

## 계층적 시각화 기법을 활용한 데이터 큐브의 탐색 방안

오미화\*, 황만모\*, 최정우\*, 최인수\*\*

### An Approach to Navigating Data Cubes with a Hierarchical Visualization Technique

Mi-Hwa Oh\*, Man-Mo Hwang\*, Jung-Woo Choi\*, In-Soo Choi\*\*

#### 요약

다량의 복잡한 데이터를 잘 분석하고자 하는 의도로 최종 사용자가 데이터 큐브 내에 있는 여러 데이터 뷰 중에서 바라는 데이터 뷰를 시각적으로 탐색하게끔 해주는 기능을 OLAP 시스템에서는 계속 마련하고 있다. 본 연구에서는 자신의 스키마가 현 OLAP 시스템에서는 구현될 수 없는 배타적 대칭 계층과 같은 것이 되는 그런 데이터 큐브 만 대상으로 하고자 한다. 본 연구에서는 추상 계층의 개념적 분류를 하였고, 본 연구에서 개발한 계층적 시각화 기법을 활용하여 데이터 큐브를 탐색해 나가는 방안을 제시하고 있다. 계층적 시각화 기법은 이항 추이폐포 개념을 활용하여 개발하였다. 국가자격관리 영역을 예로 들어 이 방안을 설명하고 있다.

▶ Keyword : 계층적 시각화, 추상 계층, 배타적 대칭 계층, 데이터 큐브 탐색, 추이폐포, 스키마

#### Abstract

To efficiently analyze complex and voluminous data, OLAP systems increasingly provide functionalities for visual exploration of the data allowing end-users to navigate the desired view of the data cube. This paper only deals with data cubes whose schemas represented like the exclusive symmetric hierarchy which is not addressed by current OLAP implementations. This paper presents a conceptual classification of abstraction hierarchies, and an approach to navigating data cubes with a hierarchical visualization technique. The hierarchical visualization technique is developed by using the transitive closure of a binary relation. The approach is exemplified using a real-world study from the domain of national license administration.

▶ Keyword : Hierarchical visualization, Abstraction hierarchy, Exclusive symmetric hierarchy, Data cube navigation, Transitive closure, Schema

• 제1저자 : 오미화    교신저자 : 최인수

• 투고일 : 2011. 01. 24, 심사일 : 2011. 01. 27, 게재확정일 : 2011. 01. 31.

\* 송실대학교 대학원 산업·정보시스템공학과 박사 과정

(In a Ph.D. program of Industrial & Information Systems Engineering, Soongsil Univ.)

\*\* 송실대학교 산업·정보시스템공학과 교수 (Professor of Industrial & Information Systems Engineering, Soongsil Univ.)

## I. 서론

OLAP (On-Line Analytical Processing) 시스템[1-3]은 오늘날 비즈니스 분야에서는 물론 비즈니스 이외의 여러 분야에서도 데이터 분석에서의 핵심 기술로 발전하고 있다. 지난 여러 해 동안 OLAP에서는 대용량 데이터 큐브에 있는 데이터에 어떻게 잘 접근하고 질의하느냐에 관련된 데이터 질의 문제와 OLAP 서버 내에서 데이터를 어떻게 잘 나타내고 처리하느냐에 관련된 데이터 모델링 문제를 주로 다루어 왔지만, OLAP 사용자와 의사 결정자에게 어떻게 잘 데이터를 제시하느냐에 관련된 데이터 시각화(visualization) 문제에 대해서는 크게 신경을 쓰지 못한 것도 사실이다[2]. 오늘날, 데이터가 복잡해지고 양이 급증함에 따라 OLAP 시스템이 발휘해야 할 기능도 변하고 있다. 종래에는 보고서 시스템(reporting system)으로서의 기능을 주로 수행했지만, 오늘 날은 탐구 분석(exploratory analysis)의 역할까지도 수행해야 한다. 다량의 복잡한 데이터를 능률적으로 분석하고 이 중에 매몰되어 있는 중요한 사실을 밝혀내는 탐구분석을 하자면 사용자가 데이터의 덩어리 중에서 바라는 덩어리를, 달리 말하면 바라는 데이터 뷰(data view)를 시각적으로 탐색해 나가도록 OLAP은 데이터 시각화 문제를 해결해 나가야만 한다[4-6].

OLAP 시스템에서는 여러 차원(dimension)으로 구성되는 다차원 모델(multidimensional model)이라는 추상 모델을 사용하여 데이터를 구성화 시키는데, 여기서 각각의 차원은 부모 수준(level)과 자식 수준이 연결되어 계층(hierarchy)을 이루는 식으로 구성 되게 된다. 그런데 차원을 계층 식으로 구성하는 이유는 다음과 같다. 실제세계에서의 복잡한 시스템을 표현하는 방법 중의 하나가 추상(abstraction)인데, 추상은 실제세계에서의 어떤 전체를 나타내는데 있어서 세세한 모든 부분까지 나타내지 않고 대표적으로 중요한 부분만을 골라서 나타내고자 하는 한 가지 방법론이다. 현실적으로 모든 세세한 부분을 다루는 것이 불가능해서 추상이란 방법을 채택하고 있지만, 이러한 추상에서 다루는 중요한 대표적인 부분까지도 너무 많아서 이들의 지적관리(intellectual management)가 어렵다는 문제가 생기게 된다. 이러한 문제를 해결하기 위해서 등장한 것이 추상 모델에서의 계층(hierarchy)이다[7]. 수많은 중요한 부분을 계층에서의 여러 수준(level)으로 나누어 분산시킴으로써 관리를 쉽게 할 수 있다는 것이 추상계층(abstraction hierarchy)의 장점이다. 이러한 연유로 본 연구에서는 차원을 추상 계층이라 부르기로 한다. 그런데 여기서 중요한 사실은 추상 계층의 수준 조합 별로 데이터 뷰가

생성된다는 것이다. 예를 들어 2 개의 계층이 있고, 이들 계층은 각각 5개와 6개의 수준으로 구성되어 있다면 모두  $30(5*6)$  개의 데이터 뷰가 생기게 된다는 설명이다. 따라서, 사용자가 이들 30개의 데이터 뷰 중 원하는 데이터 뷰를 잘 찾아 가서 들여다 볼 수 있도록 해 주는 시각적 기법의 개발이 절실히 지는 것이다.

본 연구에서는 자신의 스키마가 현 OLAP 시스템에서는 구현될 수 없는 배타적 대칭 계층과 같은 것이 되는 그런 데이터 큐브 만 대상으로 하고자 한다. 본 연구에서는 먼저 추상 계층의 개념적 분류를 하었는데, 배타적 대칭 계층은 이러한 분류에서 나온 계층의 이름을 일컫는다. 본 연구에서 개발한 계층적 시각화 기법을 활용하여 데이터 큐브를 탐색해 나가는 방안을 제시하고자 하는 것이 본 연구의 목적이다. 국가 자격관리 영역을 예로 들어 이 방안을 설명하고자 한다.

## II. 추상 계층

### 1. 추상계층의 분류

Codd[8]가 주장한 릴레이션 스키마(relational schema)에서의 릴레이션(relation) 간에는 Has-A 관계와 Is-A 관계라는 두 가지의 관계를 찾아볼 수 있다. Has-A 관계란 하위 릴레이션을 모아서 상위 릴레이션을 구성하는 관계를 말한다. Is-A 관계는 상위 릴레이션이 여러 하위 릴레이션으로 분할되는 관계를 말한다. Smith등[7]은 Has-A 관계를 추상 계층에서의 집성(aggregation)이라고 부르고, Is-A 관계를 추상계층에서의 일반화(generalization)라고 부르고 있다.

기존 여러 연구[7][9-10]에서는 분석 기준의 수에 따라 추상계층을 분류하고 있다. 분석 기준(criterion)의 수가 하나인 경우 직립성(simple), 비직립성(non-strict) 그리고 다중(multiple) 계층으로, 두 개 이상인 경우 다중기준(parallel) 계층으로 분류하기 시작했다. 그러나 본 연구에서는 추상계층의 분류를 집성이나 아니면 일반화나에 따라서 분류를 시작하고자 한다. 왜냐하면 추상 계층은 아래 여러 수준이 모여서 위 수준이 되는 Has-A 관계와 최상위 수준이 아래 수준으로 나누어지는 Is-A 관계의 두 가지만으로 궁극적으로 분류되는 것이지[8], 분류 기준의 수에 따라서 분류되는 것은 아니기 때문이다.

본 연구에서는 추상계층의 분류를 집성이나 일반화나 여부에 따라 시작하고, 다음으로 집성에서는 분석하는 기준의 수, 수준 공유 여부, 최대 카디널 수 그리고 최소 카디널 수에 따라 순차적으로 분류 하였고, 일반화에서는 최대 카디널 수에

따라서 먼저 분류하고 다음으로 최소 카디널 수에 따라 분류하였다. 이하, 집성과 일반화에 있어서의 추상계층의 분류에 관해 상세히 설명하고자 한다.

2. 집성에서의 분류

집성(集成)이라는 것은 연관성이 없는 여러 가치를 모아 체계 있는 하나를 이룬다는 뜻이다[11]. 여기서 연관성이 없다는 것은 엔티티(entity)의 식별자(identifier)가 서로 다르다는 것을 의미한다. 따라서 집성에서는 다른 정체성을 가진 여러 하위 엔티티가 모여서 상위 엔티티를 이루게 된다. 본 연구에서는 집성에서의 추상계층의 분류를 처음에는 분석 기준의 수에 따라, 수준 공유 여부에 따라, 최대 카디널 수에 따라 그리고 최소 카디널 수에 따라 순차적으로 분류하는 방법을 제시하고 있다. 본 절에서는 분석 기준의 수에 따라 제일 먼저 분류한 단일기준(single criterion)과 다중기준(multiple criteria) 계층에 대해 차례로 설명하고자 한다.

2.1 단일기준 계층

본 연구에서는 집성 계층 내에서 분석 기준의 수가 하나일 경우를 단일기준 계층이라 부르기로 한다. 단일기준 계층에 관한 예를 나타낸 것이 그림1이다.

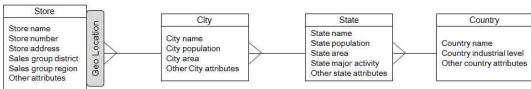


그림 1. 단일기준 계층  
Fig. 1. A Single Criterion Hierarchy

그림 1에서는 분석 기준이 Geo Location이고, 가장 하위 수준인 Store의 멤버가 여러 개 모여 하나의 City로 롤업(roll-up) 됨을 보여주고 있다. 또한, 여러 City 멤버가 모여 하나의 State로 롤업 된다. 이처럼 하나의 분석 기준을 가지며, 모든 인접한 수준에서의 멤버간의 관계가 N:1(many to one)관계로 이루어진 계층을 단일기준 계층이라 부른다.

본 연구에서는 이러한 단일기준 계층을 수준 공유 여부에 따라 단일 경로(single path)와 다중 경로(multiple paths) 계층으로 분류하고 있다.

(1) 수준 공유 여부

수준이 공유되는 계층이란 서로 다른 경로에 의해 공유되는 수준이 하나라도 있게 되는 계층을 말한다. 그림 1은 공유되는 수준이 없는 단일 경로 계층의 경우이지만, 다중 경로가 어떤 수준을 공유하고 있는 계층을 표시한 것이 그림 2이다.

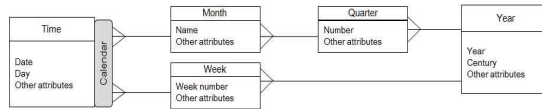


그림 2. 다중 경로 계층  
Fig. 2. A Multiple-Path Hierarchy

그림2에서는 Calendar 기준에 따라 Time 수준의 멤버가 Month, Quarter 그리고 Year 수준으로 차례로 롤업 되는 경로와 Week 그리고 year 수준으로 차례로 롤업 되는 두 개의 경로가 있음을, 또한 이 두 개의 경로가 Year 수준을 공유하고 있음을 알 수 있다. 이러한 계층을 다중 경로 계층이라 부른다.

본 연구에서는 다음으로 이러한 단순 경로와 다중 경로 계층 각각에 대해 최대 카디널 수(maximum cardinality)에 따라서 직립성(strict) 계층과 비직립성(non-strict) 계층으로 분류하고 있다.

(2) 최대 카디널 수에 따른 분류

최대 카디널 수[8]란 인접한 두 수준의 멤버간에 참여할 수 있는 최대 멤버의 숫자를 말한다.

모든 인접한 두 수준 간의 최대 카디널 수가 N:1이 되는 직립성 계층과 인접한 두 수준 간의 최대 카디널 수 중에서 최소한 하나가 N:M(many to many)이 되는 비직립성 계층으로 분류하고 있다.

그림1에서는 Store : City의 최대 카디널 수는 N:1이고, City : State의 최대 카디널 수는 N:1이며, State : Country의 최대 카디널 수도 N:1임을 알 수 있다. 즉 모든 인접한 두 수준간의 최대 카디널 수가 N:1임을 뜻한다. 본 연구에서는 이러한 특성이 있는 계층을 단일 경로를 갖는 직립성 계층이라 부른다. 물론 이러한 특성을 다중 경로에서 갖는다면 본 연구에서는 이를 다중 경로를 갖는 직립성 계층이라 부르고 있다(그림2 참조).

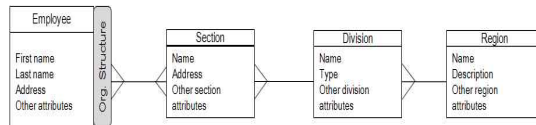


그림 3. 단일 경로 비직립성 계층  
Fig. 3. A Single-Path, Non-Strict Hierarchy

그림3에서는 Employee : Section 의 최대 카디널 수가 N:M으로 되어 있다. 세 개의 인접하는 수준(Employ : Section, Section : Division, Division : Region) 중에서 Employee : Section 하나의 최대 카디널 수가 N:M임을 알

수 있을 것이다. 본 연구에서는 인접한 두 수준 간의 최대 카디널 수 중에서 최소한 하나가 N:M(many to many)이 되는 이러한 계층을 단일 경로에서의 비직립성 계층이라 부른다.

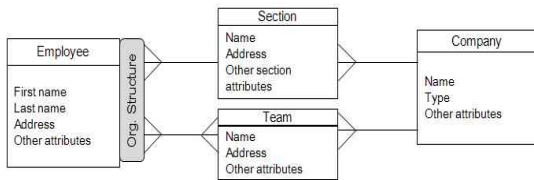


그림 4. 다중 경로 비직립성 계층  
Fig. 4. A Multiple-Path, Non-Strict Hierarchy

마찬가지로 다중 경로에서의 비직립성 계층의 한 예를 나타낸 것이 그림4이다.

단일 경로에 있어서 마지막으로 본 연구에서는 직립성 계층과 비직립성 계층 각각을 최소 카디널 수에 따라 대칭 (symmetric)과 비대칭(asymmetric) 계층으로 분류하고 있다.

(3) 최소 카디널 수에 따른 분류

최소 카디널 수[8]란 인접한 두 수준의 멤버간에 최소한으로 참여해야만 하는 멤버의 숫자를 말한다. 최소 카디널 수에는 M(mandatory, 필수적)과 O(optional, 선택적)의 두 가지가 있는데, M은 한 개의 멤버가 꼭 참여해야만 하는 경우를 나타내고, O는 한 개의 멤버가 꼭 참여하지 않아도 되는 경우를 말한다.

그림1의 경우, City와 Store와의 두 인접한 수준은 M-M의 관계에 놓여있다. City로 롤업 되는 Store의 멤버에는 최소한 한 개의 Store가 꼭 있어야함을 나타내고 있다. State와 City의 두 인접한 수준도 M-M, Country와 State의 두 인접한 수준도 M-M 관계에 놓여있다. 이와 같이 있을 수 있는 모든 인접한 수준이 M-M의 관계에 놓여 있는 계층을 본 연구에서는 대칭 계층이라고 정의하고 있다.

그림 5는 현금지급기(ATM)가 은행지점(Branch)별로 설치되어 있고, 은행 본부(Bank)에는 여러 은행지점이 있음을 나타내고 있다. 그런데 Branch : ATM 수준의 최대 카디널 수는 1:N이나, 최소 카디널 수는 M-O의 관계에 놓여있다. 여기서 M-O의 관계란 하나의 은행지점에 ATM이 한 대도 없을 수가 있다는 것을 의미한다. 이와 같이 계층 수준 간에 M-O의 관계로 놓여 있는 것이 하나 이상 있는 계층을 본 연구에서는 비대칭 계층이라고 정의하고 있다.

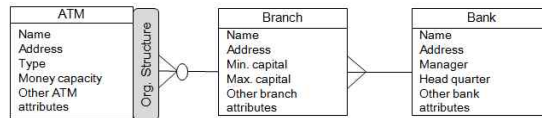


그림 5. 단일경로 직립성 비대칭 계층  
Fig. 5. A Single-Path, Strict, Asymmetric Hierarchy

단일 경로이면서 비직립성 계층을 나타내고 있는 그림3에서는 있을 수 있는 두 인접 수준은 모두 M-M의 관계에 놓여 있음을 알 수 있을 것이다. 이와 같이 모두 M-M의 관계에 놓여 있는 계층을 본 연구에서는 대칭 계층이라 정의하고 있다.

그림 6은 부서(Section)별로 여러 종업원(Employee)이 근무하고 있고, 여러 부서가 하나의 사업부서(Division)으로 롤업되며, 여러 사업부서가 하나의 지역(Region)으로 롤업됨을 보여주고 있다. 여기서 어떤 Section에는 Employee가 한명도 근무하지 않는 M-O의 관계를 보여주고 있다. 이와 같이 계층 수준 간에 M-O의 관계로 놓여 있는 것이 하나 이상 있는 계층을 본 연구에서는 비대칭 계층이라고 정의하고 있다.

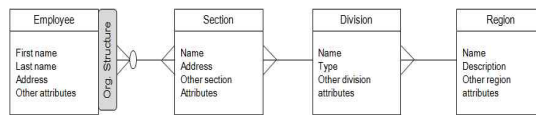


그림 6. 단일경로 비직립성 비대칭 계층  
Fig. 6. A Single-Path, Non-Strict, Asymmetric Hierarchy

지금 부터는 다중 경로 직립성 계층과 다중 경로 비직립성 계층에서의 최소 카디널 수에 대하여 설명하고자 한다. 다중 경로이면서 직립성 계층을 나타내고 있는 그림2에서는 있을 수 있는 두 인접 수준은 모두 M-M의 관계에 놓여 있음을 알 수 있을 것이다. 이와 같이 모두 M-M의 관계에 놓여 있는 계층을 본 연구에서는 대칭 계층이라 정의하고 있다.

그림 7은 부서(Section)별로 여러 종업원(Employee)이 근무하고 있고 여러 부서가 하나의 회사(Company)로 롤업되는 경로와 팀(Team)별로 여러 종업원이 근무하고 있고 여러 Team이 하나의 Company로 롤업되는 경로를 보여주고 있는데, 후자의 경로에 있어서 Team과 Employee 수준 간에는 M-O의 관계가 있음을 알 수 있을 것이다. 이와 같이 계층 수준 간에 M-O의 관계로 놓여 있는 것이 하나 이상 있는 계층을 본 연구에서는 비대칭 계층이라고 정의하고 있다.

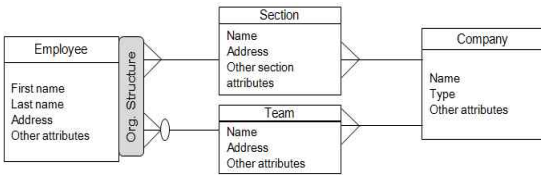


그림 7. 다중경로 직립성 비대칭 계층  
Fig. 7. A Multiple-Path, Strict, Asymmetric Hierarchy

다중 경로이면서 비직립성 계층을 나타내고 있는 그림5에서는 있을 수 있는 두 인접 수준은 모두 M-M의 관계에 놓여 있음을 알 수 있을 것이다. 이와 같이 모두 M-M의 관계에 놓여 있는 계층을 본 연구에서는 대칭 계층이라 정의하고 있다.

그림 8의 Employee-Team-Company 경로에서 Team과 Employee 수준 간에는 M-O의 관계가 있음을 알 수 있을 것이다. 이와 같이 계층 수준 간에 M-O의 관계로 놓여 있는 것이 하나 이상 있는 계층을 본 연구에서는 비대칭 계층이라고 정의하고 있다.

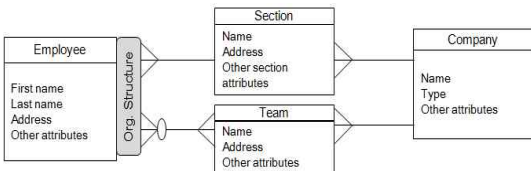


그림 8. 다중경로 비직립성 비대칭 계층  
Fig. 8. A Multiple-Path, Non-Strict, Asymmetric Hierarchy

2.2 다중기준 계층

본 연구에서는 집성 계층 내에서 분석 기준의 수가 둘 이상의 경우를 다중기준 계층이라 부르기로 한다. 다중기준 계층에 관한 예를 나타낸 것이 그림 9이다.

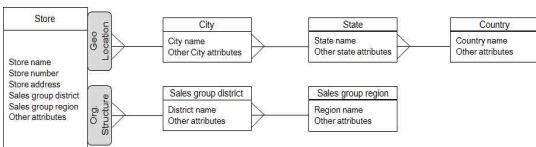


그림 9. 다중기준 계층  
Fig. 9. A Multiple Criteria Hierarchy

그림 9의 계층은 Geo Location 분석 기준에 따른 경로와 Org Structure 분석 기준에 따른 경로로 구성된 계층으로, 이 계층에서는 Geo Location 경로에 의하면 Store 수준의 멤버가 모여서 City로, City 수준의 멤버가 모여서 State로, State 수준의 멤버가 모여서 Country로 롤업 되고, Org

Structure 경로에 의하면 Store 수준의 멤버가 모여서 Sales group district로, Sales group district 수준의 멤버가 모여서 Sales group region으로 롤업 됨을 알 수 있을 것이다.

본 연구에서는 이러한 다중기준 계층을 경로의 수준 공유 여부에 따라 독립(independent)과 비독립(dependent) 계층으로 분류하고 있다.

(1) 경로별 수준 공유 여부

다중기준 독립 계층(a multiple criteria, independent hierarchy)이라는 것은 그림 9와 같이 경로별 수준 간에 공유되는 수준이 없는 계층을 말한다. 반면에 그림 10과 같이 두 개의 경로가 State 수준을 공유하는 계층을 다중기준 비독립 계층(a multiple criteria, dependent hierarchy)으로 정의하고 있다.

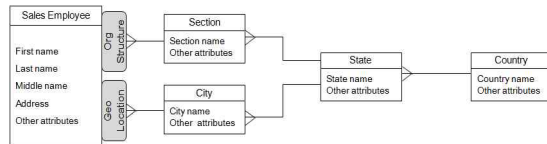


그림 10. 다중기준 비독립 계층  
Fig. 10. A Multiple Criteria, Dependent Hierarchy

여기서 중요한 것은 다중기준 독립 계층이든 다중기준 비독립 계층이든 간에 이들 각 계층을 구성하는 독립된 각각의 경로는 2.1에서 설명한 단일기준 계층에 해당된다는 사실이다. 따라서 이들 계층의 분석은 단일기준 계층의 분석과 일치되기 때문에 본 절에서는 더 이상의 설명을 하지 않기로 한다 (그림 14의 (A)부분 참조).

3. 일반화에서의 분류

일반화(一般化)는 최상위 수준에 있는 하나의 최상위 엔티티(superentity)를 이 수준에 해당되는 한 개의 분석 기준에 따라 자식 수준에서의 여러 개의 하위 엔티티(subentity)로 분할하고, 이들 각각의 하위 엔티티를 각각의 하위 엔티티에 해당되는 각각의 한 개씩의 분석 기준에 따라 다시 손자 수준의 여러 하위 엔티티로 분할해 나가는 과정을 의미한다. 그림 11을 통해 이러한 일반화를 설명하면 다음과 같다. Vehicle이라는 최상위 엔티티를 Moving Area라는 한 개의 분석기준에 따라 Air\_Vehicle, Land\_Vehicle 그리고 Water\_Vehicle이라는 세 개의 엔티티로 분할하고 있으며, Air\_Vehicle 엔티티는 다시 Takeoff라는 한 개의 분석기준에 따라 Plane과 Helicopter라는 두 개의 엔티티로, Land\_Vehicle 엔티티는

Path라는 한 개의 분석기준에 따라 Road\_Vehicle과 Rail\_Vehicle이라는 두 개의 엔티티로, Water\_Vehicle 엔티티는 Shipping Lane이라는 한 개의 분석 기준에 따라 Ocean\_Vessel, Coastal\_Vessel 그리고 River\_Craft라는 세 개의 엔티티로 분할되고 있다. Vehicle 엔티티의 식별자인 VehicleID는 모든 하위 엔티티의 식별자가 됨을 유의하기 바란다. 앞에서 기술한 집성은 보통업(bottom up) 과정인데 비해, 일반화는 톱다운(top down) 과정인 것이다.

그림 11의 설명에서 보드시피 일반화의 계층 분류에 분석 기준의 수를 적용시키는 것은 아무런 의미가 없다. 왜냐하면 집성에서는 최하위 수준에서의 분석 기준이 최상위 수준에 이르기까지 그대로 유지되고 있는데 비해, 일반화에서는 수준이 변화에 따라 각 수준에서 택하는 분석 기준이 각각 독립되게 달라지기 때문이다. 또한 일반화의 계층 분류에 수준공유 여부를 적용시키는 것도 무의미하다. 왜냐하면 톱다운 분할을 하는 일반화에서는 수준 공유가 있을 수 없기 때문이다. 달리 말하면 독립된 각각의 경로만 존재한다는 것이다. 따라서 본 절에서는 일반화 계층의 분류를 하위 엔티티를 공유하는 경로의 존재 여부에 따라 그리고 최소 카디널 수에 따라 순차적으로 분류하는 방법을 제시하고자 한다. 이를 자세히 설명하면 다음과 같다.

3.1 엔티티 공유 경로의 존재여부에 따른 분류

일반화에서는 모든 인접한 두 수준 간의 최대 카디널 수 [8]가 1:1(one to one)이 된다. 즉 일반화에서는 N:1과 N:M이 되는 경우는 존재하지 않는다. 왜냐하면 상위 수준에서 하위 수준으로 분할되는 각각의 독립된 경로만 존재하기 때문이다. 따라서 이들 각각의 독립된 경로가 어떤 엔티티를 공유하느냐 아니냐에 의해 일반화 계층의 분류를 시작하는 것이 타당하다고 본다. 이에 따라 본 항에서는 일반화의 분류를 배타적(exclusive)과 포괄적(inclusive) 계층으로 분류함으로써 시작하고자 한다.

배타적 계층을 나타낸 것이 그림11이고, 포괄적 계층을 나타낸 것이 그림12이다. 그림12를 설명하면 다음과 같다.

Vehicle - Land\_Vehicle - Amphibious\_Vehicle 경로와 Vehicle - Water\_Vehicle - Amphibious\_Vehicle 경로가 Amphibious\_Vehicle 엔티티를 공유하고 있다. 이와 같이 엔티티 공유 경로가 있는 계층을 포괄적 계층이라 부른다. 그림11에서의 ⊗는 배타적 분할을 한 표시이다[7-8]. 즉 Vehicle은 Air\_Vehicle, Land\_Vehicle, Water\_Vehicle 셋 중의 하나로 꼭 분할되는 것을 의미한다. 그림 12에서의 ○는 포괄적 분할을 한 표시이다[8]. Land\_Vehicle이 Road\_Vehicle이나 Amphibious\_Vehicle 양쪽으로 분할될

수 있고, Water\_Vehicle이 Amphibious\_Vehicle이나 Ocean\_Vessel, Coastal\_Vessel, River\_Craft 넷으로 분할될 수도 있음을 의미한다.

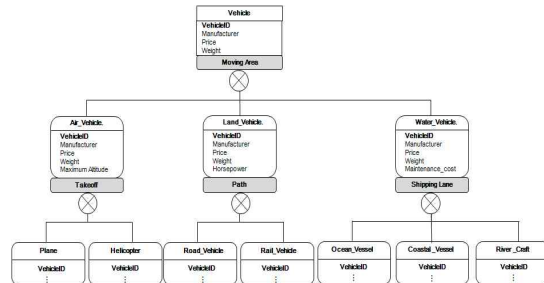


그림 11. 배타적 계층  
Fig. 11. An Exclusive Hierarchy

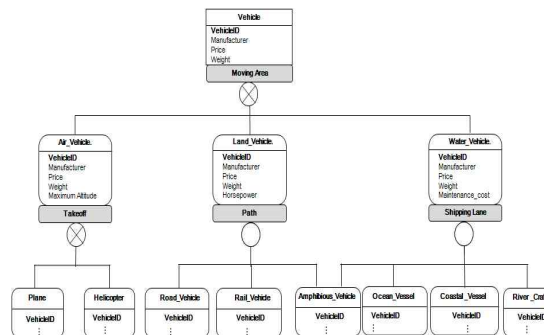


그림 12. 포괄적 계층  
Fig. 12. An Inclusive Hierarchy

본 연구에서는 마지막으로 배타적과 포괄적 계층 각각을 최소 카디널 수에 따라 대칭과 비대칭 계층으로 분류하고 있다.

3.2 최소 카디널 수에 따른 분류

본 연구에서는 최소 카디널 수[8]에 따라 배타적 계층에서 대칭과 비대칭 계층으로 나눌 수 있고, 포괄적 계층에서도 대칭과 비대칭 계층으로 분류하고 있다. 그림 11은 배타적 대칭 계층을, 그림 12는 포괄적 대칭 계층을, 그림 13은 배타적 비대칭 계층을 나타내고 있다. 포괄적 비대칭 계층의 예는 생략하기로 하겠다.

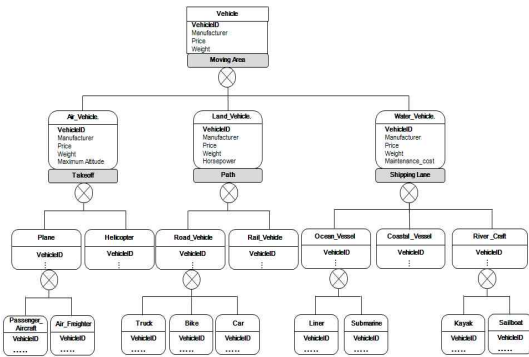
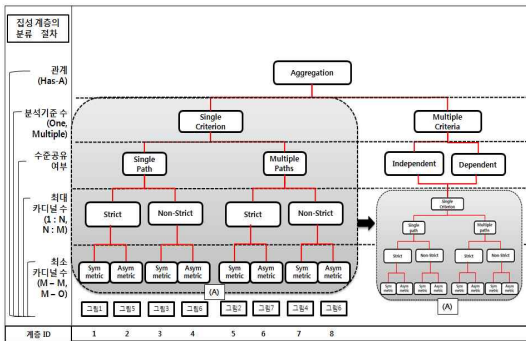
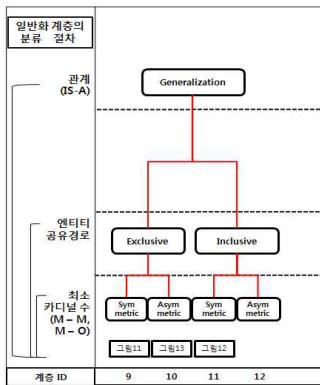


그림 13. 배타적 비대칭 계층  
Fig. 13. An Exclusive, Asymmetric Hierarchy

이상의 II에서 논한 추상 계층의 분류를 종합해 보면 그림 14와 같다. 그림 14의 (a)는 집성 계층에 관한 것이고, (b)는 일반화 계층에 관한 것이다.



(a)



(b)

그림 14. 추상 계층  
Fig. 14. Abstraction Hierarchy

본 연구에서 제시한 그림 14의 분류 방법에 관한 기존 연구를 개괄적으로 살펴보면 다음과 같다.

#### 4. 추상 계층에 관한 기존 연구

본 연구에서는 추상 계층의 분류를 J.M. Smith와 Diane Smith의 연구[7]에 기반을 두어 시작하고 있다. 확장된 E-R 모델(extended entity - relationship model)[8]에서도 계층은 논하고 있지 않지만 엔티티 간의 관계를 Has-A 관계와 Is-A 관계로 대분류하고 있음을 볼 때에 이는 타당하다고 본다. Has-A 관계는 그림 14에서의 집성(a)에 해당되며, Is-A 관계는 일반화(b)에 해당된다.

Maliowski와 Zimányi[9]는 일반화 계층까지도 분석 기준의 수가 하나인 경우로 간주하여 분류하고 있다. 그러나 이는 본 장의 3에서 논한 바와 같이 일반화에서는 각 수준마다 분석기준이 하나씩 독립되어 채택되기 때문에 잘못된 분류 방법이라고 생각한다. 또한 이들이 말한 병렬 계층(parallel hierarchy)은 그림 14(a)의 다중기준 계층에 해당되고, 비공유(non-sharing) 계층과 대안(alternative) 계층이란 단일 경로 계층에 해당되며, 다중(multiple), 공유(sharing) 그리고 포함(inclusive) 계층이란 다중 경로 계층에 해당되며, 단순(simple) 계층이란 직립성 계층에 해당하며, 균질(homogeneous), 균형(balanced), 수준기반(level-based) 그리고 후손 존립성(onto) 계층이란 모두 대칭 계층에 해당되며, 비균형(non-balanced)과 비후손 존립성(non-onto) 계층은 비대칭 계층에 해당된다.

본 연구에서는 그림 14(b)에서의 배타적 대칭 계층(그림 12 참조)을 대상으로 하고자 한다. 대상 사례를 기술하면 다음과 같다.

### III. 사례

현재 우리나라에는 노동인력에 관한 교육과정, 훈련기준 그리고 자격검정을 위한 출제기준의 제정에 출발점으로 기능할 수 있는 국가직무능력표준 시스템이 완벽하게 마련되어 있지 못하고 있다. 이로 인해 각종 교육훈련기관에서 실시하고 있는 교육훈련의 실효성이 떨어지고, 아울러 자격의 공신력이 저하되는 문제가 발생하고 있는 것이다. 따라서 교육훈련과 자격제도가 연계되지 못해 나타나는 문제를 근본적으로 해결하고 산업현장에서 요구하는 현장 직무수행능력을 갖춘 양질의 인력을 공급하기 위해서 국가직무능력표준을 개발하고 적용시킬 필요가 대두되는 것이다[12-14].

본 사례[15]를 든 목적은 국가직무능력표준 시스템 구축의 일환으로 어떠한 직업의 인력이 부족할 경우 해당 직업과

관련된 인력을 어떻게 양성할 것인가 즉, 어떠한 훈련과정을 이수시켜야 되는지 또는 어떠한 자격소지자를 활용할 것인지 등 인력수급과 관련된 분석업무를 지원할 수 있는 국가자격관리 OLAP 시스템을 구축하고자 함에 있다.

본 사례에서는 산업기사, 기사 그리고 기술사의 세 가지 국가 자격을 취득하고자 하는 응시생을 대상으로 하여 인력 수급 분석을 하고자 한다. 각 자격의 취득조건을 살펴보면 다음과 같다.

(1) 자격증 미보유자로서 현장 경력이 최소 2년 이상이면 산업기사 취득 자격시험에 응시할 수 있다. 단, 전문대학 2년 교육에 상당하는 교육을 받은 것은 현장 경력 2년에 해당하는 것으로 한다.

(2) 자격증 미보유자로서 현장 경력이 최소 4년 이상이거나, 산업기사 자격 보유자로서 현장 경력이 최소 1년 이상이 되면 기사 취득 자격시험에 응시할 수 있다. 단, 전문대학 2년 교육에 상당하는 교육을 받은 것과 대학 4년 교육에 상당하는 교육을 받은 것은 각각 현장 경력 2년과 4년에 해당 하는 것으로 한다.

(3) 자격증 미보유자로서 현장 경력이 최소 11년 이상이거나, 산업기사 자격 보유자로서 현장 경력이 최소 6년 이상이 되거나 아니면 기사 자격 보유자로서 현장 경력이 최소 4년 이상이 되면 기술사 취득 자격시험에 응시할 수 있다. 단, 전문대학 2년 교육에 상당하는 교육을 받은 것과 대학 4년 교육에 상당하는 교육을 받은 것은 각각 현장 경력 2년과 4년에 해당하는 것으로 한다.

이상의 설명을 참조하면 어느 시기에 있어서의 국가 자격 취득 응시자를 산업기사 취득 응시자, 기사 취득 응시자, 기술사 취득 응시자의 세 부분으로 완벽하게 나누어 분석하는 것이 가장 적합하다고 결론 내릴 수 있을 것이다. 이에 해당하는 추상 계층은 그림 14(b)의 배타적 대칭 계층인 것이다.

IV에서는 배타적 대칭 계층으로 구성된 국가자격 OLAP 시스템의 구축에 대하여 기술하고자 한다.

### IV 국가자격관리 OLAP 시스템의 구축

#### 1. 차원 스키마

OLAP 분야에서 데이터를 구상화시키고 캐어묻는 논리층(logical layer)의 역할을 하는 아주 중시되는 것이 다차원 모델(multidimensional model)이고, 이 다차원 모델 중에서 제일 중요한 인자가 바로 차원이다. 차원은 스키마(schema)와 인스턴스(instance)로 양분되는데, 먼저 차원 스키마는 입도(granularity)가 다른 여러 수준(level)을 DAG(directed acyclic graph)로 나타낸 것을 말한다. 차원 인스턴스는 각 수준별로 있는 멤버(member)집합과 이들 멤버간의 조상/자손 관계를 나타내는 계층관련(hierarchy relation)으로 구

성되어 있는 것을 말한다. 예로 상점의 위치를 나타내는 차원에서 수준 City의 멤버로 Atlanta를 들 수 있다. 계층관련은 롤업(rollup)으로 나타내는데, 일반적으로 계층에 있어서 수준  $L_1$ 에서 수준  $L_2$ 로 가는 테두리(edge)가 존재하면 수준  $L_1$ 이 롤업되어 다른 수준  $L_2$ 로 된다고 한다. 캐나다와 미국에 있는 상점을 나타내는 하나의 차원을 생각해 보자. 캐나다에 있는 상점은 먼저 시(city)로 롤업되고 계속해서 다시 도(province)로 롤업되며, 미국에 있는 상점은 먼저 시로 롤업되고 계속해서 다시 주(state)로 롤업된다. 세 가지 가능한 차원 스키마를 나타낸 것이 그림 15이다[16].

인접하는 어떠한 두 수준  $L_1$ 과  $L_2$ 에 있어서 수준  $L_1$ 이 롤업되어 다른 수준  $L_2$ 로 된다고 하면 이 차원 스키마는 균질(homogeneous)이라고 부른다. 균질이면서 바닥 수준이 하나인 경우, 이 차원 스키마는 직립성 균질(strictly homogeneous)이라고 부르며, 균질이 아닌 차원 스키마를 이질(heterogeneous)이라고 부른다. 그림 15에서 볼 것 같으면 차원 스키마 (a)는 균질이고, 차원 스키마 (b)는 직립성 균질이며, 차원 스키마 (c)는 이질이다.

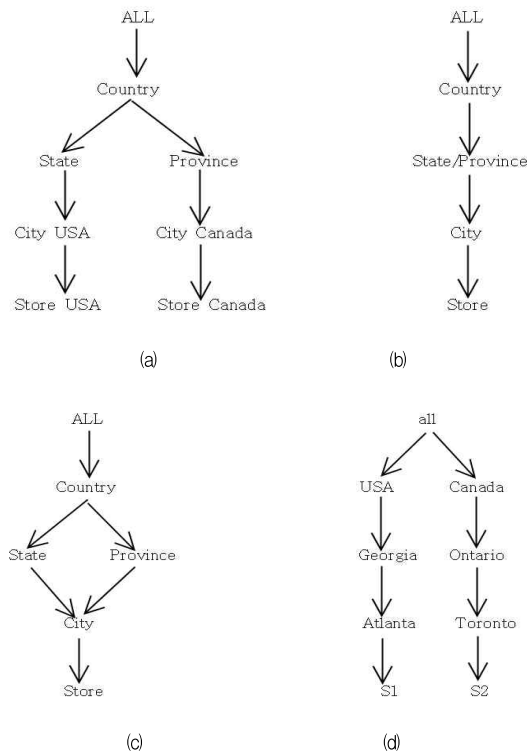


그림 15. (a), (b), (c) 세가지 차원 스키마, (d) 차원 인스턴스  
 Fig. 15. Three alternative dimension schemas(a), (b), (c) and a dimension instance(d)

그림 14의 추상 계층을 이상 기술한 차원 스키마로 나타내면 그림 16과 같다. 그림 16에서는 총 네 개의 수준을 가정하였고, 단순 DAG에다 최대 카디널리티 표시(crow's foot notation)[17]를 첨부하였으며, 계층에서의 각 수준을 노드(node)로 표시하고 있다. 앞으로 본 연구에서는 수준과 노드를 동일 개념으로 혼용하고자 한다. 또한 그림 16의 각 그림 밑에 번호가 기입되어 있는데 이는 그림 14에서의 계층 ID와 같다. 그러나 시스템 구축 시 사용하는 차원 스키마에는 관념 노드(abstract node)[18]의 개념이 추가로 채택되고 있기에 이하 이를 설명하고자 한다.

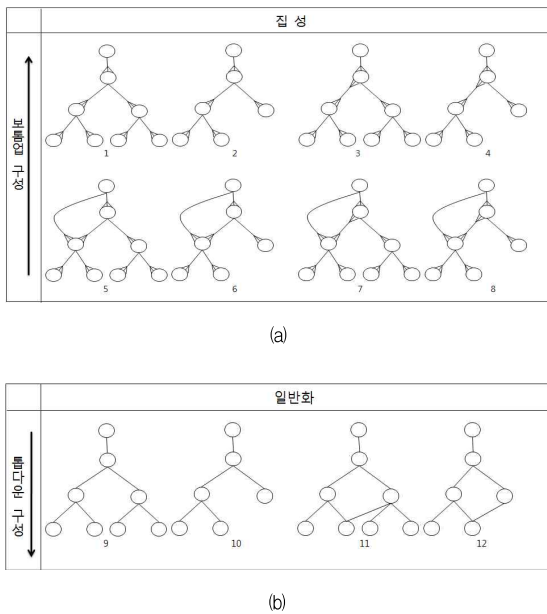


그림 16. 차원 스키마  
Fig. 16. Dimension Schemas

2. 관념 노드

OLAP 큐브는 차원이라 부르는 다음과 같은 릴레이션 스키마(relation schema)를 갖고 있다.

$$D_1 U D_2 U \dots U D_n$$

여기서  $D_i$  는 차원 속성(dimension attribute)을 말하고, 이  $D_i$  에 상응하는 릴레이션은  $d_i$  이다.

릴레이션  $d_i$  에서의 각 행을 엔티티라 부르는데, 상응하는 릴레이션과 엔티티가 존재하지 않는 노드를 관념 노드라 한다. 이 노드는 여러 자식 수준을 하나로 묶는 상부(upper) 수준을 표시하거나, 대상으로 하는 전 계층의 최상부에 있는 루트 노드(root node)를 표시할 때 사용한다.

만약에  $d_i = \emptyset$ 이면  $D_i$ 는 관념적이다.

어떤  $D_i$ 의 하위 노드  $D_k$ 를  $D_i$ 의 자식(children)이라 부르며, 노드  $D_i$ 의 계량 수(cardinality)란  $D_i$ 가 갖는 자식의 수를 말한다.

$$|D_i| = |\text{children}(D_i)|$$

이상의 개념을 바탕으로 차원 노드(dimensional node)의 몇 가지 계층 특성을 살펴보면 다음과 같다.

- (1) 단일기준-단일경로-직립성 계층

그림 1과 같이 단 하나의 분할 경로를 갖는 노드들로 구성되는 계층을 말한다. 이러한 계층에서는 다 대 일(many-to-one) 형식으로 상위수준으로 집계된다. Vinnik와 Mansmann[18]은 다음과 같은 식으로 이 계층을 설명하고 있다.

$$|D_i| = 1 \wedge d_i \neq \emptyset$$

- (2) 단일기준-다중경로 계층

그림 2와 같이 하나의 노드가 여러 경로를 갖는 계층을 말한다. Vinnik와 Mansmann[18]은 다음과 같은 식으로 이 계층을 설명하고 있다.

$$|D_i| > 1 \wedge d_i \neq \emptyset$$

- (3) 일반화 계층

그림 11과 같이 상위 수준이 여러 하위 수준으로 분할되는 IS-A 관계를 갖는 계층을 말한다.

Vinnik와 Mansmann[18]은 다음과 같은 식으로 이 계층을 설명하고 있다.

$$|D_i| > 1 \wedge d_i \neq \emptyset \wedge \forall D_k \in \text{children}(D_i): d_k \subset d_i$$

이상 기술한 관념노드 중 최상부 루트 노드로서의 관념노드만 사용하여 사례로 든 국가자격 OLAP 시스템의 차원 스키마를 그려보면 그림 17과 같다. 그림 17은 LICENSE와 TIME이라는 2개의 계층이 있고, LICENSE 계층의 관념노드는 T**응시자**, TIME 계층의 관념 노드는 T**기간**이며, LICENSE 계층에서의 응시자는 산업기사, 기사, 기술사의 세 가지 자격중 중 어느 하나에만 응시할 수 있음을 보여주고 있다. 이중 산업기사 취득 희망자는 경력만 갖고 있으면 되지만, 기사 취득 희망자는 경력만 갖고 응시할 수도 있고 산업기사 자격증을 취득한 다음 응시할 수도 있음을, 기술사 취득 희망자는 경력만 갖고 아니면 산업기사를 취득한 다음 아니면 처음으로 기사를 취득한 다음 아니면 산업기사와 기사를 차례로 취득한 다음 응시할 수 있음을 보여주고 있다. LICENSE 계층에서 상위 수준인 **응시자**는 자신의 하위 수준인 **산업기사 희망**, **기사 희망**, **기술사 희망** 수준의 밑에 놓여 있음에 유의하기 바라고, **산업기사 희망** 응시자는 아무런 자격증을 갖고 있지 않은 상태이고, **기사 희망** 응시자는 산업기사 자격

증을 갖고 있지 않은 상태이고, **기술사 희망** 응시자는 산업기사 자격증, 기사 자격증을 갖고 있지 않은 상태임에 즉 **산업기사 희망**, **기사 희망**, **기술사 희망** 응시자의 상태는 모두 자격증 미보유자의 상태임에 유의하기 바란다. 응시자 모두는 자격증 미보유 상태에서 시작하여 필요한 자격증을 취득해 나간다는 상황을 나타내고 있다는 뜻이다.

그림 17에서의 LICENSE 계층은 일반화 추상에서의 배타적 대칭 계층(그림 14(b)의 계층 ID 9)이다. 그러나 기존 OLAP 시스템에서는 일반화 추상 계층을 다룰 수 없는 것이 현실이다. 인접하는 어떠한 두 수준 L1과 L2에 있어서 수준 L1이 롤업되어 다른 수준 L2로 된다고 하면 이 차원 스키마는 균질이라고 IV-1에서 언급한 바 있는데, 현 OLAP 시스템에서는 이러한 균질화 계층의 일부만 다룰 수 있다. 그림 16(a)의 집성 계층은 모두 균질화 되어 있고, 그림 16(b)의 일반화 계층은 전부 이질화되어 있다. 따라서 현 OLAP 시스템으로 배타적 대칭 계층을 구현시키자면 이 계층을 우선 그림 16(a)에서와 같은 집성 계층으로 바꾸는 균질화 작업을 하여야 한다.

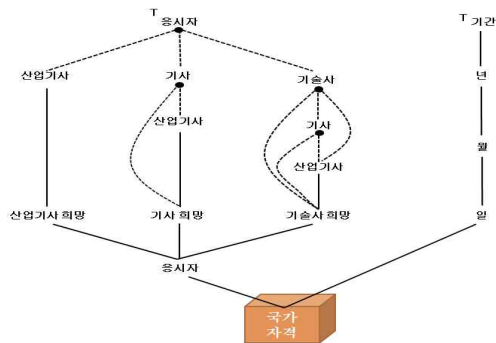


그림 17. 2차원 사실 스키마  
Fig. 17. A two-dimensional fact schema

이하 균질화에 대해서 기술하고자 한다.

### 3. 균질화

SDLC(Systems Development Life Cycle)의 분석단계인 모형작성단계에서 볼 것 같으면 일반화는 상위 수준의 인스턴스가 각자 자신의 속성과 집계 수준을 갖고 있는 하위 수준으로 나누어지는 것을 뜻한다고 볼 수 있다.

상위 수준이란 논리적으로 볼 것 같으면 자신의 하위 수준을 집계하여 만들어 진 것을 말하지만, ME/R(Multidimensional Entity/Relationship) 모형[19]에서는 일반화 엔티티를 “균질화(homogenizing)” 시키고자 상위 수준을 자신의 자식인 어떤 하위 수준의 자식으로 자리매김하게 만든다. 집계 논리로

볼 것 같으면 여러 하위 수준을 모아 집계해서 만드는 상위 수준 엔티티를 이와 같이 하위 수준의 아래에 위치시키는 것을 분명히 잘못된 것이지만, 각 차원마다 하나의 바닥 입도(bottom granularity)의 노드를 놓아야 하고, 동시에 사실 테이블에 있는 하나의 외래 키(foreign key)로 이 차원을 참조시켜야 하기 때문에 할 수 없이 이렇게 위치시키는 것이다.

Mansmann과 Scholl이 개발한 다음과 같은 이질 계층의 균질화 규칙[19]에 따라 그림 17의 LICENSE 계층을 균질화시키면 그림 18과 같이 된다. 이러한 균질화 과정을 그림 18을 참조해 가면서 자세히 살펴보고자 한다.

[규칙 1] 관념(abstract) 노드에는 루트(root) 노드와 루트 노드가 아닌 정규(normal) 노드가 있는데, **하나의 추상 계층에 있어서의 최상위 노드를 루트 노드  $T_D$  로 삼는다.** (가), (나), (다) 영역을 망라했을 때에 최상위 노드인  $T_{\text{응시자}}$  가 이에 해당된다. 반면, 수준 I의  $T_{\text{산업기사}}$ ,  $T_{\text{기사}}$ ,  $T_{\text{기술사}}$  는 정규 노드이다.

[규칙 2] 하위 노드란 부모인 상위 노드의 여러 자식 노드를 말한다. 수준 I의  $T_{\text{산업기사}}$ ,  $T_{\text{기사}}$ ,  $T_{\text{기술사}}$ 가 상위 노드  $T_{\text{응시자}}$ 의 하위 노드가 된다. 이들 하위 노드는 정규 노드이지만, **추상 계층 전체가 아닌 하위 노드 자신의 영역( $D_i$ )에 있어서는 루트 노드  $T_{D_i}$ 의 역할을 하게 한다.** 수준 I-(가)에 있는  $T_{\text{산업기사}}$ 는 (가)영역에 있어서의, 수준 I-(나)에 있는  $T_{\text{산업기사}}$  (나)영역에 있어서의, 수준 I-(다)의  $T_{\text{기술사}}$ 는 (다)영역에 있어서의 루트 노드 역할을 함을 말한다.

[규칙 3] 상위 노드는 하위 노드 하나로 분할되거나 둘 이상으로 분할된다. 본 연구에서는 둘 이상으로 분할되는 경우는 점선으로 표시하고 하나로 분할되는 경우는 실선으로 표시하고자 한다.

Step 1 :  $T_{\text{응시자}}$ 는 하위 노드  $T_{\text{산업기사}}$ ,  $T_{\text{기사}}$ ,  $T_{\text{기술사}}$ 로 나누어지므로  $T_{\text{응시자}}$ 와 이들 하위 노드들은 점선으로 연결된다.

Step 2 : 각 하위 관념 노드에 해당하는 비관념 노드를 생성한다. 수준 I-(가)의  $T_{\text{산업기사}}$  관념 노드에 비관념 노드  $\perp_{\text{산업기사}}$ 를, 수준 I-(나)의  $T_{\text{기사}}$  관념 노드에 비관념 노드  $\perp_{\text{기사}}$ 를, 수준 I-(다)의  $T_{\text{기술사}}$  관념 노드에 비관념 노드  $\perp_{\text{기술사}}$ 를 발생시킨다. 그림 18에서는  $\perp$ 를 생략하였음을 밝힌다.

Step 3 : 비관념 노드는 다시 하위 노드로 분할되기도 하고, 그렇지 않기도 한다. 분할되는 경우는 Step 4로, 그렇지 않는 경우는 멈춘다. 본 예에서는 수준 I-(가)의 비관념 노드  $\perp_{\text{산업기사}}$ 는  $T_{\text{미보유자}}$ 로 분할되고(실선으로 연결),

I-(나)의 비관념 노드  $\perp$ 기사는 하위 수준 T 미보유자와 T 산업기사로 분할된다(점선으로 연결). 그리고 수준 I-(다)의  $\perp$ 기사는 T 미보유자, T 산업기사, T 기사로 분할된다(점선으로 연결).

Step 4 : 분할된 관념 노드가 T 미보유자 로 되면 더 이상 분할을 종결하고 이 노드에  $\perp$  미보유자 비관념 노드를 발생시킨다. 수준 III-(나)의 T 산업기사, 수준 III-(다)의 T 산업기사, 수준 II-(다)의 T 기사 에 대해서는 각각 해당 되는 비관념 노드를 발생시키고 Step 3로 돌아가 이들을 다시 분할한다.

Step 3에서는 수준 III-(나)의 비관념 노드  $\perp$  산업기사는 T 미보유자로 분할되고(실선으로 연결), 수준 III-(다)의 비관념 노드  $\perp$  산업기사는 T 미보유자로 분할되고(실선으로 연결), 비관념 노드 수준 II-(다)의  $\perp$ 기사는 T 산업기사와 T 미보유자로 분할된다(점선으로 연결). Step 3로 돌아가 수준 IV-(나)의 T 미보유자 에는 비관념 노드  $\perp$  미보유자 생성시키고 분할을 종결시키며, 수준 IV-(다)의 T 미보유자 에는 비관념 노드  $\perp$  미보유자 생성시키고 분할을 종결한다. 수준 III-(다)의 비관념 노드  $\perp$  산업기사는 T 미보유자 로 분할되며(실선으로 연결) 이 노드의 비관념 노드  $\perp$  미보유자 생성시키고 분할을 종결한다.

Mansmann과 Scholl의 규칙에 따라 그림 17을 균질화시킨 그림 18에서의 LICENSE 계층은 그림 16(a)의 계층 ID 7인 집성 계층 부류에 해당한다고 볼 수 있다. 그런데 이 그림 18에서의 LICENSE 계층에는 크게 두 가지 문제점이 있다고 본다. 첫째, 이 계층에는 여섯 가지의 자격증 취득 경로만 표현되고 있다는 점이다. 그림 17의 사례 스키마에서는 일곱 가지의 자격증 취득 경로가 있음을 볼 때에 이는 잘못 표현 되었다고 본다.

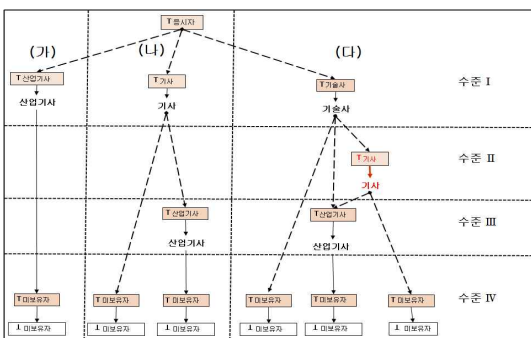


그림 18. 관념 노드를 사용한 배타적 대칭 계층의 개조  
Fig. 18. Reshaping exclusive symmetric hierarchy using abstract node

여섯 가지로 표현된 이유는 경로를 분할하는 것에 중점을 두고 분할되는 경로가 추후에 다른 경로에서의 수준을 공유하는 사실에 대해서는 전혀 신경을 쓰지 않는 일반화 추상의 톱다운 분석법에 있다고 본다. 구체적으로 **기술사 - 산업기사-미보유자** 경로와 **기술사 - 기사 - 산업기사 - 미보유자** 경로가 산업기사 수준을 공유하는 사실을 무시했기 때문에 이러한 결과가 나왔다고 본다. 둘째, 계층 ID 7과 같은 집성 계층은 현 OLAP 시스템으로는 구현시킬 수 없다는 점이다. 그림 16(a)의 계층 ID 1과 같은 직립성 대칭 계층만을 현 OLAP 시스템에서 구현시킬 수 있다.

여기서 두 번째의 문제점을 해결하자면 계층 ID 7을 계층 ID 1로 전환시켜야만 한다. 이러한 전환 작업을 정규화 작업이라 부르는데 이하 이 작업에 대해서 기술하고자 한다.

4. 정규화

어떤 계층에 대하여 망라성(covering), 후손 존립성(onto), 직립성이라는 3가지 요약 조건(summarizability condition) [9][16][19-22]을 만족시키도록 전환 시키는 작업을 정규화 작업이라 부른다. Pedersen과 Mansmann이 제안한 정규화 기법(hierarchy normalization technique)[19]에서의 세 가지 요약 조건을 설명하면 다음과 같다.

(1) 망라성

$L_i \subset L_j$ 가 되는 즉  $L_j$  수준이  $L_i$  수준의 부모가 되는 두 개의 수준  $L_i, L_j$ 에 대해서

$$\forall e_1 \in L_i : \exists e_2 \in L_j \wedge e_1 \subseteq e_2 \text{ ----- } \textcircled{1}$$

가 되면 계층이 망라되어 있다고 한다.  $e_1$ 은 수준  $L_i$ 에 속하는 인스턴스이고,  $e_2$ 은 수준  $L_j$ 에 속하는 인스턴스이다.  $\subseteq$ 는 부모자식 관계를 나타내는데,  $e_1 \subseteq e_2$ 는  $e_2$ 가  $e_1$ 의 부모라는 것을 나타낸다. 망라성을 나타내는 ①을 설명하면 다음과 같다. 자식 수준  $L_i$ 의 모든 인스턴스  $e_1$ 에 대해서 부모 수준  $L_j$ 에 속하면서 이  $e_1$ 의 부모가 되는 인스턴스  $e_2$ 가 존재하면 이 계층은 망라되어 있다고 한다는 뜻이다. 망라되어 있지 않은 이러한 계층을 ①의 조건을 만족시키게끔 변환시켜야 한다.

(2) 후손 존립성

$L_i \subset L_j$ 인 두 개의 수준  $L_i, L_j$ 에 대해서

$$\forall e_2 \in L_j : (\exists e_1 \in L_i \vee (\exists e_k \in L_k \wedge L_k \subset L_j) \wedge e_1 \subseteq e_2 \text{ ----- } \textcircled{2}$$

가 되면 후손 존립성의 계층이 된다고 한다. ②를 말로 설명하면 바닥 수준에 속하지 않는 어떤 인스턴스에 대해서도 그 후손 수준에 속하는 인스턴스가 꼭 존립되어야 한다는 뜻이다. 바닥 수준 인스턴스가 아니라면 꼭 후손 인스턴스를 갖어야 한다는 ②의 조건을 충족시키게끔 계층을 개조시켜야 한다.

(3) 직립성

$L_i \subset L_j$ 인 두 개의 수준  $L_i, L_j$ 에 대해서

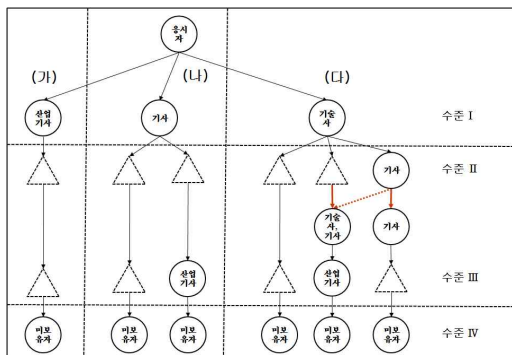
$$\forall e_1 \in L_i : e_2, e_3 \in L_j \wedge e_1 \subseteq e_2 \wedge e_1 \subseteq e_3 \Rightarrow$$

$$e_2 = e_3 \quad \text{③}$$

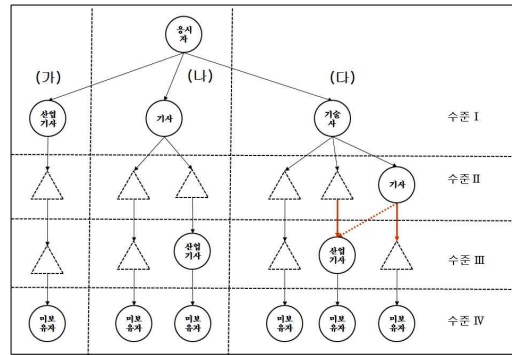
가 되면 직립성 계층이 된다고 한다. ③을 말로 설명하면 자식 수준 인스턴스는 꼭 하나의 부모 수준 인스턴스를 갖아야 한다는 뜻이다. 모든 계층이 직립성 계층이 되게끔 계층을 개조시켜야 한다.

그림 18의 계층에서는 후손 존립성 조건은 발생하지 않는다. 바닥 수준인 수준 IV가 바닥 입도의 노드로 다 채워져 있기 때문이다. 망라성 조건은 나중에 충족시켜도 되기 때문에 우선적으로 직립성 조건을 충족시켜야 한다. 그림 18의 계층이 직립성 조건을 충족시키도록 Pedersen과 Mansmann이 제안한 정규화 작업을 그림으로 소개하면 그림 19와 같다. 그림 18의 수준 II-(다)와 수준 III-(다) 사이의 N:M 관계인 비직립성 문제를 해결하기 위해서 세 가지 방법을 제안한 것이 그림 19인 것이다.

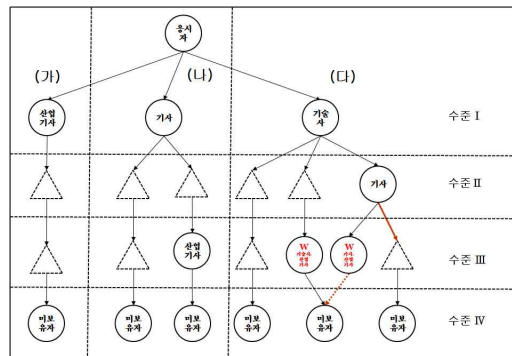
다중-부모 문제를 해결하기 위해 (1) 새로운 중간 노드(중간매체로써 관계 테이블)를 부모 수준과 자식 수준 사이에 생성시키고 이 중간 노드가 여러 부모 중 하나의 부모를 명시시키는 방법(실선으로 연결, 그림 19의 (a)), (2) 각각의 자식 값(child value)에게 우선권을 가진 단일 부모를 명시하는 방법(실선으로 연결, 그림 19의 (b)), (3) 다중 부모들 각각에 속하는 가중치를 명시하는 방법(그림 19의 (c))을 제시했다.



(a)



(b)



(c)

그림 19. 직립성 정규화  
Fig. 19. Strict Normalization

그러나 이 세 가지 방법은 저자 자신도 밝혔지만 계층 ID 7을 계층 ID 1로 전환 시킬 만큼 완벽하고 정확한 방법이 되지 못한다.

본 연구에서는 IV-3에서 언급한 첫 번째 문제점과 본 절에서 언급한 비직립성 문제를 해결하기 위해 추이페포 분석 기법을 제안하고 있다. 이하 이에 대한 설명을 하고자 한다.

### V 추이페포 분석 기법

#### 1. 추이페포 기법

집합 S에서 집합 T로의 이항관련(binary relation) R은 S×T의 부분집합이 된다.

$$R \subseteq S \times T$$

여기서 S=T 일 경우, R을 S에 관한 이항관련이라 부른다.

집합 S에 관한 이항관련 R에는 추이성(transitivity)이

결여되어 있는 경우가 있는데, 이때에 R을 확장시킴으로써 추이성을 갖추게끔 할 수 있다. R을 확장시킨다는 것은 R을 포함하면서도 R보다 큰 완벽한 추이성을 갖춘 확장된 S×S의 부분집합을 찾아낸다는 것을 말하는데 여기서 최소한의 확장에 해당되는 부분을 S에 관한 R의 추이폐포(推移閉包, transitive closure) R+라 부른다[23-25].

이하, 본 연구 사례에서의 국가자격 취득 과정을 살펴봄으로써 추이폐포를 설명하고자 한다. 자격증 미보유자(노드 ①로 표기)가 산업기사(노드 ②로 표기), 기사(노드 ③으로 표기), 기술사(노드 ④로 표기)로 되어나가는 일련의 과정을 나타내는 다음과 같은 S와 R을 고려해 보자(노드 표시를 간단히 숫자로 표기).

$$S = \{ 1, 2, 3, 4 \}$$

$$R = \{ (1,2), (2,3), (3,4) \}$$

R의 원소 중 (1,2)는 자격증 미보유자가 산업기사가 되고, (2,3)은 산업기사에서 시작하여 기사가 되며, (3,4)는 기사에서 시작하여 기술사가 되는 것을 나타내는 각 관련(relation)을 뜻한다. 즉 상기 R은 자격증 미보유자가 차례로 산업기사가 되고 기사가 되며 기술사가 되는 대표적인 국가자격 취득과정을 나타내고 있다. 그러나 어떤 사람은 산업기사를 취득하지 않고 바로 기사가 되기도 하고, 기술사가 되기도 함을 고려해 보면 상기 R에는 추이성이 결여되어 있음을 알게 될 것이다. 따라서 이 R이 완벽한 추이성을 갖게끔 R을 다음과 같이 확장시켜보기로 한다.

R의 (1,2)와 (2,3)에서 즉 자격증 미보유자가 산업기사가 되고 산업기사에서 기사가 된다는 각각의 관련으로부터 자격증 미보유자가 기사가 된다는 관련 (1,3)을 얻게 되고, 같은 방법으로 관련 (2,4)와 (1,4)를 얻게 된다.

$$(1,2) \in R \wedge (2,3) \in R \rightarrow (1,3) \in R+$$

$$(2,3) \in R \wedge (3,4) \in R \rightarrow (2,4) \in R+$$

$$(1,3) \in R \wedge (3,4) \in R \rightarrow (1,4) \in R+$$

이와 같이 얻은 관련 (1,3), (2,4), (1,4)를 기존 R에 추가함으로써 추이폐포 R+를 얻게 되는 것이다.

$$R+ = R \cup \{ (1,3), (2,4), (1,4) \}$$

$$= \{ (1,2), (1,3), (1,4), (2,3), (2,4), (3,4) \}$$

이렇게 구한 추이폐포 R+를 DAG로 나타내면 그림 20과 같다. R+에서 자격증 미보유자가 기술사의 신분에 이르는 것을 나타내는 도달관련(reachability relation)을 살펴보면 총 4가지 경로가 있음을 알게 된다.

$$(1,2) \wedge (2,3) \wedge (3,4) \rightarrow (1,4)$$

$$(1,2) \wedge (2,4) \rightarrow (1,4)$$

$$(1,3) \wedge (3,4) \rightarrow (1,4)$$

(1,4) → (1,4)  
이를 나타낸 것이 그림 20인 것이다.

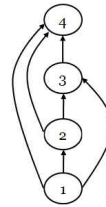


그림 20. R = { (1,2), (2,3), (3,4) }의 추이폐포  
Fig. 20. Transitive closure of R = { (1,2), (2,3), (3,4) }

같은 방법으로 자격증 미보유자가 기사로 되어나가는 과정을 자격증 미보유자(노드 ①로 표기)에서 시작하여 산업기사(노드 ②로 표기)가 되고, 다음으로 기사(노드 ③으로 표기)가 되는 전형적인 예로부터 살펴보고, 또 이에 관련된 추이폐포와 DAG(그림 21)를 살펴보면 다음과 같다.

$$S = \{ 1, 2, 3 \}$$

$$R = \{ (1,2), (2,3) \}$$

$$(1,2) \in R \wedge (2,3) \in R \rightarrow (1,3) \in R+$$

$$R+ = R \cup \{ (1,3) \}$$

$$= \{ (1,2), (2,3), (1,3) \}$$

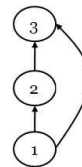


그림 21. R = { (1,2), (2,3) }의 추이폐포  
Fig. 21. Transitive closure of R = { (1,2), (2,3) }

이렇게 구한 추이폐포 R+를 DAG로 나타내면 그림 22와 같다.

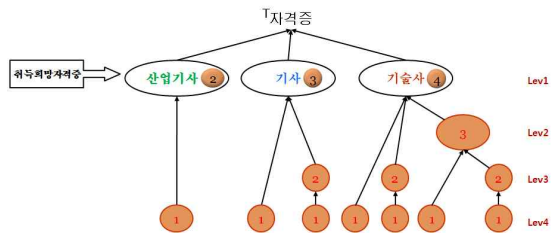


그림 22. 자격증 추이폐포  
Fig. 22. LICENSE Transitive Closure

그림 22의 자격증 추이폐포를 2차원 사실 스키마

로 변환한 것이 그림 23이다.

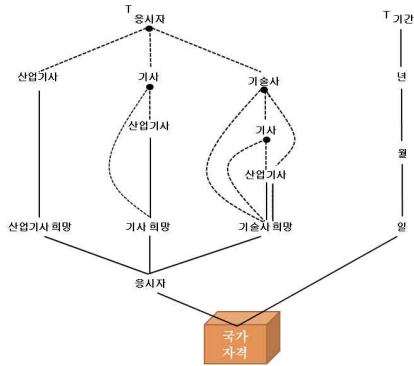


그림 23개조한 2차원 사실 스키마  
Fig. 23. Revised 2-dimensional fact schema

그림 23에서 볼 것 같으면 자격증을 취득하는 경로가 7가지가 되어서 IV-3에서 언급한 첫 번째 문제점이 해결되었고, IV-4에서 언급한 비직립성 문제도 해결되었음을 알 수 있다. 그림 22는 망라성 조건을 충족시키지 못하고 있다. 더미(dummy)[26]를 채워 넣음으로써 망라성 조건을 충족시킨 것이 그림 23이다. 결과적으로 그림 18을 계층 ID 1의 계층으로 전환시킨 것이 그림 23인 것이다.

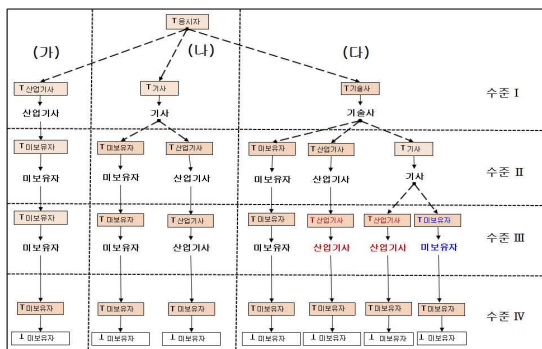


그림 24. ID 9 계층에서 유도한 ID 1 계층  
Fig. 24. An ID 1 hierarchy derived from an ID 9 hierarchy

그림 24는 현 OLAP 시스템에서 구현할 수 없는 일반화 - 배타적 - 대칭 계층(ID 9)을 현 OLAP 시스템에서 구현 가능한 집성 - 단일기준 - 단일 경로 - 직립성 - 대칭 계층(ID 1)으로 변환한 것으로 본 연구의 핵심 결과물이다. 이하 이 결과물을 구현시켜 보고자 한다.

## 2. OLAP 시스템에서의 구현

### 2.1 LICENSE 테이블의 작성

그림 24의 내용을 LICENSE 테이블로 설계를 한 것이 표 1이다. 표 1에서 LicenseID는 그림 24에서의 7가지의 경로를 나타낸 것이고, Lev4은 수준 IV, Lev3는 수준 III, Lev2는 수준II, Lev1은 수준 I을 나타낸 것이다.

표 1. LICENSE 테이블  
Table 1. LICENSE Table

LicenseID	Lev4	Lev3	Lev2	Lev1
1	미보유자I	미보유자I	미보유자I	산업기사
2	미보유자2	미보유자2	미보유자2	기사
3	미보유자3	산업기사3	산업기사3	기사
4	미보유자4	미보유자4	미보유자4	기술사
5	미보유자5	산업기사5	산업기사5	기술사
6	미보유자6	산업기사6	기사	기술사
7	미보유자7	미보유자7	기사	기술사

표 1과 EXAMINE 테이블, TIME 테이블, 그리고 경력 을 측정값으로 삼아 가상으로 설정한 100명의 응시자데이터로 구성된 사실테이블을 사용하여 국가 자격 OLAP 큐브를 구성해 보면 그림 25와 같고 이를 실행 시켜서 최상위 수준에서 마지막 잎 수준(leaf level)까지 수준 별로 전개 시킨 결과가 그림 26이다.

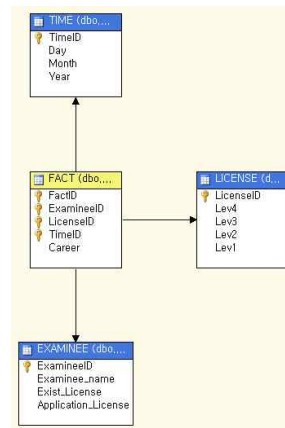


그림 25. LICENSE 차원 스키마  
Fig. 25 LICENSE Dimension Schema

Levl	Career	FACI 카운트	AvgCareer
기사		59	10
기사	미보유자2	26	4
기사	산업기사3	33	6
기사	합계	59	10
기술사		466	56
기술사	미보유자4	79	7
기술사	산업기사5	112	14
기술사	합계	466	56
산업기사		95	34
산업기사	미보유자1	95	34
산업기사	합계	95	34
총합계		620	100

(a)

Levl	Lev2	Career	FACI 카운트	AvgCareer
기사	미보유자2	미보유자2	26	4
	산업기사3	산업기사3	33	6
	합계	합계	59	10
기술사	기사	기사	275	35
	미보유자4	미보유자4	79	7
	산업기사5	산업기사5	112	14
합계	합계	466	56	
산업기사	미보유자1	미보유자1	95	34
	합계	합계	95	34
총합계			620	100

(b)

Levl	Lev2	Lev3	Career	FACI 카운트	AvgCareer
기사	미보유자2	미보유자2	26	4	6.5
		합계	26	4	6.5
		산업기사3	33	6	5.5
기사	합계	산업기사3	33	6	5.5
		합계	59	10	5.9
		합계	59	10	5.9
기술사	기사	미보유자7	169	20	8.5
		산업기사6	106	15	7.1
		합계	275	35	7.9
	미보유자4	미보유자4	79	7	11.3
		합계	79	7	11.3
		합계	79	7	11.3
산업기사5	산업기사5	112	14	8.0	
	합계	112	14	8.0	
	합계	466	56	8.3	
산업기사	미보유자1	미보유자1	95	34	2.8
		합계	95	34	2.8
총합계			620	100	6.2

(c)

Levl	Lev2	Lev3	Lev4	Career	FACI 카운트	AvgCareer
기사	미보유자2	미보유자2	미보유자2	26	4	6.5
			합계	26	4	6.5
			합계	26	4	6.5
기사	산업기사3	미보유자3	미보유자3	33	6	5.5
			합계	33	6	5.5
			합계	33	6	5.5
기술사	기사	미보유자7	미보유자7	169	20	8.5
			합계	169	20	8.5
			합계	169	20	8.5
	미보유자4	미보유자4	미보유자4	106	15	7.1
			합계	106	15	7.1
			합계	106	15	7.1
산업기사5	미보유자5	미보유자5	275	35	7.9	
		합계	79	7	11.3	
		합계	79	7	11.3	
산업기사5	합계	합계	112	14	8.0	
		합계	112	14	8.0	
		합계	112	14	8.0	
산업기사	미보유자1	미보유자1	미보유자1	466	56	8.3
			합계	95	34	2.8
총합계				95	34	2.8
				95	34	2.8
				620	100	6.2

(d)

그림 26. 데이터 큐브의 계층적 시각화  
Fig. 26 Hierarchical visualization of data cube

2.2 결과 및 분석

그림 26(a)에서는 응시자가 기사, 기술사, 산업기사 응시자로 나누어짐을 알 수 있고, 그림 26(b)에서는 그림 26(a)에서의 기사 응시자가 자격증 미보유 상태에서와 산업기사 자격증을 취득한 상태에서 응시하였음을 보여주고, 그림 26(a)에서의 기술사 응시자는 자격증 미보유 상태에서와 산업기사 자격증을 취득한 상태에서 그리고 기사 자격증을 취득한 상태에서 응시하였음을 보여주고, 그림 26(a)에서의 산업기사 응시자는 자격증 미보유 상태에서 응시하였음을 보여주고 있다.

그림 26(c)의 설명은 생략하고 그림 26(d)에서는 기사를 취득하고자 하는 응시 경로는 2개, 기술사를 취득하고자 하는 응시 경로는 4개, 산업기사를 취득하고자 하는 응시 경로는 1개가 있음을, 따라서 총 경로가 7개가 됨을 알 수 있을 것이다. 따라서 그림 17의 사례 스키마에서의 7가지 경로를 제대로 나타내었다고 결론 내릴 수 있다. 응시자의 자격 취득 이력을 전부 나타내는 이와 같은 경로 분석은 경시적 분석(longitudinal analysis)[27]이라고 할 수 있다.

OLAP에서의 탐색이란 계층별 수준의 조합별로 데이터가 집성되어 있는 데이터 큐브 속의 탐색을 말하는데, 이러한 탐색법은 크게 다음과 같은 둘로 구분할 수 있다[5].

(1) 외연 기반 탐색(extension-based navigation)

계층을 각 수준별로 전개시켜나가는 탐색, 즉 데이터 계층(data hierarchy)식 탐색을 말한다. 각 수준의 엔티티는 자신의 후손을 나타내는 식으로 반복적으로 전개된다. 그림 27(a)에서 보듯이 Quarter1을 탐색하고, 이 Quarter1을 Jan, Feb, Mar 로 전개시키면서 하는 탐색이 외연 기반 탐색이다.

(2) 내포 기반 탐색(intension-based navigation)

계층을 내포식, 즉 스키마 계층(schema hierarchy)식으로 나타낸다. 계층의 각 수준은 자신의 부모 수준의 하위 폴더(sub-folder)인 썸이다. Month 수준은 부모 수준인 Quarter의 하위 폴더라는 뜻이다. Quarter1을 탐색하자면 Quarter1, Quarter2 중 Quarter1을 지정하여 탐색하고, Quarter1을 전개시켜서 Jan, Feb, Mar를 탐색하는 것이 내포 기반 탐색이다.

집성 단일 기준 계층을 탐색하자면 단순 간단한 외연 기반 탐색법을 쓰면 충분하지만, 집성 다중 기준 계층이거나 일반화 계층과 같은 복잡한 추상 계층에서 탐색을 효과적으로 하자면 내포기반 탐색을 해야 한다고 Mansmann과 Scholl[5]은 주장하고 있으나, 이 주장은 일반화 계층의 탐색에서는 잘못 되었다고 본다. 왜냐하면 그림 22에서 내포 기반 탐색을 할 경우 전개시킬 하위 수준이 존재하지 않는 경우가 발생하기 때문이다. 따라서 일반화 계층의 탐색은 그림 26(a)와 같은 외연 기반 탐색이 더 적합하다고 볼 수 있다.

그림 26의 (a)에서는 최상위 수준의 탐색과 분석을 할 수 있고, (b)에서는 이 상위 수준의 바로 하위 수준을 탐색하고 분석할 수 있으며, 계속해서 더 하위 수준인 (c), (d) 별로 상세 탐색과 분석을 할 수 있다. 이와 같이 사용자에게 시각적으로 데이터 큐브 속의 외연 기반 탐색을 가능하게 해주는 시각화 기법을 개발한 것이 본 연구라고 할 수 있다.

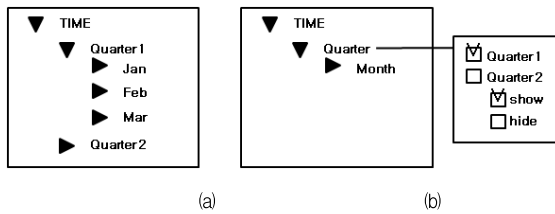


그림 27. 데이터 큐브에서의 탐색 : (a)외연 기반 탐색과 (b)내포 기반 탐색

Fig. 27. Navigating in Data Cube : (a)Extension-based vs. (b)Intension-based Approach

## VI. 결론

본 연구에서는 추상계층을 분류하는 출발점으로 집성과 일반화를 삼았다. 집성과 일반화는 그 성격이 완벽하게 다르다. 집성은 바닥에서 상부로 집계해 나가고, 일반화는 상부에서 하부로 분리해 나가는 것이기 때문에 이들 둘을 한 부류로 놓고 분류하는 것은 원칙적으로 잘못된 것이다. 따라서 본 연구에서는 집성과 일반화를 분리시켜 놓고 추상 계층의 분류를 시작한 것이다. 본 연구에서 제시한 그림 14와 같은 추상 계층의 분류는 모든 경우를 포괄하는 완벽한 분류법이라고는 볼 수 없으나 집성과 일반화를 분류의 출발점으로 삼은 원칙적인 분류법이 된다고 생각한다.

본 연구에서는 현 OLAP 시스템에서 구현 불가능한 스키마를 추이페포 개념을 활용하여 구현 가능한 스키마로 변환시키는 방안을 제시하였다. 이 방안을 따를 것 같으면 최종 사용자가 데이터 큐브 내에 있는 여러 데이터 뷰 중에서 바라보는 데이터 뷰를 시각적으로 탐색할 수 있음을 밝힌바 있다. 시각적으로 탐색할 수 있을 뿐 아니라 엔티티의 경시적 변화까지도 살펴 볼 수 있는, 다시 말하면 경시적 분석이 가능한 방안을 제시하였다고 본다.

본 연구에서 제시한 계층적 시각화 기법의 적용 분야는 아주 폭 넓은 것으로 생각한다. 기업 평가 및 심사 분야, 인사 평가 분야와 같은 경시적 분석이 필요한 분야에서는 바로 적용시킬 수 있는 기법이라고 생각한다.

본 연구에서는 그림 16의 ID 9 스키마를 ID 1 스키마로 변환시키는 것에 국한했으나, ID 1 스키마를 제외한 모든 스키마를 ID 1 스키마로 변환시키는 추후 연구가 필요하다고 본다.

## 참고문헌

[1] Surajit Chaudhuri, and Umeshwar Dayal, "An

Overview of Data Warehousing and OLAP Technology," ACM SIGMOD Record, Vol. 26, No. 1, pp. 65-74, 1997.

[2] DuckSung Lee, and InSoo Choi, "Design of an Inference Control Process in OLAP Data Cubes," Journal of The Korea Society of Computer and Information, Vol. 14, No. 5, pp. 183-193, May. 2009.

[3] DuckSung Lee, and InSoo Choi, "A Strategy for Inference Control of Official Statistics," Journal of The Korea Society of Computer and Information, Vol. 14, No. 11, pp. 200-211, November. 2009.

[4] Alfredo Cuzzocrea, Domenico Saccà, and Paolo Serafino, "A Hierarchy-Driven Compression Technique for Advanced OLAP Visualization of Multidimensional Data Cubes," Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery, LNCS 4081, pp. 106-119, 2006.

[5] Svetlana Mansmann, and Marc H. Scholl, "Exploring OLAP Aggregates with Hierarchical Visualization Techniques," Proceedings of the 22nd Annual ACM Symposium on Applied Computing, Multimedia & Visualization Track, pp. 1067-1073, 2007.

[6] Alfredo Cuzzocrea, and Svetlana Mansmann, "Models, Issues, and Techniques in OLAP Visualization," IGI Global, pp. 1-8, 2009.

[7] John Miles Smith, and Diane C.P.Smith, "Database Abstractions: Aggregation and Generalization," ACM Transactions on Database Systems, Vol. 2, No. 2, pp. 105-133, June. 1977.

[8] David M. Kroenke, "Database Processing," PEARSON, pp. 125-188, 2010.

[9] E. Malinowski, and E. Zimányi, "OLAP Hierarchies: A Conceptual Perspective," LNCS 3084, pp. 477-49, 2004.

[10] John Miles Smith, and Diane C.P.Smith, "Database Abstractions: Aggregation," ACM Transactions on Database Systems, Vol. 2, No. 2, pp. 105-133, June. 1977.

- [11] Aggregation, <http://krdic.naver.com>
- [12] EM018587, KRIVET, 2003.
- [13] EM015527, KRIVET, 2002.
- [14] EM015526, KRIVET, 2000.
- [15] SeHyeon Jang, HanJu Yu, and InSoo Choi, "Design of a Hierarchical Dimension of the Bill of Materials Type," Journal of the Korea Society of Computer and Information, Vol. 11, No. 4, pp. 244-250, Sept. 2006.
- [16] Carlos A. Hurtado, and Alberto O. Mendelzon. Jensen, "Reasoning about Summarizability in Heterogeneous Multidimensional Schemas", LNCS 1973, pp. 375 - 389, 2001.
- [17] David M. Kroenke, "Database Processing," PEARSON, pp. 182, 2010.
- [18] Svetlana Vinnik, and Florian Mansmann, "From Analysis to Interactive Exploration: Building Visual Hierarchies from OLAP Cubes," LNCS 3896, pp.496-514, 2006.
- [19] Svetlana Mansmann, and Marc H. Scholl, "Extending Visual OLAP for Handling Irregular Dimensional Hierarchies," LNCS 4081, pp. 95 - 105, 2006.
- [20] H. J. Lenz, and A. Shoshani, "Summarizability in OLAP and statistical data bases," In Proceedings of 9th International Conference on Scientific and Statistical Database Management, pp. 132 - 43, 1997.
- [21] Torben Bach Pedersen, Christian S. Jensen, and Curtis E. Dyreson, "A foundation for capturing and querying complex multidimensional data," Information Systems Vol. 26, pp. 383 - 423, 2001.
- [22] Torben Bach Pedersen, and Christian S. Jensen, "Multidimensional Data Modeling for Complex Data," A Time Center Technical Report, pp. 1 - 27, 1998.
- [23] Andreas Bauer, Wolfgang Hummer, and Wolfgang Lehner, "An Alternative Relational OLAP Modeling Approach," LNCS 1874, pp. 189 - 198, 2000.
- [24] Transitive closure, <http://en.wikipedia.org>
- [25] Xiaoyang Yo, "Transitive Closure of Binary Relation", <http://imps.mcmaster.ca>
- [26] Erik Thomsen, "OLAP Solutions," pp. 132-133, John Wiley & Sons, 2002.
- [27] Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware, "Applied Longitudinal Analysis," John Wiley & Sons, pp. 1-4, 2004.

## 저 자 소개



### 오 미 화

2002 : 성결대학교 정보통신학과 공학사  
 2004 : 숭실대학교 산업공학과 공학석사  
 현 재 : 숭실대학교 대학원산업·정보  
 시스템공학과 박사과정  
 관심분야 : MIS, DW, OLAP  
 Email : ohmihwa@empas.com



### 황 만 모

1987 : 호서대학교 정보통신공학과  
 공학사  
 1989 : 숭실대학교 산업공학과 공학석사  
 현 재 : 숭실대학교 대학원 산업·정보  
 시스템공학과 박사과정  
 관심분야 : MIS, DW, OLAP  
 Email : santa22@paran.com



### 최 정 우

1999 : 숭실대학교 산업·정보시스템  
 공학과 공학사  
 2002 : 숭실대학교  
 산업공학과 공학석사  
 현 재 : 숭실대학교 대학원 산업·정보  
 시스템공학과 박사과정  
 관심분야 : MIS, DW, OLAP  
 Email : fantom119@naver.com



### 최 인 수

1985 : 서울대학교 산업공학과 공학  
 박사  
 현재 : 숭실대학교 산업·정보시스템  
 공학과 교수  
 관심분야 : MIS, DW, OLAP  
 Email : ischoi@ssu.ac.kr