

(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。 Int. Cl. G06F 17/40 (2006.01)	(45) 공고일자 (11) 등록번호 (24) 등록일자	2006년09월06일 10-0620125 2006년08월28일
---	-------------------------------------	--

(21) 출원번호 (22) 출원일자	10-2005-0064746 2005년07월18일	(65) 공개번호 (43) 공개일자
------------------------	--------------------------------	------------------------

(73) 특허권자	인하대학교 산학협력단 인천 남구 용현동 253 인하대학교
(72) 발명자	배해영 인천 연수구 청학동 531-14 현암연구소 이재동 서울 용산구 한남동 단국대학교 정보컴퓨터학부 오영환 충남 천안시 쌍용동 456번지 나사렛대학교 정보과학부 김경배 충북 청주시 흥덕구 모충동 231 서원대학교 컴퓨터교육학과 김명근 인천 남구 주안1동 395-15 성두빌라 박순영 인천광역시 연수구 동춘2동 풍림2차아파트 110동 405호 김호석 경기도 부천시 오정구 원종1동 299-2 8/1 장용일 경기도 부천시 원미구 심곡1동 377-35번지 유병섭 인천 연수구 청학동 531-14 현암연구소 어상훈 인천 연수구 청학동 531-14 현암연구소 정영철 인천 연수구 청학동 531-14 현암연구소
(74) 대리인	이원희

(56) 선행기술조사문헌
JP2004503838 A

KR1020020038438 A

KR1020030095704 A

KR1020040066942 A

* 심사관에 의하여 인용된 문헌

심사관 : 구영희

(54) 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인재구성 시스템 및 방법

요약

본 발명은 색인의 구조에 맞게 클러스터들을 부분 색인으로 구성하여 전송하고, 전송된 부분 색인이 기존 색인에 직접 삽입될 수 있도록 한 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템 및 방법을 제공하는 것으로, 본 발명의 색인 재구성 방법은, 상기 구축기에서 상기 소스 데이터베이스로부터 추출된 데이터를 색인의 구조에 맞게 클러스터링하는 제1 단계; 상기 클러스터링에 따라 생성된 클러스터로 부분 색인을 생성하여 전송하는 제2 단계; 상기 공간 데이터 웨어하우스 서버의 미 할당된 물리적으로 연속된 공간에 상기 전송된 부분 색인을 기록하는 제3 단계; 및 상기 기록된 부분 색인을 기존 색인에 삽입하는 제4 단계;를 수행하여 달성될 수 있으며, 이러한 본 발명은 소스 데이터베이스로부터 추출된 데이터를 색인의 구조에 맞게 클러스터링하고, 생성된 각 클러스터로 부분 색인을 전송하여 기존 색인에 직접 삽입함으로써 색인 재구성을 위한 검색, 분할, 재조정 비용을 최소화함과 더불어 공간 데이터 웨어하우스 시스템이 지속적으로 안정된 처리를 가능케하여 시스템의 성능을 향상시킬 수 있게 된다.

대표도

도 1

색인어

공간 데이터 웨어하우스, 구축기, 서버, 부분 색인, 재구성, 클러스터링

명세서

도면의 간단한 설명

도 1은 본 발명을 구현하기 위한 시스템 구성도.

도 2는 본 발명의 전체적인 동작 흐름도.

도 3은 도 1의 공간 데이터 웨어하우스 서버의 자료 저장소의 구조도.

도 4는 도 1의 구축기에서의 부분 색인 생성 과정을 설명하기 위한 도.

도 5는 도 1의 구축기에서 공간 데이터 웨어하우스 서버로 부분 색인을 전송하는 과정을 설명하기 위한 도.

도 6은 도 1의 공간 데이터 웨어하우스 서버로 전송되어 기록된 부분 색인의 구성도.

도 7은 도 1의 공간 데이터 웨어하우스 서버에서 부분 색인을 삽입하는 과정을 설명하기 위한 도.

<도면의 주요 부분에 대한 부호의 설명>

100 : 소스 데이터베이스 200 : 구축기

210 : 임시 자료 저장소 220 : ETL 처리기

230 : 부분 색인 전송 관리기 240 : 적재 처리기
 300 : 공간 데이터 웨어하우스 서버 310 : 적재 중계기
 320 : 색인 처리기 330 : 자료 저장소
 340 : 질의 처리기

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 공간 데이터 웨어하우스에서의 색인 재구성 시스템 및 방법에 관한 것으로, 특히 색인의 구조에 맞게 클러스터들을 부분 색인으로 구성하여 전송함으로써 부분 색인이 기존 색인에 직접 삽입될 수 있도록 하여 색인 재구성을 위한 검색 및 분할과 재조정 비용을 최소화하여 공간 데이터 웨어하우스 시스템이 지속적으로 안정된 처리를 행할 수 있도록 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템 및 방법에 관한 것이다.

데이터 웨어하우스는 사용자의 의사 결정에 도움을 주기 위해 다양한 운영 시스템에서 추출, 변환, 통합되고 요약된 데이터 베이스로, 데이터 웨어하우스는 1980년 중반 IBM이 자사 하드웨어를 판매하기 위해 처음으로 도입했던 개념으로, IBM은 인포메이션 웨어하우스(Information Warehouse)라는 용어를 사용하였다.

데이터 웨어하우스는 원시 데이터 계층, 데이터 웨어하우스 계층, 클라이언트 계층으로 구성되며, 데이터의 추출, 저장, 조회 등의 활동을 한다. 데이터 웨어하우스의 특징을 보면, 운영 시스템은 조직 운영에 필요한 재고관리와 회계정보, 영업 시스템과 같은 특화된 기능을 지원하며, 데이터 웨어하우스는 고객과 제품, 회계와 같은 주제를 중심으로 데이터를 구축한다.

공간 데이터에 대한 공간 데이터 웨어하우스는 공간 데이터 웨어하우스 서버와 구축기로 구성되며, 구축기는 소스 데이터의 변경 사항을 구축기 내부의 저장소에 적재하며, 적재된 데이터를 일정 주기마다 일괄 처리의 형태로 공간 데이터 웨어하우스 서버에 적재한다.

또한, 데이터 웨어하우스 서버는 색인 구축을 위해 사용자 서비스를 중단하고, 구축기로부터 전송되어지는 데이터를 적재하며 사용자의 질의에 빠른 응답을 하기 위해 적재된 데이터를 이용하여 색인을 구축한다.

색인을 구성하는 종래 기법으로는 벌크 삽입 기법, Seeded 클러스터링 기법, 그리고 색인 전송 기법이 있다.

다차원 색인의 벌크 삽입 기법은 삽입 데이터를 공간상에서 근접한 데이터끼리 클러스터링하여 여러 개의 클러스터를 생성한다. 생성된 각 클러스터는 Small-Tree로 구성되며, 각 Small-Tree는 기존 R-Tree에 삽입된다.

또한, 어떤 클러스터에도 포함되지 않는 데이터들은 열외자(Outlier)로 분류하여 일반적인 삽입 알고리즘에 의해 하나씩 삽입한다. 이 기법은 데이터를 클러스터링하는 비용이 많이 들며, 기존 R-Tree 노드들과 새로 삽입되는 Small-Tree 노드들 간의 겹침이 크다. 따라서 벌크 삽입 기법은 삽입 성능 및 검색 성능이 떨어지는 단점이 있다(L.Chen, R.Choubey, and E.A.Rundensteiner, "Bulk-insertions into R-tree using the small-tree-large-tree approach", ACM GIS, 1998).

Seeded 클러스터링 기법은 기존의 공간 근접도에 따른 클러스터링 방법이 아닌 이미 구축되어 있는 색인의 구조에 맞추어 클러스터링하는 기법이다. 공간 근접도에 따른 클러스터링 기법은 기존의 색인의 구조를 고려하지 않아 색인의 영역이 확장될 가능성이 크며, 과정이 복잡하여 클러스터링 비용이 많이 든다,

그러나 Seeded 클러스터링 기법은 색인의 상위부분에 데이터를 삽입하여 클러스터링하기 때문에 기존의 클러스터링 기법보다 노드간의 겹침이 줄어들며 빠르게 클러스터링할 수 있으나, 기존의 색인의 구조를 고려하지 않아 색인의 영역이 확장될 가능성이 크며, 과정이 복잡하여 클러스터링 비용이 많이 든다(Taewon Lee, Bongki Moon, and Sukho Lee "Bulk Insertion for R-tree by Seeded Clustering", DEXA, 2003).

색인 전송을 통한 색인 구성 기법은 기 구축된 색인의 재활용을 위해 색인 구조를 직접 전송하여 색인을 구성한다. 색인 전송으로 인한 기 구축된 색인의 물리적인 사상문제는 연속적인 익스텐트(Extent)를 예약하여 해결하였다. 이 기법은 색인 전송을 통해 색인 구성을 위한 검색, 분할, MBR(Minimum Bounding Rectangles) 재구성 비용을 제거하여 색인 구성 비용을 줄이지만 주기적인 소스 데이터의 변경이 발생하는 공간 데이터 웨어하우스의 색인 재구성 기법으로는 적합하지 않다(박상근, 김호석, 이재동, 배해영, "분산 데이터베이스 시스템에서의 색인 구성비용 절감을 위한 효율적인 색인 전송 기법", 정보과학회, 2003.)

발명이 이루고자 하는 기술적 과제

본 발명은 이러한 문제점을 해결하기 위한 것으로, 본 발명의 목적은 추출된 데이터를 색인의 구조에 맞게 클러스터링하고, 생성된 각 클러스터로 부분 색인을 전송하여 기존 색인에 직접 삽입함으로써 색인 재구성을 위한 검색, 분할, 재조정 비용을 최소화함과 더불어 공간 데이터 웨어하우스 시스템이 지속적으로 안정된 처리를 가능케하여 시스템의 성능을 향상시킬 수 있도록 한 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템 및 방법을 제공함에 있다.

발명의 구성 및 작용

상기 본 발명의 목적을 달성하기 위한 본 발명에 따른 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템은, 공간 데이터 웨어하우스 시스템에 있어서, 소스 데이터베이스에서 추출되어 적재된 데이터에 대하여 색인의 구조에 맞게 클러스터링하며, 생성된 클러스터로부터 부분 색인을 생성하여 공간 데이터 웨어하우스 서버로 전송하는 구축기; 및 자신의 자료 저장소에 미 할당된 물리적으로 연속된 공간을 예약하여 예약된 공간에 상기 구축기로부터 전송된 부분 색인을 기록하며, 상기 기록된 부분 색인을 자신이 보유하고 있는 기존 색인에 삽입하는 공간 데이터 웨어하우스 서버;로 구성됨을 특징으로 한다.

상기 구축기는 페이지 단위로 부분 색인을 전송하며, 상기 부분 색인을 생성하기 위해 상기 공간 데이터 웨어하우스 서버로부터 기 구축된 색인의 상위 부분을 전송받아 전송받은 색인의 상위 부분을 통해 클러스터링하며, 사용자가 지정해 놓은 적재시점이 되면 부분 색인을 위한 공간 예약 메시지를 상기 공간 데이터 웨어하우스 서버로 전송하여 공간 데이터 웨어하우스 서버로부터 예약 완료 메시지를 전송받으면 상기 부분 색인을 전송하며, 최상위 노드부터 중간 노드, 단말 노드의 순으로 부분 색인을 전송하며, 상기 단말 노드에 대해서는 단말 노드가 가리키는 해당 레코드와 함께 전송함을 특징으로 한다.

상기 목적을 달성하기 위한 본 발명에 따른 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법, 소스 데이터베이스 및 구축기와 공간 데이터 웨어하우스 서버로 구성된 공간 데이터 웨어하우스 시스템에서의 색인 재구성 방법에 있어서, 상기 구축기에서 상기 소스 데이터베이스로부터 추출된 데이터를 색인의 구조에 맞게 클러스터링하는 제1 단계; 상기 클러스터링에 따라 생성된 클러스터로 부분 색인을 생성하여 전송하는 제2 단계; 상기 공간 데이터 웨어하우스 서버의 미 할당된 물리적으로 연속된 공간에 상기 전송된 부분 색인을 기록하는 제3 단계; 및 상기 기록된 부분 색인을 기존 색인에 삽입하는 제4 단계;를 포함하여 이루어짐을 특징으로 한다.

상기 제1 단계에서의 클러스터링은 상기 공간 데이터 웨어하우스 서버로부터 기 구축된 색인의 상위 부분을 전송받아 전송받은 색인의 상위 부분을 통해 행하며, 상기 제2 단계에서의 부분 색인의 전송은 사용자가 지정해 놓은 적재시점이 되면 부분 색인을 위한 공간 예약 메시지를 상기 공간 데이터 웨어하우스 서버로 전송하여 상기 공간 데이터 웨어하우스 서버로부터 예약 완료 메시지를 전송받으면 부분 색인을 전송하며, 상기 부분 색인의 전송은 페이지 단위로 이루어지고 또한 최상위 노드부터 중간 노드, 단말 노드의 순으로 행해지며, 부분 색인 전송시 상기 단말 노드에 대해서는 단말 노드가 가리키는 해당 레코드와 함께 전송함을 특징으로 한다.

이하, 본 발명의 바람직한 실시 예를 첨부된 도면을 참조하여 보다 상세하게 설명한다. 단, 하기 실시 예는 본 발명을 예시하는 것일 뿐 본 발명의 내용이 하기 실시 예에 한정되는 것은 아니다.

도 1은 본 발명을 구현하기 위한 공간 데이터 웨어하우스 시스템을 나타낸 것으로, 크게 소스 데이터베이스(100), 구축기(200) 및 공간 데이터 웨어하우스 서버(300)로 구성된다.

상기 구축기(200)는 소스 데이터베이스(100)에서 변경된 데이터를 추출하고 변형 및 정제하여 임시 자료 저장소(210)에 적재하는 ETL(Extraction, Transformation, Loading) 처리기(220), 상기 임시 자료 저장소(210)에 적재된 데이터에 대한 부분 색인을 생성하여 공간 데이터 웨어하우스 서버(300)로 전송하는 부분 색인 전송 관리기(230) 및 적재 처리기(240)로 구성된다.

상기 구축기(200)의 ETL 처리기(220)는 소스 데이터베이스(100)로부터 데이터를 추출하는 추출 관리기(221), 추출된 데이터를 변환 및 정제하는 변환 관리기(222) 및 적재 관리기(223)로 구성되며, 부분 색인 전송 관리기(230)는 상기 소스 데이터베이스(100)로부터 추출된 데이터에 대하여 색인의 구조에 맞게 클러스터링하는 클러스터 관리기(231), 생성된 클러스터로 부분 색인을 생성하는 부분 색인 생성기(232) 및 색인 전송 처리기(233)로 구성된다.

상기 공간 데이터 웨어하우스 서버(300)는 상기 구축기(200)의 적재 처리기(240)로부터 적재 중계기(310)를 거쳐 전송된 부분 색인을 기존 색인에 삽입하는 색인 처리기(320), 색인 및 레코드 저장소인 자료 저장소(330), 재구성된 색인을 이용하여 공간 OLAP(On-Line Analytical Processing) 연산을 지원하는 공간 OLAP 처리기(342)를 포함하는 질의 처리기(340) 및 실체화 뷰 관리기(350)로 구성된다.

공간 데이터 웨어하우스 서버(300)의 색인 처리기(320)는 구축기(200)로부터 전송된 부분 색인을 기록하는 부분 색인 기록기(321) 및 부분 색인을 공간 데이터 웨어하우스 서버(300)에 있던 기존 색인에 삽입하는 전체 색인 삽입기(322)로 구성되며, 상기 자료 저장소(330)는 색인이 저장되는 색인 저장소(331)와 레코드가 저장되는 레코드 저장소(332)로 구성된다. 또한, 질의 처리기(340)는 SQL(Structured Query Language) 처리기(341), 공간 OLAP 처리기(342) 및 질의 최적화기(343)로 구성된다.

이와 같이 구성된 본 발명을 도 2의 흐름도와 함께 설명한다.

먼저, 상기 구축기(200)의 추출 관리기(221)에서 소스 데이터베이스(100)로부터 변경된 데이터를 추출하며, 변환 관리기(222)에서 추출된 데이터를 변형 및 정제하면 적재 관리기(223)에서 이를 임시 자료 저장소(210)에 적재하게 된다.

부분 색인 전송 관리기(230)의 클러스터 관리기(231)는 상기 ETL 처리기(220)로부터 임시 자료 저장소(210)에 적재된 데이터를 공간의 근접도가 아닌 색인의 구조에 맞게 클러스터링하며(S100), 부분 색인 생성기(232)는 상기 클러스터 관리기(231)에서 생성된 각 클러스터로 부분 색인을 생성한다.

색인 전송 처리기(233)는 적재 처리기(240)를 통해 상기 부분 색인 생성기(232)에서 생성된 부분 색인을 페이지 단위로 공간 데이터 웨어하우스 서버(300)로 전송한다(S200).

상기 구축기(200)로부터 전송된 부분 색인은 적재 중계기(310)를 거쳐 색인 처리기(320)로 입력되어 색인 처리기(320)의 부분 색인 기록기(321)에서 자료 저장소(330)의 색인 저장소(331)에 기록된다.

이때, 공간 데이터 웨어하우스 서버(300)는 상기 구축기(200)에서 전송된 부분 색인의 물리적 사상 문제를 해결하기 위해 자료 저장소(330)의 색인 저장소(331)에 물리적으로 연속된 공간을 예약하여 예약된 공간에 부분 색인 기록기(321)를 통해 부분 색인을 기록하게 된다(S300).

또한, 기록된 부분 색인은 전체 색인 삽입기(322)에 의해 공간 데이터 웨어하우스 서버(300)에 있던 기존 색인에 일반적인 삽입 알고리즘을 통해 삽입된다(S400).

공간 데이터일 경우, R-tree의 삽입 알고리즘은 먼저 엔트리 E가 위치하게 될 단말노드 L을 선택한다. 이때 L은 기존 경계 사각형을 최소로 확장시키는 것을 선택하며 확장의 양이 동일한 경우는 면적이 작은 것을 선택한다. 선택한 단말노드 L에 공간이 있으면 엔트리 E를 위치시키고 그렇지 않으면 단말노드 L을 분할(Split)한다. 이때, 분할된 두 노드의 사각형 영역은 최소가 되도록 한다. 그리고, 분할이 일어나면 트리를 다시 재구성하게 된다. 이때, 부모 노드가 포함하는 사각형이 조정되어야 하며 분할의 영향은 루트방향으로 전달되게 된다.

도 3은 상기 공간 데이터 웨어하우스 서버(300)의 자료 저장소(330)의 구조를 도시한 것으로, 구축기(200)의 임시 자료 저장소(210)도 이와 동일한 구조를 가진다.

본 발명의 환경에서 색인의 노드는 파일 I/O의 기본 단위인 페이지를 사용하며, 4개의 페이지는 데이터베이스 공간 할당의 기본 단위인 익스텐트(Extent)를 구성한다.

도 3에서 프리 익스텐트 리스트(Free Extent List)는 미 할당된 익스텐트의 리스트와 데이터베이스(도 1의 자료 저장소(330))의 테이블이나, 색인구조 등에 사용되었던 익스텐트 중 반환된 익스텐트들의 리스트를 나타낸다.

도 4는 상기 구축기(200)에서의 부분 색인 생성 과정을 설명하기 위한 도로, 구축기(200)는 적재되어 있는 데이터로 부분 색인을 생성하기 위해 공간 데이터 웨어하우스 서버(300)에 있는 기 구축된 색인의 상위부분을 전송 받는다.

색인의 상위부분 레벨은 색인의 높이 및 삽입되는 데이터의 개수에 의해 결정된다. 임시 자료 저장소(210)에 저장되어 있는 데이터는 기존 색인의 상위 부분을 통해 클러스터링 된다. 클러스터 풀에 있는 각 클러스터는 일반적인 색인 삽입 방법으로 부분 색인을 생성하게 된다.

도 5는 상기 구축기(200)에서 공간 데이터 웨어하우스 서버(300)로 부분 색인을 전송하는 과정을 설명하기 위한 도로, 상기 구축기(200)에서 생성된 부분 색인은 구축기(200)의 임시 자료 저장소(210)에 저장되어 있게 된다.

그리고 사용자가 지정해 놓은 적재 시점이 되면, 구축기(200)는 부분 색인을 위한 공간 예약 메시지를 공간 데이터 웨어하우스 서버(300)로 전송한다. 공간 데이터 웨어하우스 서버(300)는 부분 색인을 저장할 연속적인 익스텐트 공간을 예약한 후, 예약 완료 메시지를 구축기(200)로 전송한다.

구축기(200)는 예약 완료 메시지를 전송받으면 최상위 노드부터 중간 노드, 단말 노드 순으로 부분 색인을 전송한다. 단말 노드에 기록된 튜플 ID 값은 물리적 특성에 의해 공간 데이터 웨어하우스에서는 사용할 수 없기 때문에, 단말 노드는 단말 노드가 가리키는 해당 레코드와 함께 전송된다.

도 6은 상기 공간 데이터 웨어하우스 서버(300)로 전송되어 기록된 부분 색인의 구성을 나타낸 것으로, 상기 구축기(200)로부터 10개의 노드로 구성된 부분 색인이 전송된다고 가정했을 시 색인 전송 전후, 공간 데이터 웨어하우스 서버(300)의 자료 저장소(330)의 구조를 나타낸다.

공간 데이터 웨어하우스 서버(300)는 구축기(200)로부터 부분 색인이 전송된다는 메시지를 받으면 미 할당된 순차적 익스텐트를 예약한다.

부분 색인의 구성 즉, 전송받은 부분 색인을 공간 데이터 웨어하우스 서버(300)의 부분 색인 기록기(321)에서 기록하는 방법은 색인 구조 전송 순서와 같이 최상위 노드부터 중간 노드, 단말 노드 순으로 색인을 구성한다.

도 6에서 12번 노드는 2개의 아이템을 가지고 있으며 각 아이템이 지닐 자식노드 ID인, 페이지 오프셋 값은 13과 14번 노드임을 예상할 수 있다. 이를 이용하여 공간 데이터 웨어하우스 서버(300)는 구축기(200)에서 전송된 부분 색인을 다시 구성할 수 있으며, 부분 색인 저장 시 기록되어야 할 자식노드 ID를 자식 노드 할당 이전에 기록함으로써 자식 노드 기록을 위한 부모 노드 접근 비용을 제거하였다.

이를 보다 상세히 살펴보면, 공간 데이터 웨어하우스 서버(300)는 예약된 저장 공간을 이용하여 자식 노드에 대한 위치 정보를 예상함으로써 부모 노드 기록과 동시에 자식 노드 위치정보를 기록하여 자식 노드의 위치정보를 기록하기 위한 차후 부모 노드로의 접근 비용을 제거한다.

자식 노드 ID의 기록은 루트 노드에서 단말 노드 방향으로 기록하며 동일 레벨에서는 왼쪽에서 오른쪽 방향으로 기록한다. 전송된 노드의 엔트리들이 지니는 자식 노드 ID는 예약된 페이지 오프셋 값 중 제일 작은 값부터 큰 값 순으로 기록되기 때문에 각 노드 내 엔트리의 기록 과정 중 자식 노드에 대한 ID인 페이지 오프셋 값을 예상할 수 있다.

즉, 도 6은 루트 노드가 2개의 엔트리를 가지고 있는 경우로, 기록을 위해 예약된 페이지 오프셋 값들 중 루트 노드를 기록하기 위해 제일 작은 값인 12를 사용하고 루트 노드 내 엔트리가 지니는 자식 노드 오프셋 값으로 다음 값인 13과 14를 순차적으로 부여하게 된다. 따라서 이 과정은 실제 자식 노드인 13번과 14번을 할당 받기 이전에 기록 할 수 있으며, 이로 인해 자식 노드 기록을 위한 부모 노드로의 차후 접근 비용은 제거될 수 있다.

도 7은 공간 데이터 웨어하우스 서버(300)에서의 부분 색인 삽입 과정을 설명하기 위한 도로, 공간 데이터 웨어하우스 서버(300)는 구축기(200)에서 전송된 부분 색인을 기존에 구축되어 있던 전체 색인에 삽입한다. 삽입을 하기 위해 노드를 찾는 방법은 일반적인 색인의 검색 방법과 동일하다.

그러나 색인의 특성상 모든 단말 노드에서 높이가 같아야 하기 때문에 부분 색인이 가능한 높이를 계산한다. 계산된 높이에 있는 노드 중 부분 색인이 삽입될 적절한 노드가 일반적인 검색 방법을 통해 선택된다. 선택된 노드에 빈 엔트리가 있다면 부분 색인은 빈 엔트리에 삽입된다. 하지만 선택된 노드에 빈 엔트리가 없다면 부분 색인은 경험적인 방법을 사용하여 삽입된다.

깊이우선 탐색이나 너비우선 탐색 등의 Blind Search 방법은 목표까지의 경로를 찾는 상당히 소모적인(Exhaustive) 방법이다. 즉 문제에 대한 해를 제공하지만 너무 많은 노드를 확장시키므로 실용적이지 못하다. 많은 문제에 있어 탐색작업을 축소시키기 위해 항상 옳은 해를 제공하지는 못하지만 대부분의 경우에 잘맞는 경험에 의한 규칙(Rules of Thumb)을 이용하는 경우가 많다. 이렇게 그래프로써 표현된 문제에 대한 특별한 정보를 이용하여 탐색(Search)작업을 빠르게 진행시키는 방식을 경험적 탐색 방법이라고 하며, 사용된 정보를 경험적 정보(Heuristic Information)라 한다.

경험적 방법은 새로이 생성된 후계 노드들을 경험적 정보에 따라 정해지는 기준에 의해 순서를 정하거나, 재조정하는 것으로서 이렇게 함으로써 탐색은 가장 바람직한 부분을 확장시켜 나가게 될 것이다. 이렇듯 순서를 재조정하기 위해서는 노드의 바람직한 정도를 평가하기 위한 척도가 필요한데, 이 척도를 평가함수(Evaluation Function)라 한다.

평가함수의 목적은 확장시킬 노드들에게 순위를 매김으로써 어떤 것이 목표노드까지의 최상의 경로에 있음직한가를 결정하는 것이다. 평가함수는 여러 가지 착상에 근거를 두고 있다. 어떤 노드에 대하여 최상의 경로에 있을 확률을 이용하는 것도 있고, 임의의 노드와 목표 노드들간의 거리(Distance)나 차이(Difference)를 이용하는 경우도 있다. 게임이나 퍼즐에서는 목표까지의 가능성과 관계된 특성들에 근거를 두는 경우도 있다.

상술한 바와 같이, 본 발명의 바람직한 실시 예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 또는 변형하여 실시할 수 있다.

발명의 효과

이상에서 살펴본 바와 같이, 본 발명에 따른 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템 및 방법은, 구축기에서 추출된 데이터를 공간의 근접도가 아닌 색인의 구조에 맞게 클러스터링하며, 생성된 각 클러스터로 부분 색인을 생성하여 페이지 단위로 전송하고, 공간 데이터 웨어하우스 서버에서는 물리적 사상 문제를 해결하기 위해 물리적으로 연속된 공간을 예약하고 예약된 공간에 부분 색인을 기록하며, 기록된 부분 색인은 공간 데이터 웨어하우스 서버에 있던 기존 색인에 일반적인 삽입 알고리즘으로 삽입함으로써 색인 재구성을 위한 검색, 분할, 재조정 비용을 최소화함과 더불어 공간 데이터 웨어하우스 시스템이 지속적으로 안정된 처리를 가능케하여 시스템의 성능을 향상시킬 수 있게 된다. 또한, 공간 객체 삽입시 부모 노드 엔트리의 재구성 비용이 높은 R-Tree 계열 색인의 재구성 비용 절감에 보다 효율적이다.

(57) 청구의 범위

청구항 1.

공간 데이터 웨어하우스 시스템에 있어서,

소스 데이터베이스에서 추출되어 적재된 데이터에 대하여 색인의 구조에 맞게 클러스터링하며, 생성된 클러스터로부터 부분 색인을 생성하여 공간 데이터 웨어하우스 서버로 전송하는 구축기; 및

자신의 자료 저장소에 미 할당된 물리적으로 연속된 공간을 예약하여 예약된 공간에 상기 구축기로부터 전송된 부분 색인을 기록하며, 상기 기록된 부분 색인을 자신이 보유하고 있는 기존 색인에 삽입하는 공간 데이터 웨어하우스 서버;

로 구성됨을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 2.

제 1 항에 있어서, 상기 구축기는

페이지 단위로 부분 색인을 전송함을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 3.

제 1 항에 있어서, 상기 구축기는

상기 부분 색인을 생성하기 위해 상기 공간 데이터 웨어하우스 서버로부터 기 구축된 색인의 상위 부분을 전송받아 전송받은 색인의 상위 부분을 통해 클러스터링하는 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 4.

제 1 항에 있어서, 상기 구축기는

사용자가 지정해 놓은 적재시점이 되면 부분 색인을 위한 공간 예약 메시지를 상기 공간 데이터 웨어하우스 서버로 전송하여 공간 데이터 웨어하우스 서버로부터 예약 완료 메시지를 전송받으면 상기 부분 색인을 전송하는 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 5.

제 4 항에 있어서, 상기 구축기는

최상위 노드부터 중간 노드, 단말 노드의 순으로 부분 색인을 전송함을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 6.

제 5 항에 있어서, 상기 구축기는

부분 색인 전송시 상기 단말 노드에 대해서는 단말 노드가 가리키는 해당 레코드와 함께 전송함을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 7.

제 1 항에 있어서, 상기 공간 데이터 웨어하우스 서버에서 상기 미 할당된 물리적으로 연속된 공간은 연속적으로 예약된 익스텐트인 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 8.

제 7 항에 있어서, 상기 연속적으로 예약된 익스텐트는 노드들의 ID인 페이지 오프셋 값을 예측 가능하게 하는 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 시스템.

청구항 9.

소스 데이터베이스 및 구축기와 공간 데이터 웨어하우스 서버로 구성된 공간 데이터 웨어하우스 시스템에서의 색인 재구성 방법에 있어서,

상기 구축기에서 상기 소스 데이터베이스로부터 추출된 데이터를 색인의 구조에 맞게 클러스터링하는 제1 단계;

상기 클러스터링에 따라 생성된 클러스터로 부분 색인을 생성하여 전송하는 제2 단계;

상기 공간 데이터 웨어하우스 서버의 미 할당된 물리적으로 연속된 공간에 상기 전송된 부분 색인을 기록하는 제3 단계; 및

상기 기록된 부분 색인을 기존 색인에 삽입하는 제4 단계;

를 포함하여 이루어짐을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 10.

제 9 항에 있어서, 상기 제1 단계에서의 클러스터링은

상기 공간 데이터 웨어하우스 서버로부터 기 구축된 색인의 상위 부분을 전송받아 전송받은 색인의 상위 부분을 통해 행하는 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 11.

제 9 항에 있어서, 상기 제2 단계에서의 부분 색인의 전송은

사용자가 지정해 놓은 적재시점이 되면 부분 색인을 위한 공간 예약 메시지를 상기 공간 데이터 웨어하우스 서버로 전송하여 상기 공간 데이터 웨어하우스 서버로부터 예약 완료 메시지를 전송받으면 부분 색인을 전송하는 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 12.

제 11 항에 있어서, 상기 부분 색인의 전송은 페이지 단위로 이루어짐을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 13.

제 12 항에 있어서, 상기 부분 색인의 전송은

최상위 노드부터 중간 노드, 단말 노드의 순으로 행해짐을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 14.

제 13 항에 있어서, 상기 부분 색인 전송시 상기 단말 노드에 대해서는 단말 노드가 가리키는 해당 레코드와 함께 전송함을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 15.

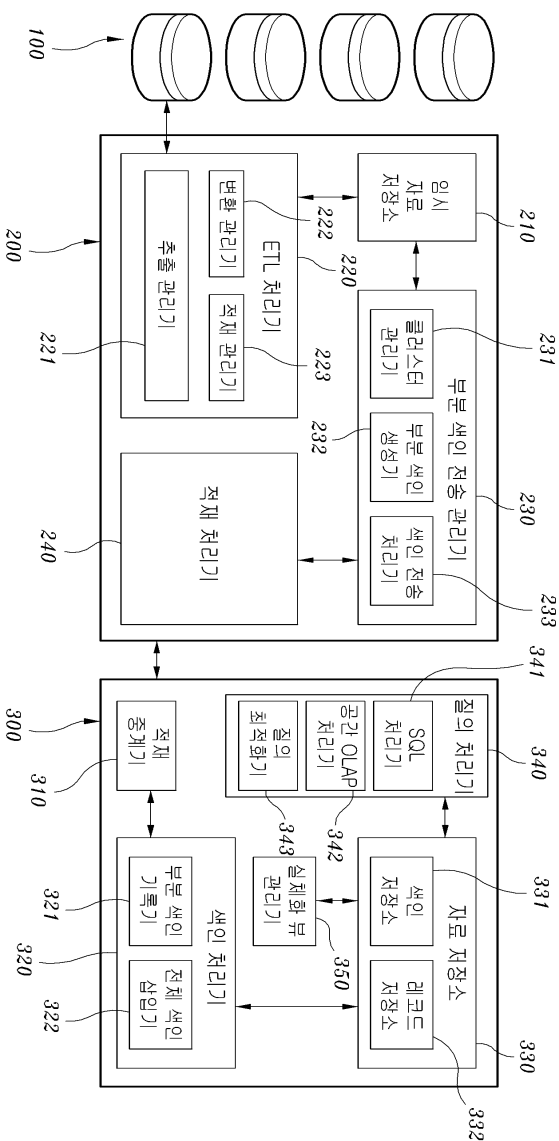
제 9 항에 있어서, 상기 미 할당된 물리적으로 연속된 공간은 연속적으로 예약된 익스텐트인 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

청구항 16.

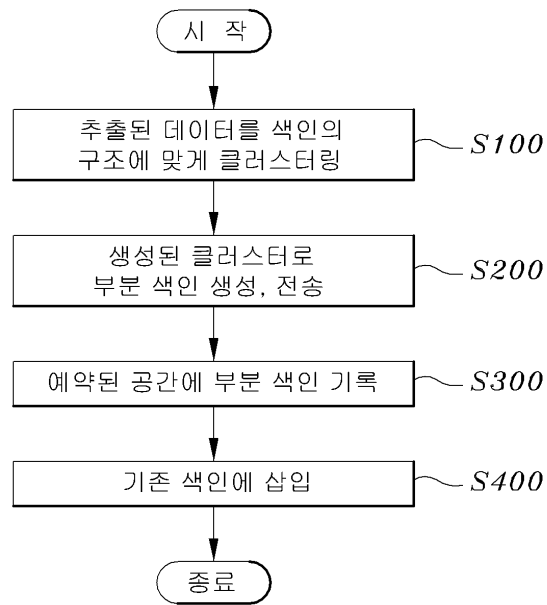
제 15 항에 있어서, 상기 연속적으로 예약된 익스텐트는 노드들의 ID인 페이지 오프셋 값을 예측 가능하게 하는 것을 특징으로 하는 공간 데이터 웨어하우스에서 부분 색인 전송을 이용한 색인 재구성 방법.

도면

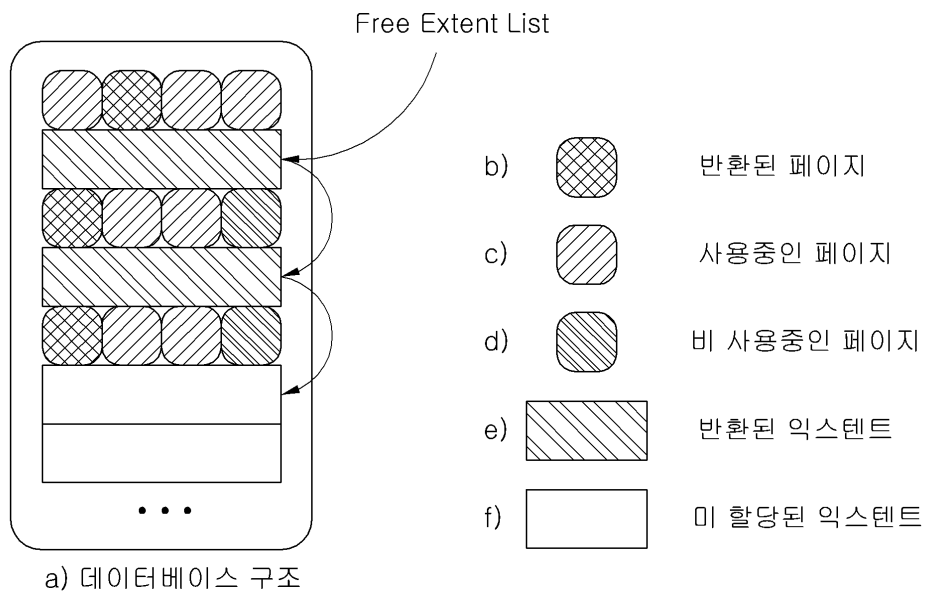
도면1



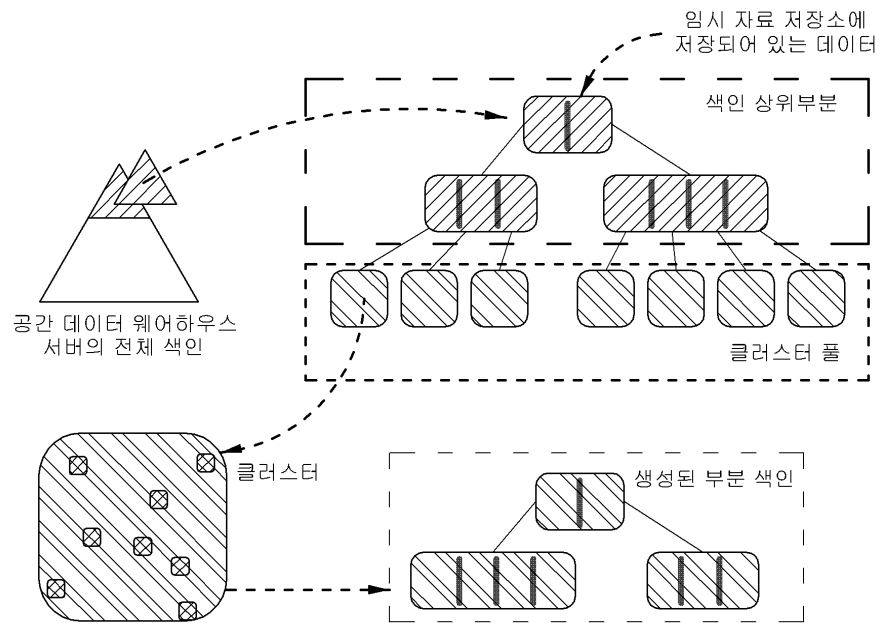
도면2



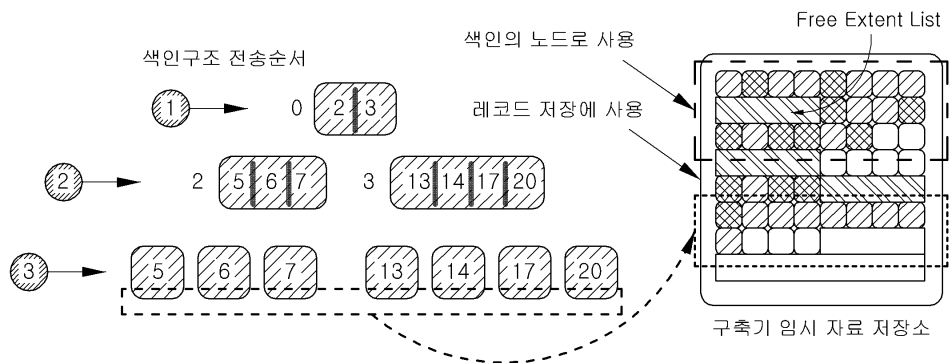
도면3



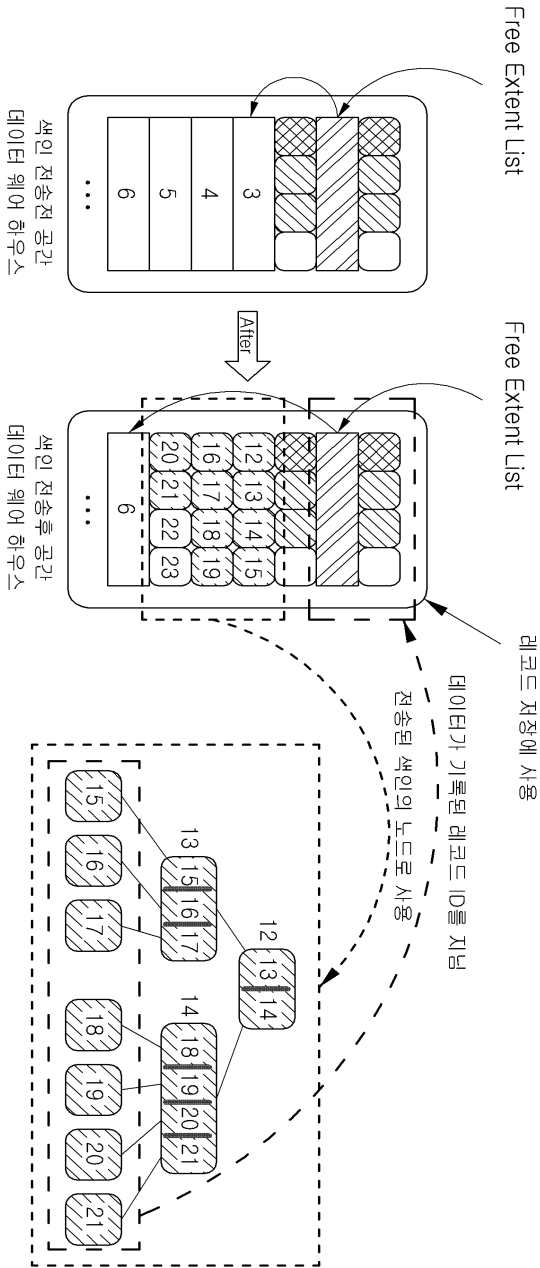
도면4



도면5



도면6



도면7

