

다중축척 데이터베이스를 위한 도구

김정자^{*0}, 강혜경^{**}, 이기준^{*}

^{*}부산대학교 전자계산학과, ^{**}지형정보협동과정
{jkim, hkkang}@isel.cs.pusan.ac.kr, lik@pnu.edu

A Tool for Derivation and Management of Multi-Scale Databases

Jung-Ja Kim^{*0} Hae Kyong Kang^{**} and Ki-Joune Li^{*}.

^{*}Dept. of Computer Science, ^{**} Dept of Geographic Information System, Pusan National Univ.

요 약

공간데이터를 변형/왜곡하여 자세한 정도를 감소시킴으로서 공간데이터를 단순화하는 기법을 일반화라고 하는데, 이미 존재하는 데이터베이스로부터 새로운 축척의 데이터베이스를 유도할 때 이 기법을 활용할 수 있다. 본 연구에서는 다중축척 데이터베이스를 생성/관리하기 위한 일반화 시스템을 소개할 것이다. 이 시스템은 기존 연구들에 의해 제시된 일반화 연산자, 일관성 유지 방법들을 통합/개선하여 하나의 시스템으로 구현되었다. 뿐만 아니라, 이 시스템에는 원시 데이터가 갱신되었을 때 이것을 유도된 다중축척 데이터베이스에 파악시킬 수 있는 방법과 사용자를 위한 모델 다이어그램 지원기능이 포함되어 있다. 본 연구에서 제시하는 다중축척 데이터베이스를 위한 도구는 다중축척 데이터를 유도(생성) 및 관리하기 위한 전체적인 틀을 제공하는 점에서 의의가 있다.

1. 서론

공간데이터의 일반화(generalization)란 데이터의 자세한 정도를 감소시킴으로서 새로운 축척의 데이터를 생성하는 기법이다. 일반화의 목적은 원시 데이터를 사용자가 요구하는 축척에 적합하도록 변형시키는 것으로 이 분야의 연구들은 데이터를 변형시키기 위한 여러가지 일반화 연산자를 제시하는데 중점을 두어왔다[1].

일반화 기법은 원시 데이터베이스로부터 새로운 소축척 데이터베이스를 유도하는데도 활용될 수 있다. 이 때 새 데이터베이스는 원시 데이터에 일반화 연산자를 적용시킴으로써 변형된 데이터를 저장하여 생성된다. 그런데 일반화 연산자에 의존한 일반화는 몇 가지 문제점을 발생시킨다. 첫째 원시 데이터베이스의 지형지물 간에 존재하던 위상 관계가 일반화된 데이터베이스에서 다른 형태로 변화할 수 있다. 둘째, 현재 일반화 연산자나 기존 연구들은 원시 데이터베이스가 갱신되었을 때 이를 유도된 데이터베이스에 전달하는 방법을 제공할 수 없다. 따라서 본 연구에서는 이와 같은 문제점을 해결하기 위한 기능들을 추가하고 기존 연구들이 제시한 일반화 방법을 하나의 시스템으로 통합한 다중축척 데이터관리시스템을 제안하고자 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 다중축척 데이터관리시스템에 사용될 용어 및 공간 데이터모델, 일반화 연산자들을 정의하고 3장에서는 다중축척 데이터관리시스템의 구조와 클래스 다이어그램을 설명하겠다. 그리고 4장에서 본 연구에서 구현한 시스템의 예를 보여주고 5장에서 결론을 맺는다.

2. 공간 데이터 모델 정의

이 장에서는 기존 연구 중에서 본 연구에서 사용할 일반화 연산자와 공간데이터 모델을 기술한다. 본 연구는 새로운 공간 데이터 모델이나 일반화 연산자를 제시하고자 하는 것은 아니다.

표1. 모델 구성요소 표기법

기호	의미	기호	의미
D	공간데이터베이스	$\forall[A_i]$	비공간축척의 값
C	지형지물(feature)	$dom(A_i)$	비공간축척의 도메인
A_i^j	비공간축척	$dom(Geo)$	공간축척의 도메인
Geo	공간축척	-	-

본 연구에서 공간(space)은 점 집합으로 구성된 2차원 벡터로 제한하며 공간데이터베이스 $D = \{C_1, C_2, \Lambda, C_n\}$, 지형지물 클래스 $C = \{A_i^1, A_i^2, \Lambda, A_i^k, Geo\}$ 로 정의한다. 또한 공간축척(Geo)의 도메인 $dom(Geo)$ 는 $\{0 \vee 1 \vee 2\}$ 이며 지형지물의 위상관계(topological relationships)는 [2]에서 정의한 8가지-포함, 겹침, 동일, 내부, 만남, 덮음(covers), 덮힘(coveredBy)-로 표현한다.

본 연구에서 사용할 일반화 연산자는 [3,4]를 바탕으로 다음과 같이 정의한다.

표2. 일반화 연산자

선택 (preselection)	유도 데이터베이스의 목적에 적합한 지형 지물을 선택한다. $D_{preselection} \leftarrow D_{subset} \subseteq D$
----------------------	---

삭제 (elimination)	사용자에 의해 주어진 조건을 만족하는 지형지물을 삭제한다. $C_{elimination} \leftarrow \sigma_{v[A_i]=predicate} C_i$
압축 (collapse)	공간속성의 차원만 축소시킨다. $C_{collapse} \leftarrow C_i$ 단, $dom(C_{collapse}.Geo) = \begin{cases} 0 \vee 1 & (C_i.Geo = 2) \\ 0 & (C_i.Geo = 1) \end{cases}$
집합 (aggregation)	전체 지형지물 중 일정한 거리내에 있는 지형지물의 공간속성을 통합한다. $C_{aggregation} \leftarrow F_{agg}(C_i.Geo)$ 단, $dist(C_i.Geo^l, C_i.Geo^m) \leq predicate$ $F_{agg}() = \prod_{i=0}^n C_k.Geo^i \quad \vee \quad \prod_{j=0, j \neq i}^n \Delta C_k.Geo^j$
단순화 (simplification)	지형지물의 공간속성의 차원은 유지하면서 해상도를 감소시킨다.
분류 (classification)	비공간속성이 동일한 지형지물의 공간속성을 통합한다. $C_{classification} \leftarrow A_i F_{merge}(Geo C_i dom(A_i))$ 단, $F_{merge}() = \prod_{i=0}^n C_k.Geo^i$

3. 다중축척 데이터베이스를 위한 시스템 구조

본 연구의 목적은 독립된 단일 시스템에서 GIS의 일반화 기법을 이용하여 다중축척 데이터베이스를 유도 및 관리하는 도구를 제시하는 것이다. 이 장에서는 다중축척 데이터 관리시스템의 시스템 요구 사항, 본 연구에서 제안하는 시스템 구조 및 각 세부모듈을 기술한다.

원시 데이터베이스로부터 새로운 데이터베이스를 유도하고, 유도된 데이터베이스를 관리하기 위해서는 다음과 같은 사항들을 고려할 필요가 있다.

1. 원시데이터로부터 새로운 데이터베이스의 유도
2. 원시 데이터베이스와 유도 데이터베이스의 연결
3. 유도된 데이터의 일관성과 정확성 평가
4. 원시 데이터의 갱신을 유도된 데이터에 반영
5. 다중축척 데이터의 유도/관리하는 과정을 사용자에게 보여줄 인터페이스

본 연구에서 제안하는 다중축척 데이터베이스를 위한 시스템의 전체 구조는 그림 1과 같다.

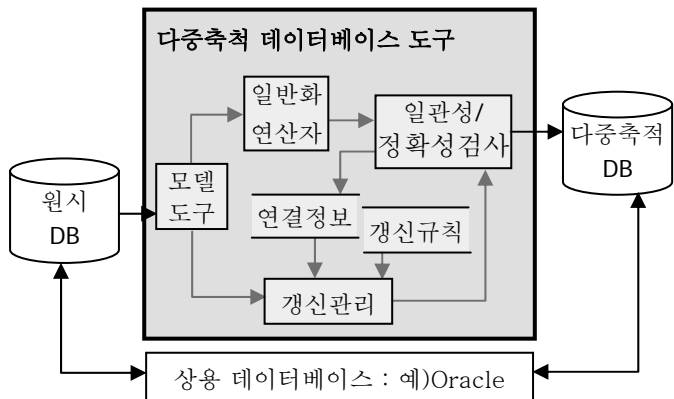


그림 1. 다중축척 데이터베이스를 위한 시스템 구조

그림 1에서 원시 데이터베이스와 유도된 다중축척 데이터베이스는 공간데이터를 지원하는 상용 데이터베이스 관리시스템과 연동됨을 보여주고 있다. 각 구성요소들의 특징은 다음과 같다.

- **모델 도구** : 사용자가 다중축척 데이터베이스의 유도 과정을 쉽게 인식할 수 있도록 인터페이스를 지원하고, 데이터와 연산자를 모델 다이어그램으로 표현한다.
- **일반화 연산자** : 원시데이터로부터 새로운 데이터베이스를 생성하기 위한 연산자들을 지원하며 원시 데이터를 변형시켜 새로운 데이터를 생성하기 위해 2장에서 정의한 여섯가지 일반화 연산자들을 제공한다.
- **연결정보** : 원시데이터, 유도된 다중축척 데이터, 사용된 연산자들을 기술한 파일이다. 연결정보는 원시 데이터베이스와 유도 데이터베이스의 연결고리(link)역할을 하여 원시 데이터베이스에 갱신이 일어났을 때 유도 데이터베이스도 함께 갱신될 수 있도록 한다. 또 다중축척에서 질의를 처리할때도 필요하다.
- **갱신규칙** : 원시 데이터가 갱신되었을 때, 원시 데이터로부터 유도된 다중축척 데이터를 갱신하기 위한 방법들을 제시한다. 본 연구의 선행연구에서 이 규칙을 제시하기 위하여 이미 작업을 하였으며 [5], 본 논문에서 세부사항을 기술하지는 않는다.
- **일관성 및 정확성 검사 연산자** : 본 연구에서는 일반화의 일관성을 위해서 원시데이터에 정의된 위상관계들이 유도된 데이터에서도 유지되는지를 검사하는 관계검사 연산자를 구현한다. 또, 일반화 정확성을 위해서 원시데이터가 갱신되었을 때 이것을 유도된 데이터에 파급시킬 수 있는 갱신파급(update propagation) 연산자를 구현한다.

4. 다중축척 데이터베이스를 위한 시스템 구현

본 연구는 ArcGIS 8.0.2에서 지원하는 ArcObjects를 이용하여, 윈도우 2000에서 VC++로 시스템을 구현하였다. 이 시스템이 지원하는 기능은 크게 네가지로 나눌 수 있는데, 첫째 다중 축척의 데이터를 유도하기 위한 일반화 연산자의 제공, 둘째 유도된 데이터와 원시 데이터간의 관계 일관성 검사, 셋째 원시 데이터의 갱신을 유도된 데이터에 반영, 넷째 일반화 과정을 사용자에게 다이어그램으로 보여주는 것이다. 그림 2는 앞에서 제안한 시스템 구조(그림 1)를 구현한 패키지 다이어그램을 보여준다.

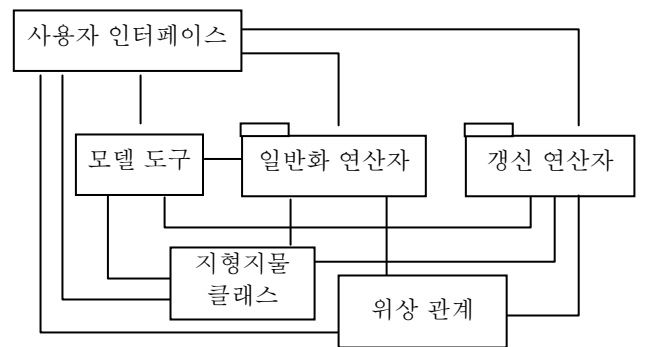


그림 2. 패키지 다이어그램

● **사용자 인터페이스** : 사용자는 다이어그램을 이용해서 데이터를 불러오거나 연산자를 사용할 수도 있고, 메뉴에서 직접 데이터와 연산자에 접근할 수 있다. 그림 3은 전체 메뉴를 보여준다.

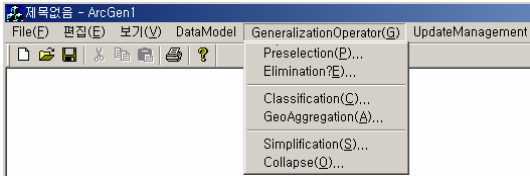


그림 3. 다중축척 데이터베이스 관리시스템을 위한 메뉴

● **지형지물 클래스** : ArcGIS에서 데이터를 처리하는 기본단위 중 하나이다. 본 연구는 ArcGIS를 기반으로 구현하였으므로 지형지물 클래스를 입력데이터 포맷으로 한다. 지형지물 클래스의 접근 및 조작은 ArcGIS-ArcObjects의 GeoDatabase라이브러리를 사용하였다. 이를 기반으로 한 이유는 ArcObjects에서 지원하는 공간데이터의 조작에 필요한 연산자들이 오라클이나 다른 데이터베이스보다 다양하고 안정적이기 때문이다. 또, 이 연산자들을 본 연구에서 다시 구현하는 것은 본 연구의 범위 밖의 과제이기 때문이다.

● **모델 도구 클래스** : 데이터 셋인 지형지물 클래스와 일반화 및 갱신연산자간의 관계를 다이어그램으로 그리고 연산자를 실행할 수 있는 기능을 제공한다.

● **일반화 연산자 패키지** : 여섯가지 일반화 연산자들을 구현한 클래스를 포함하고 있다. Preselection 클래스, Elimination클래스 등이 포함된다.

그림 4는 Aggregation 클래스를 이용해서 집합 연산자를 구현한 예를 보여준다. 그림 4에서 가장 왼쪽은 ArcGIS의 shape파일 포맷으로 저장된 원시 데이터(EDUSRC.shp)를 보여준다. 원시 데이터, Aggregation 연산자 그리고 유도될 데이터 이름을 입력하면 그림 4의 중앙과 같은 다이어그램이 보여진다. 이 다이어그램 창에서 연산자(○)를 클릭하면 Aggregation을 위한 매개변수를 입력할 수 있는 대화상자가 나타나고 매개변수 값을 입력하면 새로운 데이터(EDUSRC'.shp, 그림 4의 오른쪽 아래 그림)가 생성된다.

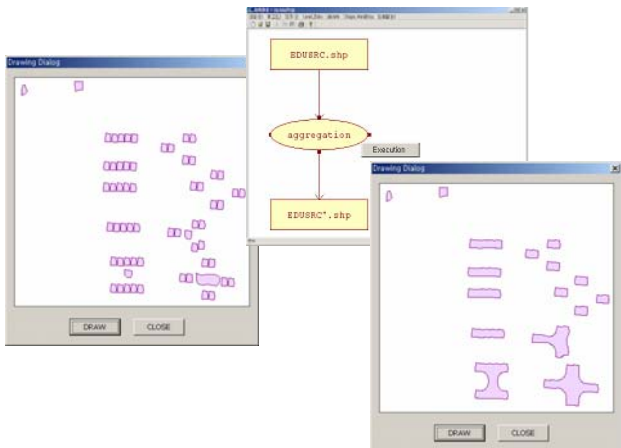


그림 4. 집합 연산자 예 : Aggregation Class

● **갱신 연산자 패키지** : 원시 데이터가 갱신되었을 때 유도된 데이터의 갱신 처리를 위한 클래스(예. ClassificationUpdate) 등이 포함되어 있다.

그림 5는 원시 데이터(그림 4)가 갱신되었을 경우 유도된 데이터에도 갱신이 반영된 예를 보여준다(그림 4와 그림 5 비교).

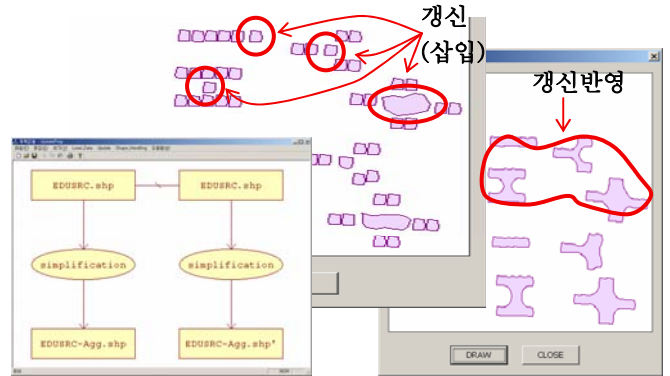


그림 5. 갱신 연산자 예 : AggregationUpdate Class

● **위상 관계** : 원시 데이터에 있는 위상 관계로부터 유도된 데이터의 위상 관계를 판단할 수 있는 기능들이 제공된다.

5. 향후 연구

본 연구에서는 다중축척 데이터베이스를 유도/관리하기 위한 도구를 소개하였다. 이 도구는 기존 연구들이 제시한 방법들을 통합/개선하여 하나의 시스템으로 구현함으로써, 다중축척 데이터를 유도하는 전체적인 틀을 제공하는 점에서 의의가 있다. 향후에는 아직 구현이 진행 중인 유도된 다중축척 데이터의 일관성 및 정확성 평가연산자들을 완성하여 실제 데이터를 대상으로 도구를 평가할 것이다.

참고문헌

[1] R. McMaster and K. Shea, *Generalization in Digital Cartography*, The Association of American Geographers, 1992, pp 1-134.
 [2] M. Egenhofer, *A Model for Detailed Binary Topological Relationships*, *Geomatica* 47 Vol3(4), pp 261-273, 1993.
 [3] H. Kang, S. Do and K. Li, *Model-Oriented Generalization Rule*, Proc.(in CD) ESRI Conf. San Diego, USA, July, 2001
 [4] H. Kang and K. Li, *Rules for Control Propagation of Geo-spatial Data Generalization*, *Journal of the Korea Open Geographic Information Systems Research Society*, Vol4(1), pp 5-14, 2002.
 [5] H. Kang and K. Li, *Incremental Update of Spatial Databases Derived by Generalization*, Technical Report, Spatiotemporal Database Lab., 2002.

