

---

# 공간 데이터웨어하우스에서 시공간 분석 지원을 위한 비중복 적재기법

## Non-Duplication Loading Method for supporting Spatio-Temporal Analysis in Spatial Data Warehouse

---

전치수\* / Chi-Soo Jeon, 이동욱\*\* / Dong-Wook Lee, 유병섭\*\*\* / Byeong-Seob You,  
이순조\*\*\*\* / Soon-Jo Lee, 배해영\*\*\*\*\* / Hae-Young Bae

### 요약

본 논문에서는 공간 데이터 웨어하우스에서 시공간 분석지원을 위한 공간 데이터의 비중복 적재 기법을 제안한다. SDW는 이기종의 다양한 서비스를 지원하는 SDBMS로부터 공간 데이터를 추출한다. 제안 기법에서는 SDW에 소스로 참여하는 SDBMS에서 변경된 부분만을 추출하고, 이를 공간연산을 통해 중복된 데이터를 제거한 후 통합된 형태로 적재함으로써 빠른 공간 데이터 분석을 지원할 수 있으며, 저장 공간의 낭비를 줄일 수 있다. 이는 공간 마이닝등의 시간에 따른 분석 및 예측 분야에 효율적인 형태로 공간 데이터를 적재 한다.

### Abstract

In this paper, we have proposed the non-duplication loading method for supporting spatio-temporal analysis in spatial data warehouse. SDW(Spatial Data Warehouse) extracts spatial data from SDBMS that support various service of different machine. In proposed methods, it extracts updated parts of SDBMS that is participated to source in SDW. And it removes the duplicated data by spatial operation, then loads it by integrated forms. By this manner, it can support fast analysis operation for spatial data and reduce a waste of storage space. Proposed method loads spatial data by efficient form at application of analysis and prospect by time like spatial mining.

**주요어 :** 공간 ETL, 공간 데이터 중복, 공간 데이터 웨어하우스

**Keyword :** Spatial ETL, Spatial Data Duplication, Spatial Data Warehouse

---

■ 논문접수 : 2007.7.18

■ 심사완료 : 2007.8.31

\* 인하대학교 정보공학과 석사 과정 (onlytifa@dblab.inha.ac.kr)

\*\* 인하대학교 컴퓨터정보공학과 박사 과정 (dwlee@dblab.inha.ac.kr)

\*\*\* 인하대학교 컴퓨터정보공학과 박사 과정 (bsyou@dblab.inha.ac.kr)

\*\*\*\* 서원대학교 컴퓨터교육학과 조교수 (sjlee@seowon.ac.kr)

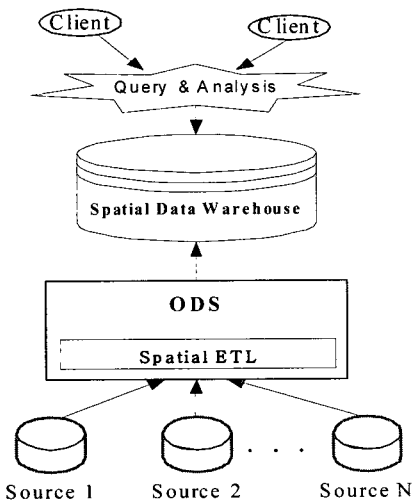
\*\*\*\*\* 교신저자 인하대학교 컴퓨터공학부 교수 (hybae@inha.ac.kr)

### 1. 서론

공간 데이터웨어하우스(Spatial Data Warehouse)는 기존의 데이터웨어하우스 시스템에 주거 지역, 도로 망 그리고 상점 위치와 같은 공간 데이터를 포함한 시스템이다. 이는 비공간 데이터와 공간 데이터를 주제 중심적이고 통합적이며 시간성을 가지는 비 휘발성 자료로 저장하여 분석 및 의사결정을 지원한다[1,2,3,4].

공간 데이터웨어하우스(SDW)는 지리정보 시스템(Geographic Information System), 교통관리 시스템(Traffic Management System) 그리고 물류관리 시스템(Distribution Management System)과 같은 응용의 지원을 위해 다양한 종류의 공간 데이터베이스 시스템(SDBMS)에서 추출된 공간 데이터를 바탕으로 구축된다.

응용 시스템들은 오랜 기간동안 축적되어온 공간 데이터의 분석을 통하여 지난 10년간 주거지역의 변화 그리고 도로망의 변화 등과 같은 공간 데이터의 변화 및 흐름에 대한 정보를 제공할 수 있으며 특히 재해 재난 관리시스템에서는 지난 몇 년간의 수해지역에 대한 정보를 통하여 수해 지역 근처의 제방둑의 건설 및 상하수도관의 확충의 필요성에 대한 정보 역시 분석 할 수 있다.



<그림 1> 공간 데이터웨어하우스 구조

<그림 1>은 공간 데이터웨어하우스의 일반적인 구조를 설명한 것으로 현재에는 다양한 공간데이터 관리시스템(SDBMS)들을 소스 데이터로 사용하여 SDW가 구축된다.

적재 시점별로 공간 데이터를 SDW에 적재하기 위해서는 소스 시스템으로부터 적재 시점별 공간 데이터에 대한 추출을 수행해야 한다. 그러나 기존에는 이러한 공간 데이터를 추출하는 데 있어 이전 적재 시점에 대한 데이터와 현재 적재 될 데이터 간 존재하는 공간 데이터에 대한 중복을 고려하지 않았다.

이러한 적재 주기별 공간 데이터의 중복 적재는 SDW의 적재 연산비용 및 데이터 분석비용의 증가를 가져온다. 또한, 여러 SDBMS로부터 추출된 공간 데이터의 적재로 인해 시스템의 저장 공간의 증가 또한 가져온다.

예를 들어, “서울시 서대문구 상수도관”에 대한 공간 데이터 적재 요구는 이 기존의 SDBMS 가 가지고 있는 같은 주제인 “서울시 서대문구 상수도관”에 대한 데이터의 중복 저장을 이끌고 있고, 이로 인한 데이터 분석시 각 SDBMS 별로 SDW에 같은 종류의 데이터를 저장하므로 저장공간의 낭비를 가져오게 된다.

본 논문에서는 SDW에서의 공간 데이터 분석지원 및 공간 데이터의 중복 저장에 의한 저장 공간의 낭비를 줄일 수 있는 이 기종으로부터 추출된 공간 데이터에 대한 통합 적재기법을 제안한다.

이는 시스템 적재 주기로 공간 데이터를 SDW 시스템에 적재하기 앞서 ODS (Operating Data Store) 내에 저장된 이전 적재 시점 이후부터 현재 적재 시점 전까지 갱신이 발생한 공간 데이터의 정보만을 주제별로 통합하는 데이터 통합 단계를 가진다. 또한 SDW의 시간의 흐름에 따른 공간 데이터 분석 지원을 위한 통합데이터 적재 단계를 지닌다.

본 제안 기법을 통해 SDW 내에 통합 저장된 하나의 공간 데이터는 공간 데이터 웨어하우스에서의 저장 공간에 대한 낭비를 줄일 수 있을 뿐만 아니라 적재 시간을 기준으로 시간 흐름에 따른 공간 데이터 분석 지원 및 공간 마이닝 툴을 이용한 다

른 공간 데이터와의 연관관계 분석을 지원할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 이 기종의 공간 데이터베이스로부터 추출, 변환 및 적재의 과정을 거쳐 생성된 공간 데이터를 SDW에 적재하는 Spatial ETL에 대해 설명하고 공간 데이터 분석에 대한 개념을 간략히 설명한다. 3장에서는 본 논문에서 제안한 공간 데이터 통합 적재 기법을 설명하고, 공간 데이터를 통합함으로써 얻어지는 저장 공간에 대한 효율성을 4장에서 성능 평가로 입증한다. 마지막 5장에서는 결론 및 향후연구에 대해 논한다.

## 2. 관련연구

기존 Spatial ETL에 대한 연구로는 이기종의 소스 시스템에서 추출한 공간 데이터를 단일화된 하나의 좌표값을 가진 데이터로 변환시키는 연구가 진행되었다.

그러나 이것은 각 공간 데이터가 지닌 좌표값의 일치율을 가져올 수는 있으나 이기종의 소스로 추출되기 때문에 동일한 좌표값을 가진 공간 데이터가 SDW 시스템내에 중복저장되는 문제점이 존재한다.

지금까지의 연구중에서는 이러한 공간데이터의 중복저장 문제를 해결하기 위한 데이터의 중복제거 및 통합 적재에 관한 연구가 구체적으로 진행된 바 없다.

이에 본 장에서는 데이터웨어하우스에서 중복제거에 대한 연산을 수행하지 않고 소스 데이터 추출, 변환 후 적재연산을 수행하는 비공간 ETL를 설명한다.

또한 비공간 ETL에 공간 데이터의 속성을 고려한 Spatial ETL에 대해 설명한 후, 적재된 공간 데이터에 대한 SDW에서의 분석에 대해 간략히 설명한다.

### 2.1. 비공간 ETL

비공간 ETL(Extract, Transform, Loading)은 데이터웨어하우스에서 판매량, 온도, 습도와 같은

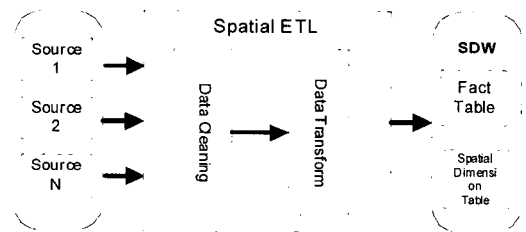
비공간 수치 데이터에 대한 추출, 변환 및 적재 과정을 통칭하여 말한다.

이중 비공간 데이터에 대한 적재는 적재 시점별로 같은 수치데이터라 할지라도 그 의미하는 바가 다르다. 예를 들어, 인천지역의 8월 7일 09시 강수량이 100mm인 것과 8월 7일 15시 강수량이 100mm인 것은 같은 강수량 수치를 가지고 있더라도 시간이라는 키 값이 다르기 때문에 같은 의미를 가진 데이터로 볼 수 없다. 그러므로 이러한 비공간 데이터 정보들은 시스템 적재주기마다 시간에 대한 키 값을 가진 테이블형태로 시스템에 적재된다 [5,6,7].

이처럼 비공간 데이터에 대한 ETL 과정에서는 중복된 수치 데이터에 대한 중복제거 연산을 요구하지 않는다.

### 2.2 Spatial ETL

공간데이터웨어하우스의 구축은 소스로 쓰이는 SDBMS(Spatial DBMS)의 공간 데이터를 추출하는 것에서부터 시작한다. 이러한 공간 데이터의 추출에는 기존의 데이터웨어하우스의 비공간 데이터 추출을 위한 ETL과정에 공간 데이터 속성을 추가한 Spatial ETL 이 사용된다.



<그림 2> Spatial ETL 과정

<그림 2>는 SDW에서의 Spatial ETL 과정을 보여주는 것으로, Spatial ETL은 이 기종 시스템으로부터 얻어지는 다양한 공간 데이터에 대해 데이터 정제 단계 및 변환 단계를 거쳐 생성된 데이터를 SDW에 적재하는 역할을 한다 [8].

다양한 소스 시스템으로부터 Spatial ETL과정

을 통해 추출되어 적재된 공간 데이터는 SDW에 대한 시간별 공간 데이터 분석요구를 지원할 수 있는 구조를 가져야 한다.

이를 위해서는 우선적으로 여러 소스 SDBMS로부터 추출된 공간 데이터에 대한 시점별 그리고 주제별 통합과정을 통해 SDW 시스템에 시점별 단일한 공간 데이터를 생성하는것이 필요하다.

또한, 공간분석 측면 이외에도 추출된 데이터 중 중복되는 공간 데이터 저장에 따른 SDW의 저장공간의 낭비를 해결하기 위해서도 공간 데이터에 대한 통합 과정이 필요하다. 이것은 비공간 ETL과정에서 중복되는 데이터를 테이블형태로 저장하는 것과는 달리 공간 데이터 속성상 동일한 공간 데이터를 적재 시점마다 중복저장할 필요가 없기 때문이다.

### 2.3. Spatial Data Analysis

공간데이터의 분석은 SDW에 축적된 공간 데이터를 사용하여 현재 시점뿐만 아니라 과거 시점에 대한 분석을 지원한다[9].

이러한 분석 지원에 사용되는 SDW내의 공간 데이터는 DW의 4가지 특징인 주제 지향성, 통합성, 비 갱신성 및 시계열성을 갖고 있기 때문에 특정 시점이 아닌 과거부터 현재까지의 공간 데이터 변화 패턴을 분석할 수 있다.

변화패턴에 대한 분석 결과는 다른 공간 데이터와의 연계 분석을 통하여 해당 공간 데이터에게 미치는 영향을 분석 할 수 있다. 이것은 공간 마이닝으로부터 향후 미래 시점에서의 해당 공간 데이터 변화 예측과 같은 추가적인 정보를 얻을 수 있다.

## 3. 이 기종의 공간 데이터 통합 적재기법

본 장에서는 Spatial ETL 과정을 통해 이 기종의 소스 SDBMS로부터 추출된 공간 데이터에 대한 통합 적재기법을 설명한다.

이는 소스 SDBMS 중 업데이트가 발생한 공간 데이터만을 추출하고 적재시점별 그리고 주제별로 통합하는 연산과정을 지닌다.

이러한 연산결과로 생성된 주제별로 통합된 공간 데이터는 SDW의 저장공간의 낭비를 줄일 수 있으며 적재 주기에 맞춰 SDW에 적재되기 때문에 시간의 흐름에 따른 공간 데이터 분석을 지원할 수 있는 구조를 가진다.

### 3.1 이 기종의 공간 데이터 비교

공간 데이터웨어하우스의 ODS(Operating Data Store) 컴포넌트는 SDW시스템 적재 주기에 맞춰 데이터를 적재하기 전 소스 시스템으로부터 추출된 갱신이 발생한 공간 데이터 또는 비공간 데이터에 대한 임시 저장소 역할을 한다.

ODS에서는 일반적으로 이 기종의 공간 데이터 교환을 위한 표준 양식인 GML (Geography Markup Language)을 사용하여 이 기종의 공간 데이터베이스 시스템으로부터 소스로 쓰이는 다양한 공간 데이터를 받아 온다. 그리고 이러한 GML로 받아들인 공간 데이터는 분석을 통하여 관계형 데이터 형태로 변환 할 수 있는데, <그림 3>은 이러한 GML 데이터에 대한 변환 예이다.

```

<?xml version="1.0" encoding="UTF-8"?>
<gml:name> Road Example </gml:name>
<gml:boundedBy>
<gml:Box srsNames="http://www.opengis.net/gml/srs.epsg.xml#4326">
<gml:coord>
<gml:X> </gml:X>
<gml:Y> </gml:Y>
<gml:Z> </gml:Z>
</gml:coord>
</gml:Box>
</gml:boundedBy>
<gml:LineString>
<gml:coord>
<gml:X> 0.000 </gml:X>
<gml:Y> 0.000 </gml:Y>
</gml:coord>
<gml:coord>
<gml:X> 3.000 </gml:X>
<gml:Y> 5.000 </gml:Y>
</gml:coord>
<gml:coord>
<gml:X> 5.000 </gml:X>
<gml:Y> 3.000 </gml:Y>
</gml:coord>
</gml:LineString>
    
```

(가) 소스로부터 추출된 GML 파일

< Example Road 1 >

Type	Data
LineString	0.000,0.000 3.000,5.000 5.000,3.000

(나) GML 파일을 관계형 데이터로 표현

<그림 3> GML 파일을 분석을 통해 얻은 관계형 데이터

GML 에서 관계형 데이터로의 변환 과정 중 이 기종 시스템에서 추출한 같은 주제의 공간 데이터는 서로 다른 공간 타입으로 표현 될 수 있다.

‘한강’ 이라는 이 기종에서 추출해 온 공간 데이터가 있을 시, 소스 시스템 A에서는 해당 데이터를 PolyLine으로 표현할 수 있지만 소스 시스템 B에서는 Line으로 표현할 수 있다. 이와같이 같은 데이터를 표현하는데 있어 이 기종이기 때문에 공간 타입의 불 일치성이 존재한다.

이러한 문제를 해결하기 위해 제안기법에서는 시스템 관리자가 정의한 공간 데이터 타입 정의 테이블을 참조한다. <표 1>은 이러한 공간 데이터 타입 정의에 관한 사전 정의 테이블이다.

<표 1> 공간 데이터 타입 정의 테이블

Spatial Data Type	Description
Point	intersection point
LineString	Streets Canals, Pipelines
Polygon	Voting district, forest, wildlife habitat
MultiPoint	Multi intersection point
MultiLineString	River system
MultiPolygon	Rural country and farms in each country

공간 타입 정의 테이블에 의거하여 만약 ‘도로’에 관한 데이터인 경우 이 기종 내에서 Line 또는

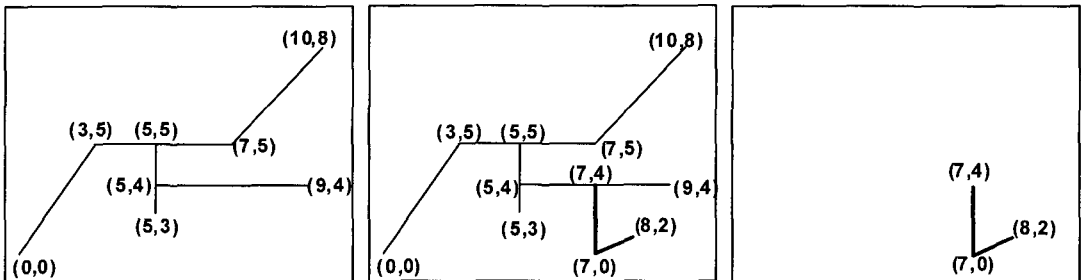
PolyLine형식으로 표현되었던 기존의 공간 데이터에 대해 LineString이라는 일관된 저장 포맷을 지닌 공간 데이터로 추출, 변환 시킨다.

SDW에 대한 저장 공간의 낭비를 막기 위해 소스 시스템으로부터 추출되는 공간 데이터는 ODS에 저장되어 있는 현재 시점에 대한 같은 주제의 공간 데이터인 이전 SDW적재 시점의 결과 데이터와 비교연산을 한 후 그 결과로 나오는 업데이트된 부분만을 추출한다.

<그림 4>는 추출된 공간 데이터의 업데이트 부분을 찾기 위한 이전 공간 데이터와의 비교 연산 과정이다. <그림 4>의 (가)는 SDW에 저장되어 있는 업데이트가 최종적으로 진행된 현재 시점에 대한 공간 데이터로 소스 시스템에서 업데이트가 반영되어 추출된 (나) 와의 Difference 공간연산을 통하여 (다)와 같은 업데이트가 발생한 공간 데이터만을 추출 하게된다.

한편, 비교연산시 사용하는 소스 시스템내의 공간 데이터 추출은 주제별 공간 데이터의 속성에 따라 ODS로의 추출 주기가 다르다. 즉, 도로 또는 강과 같은 공간 데이터는 빈번하게 업데이트가 발생하는 데이터가 아니므로 업데이트에 따른 추출 주기가 느리다. 하지만 태풍의 영향권 또는 이동체의 이동을 저장하기 위한 공간 데이터는 빈번하게 업데이트가 발생하므로 이를 추출하기 위한 시스템 비용의 소모가 크다.

이처럼 빈번하게 공간 데이터의 갱신이 일어나는



(가) 업데이트 되기이전의 도로네트워크 데이터 (나) 업데이트가 발생된 도로네트워크 데이터 (다) 업데이트 부분에 대한 데이터의 추출

<그림 4> 도로 네트워크 데이터에 대한 이전 데이터와의 비교 추출

< Source System >

Source 1	Select * form 한강; Insert into 한강
Source 2	Select * from 한강; Delete from 양계대로 where x> 100;
⋮	⋮
Source N	Update 압록강 set z=100; Insert into 섬진강;

(가) Source 시스템별 입력된 질의

< ODS >

Source 1	Insert into 한강
Source 2	Delete from 양계대로 where x> 100;
⋮	⋮
Source N	Update 압록강 set z=100; Insert into 섬진강;

(나) ODS에 적재된 업데이트 질의

<그림 5> 소스 시스템에 대한 업데이트 질의 추출

공간 데이터의 경우 업데이트가 발생 할때마다 데이터를 추출하는 것은 비 효율적이다. 또한, 이렇게 추출해온 데이터에 대해 이전 적재 시점의 데이터와 비교를 매번 수행해야하므로 이것은 많은 시스템 비용을 소모한다.

그러므로 제안기법에서는 공간 데이터 추출과정시 각 소스 시스템에 대한 Insert, Delete, Update 질의와 같이 공간 데이터에 대한 갱신을 발생시킨 질의들에 대해서만 트리거를 이용하여 따로 관리한다.

즉, 소스 시스템로부터 공간 데이터 추출 시 각 소스 시스템별로 수행된 질의 중 업데이트 질의만을 트리거를 통해 받아와 이 질의가 나타내는 공간 데이터를 추출한다. 이것은 추출 주기마다 모든 소스 시스템의 데이터를 추출하지 않아도 되므로 그만큼 시스템 비용을 줄일 수 있는 장점이 있다.

<그림 5>는 소스 시스템으로부터의 공간 데이터 추출을 하기위해 ODS의 임시 버퍼에 저장된 업데이트 질의를 보여준다. (가)는 각 소스 시스템에서 발생한 사용자 질의이고, 이 중 Select 질의와 같은 검색질의를 제외한 업데이트 관련 질의에 대해서만 트리거를 통하여 ODS내의 임시 버퍼 공간에 저장한다.

소스 시스템으로부터 전달 받은 업데이트관련 질의들은 추출을 위한 Select 질의로 변화 되어 해당 소스 시스템에 전달한다.

이 후 각 소스 시스템은 Select 질의 수행 결과로 얻은 데이터 셋을 표준 형식인 GML로 ODS에 전달함으로써 갱신이 발생한 공간 데이터에 대한 추출을 지원한다.

### 3.2 공간 데이터웨어하우스로의 통합적재

본 절에서는 이전 데이터와의 비교 연산을 하여 추출된 업데이트 셋을 SDW시스템의 적재 주기에 맞춰 하나의 데이터 셋으로 통합하는 과정 및 공간 데이터 분석 지원을 위한 통합 데이터 적재 방법에 대해 설명한다.

#### 3.2.1 업데이트 정보들의 대한 주제별 통합

이 기존의 공간 데이터 비교 단계를 거쳐 ODS에 저장된 업데이트 정보들은 소스 데이터 추출 시점에 따라 관리되며, 이렇게 모아진 업데이트 정보들은 이전 SDW적재 시점부터 현재 적재시점 전까지 업데이트가 발생한 공간 데이터들의 정보를 가지고 있다.

<그림 6>은 이러한 비교 연산에 의해 추출되어 ODS에 저장된 업데이트 정보에 대한 주제별 그리고 추출 시점별로 관리되는 모습을 보여준다.

ODS에 저장되어 있는 소스 시스템별 업데이트 정보들은 SDW 적재시기에 맞춰 주제별 하나의 업데이트 정보로 통합된다.

<그림 6>의 영동 대로의 도로 네트워크에 대한 업데이트 정보 테이블을 보면 Source 1 시스템의 추출 시점 T1과 Tn에서 업데이트 정보가 발생한 것을 볼 수 있다. 이러한 T1과 Tn에 대한 업데이트 정보는 공간 연산자인 Union을 통하여 하나의 업데이트 정보로 통합된다.

<그림 7> 은 이러한 소스 시스템별 업데이트 정

< 합강 > - MultiLineString

추출시점	Source 1	Source 2	Source 3
T1	-	-	5,1 12,7
T2	5,1 12,7 7,6	-	5,2 8,5
⋮		⋮	⋮
Tn	-	7,1 12,5	-

< 성진강 > - MultiLineString

추출시점	Source 1	Source 2	Source 3
T1	2,2 2,5 4,4 5,6 5,2	-	-
T2	-	-	2,2 2,5 4,4 5,6 5,2
⋮	⋮	⋮	⋮
Tn	-	2,9 3,11 5,9 3,6	6,4 5,7 6,9 7,5

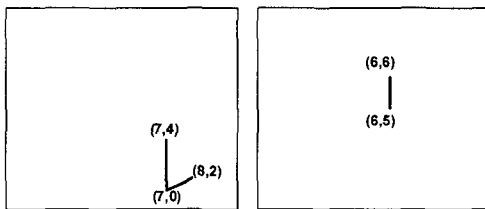
< 영동 대로 > - Linestring

추출시점	Source 1	Source 2	Source 3
T1	7,4 7,0	-	-
T2	-	6,5 6,6	7,4 7,0
⋮	⋮	⋮	⋮
Tn	7,0 8,2	-	7,3 6,3

<그림 6> ODS에 저장된 주제별, 시간별 업데이트 정보

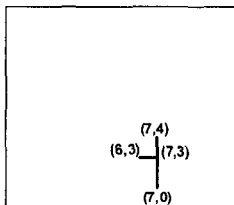
보의 통합된 결과를 보여준다.

소스 시스템별 통합된 업데이트 데이터는 SDW 적재를 위해 이 기종 시스템들 간의 통합을 필요로 한다. 기본적으로 MBR (Minimum Boundary Rectangle)로 표현되는 공간 데이터는 다른 공간 데이터와 서로 중복되는 MBR 영역에 대한 제거연산을 통하여 하나의 공간 데이터로 통합된다.



(가) Source 1의 업데이트 정보

(나) Source 2의 업데이트 정보



(다) Source 3의 업데이트 정보

<그림 7> ODS내의 시스템별로 통합된 업데이트 데이터 셋

아래의 [알고리즘 1]은 여러 공간 데이터가 중복 되었을 때, 중복된 부분을 제거한 후 한 개의 통합된 공간 데이터를 생성하기 위한 알고리즘이다.

[알고리즘 1] Spatial\_Integration 알고리즘

Input

Geom 1, Geom 2, ... Geom N : 비교되어지는 공간 데이터

Variable

OverlapRec : 두 영역사이의 중복되는 공간영역

Algorithm Spatial\_Integration(Geom1, Geom2)

Begin

```

01 : for i:= 1 to i < Geom.count-1 step 1
02 : if( Disjoint (Geom [i], Geom [i+1]) )
03 :   Geom[i+1] := Union (Geom [i], Geom [i+1])
04 : else
05 :   OverlapRec := GetOverlap (Geom [i], Geom [i+1])
06 :   Geom[i] := Difference (Geom [i], OverlapRec)
07 :   Geom[i+1] := Union (Geom [i], Geom[i+1])
08 : end if
09 : end for
10 : retrun Geom[i+1]

```

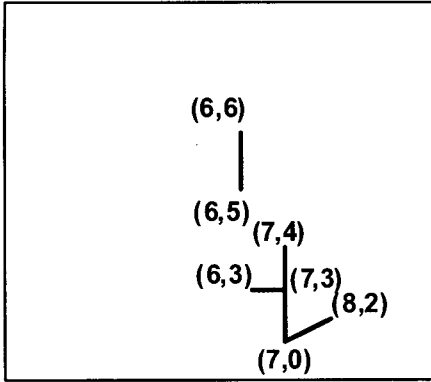
End

[알고리즘 1]의 라인 2~3은 통합되는 두 공간 데이터에 대해 Disjoint 하는지 검사하고 중복된 부분이 없으면 통합 연산을 수행한다. 라인 5~8에서는 중복되는 영역에 대한 공간 데이터를 추출하고 비교 연산되는 공간 데이터에 대해 Difference 연산을 수행한다. 이 후 다른 공간 데이터와 통합함으로써 SDW 적재를 위한 하나의 통합된 업데이트 셋을 생성해 낼 수 있다. 이러한 통합 과정은 소스 시스템별로 추출된 업데이트 데이터의 개수만큼 반복한다.

<그림 8>은 Spatial\_Integration 알고리즘에 의해 통합된 영동 대로 도로 네트워크에 대한 업데이트 데이터 셋으로 서로 다른 소스 시스템으로부터 받아온 데이터의 중복된 부분을 제거한 후 통합한 결과이다.

주제별 통합된 업데이트 셋은 SDW에 적재됨으로써 이 기종간에 있을 수 있는 데이터 중복저장에

의한 저장공간의 낭비를 줄일 수 있다.



<그림 8> 이 기종 시스템으로부터 통합된 업데이트 데이터 셋

### 3.2.2 SDW의 분석지원을 위한 적재기법

소스 시스템으로부터 추출 후 SDW에 적재된 공간 데이터는 시계열성 분석을 지원하기 위해 적재 시점 및 주제별로 통합된 데이터 형태를 가져야한다.

<표 2>는 이러한 주제별 공간 데이터의 통합 요구에 대해 <그림 8>에서 얻은 영동 대로 도로 네트워크에 대한 통합 업데이트 데이터 셋이다.

<표 2> 통합 업데이트 데이터

Subject	통합 업데이트 데이터
영동 대로	6,5 6,6 7,0 7,4 8,2 6,3 7,3

<표 2>에서와 같이 이 기종 소스 시스템으로부터 통합된 업데이트 데이터는 SDW 적재에 앞서 주제별 업데이트 정보 테이블에 저장된다.

<그림 9>는 레이어 속성별 갱신된 공간 데이터의 관리를 위한 업데이트 정보 테이블로 이전 적재 시점 이후부터 현재 적재 시점까지 레이어 속성별로 갱신이 일어난 모든 공간 데이터의 업데이트 정보가 레코드 형태로 저장되어 있다.

이러한 업데이트 정보 테이블은 속성 레이어별 객체 아이디인 OID와 OID가 가리키는 객체명 그

리고 이에 해당하는 통합된 업데이트 데이터 값 등 세 가지 필드로 나누어져 저장된다.

#### <강> - MultiLineString

OID	Subject	데이터 값
001	한강	419293.8145,524091.2452 19142.1628,654867.9885
002	압록강	328882.7140,653817.6995 328815.6889,653605.0299
⋮	⋮	⋮
N	섬진강	329162.0340,655366.3166 329185.8037,655184.3520

#### <도로> - LineString

OID	Subject	데이터 값
001	강남대로	168422.2451,21874.4652
002	영동대로	6,5 6,6 7,0 7,4 8,2 6,3 7,3
⋮	⋮	⋮
N	양재대로	329193.8983,655093.3545

<그림 9> 강, 도로 레이어 주제별 통합된 업데이트 정보

‘강’이라는 레이어안에 속한 한강, 압록강 등과 같이 갱신이 발생하여 업데이트 테이블 내에 저장된 공간 객체들은 이에 해당하는 OID를 이용하여 주제별 레이어 안의 어느 공간 객체에 갱신이 일어났는지 쉽게 알 수 있게 해준다.

또한 현재 시점의 갱신된 데이터만을 SDW에 적재하고, 이전 시점의 공간 데이터에 반영함으로써 현재 시점에 대한 공간 데이터를 생성할 수 있게 해준다.

## 4. 성능 평가

본 장에서는 제안된 이 기종의 공간 데이터 통합 적재기법에 대한 성능을 평가한다.

성능 평가에 사용된 실험환경은 <표 3>과 같다.

<표 3> 실험 환경

Computer	IBM PC 호환
CPU	Pentium(R)4 3.0GHz
Memory	2.00GB
HDD	120GB, 7200RPM, S-ATA
OS	Windows XP Professional
Development Environment	MicroSoft Visual C++ 6.0 (C/C++)

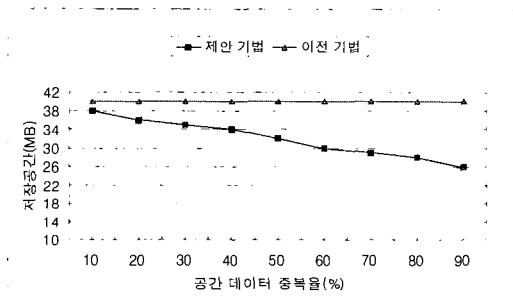


평가 방법으로는 공간 데이터 웨어하우스로 데이터 적재시 적재 시점별 추출된 데이터에 대한 일괄적인 적재방법과 업데이트 된 부분만을 추출하고 통합하는 본 제안 기법과의 비교를 수행하였다.

이중, 제안되는 기법과 비교되는 일괄적 적재방법은 지금까지 공간 데이터를 SDW로 어떻게 적재할 것인지에 대한 구체적 연구 결과가 없기 때문에 이전의 비공간 데이터에 대한 ETL 과정과 같이 추출후 적재라는 단순 적재방법을 선택하였다.

위 실험을 위한 소스 공간 데이터로는 공간 데이터 베이스 분야에서 많이 사용되는 TiGER/Line 파일을 사용하였으며 공간 객체에 대한 중복을 및 적재횟수를 단계적으로 높임으로써 이전 기법과의 비교를 진행하였다[10]. 이를 위해 소스 공간 데이터로서 사용된 데이터는 20,000개의 공간 레코드를 가진 20MB 크기의 임의로 생성된 데이터이다.

<그림 10>은 이전 적재기법과 제안 기법과의 중복율에 따른 저장공간의 비교이고 <그림 11>은 중복저장 횟수에 의한 데이터 적재 시간에 대한 비교이다.

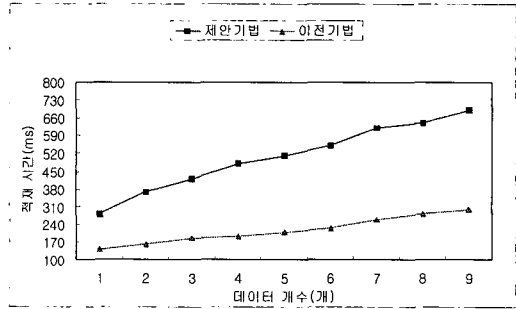


<그림 10> 중복율에 따른 저장공간 비교

<그림 10>에서 보는 것과 같이 이전 기법이 소스로부터 추출된 공간데이터에 대한 일괄적인 적재를 하는것에 비해 제안 기법에서는 비교되는 두 공간 데이터간 중복되는 공간 객체를 제거 및 통합적재함으로써 저장공간의 효율성이 중복율이 높아짐에 따라 이전 기법에 비해 우수하다는 것을 알 수 있다.

그러나 본 제안기법은 데이터 추출시 업데이트된

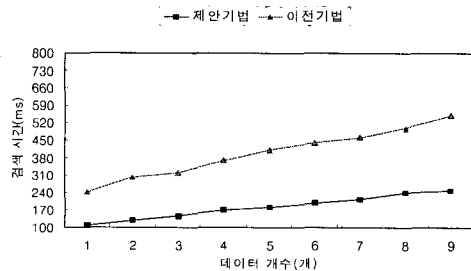
부분만을 찾기 위한 검색 연산비용 및 데이터 적재시 추출된 공간 데이터간 중복제거 연산비용을 요구한다는 측면에서 <그림 11>에서 보는 것과 같이 이전 기법에 비해 데이터 적재시간이 오래 걸린다는 단점이 존재한다.



<그림 11> 데이터 증가에 따른 적재시간 비교

다음은 공간 데이터 분석에 대한 실험평가이다. 이전 기법에서는 적재 시점별 데이터 추출 후 적재라는 일괄적인 데이터 적재방법을 수행하기 때문에 시점별 통합정보를 가지고 있는 데이터 생성하기 위해서는 추출된 데이터에 대한 통합연산 비용이 요구된다.

이것은 <그림 11>의 실험평가 결과와 비교를 통해 제안기법에서의 공간 데이터 적재 비용은 이전의 일괄 적재기법보다는 크지만 <그림 12>에서와 같이 공간 데이터 분석을 위한 데이터 검색비용 측면에서는 이전 기법보다 우수하다는 것을 알 수 있다.



<그림 12> 데이터 분석을 위한 검색시간 비교

### 5. 결론 및 향후 연구

본 논문에서는 공간 데이터 웨어하우스의 분석을 지원하기 위한 이 기종간의 통합된 공간 데이터 적재 기법을 제안하였다.

이 기법은 ODS에 임시 저장된 업데이트 데이터만을 적재 시점마다 SDW 시스템에 적재함으로써 시계열성에 따른 공간 데이터 분석을 지원을 할 수 있는 구조를 가지고 있다.

한 가지 주제별 공간 데이터에 대한 분석은 이전에 분석된 공간 데이터와 다른 공간 데이터와의 연계 분석을 통해 데이터간 서로 미치는 영향을 분석할 수 있는데, 이것은 공간 마이닝을 통해 미래 시점에 대한 공간 데이터분석 정보 또한 얻을 수 있다.

예를 들어 "2002년부터 현재까지의 서울시 중구 지역 도로망"에 대한 분석 요청이 있다고 가정하면 SDW 초기 구축 시 사용된 "중구지역 도로망"에 대한 이력 데이터에 도로망에 대한 업데이트 정보 테이블 안에 있는 중구 지역에 대한 적재 시간별 업데이트 정보를 반영함으로써 2002년에 해당하는 "서울시 중구지역 도로망"에 대한 데이터를 생성해 낼 수 있다. 이후 매 SDW적재 시점마다 ODS로부터 얻어온 업데이트 정보를 반영함으로써 시간의 흐름에 따른 해당 공간 데이터의 변화를 알 수 있게 해 준다.

또한 이와 같은 방식으로 "서울시 중구 상가지역"에 대한 같은 시간대 별 변화 흐름 역시 알 수 있는데 사용자는 이러한 도로망과 상가지역에 대한 변화 추이를 같이 분석 함으로써 도로망이 신설되고 난 후 그 주변 상가지역의 범위가 더 넓어졌다는 공간 마이닝을 통한 추가적인 분석정보를 얻을 수 있다.

제안하는 통합 적재 기법은 이전 시점 데이터에 대한 갱신된 공간 정보만을 SDW에 적재하기 때문에 SDW에서의 저장 공간의 낭비를 막을 수 있었다.

성능 평가에서는 이를 증명하기 위한 실험을 통해 제안 기법이 이전 적재 기법에 비해 저장공간 효율성 측면에서 약 11.4%의 우수성을 보여준다.

이것은 이전 기법이 적재 시점별 공간 데이터를 하나씩 저장하는 것에 비해 제안 기법에서는 업데이트 된 부분만을 저장하는데서 생기는 저장 공간의 차이이다.

향후 연구로는 ODS내에서 통합된 업데이트 데이터를 만드는데 필요한 공간 연산을 줄이는 방법에 대한 연구가 필요하다.

### 참고 문헌

1. S. Chauduri, U. Dayal "An Overview of Data Warehousing and OLAP Technology," ACM SIGMOD Record, Vol.26, 1997, pp. 65-74
2. ESRI, "Spatial Data Standards and GIS Interoperability," An ESRI White Paper, 2003.  
<http://www.esri.com/library/whitepapers/pdfs/spatial-data-standards.pdf>
3. ESRI, "Spatial Data Warehousing for Hospital Organizations," An ESRI White Paper, 1998.  
<http://www.esri.com/library/whitepapers/pdfs/sdwho.pdf>
4. ESRI, "Spatial Data Warehousing", An ESRI White Paper, 1998., <http://www.geoweb.dnv.org/Education/whitepapers/SpatialWarehousing.pdf>
5. Alkis Simisis, Panos Vassiliadis, Timos Sellis, "Optimizing ETL Process in Data Warehouse", ICDE, 2005, pp. 564-575
6. Oracle, "Integrated ETL and Modeling," An Oracle White Paper, 2003.  
[http://www.oracle.com/techno-logy/products/warehouse/pdf/OWB\\_WhitePaper.pdf](http://www.oracle.com/techno-logy/products/warehouse/pdf/OWB_WhitePaper.pdf)
7. Cass Squire, "Data extraction and transformation for the data warehouse" A Proceedings of the ACM SIGMOD international conference on Management

of data, 1995, pp. 446-447

8. Xi-qian CHEN, Zhong-xian CHI, Xiu-kun CAO, "Applying DP to ETL of Spatial Data Warehouse", 3rd International Conference on Machine Learning and Cybernetics, Vol.3, 2004, pp. 1616-1619
9. Konstantin krivoruchko, Carol A. Gotway, Alex Zhigimont, "Statistical tools for regional data analysis using GIS", A Proceedings of the ACM international symposium on Advance in GIS, 2003, pp. 41-48,
10. TIGER/Line Files, 2000 Technical Documentation, U.S. Bureau of Census, California, accessible via, [http://arcdata.esri.com/data/tiger2000/tiger\\_statelayer.cfm?s\\_tips=06](http://arcdata.esri.com/data/tiger2000/tiger_statelayer.cfm?s_tips=06)

**전치수**

2006년 단국대학교 컴퓨터학과(공학사)  
 2008년 인하대학교 정보공학과(석사과정)  
 관심분야 : 공간 데이터베이스, 공간 데이터웨어하우스, 데이터 마이닝, U-GIS

**이동욱**

2003년 상지대학교 전자계산공학과(이학사)  
 2005년 인하대학교 컴퓨터 정보공학과(공학석사)  
 2005년~현재 인하대학교 컴퓨터정보공학과(박사과정)  
 관심분야 : Spatial Database Warehouse, Ubiquitous  
 . 환경을 위한 SDBMS

**유병섭**

2002년 인하대학교 컴퓨터공학부(공학사)  
 2004년 인하대학교 컴퓨터 정보공학과(공학석사)  
 2004~현재 인하대학교 대학원 컴퓨터 정보공학과  
 (박사과정)  
 관심분야 : 공간 데이터웨어하우스, XML, GML, 클러스터 시스템

**이순조**

1985년 인하대학교 전자계산학과(이학사)  
 1987년 인하대학교 전자계산학과(이학석사)  
 1995년 인하대학교 전자계산학과(공학박사)  
 1997년~현재 서원대학교 컴퓨터정보통신공학부  
 부교수  
 관심분야 : 데이터베이스, 실시간 데이터베이스 시스템, GIS, 데이터베이스 시스템의 보안

**배해영**

1974년 인하대학교 응용물리학과(공학사)  
 1978년 연세대학교 대학원 전자계산학과(공학석사)  
 1987년 숭실대학교 대학원 전자계산학과(공학박사)  
 1985년 Univ. of Houston 객원교수  
 1992년~1994년 인하대학교 전자계산소 소장  
 1982년~현재 인하대학교 컴퓨터공학부 교수  
 1999년~현재 지능형 GIS연구센터 센터장  
 2000년~현재 중국 중경우전대학교 대학원 명예교수  
 2004년~2006년 인하대학교 정보통신대학원 원장  
 2006년~현재 인하대학교 대학원장  
 관심분야 : 분산 데이터베이스, 공간 데이터베이스, 지리정보 시스템, 멀티미디어 데이터베이스 등