

경영정보학연구  
제7권 3호  
1997년 12월

## 자료 구성표를 이용한 데이터의 생성적 의미 표현 연구\*

이 춘 열\*\*

### A Representation of Data Semantics using Bill of Data

*Data semantics is an well recognized issue in areas of information systems research. It provides indispensable information for management of data. It describes what data mean, how they are created, where they can be applied to, to name a few. Because of these diverse nature of data semantics, it has been described from different perspectives of formalization. This article proposes to formalize data semantics by the processes that data are created or transformed. A scheme is proposed to describe the structure that data are created and transformed, which is called Bill of Data. Bill of Data is a directed graph, whose leaves are primary input data and whose internal nodes are output data objects produced from input data objects. Using Bill of Data, algorithms are developed to compare data semantics.*

---

\* 본 연구는 1996년도 교육비지원 학술진흥재단의 자유공모과제 학술연구조성비 (과제번호 01C0480)에 의하여 이루어졌음.

\*\* 국민대학교 정보관리학부

## I. 서 론

데이터의 생성적 의미란 데이터들이 만들어지는 과정을 통하여 구체화된 내용을 말한다. 예를 들면, 재고량을 창고에 저장된 수량으로 계산할 경우, 이는 실제 저장 수량을 나타낸다. 이에 반하여 이월 재고량에 입고량을 가산하고 판매량을 감산하여 계산할 경우 이는 장부상에 기재된 판매 가능한 수량을 나타낸다. 즉, 서로 다른 과정에 의하여 만들어진 재고량은 서로 다른 의미를 가진다. 이와 같이 데이터들이 만들어진 과정을 통하여 구체화된 뜻을 데이터의 생성적 의미라 한다.

데이터의 생성적 의미는 데이터의 사용과 직결된다. 예를 들면 재고량이 실제 저장 수량을 나타낸다고 가정하자. 만약 판매 담당자가 이를 근거로 하여 고객의 주문을 수주하여 납품을 약속하였다고 가정할 경우, 현재 실제로 판매 가능한 수량은 저장 수량과 차이가 날 수 있다. 저장 수량은 현재 창고에 저장되어 있으나 이미 판매된 수량, 즉 창고에 저장되어 있으나 판매는 불가능한 수량들을 포함할 수 있다. 이 경우, 판매 담당자가 사용한 현재 재고량은 정확한 숫자이다. 그러나 재고량 데이터를 생성적 의미와 다르게 적용함으로써 결과적으로 부정확한 납품 약속을 초래하게 된다.

데이터 활용 오류의 주 발생원이 이와 같이 데이터의 잘못된 적용이란 점을 고려할 때 [Redman, 1995; 1996, Liepens & Uppuluri, 1990], 데이터의 기록 관리에 추가하여 데이터의 의미에 대한 연구가 시급한 실정이라 할 수 있다. 그러나 데이터의 의미를 표현하기 위한 방법은 대부분 서술적인 형태를 지니는 시험적인 수준에 머무르고 있다 [Hammer & McLeod, 1981]. 이는 '의미'라는 단어가 데이터 관리와 연관하여 의미하는 바가 비정형적인 것과 무관하지 않다 [Kent 1978; 1979]. 이와 같이 서술적

인 형태로 표현된 데이터의 의미는 데이터 관리를 위한 활용의 측면에서 매우 제약적일 수밖에 없다.

본 연구는 데이터의 의미를 생성적 차원에서 구조화하는 방안을 제시한다. 즉, 특정 데이터를 생성하기 위하여 사용된 세부 데이터들을 자료 구성표란 표현 형태를 이용하여 제시함으로써 생성적 의미가 체계적으로 표현될 수 있도록 한다. 이하 제2장에서는 본 연구에서 포함하는 생성적 의미의 범위를 설정하고, 관련 연구들을 살펴본다. 제3장에서는 생성적 의미의 정형적 표현 방안을 제시하며, 제4장에서는 제시된 표현 기법을 이용한 데이터 의미의 비교 분석 기준을 제시한다. 끝으로 제5장에서는 본 연구의 의의와 추후 발전 방향을 결론으로 제시한다.

## II. 데이터의 생성적 의미

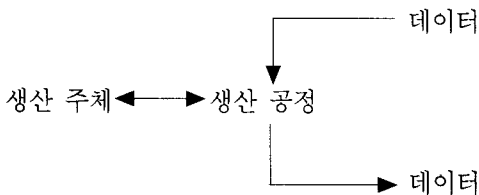
### 2.1 생성적 의미의 틀

데이터의 생성적 의미는 생성의 범위를 어떻게 설정하는가에 의하여 다르게 구체화될 수 있다. 즉, 해당 데이터가 무슨 원시 데이터들로부터 만들어 졌는지, 또는 누구에 의하여 어떠한 방법으로 만들어 졌는지 등의 데이터 생성과 연관된 사항들을 어느 관점에서 어떻게 포함하는가에 따라 다르게 설정된다. 본 연구에서는 생산 시스템을 근거로 하여 생성적 의미의 틀을 설정한다.

생산 시스템은 생산 주체와 객체 및 생산 공정으로 설명된다[이춘열, 1994]. 생산 주체란 생산 활동을 수행하는 작업자, 기계 또는 작업장 등을 나타낸다. 여기서 작업장은 개념적인 단위이며, 실제적으로는 조립 라인이나 기계군 등으로 구성된다. 생산 객체란 생산 활동의 대상으로서 제품을 의미한다. 제품에 대한 정보는 제품의 구성과 이를 구성하는 부품들에 대한 개별 정보를 포함한다. 최종적으로 생산 공정은

하위 부품이나 원료로부터 상위 부품이나 제품이 만들어지는 과정을 나타낸다.

생산 시스템의 프레임워크를 원용할 때, 데이터의 생성 또한 데이터의 생산 주체, 생산 객체 및 생산 공정들로서 설명된다. 즉, <그림 1>에 제시된 바와 같이 생산 객체는 데이터이며, 생산 공정은 데이터가 만들어지는 과정, 즉 생성 과정이며, 생산 주체는 생산 공정의 수행자 또는 컴퓨터 시스템이다.



<그림 1> 데이터의 생성 과정

데이터의 생성적 의미는 이들 생산 시스템의 구성 요소, 즉, 데이터의 생산 주체, 생산 객체 및 생산 공정에 의하여 구체화된다. 예를 들면, 생산 주체가 실사 담당자이며, 생산 공정이 재고 실사일 경우 재고량의 생성적 의미는 실제 재고량을 나타내는 것으로 구체화된다고 할 수 있다. 이에 반하여, 만약, 생산 주체가 재고 갱신 프로그램이며, 생산 공정이 이월 재고량에 입고량과 기 판매량을 가산 또는 감산하여 계산할 경우, 재고량은 장부상에 기재된 판매 가능 수량을 나타내는 것으로 의미가 구체화된다.

## 2.2 관련 연구

데이터의 생성적 의미는 데이터에 대한 명세와 이들이 생성된 프로세스에 대한 명세를 모두 포함하는 광의의 개념이라 할 수 있다. 이러한 데이터의 생성적 의미를 표현하기 위한 연구는 데이터 모델링을 중심으로 정보 시스템 개발 방법론, 데이터 처리 과정의 구조화 모형 등의 다양한 연구 분야에서 이루어져 왔다.

데이터의 생성 법칙들을 데이터 모델링에 포함시키기 위한 노력은 의미적 데이터 모형들에 의하여 시도되었다. 사실 데이터의 생성 의미는 의미적 데이터 모형들을 전통적 데이터 모형들과 구별시키는 주요 특성이라고 할 수 있다. 의미적 데이터 모형으로서 가장 대표적인 것이 Functional Data Model [Shipman, 1981]과 Semantic Data Model [Hammer & McLeod, 1981]이다. Functional Data Model은 데이터를 다른 데이터의 함수로서 정의함으로써 데이터의 생성 과정을 나타내며, Semantic Data Model은 생성 과정을 개체의 속성으로서 데이터 모델링에 포함시킨다.

정보 시스템 개발 방법론에서는 시스템 개발 노력의 일환으로 데이터의 구조적 연관 관계를 분석하고자 하는 노력이 시도되었다. 데이터 구조를 정보 시스템의 구현에 활용하기 위하여 Warnier[1974; 1981]은 Warnier Diagram을 제안하였으며, 이는 이후 Orr에 의하여 DSSD (Data Structured Systems Development)로 발전하게 된다.

Warnier-Orr Diagram은 논리 산출물 구조도를 이용하여 산출물, 즉 문서나 보고서의 구조를 표현한다. 논리 산출물 구조도를 기본으로 하여 새로운 데이터는 이전 데이터에 프로세스를 가함으로써 생성된다. 즉, 데이터 변환 프로세스는 논리 산출물 구조에 따라 데이터들을 생성하는 과정을 구조화하여 표현한 것이며, 이를 논리 프로세스 구조라 칭한다. 논리 프로세스 구조는 논리 산출물 구조의 순차적 구조, 반복 구조 및 선택 구조를 대응하는 프로그래밍 구조로 변형한 것이다. 즉, DSSD는 데이터 구조가 소프트웨어 개발의 기반을 형성한다는 가정에 기초하며, 이들 가정 하에서 정보 시스템(보다 구체적으로 소프트웨어)의 기본 구조를 설계한다 [Pressman, 1992].

데이터 모델링이나 정보 시스템 개발 방법론들은 궁극적으로 정보 시스템의 구현을 목표로 한다. 그러나 데이터의 처리 과정에 대한 명세

만을 통하여 데이터의 생성적 의미를 표현하는 방안을 생각할 수 있다.

대표적인 데이터 처리 과정의 구조화 모형으로서 AT&T의 FIP 기법을 들 수 있다. FIP 기법은 정보 처리 과정을 입력 정보 생산물에 정보 처리 함수 (FIP: functions of information processing)의 기능을 조작함으로써 출력 정보 생산물을 생산하는 과정으로 모형화 한다. 예를 들면, 다음의 데이터 생산 구조에서,

$$\text{Data\_set\_A FIP Data\_set\_B} = \text{Data\_set\_C}$$

Data\_set\_A와 Data\_set\_B는 입력 정보 생산물이며, Data\_set\_C는 출력 정보 생산물이다. 입력 정보 생산물들이 출력 정보 생산물로 변형되는 공정은 Associate, Filter, Prompt, Queue, Regulate, Transmit의 7가지 정보 처리 함수로서 설명된다 [Redman, 1996].

### Ⅲ. 생성적 의미의 구조화

#### 3.1 생성 유형

데이터의 생성적 의미를 정형화하기 위하여서는 먼저 데이터들의 생성 형태를 유형화하는 과정이 필요하다. 일반적으로 데이터를 생성 또는 처리한다고 할 때, 우리는 새로운 종류의 데이터를 만드는 것을 의미하기도 하며, 또는 기존에 존재하는 자료값을 갱신하는 것을 의미하기도 한다. 이러한 차이는 통상적으로 데이터라고 할 때, 데이터의 종류를 의미하기도 하며 또는 값을 의미하기도 함과 상응한다. 예를 들면 나이는 나이라는 데이터 종류를 의미하기도 하며, 특정인의 나이가 28세라는 값을 의미하기도 한다.

데이터의 생성이 내포하는 이와 같은 차이에 따라 본 연구는 개체적 관점과 자료값적 관점에서 데이터의 생성을 유형화한다. 개체적 관점은 새로운 종류의 데이터가 만들어지는가를 판

단하며, 자료값적 관점은 새로운 자료값이 만들어지는 가를 판단한다. 이들 관점에 따라 데이터의 생성은 <표 1>에 제시된 바와 같이 3가지 유형으로 구분된다.

<표 1> 데이터의 생성 유형<sup>1)</sup>

		자료값	
		생 성	미생성
개 체	생 성	항목 생성	개체 구성
	미생성	자료값 갱신	n.a.

<표 1>의 유형별로 데이터의 생성 형태를 살펴보면 다음과 같다.

##### 3.1.1 항목 생성

항목 생성은 기존의 데이터 또는 데이터들로부터 새로운 종류의 데이터를 창출하는 생성 형태이다. 예를 들면 재고량과 재주문점으로부터 주문량을 계산하는 것이 이에 해당한다. 항목 생성에 의하여 만들어지는 데이터 개체를 상위 데이터 개체라고 하며, 이를 위하여 사용된 개체를 하위 데이터 개체라고 한다. 앞에서 제시한 예에서, 주문량은 상위 데이터 개체이며, 재고량과 재주문점은 하위 데이터 개체이다. 그러나 이와 달리 하위 데이터 개체 없이 최초로 만들어지는 개체를 원시 데이터 개체라고 한다. 예를 들면 상품의 무게를 측정하여 기록할 경우, 이는 원시 데이터 개체이다.

항목 생성의 세부구조는 상위 데이터 개체의 유형에 따라 다른 형태를 가진다.

1) 데이터의 생성은 새로운 개체를 생성하거나 또는 자료값을 갱신하는 작업이어야 한다. 따라서 <표 1>에서 새로운 형태의 데이터 개체를 생성하지 않으며 동시에 자료값 또한 갱신하지 않는 경우, 즉 개체의 미 생성과 자료값의 미 생성에 해당하는 칸은 데이터의 생성으로 취급되지 않는다.

상위 데이터 개체가 기초 데이터일 경우, 이들은 하위 데이터 개체를 이용하거나 또는 직접 외부적으로 측정이나 판단 등의 과정을 거쳐서 만들어진다. 이에 반하여 상위 데이터 개체가 유추 데이터일 경우, 이들은 하위 데이터 개체로부터 계산이나 통계 처리 등의 방법에 의하여 생성된다.

본 연구에서 유추 데이터는 객관적 공식에 의하여 다른 데이터 개체들로부터 도출될 수 있는 데이터들을 말한다. 예를 들면, 일별 판매량으로부터 월 총 판매량을 계산하는 것이나, 사원별 판매량으로부터 부서 총 판매량을 계산하는 것이 이에 해당한다.

### 3.1.2 개체 구성

개체 구성은 데이터의 구성 관계를 통하여 새로운 데이터 개체를 만들어내는 과정이다. 즉, 개체 구성은 새로운 자료값의 생성 없이 이미 존재하는 데이터 개체들을 조합하거나 이로부터 일부를 추출하여 새로운 데이터 개체를 생성한다. 개체 구성의 세부 유형으로서는 특수화/일반화와 결합이 있으며, 이들은 데이터 모델링에서의 특수화/일반화 및 결합과 동일하다.

특수화는 하위 데이터 개체들 중 세부 유형들을 추출하여 상위 데이터 산출물을 생성하는 과정이다. 역으로, 일반화는 하위 데이터 개체들을 모두 포함하는 일반적인 유형의 데이터 개체를 생성하는 과정이다. 예를 들면, 우수거래 매출액은 매출액의 특수 형태이며, 매출액은 우수거래 매출액과 불량거래 매출액 등을 총칭하는 일반 형태이다.

결합은 하위 데이터 개체들을 결합하여 상위 데이터 개체를 구성하는 과정이다. 예를 들면 주문량과 공급 업체를 결합하여 주문서를 구성하는 것이 결합의 대표적인 형태이다.

### 3.2.3 자료값 갱신

자료값 갱신은 동일 데이터 항목의 값을 이전 자료값에 근거하여 수정하는 형태이다. 따라서 새로운 데이터 개체는 생성되지 않으며, 이미 존재하는 개체의 값이 갱신된다. 예를 들면 재고량에 입고량과 출고량을 가감하여 새로운 재고량을 계산하는 것이 이에 해당한다.

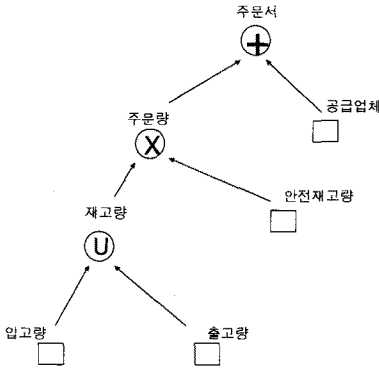
## 3.2 자료 구성표

자료 구성표는 데이터의 생성 유형을 활용하여 생성 과정을 구조화하는 표현 모형이다 [이춘열, 1996]. 예를 들어, 기간 중 입고량을 가산하고 출고량을 차감하여 재고량을 갱신한 후, 재고량과 안전 재고량을 비교하여 주문량을 계산하고, 공급업체를 선정하여 주문서를 발송한다고 가정하자. 제시된 시나리오에 포함된 데이터 개체들을 나열하면 다음과 같다.

입고량, 출고량, 재고량, 안전 재고량, 주문량, 공급업체, 주문서

이들 데이터 개체들에서 재고량은 입고량 및 출고량으로부터 갱신되며, 주문량은 재고량과 안전 재고량으로부터 생성되며, 주문서는 주문량 및 공급업체로 구성된다. 바꾸어 이야기하면, 재고량을 갱신하기 위하여서는 입고량과 출고량이 필요하며, 주문량을 생성하기 위하여서는 재고량과 안전 재고량이 필요하다. 또한 주문서를 구성하기 위하여서는 주문량 및 공급업체정보가 필요하다. 이와 같이 특정 데이터 개체를 생성하기 위하여 필요한 데이터 개체들의 구성을 도식화하여 표현한 것이 <그림 2>에 제시된 자료 구성표이다.

<그림 2>의 자료 구성표에서 상위 노드 (또는 부모 노드)는 만들어지는 데이터 개체를 나타내며, 하위 노드 (또는 자식 노드)는 이를 위하여 사용되는 데이터 개체들을 나타낸다.



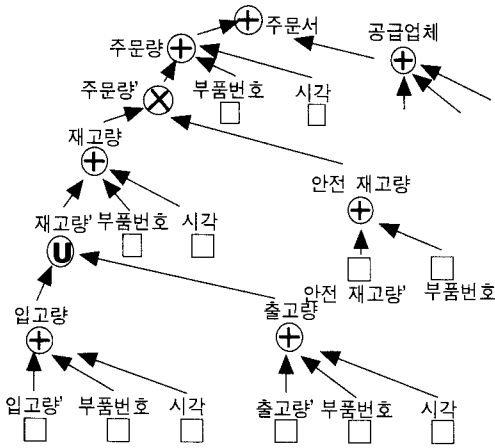
<그림 2> 주문서의 자료 구성표

자료 구성표에서 각 노드는 데이터의 생성 유형을 나타낸다. 즉, 재고량을 나타내는 노드 ⊕는 재고량의 생성 유형이 자료값 갱신임을 나타낸다. 또한 입고량과 출고량을 나타내는 노드 □는 이들이 원시 데이터 개체임을 나타낸다. 주문량을 나타내는 노드 ⊗는 주문량의 생성 유형이 항목 생성임을 나타낸다. 즉, 주문량은 안전 재고량과 재고량으로부터 생성되는 새로운 데이터 개체임을 나타낸다. 또한 주문서를 나타내는 노드 ⊕는 주문서의 생성 유형이 개체 구성임을 나타낸다. 즉, 주문서는 주문량과 공급 업체를 결합한 데이터 개체임을 나타낸다.

자료 구성표에서 데이터 개체들의 생성 구조는 이들을 연결하는 그래프로서 표현된다. 즉, <그림 2>에서 주문량의 생성 구조는 입고량, 출고량, 재고량 및 안전재고량과 주문량을 연결하는 그래프로서 표현되며, 주문량의 생성 구조는 입고량 및 출고량과 재고량을 연결하는 그래프로서 표현된다. 요약하면, 자료 구성표는 데이터의 생성 유형과 이를 생성하기 위한 하위 데이터 개체들을 나타낸다. 그러나 함수나 알고리즘과 같은 세부 과정은 표현되지 않는다.

자료 구성표에서 데이터 개체는 데이터 항목이나 레코드, 문서와 같은 임의의 형태를 모두 포함하는 개념이다. 따라서 이는 입고량이나 출고량 또는 재고량 등과 같이 데이터 항목일 수도 있으며, 공급 업체와 같이 레코드일 수도 있으며, 주문서와 같이 문서일 수도 있다. 이와 같은 데이터의 구조는 개체 구성 (⊕ 로 표현)에 의하여 구체적으로 표현된다. 즉, 주문서가 주문량과 공급 업체 정보로 구성된다는 것은 <그림 2>의 자료 구성표에서 주문서의 하위 데이터 개체들이 주문량과 공급 업체이며, 이의 생성 유형이 개체 구성임에 의하여 표현된다. 그러나 <그림 2>에 표현된 데이터 개체들 중에서 재고량 또한 {품목번호, 시각, 재고량}으로 세분화될 수 있으며, 주문량 또한 {품목번호, 시각, 주문량}으로 세분화된다. 따라서 <그림 2>의 자료 구성표는 이들 개체의 구성을 나타내기 위하여 <그림 2>와 같이 확장되어야 한다.

바꾸어 말하며, 모든 데이터 항목들은 이를 해석하기 위하여 필요한 데이터 항목들 (이들은 대부분 주키 및 외래키들임)을 포함한다. 즉, 재고량을 갱신하는 것은 특정 품목의 재고량을 갱신하는 것이며, 안전 재고량이 500개라는 것도 어느 품목의 안전 재고량이 500이라는 것이다. 그러나, 실제 데이터의 처리 과정에서 개별 항목과 이를 해석하기 위한 데이터 항목들을 분리하여 별도로 취급하는 것은 의미가 없다. 이는, 실제 데이터 처리 과정에서 재고량과 같은 데이터 개체를 칭할 때 품목 번호와 같은 주키를 포함하는 개념으로 취급하기 때문이다. 따라서 자료 구성표에서 개별 데이터 항목들은 이를 해석하기 위한 데이터 항목들과 같이 단일의 개체로 정의되며, <그림 2>는 <그림 2>로 함축하여 표현된다.



<그림 2> 주문서의 자료 구성표 (확장)

### 3.3 표현 모형

자료 구성표는 기본적으로 데이터 개체와 이들간의 생성 구조를 표현한다. 자료 구성표에서 데이터 개체들은 노드로 표현되며, 데이터들의 생성 구조는 노드를 연결하는 아크로 표현된다. 데이터의 생성 유형은 노드의 유형으로 표현된다. 요약하면 자료 구성표는 데이터 개체와 이들의 생성 구조를 나타내는 방향성 그래프이다.

**정의 1.** 자료 구성표는  $G = \langle V, E, \mu \rangle$ 로 구성되며, 데이터 개체  $v$ 의 자료 구성표는  $G(v)$ 로 표현한다.

- (1)  $V$ 는 그래프를 구성하는 노드이며, 데이터 개체를 나타낸다.
- (2)  $E$ 는 노드를 연결하는 아크를 나타내며,  $\langle V, E \rangle$ 는 방향성 그래프이다.
- (3)  $\mu$ 는 데이터 노드  $v$ 의 생성 유형을 나타내는 함수이다.
  - (a)  $\mu(v) = \oplus$ 는  $v$ 가 원시 데이터 개체임을 나타낸다.  $v$ 는 그래프의 잎 (leaf)이다.
  - (b)  $\mu(v) = \otimes$ 는  $v$ 가 항목 생성에 의한

데이터 개체임을 나타낸다.  $v$ 는 서로 상이한 자식 노드를 갖는다.

(c)  $\mu(v) = \oplus$ 는  $v$ 가 개체 구성에 의한 데이터 개체임을 나타낸다.  $v$ 는 서로 상이한 자식 노드를 갖는다.

(e)  $\mu(v) = \odot$ 는  $v$ 가 자료값 갱신에 의한 데이터 개체임을 나타낸다.  $v$ 는 서로 상이한 자식 노드를 갖는다.

이하에서,  $\mu(v)$ 를 데이터 개체  $v$ 의 생성 함수라고 한다. 즉, 생성함수는 데이터 개체의 유형이 원시 데이터 개체인지, 항목 생성에 의한 데이터 개체인지, 개체 구성에 의한 데이터 개체인지, 또는 자료값 갱신인지를 나타낸다.

$root(G)$ 와  $leaf(G)$ 는 각각 자료 구성표  $G$ 의 루트와 리프를 나타낸다. 데이터 개체  $v$ 와 이를 나타내는 자료 구성표  $G(v)$ 의 루트  $root(G(v))$  사이에는 다음의 관계가 성립한다.

$$v = root(G(v))$$

즉, 데이터 개체  $v$ 의 자료 구성표는  $v$ 를 루트로 한다.

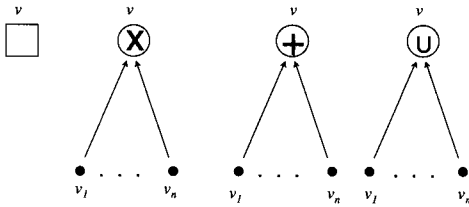
자료 구성표에서 각 노드는 데이터 개체의 생성 유형을 나타내며 하위 데이터 개체들과 아크로서 연결된다. 따라서, 이들을 정의 1과 같이 각각 분리하여 나타내기보다는 다음에 제시된 바와 같이 생성 함수에 하위 데이터 개체들을 같이 포함시킴으로써 보다 간략히 표현할 수 있다. 즉, <그림 3>에 예시된 자료 구성표의 각 노드들은 다음과 같이 표현된다.

- (1)  $\mu(v) = \square$ 는  $\mu(v) = \square$  임을 나타낸다.
- (2)  $\mu(v) = (\otimes, v_1, \dots, v_n)$ 는  $\mu(v) = \otimes$ 이며,  $v$ 의 자식 노드들이  $v_1, \dots, v_n$ 이며,  $v_1, \dots, v_n$ 과  $v$ 를  $n$ 개의 아크가 연

결함을 나타낸다.

(3)  $\mu(v) = (\oplus, v_1, \dots, v_n)$  는  $\mu(v) = \oplus$  이며,  $v$  의 자식 노드들이  $v_1, \dots, v_n$  이며,  $v_1, \dots, v_n$  과  $v$  를  $n$  개의 아크가 연결함을 나타낸다.

(4)  $\mu(v) = (\ominus, v_1, \dots, v_n)$  는  $\mu(v) = \ominus$  이며,  $v$  의 자식 노드들이  $v_1, \dots, v_n$  이며,  $v_1, \dots, v_n$  과  $v$  를  $n$  개의 아크가 연결함을 나타낸다.



<그림 3> 자료 구성표의 노드유형

이상에서 제시된 기본 요소들에 추가하여 자료 구성표는 여러 가지 관리 정보를 추가로 나타낼 수 있다. 예를 들면, 공정의 수행시점 ( $\tau(v)$ ), 공정의 수행 주체를 나타내는 ( $\rho(v)$ ) 등을 포함할 수 있다.

그러나 본 연구에서는 이들 관리 정보의 활용에 대하여서는 세부적으로 다루지 않는다.

$\tau(v) = (t = t_0) \mid (t = t + \Delta t) \mid \text{cond}$   
 $\text{cond} = v\theta a \mid v\theta a \mid \text{cond AND cond} \mid \text{cond OR cond} \mid \text{NOT cond}$

$\theta$  는 비교 연산자 ( $=, \neq, >, \geq, <, \leq$ )

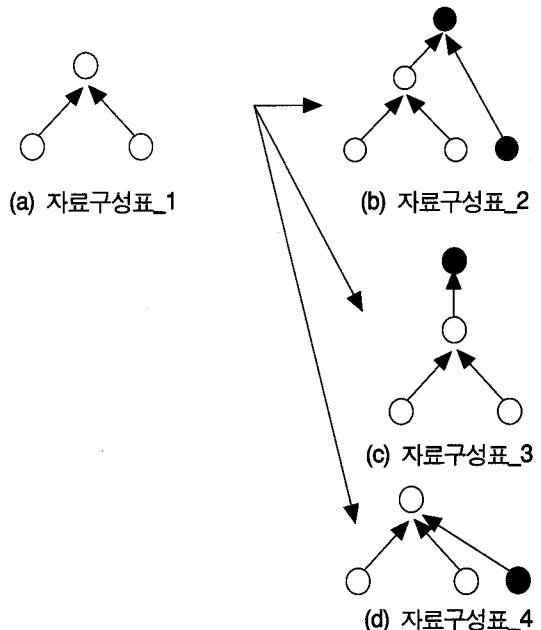
$\rho(v) = \text{person} \mid \text{program} \mid \text{system}$

이상에서 제시된 자료 구성표의 상호 비교를 위한 기준이 자료 구성표의 확장이다.

**정의 2.** 자료 구성표  $G' = \langle V', E', \mu \rangle$  가 다음의 조건을 만족할 경우 자료 구성표  $G = \langle V, E, \mu \rangle$  의 확장이라고 한다.

- (1)  $V \subseteq V'$
- (2)  $E \subseteq E'$
- (3)  $(v_1, v_2) \in E' - E$  일 경우  $v_2 = r(G')$  또는  $v_2 = r(G)$  이다. 즉, 새로운 아크는 기존 노드나 새로운노드로부터 새로운 루트를 연결하는 아크이거나, 새로운 노드로부터 기존 루트를 연결하는 아크이다.

예를 들면 <그림 4>의 자료구성표\_2, 자료구성표\_3 및 자료구성표\_4는 자료구성표\_1의 확장이다. 확장은 다시 루트 확장과 리프 확장으로 구분된다. 루트 확장은 자료구성표\_3에 예시된 바와 같이 리프의 추가 없이 루트가 확장되는 경우이며; 리프 확장은 자료구성표\_4에 예시된 바와 같이 루트의 생성 없이 리프가 첨가되는 경우이다. 자료구성표\_2와 같이 이들 두 경우가 모두 발생하는 것을 완전 확장이라고 한다. 리프 확장 또는 완전 확장에서  $v$ 를 추가된 리프 노드라 할 때, 자료 구성표  $G'$ 은 자료 구성표  $G$ 의 데이터 개체  $v$ 에 대한 확장이라고 한다.



<그림 4> 자료 구성표의 확장 유형 (I)

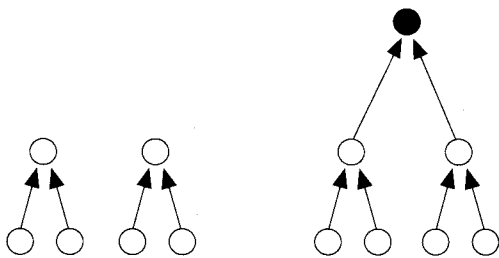


이상에서 우리는 단일 노드의 확장을 제시하였다. 그러나 <그림 5>에서 예시되는 바와 같이 단일 노드 대신 자료 구성표를 확장 노드로 포함할 수 있다. 이와 같이 확장 개념을 자료 구성표에 대하여 적용한 것이 정의 3이다.

**정의 3.** 자료 구성표  $G' = \langle V', E', \mu' \rangle$ 가 다음의 조건을 만족할 경우 자료 구성표  $G_1 = \langle V_1, E_1, \mu_1 \rangle$ 의  $G_2 = \langle V_2, E_2, \mu_2 \rangle$ 에 대한 (또는  $G_2$ 의  $G_1$ 에 대한) 확장이라고 한다.

- (1)  $V_1 \cup V_2 \subseteq V'$
- (2)  $E_1 \cup E_2 \subseteq E'$
- (3)  $(v_1, v_2) \in E' - (E_1 \cup E_2)$  일 경우  $v_1 = r(G')$ 이며,  $v_2 = r(G_1), r(G_2)$  이다. 즉, 새로운 아크는 기존 루트로부터 새로운 루트를 연결하는 아크이다.

정의 3에서 제시된 확장은 자료 구성표들 사이의 루트 확장이라고 할 수 있다. 즉, 자료 구성표  $G'$ 은 자료 구성표  $G_1$ 의 자료 구성표  $G_2$ (또는  $G_2$ 의  $G_1$ )에 대한 루트 확장이다.



<그림 5> 자료 구성표의 확장 유형 (III)

확장의 개념을 적용함으로써 자료구성표는 반복적으로 확장 생성될 수 있다. 즉,  $G_2$ 가  $G_1$ 의 확장이고,  $G_3$ 이  $G_2$ 의 확장이면,  $G_3$ 은 또한  $G_1$ 의 확장이다. 자료 구성표  $G$ 의 확장 그래프들 중 리프, 즉, 원시 데이터 개체가 동일한 자료 구성표들을 동일 확장 자료 구성표들이라고 한다.

**정의 4.** 자료 구성표의 집합  $\underline{G}$  에 포함된 임의의 두 자료 구성표  $G_1, G_2$ 에 대하여  $leaf(G_1) = leaf(G_2)$ 일 경우  $\underline{G}$  을 동일 확장 자료구성표 군이라고 한다.

임의의 두 자료 구성표  $G_1, G_2$ 가 동일 확장 자료 구성표 군에 속할 경우,  $G_1$ 의  $G_2$ 에 대한 (또는  $G_2$ 의  $G_1$ 에 대한) 확장 자료 구성표 또한  $\underline{G}$  에 속한다.

## IV. 생성적 의미 분석

### 4.1 생성적 의미의 기준

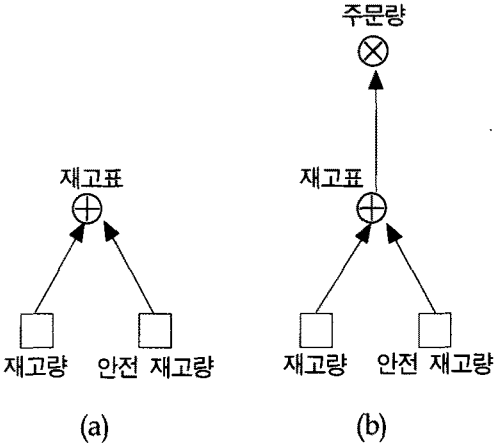
자료 구성표는 데이터의 생성적 의미를 판단하기 위한 기준을 제시한다. 즉, 자료 구성표의 비교 기준을 이용함으로써 우리가 정성적으로 이야기하는 데이터들 사이의 동질성, 유사성 또는 연관성 등의 개념들은 다음과 같이 객관적으로 조작화된다.

**정의 5.** 생성 의미적 파생: 데이터 개체  $d$ 와  $d'$ 의 자료 구성표를 각각  $G = \langle V, E, \mu \rangle$  와  $G' = \langle V', E', \mu' \rangle$  이라고 할 때, 다음의 조건이 만족될 경우 데이터 개체  $d'$ 는 생성 의미적으로 데이터 개체  $d$ 로부터 파생되었다고 한다.

- (1)  $G'$ 과  $G$ 는 동일 확장 자료 구성표 군에 속한다. 즉,  $G', G \in \underline{G}$  이다.
- (2)  $v \in V' - V$  일 경우  $\mu(v) \in \{\otimes, \odot\}$ 이다. 즉, 새로이 추가된 노드는 항목 생성 또는 자료값 갱신 노드이다.

부연하면, 특정 데이터 개체가 타 데이터 개체와 동일한 원시 데이터 개체들을 사용하여 생성될 경우, 해당 데이터 개체를 파생 데이터 개체라고 한다. 즉, 파생 데이터 개체는 동일한 원시 데이터 개체들로부터 유도된다.

예를 들면, 주문량은 생성 의미적으로 재고표의 파생 데이터 개체이다. 즉, 주문량은 재고표가 나타내는 정보로부터 생산된 데이터 개체이다.



<그림 6> 생성 의미적 파생 자료 구성표

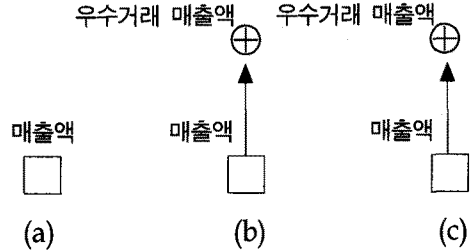
파생된 데이터 개체들 중에서 다만 형태적 구성이 변화됨으로써 기존의 데이터 개체와 동일한 정보를 전달할 경우 이를 생성 의미적으로 동등하다고 한다.

**정의 6. 생성 의미적 동등:** 두 데이터 개체  $d$ 와  $d'$ 의 자료 구성 그래프를 각각  $G = \langle V, E, \mu \rangle$ 와  $G' = \langle V', E', \mu' \rangle$  이라고 할 때, 이들 사이에 다음의 조건이 만족될 경우 두 데이터 개체  $d$ 와  $d'$ 는 생성 의미적으로 동등하다고 한다.

- (1)  $G'$ 과  $G$ 는 동일 확장 데이터 군에 속한다. 즉  $G \in \underline{G}$  이며  $G' \in \underline{G}$  이다.
- (2) 모든  $v' \in V' - V, v \in V - V'$ 에 대하여  $\mu(v), \mu(v') = \oplus$ 이다. 즉 서로 상이한 노드는 개체 구성에 의한 노드이다.

부연하면, 원시 데이터 개체들이 동일하며, 오직 개체 구성에 의한 데이터 개체들만이 추가될 경우 이들 데이터 개체들은 생성 의미적

으로 동등하다고 한다. 예를 들면, 불량 거래 매출액과 우수 거래 매출액은 생성 의미적으로 동등하다.



<그림 7> 생성 의미적 동등 자료 구성표

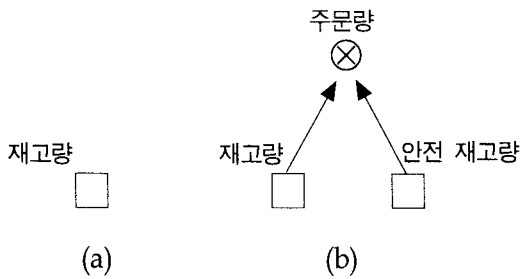
여기서 생성 의미적으로 동등하다는 것은 데이터 개체들이 전달하는 정보의 양이 서로 동일하다는 것을 시사한다. 즉, 매출액을 기준으로 하여 볼 때 (거래별 상품 매출액을 나타내는 데이터가 존재한다는 가정 하에서), 불량 거래 매출액이나 우수 거래 매출액이 전달하는 정보의 양은 서로 동일하다. 다만 개체의 구성이 변형된 것이라고 할 수 있다.

생성 의미적 파생이나 동등과 달리 새로운 원시 데이터 개체가 추가되어 생성된 데이터 개체를 생성 의미적으로 확장이라고 한다.

**정의 7. 생성 의미적 확장:** 특정 데이터 개체가 타 데이터 개체와 비교하여 추가적인 원시 데이터 개체를 포함할 경우, 해당 데이터 개체는 타 데이터 개체 보다 의미적으로 확장된 데이터라고 한다. 즉, 데이터 개체  $d$ 와  $d'$ 의 자료 구성 그래프를 각각  $G = \langle V, E, \mu \rangle$ 와  $G' = \langle V', E', \mu' \rangle$  이라고 할 때,  $G'$ 가  $G$ 의 완전 확장 또는 리프 확장일 경우 데이터 개체  $d'$ 는 데이터 개체  $d$ 의 생성 의미적 확장이다.

부연하면, 확장 데이터 개체들은 추가적인 원시 데이터 개체들을 사용한다.

예를 들면, <그림 8>의 주문량은 생성 의미적으로 재고량의 확장이다. 생성 의미적 확장은 새로운 데이터 개체가 추가됨으로써 새로운 정보가 추가되었음을 의미한다. 즉, 주문량은 재고량이라는 데이터가 나타내는 정보에 안전 재고량이 나타내는 정보가 추가됨으로써 생성된 데이터 개체이다.



<그림 8> 생성 의미적 확장 자료 구성표

이상에서 제시된 정의를 중심으로 자료 구성표의 비교 기준과 생성적 의미의 관계를 요약하면 <표 2>와 같다.

첫째, 두 그래프 사이에 확장 관계가 존재하면

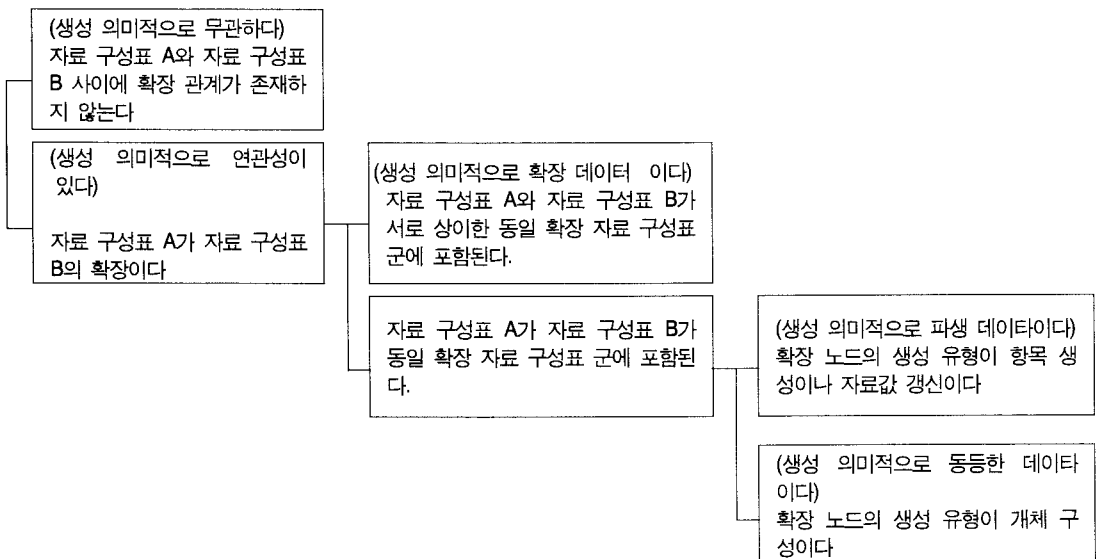
생성 의미적으로 연관성을 가진다고 하며, 확장 관계가 존재하지 않으면 연관성이 없다고 한다. 즉, 자료 구성표의 연관성은 생성적 의미의 판단 기준이 된다.

확장 관계는 다시 확장의 종류에 따라 리프 노드가 추가되는 리프 확장 또는 완전 확장과 리프 노드가 추가되지 않는 루트 확장으로 분류된다. 전자는 생성 의미적으로 추가적인 내용을 포함하는 데이터의 생성을 나타내며, 후자는 추가적인 내용을 포함하지 않는 데이터의 생성을 나타낸다. 추가적인 내용을 포함하지 않는 데이터들은 생성 유형에 따라 항목 생성과 개체 구성으로 구분되며, 후자는 개체 구성의 형태만 변경된 동등한 데이터 개체를 나타낸다. 이에 반하여 전자는 기존 데이터들로부터 산출되는 새로운 파생 데이터를 나타낸다.

## 4.2 생성적 의미의 관계

자료 구성표에 근거하여 정의된 데이터의 생성적 의미는 데이터들 사이의 의미적 관계를

<표 2> 데이터의 생성적 의미 비교표



비교 판단할 수 있는 객관적인 기준을 제공한다.

이들 기준을 적용함으로써 데이터 개체들 사이의 의미적 관계를 요약하면 다음과 같다.

**의미적 관계 1.** 생성 의미적으로 상이한 데이터들로부터 산출된 데이터 개체들은 서로 상이하다.

**<설명>** 상이한 의미의 데이터 개체  $d_1$ 과  $d_2$ 의 자료 구성표를  $G_1, G_2$ 라고 하면 정의에 의하여 이들은 서로 확장 관계가 성립하지 않는다. 즉, 이들의 동일 확장 자료 구성표 군을  $G_1, G_2$ 이라고 하면  $G_1 \notin G_2$ 이며,  $G_2 \notin G_1$ 이다. 만약  $d_1$ 으로부터 산출된 데이터 개체를  $d_1'$ 이라고 하고  $d_2$ 로부터 산출된 데이터 개체를  $d_2'$ 이라고 하면 이들의 자료 구성표  $G_1', G_2'$ 는 각각  $G_1' \in G_1$ 이며,  $G_2' \in G_2$ 이다. 그러나  $G_1, G_2$ 의 정의에 의하여 만약  $G_1' \in G_1$ 이면  $G_1' \notin G_2$ 이다. 또한  $G_2' \in G_2$ 이면  $G_2' \notin G_1$ 이다. 즉,  $d_1'$ 과  $d_2'$ 는 상이한 동일 자료 구성표 군에 소속되며, 따라서 생성 의미적으로 서로 상이하다.

**의미적 관계 2.** 데이터의 생성적 의미는 확장에 대하여 추이적이다.

**<설명>** 데이터 개체  $d$ 의 자료 구성표를  $G$ 라고 하고, 자료 구성표  $G$ 의 데이터 개체  $v$ 에 대한 확장을  $G'$ 이라고 하자. 또한  $G'$ 의 데이터 개체  $v'$ 에 대한 확장 자료 구성표를  $G''$ 이라고 하면  $G''$ 는 자료 구성표  $G$ 의 데이터 개체  $v$ 와  $v'$ 에 대한 확장이다. 따라서 데이터의 생성적 의미는 확장에 대하여 추이적이다.

동일한 방법으로 데이터의 생성적 의미는 파생과 동등에 대하여도 추이적임을 알 수 있다.

**의미적 관계 3.** 데이터의 중간 생성 과정에서 개체 구성의 변화는 최종 데이터 개체의 생

성적 의미를 변화시키지 않는다.

**<설명>** 데이터 개체  $d$ 의 자료 구성표를  $G$ 라 할 때, 이로부터 개체 구성에 의하여 산출된 데이터 개체들의 자료 구성표들을  $G_1, \dots, G_n$ 이라고 하자. 그러면  $G$ 와  $G_1, \dots, G_n$ 은 동일 확장 자료 구성표 군  $G$ 에 포함된다. 여기서  $G$ 로부터의 확장을  $G'$ 이라고 하고, 임의의  $G_i$ 로부터의 확장을  $G_i'$ 이라고 할 때, 이들 모두  $G$ 에 포함된다. 즉,  $G'$ 과  $G_i'$ 은 동일 확장 자료 구성표 군에 속하며, 모든 확장 노드들은 개체 구성에 의한 데이터 개체들이다. 따라서 이들은 생성 의미적으로 동일하다. 이상에서 제시된 생성적 의미들 사이의 관계는 일반적으로 생각하였던 데이터의 의미와 일치한다.

## V. 결 론

본 연구의 가장 큰 의의는 데이터 개체의 생성적 의미를 정형화하는 프레임워크를 제시한다는 점이다. 즉, 자료 구성표를 이용하여 데이터의 생성 관계를 구조적으로 표현하는 방안을 제시함으로써 데이터의 의미를 객관적인 관점에서 조명할 수 있도록 하였다. 이러한 관점의 확장은 객관적 판단 기준이 존재하지 않는 것으로 취급되어왔던 데이터의 의미를 데이터 관리의 타 분야와 동일하게 객관적으로 판단할 수 있는 기준을 제시하였다는 점에서 의의를 갖는다.

더불어 자료 구성표는 데이터의 관리 도구로서 여러 가지 장점을 지닌다.

첫째, 무엇보다도 데이터들간의 구성 관계를 표현할 수 있다.

이를 이용하여 데이터의 생성적 의미를 쉽게 관리하고 파악할 수 있으며, 데이터의 생성적 의미는 올바른 데이터 활용을 위한 근거 정보로서 활용할 수 있다.

둘째, 데이터의 구성 관계는 데이터들이 결합되고 변형되는 과정을 보여줌으로써 데이터의 호

를 파악할 수 있게 한다. 이와 같은 데이터의 변형 과정에 대한 정보는 데이터의 활용 실태 분석을 위한 기초 자료로서 중요한 의미를 지닌다. 본 연구에서는 구체적으로 분석되지 않았으나 이는 데이터의 이용을 분석하고 식별하며, 데이터 가공과 처리를 위한 소요 시간과 자원 등을

분석함에 근거자료로서 활용될 수 있다. 그러나 본 연구는 개념 연구로서 개념적인 예시를 중심으로 이루어졌다. 따라서, 이의 유용성에 대한 실증적 검증은 본 연구에 포함되지 않았다. 이를 위한 구체적인 실험은 본 연구의 후속 과제로서 수행되어야 할 것이다.

## 〈참 고 문 헌〉

- Hammer, M. and McLeod, D., "Database Description with SDM: A Semantic Database Model," *ACM Transactions on Database Systems*, Vol.6, No.3, 1981, pp. 351-386.
- Kent, W., "Limitations of Record-Based Information Models," *ACM Transactions of the Database Systems*, Vol.4, No.1, 1979, pp. 107-131.
- Kent, W., *Data and Reality*, North Holland Publishing Company, 1978, New York.
- Liepens, G.E. and Uppuluri, V.R.R., *Data Quality Control: Theory and Pragmatics*, Marcel Dekker, 1990.
- Pressman, R.S., *Software Engineering: A Practitioner's Approach*, 3rd ed., McGraw-Hill, Inc., 1992.
- Redman, T.C., "Improve Data Quality for Competitive Advantage," *Sloan Management Review*, 1995 Winter, pp. 99-107.
- Redman, T.C., *Data Quality for the Information Age*, Artech House, 1996.
- Shipman D.W., "The Functional Data Model and the Data Language DAPLEX," *ACM Transactions on Database Systems*, Vol.6, No.1, 1981, pp. 141-173.
- Sowa, J.F., *Conceptual Structures: Information Processing in Mind and Machine*, Addison Wesley Publishing Company, 1984.
- Warnier, J.D., *Logical Construction of Programs*, Van Nostrand Reinhold, 1974.
- Warnier, J.D., *Logical Construction of Systems*, Van Nostrand Reinhold, 1981.
- 이춘열, "생산 관리를 위한 표준 데이터 모형 개발," *국민대학교 경상 논총*, 제16집, 1994년 2월, pp. 537-554.
- 이춘열, "데이터 제조 공정 모형," '96 데이터베이스 심포지엄 및 학술대회 논문지, 1996년 11월, pp. 311-326.

## ◆ 저자소개 ◆



이 춘 열(Lee, Choon Yeul)

저자는 서울대학교에서 수학하였으며, 미국 미시간 대학교에서 Computer and Information Systems를 전공하여 경영학 박사학위를 취득하였다. 이후 한국통신 소프트웨어연구소에 근무하였으며, 현재 국민대학교 정보관리학부에 재직하고 있다. 주요 관심분야는 데이터베이스 응용 시스템, 자료 검색, 정보 공학 등이며, 현재 데이터웨어하우징 도구 및 데이터 관리 기법 등에 대한 연구를 수행 중이다.