

# MRSPAKE : A Web-Scale Spatial Knowledge Extractor Using Hadoop MapReduce

Seok-Jun Lee<sup>†</sup> · In-Cheol Kim<sup>††</sup>

## ABSTRACT

In this paper, we present a spatial knowledge extractor implemented in Hadoop MapReduce parallel, distributed computing environment. From a large spatial dataset, this knowledge extractor automatically derives a qualitative spatial knowledge base, which consists of both topological and directional relations on pairs of two spatial objects. By using R-tree index and range queries over a distributed spatial data file on HDFS, the MapReduce-enabled spatial knowledge extractor, MRSPAKE, can produce a web-scale spatial knowledge base in highly efficient way. In experiments with the well-known open spatial dataset, Open Street Map (OSM), the proposed web-scale spatial knowledge extractor, MRSPAKE, showed high performance and scalability.

**Keywords :** Spatial Knowledge Extraction, Topological Relation, Directional Relation, Hadoop MapReduce

# MRSPAKE : Hadoop MapReduce를 이용한 웹 규모의 공간 지식 추출기

이 석 준<sup>†</sup> · 김 인 철<sup>††</sup>

## 요 약

본 논문에서는 Hadoop MapReduce 병렬 분산 컴퓨팅 환경을 이용해 개발한 공간 지식 추출기를 제안한다. 이 공간 지식 추출기는 대용량의 공간 데이터 집합으로부터, 임의의 두 공간 객체들 사이에 만족되는 위상 관계와 방향 관계를 나타내는 정성 공간 지식 베이스를 생성해낸다. 본 논문에서 제안하는 MapReduce 기반의 대용량 공간 지식 추출기 MRSPAKE는 HDFS 상의 분산 공간 데이터 파일에 대한 R 트리 색인과 범위 질의들을 이용함으로써, 웹 규모의 정성 공간 지식 베이스를 매우 효율적으로 추출해낸다. 대표적인 공개 데이터 집합인 Open Street Map(OSM)을 이용한 성능 분석 실험을 통해, 본 논문에서 제안하는 웹 규모의 공간 지식 추출기 MRSPAKE의 높은 성능과 확장성을 확인할 수 있었다.

**키워드 :** 공간 지식 추출, 위상 관계, 방향 관계, Hadoop MapReduce

## 1. 서 론

공간과 관련된 지식 기반 지능형 시스템들은 객체 각각의 크기, 모양 등에 관한 세부적인 정보를 표현할 뿐만 아니라 이들 간의 물리적 관계를 표현하고 추론해낼 수 있어야 한다. 이러한 공간 지식의 표현과 추론(spatial representation and reasoning)은 지리 정보 시스템(GIS, geographical information system), 로봇 네비게이션(robot navigation), 컴

퓨터 비전(computer vision), 등의 분야에서 폭넓게 이용된다.

공간 지식을 습득하는 방법으로는 정성 공간 추론(qualitative spatial reasoning)과 정량 공간 추론(quantitative spatial reasoning)이 있다. 정성 공간 추론은 RCC(Region Connection Calculus)-8[1], CSD(Cone Shaped Direction Relations)-9[2] 등과 같이 공간 대수(spatial algebra) 이론들에 따라 초기 정성 공간 지식 베이스(initial qualitative spatial knowledge base)를 대상으로 추론 규칙들을 적용하여 새로운 위상 및 방향 관계 지식들을 유도해내는 방식이다. 대표적인 정성 공간 추론 시스템들로는 GQR[3], SparQ[4], SOWL[5], CHOROS2[6], QUSAR[7] 등이 있다. 한편, 정량 공간 추론(quantitative spatial reasoning)은 객체 각각의 고유 모양과 위치 정보를 나타내는 공간 데이터를 기초로 기하학적 연산(geometric computation)을 통해 두

\* 이 논문은 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신·방송 연구개발사업의 지원을 받아 수행한 연구 과제(No.10044494, WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발)인.

† 준 회 원 : 경기대학교 컴퓨터과학과 석사과정

†† 종 신 회 원 : 경기대학교 컴퓨터과학과 교수

Manuscript Received : October 4, 2016

Accepted : October 12, 2016

\* Corresponding Author : In-Cheol Kim(kic@kyonggi.ac.kr)

공간 객체 간의 정성적 공간 관계를 판별하는 방식이다. 대표적인 정량 공간 추론 시스템들로는 Parliament[8], QSRlib[9], SPAKE[10] 등이 있다. 정량 공간 추론은 새로운 정성 공간 지식을 얻기 위해 초기 정성 공간 지식 베이스의 구축을 요구하지 않는 대신, 정밀한 수치 데이터를 요구한다. 초기 지식 베이스의 결핍 문제를 가지고 있는 정성 공간 추론과는 달리, 정량 공간 추론은 Open Street Map[11], USGS[12], OS OpenData[13] 등과 같이 시맨틱 웹(semantic web) 상에서 잘 구축 되어 있는 웹 규모의 공개 데이터(open data)들을 활용할 수 있다. 이러한 이유로 정량 공간 추론은 정성 공간 추론이 가지고 있는 초기 정성 공간 지식 베이스의 결핍 문제를 해결하기 위한 대안이 될 수 있다. 또한, 수치 데이터에는 불확실성이 존재하기 때문에 정량 공간 추론과 정성 공간 추론은 서로 상호보완적으로도 운용이 가능하다. 하지만 현재까지 개발된 정량 공간 추론 시스템들은 모두 단일 머신 컴퓨팅 환경에서 동작하도록 개발되었기 때문에 웹 규모의 공개 데이터를 대상으로 정량 공간 추론을 수행하기에는 성능 상 한계가 존재한다.

본 연구에서는 웹 규모의 데이터로 인한 정량 공간 추론의 성능 상 한계점을 극복하기 위해 여러 대의 컴퓨터에서 작업을 분할하여 처리하는 분산 병렬 처리(distributed parallel processing) 방식을 선택하였다. 대표적인 분산 병렬 컴퓨팅 환경을 제공하는 프레임워크는 Hadoop MapReduce[14]가 있다. Hadoop은 입력 데이터를 여러 노드(node)에 분산하고 저장, 관리하기 위해 Hadoop 분산 파일 시스템인 HDFS(hadoop distributed file system)를 운용한다. 또한, MapReduce는 Map과 Reduce 함수를 이용하여 여러 노드에서 동시에 병렬로 데이터를 처리하여 대규모 데이터를 빠르게 처리할 수 있는 확장성을 가진다. 이러한 이유로 대표적인 MapReduce 기반의 추론기로는 RDFS와 OWL-Host 추론을 목적으로 개발된 Semantic 추론기인 WebPIE[15]와 QUSAR[7]를 확장한 대용량 정성 공간 추론기인 MRQUSAR[16] 등이 있다.

본 논문에서는 MapReduce의 뛰어난 확장성을 기반으로 정량 공간 추론 알고리즘을 설계하고, 이를 토대로 구현한 웹 규모의 정성 공간 지식 추출기인 MRSPAKE를 소개한다. 본 논문에서 제안하는 정량 공간 추론 알고리즘은 OGC(Open Geospatial Consortium) GeoSPARQL 표준 모델[17]을 따라 공간 객체들의 세부 정보와 이들 간의 위상 관계(topological relation)와 방향 관계(directional relation)를 표현하고 수학과 기하학적 이론을 기반으로 해당 관계들을 정량적으로 추론해낸다. 특히 본 논문에서 제안하는 방법은 HDFS 상에서 저장된 분산 공간 데이터 파일에 R-트리 색인[18]을 구축하여 지역화(localization)한 후 이를 효과적으로 이용함으로써, 대량의 정성 공간 지식을 매우 효율적으로 추출해낸다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구들을 살펴보고, 3장에서는 기초적인 공간 지식 추출 방법에 대해 소개한다. 4장에서는 MapReduce 기반의 분산 병렬 공간 지식 추출 방법 소개하고, 5장에서는 공간 지식 추출기의 구현과 Open Street Map 공개 데이터를 이용한 성능 분석 실험의 결과들을 소개한다. 마지막으로 6장에서는 결론을 기술한다.

## 2. 관련 연구

### 2.1 공간 지식 표현

공간 지식 표현은 크게 객체 각각의 고유 모양이나 위치 정보를 수치적으로 표현하는 정량 공간 표현과, 이진 관계 서술자(binary relation property)를 이용해 두 객체 간의 위상 관계(topological relation) 또는 방향 관계(directional relation)를 표현하는 정성 공간 표현이 있다. 대표적인 정량 공간 표현으로는 ISO(International Organization for Standardization)와 OGC(Open Geospatial Consortium) 표준에서 정의한 WKT(Well-Known Text)와 GML(Geography Markup Language) 등이 있다. WKT와 GML은 좌표 참조 시스템(coordinate reference system)을 토대로 점(point), 선(linestring), 면(polygon), 다중 점(multi-point), 다중-선(multi-linestring), 다중 면(multi-polygon) 등 총 13 가지의 객체 모양을 표현할 수 있다.

한편, 대표적인 정성 공간 표현으로는 위상 관계를 표현하기 위한 Randell의 RCC-8[1]과 Egenhofer의 Egenhofer-8[18], 방향 관계를 표현하기 위한 Frank의 CSD-9[2] 등이 있다. Randell의 연구에서는 두 개의 영역들(regions) 간에 가능한 모든 위상 관계들을 표현할 수 있는 DC(disconnected), EC(externally connected), PO(partially overlapping), EQ(equal), TPP(tangential proper part), TPPI(tangential proper part invers), NTPP(non-tangential proper part), NTPPI(non-tangential proper part inverse) 등 총 8 가지의 이진 관계 서술자들을 정의하였고, 이들을 기초로 새로운 관계 서술자들을 유도해내기 위한 추론 규칙(inference rule)들을 정의하였다. Egenhofer[18]의 연구에서는 영역(region) 뿐만 아니라 점(point), 선(linestring) 간의 위상 관계들도 표현할 수 있도록 disjoint, meet, equal, overlap, inside, contains, covers, coveredBy 등 8 가지의 이진 관계 서술자들을 정의하였고 이들에 기초한 추론 규칙들도 정의하였다. 한편, Frank[2]의 연구에서는 한 점(point)을 기준으로 다른 한 점(point)의 방향을 나타내는 방식으로, 두 점 간의 기본 방향(cardinal direction)을 표현할 수 있는 N(north), NE(north-east), E(east), SE(south-east), S(south), SW(south-west), W(west), NW(north-west), O(identical-to) 등 총 9 가지의 이진 관계 서술자들과 이들에 기초한 추론 규칙들을 제시하였다.

정확한 해를 요구하는 수학, 물리학 분야에서는 정량 공간 표현이 필수적이지만 사람은 정확한 해가 없어도 직관적으로 이해하기 쉬운 기호 수준의 정성 공간 표현이 더 필요한 경우가 많기 때문에 정성 공간 표현은 지리 정보 시스템이나 공간 전문가 시스템 등에서 매우 유용하게 이용된다.

### 2.2 공간 추론 시스템

앞서 소개한 공간 지식 표현들을 토대로 정성 공간 지식을 얻어낼 수 있는 방법은 크게 정성 공간 추론과 정량 공간 추론으로 나뉜다. 정성 공간 추론은 RCC-8, Egenhofer-8, CSD-9 등과 같은 공간 대수(spatial algebra) 이론들에 따라 초기 정성 공간 지식 베이스를 대상으로 관련 추론 규칙들

을 차례대로 적용함으로써 새로운 위상 및 방향 관계 지식들을 유도해내는 방식이다. 예컨대, 공간 객체 A, B, C 간에 <A contains B>, <B contains C>와 같은 정성 공간 지식을 이미 가지고 있다면, 정성 공간 추론을 통해 <A contains C>와 같은 새로운 정성 공간 지식을 유도할 수 있다. 대표적인 정성 공간 추론기들은 GQR[3], SparQ[4], SOWL[5], CHRONOS[6], QUSAR[7] 등이 있다. 정성 공간 추론은 정밀한 수치 데이터가 없어도 빠르고 직관적으로 새로운 관계를 유도해낼 수 있다. 하지만 정성 공간 추론 시스템들을 이용하기 위해서는 먼저 양질의 초기 정성 공간 지식 베이스가 존재해야 한다. 정성 공간 지식의 중요성이 부각되면서 정성 공간 지식 베이스를 구축하기 위한 노력들이 많이 진행되고 있으나, 아직까지는 범위나 규모면에서 충분한 정성 공간 지식 베이스가 구축된 사례는 거의 없는 실정이다.

한편, 정량 공간 추론은 객체 각각의 고유 모양과 위치 정보를 나타내는 공간 데이터를 기초로 기하학적 연산을 통해 두 공간 객체 간의 정성적 공간 관계를 판단하고 이를 토대로 새로운 정성 공간 지식을 유도해내는 방식이다. 대표적인 정량 공간 추론기들은 Parliament[8], QSRLib[9], SPAKE[10] 등이 있다. Parliament는 RDF/OWL 지식 저장소와 OGC GeoSPARQL 질의 처리기, 시-공간 추론 엔진을 보유한 시맨틱 웹(Semantic Web) 서버이다. Parliament는 DE-9IM 모델[19]을 기반으로 RCC-8, Egenhofer-8, Simple Feature 명세서[20]에서 정의하는 위상 관계 서술자들을 정량적으로 추론해낼 수 있다. 또한, 공간 데이터에 R-트리 색인을 구축하고 이를 토대로 공간 데이터를 지역화(localization) 함으로써 효과적으로 위상 관계를 판별할 수 있게 하였다. 하지만 Parliament는 2차원 객체들 간의 위상 관계 서술자들을 추론할 수 있으나, CSD-9과 같은 방향 관계 서술자들을 추론하지 못한다.

QSRLib은 두 객체 간의 다양한 정성 공간 지식을 정량적으로 추론해주는 Python 라이브러리이다. QSRLib에서 추론 가능한 정성 공간 지식은 IA(Interval Algebra)-13[21] 이론 기반의 시간 관계, RCC-2, RCC-4, RCC-5, RCC-8 이론 기반의 위상 관계, CSD-9, TPCC(Ternary Point Configuration Calculus)[22]이론 기반의 방향 관계 등이 있다. 하지만 QSRLib은 각 객체의 고유 모양대신에 고유 모양을 MBR(Minimal Bounding Rectangle)로 근사하고 이를 이용하기 때문에 정확성이 다소 떨어진다. 또한, QSRLib는 대규모 공간 데이터 집합으로부터 다자간의 정성 공간 관계 서술자들을 효과적으로 추론해내기 위한 특별한 색인 구조나 추론 방법을 제공하고 있지 않는다.

SPAKE[10]는 본 연구의 선형 연구로 개발된 공간 지식 추출기이다. SPAKE는 객체들 간의 공간 관계 서술자들을 정량적으로 추론하고 이를 토대로 공간 지식을 만들어내는 지식 공학 도구(knowledge engineering tool)의 역할을 수행할 뿐만 아니라 정량 추론을 통해 공간 질의에 대한 답을 구해주는 질의-응답 시스템의 역할도 수행한다. SPAKE가 추출하는 정성 공간 지식은 OGC GeoSPARQL 표준에서 정

의한 위상 관계 지식과 CSD-9에서 정의한 방향 관계 지식들이다. 또한, SPAKE는 공간 데이터들을 대상으로 R-트리 색인을 구축하고 이를 기반으로 형성된 공간 데이터의 지역성(locality)과 범위 질의(range query)를 이용하여 다자간의 위상 관계와 방향 관계를 효과적으로 분석하고자 하였다. 하지만 SPAKE와 마찬가지로 현재까지 개발된 정량 공간 추론기들은 공통적으로 단일 머신 컴퓨팅 환경(single machine computing environment)에서 동작하도록 개발되었기 때문에 웹 규모의 정량 공간 추론을 수행하기 위한 성능 상의 한계가 존재한다.

### 3. 정성 공간 지식 추출

#### 3.1 정성 공간 관계

정성 공간 지식 추출 방법을 설명하기에 앞서 공간 객체의 기하학적 데이터와 공간 지식이 어떻게 표현되는지, 즉 공간 지식 표현 체계를 먼저 정의할 필요가 있다. 본 논문에서 공간 지식은 시맨틱 웹 표준 온톨로지(ontology) 언어인 RDF/OWL로 표현되고, 특히 표준 공간 질의 언어인 OGC GeoSPARQL에서 정의한 클래스(class)들과 서술자(property)들을 사용하여 공간 객체들의 고유 성질과 관계들이 표현되는 것으로 가정한다. 하지만 GeoSPARQL에서는 두 공간 객체들 간의 위상 관계 서술자(topological property)들은 정의되어 있으나, 방향 관계 서술자(directional property)들은 정의되어 있지 않다. 따라서 본 논문에서는 GeoSPARQL의 핵심 온톨로지에 CSD(Cone-Shaped Directional relations)-9 이론에서 정의한 9가지 방향 관계 서술자들을 추가하여 Fig. 1과 같은 공간 온톨로지(spatial ontology)를 가정하였다.

Fig. 1에서 공간 객체(spatial object)는 모든 공간 객체들을 나타내는 최상위 클래스(class)이다. 두 공간 객체들 사이의 경계 및 포함 관계는 분리(disjoint), 맞닿음(touches) 등 총 7가지 위상 관계 서술자(topological property)들로 표현 가능하며, 두 공간 객체들 사이의 방향 관계는 북(north), 북동(north-east) 등 총 9가지 방향 관계 서술자(directional property)들로 표현할 수 있다. 한편, 공간 객체의 하위 클래스로 Feature와 Geometry가 있다. Feature는 실제계에서 도

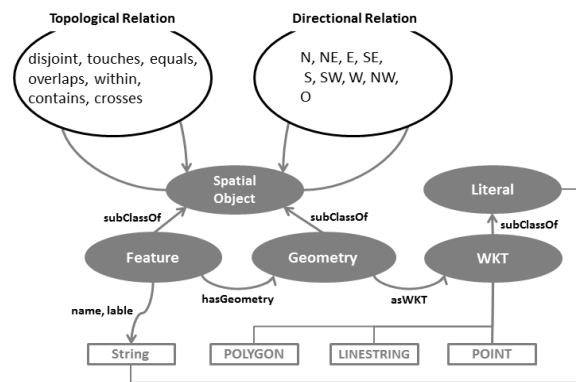


Fig. 1. Spatial Ontology

```

georesource:2297418
  rdf:type      geo:Feature ;
  rdf:name     "<Seoul>" ;
  geo:hasGeometry      geo:geom_2297418 .

georesource:200227274
  rdf:type      geo:Feature ;
  rdf:name     "HanGang" ;
  geo:hasGeometry      geo:geom_200227274 .

georesource:geom_2297418
  rdf:type      geo:Geometry ;
  geo:asWKT     "<http://www.opengis.net/def/crs/OGC/1.3/CRS84>
POLYGON((126.763987837.5549376,...))"
  ^^<http://www.opengis.net/ont/sf#wktLiteral> .

georesource:geom_200227274
  rdf:type      geo:Geometry ;
  geo:asWKT     "<http://www.opengis.net/def/crs/OGC/1.3/CRS84>
LINESTRING(127.3059215 37.5160799,...)"
  ^^<http://www.opengis.net/ont/sf#wktLiteral> .

georesource:200227274
  geo:crosses  georesource:2297418 .
    
```

Fig. 2. Example of Spatial Knowledge in Turtle Format

시, 도로, 건물과 같은 특정한 장소를 의미하고 반면에 Geometry는 점, 선, 면과 같은 Feature의 기하학적 데이터를 나타낸다. 그리고 Feature는 문자열(string) 형태의 리터럴(literal)로 표현되고 반면에 Geometry는 WKT(well-known text) 형태의 리터럴로 표현된다.

Fig. 2는 공간 지식 표현 체계에 따라 기술된 공간 지식의 한 예를 나타낸다. 이 예에서 서울과 한강은 각각 별도의 지오메트리를 통해 자신의 기하학적 데이터를 가지고 있다. 즉, 서울의 지오메트리는 2차원 좌표들로 구성된 면 형태(geom\_2297418 asWKT POLYGON)의 리터럴로 표현되어 있고, 한강의 지오메트리 역시 2차원 좌표들로 구성된 선 형태(geom\_200227274 asWKT LineString)의 리터럴로 표현되어 있다. 한편, Fig. 2는 “한강은 서울을 가로지른다(Hangang crosses Seoul)”라는 정성 공간 지식도 포함하고 있다. 본 논문에서는 이와 같은 개별 객체들의 공간 데이터로부터 둘 간의 위상 관계와 방향 관계를 분석하고 정성 공간 지식을 자동 생성하는 방법을 제시한다.

3.2 일대일 정성 공간 관계 판별

앞서 설명한 공간 지식 표현 체계를 토대로 두 공간 데이터 간에 만족하는 7 가지 위상 관계 서술자(disjoint, touches, equals, overlaps, within, contains, crosses) 중 하나를 판별하는 대표적인 종래의 방법은 DE-9IM 교차 행렬(Dimensionally Extended nine-Intersection Model)[19]을 이용하는 것이다.

$$DE-9IM = \begin{matrix} & \begin{matrix} Interior(I) & Boundary(B) & Exterior(E) \end{matrix} \\ \begin{matrix} dim(I(a) \cap I(b)) & dim(I(a) \cap B(b)) & dim(I(a) \cap E(b)) \\ dim(B(a) \cap I(b)) & dim(B(a) \cap B(b)) & dim(B(a) \cap E(b)) \\ dim(E(a) \cap I(b)) & dim(E(a) \cap B(b)) & dim(E(a) \cap E(b)) \end{matrix} & (1) \end{matrix}$$

두 공간 데이터 a와 b가 주어졌을 때, 둘 간의 DE-9IM 교차 행렬 계산하는 공식은 Equation (1)과 같다. 행렬의 가로 행과 세로 열은 기하학적 데이터의 내부(interior), 경

계선(boundary), 외부(exterior)을 의미하며 각 행렬 값은 가로행과 세로열의 교차된 면적( $X \cap Y$ )의 차원(dimension)을 의미한다.

예컨대, Fig. 3에서 객체 a의 내부와 객체 b의 내부의 교차된 면적( $dim(I(a) \cap I(b))$ )은 면이므로 차원은 2이고 a의 내부와 b의 경계선의 교차된 면적은 선이므로 차원은 1( $dim[I(a) \cap B(b)]=1$ )이다. 같은 방식으로, 교차된 면적이 점이면 차원은 0, 교차된 면적이 없으면( $\emptyset$ ) 차원은 -1이 된다. 따라서 Fig. 3의 행렬의 결과는 212101212이 되고, 불 방식(boolean)으로서 TTTTTTTTTT로 표현할 수 있다. DE-9IM 교차 행렬이 계산되면, Fig. 4의 판정표에 따라 두 공간 객체 간의 위상 관계 서술자를 판별해 낼 수 있다.

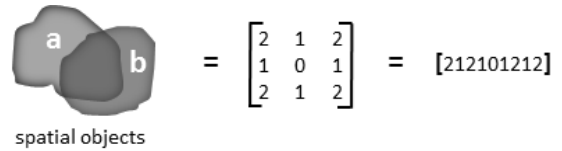


Fig. 3. Example of DE-9IM Intersection Matrix

Table 1은 DE-9IM 교차 유형(intersection pattern)별로 각각 대응하는 위상 관계 서술자들을 나타낸다. 두 공간 객체의 기하학적 데이터로부터 DE-9IM 교차 유형이 결정되면, 이것을 토대로 Table 1에 따라 두 공간 객체들 사이에 만족되는 위상 관계 서술자를 결정할 수 있다. 따라서 Fig. 3의 행렬의 결과 212101212 또는 TTTTTTTTTT는 Table 1에 의해 “객체 a는 b에 겹치다(a overlaps with b)”라는 위상 관계로 판별할 수 있다.

계산된 DE-9IM 교차 행렬의 결과는 다중으로 위상 관계 서술자를 의미하지 않는다. 즉, Fig. 3의 행렬의 결과는 overlaps라는 위상 관계만을 의미한다. 또한, DE-9IM은 두 객체가 가지는 고유 모양의 속성을 그대로 이용하기 때문에 모든 점, 선, 면 모양의 두 기하학적 데이터 간의 위상 관계를 매우 정확하게 분류할 수 있다는 장점이 있다. 하지만, 고유 모양을 그대로 이용한 방법은 그만큼 계산 시간이 오래 걸린다는 단점이 있다.

Table 1. Decision Table for Topological Relations

DE-9IM Intersection Pattern	Applies to Geometry Types	Topological Relationship
(F*F*F****)	All	<disjoint>
(F*****F, F**T*****, F***T****)	All except P/P	<touches>
(TTTTTTTTT)	All	equals
(T*T***T**) for A/A, P/P; (1*T***T**) for L/L	A/A, P/P, L/L	overlaps
(T*P**F****)	All	within
(T****FF*)	All	<contains>
(T*T***T**) for P/L, P/A, L/A; (0*****F) for L/L	P/L, P/A, L/A, L/L	crosses

Point(P), Line(L), Area(A)

앞서 설명한 공간 지식 표현 체계를 토대로 두 공간 객체 간에 만족하는 9 가지 방향 관계 서술자(north, north-east, east, south-east, south, south-west, west, north-west, identical) 중 하나를 판별하는 대표적인 종래의 방법은 두 객체의 MBR(Minimal Bounding Rectangle)을 구하고 이들의 중심점 간의 방향각을 이용하는 것이다.

$$ANGLE(p, p') = \arctan\left(\frac{p'_y - p_y}{p'_x - p_x}\right) * 180 / \pi \quad (2)$$

두 MBR의 중심점이 주어졌을 때, 둘 사이에 형성되는 방향각을 계산하는 공식은 Equation (2)와 같다. 방향각은 두 점 사이의 기울기의 역-탄젠트(arctan) 값을 이용하여 계산한다.

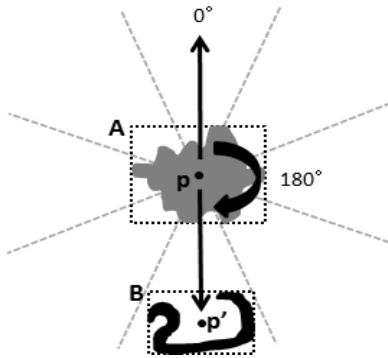


Fig. 4. Example of Direction Angle

예컨대, Fig. 4에서 두 공간 객체 A, B의 MBR과 이들의 중심점 p와 p'이 주어졌을 때, Equation (2)에 의해 p와 p' 사이에 형성되는 방향각은 180°로 계산된다. 두 MBR의 중심점이 이루는 방향각이 계산되면, Fig. 4의 판정표에 따라 두 공간 객체 간의 방향 관계 서술자를 판별해낼 수 있다.

Table 2는 두 MBR의 중심점이 이루는 방향각에 각각 대응하는 방향 관계 서술자들을 나타낸다. 두 공간 객체의 기하학적 데이터로부터 각 공간 객체를 둘러싸는 MBR의 중심점들을 구하고 이들 간의 방향각을 계산하면, 이것을 토

Table 2. Decision Table for Directional Relations

Angle	Applies to Geometry Types	Directional Relationship
[0° ~ 22.5°), [337.5° ~ 360°)	All	north
[22.5° ~ 67.5°)	All	north-east
[67.5° ~ 112.5°)	All	east
[112.5° ~ 157.5°)	All	<south>-east
[157.5° ~ 202.5°)	All	<south>
[202.5° ~ 247.5°)	All	<south>-west
[247.5° ~ 292.5°)	All	west
[292.5° ~ 337.5°)	All	north-west
P = P' (no angle)	All	identical

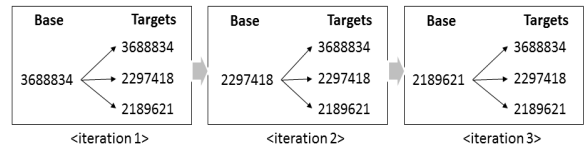
대로 Table 2에 따라 두 공간 객체 사이에 만족되는 방향 관계 서술자를 결정할 수 있다. 따라서 Fig. 4의 p와 p'이 이루는 방향각은 180°는 Table 2에 의해 객체 B는 A의 남쪽에 있다(B is south of A)는 방향 관계로 판별할 수 있다.

객체의 고유 모양 대신 MBR과 그 중심점을 이용하면 Fig. 4의 객체 B와 같이 불규칙적인 모양의 공간 객체들에 대해서도 쉽게 방향각을 구할 수 있다는 장점이 있다. 하지만, 이 방법은 고유 모양을 MBR의 중심점으로 근사했기 때문에 정확성이 다소 떨어진다는 단점이 있다.

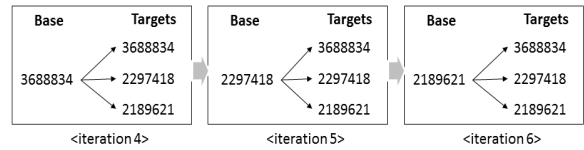
### 3.3 다자간 정성 공간 관계 판별

본 논문에서 제안하는 공간 지식 추출 방법은 다수의 공간 데이터가 주어졌을 때 이들 간의 가능한 모든 공간 관계를 판별해내야 한다. 이러한 다자간의 공간 관계를 판별하기 위해서는 입력되는 공간 데이터들 중에서 기준 데이터를 반복적으로 하나씩 선정하고 기준 데이터가 선정될 때마다 나머지 모든 공간 데이터들과의 위상 관계와 방향 관계를 분석해내야 한다. 위상 관계와 방향 관계 분석은 독립적으로 수행되기 때문에 공간 데이터 n개에 대해서 총 2n번의 분석 과정이 진행된다. 반복적으로 공간 관계를 판별하는 과정은 Fig. 5와 같다.

#### Iterative Topological extraction :



#### Iterative Directional extraction :



#### Output knowledge :

```

3688834 equals 3688834 2297418 <contains> 3688834 2189621 <disjoint> 3688834
3688834 within 2297418 2297418 equals 2297418 2189621 <disjoint> 2297418
3688834 <disjoint> 2189621 2297418 <disjoint> 2189621 2189621 equals 2189621
<iteration 1> <iteration 2> <iteration 3>
3688834 identicalTo 3688834 2297418 <south> 3688834 2189621 <south> 3688834
3688834 north 2297418 2297418 identicalTo 2297418 2189621 <south> 2297418
3688834 north 2189621 2297418 north 2189621 2189621 identicalTo 2189621
<iteration 4> <iteration 5> <iteration 6>
    
```

Fig. 5. Pair-Wise Iterative Knowledge Extraction

Fig. 5는 3개의 공간 데이터(3688834, 2297418, 2189621)에 대해서 가능한 모든 위상 관계와 방향 관계를 판별하는 과정을 나타낸다. 먼저, 위상 관계를 판별하기 위해서는 전체 데이터들 중에서 1개씩 기준(base)으로 선정하고 나머지 모든 데이터(target)와의 위상 관계를 분석하는 과정을 총 3번 반복해야 된다. 위상 관계 분석 과정이 반복될 때마다 기준으로 선정한 데이터를 기준으로 총 3개의 위상 관계가 판별되고 3번 반복되므로 총 9개의 위상 관계가 판별된다. 즉, 공간 데이터가 n개라면 n<sup>2</sup>개의 위상 관계가 판별된다.

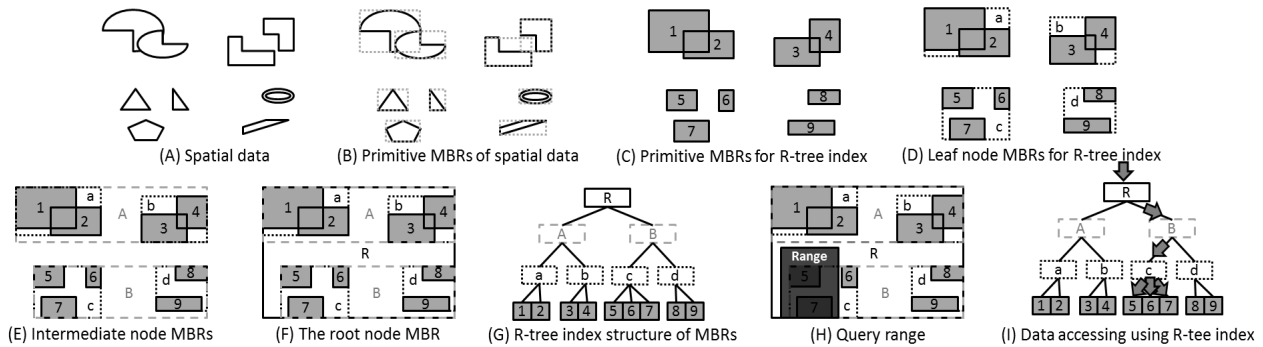


Fig. 6. Example of R-tree Index and Range Query

방향 관계 판별도 이와 같기 때문에 모든 공간 관계 분석 과정이 완료되면  $2n^2$ 개의 공간 관계가 판별된다. Fig. 5와 같은 다자간의 정성 공간 관계를 판별하는 과정은, 작업이 반복적으로 수행될 때마다 다수의 데이터들에 접근하기 위한  $2n^2$ 번의 많은 접근 비용과 위상 관계와 방향 관계를 분석하기 위한  $2n^2$ 번의 많은 분석 비용이 요구된다. 선행 연구[10]에서는 대량의 데이터에 대한 반복적인 접근과 위상 관계와 방향 관계를 효율적으로 분석하기 위해 공간 데이터를 대상으로 R-트리 색인을 구축하고 이를 토대로 형성된 데이터의 지역성을 이용하여 공간 데이터의 접근 비용과 공간 관계 분석 비용을 줄이고자 하였다. Fig. 6은 R-트리 색인과 이를 기반으로 수행하는 범위 질의의 예를 나타낸다.

공간 색인(spatial index) 중 가장 많이 쓰이는 R-트리 색인은 최소정계사각형(MBR, minimal bounding rectangle)을 기반으로 다차원 벡터 기하 데이터를 저장하고 효율적으로 접근하기 위한 트리 구조(tree structure)의 색인이다. R-트리의 핵심 아이디어는 공간 데이터들의 지역성이 보장되도록 최근접 이웃(nearest neighbor)끼리 MBR로 그룹(group)을 짓는 것이다. 예컨대, Fig. 6A와 같이 R-트리 색인을 구축하기 위한 벡터 기하 객체들이 존재한다면, 먼저 이들 각각의 초기 MBR들(primitive MBRs)을 계산할 수 있고(Fig. 6B), 초기 MBR들(1,2,3,4,5,6,7,8,9)이 계산되면(Fig. 6C), Fig. 6D의 a, b, c, d와 같이 서로 근접한 초기 MBR들을 최하위 노드 MBR들(leaf node MBRs)에 나누어 그룹을 짓는다. 본 논문에서 언급하는 최하위 노드 MBR들은 R-트리 색인에서 최하위 노드를 구성하는 MBR들을 의미한다. 그 다음 단계에서는 a, b, c, d를 다시 Fig. 6E의 A, B와 같이 중간 노드 MBR(intermediate node MBRs)들에 나누어 그룹을 짓고 이 과정을 반복하면서 모든 MBR들을 감싸는 최상위 노드 MBR(the root node MBR)까지 만든다(Fig. 6F). 이렇게 만들어진 MBR들을 단계별로 계층을 만들어 R-트리 색인을 구축할 수 있다(Fig. 6G). R-트리 색인의 뛰어난 데이터 접근성은 대표적으로 범위 질의(range query)를 수행할 때 확인할 수 있다. 예컨대, Fig. 6H에서 영역(range)에 해당하는 질의를 수행한다고 가정하면, 검색은 Fig. 6I와 같이 최상위 노드 R에서 시작하여 범위 질의의 영역에 교차(intersection)되는 노드만을 선택하면서 최하위 노드 c에 접근한다(R→B

→c). 노드 c에 접근하게 되면 노드 c가 가지는 데이터들(5, 6, 7)을 디스크에서 읽어와 범위 질의의 영역에 포함(contain)되는지 검사한 후 결과를 출력하게 되는데, 이 방법은 R-트리를 사용하지 않는 경우 디스크에서 모든 데이터(1, 2, 3, 4, 5, 6, 7, 8, 9)를 읽어 와서 범위 질의의 영역에 포함되는지 검사해야하는 방법에 비해 매우 효율적이다. 이러한 R-트리 색인으로 인한 데이터의 지역성을 이용하면 공간 데이터의 접근 비용과 위상 관계와 방향 관계 분석 비용을 효과적으로 감소시킬 수 있다. 후속 내용에서는 이들을 이용한 효과적인 위상 관계 판별 방법과 방향 관계 판별 방법에 대해 설명한다.

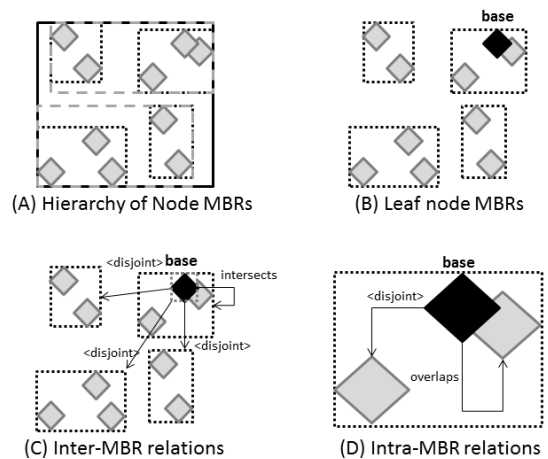


Fig. 7. Computing Topological Relations with Leaf Node MBRs

Fig. 7A는 음영으로 표기된 다이아몬드 모양의 공간 데이터들로부터 구해진 R-트리의 노드 MBR들과, 이들의 계층적 구조(hierarchy of node MBRs)를 나타낸다. 먼저, 우리는 Fig. 7A로부터 4개의 최하위 노드 MBR들(leaf node MBRs)을 분리해낸다(Fig. 7B). 이렇게 얻어진 최하위 노드 MBR들은 위상 관계 분석의 기준이 되는 데이터(base)의 MBR과의 영역이 교차(intersects)되는 MBR과 그렇지 않은(disjoint) MBR로 구분된다(Fig. 7C). 결국, 기준 데이터의 MBR과 최하위 노드 MBR들의 교차 관계를 구분함으로써, 기준 데이터와 영역이 교차되는 MBR 내의 데이터들과의

위상 관계는 실제로 Fig. 7D와 같이 떨어져 있거나(disjoint) 겹쳐져 있다(overlaps)와 같이 데이터 간의 DE-9IM 교차 행렬을 계산해야만 위상 관계를 판별할 수 있지만 그렇지 않은 MBR 내의 데이터들은 DE-9IM 교차 행렬 계산을 생략하고 서로 떨어져 있다(disjoint)는 위상 관계로 쉽게 판별해낼 수 있다. 따라서 Fig. 7의 방법은 기준 데이터의 MBR과의 영역이 교차되는 최하위 노드 MBR 내의 소수의 데이터들만 DE-9IM 교차 행렬을 계산하여 위상 관계를 판별하게 되므로 분석 비용을 상당히 감소시킬 수 있다.

앞서 설명한 위상 관계 분석 방법과 비슷하게, 선행 연구에서는 R-트리 색인의 최상위 노드 MBR(root node MBR)과 최상위 노드 MBR 내에서 형성되는 방향각의 영역(direction area)를 이용하여 방향 관계 분석 비용을 상당히 감소시켰다. 방향각의 영역(direction area)을 이용한 방향 관계 분석 방법은 Fig. 8과 같다.

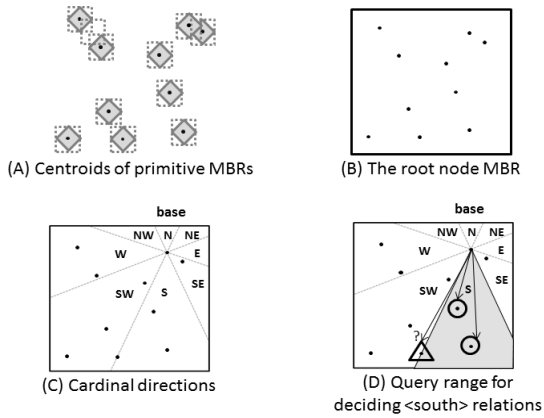


Fig. 8. Computing Directional Relations with the Top-Level MBR and Range Query

Fig. 8A는 음영으로 표기된 다이아몬드 모양의 공간 데이터로부터 구해진 초기 MBR들과 이들의 중심점들(centroids of primitive MBRs)을 나타낸다. 먼저, 우리는 구축된 R-트리 색인으로부터 모든 공간 데이터를 포함하는 최상위 노드

MBR(the root node MBR)을 분리해낸다(Fig. 8B). 이렇게 얻어진 최상위 노드 MBR 내에서 방향 관계 분석의 기준이 되는 데이터를 선정하고 기준 데이터 MBR의 중심점으로부터 8가지 기본 방향(cardinal directions)의 영역들(north(N), northEast(NE), east(E), southEast(SE), south(S), southWest(SW), west(W), northWest(NW))을 구해낸다(Fig. 8C). 결국, 기본 방향의 영역을 구함으로써, 방향각을 개별로 계산할 필요 없이 Fig. 8D의 남쪽 영역(S, south)과 같은 방향각의 영역으로 범위 질의를 수행하여 방향 관계를 판별해낼 수 있다. 따라서 Fig. 8의 방법은 각각의 데이터들과 방향각을 직접 계산하는 대신에 각 영역별로 R-트리 기반의 범위 질의를 수행하여 일괄적으로 방향 관계를 판별하기 때문에 분석 비용을 상당히 감소시킬 수 있다. 간혹, Fig. 8D의 남쪽 영역(S, south)과 같이 경계선(boundary) 상에 중심점이 존재할 수가 있는데, 이 경우에는 실제로 방향각을 계산해야만 정확한 방향 관계를 알 수 있다. 하지만 이런 경우는 매우 드물어 계산 시간에 큰 영향을 주지 않는다.

선행 연구에서는 R-트리 색인을 이용하여 다자간 공간 관계 판별 작업에서 반복적으로 요구되는 다수의 공간 데이터의 접근에 대한 부담을 줄이고 R-트리 색인을 토대로 형성된 데이터의 지역성을 이용하여 효과적인 위상 관계와 방향 관계 분석 방법을 제시하였다. 하지만 이러한 큰 장점에도 불구하고 선행 연구에서는 단일 컴퓨팅 환경에서 공간 관계 판별 알고리즘이 수행되는 것으로 가정하였기 때문에 웹 규모로 구축되어 있는 실세계 공간 데이터들을 대상으로 공간 지식을 추출하기 위한 확장성이 결핍되어 있다.

#### 4. MapReduce 기반의 대용량 공간 지식 추출

본 논문에서는 웹 규모의 정성 공간 지식을 추출하기 위해 MapReduce 기반의 정량 공간 추론 알고리즘을 설계하고 이를 토대로 웹 규모의 공간 지식 추출해낸다. 본 논문에서 제안하는 MapReduce 기반의 공간 지식 추출 작업순서도는 Fig. 9와 같다.

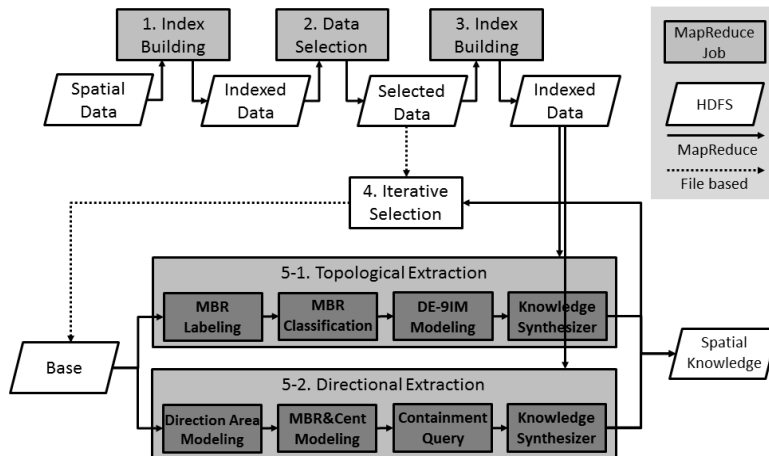


Fig. 9. Overall Process Flow for Qualitative Spatial Knowledge Extraction

첫 번째로, 우리는 전체 공간 데이터에 R-트리 색인 구축 작업을 수행한다(1. index building). 이 작업을 통해 무작위로 HDFS 상에 흩어져 있는 공간 데이터들이 HDFS 블록 크기(HDFS block size)만큼 최하위 노드 MBR 단위로 그룹을 지어 여러 노드에 재분배된다. 전체 공간 데이터에 한 번 구축된 R-트리 색인은 이후 반복되는 작업 데이터 선정 작업을 위해 계속 유지된다. 두 번째로, 구축된 R-트리 색인을 이용하여 지식 추출 작업을 위한 데이터들의 범위 선정 작업을 수행한다(2. data selection). 이 작업은 범위 질의, k-최근접 이웃(k-nearest neighbor)과 같은 공간 질의를 이용한다. 세 번째로, 선택된 데이터들을 대상으로 별도의 R-트리 색인 구축 작업을 수행한다(3. index building). 본 논문에서 제안하는 공간 지식 추출 방법은 R-트리 색인을 기반으로 재분배된 데이터의 지역성을 이용하여 위상 관계와 방향 관계를 판별하기 때문에 이 작업은 필수적이다. 첫 번째 작업에서 구축된 R-트리 색인과는 달리 한 번 지식을 모두 추출한 이후에는 지속적으로 색인을 유지하지 않는다. 네 번째는 지식을 추출하기 전에, 선정된 데이터 중 하나를 기준이 되는 데이터로 선정한다(4. base selection). 기준 데이터는 모든 선정된 데이터들이 순서대로 한 번씩 선정되고 기준 데이터가 선정될 때 마다 나머지 모든 데이터와의 지식 추출 작업이 수행된다. 기준 데이터 선정 작업은 MapReduce 작업(job)이 아니다. 마지막으로, 기준 데이터가 선정되면 나머지 모든 데이터와의 위상 관계 추출 작업(5.1 topological extraction)과 방향 관계 추출 작업(5.2 directional extraction)을 수행한다. 각각의 작업에 대한 상세 내용은 후속 절에서 설명한다.

4.1 R-트리 색인 생성

본 논문에서 제안하는 대용량 공간 지식 추출 방법은 R-트리 색인을 이용함으로써, 데이터 선정 작업과 지식 추출 작업에서 요구되는 대량의 데이터 접근 비용과 처리 비용을 효과적으로 절감하였다. 하지만 MapReduce는 병렬 처리에 좋은 함수형 프로그래밍을 채택한 반면에, 전통적인 R-트리 색인의 구조는 순차적으로 명령어들을 실행하는 절차형 프로그래밍에 최적화되어 있다. 또한, HDFS 상의 한 파일은 적절한 크기의 여러 파일로 분열(splitting)되어 여러 슬레이브 노드들에서 독립적으로 처리되지만 전통적인 R-트리 색인은 한 파일을 모두 메모리에 상주시키고 공유하기 때문에 적은 메모리 용량의 한계와 동시에 접근하여 병렬 처리할 경우에 오동작을 유발할 수 있다. 따라서 HDFS에서는 새로운 구조의 R-트리 색인이 요구된다. Fig. 10은 HDFS에서 적용 가능한 R-트리 색인의 구조를 나타낸다.

R-트리 색인의 구조는 Fig. 10의 전역 색인(global index)과 지역 색인(local indexes)과 같이 두 층(layer)으로 구성되어 있다. 각각의 슬레이브 노드들에는 R-트리 색인의 최하위 노드 MBR 단위로 그룹 지어진 공간 데이터들이 지역 색인 파일(part-xxxxx.rtree)에 저장되어 있고 이들의 메타 정보는 마스터 노드에서 하나의 전역 색인 파일(\_master.rtree)로 저

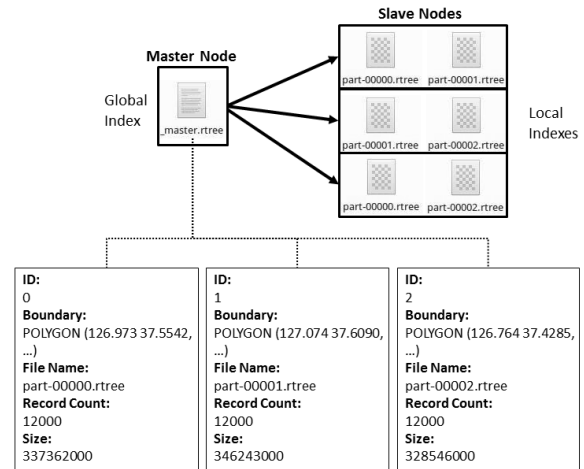


Fig. 10. R-tree Index Structure in HDFS

장된다. 전역 색인에 저장된 메타 정보는 지역 색인의 식별자(ID), 경계(boundary), 지역 색인 파일의 이름(file name), 데이터의 수(record count), 파일 크기(size)들을 포함한다. 특히, 각 슬레이브 노드들에 저장된 공간 데이터들은 부하 분산(load balancing)을 위해 HDFS 블록 크기(HDFS block size)만큼 할당되어 저장된다. 즉, R-트리의 최하위 MBR들은 HDFS 블록 크기 단위로 만들어진다. Fig. 10의 R-트리 색인을 구축하기 위한 처리 과정은 Fig. 11과 같다.

색인 구축 작업 과정은 크게 분할(partitioning), 지역 색인 구축(local indexing), 전역 색인 구축(global indexing)의 3단계로 구성된다. 먼저, 분할 단계에서는 입력 데이터를  $n$  개의 파티션(number of partition)으로 분할한다.  $n$ 을 정하는 공식은  $n = \frac{S(1+\alpha)}{B}$ 이다. 이 식에서  $B$ 는 HDFS 블록 크기(HDFS block size)이고  $S$ 는 입력 데이터의 크기,  $\alpha$ 는 내고장성을 위해 각 노드에 복제되어 저장되는 입력 데이터를 고려한 초과용량 비율(overhead ratio)이다.  $\alpha$ 의 값은 기본(default)으로 0.2로 설정한다.  $n$ 이 결정되면 각각의 파티션의 경계들(partition boundaries)을 정해야한다. 파티션의 경계들을 정하는 방법은 STR(Sort-Tile-Recursive) 알고리즘[23]을 이용한다. 하지만 STR 알고리즘은 단일 프로세서에서 수행하기 때문에 대용량의 모든 입력 데이터를 메모리에 상주시킬 수 없다. 따라서 입력 데이터의 샘플을 무작위로 선정(random sampling)하고 이를 이용하여 STR 알고리즘을 수행한다. 입력 데이터의 샘플을 무작위로 선정하는 일은 하나의 MapReduce 작업으로 진행한다. 입력 데이터는 각각 분할되어 Map 함수의 입력으로 전달된다. 만약, 입력 데이터의 모양이 점이 아니면 MBR의 중심점으로 변환되어 Map 함수에 전달된다. Map 함수에서는 입력 데이터를 1%의 확률로 결과로 출력한다. 만약, 출력된 결과의 크기가 100 MB가 넘으면 결과를 다시 입력 데이터로 하여 100 MB 이하가 될 때까지 무작위 샘플 선정 작업이 반복된다. 샘플이 선정되면,  $\sqrt{n}$ 을 STR 알고리즘의 매개 변수(parameter)  $d$ (R-tree degree)로 설정하고 샘플을 대상으로



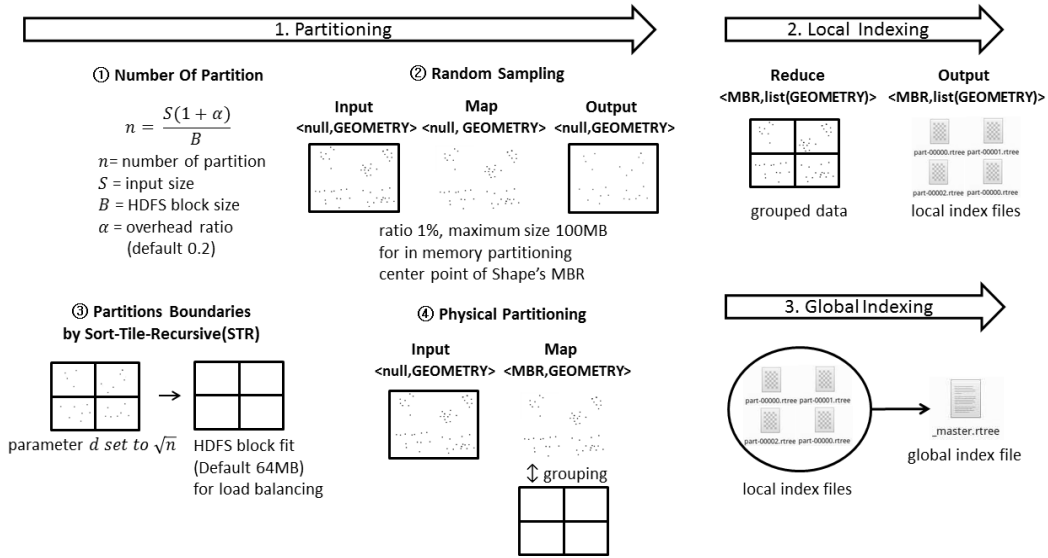


Fig. 11. Process for Building R-tree Index

STR 알고리즘을 수행한다. STR 알고리즘은 최소경계사각형(MBR, minimal bounding rectangle)으로 파티션의 경계를 정하기 때문에 각각의 파티션의 모양은 MBR이 되고 이는 곧 R-트리 색인의 최하위 노드 MBR이 된다. 추후 입력 데이터 부하분산을 위해 데이터들이 HDFS 블록 크기만큼 파티션에 포함되도록 경계를 정한다. 파티션의 경계가 결정되면, 각각의 파티션에 포함되는 데이터들을 그룹지어 재분배하는 물리적 파티션 분할(physical partitioning) 작업을 수행한다. 물리적 파티션 분할과 이후 진행되는 지역 색인 구축 단계는 하나의 MapReduce 작업이다. 먼저, 입력 데이터가 분열되어 각각의 Map 함수로 전달된다. Map 함수는 전달받은 데이터가 어느 파티션(=최하위 노드 MBR)에 속하는지 판단하고 파티션과 이에 속하는 데이터를 키-값 쌍(key-value pair)으로 만들어 Reduce 함수에 전달한다. 지역 색인 구축 단계에서는 이를 전달받아 Reduce 함수를 수행한다. Reduce 함수에서는 MapReduce의 셔플링(shuffling)을 거쳐 각각의 파티션을 키로, 이에 속하는 데이터들을 리스트(list)로 전달받아 파티션 별로 하나의 지역 색인 파일(part-xxxxx.rtree)을 결과로 출력한다. 지역 색인 파일들은 파티션에 속한 데이터들과 이 MBR의 경계에 대한 정보를 가진다. 마지막으로 전역 색인 구축 단계에서는 만들어진 모든 지역 색인들의 메타 정보를 담고 있는 전역 색인(\_master.rtree)을 만들어 마스터 노드에 저장한다.

Fig. 12는 Fig. 11의 무작위 샘플 선정 작업의 한 예를 나타낸다. 먼저, 대표 샘플 데이터들을 추출하기 위해 입력되는 모든 공간 데이터(geom\_1~6)는 각 Map 함수로 분열(datasplitting)된다. Map 함수에서는 분열되어 입력된 데이터들에 대해 사전에 정한 비율(ratio) 1%를 토대로 데이터들을 무작위로 선정(random selection)하고 선정된 데이터들(geom\_1, geom\_4, geom\_6)을 결과로 출력(data writing)한다. 만약, 결과의 크기가 100 MB보다 크면 결과를 다시 입력으로 하여 100 MB 이하가 될 때까지 작업을 반복한다.

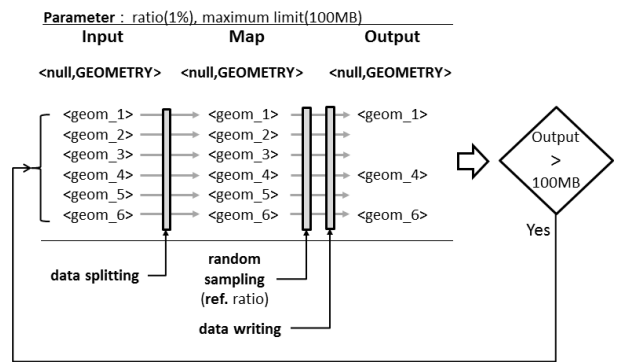


Fig. 12. Random Sampling in MapReduce

Fig. 13은 Fig. 11의 물리적 파티션 분할과 지역 색인 구축의 한 예를 나타낸다. 먼저, 초기 MBR(initial MBR)과 각 MBR로 그룹 짓기 위한 모든 공간 데이터들(geom\_1~6)이 입력되면 각각의 Map 함수로 입력 데이터를 분열(data splitting)한다. Map 함수에서는 입력된 데이터가 어느 MBR의 영역에 속하는지 검사한 후 해당 MBR과 함께 그룹을 지어(data grouping) Reduce 함수에 전달한다. 만약, 데이터가 어느 MBR에도 속하지 않으면 가장 거리가 가까운 MBR

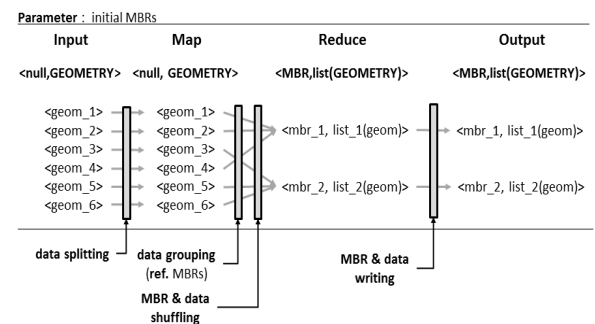


Fig. 13. Physical Partitioning and Local Indexing in MapReduce

에 속하는 것으로 간주한다. Reduce 함수는 동일한 파티션을 키로 하여 값들을 리스트로 묶어주는 셔플링(MBR & data shuffling)을 거쳐 전달받은 파티션(mbr\_1, mbr\_2)과 해당 파티션에 속한 데이터들의 리스트(list\_1, list\_2)를 토대로 파티션의 영역(boundary)을 수정하고 하나의 파티션당 해당 데이터들과 함께 하나의 이진 형태의 지역 색인 파일로 출력(MBR & data writing)한다.

#### 4.2 작업 범위 선정

모든 공간 데이터들에 대한 R-트리 색인이 구축되면 이를 이용하여 지식을 추출 작업을 위한 입력 데이터를 선정하는 작업을 진행한다. 데이터 선정 작업은 범위 질의, k-최근접 이웃과 같은 공간 질의를 이용하여 수행한다. R-트리 색인의 지역성을 이용하면 공간 질의 수행에서 데이터 접근 비용과 처리 비용을 상당히 감소시켜주기 때문에 데이터 선정 작업에 큰 도움이 된다. 본 논문에서는 범위 질의를 토대로 데이터 선정 작업을 수행하는 것으로 가정하여 설명한다. MapReduce 기반의 데이터 선정 작업 처리 과정은 Fig. 14와 같다.

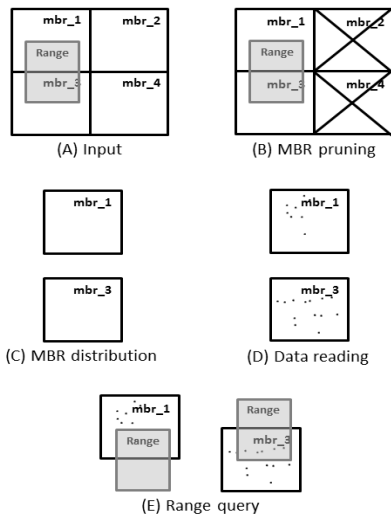


Fig. 14. Process for Data Selection

데이터 선정 작업은 R-트리 색인의 최하위 MBR들(이하 MBR)과 주요 관심의 대상이 되는 공간 데이터들이 대거 포함된 범위(range)를 입력(input)받는다. 특히, MBR들을 Map 함수로 분배(MBR distribution)하기 전에 불필요한 데이터 접근 비용을 줄이기 위해 입력 받은 범위와 MBR의 영역이 교차(intersection)되는지 검사한 후 교차되지 않는 MBR이 Map 함수로 분배되지 않도록 사전에 가지치기(MBR pruning)를 한다. 이후 가지치기 되지 않는 MBR들이 Map 함수로 분배되고 각각의 Map 함수에서 각 MBR에 속한 공간 데이터들을 읽어온다(data reading). 마지막으로 읽어 들인 공간 데이터들이 실제로 범위에 속하는지 검사(range query)한 후 범위에 속하면 결과로 출력한다. Fig. 15는 위에 설명한 데이터 선정 작업 처리 과정의 한 예를 보인다.

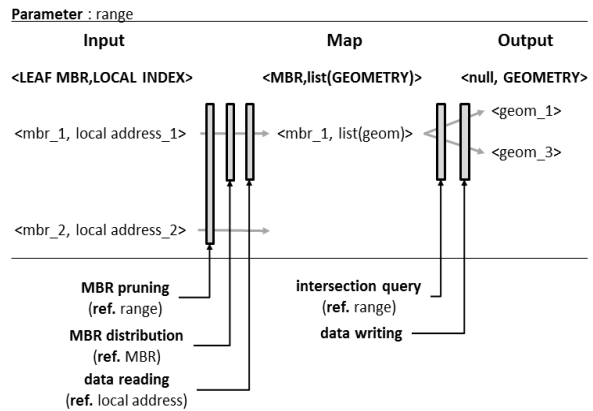


Fig. 15. Data Selection in MapReduce

먼저, 데이터 선정 작업을 진행하기 위해서 R-트리 색인의 MBR들(mbr\_1, mbr\_2)과 이에 속한 공간 데이터들이 저장된 지역 색인의 주소(local address\_1, local address\_2), 그리고 범위(range)를 입력 받는다. mbr\_1과 mbr\_2를 각각의 Map 함수로 전달하기 전에, 범위(range)와 영역이 교차되는지 검사하고 영역이 교차되는 mbr\_1은 Map 함수로 분배하고 그렇지 않은 mbr\_2은 가지치기를 하여 Map 함수로 분배하지 않는다. 즉, 범위와 관련이 없는 데이터들을 사전에 여과하여 불필요한 데이터 접근과 처리를 생략한다. Map 함수는 전달된 mbr\_1에 속한 데이터들을 모두 읽어와 리스트(list) 형태로 만들고 리스트에 속한 데이터들이 실제로 범위에 속하는지 검사하고 그 중 범위에 속한 geom\_1과 geom\_2를 결과로 출력한다.

#### 4.3 위상 관계 지식 추출

효과적인 공간 지식 추출 작업들을 위해서, 데이터 선정 작업이 완료되면 선정된 데이터들을 대상으로 재차 R-트리 색인을 구축하고 최하위 노드 MBR 단위로 데이터들을 그룹지어 재분배한다. 위상 관계를 추출하기 위해 선정된 데이터들 중에서 한 번씩 기준 데이터를 선정하고 기준 데이터를 선정할 때마다 나머지 모든 데이터와의 위상 관계 추출 작업을 수행한다. 한 번 기준 데이터를 선정했을 때 수행하는 위상 관계 추출 작업의 처리 과정은 Fig. 16과 같다.

위상 관계 추출 작업은 R-트리 색인의 최하위 노드 MBR들(이하 MBR)과 이에 속한 공간 데이터들의 주소, 그리고 위상 관계 분석의 기준(base)이 되는 데이터 1개를 입력(input)받는다. 특히, 불필요한 위상 관계 분석 비용을 줄이기 위해서 MBR들을 Map 함수들에 분배하기 전에 각 MBR들이 기준 데이터의 MBR과 영역이 교차하는지 검사하고 MBR에 '교차(intersection)' 또는 '비-교차(non-intersection)'라는 레이블을 붙인다(MBR labeling). 이후 레이블이 붙은 MBR들을 각각 Map 함수로 분배(MBR distribution)하고 Map 함수에서는 각 MBR에 속한 공간 데이터들을 읽어 온다(data reading). 기준 데이터와 읽어 들인 데이터 간의 위상 관계를 판별하기 전에 미리 MBR에 붙여 놓은 레이블을 보고 데이터 간의 DE-9IM 모델링(DE-9IM modeling)을 해야만 위상 관계를 판별할 수

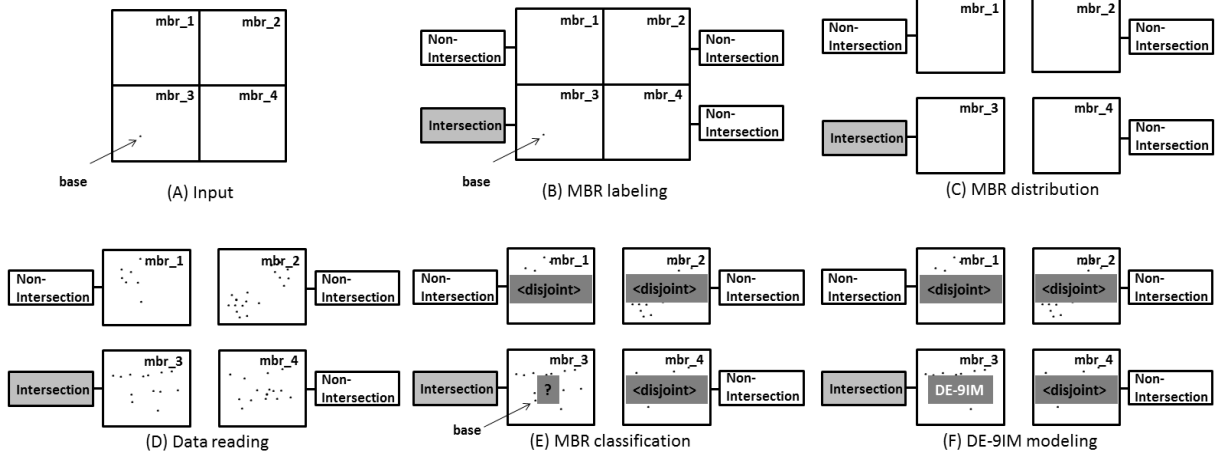


Fig. 16. Process for Topological Knowledge Extraction

있는 MBR과 DE-9IM 모델링을 생략하고 위상 관계를 판별할 수 있는 MBR들을 분류(MBR classification)한다. 결국, 기준 데이터의 MBR과 영역이 교차된 MBR 내에 속한 공간 데이터들 간에는 DE-9IM 모델링을 토대로 위상 관계를 판별한 후 지식으로 추출하고 그렇지 않은 MBR에 속한 공간 데이터들 간에는 DE-9IM 모델링을 생략하고 모두 기준 데이터와 ‘떨어져 있다(disjoint)’는 위상 관계로 판별한 후 지식으로 추출한다. Fig. 17은 위상 관계 추출 작업 처리 과정의 한 예를 보인다.

먼저, 위상 관계 추출 작업을 진행하기 위해서 R-트리 색인의 MBR들(mbr\_1, mbr\_2)과 이에 속한 공간 데이터들이 위치한 주소(local address\_1, local address\_2), 그리고 기준 데이터(base geometry)를 입력 받는다. mbr\_1과 mbr\_2를 각각의 Map 함수로 전달하기 전에, 기준 데이터의 MBR과 영역이 교차되는지 검사하고 영역이 교차되는 mbr\_1에 ‘교차’, 그렇지 않은 mbr\_2에 ‘비-교차’ 레이블을 붙이고(MBR labeling) Map 함수로 분배(MBR distribution)한다. Map 함

수는 전달된 mbr\_1과 mbr\_2에 속한 데이터들을 모두 읽어와 리스트(list) 형태로 만들고 DE-9IM 모델링을 해야 하는 mbr\_1과 이를 생략하고 위상 관계를 판별할 수 있는 mbr\_2를 분류(MBR classification)한다. MBR들이 분류되면 mbr\_1에 있는 데이터들은 기준 데이터와의 DE-9IM 모델링을 토대로 위상 관계를 판별하고 mbr\_2에 있는 데이터들은 DE-9IM 모델링을 생략하고 기준 데이터와 ‘떨어져 있다(disjoint)’라는 위상 관계로 판별한다. 위상 관계가 판별되면 위상 관계 지식을 합성(knowledge synthesizing)하여 결과로 출력(knowledge writing)한다.

#### 4.4 방향 관계 지식 추출

위상 관계와 마찬가지로, 방향 관계를 추출하기 위해서는 선정된 데이터들 중에서 한 번씩 기준 데이터를 선정하고 기준 데이터를 선정할 때마다 나머지 모든 데이터와의 방향 관계 추출 작업을 수행한다. 한 번 기준 데이터를 선정했을 때 수행하는 방향 관계 추출 작업의 처리 과정은 Fig. 18과 같다.

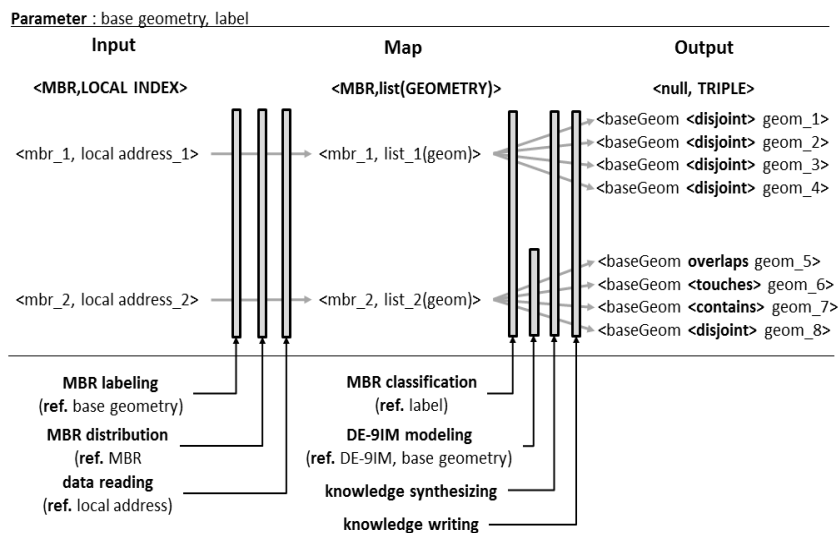


Fig. 17. Topological Knowledge Extraction in MapReduce

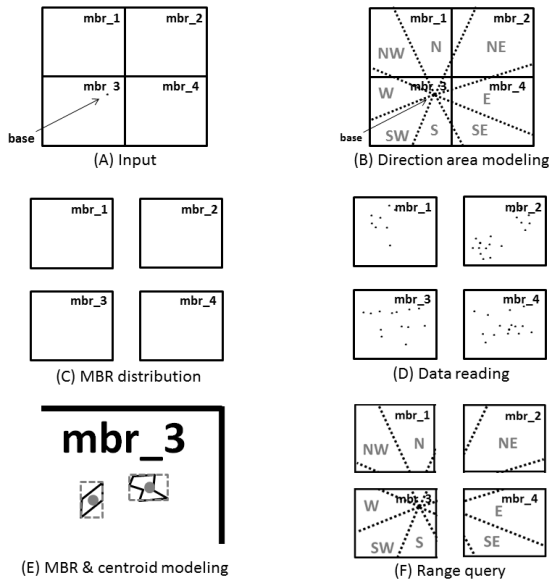


Fig. 18. Process for Directional Knowledge Extraction

방향 관계 추출 작업은 R-트리 색인의 최하위 노드 MBR들(이하 MBR)과 방향 관계 추출의 기준(base)이 되는 데이터 1개를 입력(input)받는다. 특히, 방향각을 직접 계산하는 것보다 처리 비용이 적게 드는 범위 질의를 이용하기 위해서 기준 데이터의 MBR을 구하고 그 중심점을 토대로 최상위 노드 MBR 내에서 형성되는 방향각의 영역들을 모델링(direction area modeling)한다. 이후 레이블이 붙은 MBR들을 각각 Map 함수로 분배(MBR distribution)하고 Map 함수에서는 각 MBR에 속한 공간 데이터들을 읽어 온다(data reading). 방향각 영역으로 범위 질의를 수행하기 전에 읽어 들인 데이터들의 MBR과 그 중심점을 모델링(MBR & centroid modeling)한다. 결국, 데이터의 MBR의 중심점들이 8가지 방향각의 영역 중 어느 곳에 속하는지에 따라 방향 관계를 판별한다. 만약, 중심점이 방향각 영역의 경계선 상에 존재하여 8가지 방향각의 영역 중 어느 곳에도 속해 있지 않

는다면 실제로 방향각을 계산하여 방향 관계를 판별한다. Fig. 19는 방향 관계 추출 작업 처리 과정의 한 예를 보인다.

먼저, 방향 관계 추출 작업을 진행하기 위해서 R-트리 색인의 MBR들(mbr\_1, mbr\_2)과 이에 속한 공간 데이터들이 위치한 주소(local address\_1, local address\_2), 그리고 기준 데이터(base geometry)를 입력 받는다. mbr\_1과 mbr\_2를 각각의 Map 함수로 전달하기 전에, 기준 데이터의 MBR을 구하고 그 중심점을 토대로 최상위 노드 MBR 내에서 형성되는 방향각의 영역들을 모델링(direction area modeling)하고 mbr\_1과 mbr\_2를 각각 Map 함수에 분배(MBR distribution)한다. Map 함수는 전달된 mbr\_1과 mbr\_2에 속한 데이터들을 모두 읽어와 리스트(list\_1, list\_2) 형태로 만든다. 리스트 내의 데이터들은 순차적으로 하나씩 이들의 MBR과 그 중심점으로 모델링(MBR & centroid modeling)되고 중심점이 8가지 방향각 영역으로 범위 질의를 수행하여 어느 곳에 속하는지에 따라 방향 관계를 판별한다. 방향 관계가 판별되면 방향 관계 지식을 합성(knowledge synthesizing)하여 결과로 출력(knowledge writing)한다.

### 5. 구현 및 실험

본 논문에서 제안한 웹 규모의 정량 공간 추론의 성능을 확인하기 위하여 Hadoop MapReduce 기반의 공간 지식 추출기를 구현하였다. 공간 지식 추출기는 Ubuntu 14.04.3 LTS, Java 1.7, Hadoop 2.5.0 환경에서 구현하였다. 클러스터는 1대의 마스터 노드와 9대의 슬레이브 노드들로 구성하였으며, 각 노드는 3.5GHz, 4 Core CPU와 8GB 메인 메모리, 1TB 하드 디스크로 구성하였다. Hadoop MapReduce 환경에서 R-트리 색인을 구축하고 R-트리 기반의 범위 질의를 수행하기 위해 SpatialHadoop 라이브러리[24]를 이용했고 DE-9IM 모델 계산을 위해 JTS(Java Topology Suit) 라이브러리[25], 웹 인터페이스(web interface) 구현을 위해 OpenLayer 라이브러리[26]를 이용하였다.

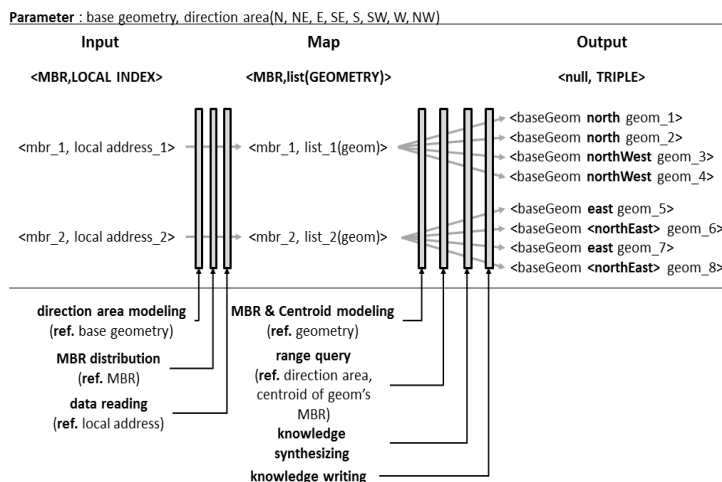


Fig. 19. Directional Knowledge Extraction in MapReduce

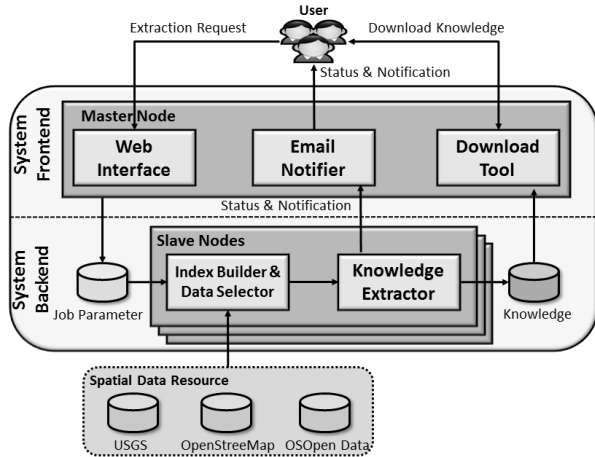


Fig. 20. System Architecture of MRSPAKE

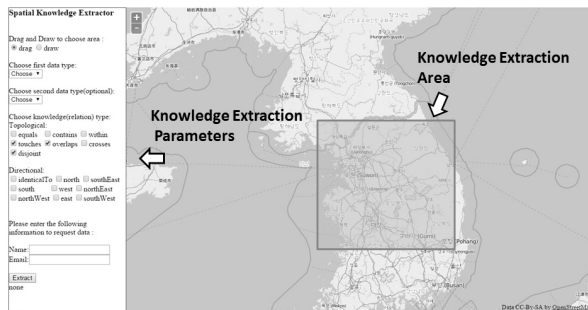


Fig. 21. Web Interface of MRSPAKE

Fig. 20은 본 논문에서 제안한 웹 규모의 정량 공간 추론 알고리즘을 바탕으로 구현한 MapReduce 기반의 공간 지식 추출기(MRSPAKE, MapReduce based spatial knowledge extractor)의 개략적인 구조도를 나타낸다. 시스템 외부적으로, 사용자는 MRSPAKE의 front-end를 통해서 MRSPAKE와의 상호 작용(interaction)을 할 수 있다. front-end는 웹 인터페이스(web interface), 이메일 통보기(email notifier), 다운로드 도구(download tool)로 구성된다. 먼저, 사용자는 Fig. 21과 같이 웹 인터페이스를 통해 지도상에서 중요도와 관심을 고려해서 사각형 모양의 작업 범위의 영역(knowledge extraction area)을 지정하고 지식 추출 작업을 위한 매개 변수들(knowledge extraction parameter)을 설정하여 MapReduce 작업을 요청한다. 중요도와 관심을 고려한 지식 추출 영역을 선정하는 일은 불필요한 지식 추출과 이로 인해 추출된 지식 용량의 과장성(expansion) 때문에 꼭 필요한 과정이다. 요청된 정량 추론 작업이 back-end에서 수행되면 사용자는 이메일 통보기로부터 작업 진행 상태를 통보받고 작업이 완료되면 사용자는 다운로드 도구를 이용하여 back-end에 저장된 추출된 지식을 N-Tuple 형태의 텍스트 파일로 받아올 수 있다. 시스템 내부적으로, MRSPAKE의 back-end는 사용자로부터 요청된 작업을 각각의 슬레이브 노드들에서 분산 병렬 처리하여 공간 지식을 추출해낸다. back-end는 색인 구축기와 데이터 선정기(index builder & data selector), 지식 추출

기(knowledge extractor)로 구성된다. 먼저, 색인 구축기는 사전에 시맨틱 웹 상에서 공개된 도시, 도로, 하천 등의 지리 정보를 포함한 공간 데이터들(USGS, OpenStreetMap, OSOpen Data)로부터 R-트리 색인을 구축하고 R-트리 색인의 최하위 MBR들을 토대로 가까운 거리에 있는 공간 데이터들을 그룹지어 재분배한다. 이후 지식 추출 작업이 요청되면 데이터 선정기(data selector)가 사용자가 입력한 사각형의 지식 추출 영역을 입력받아 범위 질의를 수행한다. 선정된 데이터는 색인 구축기에 의해 재차 R-트리 색인 구축 및 데이터 재분배가 이루어지고 사용자가 설정한 매개 변수를 토대로 지식 추출 작업을 처리하여 지식을 추출한다.

Table 3. Dataset for Experiments

Dataset	Capacity	Point(n)	Linestring(n)	Polygon(n)	Total(n)
Open Street Map Roads	24.1GB	3,529	2,111,1934	70,224,463	91,339,926

본 논문에서는 Table 3과 같이 Open Street Map[11] 공개 데이터를 이용하여 공간 지식 추출기의 성능 분석 실험을 수행하였다. Open Street Map은 전 세계 범위의 행정구역, 도로, 하천 등의 대규모 공간 데이터들을 포함하고 있으나, 실험에서는 전 세계 범위의 도로와 관련된 데이터들만 주로 이용하였다. 실험은 크게 정량 실험과 정성 실험으로 나뉜다.

먼저, 정량 실험에서는 웹 규모의 지식 생성량과 위상 관계와 방향 관계 지식 추출의 성능적인 측면에서의 효율성과 확장성을 분석해보는 실험들을 수행하였다. Fig. 22A는 지식 추출 작업이 반복적으로 수행될 때마다 추출되는 지식의 양(the number of triples)의 측정 결과를 나타낸다. 추출된 지식의 양은 위상 관계 지식과 방향 관계 지식의 양을 모두 합한 양이다. 지식의 종류별로 기준 데이터 1개에 대해서 약 9천 1백만 개의 나머지 모든 데이터와의 이진 관계가 추출되므로 선정된 기준 데이터가 10개이면 약 9억 1천만 개(19GB)의 지식이 추출되고 선정된 기준 객체가 50개이면 약 45억 6천만 개(95GB)의 대규모 지식이 추출된다. 만약, 모든 데이터를 한 번씩 기준으로 선정하여 정량 추론 작업을 진행하면 9천 1백만의 제공만큼의 지식이 추출된다. 용량은 약 172,900 TB(172 PB)로 추정된다. 이와 같은 결과는 웹 규모의 정량 공간 추론 작업이 대용량의 지식을 추출해낼 수 있음과 동시에 저장소에 담을 수 없을 만큼의 초 대규모 지식을 만들어내므로 중요도와 관심을 고려한 추론 데이터의 범위 선정이 필요함을 보여준다.

Fig. 22B는 R-트리 색인을 이용하지 않은 정량 추론 작업과 R-트리 색인을 이용한 정량 추론 작업 처리 시간의 측정 결과를 나타낸다. 처리 시간은 정량 추론이 반복적으로 수행할 때마다 측정하였으며, 위상 관계 추출과 방향 관계 추출 시간의 합으로 측정하였다. (Fig. 23B)의 결과는 모든 구간에서 R-트리 색인을 이용한 정량 추론 작업이 R-트리 색인을 이용하지 않은 정량 추론 작업보다 더 빠른 처리 시간을 보인다. 특히, R-트리 색인을 이용한 정량 추론 작업이 그래프의 경사가 더 완만하므로 성능 상 더 효과적인

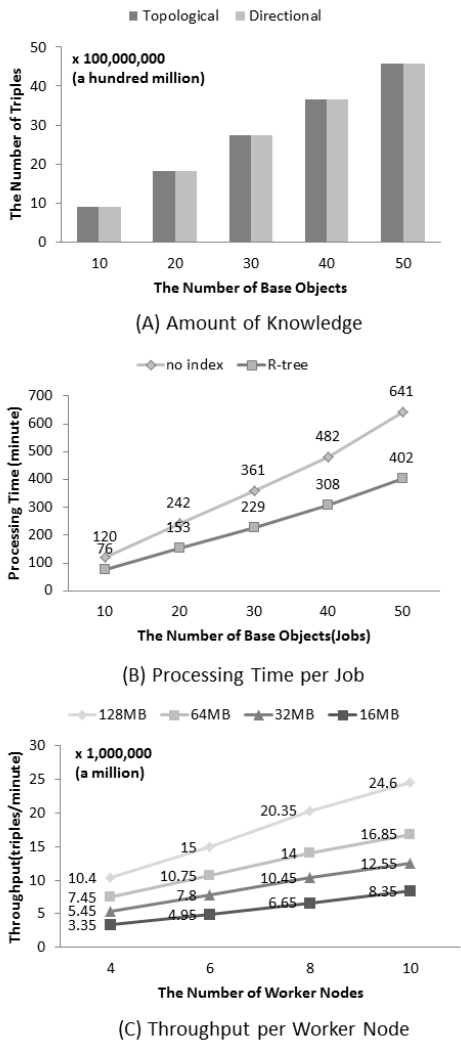


Fig. 22. Experimental Results with the Open Street Map (OSM) Roads Dataset

방법임을 입증해준다. 이러한 이유는 4장에서 기술한 바와 같이, R-트리 색인을 이용하면 위상 관계를 판별할 경우 대부분의 DE-9IM 모델 계산을 생략할 수 있고 방향 관계를 판별할 경우 개별로 방향각을 계산하는 것보다 범위 질의를 이용하여 일괄적으로 판별할 수 있기 때문이다.

Fig. 22C는 작업 노드의 수가 많아질수록 분당 추출되는 지식의 양의 측정 결과를 나타낸다. 실험을 위해 기준 데이터 1개를 임의로 선정하고 나머지 모든 데이터들과의 정량 추론 작업을 수행하였으며, 이 과정을 10번 반복하고 각각 측정된 분당 지식 추출량을 평균을 결과로 측정하였다. 분당 지식 추출량은 위상 관계 지식과 방향 관계 지식의 양을 합한 양이다. 특히, 각 작업 노드의 수별로 HDFS 블록 크기를 16MB부터 128MB까지 변경하여 측정하였기 때문에 총 4개의 그래프가 형성된다. 실험 결과는 작업 노드의 수가 많아질수록, HDFS 블록 크기가 커질수록 분당 지식 추출량이 더 많음을 보일뿐만 아니라 작업 노드의 수가 많아짐에 따라 분당 지식 추출량의 증가폭이 더 커짐을 보인다. Fig. 23C에서 보이는 최고 성능은 작업 노드의 수가 10개이고 HDFS 블록 크기가 128MB일 때 분당 약 2천 4백 6십만 개의 지식 추출량이다. 이러한 결과는 본 논문에서 제안한 웹 규모의 정량 공간 추론 방법이 매우 빠르게 지식을 추출할 뿐만 아니라 병렬 처리에 좋은 함수 프로그래밍 환경에서 매우 뛰어난 확장성을 가진다는 것을 입증해준다.

정성 실험에서는 추출된 공간 지식이 실제 지도와 일치하는지 비교해보았다. 먼저, Fig. 23A는 한강(Han Gang)과 한강 주변의 주요 도시들(Goyang, Seoul, Guri, Hanam, Namyangju)을 표기한 맵과 이들을 대상으로 추출한 일부 위상 관계와 방향 관계 지식들을 나타낸다. 먼저, “한강은 서울을 가로지른다<Han Gang> <crosses> <Seoul>”, “서울은 고양과 맞닿아 있다<Seoul> <crosses> <Goyang>”, “서울은 남양주와 떨어져 있다<Seoul> <disjoint> <Namyangju>”

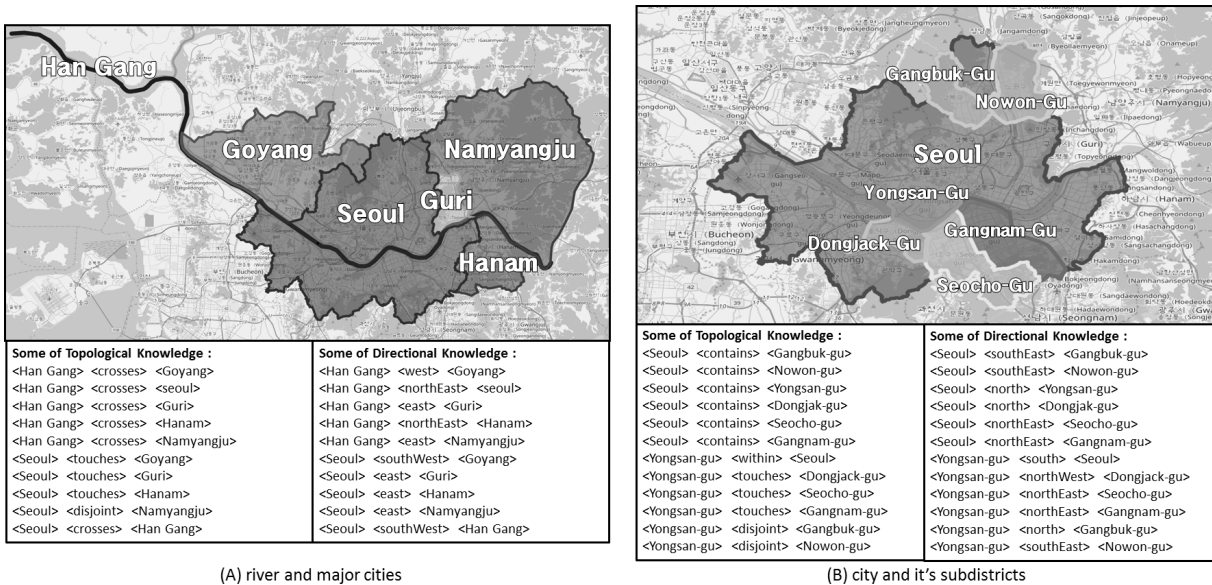


Fig. 23. Sample Spatial Knowledge Produced by MRSPAKE and the Corresponding Real Map

등과 같이 위상 관계 지식은 지도와 완벽히 일치하는 결과를 보인다. 반면, 방향 관계 지식은 대부분 지도와 일치하지만 그렇지 못한 경우도 있었다. 예컨대, “서울은 고양의 남서쪽에 있다<Seoul> <southWest> <Goyang>”와 “서울은 구리의 동쪽에 있다<Seoul> <east> <Guri>” 등과 같이 도시 간의 방향 관계 지식은 지도와 정확히 일치하지만 “한강은 고양의 서쪽에 있다<Han Gang> <west> <Goyang>”, “한강은 서울의 북동쪽에 있다<Hang Gang> <northEast> <Seoul>”와 같이 한강과 도시 간의 방향 관계 지식은 지도 불일치하였다.

다음으로, Fig. 23B는 주요 도시 중 서울과 서울에 포함된 하부 행정구역(subdistrict)들을 표기한 맵과 이들을 대상으로 추출한 일부 위상 관계와 방향 관계 지식을 나타낸다. Fig. 23A와 마찬가지로, “용산구는 서울에 포함된다<Yongsan-gG> <within> <Seoul>”, “용산구는 동작구 맞닿아 있다<Yongsan-Gu> <crosses> <Dongjack-gG>” 등과 같이 위상 관계 지식은 지도와 완벽히 일치하지만 방향 관계 지식은 그렇지 못한 경우도 있었다. 예컨대, “용산구는 강북구의 남쪽에 있다<Yongsan-Gu> <south> <Gangbuk-Gu>”와 같이 서로 분리(disjoint) 관계이거나 “용산구는 동작구의 북서쪽에 있다<Yongsan-Gu> <northWest> <Dongjack-Gu>”와 같이 서로 맞닿음(touches) 관계일 경우에는 방향 관계가 지도와 일치하지만 “서울은 용산구의 북쪽에 있다”와 같이 서로 포함(contains, within) 관계에 있는 객체들 간의 방향 관계는 납득하기 힘든 지식이었다.

위와 같은 정성 실험을 통해 객체의 고유 모양을 그대로 이용하여 추출하는 위상 관계 지식은 모두 지도와 일치하는 것을 보이지만 객체의 고유 모양을 MBR과 그 중심점으로 근사하고 이를 이용하여 추출하는 방향 관계 지식은 한강과 같은 불규칙적인 모양의 객체를 대상으로 할 경우에 일부 맞지 않는다는 것을 보여준다. 또한, 서로 분리(disjoint)되거나 맞닿음(touches) 관계인 객체들 간의 방향 관계 지식은 올바른 지식을 추출하지만 서로 포함(contains-within) 관계에 있는 객체들 간의 방향 관계 지식은 납득하기 힘든 것을 통해 방향 관계 지식의 질(quality)이 위상 관계에 영향을 받는다는 것을 보인다.

## 6. 결 론

본 논문에서는 Hadoop MapReduce 분산 병렬 컴퓨팅 환경을 이용해 개발한 웹 규모의 공간 지식 추출기 MRSPAKE를 제안하였다. MRSPAKE는 HDFS 상의 분산 공간 데이터 파일에 대한 R 트리 색인과 범위 질의들을 이용함으로써, 웹 규모의 정성 공간 지식 베이스를 매우 효율적으로 추출해낸다. 대표적인 공개 데이터 집합인 Open Street Map(OSM)을 이용한 성능 분석 실험을 통해, 본 논문에서 제안하는 웹 규모의 MRSPAKE의 높은 성능과 확장성을 확인할 수 있었다.

본 논문에서 진행한 연구는 정량 공간 추론을 Hadoop

MapReduce 분산 병렬 컴퓨팅 환경에서 잘 적용하였으나 아래와 같은 이유로 향후 연구를 진행할 계획이다. 첫 번째로, 위상 관계와는 달리 방향 관계 계산 방법은 아직 잘 정립된 이론이 없는 실정이다. 이러한 이유로 본 논문에서는 각 공간 데이터의 모양을 MBR의 중심점으로 근사하여 기본 방향을 판단하고 있다. 이 방법은 공간 데이터의 고유 모양을 MBR의 중심점으로 근사했기 때문에 정밀도와 정확성이 다소 떨어진다. 따라서 공간 데이터의 고유 모양을 유지한 방향 관계 계산 방법이 필요하다. 두 번째로, 본 논문에서 제안한 방법은 대량의 공간 데이터에 대해서 모든 가능한 이진 관계를 전부 구해내기 때문에 저장소가 감당하기 힘든 양의 지식을 추출한다. 따라서 중요도와 관심을 고려해서 필요한 만큼의 현실적인 용량의 공간 지식을 추출하기 위한 지식 추출 대상 선정에 관한 연구가 진행되어야 한다. 현재로서는 범위 질의를 이용하여 관심 데이터를 선정하고 있지만 행정 구역, 도로, 하천 등의 공간 데이터 유형, 크기, 모양과 같이 좀 더 구체적인 데이터의 속성을 고려하여 관심 데이터를 선정하는 일이 필요하다. 마지막으로, Hadoop MapReduce는 연속되는 작업들 사이에 발생하는 데이터들을 모두 파일 단위로 디스크에 입출력하기 때문에 많은 반복적인 작업들로 이루어진 정량 공간 추론 작업의 성능 향상에 한계가 있다. 따라서 Apache Spark와 같이 메모리 캐싱(caching) 기법을 이용하여 작업 간의 입출력을 최적화한 인메모리(in-memory) 기반의 분산 병렬 컴퓨팅 환경에서 정량 공간 추론 작업의 처리 방법을 연구할 계획이다.

## References

- [1] J. Renz, “Maximal Tractable Fragments of the Region Connection Calculus: A Complete Analysis,” *Proc. of IJCAI*, 1999.
- [2] D. J. Peuquet and C. X. Zhang, “An Algorithm to Determine the Directional relation between Arbitrarily-Shaped Polygons in the Plane,” *Pattern Recognition*, Vol.20, No.1, pp.65-74, 1987.
- [3] Z. Gantner, M. Westphal, and S. Wolf, “GQR: A Fast Reasoner for Binary Qualitative Constraint Calculi,” *Proceedings of AAAI*, Vol.8, 2008.
- [4] D. Wolter, “SparQ-A Spatial Reasoning Toolbox,” *AAAI Spring Symposium: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*, p.53, 2009.
- [5] S. Batsakis and E. G. M. Petrakis, “SOWL: a framework for handling spatio-temporal information in OWL 2.0,” *International Workshop on Rules and Rule Markup Languages for the Semantic Web, Springer Berlin Heidelberg*, pp.242-249, 2011.
- [6] N. Mainas and E. G. M. Petrakis, “CHOROS2: Improving the Performance of Qualitative Spatial Reasoning in OWL,” *IEEE 26th International Conference on Tools with Artificial Intelligence*, pp.283-290, 2014.

[7] S. Nam and I. Kim, "A Qualitative Spatial Reasoner Supporting Cross-Consistency checks between Directional and Topological Relations," *Journal of KIISE: Computing Practices and Letters*, Vol.20, No.4, pp.248-252, 2014.

[8] R. Battle and D. Kolas, "Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL," *Semantic Web Journal*, Vol.3, No.4, pp.355-370, 2012.

[9] Y. Gatsoulis, M. Alomari, C. Burbridge, C. Dondrub, P. Duckworth, P. Lightbody, M. Hangeide, N. Hawes, D. C. Hogg, and A. G. Cohn, "QSRLib: A Software Library for Online Acquisition of Qualitative Spatial Relations from Video," *International Workshop on Qualitative Reasoning*, pp.36-41, 2016.

[10] S. J. Lee, et al. "Spatial Knowledge Extraction from Geometric Data," *Proc. of KIISE Winter Conference*, pp.1337-1339, 2014.

[11] Open Street Map [Internet], <https://www.openstreetmap.org/>.

[12] USGS [Internet], <http://www.usgs.gov/>.

[13] OS OpenData [Internet], <http://www.ordnancesurvey.co.uk/>.

[14] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on large Clusters," *Communications of the ACM*, Vol.51, No.1, pp.107-113, 2008.

[15] J. Urbani, S. Kotoulas, J. Maassen, F. Van Harmelen, and H. Bal, "WebPIE: A Web-Scale Parallel Inference Engine using MapReduce," *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol.10, pp.59-75, 2012.

[16] S. Nam and I. Kim, "Design and Implementation of Large-Scale Spatial Reasoner using MapReduce Framework," *Transactions on KIPS: Software and Data Engineering*, Vol.3, No.10, pp.397-406, 2014.

[17] M. Perry and J. Herring, "GeoSPARQL - A geographic query language for RDF data," OGC Implementation Standard, 2012.

[18] M. Egenhofer, "Qualitative spatial-relation reasoning for design," *Studying Visual and Spatial Reasoning for Design Creativity*, pp.153-175, 2015.

[19] M. Egenhofer and J. Herring, "A Mathematical Framework for the Definition of Topological Relationships," in *Fourth International Symposium on Spatial Data Handling*, pp.803-813, 1990.

[20] J. Herring, "OpenGIS Implementation Standard for Geographic information-Simple feature access-Part 1: Common architecture," OGC Document, 2011.

[21] J. F. Allen, "Maintaining Knowledge about Temporal Intervals," *Communications of the ACM*, Vol.26, pp.832-843, 1983.

[22] R. Moratz and R. Marco, "Qualitative Spatial Reasoning about Relative Point Position," *Journal of Visual Languages & Computing*, Vol.19, No.1, pp.75-98, 2008.

[23] S. Leutenegger, M. Lopez, and J. Edgington, "STR: A Simple and Efficient Algorithm for R-tree Packing," in *Data Engineering, Proceedings, 13th International Conference on IEEE*, pp.497-506, 1997.

[24] A. Eldawy and M. F. Mokbel, "SpatialHadoop: A MapReduce Framework for Spatial Data," in *Proceedings of the IEEE International Conference on Data Engineering(ICDE)*, 2015.

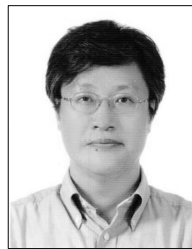
[25] JTS Topology Suit [Internet], <http://www.vividsolutions.com/jts/JTSHome.htm/>.

[26] Openlayers [Internet], <http://openlayers.org/>.



**이 석 준**

e-mail : 20151101162@kyonggi.ac.kr  
 2015년 경기대학교 컴퓨터과학과(학사)  
 2015년~현 재 경기대학교 컴퓨터과학과  
 석사과정  
 관심분야: 인공지능, 공간 표현 및 추론,  
 지능로봇



**김 인 철**

e-mail : kic@kyonggi.ac.kr  
 1985년 서울대학교 수학과(학사)  
 1987년 서울대학교 전산과학과(이학석사)  
 1995년 서울대학교 전산과학과(이학박사)  
 1996년~현 재 경기대학교 컴퓨터과학과  
 교수

관심분야: 인공지능, 지식 표현 및 추론, 지능로봇