

논문 2006-43CI-6-3

기업 전사 자료 설계에서 역정규화 중심 데이터 모델링의 부작용

(Negative Side Effects of Denormalization-Oriented Data Modeling in
Enterprise-Wide Database Design)

이 혜 경*

(Hae Kyung Rhee)

요 약

전산화 대상 규모가 확대되면서 새로운 방식의 데이터 모델링 방법론이 필요하게 됨에 따라 자료 설계라는 것에 최근 들어 1980년대 초반 못지않게 이른 바 데이터 거버넌스 혹은 데이터 질 등의 화두의 모습으로 큰 관심이 재연되고 있다. 그러나 데이터 모델링에 대한 확고한 견해를 가지고 있지 못하기 때문에 막상 데이터베이스 설계를 실무적으로 소화해 내기에는 적지 않은 어려움을 겪게 된다. 본 논문에서는 업무의 요구 형성/분석 단계부터 데이터베이스 구현까지 최종 단계까지를 체계적인 방법론을 통해 정보시스템의 데이터 모델을 창출하는 정규화 중심의 NS방법론을 적용한 설계방법과 기존의 관행에 의해 역정규화를 어쩔 도리 없이 용인하고 마는 정보공학(Information Engineering, 약칭 IE)방법론 기반의 설계방법과 비교한다. 동일한 업무를 대상으로 설계된 데이터 설계 산출물을 기준으로 각종 측면에서 비교해 봄으로써 정규화 중심 기반의 체계적 방법론이 기업의 정보시스템의 성능 향상에 어느 정도 기여하는지를 정량 파악하였다. 비교 실험 결과, 개체 설정의 타당성, 설계의 균형성 지표인 사각고리형 경로 유무, 그리고 속성중복률 면에서 모두 NS방법론적 설계가 우수하였다.

Abstract

As information systems to be computerized get significantly scaled up, data modeling issues apparently considered to be crucial once again as the early 1980's under the terms of data governance, data architecture or data quality. Unfortunately, merely resorting to heuristics-based field approaches with more or less no firm theoretical foundation of knowledge with regard to criteria of data design lead quite often to major failures in efficacy of data modeling. In this paper, we have compared normalization-critical data modeling approach, well-known as the Non-Stop Data Modeling methodology in the literature, to the Information Engineering in which in many occasions the notion of de-normalization is supported and even recommended as a mandatory part in its modeling nature. Quantitative analyses have revealed that NS methodology ostensibly outperforms IE methodology in terms of efficiency indices like adequacy of entity judgement, degree of existence of data circulation path that confirms the balancedness of data design and ratio of unnecessary data attribute replication.

Keywords : Data Modeling, Non-Stop(NS) Data Modeling Methodology, Information Engineering Methodology, De-normalization, Data Circulation Path

I. 데이터 모델링의 개요

최근 들어 자료 설계(data modeling: DM)라는 것이 큰 쟁점으로 부각되기 시작했다. 그러나 실무적으로 실행하기에는 여러 어려움을 겪게 된다. 그 중 하나는 데

이터 모델링에 대한 확고한 견해를 가지고 있지 못하기 때문이다. 무엇보다도 전산화를 원하는 실무 부서의 요구를 잘 파악하는 일이 중요한데, 이것이 잘 이루어지지 않는 경우가 많다. 만일 잘 파악이 된다 하더라도 다음 단계에서 이를 활용할 수 있는 형식이 갖추어지지 않는 경우가 많다. 요구가 잘 파악되면, 이를 업무기술서라는 정형화된 형태로 나타나야만 한다. 업무처리 규정을 완벽하게 준비하고 정비하는 일은 데이터 모델링을 성공적 수행할 수 있는 지름길이다.

* 정회원, 용인송담대학 컴퓨터게임정보과
(Yong-In Songdam College Dept. of Computer
Game Information)

접수일자: 2006년8월21일, 수정완료일: 2006년10월30일

두 번째 어려움은 업무에서 발생하는 데이터들 간의 관계에 대한 집중적인 분석이 미흡하다는 데에 있다. 또한 데이터 들인 개체가 갖는 속성들 간의 관계에 대한 집중적인 분석이 미흡한 점에도 그 원인이 있다. 개체 간의 관계를 파악한다는 것은 개체들 중에서 어느 개체가 더 영향력이 있는지 결정하는 것인데, 마찬가지로 속성들 간에도 어느 것이 더 결정력이 있는지를 파악하는 것이다.

세 번째 어려움은 관계형 데이터베이스를 설계하는 것을 전제로 했을 경우, 표준형 테이블 설계기술의 핵심에 대한 근본적 이해 부족이다. 데이터베이스 설계를 현실과 이상 두 가지 사이에서 가장이상적인 시스템을 구축하되, 현실적으로는 현장 감각에 맞게 쓸모 있게 설계를 해야만 한다. 너무 이상적으로 응답 시간은 고려하지 않는다는든지, 또 반대로 너무 현실에 치우쳐서 무작정 무제한적으로 기억공간을 사용하면서 응답시간만 빨리 할 수는 없는 것이다. 즉 기억공간도 적절히 사용하면서 동시에 응답시간도 만족스런 설계 결과가 나오도록 해야만 하는데, 테이블의 설계에 있어서 어떤 정규형이 모두를 만족시킬 수 있는지 잘 파악해야 한다.

대다수의 기업에서 구축해 놓은 데이터베이스 시스템을 분석해 본 결과 개체나 관계가 아닌 속성들이 많게는 전체의 속성 중 65퍼센트 이상 중복되어 있다는 사실을 알 수 있었다^[1]. 만일 기업에서 관리하는 데이터의 양이 방대하면 할수록, 불필요하게 중복된 데이터를 처리하는 비용 또한 무시할 수 없는 수준일 것이다. 외래 키로 인한 중복인 경우는 각 테이블 간의 조인을 위한 것이기 때문에 예외로 간주한다. 왜냐하면 이와 같은 종류의 중복은 전체 데이터의 아주 작은 부분에 불과하기 때문이다. 데이터 중복의 커다란 부분을 차지하고 있는 것은 데이터의 질의에는 아무런 도움도 되지 않는 불필요한 중복이다^[2].

정규화(normalization)는 속성 간의 불필요한 중복을 최소화할 뿐만 아니라 데이터의 무결성을 높이는데 기여하기 때문에 데이터베이스의 분석 및 설계에 있어서 매우 중요한 부분을 차지한다. 정규화가 다른 한편으로는 질의응답 시간을 최소화하려는 이중적 목표를 추구해야 하기 때문에 설계 과정상의 복잡성이 불가피하게 유발된다. 비록 정규화를 함으로써 데이터의 일관성을 향상시키는 효과는 볼 수 있지만, 일부에서는 자료 검색 시 조인을 추가적으로 해야 하기 때문에 질의 응답 시간 측면에서 성능을 저하시킬 수도 있다는 것을 제기

하고 있다^[3]. 조인은 응답시간을 지연시킬 수 있는 가장 중요한 요소 중 하나이므로 보다 정교하게 정규화된 테이블은 응답시간 측면에서 볼 때 성능이 저하시킬 수도 있기 때문이다. 이러한 이유로 비록 역정규화가 정규화 이론에는 반기를 들고 있지만, 성능을 향상시킬 수는 있다^[4].

역정규화의 목적은 질의응답 시간을 줄이는 효과를 얻을 수 있기 때문에 실제 업무에서 널리 적용이 되지만 이러한 역정규화 작업이 데이터베이스 설계자의 감각에 의존하여 아무런 기준 없이 이루어진다면 데이터베이스 시스템의 성능 또한 향상되리라 기대할 수가 없는 것이다. 따라서 시스템을 설계할 때 역정규화를 해서 데이터의 중복을 허용할 것인가, 아니면 데이터의 중복을 피할 것인가를 결정하기 위해 실제적이고 객관적인 비교 실험을 통해 성능 평가를 할 필요가 있다.

본 논문에서는 업무의 요구 형성 과정부터 개체관계도 도출까지 체계적인 방법을 통한 NS 방법론적 설계 방법^[5]과 관행에 의해 역정규화를 시도하는 정보 공학 방법론적 설계 방법과의 성능을 비교 하고자 한다. 성능 평가를 위해 인터넷서비스 업체의 도서 거래 업무 시스템에 대해 설명하고 이를 바탕으로 두 가지 방법을 적용하여 데이터베이스 시스템 설계에 있어서 가장 핵심 부분인 ERD를 설계한다. 동일한 업무기술서를 바탕으로 설계된 각각의 ERD를 4가지 측면에서 비교해 봄으로써 체계적인 절차를 통한 시스템 구축이 기업의 정보시스템의 성능에 기여하고 있는지 여부를 파악하고자 한다.

본 논문의 구성은 다음과 같다. 제 II장 관련 연구에서는 요구 형성 과정과 개략 설계와 상세 설계에 대한 간략한 개념 소개와 그 방법에 대해 소개하며, 제III장에서는 성능 평가를 위한 실제 표준 업무를 설정하여 자세한 업무기술서를 바탕으로 두 가지 방법에 의해 설계된 ERD를 소개한다. 또한 작성된 ERD 들을 기반으로 4가지 다른 측면에서 비교, 분석한다. 제 IV장에서는 본 연구의 결론과 한계점을 다루었다.

II. 관련 연구

데이터를 전산화할 때에는 두 가지 각도에서 전산화 요구를 하게 된다. 즉, 요구를 해오는 부서들을 의뢰자(client)라 할 때 전산 담당 부서는 제공자(server)역할을 하게 된다. 요구 형성 과정에서는 의뢰자의 역할이 커지며, 개략 설계 및 상세 설계 과정에서는 요구 형성

과정에서 도출된 자료를 가지고 제공자가 전적으로 역할을 맡아야 한다. 이전에는 처음 초기 작업인 업무 분석 단계부터 전산담당 부서에서 전담하여 의뢰자/제공자 역할을 모두 겸했지만 이제는 역할을 분담하는 정보 시스템 시대로 접어들게 된 것이다.

2장에서는 업무의 분석 단계인 요구 형성 과정과 개념적 설계 과정 위주로 연구하였다. 개념적 설계 과정에서 ERD를 도출할 때 개체나 관계, 또는 불필요한 속성들의 중복을 배제하였다면 제3정규형을 만족하므로 설계의 90%이상을 완성한 것으로 본다. 그러므로 논리적 설계 과정과 물리적 설계 과정은 상대적으로 개념적 설계 과정에 비해 업무의 비중이 그리 크지 않기 때문에 본 논문에서는 자료 설계의 핵심 부분인 개념적 설계에 초점을 맞추었다.

1. 요구 형성

어떤 형태의 의뢰자라 할지라도 요구의 형식은 "업무 기술서"라는 형태로 나타낼 수 있어야 한다. 한 가지 예를 들어 보면 "대학의 학사 행정 업무 시스템"^[6]에 관한 업무 처리 규정을 그림 1의 아홉 가지 항목으로 정리하여 나타낼 수 있다.

1. 대학은 여러 개의 단과대학으로 나뉜다. 각 단과대학은 학장이 반드시 한 명 있어야 하며 각 학장은 반드시 한 단과대학만 관리할 수 있다. 각 단과대학은 여러 개의 학과로 나뉘며 각 학과는 반드시 한 단과대학에만 소속되어 있다.
2. 각 대학은 여러 개의 학과로 구성된다. 예를 들어, 상경대학은 회계학과, 경영학과, 경제학과, 경영정보학과가 있다. 카디날리티 규칙은 한 대학은 적어도 한 학과가 있으며, 최대 학과 수는 정해져 있지 않다. 반대로 각 학과는 오직 한 대학에 속한다.
3. 각 학과에는 반드시 여러 개의 과목들이 제공된다. 예를 들어, 경영학과에서는 마케팅 원리, 생산경영, 경영의 소개와 같은 과목이 제공된다.
4. 각 학과는 같은 과목에 대하여 여러 반을 개설할 수 있다. 각 반은 주어진 장소에서 주어진 시간에 한 교수에 의해 가르쳐진다. 그러나 대학 카탈로그에는 현재의 강의 일정에 없는 과목을 가질 수 있다. 따라서 CLASS는 COURSE에 선택적이다.
5. 각 학과는 반드시 여러 명(1명 이상)의 교수들이 있고, 각 교수는 반드시 하나의 학과에만 속한다. 각 학과는 단 한 명의 학과장이 있어야 하는데, 여

- 기서 모든 교수가 다 학과장이 될 수 없다. 따라서 DEPARTMENT는 PROFESSOR에 선택적이다.
6. 각 교수는 4반까지 강의할 수 있으며, 어떤 교수는 그 학기에 강의를 안 할 수도 있다. 한 반은 단 한 사람의 교수에 의해서만 강의된다.
 7. 한 학생은 반드시 등록을 하되 7개의 반까지 등록할 수 있고 한 반은 등록 기간에 반드시 한 번만 선택될 수 있다. 각 반은 40명까지 등록할 수 있다.
 8. 각 학과에는 여러 명(최소한 1명)의 학생이 있으며 각 학생의 전공은 한 학과에만 속해야 한다.
 9. 각 학생은 반드시 소속된 학과의 지도 교수가 있다. 각 지도 교수는 여러 학생을 지도할 수 있으나 반드시 지도해야 되는 것은 아니다.

그림 1. 대학의 학사 행정 정보 시스템을 위한 요구형성

Fig. 1. Requirements Formulation for A University Academic Affairs Information System.

의뢰자 마다 다양한 요구를 해오는 경우에는 우선 요구들을 취합하는 일이 필요하다. 중복되는 것이 없도록 최적화하는 작업과 전체적으로 일관성을 유지하는 표현되도록 조정하는 작업이 수반된다. 다양한 요구들을 취합하는 과정을 거친 후 최종적으로 단일화된 업무기술서가 나오게 되는데, 이 단일화된 규정을 토대로 다음 단계의 요구 분석이 적용될 수 있는 것이다. 의뢰자의 요구 형성과정을 전산 부서에서 주도를 한다면, 다음 단계에서 데이터 간의 "관계"가 자연스럽게 이루어질 수 있도록 유도할 수 있다.

요구 형성 과정에서 업무기술서를 작성할 때, 제공자가 의도적으로 의뢰자 에게 협조를 구할 수도 있을 것이다. 즉, 실무를 담당하는 의뢰자와의 면담 및 설문문을 통하여 조목조목 업무에 대한 분석 결과를 문서 형태로 작성하는 일이 필요하다. 이러한 업무기술서가 최종적으로 완성이 되면 다음 단계로 제공자에 의한 개략설계에 착수한다.

2. 개략 설계

요구 형성 과정에서 업무기술서가 도출되면 이를 토대로 개략 설계에 들어간다. 개략적 설계란 개념적 설계를 말하며 정형화된 도구가 없기 때문에 데이터베이스 설계자의 역량에 따라 설계의 성공 여부가 좌우된다. 개념적 설계의 목표는 요구 형성 과정에서 밝혀낸 의미를 도식화하는 것인데 이러한 의미 도출 작업이

쉽지 않다.

개념적 설계 단계에서는 업무기술서를 바탕으로 개체와 개체 간의 관계를 나타내는데 초점을 둔다. 개체 간의 연관성을 보통 "관계(relationship)"라고 일컫는다. 이러한 개체 간의 관계를 표현하기 위한 도구로서 "개체 관계도(Entity-relationship diagram: ERD)를 주로 이용한다. 요구 형성 과정에서 도출된 주요 데이터들에 대해서 관계를 추가한 것이 바로 개념적 설계의 결과물이다.

개념적 설계의 결과물이 타당한 것인지는 이전 단계의 업무기술서와의 상용성을 점검해 봄으로써 알 수 있다. 만일 상용하지 않는 부분이 한 군데라도 있을 경우에는 반드시 1대 1로 검토하여 대응이 되도록 해 놓아야 한다. 만약 업무처리 규정이 ERD로 대응되는 과정에서 오류가 발생했다면 즉각 업무처리 규정과 ERD양자 간의 양방향 대응 관계를 명확히 짚고 넘어가는 작업을 반복적으로 수행할 수도 있다. ERD의 효용성은 다음 두 가지 면에서 확인된다. 첫째, 문장 형식의 업무기술서를 그림 형태로 나타내어 무엇인가 흘러 들어가고 흘러 나오는 모양으로 업무를 이해한다는 의미에서 충분한

가치가 있다. 둘째, 데이터베이스의 구현 단계인 논리적 설계단계에서 반드시 필요로 하는 정규화(normalization)된 테이블을 만들기 위한 전 단계로서 ERD가 요구되기 때문이다.

동일한 요구 분석 결과물을 토대로 개념적 설계를 했더라도 설계자의 주관에 따라 서로 다른 설계 결과가 나올 수 있다. 그것이 잘못된 일은 아니지만 설계에 있어서 너무 개인의 주관이 개입되면 아주 편이한 설계 결과가 나올 수 있으므로 향후 적응력 등을 고려해서 설계가 바람직하게 되었는지를 살펴봐야 한다. 예를 들어, 제품 판매 체계에서 개체와 관계를 도출해 보면 그림 2와 그림 3에서와 같이 두 가지 방향으로 설계를 할 수 있다. 그림 2에서는 수강이라는 항목을 개체화하였고, 그림 3에서는 관계로 설계하였다.

두 가지 경우 모두 타당한 설계라고 할 수 있다. 그림 2에서는 학생, 교수, 수강, 과목이라는 4개의 개체와 학생과 수강 간에는 1:다 관계가 성립하며, 과목과 수강 간에도 역시 1:다 관계가 성립한다. 교수와 과목 간에는 1:다 관계가 존재한다. 그림 3에서는 학생, 과목, 교수라는 3개의 개체와 3개의 관계로서 수강, 강의, 지도가 있다. 이 두 가지 설계의 예로 볼 때 업무기술서에서 도출된 업무처리 규정에 어느 설계가 더 잘 어울리는지 잘 따져본 다음에 판단을 해야 한다.

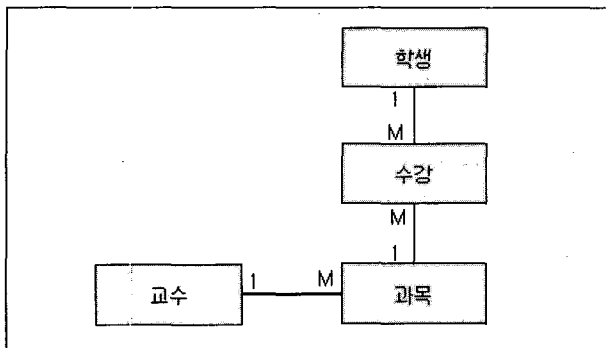


그림 2. 수강을 개체로 설정한 설계
Fig. 2. Conceptual Data Modeling In Case Enrollment Treated as Entity.

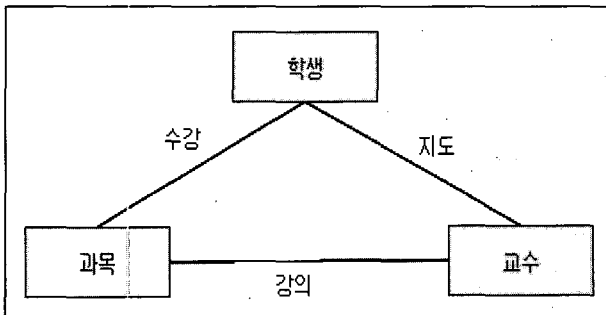


그림 3. 수강을 관계로 설정한 설계
Fig. 3. Conceptual Data Modeling In Case Course Registration Treated as Relationship.

III. 업무 설정 및 방법 비교

현재 인터넷상에는 수많은 전자 상거래 사이트가 존재하지만 시스템의 결함 등으로 인해 서비스를 중단하는 경우가 종종 있다. 또 데이터가 정확하지 않든지, 서비스 응답 속도가 너무 느려 사용자가 사용하기에 어려움을 겪곤 한다. 따라서 본 논문에서는 다양한 전자 상거래 중 인터넷을 이용한 도서 판매 업무 시스템을

인터넷 도서 판매 업체의 영업 정보 시스템은 웹을 이용하여 도서를 판매하는 업무를 주로 하고 있다. 고객 등록 및 입금 확인 등 고객에 관련된 업무와 도서를 출판사로부터 공급받아 공급하는 업체에 관련된 업무로 분류되며, 고객이 주문한 도서를 배송하는 배송 관련 업무를 포함한다.

그림 4. 웹 도서 판매 업무 시스템의 표준 업무기술서 일부 내용

Fig. 4. Excerpt of Standard Job Descriptions for Web-Based Bookstore Sales Information System

설정하여 체계화된 데이터베이스 설계 방법을 사용한 경우와 현장 경험에 의한 방법으로 설계한 방법의 효율성을 비교 분석한다. 인터넷 도서 판매 업무 시스템은 그림 4와 같은 체계를 가지고 있다.

데이터베이스를 구축하기 위해 적용하는 과정 중 데이터베이스 설계자의 역량에 따라 영향을 많이 받는 부분이 개념적 설계 과정이다. 요구 형성 과정의 결과물로 산출된 업무 기술서는 의뢰자의 요구를 전산 실무자의 시각에서 현실 업무를 요약해 놓은 상태이므로 특별히 기술적인 면이 요구되지는 않는다. 단지 업무 담당자와 전산 담당자가 업무의 처리 규정 등 업무 내용을 치밀하게 분석해야 하는 노력이 필요하다. 그러나 개념적 설계 과정을 통해 산출해야 하는 ERD는 마치 병아리 감별사가 갓 태어난 병아리의 성별을 순식간에 파악 하듯이 설계자가 업무기술서를 토대로 개체와 개체 간의 관계를 구별해 내야만 한다. 개체와 관계를 구분해 내는 정형화된 도구가 없기 때문에 데이터베이스 설계 과정 중 제일 과정이라 할 수 있다. ERD가 나온 다음, 이를 바탕으로 논리적 설계 단계를 수행해야 하는데, 이 때 정규화 과정을 거쳐야 한다. 정규화 과정은 데이터베이스 설계 시 데이터가 불필요하게 중복되는 것을 막고, 질의에 대한 응답 시간을 줄이기 위한 수단이다. 최적화된 ERD를 얻었다면 이미 제3 정규화까지 만족한 단계까지 갔다고 할 수 있으므로, 정규화 작업은 그리 큰 의미를 갖지 못한다. 즉, 가장 힘들고, 설계자의 역량이 많이 요구되는 부분이 개념적 설계 단계이기 때

회원등록 및 주문

- 인터넷에서 도서를 구입하려면 반드시 회원으로 등록을 해야 구매가 가능하다.
- 인터넷의 회원 등록 정보에는 회원번호, 주민등록번호, 주소, 전화번호, 전자메일ID, 휴대폰번호, 결제방법이 있다.
- 각 회원마다 담당 직원이 관리를 하며 주문 완료시 까지 모든 절차를 관리한다.
- 등록을 마친 회원은 구매할 도서목록을 선택하고, 인터넷 주문서 양식에 주문 내용을 입력하여 주문한다.
- 주문서 양식에는 주문목록에 대한 상세정보와 주문일자, 주문수량, 결제가격, 배송비가 있다.
- 등록된 회원에게는 도서를 구매하는 실적에 따라 구매 포인트를 부여하여 다음 구매 시 혜택을 보게 한다.

- 포인트 부여 기준은 구매가의 1%로 한다. 회원이 책을 구매하기 위해 체크하였으나 당장 구매하지 않고 다른 시간에도 구매가 가능해야 한다.

공급업체와출판사의거래

- 도서를 공급하는 공급업체는 출판사와 별도의 계약을 맺고 있으며 계약을 맺지 않은 출판사와는 거래를 하지 않는다.
- 한번 계약을 맺었다가 해지할 수 있으며 다시 계약을 맺을 수도 있다. 이때 이전 계약 상태에 대한 정보도 필요하다. 계약의 갱신은 여러 번 가능하다.
- 계약서에는 계약번호, 계약일자, 최저가비율, 계약상태코드가 포함된다.
- 출판사 별로 원가에 대한 최저가 비율을 지정하여 비율에 의해 산정된 최저 금액 이하로는 도서를 판매할 수 없다.
- 계약 수정에 의해 최저가 비율이 변경 가능하다.
- 출판사 정보는 출판사사업자번호, 출판사명, 출판사주소, 출판사전화번호, 거래은행코드, 계좌번호를 관리한다.
- 공급업체에 대한 정보는 공급업체사업자번호, 대표자명, 대표자전화번호, 팩스번호, 업체주소가 포함된다.
- 관리하고자 하는 도서의 정보는 도서번호, 도서명, 도서구분, 수량, 공급단가, 저자, 발행일자, 정가, 적용할인율이다. 적용 할인율은 출판사별 최저가 비율을 초과할 수 없다.
- 출판사에서 신간 서적을 발간하면 공급받아서 인터넷에 신간 광고를 한다.
- 인터넷에 접수되는 도서 주문을 받아 재고를 파악하고, 만약 재고가 남아 있지 않은 경우에는 도서를 공급한 회사에 연락을 하여 공급받는다.

공급제휴

- 계약 담당자는 여러 공급 업체 중 제품을 납입 받을 공급 업체와 계약제목, 제휴일, 제휴 만기일, 제휴 조건에 합의하여 제휴를 맺는다.
- 선정된 공급업체의 사업자번호, 대표자명, 대표자전화번호, 팩스번호, 업체주소를 요청하고 납품 받을 때, 또는 문제가 발생했을 때를 대비해 기록한다.
- 계약자는 담당 직원으로 하며 사원은 사원번호, 주민등록번호, 사원명, 직위, 소속부서, 연봉, 사원전화번호, 전자메일ID, 사원휴대폰번호, 입사일의 정보를 가진다.

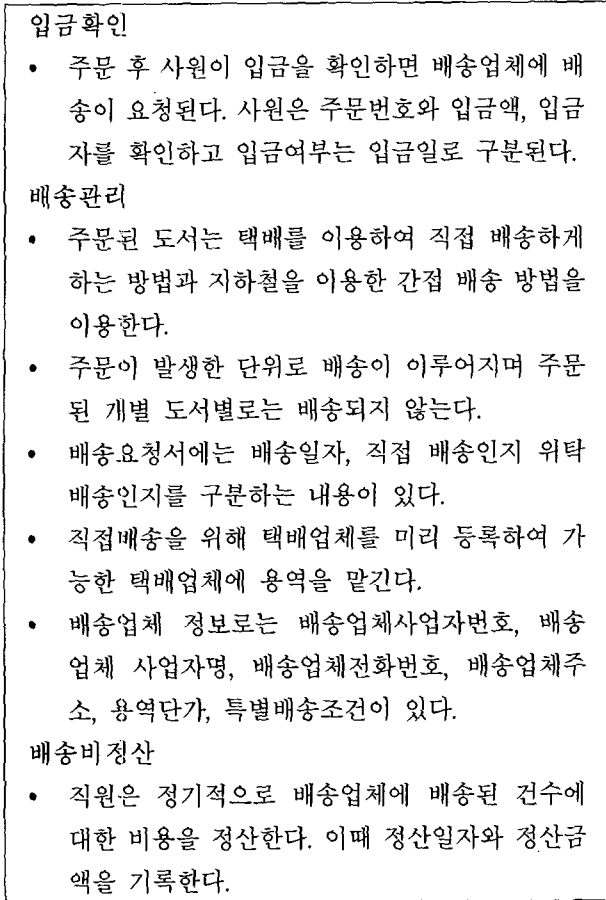


그림 5. 웹 도서 거래 업무에 관한 업무기술서
Fig. 5. Job Descriptions for Web-Based Bookstore Transactions.

문이 산출된 ERD에 따라 데이터베이스의 성능이 좌우되기 십상이다. 그래서 시스템 성능 측면에서 보다 객관적인 판단을 위해 최적화된 ERD를 바탕으로 한 테이블과 역정규화를 허용한 테이블과의 비교 실험이 필요하다.

1. 업무 기술서 및 업무 환경

인터넷을 이용한 도서 거래 업무는 크게 공급 업체, 인터넷 쇼핑몰 업체, 배송 업체, 구매 고객 관리 업무로 구분한다. 업무 분석 결과 그림 5와 같은 업무기술서가 산출되었다.

2. 개체 관계도 비교

그림 5에서 기술한 업무기술서를 바탕으로 그림 6과 그림 7을 설계하였다. 동일한 업무기술서를 가지고 각기 다른 방법으로 설계된 ERD를 가지고 개체 설정 측면, 개체와 관계의 연결 측면, 그리고 고리 형 경로 측면에서 비교해 보았다. 그림 6은 실제 인터넷 쇼핑몰인

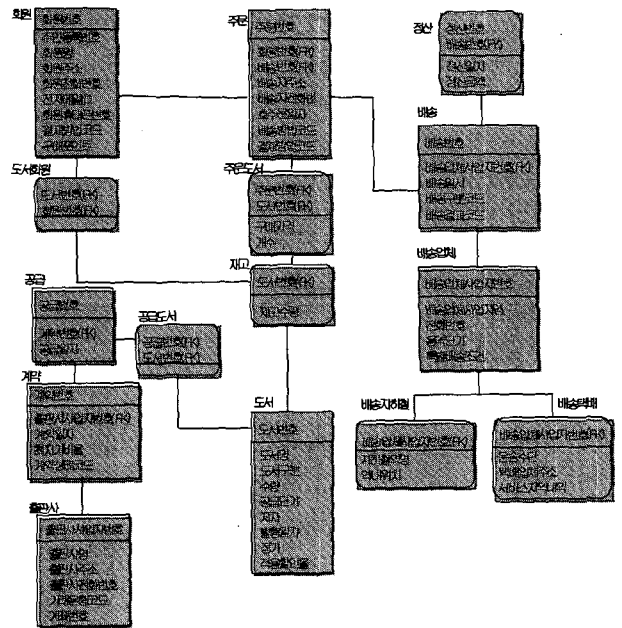


그림 6. 정보공학 방법론을 이용하여 설계한 ERD
Fig. 6. ERD Outcome on Basis of The Information Engineering Methodology.

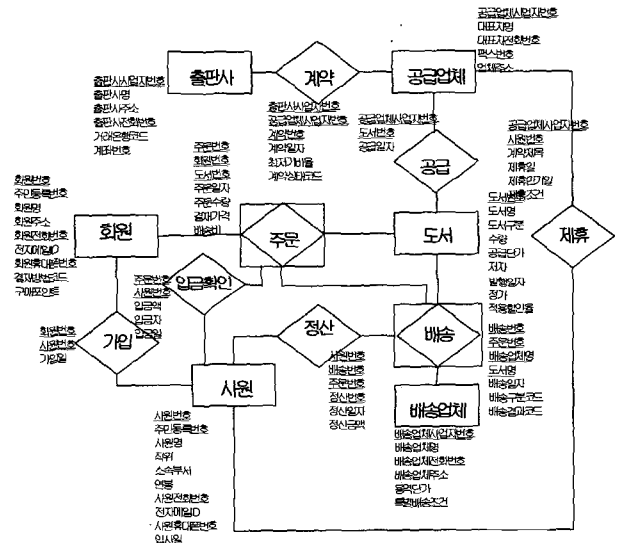


그림 7. NS방법론을 이용하여 설계한 ERD
Fig. 7. ERD Outcome on Basis of The Non-Stop Data Modeling Methodology.

N사의 관행에 의한 정보 공학 방법론적 ERD로서 그림에서 네모 박스는 개체를 의미하며 둥근 네모 박스는 행위를 나타낸다. 그림 7은 NS방법론적 ERD로서, 네모 박스는 개체를 의미하며, 마름모 모양은 두 개체 간의 행위를 나타낸다.

1) 개체 설정 측면에서의 비교

그림 7에서 보면 6개의 주요 개체가 뚜렷하게 나타나

있어서 업무의 전체적인 윤곽을 명료하게 표현하고 있다. 영업 정보 시스템은 인터넷을 이용하여 도서를 판매하는 업무이기 때문에 고객 관련 업무와 도서를 출판사로부터 공급받아 공급하는 업무와, 그리고 고객이 주문한 도서를 배송하는 배송 관련 업무로 나뉘어 진다. 특히 회원은 도서를 주문하고 사원이 회원, 도서공급업체 그리고 배송업체를 관리한다. 사원이 업무의 중추적인 역할을 하게 되며 모든 업무에 관련이 되어 있다. 즉, 회원, 사원, 그리고 도서 이 세 개체를 중심으로 각 개체 간의 행위 위주의 관계들이 형성됨으로써 ERD가 완성되었다.

그러나 그림 6에서는 8개의 개체로 구성되었는데, 우선 개체와 개체간의 관계가 일관성이 없이 연결되어 있으며, 회원 관리, 배송업체와의 업무, 그리고 서적 공급업체와의 제휴 업무가 전혀 언급되어 있지 않다. 특히 업무의 핵심 주체가 되는 사원을 개체로 설정하지 않았기 때문에 회원, 배송, 그리고 공급업체 업무를 처리할 핵심 주체를 알 수 없게 된다. 즉, 회원 가입, 주문한 서적에 대한 입금 여부 확인 작업 등이 어느 사원에 의해 체계적이고 책임감 있게 처리되는지 알 수가 없다. 마찬가지로 서적 공급업체와의 제휴관계, 배송업체와의 정산관계를 누가 담당하는지 전혀 알 방법이 없다.

그림 6에서 개체 '주문'과 '도서' 사이에 두 개의 관계가 설정되어 있는데, 회원이 주문한 서적에 대한 구매가격, 개수는 개체 주문의 속성으로 들어가야 한다. 또한 재고량을 저장하기 위해 따로 관계를 설정하였는데, 이는 응용프로그램을 이용하여 얼마든지 알아 낼 수 있는 정보이기 때문에 불필요하다. 관계가 많을수록 데이터베이스 구현 단계에서 테이블의 개수가 늘어나는 것은 자명하기 때문에 반드시 개체 간의 행위 위주의 관계를 설정해야만 한다.

도서 배송업무에서 그림 6은 배송 수단에 따라 개체 배송업체에 2개의 관계를 연결해 놓았다. 배송 수단은 배송업무의 속성으로 구성될 수 있으며 따로 분류할 필요가 없다. 또한 도서를 공급하는 공급업체와 출판사와의 계약 관계만 존재 할 뿐 막상 인터넷 서점과 공급업체와의 관계는 설정되지 않아서 이에 대한 자료는 저장할 수가 없다.

2) 개체와 관계의 연결 측면에서의 비교

그림 7은 철저하게 업무기술서에 근거한 개체와 개체 간의 행위 위주의 설계를 하였다. 또한 선 후행 관계를 표현해 놓음으로써 업무의 흐름을 체계적으로 수행

할 수 있도록 하였다. 예를 들어 개체 '주문'은 사원이 회원으로부터 도서 대금이 입금이 되었는지 확인한 후에 배송 업무에 돌입을 할 수 있도록 하였다. 그러나 그림 6은 회원이 도서를 구매할 경우 대금 결제 수단만 속성으로 설정되었을 뿐, 대금의 입금 여부를 확인 할 수 있도록 설계되지 않았다. 그러므로 일일이 회원이 입금했는지는 수작업에 의해 확인 한 후에 배송을 하게 되므로 업무의 흐름이 끊어지게 되며, 또한 업무의 선 후행 관계는 나타낼 수 없게 된다.

그림 6은 개체와 관계의 연결 방법이 일관성 없이 기존의 관행대로 업무의 편리성에 의해 결정되었다. 즉, 개체 '주문'과 '도서' 사이에 두 개의 관계가 연결되었는가 하면 도서를 공급하는 공급, 계약, 출판사 간에는 개체끼리 무작정 연결하고 있다.

3) 고리 형 경로 유무측면에서의 비교

그림 7을 살펴보면, 회원/가입/사원/입금확인/주문으로 이어지는 고리와 같은 경로가 여러 개 존재한다. 주문/입금확인/사원/전산/배송과 같은 경로와 '사원'과 '공급업체', 그리고 '도서' 간의 고리 형 경로를 찾아볼 수 있다. 즉, ERD자체가 여러 개의 고리 형 경로로 구성되어 있다고 할 수 있다. 그러나 그림 6에서는 고리 형 경로를 단 한 곳에서만 찾아볼 수 있다. 이와 같은 고리 형 경로가 존재하는 ERD를 일컬어 '균형 잡힌' ERD로 부를 수 있는데, 고리 형 경로가 존재하지 않는 ERD는 한 방향으로 치우친 편향적 모습을 가지게 되므로 편향나무 구조(skewed tree)가 균형나무 구조(balanced tree)에 비해 상대적으로 지니는 단점을 그대로 다 지니는 부담이 있다. 고리 형 경로가 많을수록 보다 나은 성능을 기대할 수 있다.

질의 처리 측면에서 고리 형 경로는 한 방향으로 치우친 경로에 비해 항상 일정한 응답 시간을 보장한다. 편향나무 구조인 경우 질의를 처리할 때 최악의 경우 끝에서 끝까지의 조인을 통해서 결과를 얻을 수 있기 때문이다. 예를 들어 그림 6이나 그림 7에서 주어진 도서명에 대한 출판사명을 검색하고자 할 때 개체 도서부터 출판사까지 5개의 테이블을 조인해야 그 결과를 얻을 수 있게 된다.

4) 속성 중복률 비교

속성 간의 불필요한 중복은 데이터베이스의 일관성 유지에 큰 장애가 된다. 본 논문의 앞에서도 언급한 바와 같이 데이터베이스 설계의 세 번째 단계인 논리적

표 1. 속성 중복률 비교
Table 1. Comparison of Attribute Redundancy Ratio between IEM and NSM.

비교 관점 \ 방법론	그림 6 정보공학방법론	그림 7 NS 방법론
중복되는 불필요한 테이블	26.7%	0%
중복되는 속성의 개수	14%	0%
외래 키로 인한 중복 속성 개수	21.1%	17.5%

설계 단계에서는 정규화 작업을 통하여 데이터의 불필요한 중복을 배제하고 있다. 본 논문에서 설정한 업무에 대한 속성 중복을 표 1과 같이 나타낼 수 있다.

외래 키로 인한 중복 속성 개수 21.1% 17.5% 정보공학 방법론을 기초로 한 그림 6인 경우 전체 개체와 관계의 개수가 15개인데, 불필요한 것은 모두 4개이다. 개체 '주문'과 '도서' 사이의 관계 '주문도서'에서 속성 구매가격과 개수는 '주문'에 포함되어야 하기 때문에 삭제될 수 있고, 관계 '재고'에서 속성 재고수량은 응용 프로그램을 통하여 얻을 수 있는 정보이기 때문에 제거될 수 있는 것이다. 개체 '회원'과 관계 '재고' 사이의 '도서회원'은 속성 도서번호, 회원번호로만 구성되었는데, 개체 '주문'에 회원이 주문할 도서번호를 추가한다면 존재할 필요가 없다. 관계 '공급도서'도 개체 '공급'에 공급할 도서번호를 추가한다면 불필요하다. 그러나 그림 7에서는 모든 개체가 관계들로 철저하게 연결되어 있어서 불필요한 개체나 관계는 찾아볼 수 없다. ERD에서 도출된 개체, 관계를 모두 테이블로 변환할 수 있는 제3정규형을 만족하기 때문이다.

그림 6에서는 이미 불필요한 개체나 관계 4개가 존재하기 때문에 전체 속성 71개 중에서 이들 4개의 관계를 구성하는 속성 10개가 중복되었다. 그러나 그림 7에서는 제거 할 개체나 관계가 없기 때문에 불필요한 속성 또한 찾을 수 없다.

외래 키로 인해 중복된 속성의 개수는 그림 6인 경우 15개로 전체의 21.1%를 차지하였고, 그림 7인 경우 관계 8개에 대한 외래 키로서 17개의 속성이 중복되어 전체 97개의 속성 중 17.5%를 나타내었다.

실제 중소기업 영업정보시스템에서 도출된 속성의 개수는 대략 1000개에서 1500개이므로 본 논문에서 설정한 업무의 범위는 약 1/10 정도 이므로 본 논문의 연구 결과에 대해서 신뢰도가 있다고 본다. 또한 그림 7

은 그림 6에서 간과하지 못한 업무도 모두 포함시켰기 때문에 전체적인 업무의 범위가 훨씬 광범위한데도 표 3.1과 같은 성능을 유지한다는데 큰 의미가 있다. 즉, 개체 '사원', '공급업체', 그리고 관계 '가입', '제휴', '입금확인'은 그림 6에서는 고려하지 않은 부분들이다.

IV. 결 론

본 논문에서는 흔히 실제 업무에서 데이터베이스를 설계할 때 관행적으로 하고 있는 역정규화의 효용성을 측정하기 위해 NS방법에 의한 설계 방법과 비교 분석하였다. 업무의 요구 형성 과정을 통해 도출된 업무기술서를 토대로, 체계적인 방법을 통한 NS 방법론과 관행에 의한 정보 공학 방법론으로 설계된 ERD를 가지고 개체설정, 개체와 관계와의 연결, 고리 형 경로 유무, 그리고 속성 중복률 측면에서 비교 분석하였다. 비교 분석을 위해 인터넷 서비스 업체의 도서 거래 업무 시스템에 대해 설명하고 이를 바탕으로 두 가지 방법을 적용하여 데이터베이스 시스템 설계에 있어서 가장 핵심 부분인 ERD를 설계하였다. 이 비교 분석을 통해 관행적인 절차에 의한 정보공학 방법론적 시스템 구축이 기업의 정보시스템의 성능에 기여하고 있는지 여부를 파악하고자 하였다.

비교 실험에 의하면 개체의 설정에 있어서 정보공학 방법론인 경우 반드시 주체가 되어야 하는 개체들이 누락이 되었으며, 개체들 간의 관계에 있어서도 철저히 행위위주가 아닌 주 개체의 보안을 위한 수단으로 관계를 설정해 놓았다. 필요하지 않은 관계로 인해 속성들이 중복되는 현상을 발생시켰으며, 외래 키로 인한 속성 중복도 NS방법론에 의한 설계보다 높았다. NS 방법론적 ERD에서는 ERD자체가 제3정규형을 만족하기 때문에 속성 간의 불필요한 중복은 없었고, 외래 키로 인한 속성 중복도 정보공학방법론적 ERD보다 적게 나타났다. 사실 외래 키로 인한 속성 중복은 불가피한 것으로 대부분의 기업에서 약30%정도를 상회하고 있는 실정이므로 실험에서 나타난 수치는 지극히 자연스러운 현상이라 할 수 있다. 그러나 외래 키에 의한 속성 중복이 아니라 불필요한 개체나 관계들의 중복이 큰 문제를 야기 시킨다. 이는 속성 연결이 되기 때문이다. 고리 형 경로는 NS방법론적 ERD인 경우 여러 곳에서 나타나고 있다. 그러나 정보공학 방법론적 ERD인 경우 단 한 곳에서만 발견할 수 있었다. 고리 형 경로가 많이 나타나도록 시스템을 설계하면, 데이터베이스의 생명인 검색

요구에 대한 일정한 응답시간을 유지할 수 있으므로 NS방법론적 ERD가 보다 효율적이라 할 수 있었다.

본 논문에서는 대상업무가 실제 인터넷 도서 판매 업무 시스템을 대상으로 하였으나, 시스템을 보다 간략화 하여 A/S, 반품, 교환 등의 부수적인 업무들을 포함시키지 않았다. 실험에 사용된 속성들도 가상으로 생성한 것으로 실제 업무 상황과는 다소 다를 수 있다. 본 논문에서 수행한 비교 연구를 많은 현장 업무에 적용한 후의 보다 객관화된 정량적, 정성적 분석과 비교 평가가 기대된다.

앞으로 수행되어야 할 연구는 시스템의 규모를 확대 시켜서 전체 정보시스템을 구축한 상태에서 다수의 이용자들이 자유자재로 시스템을 이용할 수 있는 여건을 조성하여 시스템의 성능 비교를 하는 것이다. 보다 현실감 있는 시스템의 성능 분석을 위해 속성의 개수를 증가시킨 후 다양한 질의에 대한 응답시간 분석 및 저장 공간의 이용 측면에서의 분석이 필요하다.

참 고 문 헌

- [1] S. Moon, "Unclassified data is merely garbage: data modeling is more crucial than programming," Hitech Information, vol. 14, pp. 50-51, 2003.
- [2] R. Y. Wang, V. C. Storey and C. P. Firth, "A framework for analysis of data quality research," IEEE transactions on Knowledge and Data Engineering, vol. 7, no. 7, pp. 623-640, 1995.
- [3] G. L. Sanders and S. Shin, "Denormalization Effects on Performance of RDBMS," Proceedings of the 34th International Conference on System Sciences, Hawaii, pp. 1-9, 2001.
- [4] D. B. Bock and J. F. Schrage, "Denormalization guidelines for base and transaction tables," ACM Special Interest Group on Computer Science Education, vol. 34, no. 4, pp. 1, 2002.
- [5] 문송천, "데이터 아키텍처, 형설출판사," 2004
- [6] Peter Rob, Carlos Coronel, "Database Systems: Design, Implementation, & Management, 4/e", Thomson, 2000.

저 자 소 개



이 혜 경(정희원)

1979년 2월 숭실대학교 전자계산학과 졸업
 1985년 4월 University of Illinois (Urbana-Champaign) 전산학과 석사
 2000년 2월 성균관대학교 정보공학과 박사
 1988년 3월~1989년 2월 국립천안공업전문대학 전자계산과 전임강사
 1992년 3월~2001년 8월 경인여자대학 멀티미디어정보 전산학부 조교수
 2001년 9월~현재 용인송담대학 컴퓨터게임정보과 부교수

<주관심분야 : 데이터베이스, 데이터 모델링, 동시성제어>