

# 정규화된 논리적 데이터 모델의 생성을 위한 사건 기반 개체-관계 모델링 방법론

유재건<sup>†</sup>

한성대학교 산업경영공학과

## An Event-Driven Entity-Relationship Modeling Method for Creating a Normalized Logical Data Model

Jaegun Yoo

Department of Industrial and Management Engineering, Hansung University, Seoul 136-792, Korea

A new method for creating a logical data model is proposed. The logical data model developed by the method defines table, primary key, foreign key, and fields. The framework of the logical data model is constructed by modeling the relationships between events and their related entity types. The proposed method consists of a series of objective and quantitative decisions such as maximum cardinality of relationships and functional dependency between the primary key and attributes. Even beginners to database design can use the methodology as long as they understand such basic concepts about relational databases as primary key, foreign key, relationship cardinality, parent-child relationship, and functional dependency. The simple and systematic approach minimizes decision errors made by a database designer. In practical database design the method creates a logical data model in Boyce-Codd normal form unless the user of the method makes a critical decision error, which is very unlikely.

**Keywords:** Relational Database, Data Modeling, Logical Data Model, Entity-Relationship

### 1. 서론

개념적 데이터 모델인 개체-관계 다이어그램을 완성하고, 이를 논리적 데이터 모델로 변환하는 것이 일반적인 관계형 데이터베이스 설계 과정이다. 본 논문에서는 개념적 설계를 생략하고 직접 논리적 데이터 모델을 생성하는 개체-관계 모델링 방법론을 제시한다. 생성된 논리적 데이터 모델은 테이블, 기본키 필드, 외래키 필드, 일반 필드의 정의를 포함한다. 현실적인 데이터베이스 설계에 적용할 경우 정의된 테이블은 Boyce-Codd 정규형이 된다.

논리적 데이터 모델링은 데이터베이스 개념에 대한 이해만이 아니라 데이터베이스 응용 분야에 대한 인간의 지식과 경험이 요구되는 매우 복잡한 사고 과정이다(Ra, 2009). 개체-관

계 모델은 논리적 데이터 모델링 도구로서 오랫동안 사용되어 왔다(Chen, 1976). 최근 객체지향 방법론들이 각광을 받고 있으나 개체-관계 데이터 모델과 객체지향 데이터 모델의 설계 품질을 비교한 연구에 의하면 개체-관계 모델이 정확성, 소요 시간, 설계자의 선호도 측면에서 더 우수함이 실험적으로 입증되었다(Shoval *et al.*, 1997). 개체-관계 다이어그램은 이러한 사고 과정의 결과를 다이어그램으로 나타낸 것이다. 개체-관계 다이어그램은 Information Engineering, Barker Notation, IDEF1X, UML 등 다양한 표기 방법을 가지고 있으나, 결국 개체, 관계, 속성을 나타낸다는 점에서 크게 다르지 않다. 개체-관계 다이어그램은 개체, 관계, 속성을 표현하는 도구일 뿐이며, 그들을 도출하고 식별하는 사고 과정은 온전히 설계자의 지식과 경험에 의존한다. 이 사고 과정을 지원하는 방법론은 어떠한 참고

본 연구는 2010년도 한성대학교 교내연구비 지원으로 수행되었음.

<sup>†</sup> 연락처 : 유재건 교수, 136-792 서울시 성북구 삼선동2가 389번지 한성대학교 산업경영공학과, Tel : 02-760-4393, Fax : 02-760-4490,

E-mail : jgyoo@hansung.ac.kr

2011년 6월 27일 접수; 2011년 8월 17일 게재 확정.

문헌에서도 발견되지 않는다. 개체, 관계, 속성을 도출하고 식별하는 구체적인 방법론이 없기 때문에 초보 설계자들은 개체, 관계, 속성에 대한 개념만 습득한 상태에서 모델링 문제를 접하게 되고, 숙련되기 위해서는 많은 시행착오를 겪게 된다.

본 논문이 제시하는 모델링 방법론은 복잡한 사고 과정을 단순하고 구체적인 판단 과정으로 분해하고 체계화함으로써 설계자의 판단 오류를 최소화한다. 특히 초보 설계자의 경우 매우 난해하게 느껴질 수 있는 사고 과정을 안내함으로써 개체, 관계, 속성의 개념을 좀 더 쉽게 이해하는 데에도 도움이 된다. 이는 저자가 대학교 2학년을 대상으로 한 강의에서 확인 할 수 있었다.

본 논문에서 제시하는 방법론은 개체 유형(entity type)과 관계(relationship)를 정의한 후에 속성을 정의하는 하향식 방법론이다. <Figure 1>에 나타난 바와 같이 방법론은 크게 테이블, 기본키 및 외래키 필드를 결정하는 과정과 일반 필드를 결정하는 과정으로 나누어진다.

명사 및 동사 개체 유형의 정의와 관리 대상(데이터베이스에 기록되고 유지되어야 할 대상) 사건(event)의 관계 모델링의 결과로 테이블, 기본키 및 외래키 필드를 결정한다. 이는 논리적 데이터 모델의 틀 또는 뼈대에 해당한다. 이 과정에서 사용되는 주된 판단 기준은 관계의 최대 사상수이다. 다음으로 관리 대상 속성의 수집과 주인 개체 유형(속성을 소유하는 개체 유형)의 결정 과정의 결과로 테이블의 일반 필드를 결정한다. 이 과정에서는 기본키와 속성의 함수 종속 여부가 주된 판단 기준이 된다. 이렇게 완성된 논리적 데이터 모델은 개체 유형의 통합과 분리, 기본키의 확정, null 값의 최소화를 위한 관계의 개체화, 데이터베이스의 성능 향상을 위한 반정규화 등의 세부 조정이 필요하지만 이 과정은 본 논문의 범위를 벗어난다.

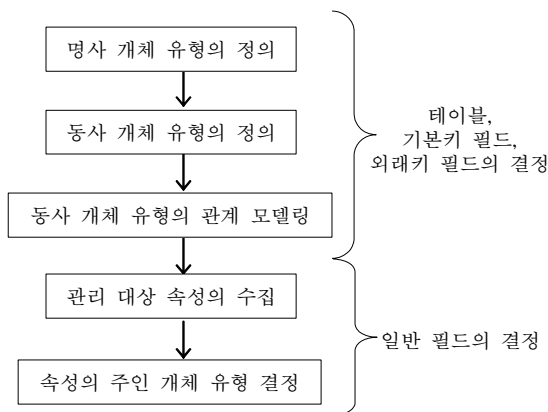


Figure 1. Major steps of the method

관계형 데이터베이스의 개체-관계(entity-relationship) 모델에서 개체(entity)는 하나의 독립된 존재로서 식별(identification)될 수 있는 사물로 정의된다. 개체들의 집합을 개체 유형이라고 한다. 개체 유형은 <Table 1>과 같이 분류될 수 있다(Lee, 2003).

Table 1. Classification of entity types

| 개체 분류  | 개체 유형의 예                           |
|--------|------------------------------------|
| 사람     | 학생, 교원, 직원, 회원, 주민, 의사, 환자 등       |
| 장소     | 강의실, 창고, 생산라인, 행정구역, 병실, 하천, 선거구 등 |
| 조직     | 학과, 부서, 거래처, 대리점, 병원, 정부기관 등       |
| 유형의 재화 | 기자재, 상품, 원자재, 부품, 약, 장비 등          |
| 무형의 재화 | 계좌, 특허권, 강좌, 주식 등                  |
| 사건     | 강의, 수강, 입금, 대여, 주문, 전입, 진료, 처방 등   |

<Table 1>에서 사람, 장소, 조직, 유무형의 재화는 명사이고, 사건은 동사이다. 또한 후자는 하나 이상의 개체 유형이 연관되어 발생하므로, 개체 유형 간의 관계가 고려되어야 한다는 점에서 전자와 구별된다. 앞으로 전자를 ‘명사’ 개체 유형이라고 하고, 후자를 ‘동사’ 개체 유형이라고 한다.

사건은 사람, 장소, 조직, 유무형의 재화, 또는 다른 사건이 연관되어 발생한다. 이와 같이 사건은 하나 이상의 개체 유형이 연관되어 발생하므로 개체 유형 간의 관계가 모델링되어야 한다. 사건의 집합을 동사 개체 유형 V라고 하고, 관련된 개체 유형을 E1, E2, ..., En이라고 하자. 사건과 관련된 개체 유형의 개수 n을 관계의 차수(degree)라고 한다. 차수에 따라 일진(unary), 이진(binary), 삼진(ternary), ..., n-ary 관계라고 한다. 동사 개체 유형 V를 정의하기 위해서는 관련 개체 유형 E<sub>j</sub>(j = 1, 2, ..., n)이 먼저 정의되어야 한다. 명사 개체 유형은 동사 개체 유형의 관련 개체 유형으로만 존재하므로 <Figure 1>에 나타난 바와 같이 명사 개체 유형의 정의가 동사 개체 유형의 정의에 선행된다.

사건과 관련된 개체 유형 간의 모든 관계는 V와 E<sub>j</sub>(j = 1, 2, ..., n)의 관계에 의해 결정된다. 따라서 V와 E<sub>j</sub>(j = 1, 2, ..., n)의 관계를 개별적으로 모델링한다. 여기서 관계의 개별적 모델링은 최대 사상수에 따라 부모와 자식 개체 유형을 결정하고, 이에 따라 자식 테이블에 부모 테이블의 기본키를 외래키로 추가하는 것을 말한다.

하나의 관리 대상 사건에 대한 데이터 모델링은 다음과 같은 두 가지 과정으로 나누어진다.

- 동사 개체 유형 V의 정의-V의 관련 개체 유형 E<sub>j</sub>(j = 1, 2, ..., n)을 선택하고, V에 대한 E<sub>j</sub>(j = 1, 2, ..., n)의 최대 사상수를 결정한다.
- 동사 개체 유형 V의 관계 모델링-관련 개체 유형 E<sub>j</sub>(j = 1, 2, ..., n)에 대한 V의 최대 사상수를 결정한다. V와 E<sub>j</sub>(j = 1, 2, ..., n)간의 관계의 최대 사상수에 따라 외래키가 결정된다.

문서, 협업과의 인터뷰 등 다양한 경로를 이용하여 관리 대상 속성을 수집한다. 수집된 속성의 주인 개체 유형을 결정함

으로써 테이블의 일반 필드가 결정된다. 속성의 주인 개체 유형 결정 과정은 (Yoo, 2007)에서 제시된 방법론이 사용된다. 이는 주인 개체 유형이 될 만한 후보 개체 유형을 선택하고, 기본 키와 속성의 함수 종속 여부를 판단하는 과정으로 이루어진다. 함수 종속 여부에 따라 후보 개체 유형의 조상 또는 후손 개체 유형을 다음 후보 개체 유형으로 선택한다. 조상 개체 유형은 부모 개체 유형과 부모의 부모 개체 유형 등을 말하며, 후손 개체 유형은 자식 개체 유형과 자식의 자식 개체 유형 등을 말한다. 개체 유형  $P_0$ 의 조상 개체 유형은  $\{P_i \mid P_i \text{는 } P_0 \text{의 부모 개체 유형, } i = 1, 2, 3, \dots\}$ , 개체 유형  $C_0$ 의 후손 개체 유형은  $\{C_i \mid C_i \text{는 } C_0 \text{의 자식 개체 유형, } i = 1, 2, 3, \dots\}$ 와 같이 표현될 수 있다. 최초의 후보 개체 유형은 거의 대부분 주인 개체 유형이나 주인 개체 유형의 조상 또는 후손 개체 유형에서 선택되므로 제시된 방법론에 의해 주인 개체 유형이 발견될 가능성은 매우 높다.

본 논문이 제시하는 방법론은 관계의 최대 사상수나 함수 종속과 같이 객관적이고 정량적인 판단 기준을 사용함으로써 데이터 모델링 초보자가 사용하기에도 적합하다.

### 2. 명사 개체 유형의 정의

업무와 관련된 사람, 장소, 조직, 유무형의 재화를 대상으로 유사한 속성을 가진 개체의 집합을 정의한다. 하나의 명사 개체 유형은 하나의 테이블과 일대일로 대응된다. 명사 개체 유형에 대응되는 테이블은 일반적으로 단일 필드로 이루어진 기본 키를 갖는다. 이 단계에서는 속성을 구체적으로 정의하지 않으므로 개체의 명칭에 의존하여 집합을 정의하는 것으로 충분하다. 추후에 개체 유형의 관리 대상 속성을 정의하는 과정에서 개체의 집합을 통합하거나 분리하는 작업이 이루어진다. 관리 대상 속성이 유사하면 하나의 개체 집합 또는 개체 유형으로 통합하고, 매우 상이하면 서로 다른 개체 유형으로 분리하여 정의하는 것이 일반적인 원칙이다. 개체 유형의 통합과 분리 과정은 본 논문에서 제시하는 방법론의 적용 범위를 벗어난다.

### 3. 동사 개체 유형의 정의

문장에서 하나의 서술어는 하나의 사실을 표현한다. 둘 이상의 서술어를 가진 문장은 둘 이상의 사실을 표현하는 것이다. 관리 대상 사건이 한 가지 사실에 국한되기 위해서는 주어와 서술어가 각각 하나씩 있어서 둘 사이의 관계가 한 번만 이루어지는 문장, 즉 단문으로 서술되어야 한다. 그리고 관리 대상 사건과 직접적으로 관련된 개체 유형이 가능한 모두 단문에 언급되도록 한다.

하나의 관리 대상 사건이 다른 관리 대상 사건에 종속되어

발생하고, 사건 간의 종속성이 관리 대상일 경우 후자의 사건 집합인 동사 개체 유형  $V'$ 가 전자의 사건 집합인 동사 개체 유형  $V$ 의 관련 개체 유형이 된다. 그런데 일반적으로 사건의 종속성은 사건의 발생 순서에 의해 결정된다. 예를 들어 “의사가 환자를 진료하다”라는 사건이 “의사가 약을 처방하다”라는 사건보다 먼저 발생하고, 후자는 전자에 종속된다. 전자의 사건 집합인 동사 개체 유형이 먼저 정의된 후에 후자의 사건 집합인 동사 개체 유형이 정의되어야 한다. 따라서 관리 대상 사건을 서술한 단문들을 사건 발생순으로 정렬해야 한다.

관리 대상 사건을 서술한 단문을 준비하는 과정은 다음과 같다.

- 관리 대상 사건을 관련 개체 유형이 포함된 단문으로 서술한다.
- 서술된 단문들을 사건이 발생하는 시간 순으로 정렬한다.
- 정렬된 단문들을 대상으로 하나씩 순서대로 모델링한다. <Figure 2>에 관리 대상 사건의 모델링 흐름도가 나타나 있다.

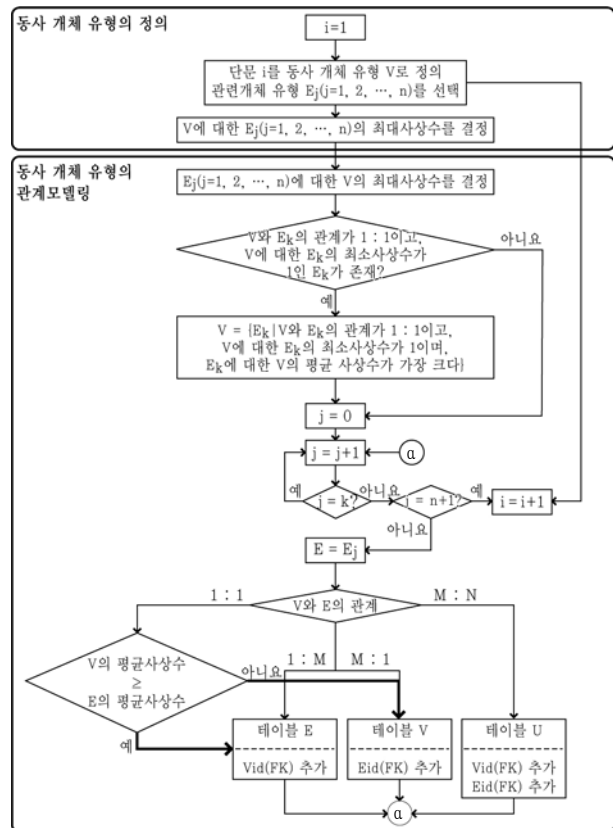


Figure 2. Flow diagram for modeling target events

하나의 관리 대상 사건의 집합을 동사 개체 유형  $V$ 라고 하면, 동사 개체 유형  $V$ 의 정의에서 결정되어야 할 사항은 다음과 같다.

- 관련 개체 유형  $E_j(j = 1, 2, \dots, n)$ 의 선택
- 관련 개체 유형의 중복 적용

- 동사 개체 유형  $V$ 에 대한  $E_j(j = 1, 2, \dots, n)$ 의 최대 사상수

### 3.1 관련 개체 유형 $E_j(j = 1, 2, \dots, n)$ 의 선택

관리 대상 사건이 이전에 발생한 다른 사건에 종속되어 발생하고, 사건 간의 종속성이 관리 대상이라면, 이전 사건 집합인 동사 개체 유형  $V'$ 가 종속 사건 집합인 동사 개체 유형  $V$ 의 관련 개체 유형으로 선택되어야 한다. 그리고 관련 개체 유형  $V'$ 가 포함되도록 종속 사건을 서술하는 단문을 수정한다.

동사 개체 유형  $V'$ 가 동사 개체 유형  $V$ 의 관련 개체 유형이고, 동사 개체 유형  $V$ 와  $V'$ 가 모두 동일한 개체 유형  $E$ 를 관련 개체 유형으로 갖는다고 하자. 이 경우 다음 조건을 모두 만족하면 개체 유형  $E$ 는 동사 개체 유형  $V$ 의 관련 개체 유형에서 제외되어야 한다.  $V$ 와  $E$ 의 관계는  $V$ 와  $V'$  그리고  $V'$ 와  $E$ 의 두 가지 관계에 의해 모델링되기 때문이다.

- $V$  개체 인스턴스  $v$ 와  $V'$  개체 인스턴스  $v'$ 가 관련된다.
- $v'$ 와  $E$  개체 인스턴스 집합  $e$ 가 관련된다.
- $v$ 와  $e$ 가 관련된다.

예를 들어, “의사가 환자를 진료하다”라는 관리 대상 사건의 집합인 동사 개체 유형을 ‘진료’라고 하고, “진료 결과 집도의로서 수술하다”라는 관리 대상 사건의 집합인 동사 개체 유형을 ‘집도의 수술’이라고 하면, 진료는 집도의 수술의 관련 개체 유형이다. 그런데 진료한 의사가 항상 집도의가 된다고 가정하면, 의사는 집도의 수술 개체 유형의 관련 개체 유형에서 제외되어야 한다. 집도의 수술과 의사의 관계는 집도의 수술과 진료 그리고 진료와 의사의 두 가지 관계에 의해 모델링되기 때문이다.

동사 개체 유형  $V$ 와 관련 개체 유형  $E$ 가  $M:N$  관계를 가지면, 관계의 개체화에 의해 관계 개체 유형이 생성되어야 한다. 관계 개체 유형을  $U$ 라고 하면,  $U$ 는  $V$ 로 정의되는 사건의  $E$  관련 내역이 된다.  $V$ 는 사건,  $U$ 는 사건의 내역이 되며,  $V$ 와  $U$ 에 대한  $E$ 의 최대 사상수는 각각  $M$ 과  $1$ 이 된다.  $V$  또는  $U$ 로 정의되는 사건 또는 사건 내역에 종속되어 발생하는 관리 대상 사건의 집합을 동사 개체 유형  $W$ 라 하고 사건 간의 종속성이 관리 대상이라면,  $V$ 와  $U$  중에서  $W$ 와 관련된 개체 유형을 선택해야 한다.  $W$  인스턴스  $w$ 에  $V$  인스턴스  $v$ 가 관련되고,  $v$ 에  $E$  인스턴스 집합  $e$ 가 관련된다면 선택 기준은 다음과 같다.

- 집합  $e$ 가 항상 하나의 단위로  $w$ 와 연관된다면  $V$ 를  $W$ 의 관련 개체 유형으로 선택해야 한다.
- 집합  $e$ 의 원소가 개별적으로  $w$ 와 연관된다면  $U$ 를  $W$ 의 관련 개체 유형으로 선택해야 한다.

예를 들어, “진료 결과 주사를 처방하다”라는 관리 대상 사건의 집합을 동사 개체 유형 ‘주사처방’이라고 하고 ‘주사처방’의 관련 개체 유형을 진료와 주사라고 하자. 주사처방과 주

사가  $M:N$  관계를 가진다고 하고, 관계의 개체화에 의해 생성된 관계 개체 유형을 ‘주사처방내역’이라고 하면, 주사처방내역은 주사처방 사건의 주사 관련 내역이 된다. “주사처방을 간호사가 실시하다”라는 사건이 주사처방 사건에 종속되어 발생하고, 두 사건 간의 종속성이 관리 대상이라면, ‘주사처방’과 ‘주사처방내역’ 중에서 ‘주사실시’와 관련된 개체 유형을 선택해야 한다. 하나의 주사실시 인스턴스  $w$ 에 하나의 주사처방 인스턴스  $v$ 가 관련되고,  $v$ 에 주사 인스턴스 집합  $e$ 가 관련된다면, 주사 인스턴스 집합  $e$ 가 항상 하나의 단위로 주사실시 인스턴스  $w$ 와 연관된다면, 주사처방이 주사실시 개체 유형의 관련 개체 유형으로 선택되어야 한다. 그러나 주사실시 인스턴스  $e$ 의 원소가 개별적으로 주사실시 인스턴스  $w$ 와 관련된다면 주사처방내역이 주사실시 개체 유형의 관련 개체 유형으로 선택되어야 한다. 달리 표현하면 처방된 주사 모두가 항상 한 명의 간호사에 의해 하나의 단위로 실시된다면 주사처방을, 여러 명의 간호사에 의해 개별적으로 실시될 수 있다면 주사처방내역을 주사실시의 관련 개체 유형으로 선택해야 한다.

### 3.2 관련 개체 유형의 중복

사건에 관련된 하나의 개체 유형의 역할이 구별되어 관리되어야 할 경우, 해당 개체 유형을 서로 다른 의미의 관련 개체 유형으로써 중복 사용한다. 그러나 이것이 개체 유형의 분리를 의미하지는 않는다.

예를 들어 “의사가 환자를 수술하다”라는 수술 사건에 관련된 의사의 역할이 집도의와 보조시술의로 구별되어 관리되어야 한다면, 의사 개체 유형을 집도의와 보조시술의의 서로 다른 의미의 관련 개체 유형으로써 중복 사용한다. 그러나 이것이 의사 개체 유형을 집도 의사와 보조시술 의사의 두 가지 개체 유형으로 분리하는 것을 의미하지는 않는다.

### 3.3 동사 개체 유형 $V$ 에 대한 관련 개체 유형 $E_j(j = 1, 2, \dots, n)$ 의 최대 사상수

동사 개체 유형  $V$ 에 대한 관련 개체 유형  $E_j(j = 1, 2, \dots, n)$ 의 최대 사상수는 “하나의  $V$  개체 인스턴스와 관련된  $E_j(j = 1, 2, \dots, n)$  개체 인스턴스의 수가 많아야 하나인지 아니면 둘 이상일 수 있는지”에 따라  $1$  또는  $M$ 이 된다. 하나의  $V$  인스턴스에  $E_j$  인스턴스가 둘 이상 관련될 수 있고, 이들이 하나의 단위로 취급되어 다른 관리 대상 사건에 연관된다든지 또는 하나의 단위로 관리 대상 속성을 가질 경우  $V$ 에 대한  $E_j$ 의 최대 사상수를  $M$ 으로 정의한다.

예를 들어 “고객이 상품을 주문하다”로 표현된 사건의 집합을 동사 개체 유형 ‘주문’이라고 하고, 관련 개체 유형이 고객과 상품이라면, 하나의 주문 인스턴스에 대한 고객과 상품 인스턴스의 최대 사상수를 결정한다. 하나의 주문 사건에 하나 이상의 상품이 포함되는 경우는 일반적이다. 이 상품들이 하나

의 단위로 취급되어 수주되고 배송되는 사건이 관리 대상이고, 하나의 단위로서 주문일, 배송일, 배송지주소 등을 관리 대상 속성으로 가지는 경우도 일반적이다. 따라서 '주문' 동사 개체 유형에 대한 '상품' 개체 유형의 최대 사상수를  $M$ 으로 정의한다.

공동 구매의 경우 하나의 주문 사건에 한 명 이상의 고객이 포함될 수 있다. 그러나 이 고객들이 하나의 단위로 취급되어 다른 관리 대상 사건에 연관된다든지 또는 하나의 단위로서 관리 대상 속성을 가지지 않는다고 가정하면 '주문' 동사 개체 유형에 대한 '고객' 개체 유형의 최대 사상수는 1로 정의한다.

#### 4. 동사 개체 유형의 관계 모델링

동사 개체 유형의 정의에서 동사 개체 유형  $V$ 에 대한 관련 개체 유형  $E_j(j = 1, 2, \dots, n)$ 의 최대 사상수를 결정하였으므로 관련 개체 유형  $E_j(j = 1, 2, \dots, n)$ 에 대한 동사 개체 유형  $V$ 의 최대 사상수를 결정하면  $V$ 와  $E_j(j = 1, 2, \dots, n)$  관계의 최대 사상수가 결정된다.

$V$ 와  $E_j(j = 1, 2, \dots, n)$ 의 관계 중에서 다음 조건을 모두 만족하는  $E_k$ 가 존재하면,  $E_k$ 는  $V$ 가 갖는 모든 관계와 속성을 대신 가질 수 있다. 따라서  $V$ 의 역할을  $E_k$ 가 대신하고,  $V$ 는 삭제한다.

- $V$ 와  $E_k$ 의 관계가 1:1이다.
- $V$ 에 대한  $E_k$ 의 최소 사상수가 1이다.

위의 조건은 만족하는  $E_k$ 가 둘 이상 존재할 경우  $V$ 의 관계를 나타내는 외래키 필드의 Null 발생을 최소화하기 위하여  $E_k$ 에 대한  $V$ 의 평균 사상수가 최대인  $E_k$ 가  $V$ 를 대신한다.

예를 들어 "고객이 상품을 주문하다"라는 사건의 집합을 동사 개체 유형 '주문'이라고 하고, 관련 개체 유형이 고객, 상품, 직원이라고 하자. 하나의 주문 개체 인스턴스와 관련된 고객과 직원 개체 인스턴스가 항상 하나이고, 한 명의 고객 또는 직원 개체 인스턴스와 관련된 주문 개체 인스턴스는 많아야 하나라고 가정하자. 이는 한 명의 고객이 많아야 한 번 주문할 수 있고, 한 명의 직원이 많아야 한 건의 주문을 처리할 수 있음을 의미한다(매우 비현실적인 가정임). 이 경우 고객이나 직원 개체 유형이 주문 개체 유형이 갖는 관계와 속성을 대신 가질 수 있다. 고객이나 직원 개체 유형이 주문 개체 유형을 대신하고, 주문 개체 유형은 삭제한다. 그런데 고객의 수가 직원의 수보다 많다면 고객 개체 유형에 대한 주문 개체 유형의 평균 사상수가 직원 개체 유형에 대한 주문 개체 유형의 평균 사상수보다 작다. 이 경우 직원이 주문 개체 유형을 대신하는 것이 외래키 필드의 Null 발생을 최소화한다. 예를 들어 직원 테이블에 '주문고객번호' 외래키 필드를 두는 것이 고객 테이블에 '주문처리직원번호' 외래키 필드를 두는 것보다 외래키 필드의 Null 발생이 적다.

$V$ 와  $E_j$  관계의 최대 사상수에 따라 부모와 자식 개체 유형은 다음과 같이 결정된다(Yoo, 2010; 52-58). 자식 개체 유형은 부모의 기본키를 외래키 속성으로 갖는다.

- $V$ 와  $E_j$  개체 유형 간의 관계의 최대 사상수가 1:M인 경우 1쪽인  $V$  개체 유형이 부모,  $M$ 쪽인  $E_j$  개체 유형이 자식이 된다.  $E_j$  개체 유형은  $V$ 의 기본키를 외래키 속성으로 갖는다.
- $V$ 와  $E_j$  개체 유형 간의 관계의 최대 사상수가 M:1인 경우 1쪽인  $E_j$  개체 유형이 부모,  $M$ 쪽인  $V$  개체 유형이 자식이 된다.  $V$  개체 유형은  $E_j$ 의 기본키를 외래키 속성으로 갖는다.
- $V$ 와  $E_j$  개체 유형 간의 관계가 M:N인 경우  $M$ 쪽인  $V$ 와  $E_j$  개체 유형은 모두 부모 개체 유형이 된다. M:N 관계는 개체 화하여 새로운 관계 개체 유형  $U$ 가 생성되고, 이는 자식 개체 유형이 된다.  $U$  개체 유형은  $V$ 와  $E_j$ 의 기본키를 외래키 속성으로 갖는다.
- $V$ 와  $E_j$  개체 유형 간의 관계의 최대 사상수가 1:1인 경우 평균 사상수가 큰 개체 유형이 부모, 작은 개체 유형이 자식이 된다. 만약  $V$  개체 유형의 평균 사상수가  $E_j$  개체 유형의 평균 사상수보다 크면,  $V$  개체 유형이 부모,  $E_j$  개체 유형이 자식이 된다.  $E_j$  개체 유형은  $V$ 의 기본키를 외래키 속성으로 갖는다. 반대로  $E_j$  개체 유형의 평균 사상수가  $V$  개체 유형의 평균 사상수보다 크면,  $E_j$  개체 유형이 부모,  $V$  개체 유형이 자식이 된다.  $V$  개체 유형은  $E_j$ 의 기본키를 외래키 속성으로 갖는다. 외래키 속성은 고유 값 제약조건이 설정되어야 한다. 그렇지 않으면 자식 개체 유형의 최대 사상수가  $M$ 이 될 수 있다.

#### 5. 관리 대상 속성의 정의

업무 관련 문서의 조사, 현업과의 인터뷰 등 다양한 경로를 이용하여 관리 대상 속성을 수집한다. 속성이 가질 수 있는 값의 집합을 원소나열법이나 조건제시법을 이용하여 가능한 상세하고 구체적으로 정의한다.

속성의 주인 개체 유형을 결정한다. 속성의 주인 개체 유형이란 속성을 소유하는 개체 유형을 말한다. 속성의 주인 개체 유형을 결정하기 위해서는 우선 주인 개체 유형이 될 만한 후보 개체 유형을 선택한다. 최초의 후보 개체 유형이 잘못 선택될 수 있지만 전혀 무관한 후보 개체가 선택될 가능성은 높지 않다. 적어도 주인 개체 유형의 조상이나 후손 개체 유형에서 후보 개체 유형이 선택될 가능성이 높다. 최초의 후보 개체 유형으로 전혀 무관한 후보 개체가 선택되더라도 기본키와 속성 간에 함수 종속이 성립하지 않을 확률이 높으므로 주인 개체 유형을 잘못 결정할 가능성은 매우 낮다. 예를 들어 <Figure 3>에서 '고객', '주문', '주문상품내역', '상품' 등의 개체 유형은 서로 부모-자식 관계를 갖는다. '주문량'이라는 속성의 주인 개체 유형은 '주문상품 내역'이지만, 후보 개체 유형으로 '주문상품'의 조상 개체 유형인 '주문', '고객', '상품'을 선택할 가능

성이 높다. 따라서 Yoo(2007)에 제시된 명제 1~5에 따라 다음과 같은 기준으로 속성의 주인 개체 유형을 찾을 경우 오류가 발생할 확률은 매우 낮다.

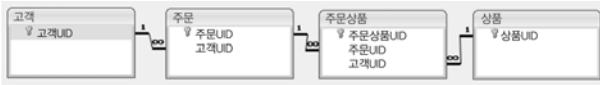


Figure 3. Parent-child relationship between tables associated with 'order' event

- 후보 개체 유형의 기본키와 속성 간에 함수 종속이 성립하지 않으면 속성의 주인 개체 유형은 후보 개체 유형의 자손 개체 유형에 존재한다. 따라서 후보 개체 유형을 자식 개체 유형으로 변경한 후 함수 종속 여부를 조사한다.
- 후보 개체 유형의 기본키와 속성 간에 함수 종속이 성립하면 속성의 주인 개체 유형은 후보 개체 유형일 수도 있지만 후보 개체 유형의 조상 개체 유형에 존재할 수도 있다. 따라서 후보 개체 유형을 부모 개체 유형으로 변경한 후 함수 종속 여부를 조사한다.
- 후보 개체와 부모 또는 자식 개체 간의 관계가 1:1이라면 부모 또는 자식 개체의 기본키와 속성 간에는 항상 함수 종속이 성립하기 때문에 함수 종속 여부로는 부모 또는 자식 개체가 속성의 주인 개체인지 검증할 수 없다. 이는 모델러의 경험과 지식에 의존할 수밖에 없다.

최초의 후보 개체 유형으로 전혀 무관한 개체 유형을 선택할 경우 후보 개체 유형이나 그 조상/후손 개체 유형을 주인 개체 유형으로 결정할 확률은 낮다. 대부분의 경우 기본키와 속성 간의 함수 종속이 성립하지 않을 것이기 때문이다.

<Figure 4>에 속성의 주인 개체 유형을 결정하기 위한 흐름도가 나타나 있다.

### 6. 결론 및 추후 연구 방향

기본키와 외래키, 관계의 최대 사상수, 개체 유형 간의 부모-자식 관계, 그리고 기본키와 속성 간의 함수 종속 등 데이터베이스의 기본 개념만 습득하면, 데이터베이스를 처음 접하는 초보자들도 쉽게 이해하고 사용할 수 있는 방법론을 제시하였다. 참조 무결성 원칙의 개념을 적용하여 조상 개체 유형에서 후손 개체 유형의 순서로 개체 유형의 정의가 이루어지도록 하였다. 데이터 모델링은 데이터베이스 개념 뿐 아니라 적용 분야에 대한 지식과 경험에 크게 의존하는 매우 추상적인 작업이다. 따라서 어떠한 모델링 방법론도 모델러의 판단 오류를 배제할 수는 없다. 그러나 본 논문에서 제시된 방법론은 객관적이고 정량적인 판단 방법을 사용함으로써 오류를 최소화한다. 데이터베이스 교과목을 수강하는 대학교 2학년 학생들에게 방법론을 설명하고 데이터베이스 설계 과제를 수행하게 한 결과 논리적 데이터 모델의 수준이 높아지고, 설계 오류가 감소하여 수정 요구가 최소화되었음을 관찰할 수 있었다. 방법론을 좀 더 쉽게 적용할 수 있도록 모델링 소프트웨어를 개발하는 것이 추후 목표이다.

### 참고문헌

Chen, P. (1976), The Entity Relationship Model: Towards a Unified View of Data, *ACM Transactions on Database Systems*, 1(1), 9-36.

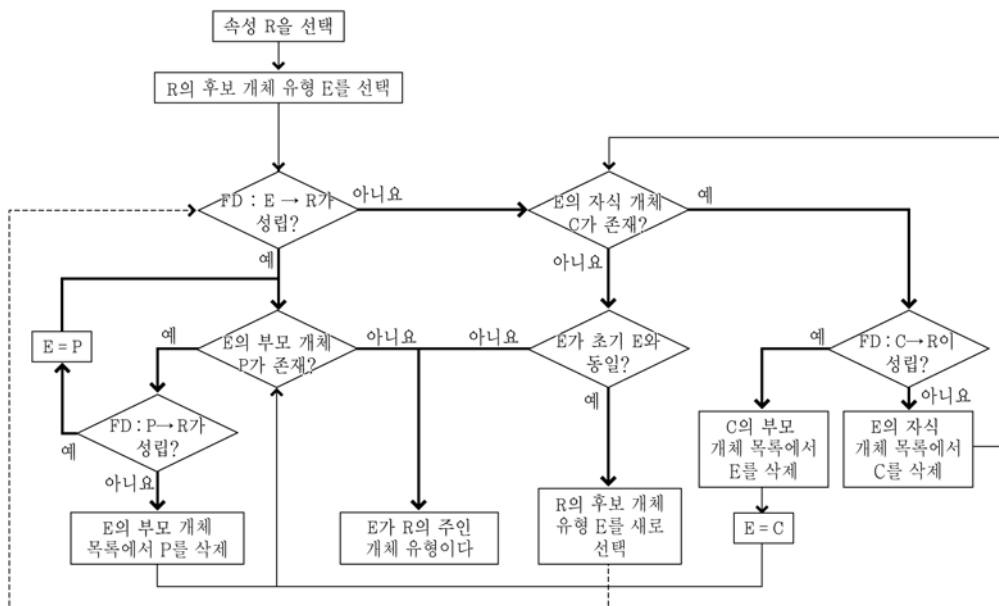


Figure 4. Flow diagram for finding the owner entity type of an attribute

- Lee, W. S. (2003), *Data Architecture Solution I*, En-Core Consulting, Seoul, Korea.
- Ra, Y. (2009), Executional Form Prototyping System for Data Modeling Review, *Journal of KIISE Database Society*, 25(1), 19-37.
- Shoval, P. and Shiran, S. (1997), Entity-relationship and object-oriented data modeling-An experimental comparison of design quality, *Data and Knowledge Engineering*, **21**, 297-315.
- Yoo, J. (2007), A Study for Consolidating the Normalizing Function of Computer-Aided Database Design System, *Engineering Research*, **5**(1), 131-143.
- Yoo, J. (2010) *Step by Step : Access 2007 Database Basics and Applications*, Hansung University Press, Seoul, Korea.