

□ 기술해설 □

정보공학과 데이터 모델링

(주)위세정보기술 김 중 현

| ● 목 | 차 ● |
|-----------------|-------------------|
| 1. 데이터모델의 특징 | 4.1 기본데이터모델 작성 |
| 2. 데이터의 계층별 모델링 | 4.2 데이터모델 상세화 |
| 3. 엔터프라이즈데이터모델링 | 4.3 데이터 통합 |
| 3.1 주제영역의 정의 | 4.4 상관분석 |
| 3.2 기본실체의 정의 | 5. 데이터모델링의 주요성공요소 |
| 4. 논리적 데이터모델링 | |

1. 데이터모델의 특징

모델링이란 대상 비즈니스를 이해하여 모두에게 공통적으로 이해될 수 있도록 약속된 다이어그램으로 표현하는 작업이다. 모델링에서 가장 중점적으로 파악해야 할 엔터프라이즈 대상 객체는 데이터와 프로세스라고 할 수 있으며 이 둘을 비교해 볼 때 데이터는 정적임에 반해 프로세스는 동적이고 데이터보다 더 추상적이다. 토마스 데이븐포트는 프로세스모델링에 대하여 “프로세스를 명확하게 규정하고 설정하기 위하여 아무리 많은 노력을 기울여도, 프로세스 조감은 시간과 경험에 따라 변한다. 프로세스 정의는 과학(Science)이라기 보다는 예술(Art)에 가깝다.”라고 실토하고 있다.

여기에 비해 비즈니스에서 알고 싶어하는 데이터는 거의 영속적으로 존재한다. 조직과 기술에 비해 상대적으로 안정적인 프로세스를 대폭적으로 재설계하더라도 비즈니스 관점의 데이터는 거의 변화하지 않는다. 즉 데이터는 엔터프라이즈 객체 중 가장 안정적임에 따라 데이터 중심의 모델링은 변화에 가장 안정적인 시스템을 구축하는 첫걸음이 될 수 있다.

데이터와 프로세스의 물리적인 구현이 데이터



그림 1 프로그램의 유형

베이스와 프로그램이라고 할 때 프로그램은 그림 1과 같이 6가지 유형으로 분류할 수 있다.

이런 점에서 “프로그램은 데이터베이스를 유지보수하는 로직 또는 절차”라고 정의될 수 있으며 데이터베이스가 프로그램보다 더 중심에 위치해 있음을 알 수 있다. 바호만은 더 나아가 프로그램을 단지 “데이터의 바다에 떠 있는 항해사”로 비유하고 있다. 따라서 데이터모델은 안정적이고 가장 중심에 위치해 있다라고 하겠다.

2. 데이터의 계층별 모델링

데이터모델링의 주기는 정보전략계획 단계의 엔터프라이즈데이터모델링과 업무영역분석 단계의 논리적데이터모델링으로 나눌 수 있으며 양자의 차이는 데이터 객체에 대한 구체성의 정도

표 1 데이터모델의 용어 정의

| 용어 | 개념 |
|---|---|
| <ul style="list-style-type: none"> 주제영역(Subject Area) 실체(Entity) | <p>주로 자원이나 활동을 중심으로 한 기업의 자연스러운 분할로 하나 이상의 실체로 분할됨</p> <p>기록하고자 하는 사실의 대상인 사람, 사물 또는 개념으로 데이터의 물리적인 표현 형태인 레코드(Record)와 비교되는 논리적인 표현</p> <ul style="list-style-type: none"> 기본실체 : 전사적으로 공유되거나 독립적으로 존재하는 실체 2차 실체 : 단일 부서에서 사용되거나 기본실체의 구체화 과정에서 추가적으로 정의됨 |
| <ul style="list-style-type: none"> 관계(Relationship) 속성(Attribute) | <p>2개 실체간의 연관성(Association)</p> <p>실체를 구성하는 자료항목으로 더이상 쪼개질 수 없는 최소 단위의 정보로, 물리적인 표현 형태인 데이터 필드와 비교되는 논리적 표현</p> |
| <ul style="list-style-type: none"> 식별자(Identifier) | <p>실체의 단일 경우(Occurrence)를 유일하게 식별할 수 있는 속성 또는 최소한의 속성집합</p> |

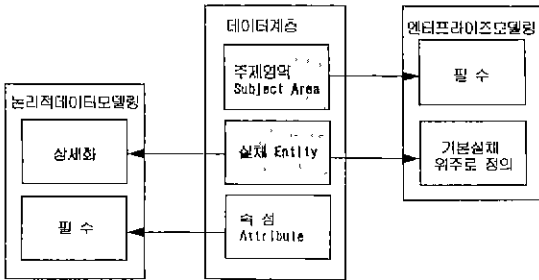


그림 2 데이터 계층

하여 이를 알기 쉽고 체계적으로 표현한 기업의 조감적 정보요구”로 정의될 수 있으며 주제영역도 작성과 실체관계도 작성의 2단계로 모델링 작업이 진행된다. 해운산업을 대상으로 2단계 작성된 엔터프라이즈데이터모델을 예시하면 그림 3과 같다.

엔터프라이즈모델링에 대한 지침을 간단히 소개하면 다음과 같다.

3.1 주제영역의 정의

- 주된자원, 제품과 서비스를 중심으로한 데이터의 자연스러운 묶음
- 6~10개의 기본실체로 구성
- 경영목표와 CSF(필수성공요소)에서 등장되는 명사에 주목

3.2 기본실체의 정의

- 기업의 주된자원을 기록
- 경영환경의 객체를 기록
- 주요성과지표를 실체에 포함시킴

로 나타난다. 데이터모델링은 하향식으로 작성되어 정보전략계획의 엔터프라이즈데이터모형은 주제영역과 기본실체를 중심으로 구축되고 업무영역분석에서는 데이터모델의 모든 요소가 상세하고도 완전히 정의된다. 데이터모델의 구성요소에 대한 개념은 간단히 설명하면 표 1과 같다. 구체성의 기준을 데이터 계층으로 그림 2와 같이 나눌 때 엔터프라이즈모델에서는 상위계층인 주제영역과 전사적으로 공유되는 기본실체 위주로 정의되며 논리적 데이터모델링을 통해 이들은 완전히 구체화되어 추가적인 실체와 속성이 결정된다.

3. 엔터프라이즈데이터모델링

엔터프라이즈데이터모델링은 “비즈니스에 대한 종합적인 이해를 바탕으로 기업내에 존재하는 데이터에 대하여 프로세스와는 독립적으로 인식

4. 논리적 데이터모델링

논리적데이터모델링은 일정 업무영역에 대하여 나중에 참조하기 위하여 저장되어야 할 데이터와 데이터들간의 연관성을 실체관계도로 도시

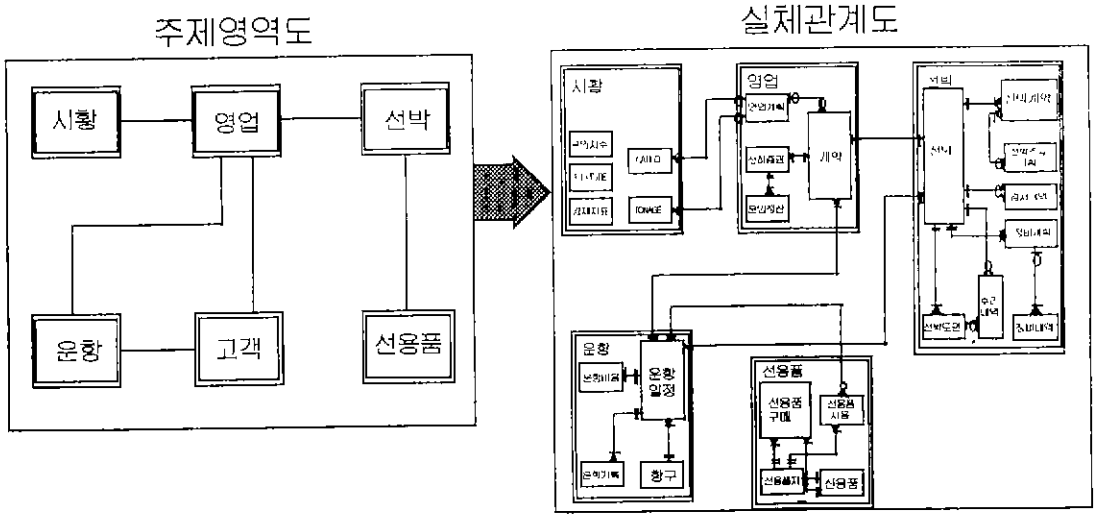


그림 3 2단계로 작성된 엔터프라이즈모델

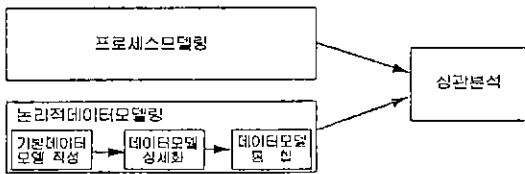


그림 4 논리적 데이터모델링의 세부단계

하고 실체의 구체적인 내용인 속성과 데이터의 정확성을 유지하기 위한 업무규칙을 구체적으로 정의하는 데이터중심의 분석기법이다. 논리적 데이터모델링의 작업단계는 그림 4와 같이 프로세스모델링과 동시에 진행이 되며 이 때 데이터 객체와 프로세스 객체와의 구체성이 서로 격에 맞도록 작업을 진행하며 양 객체별 모델링이 종료된 후 서로의 상관관계를 분석하게 된다.

4.1 기본데이터모델 작성

실체와 관계만을 정의하여 기본적인 데이터 모델을 작성하는 단계로 세부작업을 살펴보면 다음과 같다.

4.1.1 실체 정의

사용자의 관심 대상이 되는 실체란 구체적이고 물리적인 것도 있지만 추상적인 개념과 서비스도 포함될 수 있기 때문에 해당 업무영역에 대한

어느 정도의 업무지식과 현업 사용자와의 긴밀한 공동작업이 요구된다. 실체를 식별하는 방법을 구체적으로 소개하면 다음과 같다.

- 인터뷰에서 사용자가 관심을 갖고 있는 데이터에 주의를 기울이며 사용자가 이야기하는 명사, 자원, 제품, 개념과 서비스에 주목한다.
- 현업에서 사용되는 문서와 대장을 조사한다. 문서와 대장에는 복수개의 실체가 복합적으로 존재하는 실체 추출의 보고서이다.
- 기존시스템의 데이터베이스와 화일은 대부분 실체가 비정규화된 형태로, 즉 복합적으로 존재하기 때문에 실체 추출의 좋은 대상이다.
- 대상 실체가 업무에의 중요성, 가치 그리고 안정성을 가지고 식별자 이외의 속성이 존재하는지 검증한다.¹⁾

4.1.2 관계 정의

두 실체간의 관련 사실이나 연관성으로부터 도출되는 관계는 기수(Cardinality)와 선택성으로 표현된다. 이들은 업무의 정책이나 규칙을 자연스럽게 나타내는 수단이다. 기수와 선택성에 대하여 예시하여 설명하는 다음과 같다.

¹⁾식별자와 속성의 구체적인 정의는 "데이터모형 상세화" 단계에서 수행된다. 여기에서는 단지 존재 유무만 판단하면 된다.

• 기수

| 1 대 1 | 1 대 다 | 다 대 다 |
|---|---|---|
| | | |
| 하나의 공장은 하나의 유형 제품을 생산한다. 하나의 제품유형은 하나의 공장에서만 생산된다. | 하나의 공장은 하나 이상의 제품을 생산한다. 하나의 제품유형은 하나의 공장에서만 생산된다. | 공장들은 여러제품을 생산한다. 각각의 제품유형은 하나 이상의 공장에서 생산된다. |

그림 5 기수의 예시

• 선택성(Optionality)

| 필수적 | 선택적 |
|--|-------------------------------|
| | |
| 모든 고객은 반드시 계약을 체결해야 한다. 계약을 체결한 고객만 서비스를 받을 수 있다. | 계약을 체결하지 않은 고객도 서비스를 받을 수 있다. |

그림 6 선택성의 예시

관계를 추측하기 위해서는 다음과 같은 의문을 가져야 한다.

- 두 실체간의 업무에 중요한 연관이 있는가?
- 두 실체에 대한 정보의 조합이 필요한가?
- 두 실체의 연관을 입증하는 문서, 매뉴얼 또는 양식이 존재하는가?
- 두 실체를 연결짓는 동사가 있는가?

4.1.3 실체관계도 작성

정의된 실체와 관계를 그림으로 일목요연하게 표현하는 단계이며 아래의 순서로 진행된다.

- 관계가 많이 생기게 되는 거래나 사건으로부터 파생되는 개념적 실체를 중심에 배열한다.
- 관계선이 겹치지 않도록 관계를 도식한다.
- 관계의 특성인 기수와 선택성을 표기한다.

제품구매 업무영역을 대상으로 한 기본데이터 모델 형태를 예시하면 그림 7과 같다.

4.2 데이터모델 상세화

기본데이터모델을 토대로 실체의 특정 경우를 식별해 주는 식별자를 결정하고 각각의 실체에 대해 추가적인 세부 속성을 찾아내 실체 항목을 구성한 다음, 실체 항목의 중복성과 불일치성 배제를 위해 정규화 검증을 수행하게 된다. 그러나 정규화 검증만으로는 해당 업무영역에 대한 제반규칙을 데이터모델에 충분히 반영시킬 수 없기 때문에 이를 위해 추가적인 속성에 대한 업무규칙인 도메인을 정의하게 된다.

4.2.1 주·부식별자 결정

실체의 특정 경우(Occurrence)를 유일하게 식별해주는 속성 또는 최소한의 속성 집합을 모두 찾아내어 그중 하나를 주식별자(PK: Primary Key)로, 나머지를 부식별자로 결정하게 된다. 이때 주·부식별자가 되는 속성은 어떤 경우에

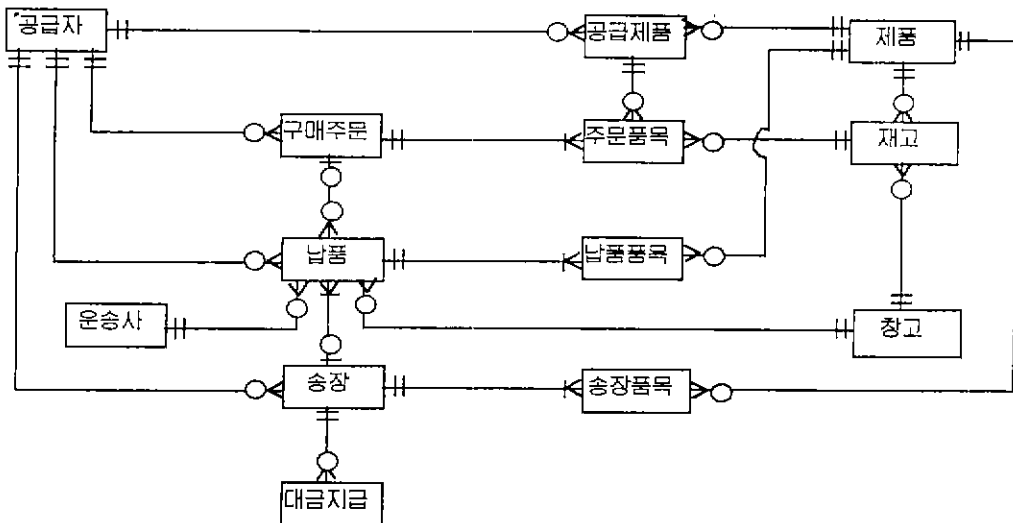


그림 7 제품구매 업무영역 기본데이터모델

있어서도 널(Null) 값을 가져서는 안된다.

4.2.2 외부식별자 결정

외부식별자(FK: Foreign Key)란 두 실체의 관계에 있어 상위실체(Parent Entity)의 특정 행을 가리키는 속성(또는 속성집합)이다.²⁾ 항상 하위 실체(Child Entity)에 위치하며 대부분 상위 실체의 주식별자와 같은 값을 가진다. 따라서 실체간의 관계를 나타내는 속성을 찾아내면 쉽게 결정할 수 있다.

4.2.3 속성 정의

각 실체에 대해 이미 찾아낸 주식별자에 종속적인 속성들을 해당 실체에 포함시키며, 이런 류의 속성을 비식별자 속성(Nonkey Attributes)이라고 한다. 비식별자 속성은 가능한 한 데이터모형내에서 관계를 갖는 상위 실체쪽으로 위치시켜야 하며, 만일 한 실체내의 속성이 여러 개의 값을 가질 수 있는 경우에는 새로운 하위 실체의 속성으로 재분류시켜야 한다. 또한 두 실체간의 관계로부터 파생되는 속성은 두 실체와 관계를 갖는 하위 실체를 추가하여 위치시킨다. 속성은 표 2와 같이 3가지로 분류될 수 있다.

제품구매 데이터모형을 토대로 제품과 재고 실체의 세부속성을 정의하면 표 3과 같다.

표 2 속성의 구분

| 속성구분 | 개 | 비 |
|-------|-------------------------------------|---|
| 기 초 | 속성의 값이 다른 속성의 값에서 추출될 수 없고 업무에서 재공급 | |
| 실 · 계 | 업무의 제약이나 시스템운영의 단순화를 위해 속성의 값이 정조됨 | |
| 추 출 | 속성의 값이 다른 속성의 값에 의해 추론되거나 계산됨 | |

표 3 속성정의표

| 실 체 | 속 성 | 속성유형 | 식별자 구분 | 비 고 |
|-----|-------------------------------------|----------------------------|------------------|----------|
| 제품 | 제품코드 재종명 기대수량 재주문요구량 | 설계 기초 기초 기초 | PK | |
| 재고 | 제품코드 창고코드 입고일 적정재고 현황일량 | 설계 설계 기초 기초 추출 | FK FK1 FK2 | 초과적 재고수준 |

²⁾ 데이터모형을 완전히 업무적인 측면에서 도시한다면 외부식별자를 데이터모형에 표기하지 않는 것이 원칙이다. 그러나 현실적으로 관계형기술을 사용하는 환경이 대부분이기 때문에 바로 외부식별자로 표기하는 방법을 선택했다. 이러한 점은 CASE 도구에서도 극명하게 드러나서 CASE 도구도 데이터모델링 단계에서 외부식별자를 표기하는 것과 표기하지 않는 것의 2분류로 분류되고 있다.

4.2.4 속성의 도메인 정의

각 속성이 지나는 데이터 값의 무결성을 관리하는 업무규칙으로부터 속성에 대한 데이터 형태, 길이, 포맷, 허용범위, 유일성구분, 널구분, 초기값 등의 도메인 특성을 정의하게 된다. 특히 식별자(주, 부식별자)에 해당하는 속성과 다소간의 중복성을 내포한 추출속성(Derived Attribute)에 대해서는 관련기본조건과 추출알고리즘을 추가한다. 구매주문 실체의 속성별 도메인을 예시하면 표 4와 같다.

4.3 데이터 통합

단위업무영역별로 작성된 데이터모형에 대한 정규화 검증과 정의된 데이터모형을 조합하여 하나의 통합된 데이터모형을 구축하는 단계이다.

4.3.1 정규화 검증

정규화란 데이터의 정확성, 중복의 최소화, 데이터 구조의 안정성을 높이기 위한 일련의 직관적인 규칙으로 완전히 정규화된 데이터모형은 논리적데이터모델링의 목적인 데이터의 정확성, 일치성, 단순성, 비중복성과 안정성을 보장해 준다.

여기에는 일반적으로 반복적이거나 여러 개의 값을 지닐 수 있는 속성을 제거하는 1차 정규화, 주식별자에 완전히 종속되지 않는 속성을 제거하는 2차 정규화, 그리고 비식별자 속성에 종속되는 속성을 제거하는 3차 정규화와 특정 속성 값에 따라 선택적인 속성을 제거하여 새로운 실체를 추출하는 4차 정규화가 있다.

4.3.2 데이터모형 조합

서로 다른 업무영역별 데이터모형을 대상으로 불일치성을 해소하고, 중복된 부분을 단일화하며, 데이터모형간에 새로운 관계나 업무규칙을 추가하여 통합된 데이터모형을 만들어 내는 단계이다. 여기에는 실체와 관련 업무규칙(주,부식별자) 통합, 관계 통합, 속성과 관련 도메인 통합의 순으로 진행된다. 이는 동일한 사실에 대한 서로 다른 데이터 표현을 밝혀내기 위한 고도의 분석력과 원활한 의사소통을 위한 협상기법이 요구

표 4 도메인 정의표

실체명 : 구매주문

| 속성명 | 타입 | 길이(n,m) | 포맷 | 범위 | 유일성 | 널여부 | 초기값 |
|-------|------|---------|--------------|------------------|--------|----------|-------|
| 주문번호 | NUM | 7 | YYMM999 | | Unique | Not Null | |
| 공급자번호 | CHAR | 5 | A9999 | | | Not Null | |
| 주문일자 | DATE | 8 | YY.MM.DD | | | | Today |
| 납기일 | DATE | 8 | YY.MM.DD | | | Not Null | |
| 총금액 | NUM | 9,0 | ₩999,999,999 | | | Not Null | |
| 상태 | CHAR | 1 | A | IN ('C','E','D') | | Not Null | 'C' |
| 최근접촉일 | DATE | 8 | YY.MM.DD | | | | |

되기 때문에 적은 규모로 각각의 통합 단계를 순서대로 수행하고 반복하는 것이 좋다.

4.3.3 엔터프라이즈모델로 통합

업무영역별 데이터모델을 통합된 엔터프라이즈모델로 발전시키는 작업이다. 여기에는 막대한 양의 데이터 객체와 공유되는 객체의 관리가 이루어져야 하기 때문에 CASE 도구의 사용과 CASE의 정보저장소(Repository)를 중앙에서 관리할 수 있는 중앙정보저장소관리자의 존재가 필수적이다.

4.4 상관분석

상관분석에서는 데이터모형에서 정의된 실체와 관계에 대하여 단위프로세스와 데이터의 2가지 관점에서 검토하게 된다.

4.4.1 실체생명주기분석

실체는 몇개의 상태로 구분이 가능하며 이를 구체화하면 생성상태, 중간상태, 종료상태로 나타낼 수 있다 분석가는 실체의 생명주기를 분석하여 실체를 일정 상태에서 다음의 상태로 전환시키는 단위프로세스의 실행을 통제하는 조건을 식별하여 프로세스나 실체의 누락을 찾아낼 수 있다. 실체생명주기 분석은 실체상태전환도(Entity State Transition Diagram)로 나타내며 세부 지침을 소개하면 다음과 같다.

- 선정된 실체의 가능한 생명주기 상태를 정의
- 실체의 상태를 변화시키는 업무활동을 식별

4.4.2 프로세스논리분석(Process Logic Analysis)

분석가는 프로세스논리분석에서 각각의 프로세스에 대한 데이터요건을 검토하여 프로세스의 실체와 관계에 대한 사항을 구체화시키고 실체관계도가 모든 프로세스를 지원할 수 있는가를 확인하게 된다. 프로세스의 논리를 분석하여 프로세스논리도(Process Logic Diagram)를 작성하는 절차를 소개하면 다음과 같다.

- 단위프로세스를 선정하고 여기에 대한 일반적 설명, 입력데이터, 출력 정보를 파악
- 선정된 단위프로세스에 연관된 실체를 식별
- 실체에 가해지는 동작(Read, Update, Delete)을 정의하고 데이터모델에 이를 표기
- 위의 활동에 순서를 매김

5. 데이터모델링의 주요성공요소

여태까지 기술한 논리적데이터모델링 방법론이 포괄적이고 효과적이지만, 실제작업을 할 때 이 방법론을 그대로 적용할 필요는 없다. 즉 프로젝트의 상황에 맞게 방법론을 단순화, 또는 확대하여 적용할 수 있다. 그러나 적용 방법론의 성격은 적절한 논리적데이터모델을 성공적으로 작성하는 데에 중요한 영향을 미친다는 점을 인식해야 한다. 논리적데이터모델링의 주요성공요소를 일곱가지로 정리하면 다음과 같다.

첫째, 사용자와 가능하면 공동으로 작업할 수 있도록 모임을 갖고 데이터모델에 대해 사용자와 이해를 공유하도록 한다.

둘째, 작업하기 전에 체계화된 방법론을 채택하고 논리적데이터모델링 전 단계에 걸쳐서 이를 따른다.

셋째, 논리적데이터모델이 특정 처리요건이나 기술적 관점에서 벗어나 어떻게 현업에서 정보를 사용하고 관리하는지를 나타낼 수 있도록 데이터 중심으로 접근한다.

넷째, 데이터베이스의 레코드 구성인 데이터 구조와 업무규칙의 2가지 관점을 논리적데이터베이스의 레코드 구성인 데이터구조와 업무규칙의 2가지 관점을 논리적데이터모델에 융합한다.

다섯째, 업무에서 정보를 사용하는것을 실체, 관계, 식별자로 나타내는 개념화와 더불어 속성을 적절한 실체에 귀속시키는 방법인 정규화를 방법론에 접목시킨다.

여섯째, 논리적데이터모델을 명확하고 단순하게 나타낼 수 있도록 도형(실체관계도)을 사용한다.

일곱째, 실체관계도를 보충하기 위해 실체/관계/속성에 대한 정의와 업무규칙을 자료사전을 이용하여 명세화한다.

참고문헌

[1] C. Finkelstein. An Introduction to Information Engineering from Strategic Planning to Information System. Addison Wesley, 1989.
 [2] C. Finkelstein. Information Engineering Strategic Development. Addison Wesley, 1993.
 [3] C. Fleming & B. Halle. Hand Book of Relational DB Design Addison Wesley, 1989.

[4] Richard Barker. CASE * METHOD Entity Relationship Modelling, Addison Wesley, 1990.
 [5] TEXAS INSTRUMENT. Information Strategy Planning Workshop. Addison Wesley, 1990.
 [6] TEXAS INSTRUMENT. Business Area Analysis Student Guide, 1993.
 [7] TEXAS INSTRUMENT. Advanced Data Modeling Student Guide, 1992.
 [8] 토마스 데이븐포트. 프로세스 이노베이션, 1993.
 [9] D. Vaskevitch. Client/Server Strategies. IDG Books, 1993.
 [10] 김종현, "업무영역분석", 경영과 컴퓨터 92년 12월.
 [11] 김규동, "논리적데이터모델링", 경영과 컴퓨터, 1994.
 [12] (주)위세정보기술, WISE * METHOD 정보전략 계획, 1994.
 [13] (주)위세정보기술, WISE * METHOD 논리적데이터모델링, 1994.



김 종 현

- 연세대 경제학과
- 연세산업대학원 전산
- 과학기술원 시스템공학센터
- (주)딜로이트 컨설팅
- 현재 (주)위세정보기술 대표 컨설턴트