

대학 취업정보시스템에서 데이터모델링 기본이론 경시에 관한 연구

이혜경*, 김희완**

A Study on Negligence of Data Modeling Fundamentals at the University Job Information System

Hae-Kyung Rhee*, Hee-Wan Kim**

요약

본 논문은 대학 취업정보시스템에서 데이터베이스 시스템이 얼마나 잘 운용되지 않는지 살펴보고자 하였다. 왜냐하면 현업에서는 데이터 디자인 기본 원칙에 미달하는 점들이 보이고 있기 때문이다. 다행히도 우리의 연구를 위해 대학취업지원 데이터베이스를 선택할 수 있었다. 본 연구에서는 현행 시스템으로써 대학 취업프로그램 이수시스템을 선택하였고 데이터 모델링에 있어서 기본 원칙을 엄격하게 따르는 새로운 대학취업지원 데이터베이스라는 이름으로 개발해 보았다. 기존의 대학취업지원 데이터베이스가 새로운 대학취업지원 데이터베이스에서 발견할 수 있는 표준 균형성으로부터 얼마나 기준에서 벗어나 있는지 비교해 볼 수 있었다. 데이터 비만도 관점에서 볼 때 기존의 대학취업지원 데이터베이스가 얼마나 많은 불필요한 데이터를 내포하고 있는지 계산해 보았다. 새로운 대학취업지원 데이터베이스는 불필요한 데이터를 전혀 포함하지 않았다. 즉, 데이터 모델을 형성하기 위해 필요한 데이터들만으로 구성되어 있다는 것을 의미한다. 그러나 기존의 대학취업지원 데이터베이스 시스템의 비만도는 50%가 넘었다. 그러므로 본 논문에서 제안한 새로운 대학취업지원 데이터베이스 시스템의 데이터 비만도가 19%로 나타나 기존의 시스템의 비만도가 37%나 높다는 것은 매우 중요한 차이인 것이다.

▶ Keywords : 데이터 모델링, 데이터 비만도, 데이터베이스 시스템, 취업 프로그램

Abstract

In this essay, we hoped to see how the operational university job database systems perform badly, since there are a lot of signs that they fall short of data design fundamentals. Fortunately, we are able to choose a university employment support database(UESD) application as a candidate for our study. We have taken UESD system as it is and we have developed an enhanced version of

•제1저자 : 이혜경 •교신저자 : 김희완

•투고일 : 2014. 6. 18, 심사일 : 2014. 7. 7, 게재확정일 : 2014. 8. 19.

* 용인송담대학교 컴퓨터게임과(Dept. of Computer Game, Yongin Songdam College)

** 삼육대학교 컴퓨터학부(Division of Computer Science & Engineering, Shamyook University)

the UESD system, which strictly followed the basic fundamentals in data modeling and named it as NewUESD system. Then, we have conducted a comparison to see how much UESD system is deviated from the standard balancedness found in NewUESD. From the perspective of data obesity, we computed how much UESD contains unnecessary data. The NewUESD contained no unnecessary data at all, which implies that it has only data that is necessary to form a data model. But the data obesity of UESD system is found to be more than 50%. Therefore, it is significantly higher, 37% higher in this case, than the obesity judged to be normal, which is 19% in NewUESD system.

▶ Keywords : Data Modelling, Data obesity, Database System, Job Program

I. 서 론

1. 연구배경

현업에서 운영 업무 전체를 구성하는 크고 작은 업무들은 하나의 유기체처럼 연결돼 있다. 어느 한 데이터에서 출발하여 다른 목표 데이터로 향하여 목표지점에 도달하기까지의 경로를 Data Path라고 부른다면, 업무 세부 절차 관련 데이터 경로가 모여 단위 업무 데이터 지도를 형성하게 된다. 하지만 데이터 지도가 잘못돼 목표지점을 표시도 못한 채 나타난다면 업무의 운영은 혼선을 빚을 수밖에 없을 것이다. 과학적 업무 운영을 위해서는 불필요한 정보를 빼고, 체계적인 데이터 지도를 수립하는 과정이 필수적이다.

교통지도가 있어야만 내비게이터라는 자동화 도구를 제작할 수 있듯이 현업에서도 데이터 지도가 존재해야만 업무자동화 내비게이터라는 자동화 도구를 제작할 수 있다. 업무자동화 내비게이터를 '데이터 모델'이라고 부르고 있으나 심각한 문제는 바로 여기에 있고 각종 문제의 출발점이 그 모델에 기인하곤 한다.

오늘날 현업에서 보유하고 있는 정보 시스템의 데이터 모델은 기본 수준에서 벗어난 정도가 도를 넘은 수준으로 심각하다. 교통지도는 물론, 데이터 지도도 지도이므로 지도의 본질과 성격의 상식을 준수해야만 지도로서의 자격이 있는 것이다. 교통지도에는 특정 지명(地名)이 지도 전체에 한 번만 등장하는 것이 정도(正道)이므로 여러 번 이곳저곳에 나타나는 안 된다.

기본에서 보면 오늘날 기업이 보유하고 있는 데이터 지도들은 같은 지명이 지도 전체에 수시로 편의에 따라 중복 등장하는 모습을 지니고 있어서 도저히 지도라고 부를 수조차 없는 형태로 되어 있다. 데이터가 쓸데없이 중복되어 있게끔 나타난 모양의 심각성 수준을 데이터 비만도로 측정할 수 있다.

쓸데없는 데이터 중복은 쓸데없는 업무 절차 중복과 의미가 같다. 하나의 업무로 포장돼 단위 업무로 관행적으로 용납되어 있어서 그렇지 만약 업무 리엔지니어링을 실시한다면 반드시 제거해야 할 비만 요소들인 것이다. 이렇게 보면 업무 간 세부 절차상의 중복은 데이터 지도라는 현미경 프리즘을 통해서만 가시화될 수 있는 것이지 그렇지 않고는 절대로 면화되기 불가능하다.

데이터가 중복 없이 정확성과 신속성을 겸비한 품질 좋은 데이터베이스를 개발하기 위한 개발 방법론에 대한 연구[1]에서는 데이터베이스 설계 과정에서 세심하게 데이터를 수집하며 데이터베이스를 설계하는 과정을 소개하고 있지만 데이터 중복에 대한 중요성은 잠깐 동안 언급만 했을 뿐 어떻게 설계를 해야 중복을 피하는 품질 좋은 데이터베이스를 설계하는지는 연구하지 않았다. 기껏해야 데이터의 불일치(2, 3, 4)로 인한 문제점을 지적하거나 단순히 데이터 중복 발생(5, 6), 또는 데이터 과적현상(7) 문제만을 지적하는 연구에 그쳤다.

2. 동기

데이터베이스 설계에 대해 연구를 하게 된 동기로는 과연 현업에서 데이터 모델링을 할 때 균형감을 갖춘 설계를 하고 있는지에 대한 의구심을 갖게 되었기 때문이다. 데이터 모델링은 어떤 정형화된 도구에 의해 행해지는 것이 아니라 완전

히 데이터베이스 설계자의 역량에 따라 데이터베이스의 품질이 좌우되는 것이 현실이기 때문이다.

데이터 모델링이란 기업의 '전사적 데이터 지도'(8, 9)를 만드는 일련의 방법론을 말한다. 많은 양의 데이터를 효과적으로 운용하기 위해 불가피한 중복은 예외인 상태에서 데이터 중복을 최대한 제거하고 속도와 질을 모두 만족시키는 데이터 지도만이 현업의 데이터 무결성을 보장하는 동시에 정보시스템의 응답 속도 및 서비스 품질을 향상시킬 수 있기 때문이다.

현행하는 대학 취업정보시스템에서는 데이터 모델링은 전무하였고 심지어 데이터베이스의 구성 요건조차 갖추고 있지 않은 실정이었다. 무엇이 개체인지, 무엇이 행위인지 구분이 안 된 채 하나의 테이블에 모든 필요한 데이터를 모아놓은 데이터 덩어리였다. 엄청난 데이터 중복과 널(null)값의 발생으로 생길 수 있는 문제점들을 무시한 채 데이터만 채워 넣고 취업프로그램관리 정보시스템이라는 이름으로 운영하고 있었다.

이런 근원적 문제점을 개선하지 않고는 데이터 일관성이 시스템에 의해 자동 유지 보장될 수 없다. 데이터베이스의 중요 요건인 일관성 유지 문제는 반드시 시스템이 자동으로 처리할 수 있어야 하며 일관성 체크를 어떤 경우에도 시스템 자동으로 처리하지 않을 경우 그로 인한 시스템의 정확성 유지와 응답 속도의 저하 현상은 심각해질 수 있다.

논리데이터모델에서 데이터의 유형은 개체(entity)와 개체 간의 관계를 형성해주는 행위(relationship)로 반드시 구분되어야 한다. 개체와 개체 간의 관계를 설계도면 상에서 명시적으로 구분해낼 수 없는 경우에는 설계의 균형성이 파괴될 수 있다. 균형성이란 개체와 관계간의 균형 감각을 가리키는 것으로서 균형 파괴 설계가 나타나는 원인은 주로 설계를 관계(행위) 중심으로 하지 않고 개체 중심으로 하는 잘못된 관행에 기인한다.

데이터베이스 설계 구조의 균형이 파괴되면 데이터 질의에 대해 응답하는 경우 응답속도가 항상 일정 시간(예: 3초)내에 불가능해지며 데이터베이스 비만도가 높아지게 되는 것도 데이터베이스 설계에서 방사형 설계가 그 원인으로 추측이 된다. 본 논문에서는 하나의 구조로 되어있는 것을 정방향 구조로 다시 설계해 봄으로써 데이터베이스 설계의 중요성을 확인하고자 한다. 균형 파괴의 사례는 수없이 많으나 독자들의 이해를 돕기 위해 전형적 사례 하나를 보임으로써 이해를 증진시키고자 한다.

3. 목적

데이터베이스 시스템에서 데이터 품질제고를 위해서는 데이터의 정확성, 완전성 및 일관성을 보장하여야 한다. 따라서

본 연구에서는 현장 데이터베이스 시스템에서의 데이터 모델링의 완전성을 측정하고, 실제 대학 취업관리 데이터베이스를 사례로 기존의 데이터베이스에 대한 평균 데이터 중복률을 조사하였다. 또한 데이터베이스 설계에 따른 비만도가 어떻게 달라지는지 살펴보기 위해 균형감 있는 설계를 통해 정방향 구조로 데이터 모델링을 통한 데이터베이스를 설계하였다.

사례로 든 데이터베이스는 실제 4년제 대학의 취업프로그램이수관리 데이터베이스로서 재학생들의 취업을 고취시키기 위해 여러 가지 취업프로그램을 개설하여 학생들이 취업프로그램을 신청하여 이수하는 과정을 나타낸다. 새로 설계한 데이터베이스는 정방향 구조를 가진 데이터베이스로서 대학 전체 업무로 데이터베이스의 규모가 확장하여도 동일한 구조를 가지게 될 것이다.

본 논문의 목적은 정방향 구조를 가진 데이터 모델링을 통하여 데이터의 정확성과 완전성을 보여주고자 한다. 또한, 기존 데이터베이스와 개선된 데이터베이스의 평균 데이터 비만도를 비교 분석하고자 한다. 기존 현존하는 대학 학사의 일부인 취업프로그램 관리 정보시스템이 얼마나 정보시스템 개선을 위한 목표 의식이 부재하고 정교함이 갖추어지지 않은 체계화성 없이 기획 운영되어 왔는지 알 수 있었다. 데이터 지도에 지도의 일부분이 조각 혹은 파편 형태로 고립된 상태가 존재해서는 지도로서의 자격이 상실된다. 그러나, 현행 취업프로그램이수관리 정보시스템의 경우, 이런 고립된 파편 형태로 존재함으로써 대학운영시스템 전반적으로 정보시스템 설계의 품질이 격하되어 있는 상태이다. 이러한 설계의 폐단은 데이터 중복 및 널 값의 과도함으로 드러나게 된다. 3장에서는 현존하는 취업프로그램이수관리 업무를 분석함으로써 문제점을 도출하였고 4장과 5장에서는 새로 개선된 취업관리 데이터베이스를 설계함으로써 이전 취업프로그램이수관리 데이터베이스의 문제점을 해결하고자 하였다. 과거 데이터 비만도가 과도하게 나타나는 방사형 및 고립형 설계를 지양하고 사각정방향으로 설계함으로써 데이터 경로가 어느 경우에도 선명하게 가시화되게끔 현행 시스템을 개선하는 것을 목적으로 한다.

II. 관련 연구

DB품질은 DB의 바람직한 정도 또는 우수성이라고 정의되며, DB 품질기준은 데이터 품질과 서비스 품질로 양분하여 접근하고 있다. 모두 7개 기준 중 DB데이터 품질을 구성하고 있는 요소는 정확성(Accuracy), 완전성(Completeness), 현행성(Currentness), 일관성(Consistency)이다. DB서비

스 품질은 검색성(Searching), 사용용이성(Ease of Use), 그리고 사용자지원성(Customer Support) 세 가지로 구성 되어 있다[11].

본 연구에서는 7개 기준 중 표현하고자 하는 실세계의 중요한 객체들과 속성들이 모두 담겨있어야 하는 점이 관건인 데이터 완전성에 대해 살펴보고자 한다. DB데이터의 완전성은 데이터 구조(structure)의 완전성, 데이터 값(value)의 완전성, 데이터 표현(representation)의 완전성 세 가지로 나누어진다.

1. 데이터 구조의 완전성

데이터 구조의 완전성이란 데이터베이스의 데이터가 실제계의 중요한 객체그룹을 모두 포함하고 있는지, 또한 객체에 관한 중요한 속성들을 모두 담고 있는지를 분석함으로써 데이터베이스 품질을 평가하는 도구이다. 데이터 구조가 완전하지 못하다 함은 사용자의 정보요구 분석단계, 데이터모델링 단계, 혹은 데이터베이스의 논리설계 단계에서 치명적 결함이 발생했음을 의미하며, 이러한 결함은 데이터베이스 품질뿐 아니라 데이터베이스 존재가치에 까지 영향을 줄 수 있다. 그러나 무엇이 중요한 속성이며 그것이 왜 빠졌는가? 라는 문제는 해당 분야 전문가의 심층적 분석과 판단이 요구된다.

2. 데이터 값의 완전성

데이터 값의 완전성이란 누락된 객체는 없는지, 또는 레코드의 필드 값이 비어있지는 않는지 평가하는 도구이다. 이러한 문제는 원시 데이터 수집 또는 입력단계에서 오류가 발생한 결과이다. 전체 레코드 중에서 누락된 레코드의 비율에 의해 데이터의 완전도를 계량화할 수 있다. 예컨대 1000명 학생 중 900개 레코드만 존재한다면 완전도 90%로 표현할 수 있다. 때로는 실세계에 존재하는 모든 객체들의 수를 정확히 파악할 수 없는 경우도 있으며, 이런 경우 완전성 기준이 100% 측정되지 못한다.

한편 필드 값이 비어있다 함은 널 값이 있다는 말이다. 언뜻 보기에는 널 값이 들어있는 레코드는 불완전한 레코드이고, 전체 레코드 중에서 불완전한 레코드의 수에 의해 완전도를 나타낼 수 있는 것처럼 판단할 수 있다. 널 값이 가지는 문제점은 다음과 같다.

우선 속성 값이 널 인 경우 제1정규형에 위배가 된다. 제1 정규형이란 테이블 내에 반복되는 속성이나 속성 그룹이 없는 형태를 말한다. 예를 들어, [그림 1]은 다수 개의 전화번호 행을 가지고 있는 고객들에 대한 테이블이다

Customer					
Customer_ID	First Name	Surname	Tel_No_1	Tel_No_2	Tel_No_3
123	Robert	Ingram	555-861-2025	null	null
456	Jane	Wright	555-403-1659	555-776-4100	555-777-5100
789	Maria	Fernandez	555-808-9633	null	null

그림 1. Customer 테이블
Fig. 1. Customer Table

[그림 1]의 Customer 테이블은 제 1 정규형이 아니다. Tel_No_1, Tel_No_2, Tel_No_3 행은 완전히 동일한 도메인과 의미를 가진다. 위와 같이 전화번호 행을 3개로 인위적으로 나누게 된다면 아래와 같은 논리적인 문제들을 낳게 된다.

테이블 질의 시 어려움 발생 : “어느 고객이 전화번호 X를 가지고 있는가?”, “어떤 고객들끼리 같은 전화번호를 공유하는가?” 질의들은 답하기 어렵다.

데이터베이스에서 고객-전화번호의 유일성을 확보하기 어렵다. 고객 789가 실수로 Tel_No_2 값이 Tel_No_1 값과 동일한 값이 들어갈 수 있다.

전화번호 개수의 제한. 어떤 고객이 4개의 전화번호를 가진다면 4번째 전화번호는 기록될 수 없다. 이는 데이터베이스 디자인이 완전한 사용자의 편의를 제한하고 있다는 의미인데, 사실 이상적으로는 반대가 되어야 한다.

이러한 문제점들을 해결하기 위해 한 개의 전화번호 행을 두면서 전화번호 행 길이를 여러 개의 전화번호를 기록하기 충분한 길이로 [그림 2]와 같이 변경하였다.

Customer			
Customer_ID	FirstName	Surname	Telephone_Numbers
123	Robert	Ingram	555-861-2025
456	Jane	Wright	555-403-1659, 555-776-4100, 555-777-5100
789	Maria	Fernandez	555-808-9633

그림 2. 변경된 Customer 테이블
Fig. 2. Modified Customer Table

[그림 2]는 제 1정규형인 것처럼 보이지만 문제점이 여러 개 있다. 전화번호는 의미상으로 모호해져서 전화번호를 표현

할 수도, 전화번호들의 리스트를 표현할 수도, 심지어는 아무거나 표현할 수도 있다. "어떤 고객들이 같은 전화번호를 공유하는가?" 질의는 답하기 더욱 어려워졌으며, 전화번호와 전화번호들의 리스트를 관리해야 할 필요성이 생겼다. 데이터베이스 내에서 전화번호에 대한 의미적인 제한을 정의하는 것도 또한 어려워졌다. 즉 데이터 원자성(atomicity)이 충족되지 못하는 것이다. 최종적으로 제 1정규형을 만족하는 길은 [그림 3]과 [그림 4]와 같이 2개의 테이블로 나누는 것이다.

[그림 3]의 Customer_Name 테이블에서는 Customer_ID가 기본키이고, [그림 4]에서의 Customer_TelephoneNumber 테이블에서는 Customer_ID와 Serial_Number가 복합키이면서 기본키이다. 이와 같이 두 개의 테이블로 나누 후에야 비로소 널 값의 삽입으로 인한 문제점을 해결한 제 1 정규형이 되는 것이다.

Customer_Name		
Customer_ID	FirstName	Surname
123	Robert	Ingram
456	Jane	Wright
789	Maria	Fernandez

그림 3. Customer_Name 테이블
Fig. 3. Customer_Name Table

Customer_TelephoneNumber		
Customer_ID	Serial_Number	Telephone_Number
123	1	555-861-2025
456	1	555-403-1659
456	2	555-776-4100
456	3	555-777-5100
789	1	555-808-9633

그림 5. Customer_TelephoneNumber 테이블
Fig. 5. Customer_TelephoneNumber Table

3. 데이터 표현의 완전성

데이터 표현의 완전성이란 가공된 데이터가 원문(원시 데이터)에 담긴 정보를 완전하게 담고 있는지를 의미한다. 분류, 재배열, 또는 재구성된 데이터의 내용이 원시 데이터의 일부를 누락하고 있지 않는지, 또는 추출된 키워드(keyword)는 해당 주제 분야의 주요 개념을 완전히 커버하는지, 혹은 기타 형태로 가공된 데이터가 원시 데이터의 일부

를 누락하고 있지 않는지 여부를 분석함으로써 데이터 표현의 완전성을 평가한다.

III. 표준업무

1. 가정

본 논문에서는 기존의 데이터베이스 설계에 대해 문제점을 도출하고 새로 설계한 데이터베이스 설계도를 비교하고 시스템 구축 설계를 위해 반드시 필요한 몇 가지 가정을 도입하기로 한다.

가정 1(업무의 종류): 대상 업무는 현재 4년제 대학에서 운영하고 있는 취업프로그램 이수관리 시스템이라고 가정한다.

가정 2(설계방법론): 데이터모델링에 도입되는 데이터베이스 설계 방법론으로는 요구 형성 과정을 업무의 시작으로 하여 개체관계도를 작성하는 것으로 한다.

가정 3(개체관계도의 표현): 개체관계도를 작성할 때 정보공학방법론[12]에 의하면 쌍방향 명령법을 사용하고 선택사항(Optionality)을 정의하며 관계성(Cardinality)의 표현으로 새발표기법(Crow's foot)을 사용하지만 개체 관계도 제작에 있어서 공간의 부족으로 인해 표현상의 어려움을 감소시키고자 하여 Chen 방식[13]을 사용하여 표현한다.

가정 4(업계관행 수용): 검색 속도 증진을 위해 업계에서 널리 관행적으로 수용되는 방식인 테이블 통합이 이루어진다고 가정한다. 업무기술서 작성을 시작으로 하여 데이터모델링을 하여 완성된 데이터베이스 설계도와 업계에서 설계된 설계도와의 데이터 비만도를 측정할 때, 업계의 관행적 설계방식이 허용된다고 가정한다.

가정 5(데이터베이스 관리 시스템의 종류): 설계된 정보 시스템은 표준화된 인터페이스인 SQL을 지원하여 현재 가장 보편적으로 사용하고 있는 RDBMS로 구축된다고 가정한다.

2. 취업프로그램 이수관리 시스템

실제 대학 취업지원팀이 학생들의 취업프로그램 수강 및 이수를 관리하는 데이터베이스는 "취업프로그램 이수관리"라

는 이름으로 하나의 테이블로 구성되어 있었다. 업무기술서 초차 먼저 작성하지 않고 취업관련 업무에 필요할 만한 데이터만 테이블 형태를 갖춰 모아놓고 사용하고 있었기 때문에 데이터베이스 설계도면이 부재했다. 본 논문에서는 현재의 테이블을 구성하고 있는 데이터를 가지고 취업지원팀의 업무를 파악하여 처음부터 업무기술서를 작성해 보았다. 취업프로그램 이수를 관리하는 업무기술서는 다음과 같다.

2.1 취업프로그램 이수관리 업무기술서

취업프로그램 이수관리는 해당 연도와 학기에 학생이 취업 프로그램을 신청하고자 할 때 각각 일련번호를 부여하며 각 취업프로그램마다 사업명이 주어지며 취업프로그램은 프로그램명, 프로그램 시작일자, 프로그램 종료일자, 프로그램 참여일수로 구성되어 있다. 프로그램은 산업체들이 참여하며 산업체 정보로서 업체구분, 업체명, 업체담당자, 업체담당자 전화번호, 업체담당자 이메일, 업체 팩스번호로 구성되어 있으며 프로그램을 이수 완료하였을 경우 국고지원금, 대응지원금, 교비지원금 중 하나의 지원금을 학생에게 지원하게 되는데 지원금을 지급하기 위해 은행구분, 계좌번호, 예금주명 등의 정보를 입력해야 하며 취업프로그램을 이수한 학생들에게 마일리지 부여하기 위해 마일리지 식별ID를 입력한다.

2.2 취업프로그램 이수관리 테이블

취업프로그램 이수관리 테이블은 [그림 5]와 같이 해당 연도와 학기에 학생이 취업프로그램을 신청하고자 할 때 각각 일련번호를 부여하며 각 취업프로그램마다 사업명이 주어지며 취업프로그램은 프로그램명, 프로그램 시작일자, 프로그램 종료일자, 프로그램 참여일수로 구성되어 있다. 프로그램은 산업체들이 참여하며 산업체 정보로서 업체구분, 업체명, 업체담당자, 업체담당자 전화번호, 업체담당자 이메일, 업체 팩스번호로 구성되어 있으며 프로그램을 이수 완료하였을 경우 국고지원금, 대응지원금, 교비지원금 중 하나의 지원금을 학생에게 지원하게 되는데 지원금을 지급하기 위해 은행구분, 계좌번호, 예금주명 등의 정보를 입력해야 하며 취업프로그램을 이수한 학생들에게 마일리지 부여하기 위해 마일리지 식별ID를 입력한다.

3. 취업프로그램 이수관리 테이블의 문제점

현재 사용 중인 취업프로그램이수관리 테이블을 사용 시 발생할 수 있는 문제점은 우선 꼭 필요한 데이터들이 누락되어 있는 점이다. 또한 널 값이 많이 발생하는 문제와 데이터

중복 문제, 업무 담당자가 데이터를 입력 시 발생할 수 있는 번거로움, 그리고 유용하게 사용될 수 있는 통계자료를 산출하기 어려운 문제점이 내재되어 있다.

3.1 필요데이터의 누락

기존의 취업프로그램이수관리 테이블은 모든 데이터를 하나의 테이블에 모아놓은 형태로서 사용자의 정보요구 분석 단계, 데이터 모델링단계를 거치지 않은 원시상태의 데이터를 모아놓은 집합에 불과하다. 즉 데이터베이스의 데이터가 중요한 객체그룹들을 포함하고 있지 못하며 객체에 관한 중요한 속성들도 누락하고 있다. 누락된 데이터는 다음과 같다.

첫째, 과목개설관련 데이터의 누락이다. 취업프로그램 중 인턴십, 청년강소기업체협 프로그램을 학생이 신청했을 경우 취업지원팀에서는 교무처에 교과목 개설을 요청해야 하는데 과목 개설 요청에 대한 데이터가 없다. 단순 취업프로그램 이수인지, 정식과목 이수인지를 구분하지 않았다.

둘째, 과목이수 및 프로그램 이수 관련 데이터의 누락이다. 취업프로그램은 두 종류로 구분한다. 첫째는 인턴십, 청년강소기업체협 프로그램인 경우 정식 과목으로 신청하여 이수를 하게 되면 학점이 부여되고, 둘째로 인턴십, 청년강소기업체협 프로그램이 아닌 구직스킬강화 프로그램, 매경TEST모의 시험, 선배와의 대화, 이미지 메이킹, 인사담당자 취업특강, 직무적성검사 교육, 직무적성검사 모의고사, 모의토의시험, 토익사관학교, 한류면접집중반, 한류입사지원반, YES리더십 특강 같은 경우 프로그램을 이수하였을 경우 마일리지만을 부여하게 된다. 이와 같이 프로그램 이수 절차가 다른 두 종류의 프로그램에 대한 분류도 되어 있지 않았다.

셋째, 취득한 학점에 대한 데이터 누락이다. 과목을 이수하였을 경우 취득하게 되는 학점과 성적을 나타내는 속성을 포함하지 않았다.

넷째, 산업체 평가 및 통보관련 데이터가 누락되어 있다. 산업체에서 인턴십, 청년강소기업체협 프로그램을 실습하게

학번
일련번호
년도
학기
사업명
프로그램명
프로그램시작일자
프로그램종료일자
프로그램참여일수
활동시간
수료여부
수료번호
업체구분
업체명
업체담당자
업체담당자전화번호
업체담당자이메일
업체팩스번호
국고지원금 1
대응지원금 2
교비지원금 3
은행구분
계좌번호
예금주명
비고
마일리지식별 ID

그림 5. 취업프로그램 이수관리 테이블

Fig. 5. Job Program Completion Table

되던 산업체에서는 해당 학생에 대한 실습 결과를 평가하며 평가결과를 취업지원팀에 통보를 하게 되어 있는데 이에 대한 정보가 결여되어 있다.

이와 같은 중요한 데이터를 누락하였고 무엇이 객체인지, 속성인지 전혀 구분해 놓지 않아 차마 데이터베이스의 테이블이라는 용어조차 사용할 수 없을 정도이다.

3.2 널(null) 값 문제

2장의 [그림 1]에서 살펴본 같이 테이블 내에 존재하는 널 값은 여러 문제점을 발생시키고 있다. 널 값의 문제점을 취업 프로그램이수관리 테이블에서 찾아보면 다음과 같다.

첫째로 학생이 단순히 취업프로그램만 이수하고자 할 경우 수많은 속성들이 널 값을 갖는다. 총 26개의 속성 중 절반에 해당되는 학번, 일련번호, 년도, 학기, 사업명, 프로그램명, 프로그램 시작일자, 프로그램 종료일자, 프로그램 참여일수, 활동시간, 수료여부, 수료번호, 마일리지식별ID 만이 입력될 수 있다. 나머지 업체구분, 업체명, 업체담당자, 업체담당자 전화번호, 업체담당자 이메일, 업체팩스번호, 국고지원금1, 대용지원금2, 교비지원금3, 은행구분, 계좌번호, 예금주명, 비고와 같은 13 개의 속성에는 널 값이 입력된다. 대다수의 학생들은 취업프로그램을 단순히 이수하는 경우가 많다고 판단되므로 널 값으로 인한 문제점은 심각한 수준이다.

둘째로 학생이 취업프로그램 중 과목으로 학점을 따고 싶은 경우에도 널 값들이 속출한다. 인턴쉽, 청년강소기업 체험 프로그램을 과목으로 이수한 학생들인 경우에만 국고지원금 1, 대용지원금2, 교비지원금3 중 한 가지 종류의 지원금을 받게 됨으로써 나머지 두 가지 지원금에는 널 값이 들어갈 수밖에 없다. 또한 프로그램 수료여부, 프로그램 참여일수, 프로그램 시작일자, 프로그램 종료일자와 같은 속성에는 널 값이 들어간다. 이와 같이 취업프로그램 이수관리 테이블은 속출되는 널 값으로 인하여 제1정규형의 조건에도 미치지 못하는 형편이다.

셋째로 테이블 질의 시 어려움 발생할 수 있다. “취업지원 프로그램에서 국고지원금이 얼마나 소진되었는가?”에 대한 질의가 주어졌을 때, 테이블 내 국고지원금 속성에 존재하는 수많은 널 값을 어떻게 연산할 수 있겠는가? 과연 그 결과를 신뢰할 수 있을지 의문이다. 또한 “어떤 학생들이 국고지원금을 지원받았는지?”에 대한 질의들은 답하기 어렵다.

3.3 데이터의 중복

데이터의 중복은 이미 널리 알려진 바와 같이 여러 가지

문제점들을 야기 시키기 때문에 데이터베이스의 정확성과 효율성을 높이기 위해 중복을 피하기 위해 노력해야 한다. 통제되지 않는 데이터의 중복은 데이터의 일관성을 유지하는데 있어서 문제가 야기될 수 있고 데이터 이상 현상이 발생할 수 있다.

우선, 데이터가 일관성을 유지할 수 없다는 것은 데이터의 무결성을 지킬 수 없다는 것을 의미한다. 예를 들어 아래의 취업프로그램이수관리 테이블에서 동일한 프로그램에 대한 프로그램명, 프로그램 시작일자, 프로그램 종료일자는 여러 행에 걸쳐 동일하게 중복된다. 또한 동일한 업체에 대한 속성으로서 업체구분, 업체명, 업체담당자, 업체담당자 전화번호, 업체담당자이메일, 업체 팩스번호는 동일한 업체에 대해 완벽히 중복이 된다. 만일 동일하게 중복 입력되어야 하는 상황에서 데이터 입력 오류로 인해 정확한 값이 입력되지 못했다면 일관성 없는 결과를 초래할 수 있다. 특히 자주 입력될 경우 데이터 입력 오류는 심해질 수 있다.

데이터가 중복되면 여러 곳에서 필드 값의 변경이 이루어져야 한다. 그런데 필드 값의 변경으로 인해 데이터 이상 현상이 발생할 수 있다. 한 필드의 값을 변경하면 일관성을 유지하기 위해 다른 여러 곳의 값도 수정되어야 한다. ‘취업프로그램 이수관리’테이블에서 프로그램명인 ‘인턴쉽’의 프로그램 시작일자와 종료일자가 변경되었다면 이 ‘인턴쉽’프로그램을 신청한 모든 행에 대해서 일관성 있게 수정이 이루어져야 한다. 마찬가지로 업체 담당자가 변경되었다면 일일이 업체 담당자와 업체담당자 전화번호, 업체담당자 이메일, 업체 팩스번호 등을 해당 행에 대해 일관성 있게 수정해야만 한다. 학교와 협력관계에 있던 업체가 계약을 종료하고자 했을 경우 그 업체에 대한 모든 데이터를 일일이 찾아서 삭제해야만 한다. 학생들의 취업프로그램이수관리를 위해 수 년 동안 많은 레코드가 축적되어 있다면 아주 번거로운 작업이 될 것이다.

3.4 데이터 입력의 번거로움

취업프로그램은 동일한 학생이 여러 차례에 걸쳐서 취업프로그램을 신청하여 이수할 수 있게 되어 있다. 그런데 기본 키로서 ‘학번’과 ‘일련번호’, ‘년도’, ‘학기’를 복합키로 정해 사용하고 있다. 그런데 각 학생마다 일련번호를 부여하기 위해 각 학생 별로 몇 번째 신청인지 알아야만 입력을 할 수 있다. 예를 들어 한 학생이 한 학기에 여러 번 취업프로그램을 신청하고자 할 경우 일련번호만이 식별될 수 있는 속성이 되기 때문에 그 학생의 일련번호가 몇 번인지 검색해 내야만 한다. 이를 위해서는 일련번호를 알아내기 위해 다시 쿼리를 실행해야만 하는 번거로움이 발생할 수 있다. 그러나 새로운 모델링에 의하면 신청번호가 일

련번호로 주어짐으로써 동일한 학생이 취업프로그램을 여러 번 신청하는데 문제가 발생하지 않는다.

3.5 통계자료 산출의 어려움

학교에서 취업관리는 매우 중요한 업무 중 하나이다. 학생들의 취업을 도와주고 독려하기 위해 여러 가지 학생들에게 도움이 되는 취업프로그램을 개발하여 관리하는 것도 이 때문이다. 학생들의 요구와 필요성을 충족시키기 위해 취업지원팀에서는 학생들의 필요성을 즉각적으로 파악하고 실현하기 위해 취업프로그램관련 통계적 자료를 활용해야만 한다. 통계 자료를 바탕으로 학생들에게 꼭 필요한 맞춤형 취업프로그램을 개발, 운영할 수 있는 것이다. 그런데 취업프로그램 이수 관리 테이블에서는 취업프로그램 이수관련 통계자료를 산출하기 어렵다. 만일 동일한 취업프로그램을 두 번 이상 이수한 학생에 대한 정보는 산출해 내기 힘들다. 그러나 새로운 모델링에 의하면 이러한 통계자료는 질의를 통해 쉽게 얻어낼 수 있게 된다. 또한 어떤 업체가 '인턴쉽'을 많이 수행 했는가에 대한 통계자료와 실습을 이수한 후 평가되는 학점에 대한 정보도 취업프로그램이수관리 테이블에서는 누락되어 있음으로써 예를 들어 학생들의 '인턴쉽' 이수 평균 성적을 산출해 낼 수 없는 것이다.

IV. 새로운 취업관리 데이터베이스 설계

3장에서 살펴본 바와 같이 현재 사용 중인 취업프로그램이수관리 테이블은 많은 문제점을 가지고 있기 때문에 제대로 된 방식대로 절차를 거쳐서 데이터베이스를 설계하고자 한다. 우선 현업 담당자의 인터뷰 및 업무기술서를 통하여 업무를 파악하고 분석한 후 업무기술서를 다시 작성하였으며, 이를 바탕으로 새로운 취업관리 데이터베이스를 설계하고자 한다.

1. 새로운 취업관리 업무기술서

새로운 업무기술서는 업무 매뉴얼, 업무 담당자와의 인터뷰를 통해 업무를 분석한 후 작성하였다. 취업관리 업무는 크게 9가지 유형의 행위로 구성되어 있다.

(1. 프로그램 신청) 학생이 부서(취업지원팀)에 취업프로그램을 신청한다. 신청 결과 학점을 인정받는 정식 과목을 이수하거나 프로그램만 이수하게 됨으로써 마일리지점수를 얻게 될 수 있다.

(2. 과목개설) 담당 부서에서는 교무처에 취업프로그램 중

인턴쉽, 청년강소기업체험 프로그램인 경우 교과목 개설을 요청한다. 과목을 개설하게 되면 학생이 과목으로 이수할 수 있게 되며 산업체에서 실습을 통해 성적을 평가받아 학점을 받을 수 있게 된다.

(3. 과목이수) 학생은 신청한 취업프로그램(인턴쉽, 청년강소기업체험 프로그램, 구직스킬강화 프로그램, 매경TEST 모의시험, 선배와의 대화, 이미지 메이킹, 인사담당자 취업특강, 직무적성검사 교육, 직무적성검사 모의고사, 모의토의시험, 토익사관학교, 한큐면접집중반, 한큐입사지원반, YES리더쉽특강)중 하나를 이수한다. 취업프로그램을 이수하면, 성적을 취득할 수 있으며 해당 프로그램당 마일리지 가 부여된다. 마일리지를 누적하게 되면 점수에 따라 소정의 장학금을 지급한다. 취업프로그램 중 인턴쉽, 청년강소기업직장 프로그램을 이수했을 경우 취업지원팀에서는 해당 지원금(국고, 대응, 교비지원금)을 지급한다. 해당 지원금 중 국고는 노동부 지원사업이고 대응지원금은 정부지원 취업분야 특성화지원금 중 대학에서 대응투자비이다. 교비지원금은 순수 교비에서 지원되는 비용이다.

(4. 프로그램이수) 학생은 신청한 취업프로그램 중 과목개설이 되지 않은 것들(인턴쉽, 청년강소기업체험 프로그램, 구직스킬강화 프로그램, 매경TEST모의시험, 선배와의 대화, 이미지 메이킹, 인사담당자 취업특강, 직무적성검사 교육, 직무적성검사 모의고사, 모의토의시험, 토익사관학교, 한큐면접집중반, 한큐입사지원반, YES리더쉽특강) 중 하나를 이수한다. 취업프로그램을 이수하면, 해당 프로그램당 마일리지를 부여받을 수 있으며 마일리지가 누적하게 되면 점수에 따라 소정의 장학금이 지급된다.

(5. 실습) 학생은 산업체에서 실습을 이행한다. 실습을 이행하기 전 산업체 담당자와의 인터뷰를 통해 실습여부를 결정하게 된다. 실습을 함으로써 성적을 평가받는다.

(6. 평가) 학생이 취업프로그램(인턴쉽, 직장체험)실습을 수행하고 나면 산업체는 학생의 실습성적을 평가한다. 실습평가가 이루어지면 학생의 성적이 산출된다.

(7. 통보) 취업프로그램(인턴쉽, 직장체험) 성적을 평가한 후 해당 산업체에서는 부서에 실습성적을 통보한다. 실습성적을 통보하고 나면 산업체에서는 해당 학생에 대한 실습업무가 종료된다.

(8. 계약) 학교와 산업체 간의 취업지원프로그램에 관한 계약이 이루어진다. 산업체와의 취업프로그램에 계약에 따라 학생들이 실습을 수행할 수 있게 된다.

(9. 운영) 산업체와의 계약이 이루어지면 취업지원팀에서는 취업관련 프로그램으로 인턴쉽, 청년강소기업체험 프로그램,

구직스킬강화 프로그램, 매경TEST모의시험, 선배와의 대화, 이미지 메이킹, 인사담당자 취업특강, 직무적성검사 교육, 직무 적성검사 모의고사, 모의토익시험, 토익사관학교, 한류면접 집중반, 한류 입사지원반, YES 리더쉽 특강을 운영한다.

2. 취업관리 데이터베이스 설계모형

작성된 취업관리 업무기술서를 바탕으로 10가지 유형의 행위를 기본으로 다음과 같은 설계도 [그림 6]을 도출하였다.

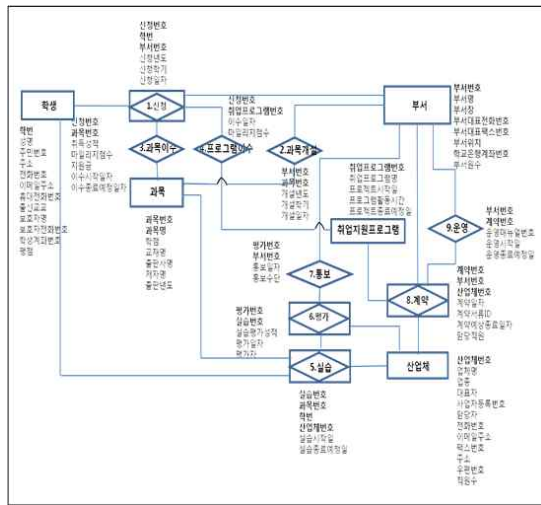


그림 6. 취업관리 데이터베이스 리모델링 모형
Fig. 6. Job Management Database Remodelling Model

3. 문제점이 해결된 새로운 데이터모델링

3장에서 기존 취업프로그램 이수관리 데이터베이스의 문제점을 5가지 유형으로 분석하였는데, 이러한 문제점들을 새로운 취업관리 데이터 모델링에서는 다음과 같이 해결할 수 있었다.

3.1 필요데이터 누락문제 해결

우선 과목개설관련 데이터의 누락문제가 해결되었다. 이전에는 학생이 프로그램을 신청할 때 단순 취업프로그램이수인지, 정식과목 이수인지를 구분하지 않았다. 그러나 새로운 설계에 의하면 취업프로그램 중 인턴쉽, 청년강소기업체험 프로그램을 학생이 신청했을 경우 취업지원팀에서는 교무처에 과목 개설을 요청하여 과목 개설이 이루어질 수 있다.

둘째, 과목이수 및 프로그램이수 관련 데이터의 누락 문제가 해결되었다. 취업프로그램은 두 종류로 구분한다. 하나는 인턴쉽, 청년강소기업체험 프로그램인 경우 정식 과목으로 신청하여

이수를 하게 되면 학점이 부여되는 것이 있고 또 다른 하나는 인턴쉽, 청년강소기업체험 프로그램이 아닌 기타 취업프로그램을 이수하였을 경우 마일리지만을 부여하게 된다. 이와 같이 프로그램 이수 절차가 다른 두 종류의 프로그램에 대해 과목이수인지 단순 프로그램이수인지 구분해서 설계하였다.

셋째, 취득한 학점에 대한 데이터 누락문제가 해결되었다. 과목을 이수하게 되면 취득하게 되는 학점과 성적을 추가로 포함하였다.

넷째, 산업체 평가 및 통보관련 데이터 누락 문제가 해결되었다. 산업체에서 인턴쉽, 청년강소기업체험 프로그램을 실시하게 되면 해당 산업체에서는 학생에 대한 실습 결과를 평가하며 평가결과를 취업지원팀에 통보를 하게 되어 있는데 이에 대한 데이터를 추가하였다.

이외에도 취업프로그램에 참여하는 업체의 참여일, 취업프로그램업무를 담당하는 담당 직원에 대한 데이터도 추가함으로써 보다 책임감 있는 취업업무를 할 수 있도록 데이터베이스를 설계하였다. 또한 누락된 데이터를 추가함과 동시에 개체와 속성의 구분도 확실하게 정의하였다.

3.2 널 값 및 불필요한 데이터 중복 문제 해결

기존의 취업프로그램이수관리 테이블은 많은 널 값을 내포할 수밖에 없었다. 예를 들어, 우선 학생이 단순히 취업프로그램만 이수하고자 할 경우 무려 26개의 필드 중 13개의 필드 값이 널 값이었다. 그러나 새로운 취업관리 데이터베이스 설계에 의하면 이 경우 단지 신청번호, 취업프로그램번호, 수료완료일자, 마일리지점수 와 같은 필드 값만 입력되면 되므로 이전의 업체구분, 업체명, 업체담당자, 업체담당자전화번호, 업체담당자이메일, 업체팩스번호, 국고지원금1, 대응지원금2, 교비지원금3, 은행구분, 계좌번호, 예금주명, 비고와 같은 수많은 속성들을 입력할 필요가 없기 때문에 널 값이 발생하지 않는다.

둘째로 학생이 취업프로그램 중 과목으로 이수하여 학점을 따고 싶은 경우에도 널 값들이 속출했었다. 즉, 인턴쉽, 청년강소기업 체험프로그램을 과목으로 이수한 학생들인 경우에만 국고지원금1, 대응지원금2, 교비지원금3 중 한 가지 종류의 지원금을 받게 됨으로써 나머지 두 가지 지원금에는 널 값이 들어갈 수밖에 없었는데 새로운 취업관리 데이터베이스 설계에서는 학생이 과목을 이수함으로써 받게 되는 지원금은 하나의 속성으로만 부여함으로써 한 가지 종류의 지원금만을 입력한다. 또한 단순 프로그램이수 시 입력되는 속성인 프로그램 수료여부, 프로그램참여일수, 프로그램시작일자, 프로그램 종료일자와 같은 속성을 입력할 필요가 없기 때문에 널 값이

발생되지 않는다. 셋째로 테이블 질의 시 널 값으로 인해 발생하는 오류들도 자연히 소멸될 수 있게 된다.

취업프로그램이수관리 테이블에서는 동일한 취업프로그램에 대한 프로그램번호, 프로그램명, 프로그램 시작일자, 프로그램종료일자는 여러 행에 걸쳐 동일하게 중복된다. 그러나 새로운 취업관리 데이터베이스 설계에서는 취업지원프로그램을 하나의 개체로 설계함으로써 이 중복 문제가 해결된다. 또한 산업체도 하나의 개체로 설계함으로써 업체에 대한 정보는 중복되지 않고 한 곳에 모아놓음으로써 업체번호, 업체구분, 업체명, 업체담당자, 업체담당자전화번호, 업체담당자이메일, 업체팩스번호는 매번 중복되지 않는다. 데이터 변경이 일어나더라도 한 곳에서만 변경이 발생함으로 중복으로 인한 데이터 일관성 오류는 발생하지 않는다. 데이터 변경 시 취업프로그램이수관리 테이블에서 프로그램명인 '인턴쉽'의 프로그램시작일자와 종료일자가 변경되었다면 이 '인턴쉽' 프로그램이 저장되어 있는 취업지원프로그램 테이블 한 곳에서만 변경이 이루어지면 된다. 마찬가지로 업체 담당자가 변경되었다라도 산업체 테이블 한곳만 수정하면 된다. 학교와 협력관계에 있던 업체가 계약을 종료하고자 했을 경우 그 업체에 대한 모든 데이터를 일일이 찾아서 삭제해야만 했던 것도 산업체 테이블에서 간단히 하나의 행만 삭제하면 되는 것이다.

3.3 데이터의 중복률 비교

새로 설계한 취업관리 데이터베이스와 현업에서 사용 중인 취업프로그램 이수관리 데이터베이스와의 데이터 중복률을

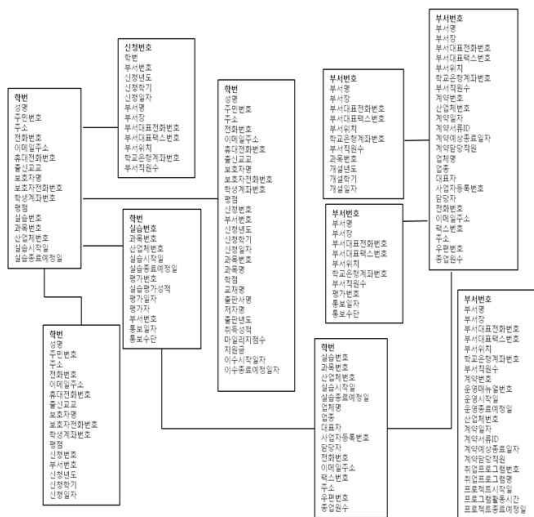


그림 7. 현업 취업프로그램 데이터베이스 모형도
Fig. 7. Job Program Database Diagram in the Work

비교해 보고자 한다. 그래서 새로 설계된 데이터베이스와 현업에서 관습적으로 설계된 데이터베이스와의 정량적인 데이터 중복률 비교를 위해 현업에서 통상적으로 검색속도를 높이기 위해 실시하는 테이블 통합기법을 이용하여 [그림 7]과 같이 현업에서 운영 중인 취업프로그램 데이터베이스를 설계해 보았다.

다음은 데이터모델링 설계 절차를 거쳐 새롭게 설계된 모델링 취업관리 데이터베이스인 [그림 6]의 데이터 중복률과 데이터 모델링을 전혀 고려하지 않고 설계되어 테이블 통합기법으로 운영되는 취업프로그램 이수관리 데이터베이스인 [그림 7]의 데이터 중복률을 구하였다.

데이터 중복률은 데이터베이스 스키마에 포함된 총 속성(컬럼)의 갯수 중에서 부모-자식 관계에서 기본키가 자식 개체(테이블)에 상속되어 참조하는 외래키와 같이 스키마 내에서 서로 중복되는 속성들의 개수의 비율을 말한다.

따라서, 현업에서 운영 중인 취업프로그램 이수관리 데이터베이스의 데이터 중복률(data redundancy ratio)

$$\frac{\text{중복된 속성들의 총 빈도수}}{\text{DB Schema 에 포함된 총 속성 종류의 갯수}} = \frac{98}{175} = 56\%$$

이며, 새로 설계된 취업관리 데이터베이스의 데이터 중복률

$$\frac{\text{중복된 속성들의 총 빈도수}}{\text{DB Schema 에 포함된 총 속성 종류의 갯수}} = \frac{18}{93}$$

= 19% 로 나타났다.

이와 같이 두 데이터베이스간의 중복률이 37% 차이가 남을 알 수 있었다. 즉, 새로 설계된 데이터베이스의 중복률이 현저히 감소되었다. 데이터베이스 정규형인 제1 정규형에서 제3정규형으로 변환함으로써 데이터 중복을 방지하고 저장 가능공간을 최적화할 수 있었다. 또한 적절한 외래키를 사용함으로써 데이터 중복뿐만 아니라 연쇄 이상현상(anomalies)의 가능성을 최소화할 수 있었다.

현대의 기술을 기준으로 할 때 정보기술 자체의 한계선인 중복률 15%를 상회한다는 이야기 자체는 기술적으로는 주로 기본키 속성을 부주의하게 외래키 속성으로 빈번히 여러 군데에서 방만하게 중복되게끔 허용했다는 뜻이다. 물론 이외에도 중복이 유발되는 경우는 존재한다. 이는 현대의 데이터 설계의 규범과 기초를 무시한 편법 설계로서 반드시 개선해야 할 대상이다. 인체 기준으로 해도 인체 체지방율이 30%를 넘는

다는 것은 고도 비만 상태의 심각한 지경으로 누구나 상식적으로 평소 들어서 알 내용이다.

중복율이 최적 중복율 15%를 기준으로 삼을 때 심지어 41%만큼이나 과다하게 상회 초과하고 있다는 점이 시사하는 바는 취업프로그램 이수관리 데이터베이스가 현재 전체 보유하고 있는 데이터 중 41%를 과감히 도려내더라도 취업프로그램 이수관리 데이터베이스는 전혀 문제없이 오히려 효율적으로 작동할 것이라는 점이다. 뿐만 아니라 불필요 부분 제거 작업을 수행함으로써 신체의 경우처럼 몸이 가벼워져 질의에 대한 응답속도 또한 제거한 만큼에 자동 비례하여 훨씬 더 신속하게 나오게 되므로 중복 데이터 제거 작업 계획은 조기에 마련될수록 바람직하다고 본다.

3.4 데이터 모델의 균형성

[그림 6]에서와 본 바와 같이 새로운 취업관리 데이터베이스는 개체 수 대 행위 수의 비율이 9 : 10 으로 매우 균형적인 구조를 나타내고 있다. 개체 수가 행위 수가 거의 비슷하다는 것은 데이터베이스 설계구조가 균형성을 만족한다는 뜻이다. 또한 원형(circle)구조를 가짐으로써 질의에 대한 응답속도가 적정하게 유지된다. 그러나 [그림 7]의 현업 취업관리 데이터베이스 모형도에서는 설계의 균형성이 유지되지 못하였다. 즉 개체 간의 구조가 한쪽으로 치우쳐 뻗은 구조로서 질의에 대한 응답 속도라든지 특정 질의에 대한 정확한 결과를 예측할 수 없다.

3.5 데이터 입력의 번거로움 해결

취업프로그램에서는 동일한 학생이 여러 번 취업프로그램을 신청하여 이수하고자 할 경우 기본 키의 일부로 '일련번호'를 사용함으로써 이 일련번호를 알아내기 위해 검색이 우선되어야만 하나, 새로운 취업관리 데이터베이스 설계에서는 취업프로그램을 신청할 때 '신청번호'속성을 기본 키의 일부로 설정함으로써 자동적으로 신청번호가 발생된다. 일련번호를 알아내기 위해 다시 쿼리를 실행해야만 하는 번거로움이 해소됨과 동시에 동일한 학생이 취업프로그램을 여러 번 신청하는데 문제가 발생하지 않는다.

3.6 통계자료 산출의 어려움 해결

새로운 취업관리 데이터베이스 설계에서는 상시 3초내 응답속도 보장을 위해서는 특히 방사형(skewed) 설계를 탈피하는 형태로 논리데이터모델로 설계하였다. 이에 더하여 고립형 설계까지 탈피함으로써 자동적으로 정방형 설계 혹은 사각

경로 형 설계의 형태를 지니게 되었다. 또한 취업프로그램 이수관련 통계자료를 다양한 측면에서 검색할 수 있기 때문에 취업지원팀의 운영 전략에 큰 도움이 되는 자료를 산출해 낼 수 있다. 사각정방형 설계로 인해 응답시간 3초를 원천적으로 상시 보장하며 고 수준의 통계자료를 언제든지 알아낼 수 있는 장점이 있다.

V. 결론

데이터가 불필요한 중복 없이 정확성과 신속성을 갖춘 품질 좋은 데이터베이스를 개발하기 위한 개발 방법론이 많이 연구 발표되었으나 아직도 대학의 취업정보시스템에서는 데이터베이스를 설계하는 과정에서 이를 완전히 무시한 채 업무의 효율만을 지나치게 강조하여 실제 데이터베이스 설계하고 있다. 대학 취업정보시스템에서는 데이터 모델링은 전무하였고 심지어 데이터베이스의 구성 요건조차 갖추고 있지 않은 실정이었다. 무엇이 개체인지, 무엇이 행위인지 구분이 안 된 채 하나의 테이블에 모든 필요한 데이터를 모아놓은 데이터 덩어리였다. 엄청난 데이터 중복과 널 값의 발생으로 생길 수 있는 문제점들을 무시한 채 데이터만 채워 넣고 데이터베이스라는 틀을 씌워 운영하고 있었다.

본 논문에서는 정방형 구조를 가진 데이터 모델링을 통하여 데이터의 정확성과 완전성을 보여주고자 하였다. 또한, 기존 데이터베이스와 개선된 데이터베이스의 평균 데이터 비만도를 비교 분석하였다. 기존 현존하는 대학 학사의 일부인 취업정보시스템이 얼마나 정보시스템 개선을 위한 목표 의식이 부재하고 정교함이 갖추어지지 않은 채 계획성 없이 기획 운영되어 왔는지 알 수 있었다. 이러한 설계의 폐단은 데이터 중복 및 널 값의 과도함으로 드러났다. 과거 데이터 비만도가 과다하게 나타나는 방사형 및 고립형 설계를 지양하고 사각정방형으로 설계함으로써 데이터 경로가 어느 경우에도 선명하게 가시화되게끔 현행 시스템을 개선하였다. 그 결과 기존의 대학 취업정보시스템의 비만도는 50%가 넘었으나, 본 논문에서 제안한 새로운 대학 취업데이터베이스 시스템의 데이터 비만도는 19%로 나타나 기존 시스템의 비만도가 37%나 높게 나타났다.

본 연구는 하나의 대학취업지원 데이터베이스를 선택하여 제안한 대학취업지원 데이터베이스의 비만도를 비교 분석하여 대학 취업정보시스템으로 일반화하기에는 한계점이 있다. 차세대 정보시스템을 구축하여 운영 중인 대학들을 비교하지 못한 한계점이 있다. 본 연구를 기반으로 향후에 개발하는 정보시스템은 업무분석을 통하여 데이터 중심의 데이터모델링

을 개발되어야 할 것이다. 데이터를 저장할 공간을 효율적으로 확보하고 데이터의 중복성을 최소화하기 위해서는 효율적인 데이터모델링이 선제 조건인 것이다.

참고문헌

[1] Richard Y. Wang, Henry B. Kon, and Stuart E. Madnick, "Data quality requirements analysis and modeling", Proceedings. IEEE Ninth International Conference on Data Engineering, pp. 670 - 677. April 1993.

[2] C. B. Cinzia Cappiello, and C. Francalanci, "A. Maurino, Methodologies for data quality assessment and improvement", ACM Computing Surveys, Vol. 41, No. 3, pp. 52-55, 2009.

[3] C. W. Fisher, and B. R. Kingma, "Criticality of data quality as exemplified in two disasters", Information Systems, Vol. 39, pp.109-116. 2010.

[4] "[DB]Elements of a good data model", <http://blog.daum.net/fmddn/1787002>.

[5] D. Katz, M. Bommaroti, and J. Zelner, "The data deluge, The Economist", March 2010.

[6] Min-Kyu Lee, "Data Performance Case Study through Removing of Data Duplicate Relationships", Soongsil University, 2010.

[7] T. Shanker, and M. Richtel, "Data overload can be deadly", The New York Times, Jan 16, 2011.

[8] I. Davies, P. Green, M. Rosemann, M. Indulska, and S. Galo, "How do practitioners use conceptual modeling in practice?", Data and Knowledge Engineering, Vol 58, pp. 358-380, 2006.

[9] "Practical Projects Application of Data Model Normalization / De-normalization", <http://blog.naver.com/jooyong3/40035951092>.

[10] Hye-Kyung Rhee, Hee-Wan Kim, "A Study on the Data Modeling decreasing the Data Obesity", The Society of Digital Policy & Management, The Journal of Digital Policy & Management The Society of Digital Policy & Management, Vol. 11, No. 11, pp.359-366, Nov 2013.

[11] Kook-Hee Lee, "A Study on the Database Quality Assessment", Korea Database Agency, Dec, 1995.

[12] James Martin, "Information Engineering", Vol. 1, Vol.2, Vol.3, Prentice-Hall, 1989.

[13] Peter Chen, "The Entity Relationship Model -Toward a Unified View of Data", ACM, Vol.1, No.1, 1976.

저 자 소 개



이 혜 경
 1972: 숭실대학교 전자계산학과 공학사
 1985: University of Illinois (Urbana-Champaign) 공학석사(전산학)
 2000: 성균관대학교 공학박사(컴퓨터공학)
 현 재: 용인송담대학교 컴퓨터게임과 교수
 관심분야: 데이터베이스, 데이터모델링, 정보보호
 Email : leehk@ysec.ac.kr



김 희 완
 1995: 성균관대학교 공학석사(컴퓨터공학)
 2002: 성균관대학교 공학박사(컴퓨터공학)
 1996: 정보관리기술사 취득
 현 재: 삼육대학교 컴퓨터학부 교수
 관심분야: 정보보호 및 보안, 데이터베이스, 정보시스템 감리, 데이터모델링
 Email : hwkim@syu.ac.kr