

데이터베이스 품질 평가에 관한 사례 연구

이 춘 열* · 박 현 지**

A Case Study on Database Quality and Quality Factors

Choon Yeul Lee* · Hyun Jee Park**

Abstract

This paper presents case studies on database quality using an extended database quality model developed by Korea Database Promotion Center, which shall be called KDPC2003. The model is applied to evaluate two kinds of databases; one is an operational database, the other is an information service database. The purpose of this research is two-folded. One is to evaluate database quality and assess current status of database quality management; the other is to assess usefulness of KDPC2003 and to propose ideas for its augmentation. The findings in this study will provide basic facts to test database quality evaluation models.

Keywords : Database Quality Evaluation Model, Data Quality, Data Evaluation Guidelines

1. 서 론

정보화의 발달에 따른 가장 눈에 띄는 현상은 데이터베이스의 확산이라고 할 수 있다. 이전에는 개인이나 부서 단위로 이용되는 데이터들이 이제는 전사적으로 그리고 공개적으로 공유되게 되었으며, 이러한 현상은 일상적인 활동이 되고 있다.

이와 같이, 데이터베이스가 확산됨에 따라 데이터베이스에 대한 품질 관리 또한 점점 중요한 이슈로 부각되고 있다. 많은 사용자들이 데이터를 공유하게 됨으로써 데이터의 내용이나 의미 등에 친근하지 않은 일반 사용자들이 데이터를 사용하게 될 기회가 늘어나게 되었다. 그리고 이러한 일반 사용자들의 정보 요구를 보다 효과적으로 지원하기 위해서는 높은 품질의 데이터 서비스가 필요하게 되었다.

데이터베이스 품질 향상을 위하여서는 데이터베이스 품질을 체계적으로 평가하고 이를 개선하기 위한 평가 모형이 필요하다. 외국의 경우 데이터베이스 품질평가 모형 및 지침들이 보편화되어 있으며, 동태적 의사결정 환경에서의 효과적인 데이터 품질 관리 방안 등과 같은 다양한 연구들이 활발하게 이루어지고 있다 [English, 1999 ; Kahn et al., 2003 ; Yair & Richard, 1996]. 그러나 국내에서는 대부분의 연구가 평가 척도에 대한 연구들로서 데이터베이스 품질에 대한 실증 연구는 아직 부족한 점이 많다. 국내에서는 한국데이터베이스진흥센터가 1998년 데이터베이스 품질 평가 지침, 데이터베이스 구축방법론, 2002년 데이터베이스 품질 평가모델 개발, 그리고 2003년 데이터베이스 품질평가확장모델 연구까지 주도적으로 관련 연구를 수행하여 왔다[한국데이터베이스진흥센터, 1998a ; 1998b ; 2002 ; 2003].

본 연구는 한국데이터베이스진흥센터의 데이터베이스 품질평가확장모형을 실제 데이터베이스

에 적용함으로써 모형의 실용성을 검토하고, 데이터베이스 품질 현상들을 평가한다.

이를 위하여, 제2장에서는 데이터베이스 품질 평가확장모형을 소개한다. 제3장에서는 외부 사용자를 위한 정보서비스 데이터베이스와 내부 업무를 위한 거래처리 데이터베이스를 대상으로 데이터베이스 품질에 대한 실증 분석을 실시한다. 그리고 분석된 결과를 토대로 제4장에서는 데이터베이스 품질평가확장모형을 적용할 때 고려되어야 할 요인들 및 모형의 실용성을 평가하고, 데이터베이스 품질 향상을 위한 방안을 본 연구의 결론으로 제시한다.

2. 데이터베이스 품질평가확장모형 및 절차

데이터 품질에 대한 연구는 오래 전부터 정보화와 관련하여 주요 이슈로 다루어져 왔으며, 데이터베이스의 품질 향상을 위한 노력 또한 활발하게 이루어지고 있다. 특히 최근에는 데이터 정제, 표준화 및 매칭 등의 작업을 통해 데이터의 품질을 향상시키는 솔루션들이 시장에 출시되고 있으며, 국내에도 소개되고 있다[Haward, 2004].

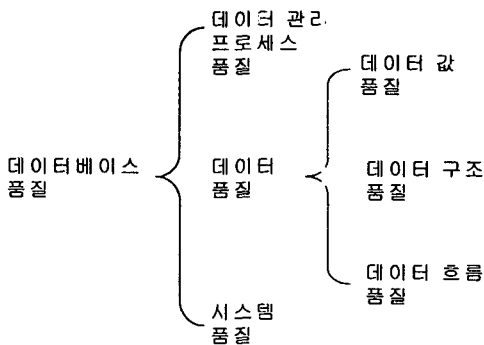
데이터 품질에 대한 초기 연구들은 대부분 정확성 측면에서 데이터 품질을 논하고 있다. 즉 실제 활용에서 문제가 되는 데이터베이스의 형태나 구조 및 데이터 관리 프로세스 등은 포함하지 않고 있다. 그러나 최근에는 실제 데이터 활용시 주요 이슈들인 데이터의 유용성, 적시성 및 완전성 등이 데이터의 품질 요소로 포함되고 있다. Wang은 “데이터의 품질은 해당 데이터가 데이터 이용자들이 사용하기에 얼마나 적합한가를 나타내어야 한다”라고 정의하고, 실제 사용자가 중요하게 생각하는 데이터 품질 요인에 관한 연구를 수행하여 총 20여개의 요소를 사용자 관점에서의 데이터 품질 요소로서 제시하였

다[Wang, 1998].

이러한 데이터베이스 품질에 대한 국내 연구로서 한국데이터베이스진흥센터는 데이터베이스 품질기준을 정확성(Accuracy 또는 Correctness), 완전성(Completeness), 현행성(Currentness), 일관성(Consistency), 검색성(Searching), 사용 용이성(Ease of Use) 및 사용자 지원성(Customer Support) 등 7가지로 분류하였다[13]. 또한 이러한 품질기준을 확장하여, 데이터베이스 품질을 데이터 품질, 데이터 관리 프로세스 품질 및 시스템 품질로 정의하였으며, 이러한 기준에 의해 데이터베이스 품질평가확장모델을 개발하였다[한국데이터베이스진흥센터, 2003].

2.1 데이터베이스 품질평가확장모형

한국데이터베이스진흥센터에서 제시한 데이터베이스 품질평가확장모델은 데이터베이스 품질을 데이터 품질과 시스템 품질뿐만 아니라 데이터 관리 프로세스 품질 또한 포함하도록 확장한다. 즉 <그림 1>에서 제시되는 바와 같이, 데이터 품질, 데이터관리 프로세스 품질 및 시스템 품질을 데이터베이스 품질의 구성요소로 포함한다.



<그림 1> 데이터베이스 품질평가확장모형의 프레임워크

이러한 관점은 최종 산출물의 품질만이 아니

라 이에 영향을 미치는 제반 요인들을 같이 평가하는 종합품질관리(Total Quality Control) 프레임워크와 크게 다르지 않다.

데이터베이스 품질평가확장모형에서 데이터 품질은 데이터값, 데이터구조 및 데이터흐름의 품질을 포함한다. 여기서 데이터 값은 전산화된 데이터를 말하며, 데이터 구조는 데이터를 효과적으로 저장, 관리, 서비스하기 위한 표준 및 모델을 의미한다. 그리고 데이터 흐름은 원시 데이터(문서, 데이터베이스 등)로부터 데이터를 생성하거나 추출, 변환, 적재하는 것을 말한다. 데이터에 대한 품질평가는 이들 데이터의 값, 구조 및 흐름에 대하여 완전성, 일관성, 최신성, 정확성을 평가한다. 데이터 품질을 평가하기 위한 평가 항목들을 평가 대상과 평가 기준별로 정리하면 <표 1>과 같다.

데이터관리 프로세스란 데이터베이스 품질관리와 관련된 활동을 포함하는 개념으로, 데이터 관리 정책, 요구사항관리, 표준관리, 모델관리, 데이터베이스관리 및 데이터 생성/가공관리 등을 포함한다. 데이터 관리 프로세스에 대한 평가는 이들 프로세스들이 적절히 잘 정의되어 있는가와 이들이 실제로 잘 운영되는가를 나타내는 척도인 프로세스의 적절성과 운영성을 평가한다. 데이터 관리 프로세스를 평가하기 위한 평가 항목들을 평가 대상과 평가 기준별로 정리하면 <표 2>와 같다.

시스템이란 데이터품질을 안정적으로 유지하기 위한 시스템을 말하며, 서버 등과 같은 하드웨어, 데이터베이스 서비스를 위한 소프트웨어 및 네트워크를 포함한다. 시스템에 대한 평가는 이들이 안정되게 잘 운영되고, 보안되며, 사용하기에 편리한가를 나타내는 척도인 안전성, 보안성 및 편의성을 평가한다.

〈표 1〉 데이터 부문 품질평가 모델

평가 대상 평가 기준	데이터 평가		
	데이터 값	데이터 구조	데이터 흐름
완전성	<ul style="list-style-type: none"> 데이터 크기 데이터 범위 데이터 값 누락 여부 	<ul style="list-style-type: none"> 중요속성 누락 여부 필수 속성의 설계 반영 여부 	<ul style="list-style-type: none"> 데이터 생성/가공시 누락
일관성	<ul style="list-style-type: none"> 데이터 속성정의와 값의 일치 여부 데이터 제약조건과 값의 일치 여부 테이블 정의와 값의 일치 여부 동일 데이터의 상호 일관성 	<ul style="list-style-type: none"> 데이터 표준 정의의 적절성 도메인 정의의 적절성 코드 정의의 적절성 	<ul style="list-style-type: none"> 데이터 생성/가공시 데이터 표준 적용 여부
최신성	<ul style="list-style-type: none"> 최신 데이터의 제공 여부 	-	<ul style="list-style-type: none"> 데이터 갱신 주기
정확성	<ul style="list-style-type: none"> 데이터 오탈자 실제 사실과 일치 여부 레코드 중복 	<ul style="list-style-type: none"> 참조 무결성 속성 중복, 유일성 보장 여부 	<ul style="list-style-type: none"> 원천 데이터의 신뢰성 데이터 생성/가공 시 오류 및 중복

〈표 2〉 데이터 관리 프로세스 부문 품질평가 모델

평가 대상 평가 기준	데이터 관리 프로세스 (데이터 관리 정책, 데이터 표준 관리, 데이터베이스 관리, 요구사항 관리, 데이터 모델 관리, 데이터 생성/가공 관리)
적절성	<ul style="list-style-type: none"> 정책의 적절성 프로세스 규정의 적절성
운영성	<ul style="list-style-type: none"> 인력의 운영성 교육의 운영성 프로세스의 운영성 개선작업의 운영성

그러나 데이터관리프로세스 부문의 평가 기준인 적절성과 운영성은 기존의 평가 모형에는 포함되어 있지 않은 새로운 평가 기준들을 알 수 있다. 이와 같이 데이터관리프로세스 부문의 평가가 포함된 이유는, 앞에서 언급된 바와 같이 품질평가확장모형은 최종 산출물인 데이터의 품질만이 아니라 이에 영향을 미치는 제반 요인들을 같이 평가하는 종합품질관리 프레임 워크에 기반한 평가모형이기 때문이다.

시스템을 평가하기 위한 항목들을 평가 대상과 평가 기준별로 정리하면 <표 3>과 같다.

이상에서 설명한 데이터베이스 품질평가확장모형을 기존의 평가모형들과 비교하여 보면 <표 4>와 같다. <표 4>는 기존 데이터 품질 평가 모형들 중의 하나인 Wang의 품질 모형과 품질평가확장모형을 비교한 비교표이다[Wang, 1998].

<표 4>에서 나타난 바와 같이, 품질평가확장모형에서 제시되는 평가 기준들 중에서 데이터 부문의 품질평가 기준과 시스템 부문의 품질평가 기준들은 기존의 품질 모형들에 직·간접적으로 포함되어 있음을 알 수 있다.

〈표 3〉 시스템 부문 품질평가 모델

평가 대상 평가 기준	시스템 (하드웨어, 소프트웨어, 네트워크)
안전성	<ul style="list-style-type: none"> 운영 현황관리의 적절성 시스템 유지보수 방안의 적절성 재해관리 방안의 적절성 전담 관리자 지정 여부
보안성	<ul style="list-style-type: none"> 보안관리의 적절성 물리적 접근 통제의 적절성 사용 통제의 적절성
편의성	<ul style="list-style-type: none"> 사용 용이성 검색성 성능

〈표 4〉 기존 데이터 품질평가 모형과 데이터베이스 품질평가확장모형과의 비교표

Wang의 데이터 품질 평가 모형		데이터베이스 품질평가확장모형						
품질 카테고리	품질카원	데이터 부문 (값/구조/흐름) 품질 평가 기준	데이터관리 프로세스 부문 품질평가기준	시스템 부문 품질 평가 기준				
내재적 품질	정확성	정확성						
	무결성	"						
	일관성	일관성						
표현적 품질	표현일관성	"				편의성		
	해석용이성						안정성	
접근적 품질	접근성						편의성	
	성능							보안성
	보안							
상황적 품질	유용성	최신성						
	적시성	완전성						
	완전성							
			적절성					

2.2 데이터베이스 품질 평가 절차

데이터베이스 품질평가확장모형은 샘플링을 이용한 객관적인 측정, 전문가의 주관적 평가 및 사용자를 대상으로 한 설문 평가를 포함한다. <표 1>부터 <표 3>의 평가 모델에서 소개된 평가 항목들을 평가 방법에 따라 분류하면 <표 5>에 제시된 바와 같이 3가지 유형으로 분류된다.

2.2.1 표본 데이터를 이용한 측정에 의한 평가

데이터베이스 품질평가에서 가장 중심을 이루는 것이 평가 대상인 데이터를 직접 측정하고 비교하는 방법이다. 이를 위하여 먼저 무작위 추출법에 따라 표본 데이터를 추출하였다. 이렇게 생성된 표본 데이터를 이용하여 평가항목별로 평가절차에 따라 데이터베이스의 완전성, 일관성, 최신성 및 정확성을 평가하였다.

2.2.2 자료 실사와 면담에 의한 평가

데이터베이스의 관리 절차나 시스템의 운영 상황 등을 평가하기 위하여 먼저 관련 자료를 수집하였다. 이렇게 수집된 자료를 검토한 후 면담 일정에 따라 담당자와의 면담을 수행하여 데이터 관리 프로세스 및 시스템 품질을 평가하였다.

2.2.3 사용자 설문 조사에 의한 평가

시스템 편의성에 대한 평가항목 중에서 최종 사용자가 느끼는 사용 용이성은 인터페이스, 도움말 기능 등에 대한 설문을 통하여 평가하였다. 그리고 데이터베이스의 성능과 검색성 또한 사용자의 주관적 판단을 설문을 통하여 분석하였다.

사용자 설문을 위하여 사용된 문항은 다음과 같다.

〈표 5〉 평가 항목별 평가 방법 분류

평가 방법	평가 항목
표본 데이터를 이용한 측정에 의한 평가	<ul style="list-style-type: none"> ◦ 데이터 범위 ◦ 데이터 값 누락 ◦ 데이터 속성정의와 값의 일치 여부 ◦ 데이터 제약조건과 값의 일치 여부 ◦ 테이블정의와 레코드의 일치 여부 ◦ 동일데이터의 상호일관성 ◦ 최신 데이터 제공 여부 ◦ 데이터 오/탈자 ◦ 실제사실과의 일치 여부 ◦ 레코드 중복 ◦ 데이터 생성/가공시 누락 ◦ 데이터 생성/가공시 표준 적용 여부 ◦ 데이터 생성가공시 오류 및 중복
자료 실사와 면담에 의한 평가	<ul style="list-style-type: none"> ◦ 데이터 크기 ◦ 중요속성 누락 여부 ◦ 필수속성의 설계 반영 여부 ◦ 데이터 표준 정의의 적절성 ◦ 도메인 정의의 적절성 ◦ 코드 정의의 적절성 ◦ 참조무결성 ◦ 속성 중복 및 유일성 보장 여부 ◦ 데이터 갱신 주기 ◦ 원천 데이터의 신뢰성 ◦ 정책의 적절성 ◦ 프로세스 규정의 적절성 ◦ 인력의 운영성 ◦ 교육의 운영성 ◦ 프로세스의 운영성 ◦ 개선작업의 운영성 ◦ 운영 현황 관리의 적절성 ◦ 시스템 유지보수 방안의 적절성 ◦ 재해관리방안의 적절성 ◦ 전담관리자 지정 여부 ◦ 보안관리의 적절성 ◦ 물리적 접근 통제의 적절성 ◦ 사용통제의 적절성
사용자 설문 조사에 의한 평가	<ul style="list-style-type: none"> ◦ 사용용이성 ◦ 검색성 ◦ 성 능

- ① 화면상에 나타나는 각 메뉴들이 각 주제와 기능에 적합하게 구성되어 있다고 생각하십니까?
- ② 화면상에 나타나는 각 메뉴들이 간단명료하고 이해하기 쉽게 구성되어 있다고 생각하십니까?
- ③ 화면상의 타이틀명 및 안내 메시지가 사용상 불편이 없도록 적절히 제공되고 있다고 생각하십니까?
- ④ 이용 목적에 맞는 다양한 검색조건이 제공되고 있다고 생각하십니까?
- ⑤ 검색조건에 따른 검색결과에 대해 만족하십니까?
- ⑥ 검색결과의 출력방식에 만족하십니까?
- ⑦ 접속속도는 적당하다고 생각하십니까?
- ⑧ 사용시 과도한 접속으로 인한 서비스 중단을 경험한 적이 있습니까?

- ⑨ 사용시 검색에 따른 응답속도에 만족하십니까?
- ⑩ 일정한 속도로 안정적으로 서비스되고 있다고 생각하십니까?

이들 각 문항에 대하여 (1) 매우그렇다 (2) 다소그렇다 (3) 보통이다 (4) 다소그렇지않다 (5) 전혀 그렇지않다의 5단계 척도로 측정하였다.

3. 데이터베이스 품질평가

본 장에서는 제2장에서 소개된 품질평가확장모형을 적용하여 기업 내부 업무 처리를 위한 고객포인트관리 데이터베이스와 외부 정보 서비스를 위한 인물정보 서비스 데이터베이스를 평가한다.

3.1 고객포인트관리 데이터베이스 개요

평가 대상으로 선정한 고객포인트관리 데이터베이스는 국내외 수백만 명의 회원에 대한 고객 포인트 정보를 관리하는 대용량 데이터베이스이다. 고객들에 대한 포인트 정보를 관리함과 더불어 전략적인 제휴사들과 회원의 이용 내역에 대한 자료를 교환하며, 포인트 교환 자료를 관리한다. 회원들은 인터넷을 통하여 포인트 정보를 직접 조회할 수 있다.

고객포인트관리 시스템은 회원이 서비스를 활용할 때마다 해당 기준에 따라 포인트를 적립하는 거래처리시스템, 각종 제휴사와 포인트 정보를 제휴하고 회원별로 상세 정보를 관리하는 데이터베이스 서버, 인터넷을 통하여 회원가입과 포인트 관련 정보 및 회원정보 수정을 수행하는 웹 서버로 구성되어 있다.

본 연구의 평가 대상인 고객포인트관리 데이터베이스 시스템은 전형적인 중일집중형 시스템으로 대용량의 데이터 처리에 장점을 가지고 있다.

3.2 고객포인트관리 데이터베이스의 품질평가 결과

고객포인트관리 데이터베이스의 품질과 관련된 자료의 수집 및 분석은 2003년 9월초부터 약 5개월간 이루어졌다[박현지 & 이춘열, 2004]. 2003년 9월초부터 2003년 12월 말까지 약 4개월간 데이터 품질향상을 위한 데이터 정제(Data cleansing) 계획 수립 프로젝트에 참여하여 시스템의 전반적인 구조를 상세하게 파악하였다. 그리고 2004년 1월초에는 실제 고객포인트관리 데이터베이스에 접근하여 여러 SQL 문장 및 분석도구를 활용하여 데이터베이스 품질평가확장모형에서 설정한 평가항목에 따라 평가하고 이를 종합하였다.

3.2.1 데이터 품질 평가

고객포인트관리 데이터베이스에 저장된 데이터들의 품질 평가를 위하여 전체 회원의 약 1%인 90,000명의 자료를 무작위로 추출하여 데이터베이스 품질평가확장모형에서 설정한 평가항목에 따라 품질을 측정하였다.

(1) 정확성

고객포인트 데이터가 정확히 관리되고 있는가를 평가하기 위하여 데이터베이스 및 타 시스템에서 관리되는 자료가 실제 값과 동일한지를 측정하였다. 즉, 회원의 통합포인트 정보가 개인의 이용이력 혹은 제휴사 실적과 정확히 일치하는지와 거래처리시스템에서 관리하는 값과 고객포인트관리 데이터베이스에서 관리하는 값이 서로 일치하는가를 측정하였다.

<표 6>에서 정리된 바와 같이 마스터와 이력 테이블 사이에 데이터 불일치가 발생하고 있는 것을 알 수 있다. 이러한 데이터 불일치가 발생하는 것은 프로그램의 비정상적인 종료 혹은 오류가 발생할 경우 근본적인 문제해결보다는 마스터 테이블만을 비정상적으로 갱신한 것이 그 원인으로 사료된다. 이러한 오류는 프로그램에 ROLLBACK 문장을 철저히 기술하여 프로그램의 비정상적인 종료에 대비하고, 검색 기능 이외의 SQL 문장 사용을 지양하고 수작업에 의한 데이터 갱신을 방지함으로써 근본적인 원인들을 해결할 수 있다. 현재 고객포인트관리 데이터베이스에서 이러한 불일치가 발생하는 이유는 수작업으로 마스터 테이블을 정정하기 때문으로 유추된다.

제휴사와의 포인트 교환 데이터는 상대적으로 부정확한 정도가 심한 편이다. 이와 같은 오류가 발생하는 이유는 포인트 데이터를 교환하는 과정에서 오류가 발생하기 때문으로 추정된다. 즉 문자열 형태로 관리되는 회원번호와 성

명이 잘못 기재되어 자료가 교환될 경우 시스템에 포인트를 누적할 방법이 없게 된다. 이러한 오류는 각 회사가 상대방 시스템에 접속하여 회원자료를 공유한다면 근본적인 해결이 가능하다. 그러나 현재는 수작업 검증(eye check)으로 데이터를 보정하고 있으며, 이러한 과정에서 오류들이 발생하고 있다.

<표 6> 고객포인트관리 데이터베이스의 정확성 평가

측정 항목		평가대상 건수	불일치 건수	비율(%)	평가 점수 ¹⁾
마스터와 이력 데이터의 일치 여부		9,237,351	499	0.0054	미흡 (2)
제휴사 포인트 교환 데이터와의 일치 여부	청구 건수	9,729	94	0.9669	
	청구 포인트	7,089,292	150,211	2.1188	

(2) 일관성

시스템에서 제공하는 정보의 일관성이 유지되는가를 평가하기 위하여 <표 7>에서 보는바와 같이 명명규칙(naming rule)의 준수, 데이터 형식 등을 측정하였으며, 데이터 값의 저장 형식도 이 범주에 포함하였다.

명명 규칙에 대한 측정 결과는 미흡한 점이

1) 데이터베이스 품질평가확장모형의 평가 점수는 4점 기준으로 구성되어 있다. 가장 우수한 평가가 4점이며, 가장 미흡한 평가가 1점이다. 일례로 정확성의 평가 항목 중 실제 사실과의 일치 여부에 대한 평가 기준은 다음과 같다.

- 4점 : 실제 사실과 일치하지 않는 비율이 0%일 경우
- 3점 : 실제 사실과 일치하지 않는 비율이 0% 초과~0.5% 이하일 경우
- 2점 : 실제 사실과 일치하지 않는 비율이 0.5% 초과~1% 이하일 경우
- 1점 : 실제 사실과 일치하지 않는 비율이 1% 초과일 경우

본 논문에서는 이들 숫자로 표시된 평가 점수에 대한 이해를 돕기 위하여 [매우 우수], [우수], [미흡], [매우 미흡]의 평어를 추가하여 같이 표시하였다.

매우 많다. 이는 10여 년 전 개발을 시작하여 서로 다른 개발 업체 및 개발자가 참여하여 조금씩 기능을 확장한 이력을 고려할 때 충분히 이해할 만한 상황이다. 즉 1995년 이후 서로 다른 3개 업체가 개발 문서를 참조하지 않고 각각에 의지하여 업무를 수행한 것이 그 원인으로 추정되고 있다²⁾.

데이터 형식 또한 일관성이 결여된 경우들이 발생하고 있다. 특히 날짜 항목의 경우 문자(Character : 8Bytes)와 숫자(Fixed Decimal : 9 Bytes)를 각 테이블 별로 혼용해서 활용하고 있으며, 시스템 날짜 관리 항목의 경우도 문자(YYYYMMDD : 8Bytes) 및 Date(YYYY-MM-DD : 10Bytes)를 혼용해서 활용하고 있는 실정이다. 또한 소수점 자리수도 일정한 규정을 지키지 않아 동일한 항목인 경우에도 전체 길이가 상이함은 물론 소수점 자리수도 다르게 표현되고 있는 경우들이 발생하고 있다.

내부저장 데이터 값 또한 각 항목에 대한 제약 조건이 공통적으로 적용되고 있지 않다. 공통 항목들에 대하여 각 프로그램에서 표준적인 관리 프로그램을 공동으로 활용하는 방식을 도입하지 않고 있는 것이 주요 원인으로 추정되고 있다. 특히 메인프레임 컴퓨터가 EBCDIC 코드체계를 활용함으로써 한글과 영문자가 혼합하여 입력되는 경우 입력 규정을 점검하고 준수하는 것이 매우 어려운 것도 하나의 원인으로 판단된다.

2) 고객포인트관리 데이터베이스의 약어 규정은 영문 이니셜을 따르는 것을 원칙으로 하며 세 문자로 구성한다. 두 단어로 구성될 경우는 앞 단어에서 두 문자를 추출하고, 다음 단어에서 한 문자를 추출하는 방식을 취하며, 영문자의 모음은 제거하나 모음이 연이어 나오는 경우에는 이를 택한다. 따라서 Booking Class를 나타내는 컬럼명은 BOC(BOOking Class)가 규정을 준수하여 정확히 명명된 것이며, BKC(BooKing Class)는 그 규정을 정확히 이해하지 못하고 잘못 명명한 것이다.

〈표 7〉 고객포인트관리 데이터베이스의 일관성 평가

측정 항목	세 부 내 용	의견
명명규칙 (테이블, 컬럼, 인덱스 등)	필드명의 혼용 <ul style="list-style-type: none"> CLS, BOC, BKC : 거래 등급에 대한 표현 DFR, DTO, DTE : 거래 날짜에 대한 표현 	미흡 (2)
데이터 형식	<ul style="list-style-type: none"> 날짜 : 숫자, 문자를 혼용 시스템날짜 : 문자, Date 혼용 컬럼길이 : 전체 길이 및 소수점자리수 	
내부저장 데이터 값	<ul style="list-style-type: none"> 데이터의 표현 방법 : 규정 미비 ('747', '747-400', '747.100' 등 '0413', '413' 등) 띄어쓰기 규정 준수 결여(주소) 각종 입력 항목의 지양 조건 미비 	

(3) 완전성

완전성은 실제 중요한 객체들과 속성들이 모두 갖추어져 부족함이나 결함기 없는가를 나타내는 기준으로 데이터의 구조, 값 등에 대한 평가를 포괄한다. 거래처리 데이터베이스에서 모든 거래들은 데이터베이스에 포함된다. 따라서 중요한 객체들이 누락될 경우들은 거의 발생하지 않는다. 다만 동일한 객체를 다른 객체로 인식하거나 삭제된 객체를 존재하는 것으로 표시하는 등의 오류로 인하여 데이터베이스가 중요한 객체들을 완전히 나타내지 못하는 경우들이 발생할 수 있다. 이러한 점에서 본 연구에서는 키정의 및 제약조건, 삭제 레코드의 표현, 논리적 무결성 등을 완전성에 포함하여 <표 8>과 같이 조사하였다.

완전성에 대한 평가 중에서 삭제 구분자의 경우 각 서브시스템별로 약간씩 다르게 활용하고 있어 단위 프로그램에서 서로 상이하게 자료를 활용하는 불편함을 감수하고 있으며, 시스템 유지보수에 상당한 어려움을 겪고 있다.

키(Key) 제약에 대한 준수 여부에 대해서는 전체 104개의 테이블 중 최근에 추가된 몇 개의 테이블을 제외하고는 성격상 반드시 주키(Primary Key)와 외래키(Foreign Key)를 정의하고 참조무결성법칙(Reference Integrity)을 정의하여 데이터의 정합성을 유지해야 함에도 불구하고, 별도의 독립된 테이블 구조를 갖도록 구성되어 있다. 전체 104개 테이블이 4개 테이블을 제외하고 주키가 정의되지 않았으며, 단순히 유일키(Unique Key)만을 정의하고 있다. 이에 대한 원인은 다음과 같이 세 가지 사유로 판단된다.

첫째, 고객포인트관리 업무의 특성상 관리되는 데이터의 변동으로 인한 수정이 자주 발생한다. 이러한 키 값에 변동이 생길 경우, 삭제 후 다시 생성하는 절차 없이 개발자의 편의만을 추구하여 단순히 갱신 처리만으로 해결할 수 있도록 테이블을 설계하였다.

둘째, 수년간에 걸쳐 여러 담당자에 의하여 기능들이 추가됨으로써 전체적인 시각을 갖지 못하고 단순히 순간적인 필요에 따라 테이블들을 독립적으로 추가하였기 때문이다.

셋째, 인터넷을 활용한 정보 조회의 증대와 글로벌 업무로 인하여 고객포인트관리 데이터베이스는 24시간 × 365일 운영되고 있다. 이러한 업무의 특성상 테이블 변경 등에 따른 시스템 가동 중단을 고려하지 않을 수 없다. 통상적으로 테이블에 새로운 필드를 추가하고자 할 경우, 데이터 추출, 테이블 재정의, 데이터 적재 과정 등을 거쳐야 하는데, 관리되는 데이터의 양이 많아질수록 상당시간동안(마스터테이블 작업시 최소 6시간 소요) 시스템 운영을 중단해야 하는 상황이 발생한다. 따라서 보다 신속히 테이블 변경을 실행할 수 있는 방법을 선택하였다고 판단된다.

<표 8> 고객포인트관리 데이터베이스의 완전성 평가

측정 항목	평가대상 건수	불일치 건수	의견
시스템 내부 데이터 값	◦ 삭제 Indicator : 'D', '*', 'E' 등		미흡 (2)
주키 정의 및 제약조건	◦ 전체 테이블 : 104개	◦ 주키 미정의 테이블 : 100개 ◦ 주키 정의 테이블 : 4개	
데이터의 중복률	◦ 전체 회원수 : 9,237,531	◦ 중복 회원수 : 21,338 (0.23%)	

(4) 최신성

고객포인트관리 데이터베이스는 거래처리시스템과 별도로 이중으로 관리되고 있다. 이에 따라 포인트 누적 시점에서 약간의 차이가 발생한다. 그러나 최종 포인트 정보가 필요할 때마다 거래처리시스템의 정보가 갱신되어 고객이 느끼는 차이는 발생하지 않는다.

<표 9> 고객포인트관리 데이터베이스의 최신성 평가

측정 항목	평가대상 건수	불일치 건수	비율(%)	평가 점수
거래처리 시스템과 마스터 테이블의 일치 여부	9,237,351	5,202	0.0563	우수 (3)

3.2.2 데이터 관리 프로세스 품질 평가

고객포인트관리 시스템은 데이터베이스 관리 규정 및 프로세스 정의가 미흡한 것으로 평가되었으며, 요구사항 관리 및 표준 관리 또한 미흡한 것으로 평가되었다. 따라서 비효율적인 프로세스를 재정의하고 정기적으로 문제점을 파악하며 개선작업을 수행하지 않을 경우 데이터 품질수준을 향상시키기 어려울 것으로 사료된다.

고객포인트관리 데이터베이스의 업무 절차도 프로젝트의 시행 단계에서는 개발자와 데이터베이스관리자(DBA)가 상호 협력하여 각종 자동화도구(CASE Tool)를 활용하여 상당한 수준

의 시스템 품질을 유지하려고 노력하고 있다. 반면, 유지보수 단계부터는 추가로 관리 항목의 변동이 생기는 경우 자동화 도구가 거의 활용되지 않는 실정이며 개발자의 감각에 의존하고 있다. 데이터베이스 관리자의 역할 또한 단순히 데이터베이스를 생성하고 디스크 공간을 할당하며, 처리권한을 부여하는 등의 단순한 운영업무만을 관장하며, 근본적인 데이터베이스 품질향상을 위한 별다른 노력을 하고 있지 않는 실정이다.

<표 10> 고객포인트관리 데이터베이스의 데이터 관리 프로세스 품질 평가

평가기준	데이터관리정책, 요구사항관리, 데이터표준관리, 데이터모델관리, 데이터베이스관리, 데이터 생성/가공 관리
적절성	미흡 (2)
운영성	미흡 (2)

3.2.3 시스템 품질 평가

시스템 품질은 데이터베이스 시스템이 안정되게 잘 운영되고, 보안되며, 사용하기에 편리한가를 나타내는 척도인 안전성, 보완성 및 편의성을 평가하였다. 이 중에서 안전성과 보완성은 메인프레임의 특성상 매우 우수한 것으로 평가되었다.

(1) 편의성

자료접근의 편의성이란 이용자가 데이터베이스에 접속하고, 관련 정보를 검색하며, 검색된 결과를 제공받는 인터페이스 및 유틸리티 기능의 편리성, 친절성, 이해가능성, 적합성 및 다양성 등을 망라한 개념이라고 할 수 있다. 자료접근의 편의성에 대한 평가 결과는 <표 11>과 같다.

분석 대상시스템이 SNA3270 에뮬레이터를 활용하는 텍스트 자료처리 위주의 시스템으로서 단순히 규정된 정보의 처리에 있어서는 탁월

한 능력을 발휘하는 거래처리시스템이라고 할 수 있다. 그러므로 단축키 및 메뉴를 활용하여 처리하는 방식을 제외하고는 웹 브라우저를 활용한 최근의 웹 환경에서의 응용시스템과는 편의성 측면에서 비교의 대상이 될 수 없다. 가공하지 않고 단순히 규정된 자료의 검색 및 관리의 측면만 본다면 최상의 조건을 구비하고 있으나, 사용자의 요구에 제대로 부응할 수 없는 한계를 지니고 있다.

〈표 11〉 고객포인트관리 데이터베이스의 편의성 평가

측정 항목	평가 점수	비 고
검색화면 인터페이스	미흡 (2)	24*80 칼럼
제공 정보 표현 형식의 다양성	미흡 (2)	
단축키 활용	매우 우수 (4)	
마우스 및 바코드 등의 활용	매우 미흡 (1)	Dummy형 회원카드

(2) 성능(신속성)

자료접근의 신속성은 사용자가 해당 자료를 얼마나 신속하게 취득하여 활용할 수 있는가에 대한 평가 항목으로 메인프레임 컴퓨터의 특성 상 각종 규정된 정보는 신속하게 제공하고 있다³⁾. 그러나 각종 조건의 조합에 따른 정보의 취득은 시스템 담당자를 통하지 않고는 매우 어려운 실정이다. 물론 일부 사용자는 SQL 문장을 활용하여 직접 자료를 확보하는 경우도 있으나 이는 매우 제한적인 경우이다.

3) 한국데이터베이스진흥센터에서 개발 제시한 데이터베이스 품질평가확장모형은 시스템의 신속성과 성능을 시스템의 편의성을 평가하는 항목으로 포함하고 있다. 그러나 본 연구에서는 이의 중요성을 고려하여 자료접근의 신속성을 나타내는 성능을 편의성과 구분하여 별도로 분석하였다.

고객포인트관리 데이터베이스는 회원수에 비하여 그다지 많은 이력자료를 보유하고 있지 않으며, 관리되는 레코드의 길이도 그다지 길지 않은 실정이다. 반면에 메인프레임 컴퓨터 환경에서 데이터가 처리되기 때문에 온라인 트랜잭션 처리 및 대용량의 배치파일 처리에 있어서 이상적인 능력을 발휘하고 있다. 하지만 운영계와 정보계 데이터베이스가 구분되어있지 않기 때문에 온라인 트랜잭션이 주로 처리되는 주간 업무시간에는 정보 획득을 위한 질의 문장의 활용을 제한하고 있다.

〈표 12〉 고객포인트관리 데이터베이스의 신속성에 대한 평가 내용

측정 항목	세부내용	의견
온라인 프로그램	3초 이내가 대부분이나 특정 프로그램의 경우 10초 정도 수행	미흡 (2)
배치 프로그램	메인프레임 컴퓨터의 활용으로 일부 작업을 제외하고 10분 이내에 수행됨	
SQL문장의 활용	시스템에 부하를 가하기 어렵고 업무시간에 대용량의 자료 추출을 금함	
사용자의 요청 후 정보가 제공되기까지 소요시간	자료 요청부터 수령까지 평균 2주 소요됨	

데이터베이스 구성에 대한 성능을 평가하기 위하여 이력 테이블들이 전체적으로 균형을 유지하는지를 조사하였다. 발행된 회원번호별 분류(지역별 번호 할당)된 이력 테이블은 <표 13>에서 보는 바와 같이 특정 테이블에 50%의 자료가 집중되어 관리되는 심한 불균형을 보이고 있으며, 이로 인한 검색 성능의 저하가 발생하고 있다.

〈표 13〉 고객포인트관리 데이터베이스 이력 테이블의 Load Balancing 측정

측정 항목	테이블명	레코드수	점유비 (%)
포인트 이력 데이터 Load Balancing 측정	포인트이력-1	4,347,335	5.43
	포인트이력-2	5,638,005	7.04
	포인트이력-3	4,616,044	5.76
	포인트이력-4	5,439,888	6.79
	포인트이력-5	5,708,432	7.13
	포인트이력-6	8,342,308	10.41
	포인트이력-7	40,027,121	49.97
	포인트이력-8	5,765,648	7.20
	포인트이력-9	214,663	0.27

3.3 인물정보 서비스 데이터베이스 개요

정보 서비스 데이터베이스는 일반 외부 고객들을 상대로 뉴스와 같은 각종 정보들을 서비스하는 데이터베이스이다. 이러한 정보 서비스 데이터베이스들 중에서 본 사례에서는 인물정보 서비스를 위한 데이터베이스를 평가하였다.

인물정보 서비스 데이터베이스는 국내 주요 인사 15만 여명의 인물정보를 제공한다. 이름, 소속기관, 직업, 출생지, 출생년도, 학력 경력 등의 기본 정보와 연락처 및 관심사 등의 개인 정보들을 유료로 제공한다.

3.4 인물정보 서비스 데이터베이스의 품질평가 결과

인물정보 서비스 데이터베이스의 평가는 2003년 8월부터 약 3개월에 걸쳐 품질평가확장모형의 시범적용을 위하여 이루어졌다[이춘열외, 2003]. 2003년 8월말부터 실무자 면담을 통하여 시스템의 전반적인 구조를 상세하게 파악하였으며, 실제 인물정보 서비스 데이터베이스에 접근하여 여러 SQL 문장 및 분석도구(Tool)를 활용하

여 데이터베이스 품질평가확장모형에서 설정한 평가항목에 따라 평가하고 이를 종합하였다.

3.4.1 데이터 품질 평가

인물정보 데이터의 품질 평가를 위하여 전체 데이터의 약 1%에 해당하는 100명의 자료를 무작위로 추출하여 이들의 값을 측정하였다. 표본 데이터는 데이터베이스 이용자들이 관심 있을 사람들로 구성하고자 하였다. 여기서 관심 있다고 생각하는 기준으로는 신문에 게재된 인물 중 중요도 있게 실린 인물들을 사용하였다.

표본 데이터 추출 과정을 보다 자세히 설명하면, 일간지를 무작위로 선정하여 게재일을 추출하였다. 추출 방법은 1991년부터 2003년까지 국내 주요일간지를 무작위로 추출하여 해당 일간지에 게재된 인물들을 선정하였으며, 직업분류를 기준으로 특정 직업에 편중되지 않도록 하였다. 이렇게 추출된 표본 데이터를 사용하여 데이터의 품질을 평가하였다.

(1) 정확성

인물정보 데이터가 정확히 관리되고 있는가를 측정하기 위하여 데이터 값과 데이터 표현이 실제 값과 동일한지를 측정하였다. 정확성은 3가지 방법으로 이루어졌는데, 먼저 오자나 탈자가 있는가를 검토하였으며, 이들 저장된 데이터들이 원본 데이터와 같은 가를 비교하였다. 그리고 실제 데이터와의 확인을 위하여 인터넷 등과 같은 다른 정보 검색 채널에 의하여 검색된 값과 비교 평가하였다.

〈표 14〉에서 정리된 바와 같이 저장된 데이터들은 오류가 거의 없는 정확한 값을 유지하고 있다. 다만 실제 데이터와의 일치여부는 대상 인물들을 실제로 추적하는 것이 쉽지 않아 인터넷과 같은 다른 매체에서 제공되는 정보를 실제 데이터로 대용하여 비교하였다.

〈표 14〉 인물정보 서비스 데이터베이스의 정확성 평가

측정 항목	평가대상 건수	불일치 건수	비율 (%)	평가 점수
오탈자	371	1	0.2176	우수 (3)
원본 데이터와의 일치 여부	323	0	0.0	
실제 데이터와의 일치 여부	266	0	0.0	

(2) 일관성

시스템에서 제공하는 정보의 일관성이 유지되는가를 측정하기 위하여 <표 15>에서 보는 바와 같이 데이터베이스 명명규칙의 준수 여부, 데이터 항목의 표현 형식 등을 평가하였으며, 데이터 값의 내부 저장 형식 등도 이러한 범주에 포함하였다.

일관성 유지에 대한 측정 결과는 대부분의 경우 모두 만족스러우나 직업 등을 코드로 관리하지 않는 점이 약간 미흡하였다.

〈표 15〉 인물정보 서비스 데이터베이스의 일관성 평가

측정 항목	세부 내용	평가 점수
명명규칙 (테이블, 컬럼, 인덱스 등)	일관성있는 명칭을 사용하고 있으나 명시적인 명명 규칙은 없음	우수 (3)
데이터 형식	성별, 직업 등에 대하여 도메인을 정의하고 입력시 풀다운 메뉴를 이용하여 위배되는 속성이 입력되지 않도록 하고 있음	
내부 저장 데이터 값	직업 등을 코드로 표현하지 않음	

(3) 완전성

완전성은 실제 데이터들이 중요한 객체들과 속성들을 모두 포함하여 부족함이나 결함이 없는가를 평가하는 기준이다. 인물정보 서비스 데이터베이스의 완전성을 평가하기 위하여 찾고자 하는 인물에 대한 정보가 데이터베이스에 모두 포함되어 있으며, 이들에 대한 정보들이 모

두 포함되어 있는가를 측정하였다.

<표 16>에 표시된 바와 같이 선정된 100명의 표본을 기준으로 현재 보유 비율을 측정하였는데, 80명에 대한 정보가 데이터베이스에 포함되어 있었다. 그리고 이들 포함된 인물들의 성별, 생년월일, 양력음력구분, 띠, 출생지, 본관, 종교, 본적, 직업, 현직기관, 자택주소, 병역, 신장/체중, 취미, 사망일, 학력사항, 경력사항, 가족사항, 저서/논문사항 등과 같은 세부 항목을 분석하기 위하여 데이터베이스에 포함된 80명중 50명을 임의 추출하여 평가하였다. 그 결과 전체 1000개 항목 중 168개 항목이 누락된 것으로 평가되었다.

〈표 16〉 인물정보 서비스 데이터베이스의 완전성 평가

측정 항목	평가대상 건수	누락 건수	비율 (%)	평가 점수
인물에 대한 보유 비율	100 명	20 명	20.00	매우 미흡 (1)
인물에 대한 상세 정보 누락	1000 레코드	168 레코드	16.80	

(4) 최신성

인물에 대한 정보는 크게 신문, 잡지 및 공문서를 참조하여 필요한 인물에 대해 정보를 매일 갱신하는 방법과 정기적으로 해당 인물들에게 공문을 보내어 갱신에 필요한 자료를 수집하는 방법 및 매년 관련 인터넷 사이트를 방문하여 변경된 정보들을 수집하여 갱신하는 작업을 병행한다. 이 중에서 우편을 이용한 정보 수집은 약 20%에 그치고 있다.

실제로 저장된 인물들에 대하여 현직기관, 부서/직위, 직장주소, 자택주소, 사망일, 경력사항 등과 같은 주요 항목들에 대하여 저장된 데이터와 인터넷과 같은 다른 매체에서 제공되는 정보들과 최근 수정일들을 비교하였다. 이 결과 전체 167개 항목들 중에서 1개의 항목 (1.83%)에

대하여 최신값이 반영되지 않은 것으로 조사되었다.

〈표 18〉 인물정보 서비스 데이터베이스의 최신성 평가

측정 항목	평가대상 건수	미 반영 건수	비율 (%)	평가 점수
최신값 미반영 항목 수	167 항목	1 항목	1.83	미흡 (2)

3.4.2 데이터 관리 프로세스 품질 평가

대부분의 다른 데이터베이스 관리와 마찬가지로 인물정보 서비스 데이터베이스 또한 개념상의 정책은 있으나, 문서화 되어 관리되지 않고 있다. 즉 대부분의 관리 절차들이 명문화된 문서나 절차에 의존하기 보다는 담당자들의 의견 공유로 진행 된다. 이에 반하여 데이터 수집이나 갱신 등의 프로세스 규정은 잘 정의되어 문서화 되어 있으며, 작업 목록 및 단계별 절차, 세부 방법들은 명확히 명시되어 있다.

3.4.3 시스템 품질 평가

시스템 품질은 데이터베이스 시스템이 안정되게 잘 운영되고, 보안되며, 사용하기에 편리한가를 나타내는 척도인 안전성, 보완성 및 편의성을 평가하였다. 이 중에서 안전성과 보안성은 문서화된 절차는 없으나 경험에 의한 절차가 잘 수행되고 있는 것으로 평가되었다.

(1) 편의성

자료접근의 편의성이란 이용자가 데이터베이스에 접속하고, 관련 정보를 검색하며, 검색된 결과를 제공받는 S/W 인터페이스 및 유틸리티 기능의 편리성, 친절성, 이해가능성, 적합성 및 다양성 등을 망라한 개념이라고 할 수 있다. 자료접근의 편의성에 대한 평가 결과는 <표 19>와 같다. 대체적으로 사용하기 편리한

시스템으로 평가되고 있으나 검색 결과의 출력방식에 대하여서는 불만이 있는 것으로 조사되었다.

〈표 19〉 인물정보 서비스 데이터베이스의 편의성에 대한 평가

측정 항목	평가 점수
검색 화면 인터페이스	매우 우수 (4)
정보제공의 다양성	매우 우수 (4)
검색조건에 따른 검색결과	우수 (3)
검색결과 출력방식	미흡 (2)

(2) 성능(신속성)

시스템의 성능에 대하여서도 대체적으로 만족하는 것으로 조사되었다.

〈표 20〉 인물정보 데이터베이스의 성능에 대한 평가

측정 항목	평가척도
접속속도	매우 우수 (4)
과도한 접속으로 인한 서비스 중단	우수 (3)
사용시 검색속도	매우 우수 (4)

4. 데이터베이스 품질평가 모형의 수정 및 활용 방안

본 사례 연구는 한국데이터베이스진흥센터에서 개발한 데이터베이스 품질평가확장모형의 실용성 평가와 데이터베이스의 품질 현상과 원인을 진단하는 2가지 목적으로 수행되었다. 본 사례에서 분석한 고객포인트관리 데이터베이스는 전형적인 온라인 거래처리 데이터베이스이며, 인물정보 서비스 데이터베이스는 전형적인 정보서비스 데이터베이스이다.

먼저 데이터베이스의 품질 현상에 대한 평가 결과들을 요약하면, 고객포인트관리 데이터베이스의 경우, 데이터 품질평가의 기준으로 설정한 정확성, 일관성에 있어서 상당한 문제

점을 지니고 있는 것으로 평가되었다. 정확성 평가의 경우, 고객포인트관리 데이터베이스는 마스터 테이블과 이력 데이터의 합이 서로 상이한 경우가 발생하여 수시로 이력정보를 근거로 마스터 정보를 수작업 갱신하는 상황이 발생하고 있다.

일관성 평가의 경우 테이블 명명규칙을 따르지 않아 동일한 값을 갖는 컬럼이 서로 다른 이름과 형식을 갖는 경우가 다수 있으며, 축약처리 시 일관성이 없는 경우가 빈번히 나타나고 있다. 이러한 상황으로 인하여 특정 회원이 포인트 이력을 조회하는 경우 시간제한으로 정보가 제공되지 못하는 경우가 발생하고 있다.

데이터베이스 시스템의 평가 기준인 검색의 편의성 및 자료접근의 신속성은 대상시스템이 메인프레임 환경에 기반한 시스템이라는 한계로 인하여 사용자의 욕구를 충족하기에는 근본적인 한계를 가지고 있는 실정이다.

이에 반하여 인물정보 서비스 데이터베이스는 정확성과 일관성에서는 모두 보통 이상의 평가를 받았다. 그러나 사용자들이 찾고자하는 인물이 포함되어 있지 않거나 있더라도 주요 정보들이 누락되어 있는 경우가 많아 완전성의 관점에서 볼 때 매우 미흡한 평가를 받았다. 이러한 문제가 발생하고 있는 이유는 인물정보는 개인적인 자료로서 수집시 어려움을 겪고 있기 때문인 것으로 사료된다.

이상에서 요약한 시범 평가 결과를 근거로 판단해 볼 때, 현재의 데이터베이스 품질평가 확장모형은 데이터베이스의 종류에 따라 평가 결과가 나누어지는 단점을 가지는 것으로 나타난다. 특히 인물정보와 같은 정보서비스 데이터베이스의 경우에는 완전성과 같은 항목이 낮게 평가된다. 이러한 문제점들을 개선하고 실용성을 향상시키기 위한 방안을 요약하면 다음

과 같다.

첫째, 본 사례연구의 대상이 되었던 인물정보 데이터베이스의 경우, 사용자들이 찾고자 하는 인물에 대한 기대치는 거래처리 데이터를 사용하는 사용자의 기대치보다 훨씬 낮다. 이러한 데이터베이스의 완전성을 고객포인트관리 데이터베이스와 같이 거래가 발생함에 따라 자동으로 수집되는 데이터베이스와 직접 비교하는 것은 논리적으로 불합리하다. 그러나 현재의 데이터베이스 품질평가확장모형에서 항목별 평가 기준은 모든 데이터베이스에 대하여 동일하게 설정되어 있다. 이와 같이 모든 데이터베이스에 대하여 동일한 품질 평가 기준을 적용한다는 것은 불합리한 것으로 평가된다. 따라서 데이터베이스 품질 평가는 용도에 따라 평가 기준을 세분화 하는 것이 필요하다.

데이터베이스들은 용도에 따라 보통 정보서비스 데이터베이스와 거래처리 데이터베이스로 구분되고 있다. 거래처리 데이터베이스는 다시 제공하는 데이터가 거래 처리를 위한 필수 데이터인지 또는 보조적인 분석 정보인지에 따라 구분된다. 이러한 점에서 데이터베이스들은 용도와 중요도에 따라 <표 21>과 같이 구분하는 것이 필요하다. 그리고 데이터베이스의 평가기준은 이들 평가유형별로 세분화하는 것이 바람직하다.

<표 21> 용도와 중요도에 따른 데이터베이스 구분

구분		중요도	
		필수 데이터	보조 데이터
용도	거래처리	업무 처리를 위하여 반드시 있어야 하는 데이터(예: 제품 단가, 현 재고량 등)	업무처리를 위한 분석 정보(예: 고객포인트 등)
	정보분석	-	정보서비스 DB(예: 인물 정보, 지역 특산물 정보)

둘째, 현재의 데이터관리 프로세스에 대한 평가는 문서나 절차의 문서화와 전담조직 보유를 위주로 이루어지고 있다. 따라서 대규모의 시스템운영 및 지원인력(데이터베이스 관리자 및 시스템 유지보수인력) 등을 보유하고 있는 조직은 관리가 원활하지 않다고 하더라도 그러한 조직을 보유하고 있는 것만으로도 높은 평가를 획득할 수 있다. 그러나 관리 프로세스의 품질을 결정하는 가장 중요한 요소는 담당자들의 의견공유와 원활한 지원이라고 할 수 있다. 따라서 데이터베이스의 품질 향상을 위해서는 규정 준수와 변경 사항 발생시 형상관리 능력에 대한 평가가 보다 강조되어야 한다.

셋째, 현재의 데이터베이스 품질평가확장모형은 사용자의 관점보다는 운영자의 관점에서 데이터베이스의 품질을 평가하고 있다. 물론 시스템의 편의성 평가에서 사용자들이 느끼는 사용용이성, 검색성, 성능(신속성)을 포함하고 있으나, 이들의 중요성에 견주어볼 때 하나의 평가 기준으로 평가하기에는 미흡하다. 따라서 본 연구에서는 성능(신속성)을 시스템의 편의성과 구분하여 별도의 평가기준으로 분류하여 적용하였다.

보다 적극적으로 사용자들이 느끼는 데이터베이스 품질 평가를 반영하기 위하여서는 본 연구에서 제시한 바와 같이, 성능(신속성)을 시스템의 편의성과 구분하여 별도의 평가기준으로 분류하는 등의 평가 기준 세분화가 필요할 것으로 사료된다.

이상에서 제시한 바와 같이, 서로 다른 품질 특성을 가지는 서로 다른 형태의 데이터베이스들을 평가하기 위하여서는 품질평가의 기준과 절차를 세분화하는 것이 필요하다.

본 사례연구는 데이터베이스 품질평가확장모형을 이용하여 시범적으로 시도한 내용으로서 이러한 사례 연구는 품질평가 모형과 절차를 발

전시키는데 필요한 기초정보들을 제공한다. 또한 최근 국내에 소개되기 시작한 여러 데이터 품질관리 솔루션(데이터 정제, 표준화, 매칭 등의 작업을 통해 데이터 품질을 향상시키는 도구)을 활용하여 품질을 평가한다면 좀더 다양한 분석 결과를 얻을 수 있을 것이다. 그리고 이러한 결과는 데이터베이스의 품질 향상과 품질평가 모형의 발전에 더욱 유용한 정보를 제공할 것이다.

참고 문헌

- [1] 박현지, 이춘열, 고객DB 품질평가 및 개선 방안에 관한 연구 - 고객 포인트관리 DB에 대한 실증적 평가, 한국데이터마이닝학회 2004 춘계학술대회 논문집, 2004년 4월, pp. 103-123.
- [2] 이춘열외, 데이터베이스 품질평가의 실증연구; 한국데이터베이스진흥센터의 품질평가 프레임워크를 중심으로, 2003 데이터베이스 그랜드 컨퍼런스, 2003년 11월, pp. 455-461.
- [3] 한국데이터베이스진흥센터, 데이터베이스 표준 DPCS 006-012, 1998년.
- [4] 한국데이터베이스진흥센터, 데이터베이스 구축 방법론, 1998년.
- [5] 한국데이터베이스진흥센터, 데이터베이스 품질평가 모델 개발 연구, 2002년.
- [6] 한국데이터베이스진흥센터, 데이터베이스 품질평가 모델 확장 개발, 2003년.
- [7] English, L., *Improving Data Warehouse and Business Information Quality*, John Wiley & Sons, Inc., 1999.
- [8] Howard, P., *Data Quality Products : an evaluation and comparison*, Bloor Research, 2004.
- [9] Kahn, B., Strong, D., and Wang, R., "Infor-

mation Quality Benchmarks : Product and Service Performance”, *Communications of the ACM*, Vol. 45, Apr, 2003.

- [10] Pipino, L., Lee, Y., and Wang, R., “Data Quality Assessment”, *Communications of the ACM*, Vol. 45, Apr, 2000.
- [11] Redman, T., *Data Quality for the Information Age*, Artech House, 1996.
- [12] Shankaranarayan, G., Zia, M., and Wang, R., “Managing Data Quality in Dynamic Decision Environments An Information product Approach”, *Journal of Database Management*, Vol. 14, No.4, pp. 14-32, Oct-Dec, 2003.
- [13] Wang, R., “A Product Perspective on Total Data Quality Management”, *Communications of the ACM*, Vol. 41, No. 2, Feb., 1998, pp. 58-65.
- [14] Yair, Wand, and Warg, Richard Y., “Anchoring data quality dimensions in ontological foundations”, *Communications of the ACM*, November, Vol. 39, 1996.

저자소개



이 춘 열

저자는 서울대학교 산업공학과 학사, University of Michigan에서 경영정보학박사(Computer & Information Systems 전공)를 수여받았으며, 이후 한국통신연구개발단을 거쳐 현재 국민대학교 비즈니스IT전문대학원에 재직중이다. 주 관심분야는 데이터 관리, 데이터웨어하우스, 비즈니스 인텔리전스, 경영성과관리, Business Process Management 등이다.



박 현 지

공동 저자는 한양대학교 이학사, 연세대학교 경영정보학석사를 취득하고, 국민대학교 BIT전문대학원 박사과정에 재학중이다. 아시아나항공 운항시스템팀 과장, (주)에어소프트 SI사업부장을 거쳐, 현재는 한국기업평가원 기업컨설팅부 부장으로 재직하고 있다. 주 관심분야는 데이터품질, 고객관계관리(CRM), 데이터웨어하우스, 정보시스템감리 등이다.