1 절에서 각 행정구역 단위 별로 위치정보를 하나의 이진 스트링으로 표현하는 방법을 살펴 보았다. 하지만 이동 객체를 “서울특별시 서대문구 연희동 연희로”와 같이 영역 정보만 이용하면 많은 정보 손실이 발생한다. 따라서 이동 객체의 위치정보를 표현 할 때는 기존의 영역 정보뿐만 아니라 그 영역의 어디쯤에 있는지를 알 수 있는 상대 위치정보를 포함한다. 이때 도로 안에서 위치를 (X, Y)가 아니라 도로의 OO%처럼 상대적인 위치로 표현해 기존의 영역 정보에 상대

적인 위치정보까지 연결해서 사용한다.

이와 같이 이동 객체를 이진 스트링을 변환한 하나의 숫자 형태로 표현할 경우, 이동 객체의 궤적은 연속적인 위치정보를 표현하므로 이동 객체의 궤적 정보는 숫자 배열 형태로 표현이 가능하다.

5. 질의 처리의 예

위치 기반 서비스에서 주로 사용되는 질의를 처리하는 예를 살펴 보면 다음과 같다.

예 1) 월수가 8시부터 8시 10분까지 움직인 궤적은?

위 질의를 처리하기 위해서는 일정 시간 간격마다 생성된 각 이동 객체의 위치정보를 필요로 한다. 예를 들어 객체의 정보를 1분마다 저장하고, 월수의 위치정보를 7시부터 저장했다고 가정한다면 61번부터 71번째 이진 스트링들을 검색하여 이를 행정구역 단위의 정보로 변환해서 제시하면 된다.

예 2) 월수가 서울시 서대문구 신촌동에 있었던 시간은?

위 질의를 처리하기 위해서는 위치정보를 키로 사용하는 인덱스 구조가 필요하다. 이 때 그림 3 과 같이 간단히 B+트리를 이용하는 경우를 생각해 볼 수 있다. "서울시 서대문구 신촌동"을 나타내는 숫자의 범위를 접두어로 하는 이진 스트링의 최소값과 최대값으로 영역을 표시할 수 있다. 이 정보를 이용해서 B+트리에 영역질의를 하면 원하는 결과를 쉽게 검색할 수 있다.

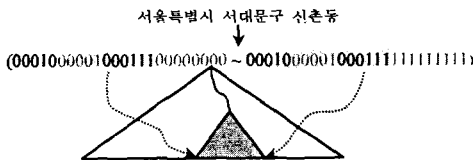


그림 3. B+트리를 이용한 질의 처리 예

예 3) 7시부터 7시 20분까지 서울시 서대문구 신촌동에 머물렀던 사람은?

위 질의를 처리하기 위해서는 일정 시간 간격 동안 움직인 위치정보를 하나의 영역으로 변환해서 그 정보를 키로 사용하는 B+트리를 이용할 수 있다. 예를 들어 10분 단위로 B+트리를 생성한다면 7시~7시 10분, 7시 10분~7시 20분의 B+트리를 모두 검색하여 서울시 서대문구 신촌동에 있던 사람들을 검색한 후에 얻어진 중간 결과들을 정렬병합(sort-merge)하여 원하는 최종 결과를 얻을 수 있다.

6. 기대효과

본 논문에서 제시된 방법으로 위치정보 표현 방식을 대체하면 다음과 같은 5 가지 효과를 기대할 수 있다. (1) 이동 객체를 표현하기 위해 사용되던 기존의 4 차원의 데이터를 3 차원으로 줄임으로써 데이터의

차원 감소로 인한 알고리즘의 복잡도를 감소시킬 수 있다. (2) 데이터를 구성하는 차원의 감소로 인하여 작은 물리적 저장공간을 사용할 수 있다. (3) 궤적 검색 시 좌표 데이터를 다룰 때 쓰이는 기하학 기반의 알고리즘 대신에 숫자 배열을 이용할 수 있는 문자열 관련 알고리즘을 이용할 수 있다. (4) 궤적을 쉽게 보존할 수 있다. (5) 숫자 배열로 표현된 궤적 정보에는 연속적으로 같은 숫자가 출현하기 때문에 과거 데이터를 압축할 때 높은 압축 효율을 얻을 수 있다.

7. 결론 및 향후 연구과제

본 논문에서는 위치 기반 서비스에서 주로 사용되는 위치정보를 (X, Y)의 2 차원 공간 데이터에서 계층적인 행정구역 정보를 이용한 이진 스트링 형태의 1 차원의 데이터로 변환하여 표현하는 방식을 제안하였다. 제안된 방식에서는 데이터의 손실이 발생할 수 있으나, 행정구역에 기반한 근사적인 위치정보만을 필요로 하는 위치 기반 서비스에서는 효율적인 표현 방식이 될 수 있다.

본 논문에서 제안한 위치정보 표현 방식을 통해서 이동 객체를 표현하기 위한 기존의 4 차원 데이터를 3 차원의 데이터로 차원을 감소하여 표현할 수 있으므로 기존의 위치 기반 서비스를 위한 인덱스 방식들에 비해서 알고리즘을 단순화시킬 수 있다. 또한 궤적 정보를 표현하는 방식이 기존의 기하학적인 데이터에서 연속된 숫자의 배열 형식으로 바뀐에 따라 궤적 보존 및 검색 시 보다 효율적이다.

향후 연구로는 제안된 새로운 위치정보 표현 방식에 기반한 인덱싱 방법 및 질의 처리 방법에 대한 연구를 수행하고자 한다.

참고문헌

[1] A. Guttman, "R-trees: a dynamic index structure for spatial searching," In Proc. ACM SIGMOD, pp. 47-57, 1984.
 [2] Y. Theodoridis, M. Vazirgiannis and T. Sellis, "Spatio-Temporal Indexing for Large Multimedia Applications," In Proc. ICMCS, pp. 441-448, 1996.
 [3] M.A. Nascimento, J.R.O. Silva, "Towards historical R-trees," In Proc. ACM SAC, pp. 235-240, 1998.
 [4] 조형주, 정진완, "시공간 질의를 위한 인덱싱 기법," 한국데이터베이스 학술대회 논문집, Vol.18, No. 2, pp. 93-100, 2002.
 [5] D. Pfoser, C.S. Jensen, Y. Theodoridis, "Novel Approaches in Query Processing for Moving Objects," In Proc. VLDB, pp. 395-406, 2000.
 [6] D. Papadias, J. Zhang, N. Mamoulis, Y. Tao, "Query Processing in Spatial Network Databases," In Proc. VLDB, pp. 802-813, 2003.
 [7] D. Pfoser, C.S. Jensen, "Indexing of Network Constrained Moving Objects," In Proc. ACM GIS, pp. 25-32, 2003.