

데이터베이스 품질에 관한 연구

이국희*

An Exploratory Study of Database Quality

Kuk Hie Lee

〈요 약〉

데이터베이스 품질문제는 학문적으로나 실무적으로 상당한 관심을 모으고 있지만 아직까지 많은 문제들이 풀리지 않은 상태로 남아있다. 본 연구는 DB 품질에 관한 다양한 관심과 문제들을 분석하고, 이 분야에 관한 체계적 연구흐름의 출발선을 찾고자하는 노력의 일환으로 시도되었다. 우선 지금까지 발표된 관련 연구와 현재 실무에 종사하는 전문가들의 견해를 종합하여 DB 품질에 관한 연구 프레임워크를 고안하고, 그 속에서 다섯가지 연구영역(품질기준, 품질측정, 품질등급, 품질관리, 품질인증)을 제안하였다. 그리고 DB 품질을 분석할 수 있는 일곱개의 품질기준(정확성, 완전성, 일관성, 현행성, 검색성능, 사용용이성, 사용자지원성)을 추출하고, 각 기준별로 이론적 정의를 시도하였다. 그리고 현실적으로 가장 논란의 여지가 많은 계량적 측정도구, 즉, 설문조사에 사용될 68개 문항과 척도를 개발하였다.

I. 서론

데이터베이스(이하 DB라 약칭함)기술과 응용 범위가 확산됨에 따라 온라인 또는 CD ROM의 형태로 다양한 DB가 개발되고 있지만 이러한 양적 팽창현상 이면에는 DB로부터 제공되는 정보의 낮은 신뢰성이나 서비스 기능의 부실함이 점점 심각해지는 문제로 부각되고 있다. 현재 국내에서 대표적인 것으로 인식되는 온라인 DB를 조사한 결과 레코드 70% 이상이 정확하지 못한 정보를 담고 있었으며, 미국 FBI 범죄기록 DB에서 오류 레코드가 50% 이상이라는 보고도 있고 [Laudon, 1986], DB가 담고 있는 불량 데이터로 인하여 업무장애가 초래된 적이 있다고 응답한 사용자가 전체 70%에 달한다는 조사 결과도 발표되고 있다[Wilson, 1992].

소위 "dirty, old, wrong, incomplete, bulky, or unusable" 데이터로 대변되는 DB 품질문제는 어제 오늘의 일이 아니라 컴퓨터가 탄생한 이래 줄곧 이어져 온 고질적인 문제이며, 최근에 이르러 DB사용이 보편화, 대중화 추세에 접어들면서부터 더욱 관심을 모으고 있다. 인터넷 등 통신망 확산과 상용 또는 공공DB의 보급으로 인해 DB 사용자층이 종전의 전문 사용자로부터 PC 앞의 초보사용자, 즉, 일반대중으로 이동되었으며, 이러한 사용자들은 불량 또는 오류 데이터의 위협앞에 거의 속수무책으로 노출되고 있기 때문이다. 전문 사용자는 불량 데이터를 예방하거나 발견하는 나름대로의 경험칙을 가지고 있으며, 스크린에 나타난 데이터가 어떻게 입력이 되었으며 따라서 그 정확도가 어느 정도 될 것인지를 한번 쬐은 생각한다. 그러나 일반 사용자는 스크린 상의 데이터를 그것이 명백한 철자 오류가 아니라면 쉽게

믿어 버리는 경향이 있으며 그만큼 피해의 빈도와 심도가 커질 수 밖에 없다.

일반 사용자가 겪는 이러한 피해는 단순히 개인적인 문제가 아니라 넓게는 정보화 사회의 성공 여부에까지 영향을 미치게 된다. 우리나라의 경우 많은 수의 DB가 급속히 만들어지고 있으나 이러한 DB의 현행화(update) 등 유지보수에 대한 대책은 거의 전무한 실정이며, 따라서 대부분 DB의 품질은 몇년 내로 형편없는 수준으로 떨어질 우려가 있다. 호기심으로라도 이러한 DB에 접근한 사용자는 저품질 데이터에 실망하게 되고, 한번 실망한 사용자는 두번 다시 비싼 사용료를 내고 들어오지 않을 것이며, 사용자층이 메말라 가는 DB의 가치나 수익성은 더욱 쇠퇴해 질 것이다. 그러므로 사용성이나 수익성이 없는 DB에 대한 재투자는 줄어들 것이고, 재투자없이 DB의 현행화 및 발전적 갱신(enhancement)은 커녕 기본적인 유지보수 활동마저 어려워지며, 따라서 DB의 쓰레기화는 더욱 가속화되고 실패의 악순환은 반복될 것이다.

미국이나 유럽의 경우, DB 품질문제는 오래전부터 관심의 대상이 되어 왔으며, 품질을 평가 또는 향상하기 위해 쏟는 노력이 괄목할 만 하다. 컴퓨터 관련 저널에는 새로이 만들어지는 DB를 소개하면서 그 품질에 대한 나름대로의 감정(assessment)을 빠뜨리지 않고 있으며, 여러 학자들에 의해 "DB품질을 어떻게 평가할 것인가?"라는 문제가 심층적으로 다루어지고 있다[Harry & Oppenheim, 1993; Herther, 1988; Large, 1989]. 미국 남부 캘리포니아 온라인 사용자 그룹(SCOUG; Southern California Online User's Group)에서는 DB의 품질정도를 나타내는 "점수"를 개발하는 과정에 있으며(Granick, 1991),

유럽에서는 EQUIP(European Quality in Information Programme) 등의 프로젝트를 통해 품질에 관한 연구를 조직화, 전역화하고 있는 중이다[Lester, 1994].

그러나 현재까지의 많은 노력에도 불구하고 아직까지 DB 품질문제에 관한 많은 부분이 연구 대상으로 남아있는 상태이다. “무엇이 DB 품질인가?”에 관한 정의도 일반적으로 정립된 상태에 있지 않으며, DB 품질에 관한 여러 문제와 아이디어는 혼란스러울 만큼 다양한 각도와 관점에서 거론되고 있다. 본 연구는 DB 품질에 관한 여러 이슈와 관심을 분석하고 DB 품질문제에 대한 체계적 연구노력의 출발선을 찾고자하는 시도에서 시작되었으며, 그 구체적인 목적을 다음과 같이 세가지로 나누어 볼 수 있다.

① DB 품질에 관한 연구 프레임워크의 제안. DB 품질에 관한 여러 관심과 의문들을 전체적으로 바라보고, 체계적으로 이해할 수 있는 연구 프레임워크를 제안한다. 연구 프레임워크란 어떤 연구분야에서 한 그루 나무가 아니라 숲 전체를 바라볼 수 있는 조감도와 같은 역할을 수행한다. 연구 프레임워크를 통해 DB 품질에 관한 이슈가 무엇인지, 현재까지 어떠한 연구가 수행되어 왔으며, 앞으로 어떠한 연구가 필요한지를 이해할 수 있다.

② DB 품질기준의 제안. DB 품질이 무엇인지에 대한 정의를 시도하고, 그것을 평가할 수 있는 주요한 기준들을 제안한다. 이러한 기준을 통하여 DB 품질에 관한 이해를 도모하고, 제반 품질문제에 대한 체계적인 접근방법의 정립을 기대할 수 있다.

③ 품질 측정문항의 개발. 위에서 제시된 품

질기준에 의하여 실제로 DB 품질을 측정할 수 있는 문항(instruments)을 개발한다. 이러한 문항에 대하여 여러가지 비판과 다양한 의견이 개진될 것이며 그 과정에서 DB 품질문제에 관한 관심과 연구가 보다 활성화될 수 있을 것이다.

II. DB 품질에 관한 연구 프레임워크

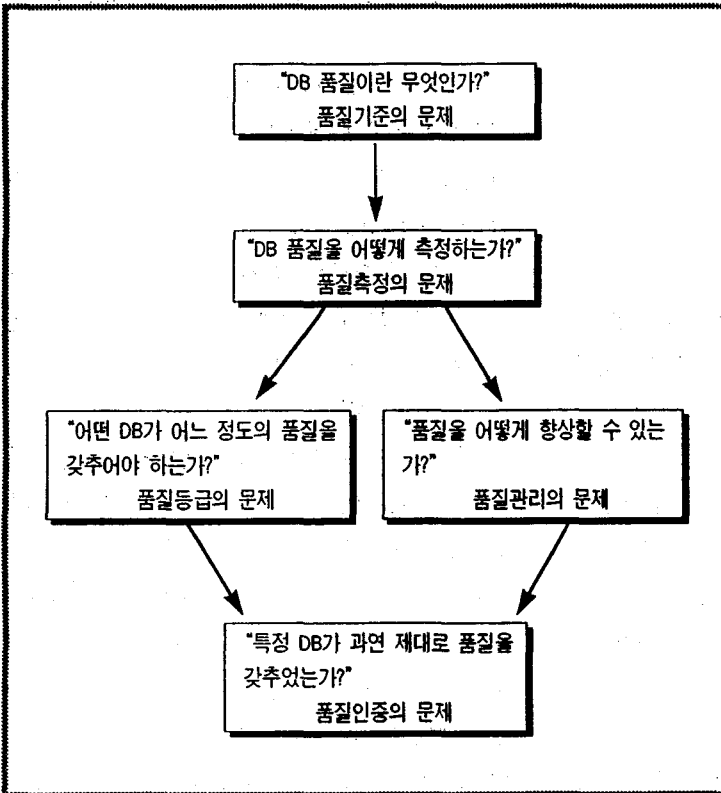
DB 품질은 그것을 인식하는 관점의 차이, 용어의 차이, 의미해석의 차이 등으로 인해 기본 개념조차 명료하게 정립되지 않은 상태에 놓여 있다. 예컨대 데이터가 무엇인지에 대해 다양한 정의들이 있으며, 어떤 정의를 취하느냐에 따라 그 품질을 인식하고 이해하는 시각이 상당히 달라진다[Fox et. al, 1992]. 많은 사람들이 DB 품질에 관한 여러가지 문제와 이슈들을 거론하고 있지만 동일한 문제를 서로 다른 용어로 표현하고 있거나, 서로 다른 문제를 동일한 용어로 부르고 있는 경우가 빈번하다. 예로서 품질을 이야기 할 때 가장 흔하게 등장하는 “reliability”의 의미는 보는 사람이나 상황에 따라 달라진다. 따라서 어떤 사람이 자신의 의사를 정확히 전달하고, 남이 의미하는 바를 정확히 이해하기 위해서 품질문제에 관한 용어나 관점을 표준화 또는 체계화한 틀(framework)이 필요하다.

연구 프레임워크에는 해당 분야의 중요한 의문과 관점들을 포함하는 포괄성, 각 관점들의 상호 유기적 연계관계를 나타내는 체계성, 그리고 전체적으로 한 눈에 이해될 수 있는 명료성이 갖추어져야 한다. 이러한 연구 프레임워크를 통하여 어떠한 주제가 연구대상인지를 파악하고, 여러 연구들이 어떻게 서로 연결되고 있는지를 분석하며, 앞으로의 연구방향이 어떻게 전개될 것인지를 이

해할 수 있다. 그리고 연구노력의 중복을 방지하고, 중요한 연구대상이 누락되는 것을 방지하는 효과도 기대할 수 있다.

본 연구에서는 지금까지 발표된 수십여편의 연구논문과 문헌자료를 분석하고, DB분야 연구자, 전문기술자, 실무자들과의 토론을 거쳐 DB품질에 관한 연구 프레임워크를 그림 1과 같이 도출하였다. 그림 1에서 보는 바와 같이 DB 품질문제는 크게 다섯가지로 구분되고 있으며, 각 문제들 사이의 관계가 화살표로 표현되고 있다. 화살표 방향은 "화살표가 나가는 문제"가 먼저 해결되어야 "화살표가 들어가는 문제"가 풀릴 수 있다는 의미를 가진다. 다섯가지 문제들의 중요성을 다음과 같이 요약할 수 있다.

(그림 1) 품질문제에 관한 연구 프레임워크



(1) 품질기준의 문제

"DB 품질이란 무엇인가? 어떠한 기준에 의해 그것을 정의할 수 있는가?"라는 문제는 DB 품질 문제를 이해하고 연구하기 위한 과정에서 첫번째로 부딪치는 문제이다. 그러나 DB라는 용어가 워낙 광범위한 의미로 사용되고 있고, 품질이라는 용어 또한 모호한 개념을 담고 있으므로 DB 품질이 무엇인지에 대한 정의가 쉽지 않은 실정이다. 지금까지 발표된 여러 연구논문과 보고서들도 각기 나름대로 DB 품질에 대한 정의를 내리고 있으며, 연구자의 입장에 따라 조금씩 강조하는 바를 달리 잡고 있다. 이러한 연구들의 공통점 중 하나는 DB 품질을 단일 기준으로 파악하지 않고 여러 기준들의 복합체로 인식하고 있다는 점이다. 이러한 기준들은 경우에 따라 품질영역, 차원, 요소, 요인, 속성 등으로 불리우고 있으나 본 연구에서는 기준이라는 용어를 사용하기로 한다.

(2) 품질측정의 문제

앞 단계에서 DB 품질에 대한 정의가 내려지고 나면 당장 그 다음에 떠오르는 것이 "DB 품질을 어떻게 측정할 것인가?"라는 문제이다. 측정될 수 없는 것은 결코 이해되지도 통제되지도 않는다는 말과 같이 측정될 수 없는 품질기준의 정의는 공허한 언어의 나열에 지나지 않는다. 따라서 "DB 품질이란 무엇인가?"와 "어떻게 측정할 것인가?"라는 두 문제

는 동전의 양면과 같은 불가분의 관계를 지니고 있다. 품질측정은 학계 뿐 아니라 산업계에서도 가장 관심있는 연구주제로 제기되고 있으며, 구체적으로 다음과 같은 문제들을 포함하게 된다.

① 왜 측정하는가? 측정의 목적과 동기는 여러가지가 있을 수 있으며 그 목적에 따라 다음에 기술되는 제반 문제에 대한 선택이 달라질 수 있다. 측정목적은 공식적인 것과 비공식적인 것으로 구분할 수 있다.

② 누가 측정할 것인가? 누가 측정하는지에 따라 측정의 주안점이 달라지고 아울러 그 결과도 달라질 수 있다. 개발자, 사용자, 투자자 중 누구의 입장과 관심을 강하게 반영할 것인지 또는 객관적인 제삼자를 어느 정도 개입시킬 것인지는 측정목적에 맞게끔 선택되어야 한다.

③ 어떻게 측정할 것인가? 데이터 관찰, 샘플링, 설문조사 등 다양한 방법이 있으나 현실적으로 측정작업에 주어지는 시간과 비용의 제약으로 인해 측정자에 대한 설문조사를 채택하는 것이 일반적이다.

④ 어떠한 도구가 사용되는가? 측정방법을 수행하기 위한 효율적인 측정도구가 필요하다. 설문조사의 경우 문항과 척도의 개발이 요구된다.

⑤ 어떠한 절차를 밟아야 하는가? 측정작업을 단시간 내에 정확하고 일관성있게 수행할 수 있는 절차와 가이드라인이 요구된다. 측정 결과의 공정성을 확보하기 위해서는 표준화된 절차가 미리 정립되어 있어야 한다.

⑥ 측정결과를 어떻게 나타내어야 하는가? 측정 결과에 “좋거나 나쁘다” 또는 “우수하거나 불량하다”라는 식의 가치판단을 부여하는 수 있고, 오류율 95% 또는 응답시간 5초 등과 같이 정량

적인 값만을 나타내는 수도 있다. 측정의 목적을 고려하여 선택되어야 한다.

(3) 품질등급의 문제

“어떠한 DB가 어느 정도의 품질을 갖추어야 하는가?”라는 문제는 다양한 DB의 개발보급과 함께 등장한 비교적 새로운 이슈이다. 물론 모든 DB가 100% 완벽한 품질상태를 갖추는 것이 가장 이상적이지만 방대한 량의 데이터를 한자의 오자도 없이 입력한다거나, 실세계에서 지속적으로 발생하는 데이터를 한치의 오차도 없이 제때 갱신한다거나, 수많은 사용자들을 모두 흡족케하는 완벽한 검색서비스를 제공하는 것은 현실적으로 불가능하다.

즉, 개발자가 최선의 노력을 다 한다해도 현실적으로 어느 정도의 품질이 떨어지는 현상을 피할 수 없다. 문제는 DB의 저품질 데이터로 인해 사용자가 선의의 피해를 입거나 개발자와의 분쟁이 발생할 가능성이 점점 높아지고 있다는 점이다. 물론 이러한 피해나 분쟁은 궁극적으로 공급자와 수요자 사이의 자유계약원리에 의거하여 처리되는 것이 원칙이지만 정보화시대에서는 일반대중에게 제공되는 DB의 공공성이 강조되고, 그러한 공공성에 걸맞는 최소한의 품질수준을 사회적으로 정립할 필요성이 부각된다.

사회적으로 요구되는 DB의 품질수준을 정립하기 위해서는 DB의 특성에 따라 품질수준을 차별적으로 등급화하는 것이 필요하다. 즉, 모든 DB가 동일한 품질수준을 확보하는 것이 아니라 DB 종류나 특성에 따라 각기 다른 수준을 갖추도록 하는 것이다. 예컨대 “신간도서정보 DB”와 “증권정보DB”는 사용자 층도 다르고, 의사결정의 중요성도 틀리며, 오류 데이터로 인한 피해의

심각도도 현저한 차이가 있다. 따라서 이 두 DB는 각기 다른 품질수준이 요구되는 것이 바람직하다. 유사한 예를 들자면 호텔에 부여된 무궁화 마크의 갯수(5개, 4개, 3개...)에 따라 각기 해당 등급에 맞는 시설과 서비스를 기대하는 것에 비유할 수 있다. 품질등급에 관한 구체적인 문제는 다음과 같이 나누어 질 수 있다.

- DB 품질을 몇 등급으로 차등화할 것인가?
- 어떤 특성을 가진 DB가 어떠한 품질등급에 속하는가?
- 각 등급에 속하는 DB는 최소한 어느 정도의 품질을 갖추어야 하는가?

여기서 주의할 것은 “실제 특정 DB가 과연 제대로의 품질을 갖추고 있느냐?” 하는 문제는 나중에 기술되는 품질인증에 관련되는 문제로서 여기에 직접적으로 해당되지 않는다는 점이다. 품질 등급에서는 특정 DB가 실제로 지니고 있는 품질 수준과는 상관없이 사회적으로 기대되는 또는 당연히 보유해야 할 품질수준을 정의하는 문제이다.

(4) 품질관리의 문제

“DB개발작업의 생산성을 어떻게 높일 것인가?”가 DB 프로젝트 관리자의 중전 관심사였다면 최근에는 “DB 품질을 어떻게 향상할 것인가?”라는 문제가 DB 성공여부를 결정하는 중요한 이슈로 등장하고 있다. DB는 제조상품이나 통상적 서비스와는 달리 사용에 의한 마모가 일어나지 않으며, 시간 또는 장소에 따라 동일한 데이터의 품질이 급격히 변화할 수 있고, 개별 사용자의 경험이나 판단에 의해 그 가치정도가 크게 변화하는 등의 특성을 가지고 있다. 즉, 품질 자체가 정형화, 반복화, 구조화, 획일화되는 것이 아니라 비정

형화, 비구조화, 다양화, 개별화되어 있다고 볼 수 있다. 따라서 좋은 품질을 창출하려는 품질관리노력은 다른 산업에서 찾아볼 수 없는 독특한 문제를 지니고 있다.

DB 품질관리를 위한 패러다임은 정형성(formalism)과 비정형성(informalism)으로 크게 구분될 수 있다. 정형성에서는 품질관리를 품질을 향상하려는 개발자의 제반 활동과 제도로 이해한다. 따라서 DB를 제작하고 서비스하는 전 과정에서 일어나는 작업을 기획하고, 조직하며, 통제하고, 그것을 위해 필요한 방법과 각종 도구의 제시를 요구한다.

특히 다음과 같은 단위 프로세스마다 무엇을 어떻게 해야 하는지의 규칙과 절차 그리고 적용되어야 할 품질개념과 기준의 설정을 필요로 한다.

- DB가 담아야 할 정보의 내용, 형식, 구조, 의미의 분석
- 사용자의 정보서비스 요구 분석
- DB의 논리적, 물리적 구조의 설계
- 검색 소프트웨어(또는 DBMS)의 기능 및 성능요구 분석
- 원시 데이터의 수집
- 원시 데이터의 가공(분석, 편집, 요약, 발췌, 재배열, 분류, 통합, 구조화)
- 가공된 데이터의 입력
- 테스트 및 시범운영
- 사용자 교육훈련
- DB 운영(정보서비스 제공, Help Service)
- 현행화 및 발전적 유지보수

그러나 DB를 만드는 과정 자체가 다른 공산품 제조 과정과는 달리 잘 정형화되어 있지 않고 DB마다 그 구조, 내용, 표현방식이 각기 다르기

때문에 어느 상황에서나 명료하게 적용될 수 있는 방법과 도구의 개발이 쉽지 않다.

한편 비정형성에서는 품질관리를 제도보다는 철학과 문화 그리고 의사소통의 문제로 인식한다. 그리고 표준화된 절차의 확립 및 준수, 서류화, 도구활용, 조직화에서 드러나는 경직성을 경계하고 상황에 따라 탄력적으로 대응할 수 있는 유연성과 창의성을 강조한다. 품질관리란 딱딱한 것이 아니라 품질을 인식하고 품질을 소중히 여기는 정신과 마음가짐의 문제이므로 철학과 신념의 변화에 의해 진정한 의미의 품질향상을 이룰 수 있다고 믿는다. 예컨대 품질향상을 "the way things can be done"이 아니라 "the way things should be done"으로 인식함을 요구한다.

그러므로 조직구조 및 업무절차의 변화를 뒷받침할 수 있는 기업문화의 변화가 요구되며, 계층구조보다 팀 위주, 개인간/팀간 의사소통 및 협력이 강조된다. 특히 고객개념의 도입으로 인해 모든 구성원은 어떤 서비스의 공급처이자 동시에 다른 서비스의 고객이 된다는 책임의식과 주인의식을 동시에 요구한다. 품질관리를 위하여 서클모임에 의한 열린 대화, 고객과의 서비스 계약 등이 주요한 수단으로 등장한다.

(5) 품질인증의 문제

"특정 DB가 과연 제대로 품질을 갖추고 있는가?"라는 문제를 취급한다. 이것은 객관적인 인증자에 의해 해당 DB의 품질을 "합격 또는 불합격", "Yes or No", 또는 "Good or Bad"로 평가하는 과정으로 이해된다. 이 문제를 다루기 위해서는 앞에서 기술한 바와 같이 DB품질이 무엇인지, 어떻게 측정하는지, 해당 DB가 어느 정도의 품질을 갖추고 있어야 하는지, 그리고 어떻게 품

질을 향상하는지에 대한 대답이 선행되어야 한다. 특히 본 연구에서 말하는 측정과 평가의 차이를 명백히 할 필요가 있다. 측정이란 실제 값이 무엇인지를 구하는 작업이고, 평가란 측정에 의해 구해진 값이 미리 정해진 어떤 기준값을 충족하는지 아닌지를 판단하는 일이다. 예컨대 전체 레코드 중 오류가 담긴 레코드 비율이 5%라는 것을 구하는 것이 측정이고, 이 5%라는 데이터 오류율을 불합격시켜야 하는지 아닌지를 판단하는 것이 평가이다.

물론 현실적으로는 이 두가지 작업이 하나로 이해되는 경우도 많지만 이론적으로 두 작업은 분리되어 있다. 품질인증은 주체에 따라 개발자에 의한 자체평가, 감리자 또는 감독자에 의한 평가, 발주자에 의한 검수, 또는 권위있는 전문/공공기구에 의한 인증 등으로 구분된다. 품질인증의 경우 품질을 실제로 평가하는 방법과 도구의 객관성, 결과의 공정성, 인증 주체의 권위 등을 어떻게 확보하고 제도화할 것이냐가 관건이 된다.

III. 데이터베이스 품질기준

1. 관련 개념

(1) 품질과 가치

DB 품질은 우리에게 익숙하지 않은 용어이다. 그 의미를 이해하기 위해서는 우선 흔히 거론되는 생산성, 가치 등의 용어들과 비교해서 살펴 볼 필요가 있다. 품질이란 상품이나 서비스의 우수성 또는 바람직한 정도를 나타내기 위해 사용되는 용어이다. 어떤 상품이나 서비스를 산출하는 과정 또는 방법의 우수성을 생산성 또는 능률성이라 부른다면, 만들어진 산출물의 우수성을 품질이라

부른다. 품질은 만드는 과정이 아니라 만들어진 산출물에 초점을 둔다는 점에서 효과성 또는 가치라는 용어와 혼용되기도 한다. 그러나 품질은 산출물의 객관적 성능과 기능의 바람직한 정도를 나타낼 뿐이며, 그 산출물이 사용자에게 어떻게 활용되어 어떠한 변화가 얼마만큼 발생하는지에 대한 관점을 배제하고 있다는 점에서 가치 개념과 구분된다.

DB는 개발이 완료된 시점에 객관적으로 관찰될 수 있는 특정 수준의 품질을 보유하고 있다고 볼 수 있다. 그러한 품질을 갖춘 DB가 사용자에게 의해 어떻게 인식되고, 업무처리활용이나 의사결정지원을 통해 어느 정도의 가치가 발생하는지는 사용자에게 따라 달라진다. 흔히 거론되는 비용-효과분석, 정보경제, 기업가치 등의 가치측정방법론은 사용자가 판단하는 가치의 종류와 규모에 대한 평가에 주안점을 두고 있으며, 따라서 사용자의 주관적 인식도에 따라 가치가 달라진다는 비판을 면할 수 없다. 또한 가치의 대체지표로서 간주되는 사용자만족도 조사에서 볼 수 있는 “무엇이 얼마나 유용합니까?”를 묻는 문항은 사용자의 주관적 인식도를 묻는 것이며 그 질문에 대한 대답은 사용자에게 따라 달라진다.

이와 달리 품질은 객관적으로 존재하고 있고, 그 상태를 직접적으로 측정할 수 있고, 그 결과가 누가 측정하는지에 따라 현저히 다르지 않다는 점에서 가치와 구분될 수 있다. DB 가치를 사용자에게나 기업에 미치는 영향이라 한다면 DB 품질이란 사용자에게 의해 사용되기 이전에 이미 존재하는 바람직한 정도이다. 물론 높은 품질의 DB가 높은 가치를 발생시킬 수 있을 것으로 추정되지만 아직까지 이러한 상관관계가 본격적으로 검증된 바는 없다.

(2) DB 품질의 이차원성

DB는 특정 분야 또는 조직의 관련된 데이터를 컴퓨터 저장매체에 통합적, 체계적으로 저장시켜 놓은 상태이다. 여러 곳에 산재하는 데이터를 통합함으로써 데이터 중복을 배제하고, 관리의 효율성을 제고하며, 여러 사용자에게 데이터를 제공할 수 있다는 장점을 지니게 된다. 이처럼 DB의 존재는 방대한 량의 데이터를 저장했다는 차원과 그렇게 저장된 데이터가 여러 사용자에게 편리하게 제공된다는 차원에서 찾을 수 있다. 즉, DB를 형성하는 것은 DB 데이터 구축과 DB 서비스 구현이며, 그러므로 DB품질은 다음과 같이 DB 데이터에 관한 품질과 DB 서비스에 관한 품질로 양분될 수 있다.

$$\text{DB품질} = \text{DB 데이터 품질} + \text{DB 서비스 품질}$$

2. DB 품질기준에 관한 연구동향

본 연구의 DB 품질기준을 설명하기 전에 우선 국내외에서 발표된 몇가지 주요한 연구결과를 간단히 살펴볼 필요가 있다. Brodie[1980]는 DB 품질 중에서 데이터 부분에만 주안점을 두고 있으며 데이터 신뢰성, 논리적 무결성, 물리적 무결성의 세가지로 구분하였다. 데이터 신뢰성이란 DB가 표현하고자 하는 실세계의 구조적 속성을 제대로 표현하는 정도로 정의된다.

그리고 논리적 무결성은 사용자에게 인식되는 개념적 데이터의 일관성을, 물리적 무결성은 컴퓨터에 물리적으로 저장된 상태로서의 품질을 의미한다. 이러한 구분은 DB가 만들어지고(물리적), 사용자에게 제공되며(논리적), 업무처리에 활용되는(신뢰성) 과정의 단계별 구분에 근거하

고 있다.

Doll and Torkzadeh[1988]는 최종사용자환경의 사용자만족도에 대한 연구에서 Factor analysis를 통해 다음과 같이 5가지 품질요소를 추출하고 있다: 내용상 정보요구의 충족성과 수록된 정보의 완전성, 정보의 정확성, 제공되는 정보의 표현형태의 유용성과 사용자의 이해용이성, 인터페이스의 친절성과 사용용이성, 신속성과 데이터의 현행성. 이 연구는 수백명 응답자의 데이터 분석을 통해 비교적 탄탄한 통계적 근거를 제시하고 있으나 DB 데이터가 지니는 품질과 검색 소프트웨어가 지니는 서비스 기능의 품질을 다소 혼동하고 있는 등의 논리적 취약함을 보이고 있다. 예컨대 검색속도의 신속함과 데이터의 현행화 정도가 어떻게 하나의 요인으로 취급되고 있는지는 의문이다.

Herther[1988]는 CD ROM 평가에 사용되는 항목을 열거한 checklist에서 다음과 같이 8개의 관점을 제공하고 있다: Accuracy and documentation, Appropriateness, Ease of use/comprehension, Authority and credibility, Content analysis, Comparisons to similar products, Revision and updates, Overall interest and worth. 이러한 관점들은 DB 품질에 관한 이론적 분석에 근거했다기 보다는 실무적으로 도서관 관계자들이 CD ROM의 여러가지 기능과 사양을 빠뜨리지 않고 살펴보기 위해 사용되는 일종의 점검목록적 성격을 강하게 지니고 있다. 따라서 용어의 정의도, 관점간의 구분 근거도 명료하지 않다.

Large[1989]는 DB를 평가하기 위하여 다음과 같은 4가지 기준을 제시하고 있다: 범위와 권위, 현행성, 정확성, 유일성. 범위와 권위는 DB가

관련 주제의 해당 데이터를 어느 정도 넓게 그리고 얼마나 심층적으로 커버하고 있는지를 의미한다. 현행성은 데이터 갱신 정도를, 정확성은 데이터의 오류 정도를 나타낸다. 그리고 네번째 기준인 유일성은 다른 유사한 DB와 얼마나 색다른 정보를 제공하는지를 나타내고 있다. DB품질 중에서도 서비스에 관한 부분을 배제하고 데이터 부분에 초점을 맞추고 있다.

Granick[1991]는 온라인DB 사용자그룹에서 채택된 DB 평가를 위한 10가지 카테고리를 다음과 같이 소개하고 있다: Consistency, Coverage/Scope, Timeliness, Accuracy/Error Rate, Assessibility/Ease of Use, Integration, Output, Documentation, Customer Support and Training, Value-to-Cost Ratio. 마지막 카테고리인 Value-to-Cost Ratio란 DB가 창출하는 가치와 DB가격(또는 사용료)과의 비율을 의미하는 개념이다. 그러나 앞서 기술한 바와 같이 가치와 품질은 서로 다른 의미를 지니고 있으며, 실제적으로 그 가치를 어떻게 평가할 것인가가 문제로 남는다. Granick 역시 이 부분에 관해서 "cannot be helpfully discussed here"라는 귀절로 언급을 피하고 있다.

AT&T Bell 연구소에서는 DB품질을 개선하기 위한 일련의 방법, 기법, 도구 등의 개발에 연구노력을 기울이고 있으며, 그 과정의 연구에서 다음과 같이 네가지 데이터 품질기준을 설정하고 있다[Fox et. al, 1992]:

- Accuracy, Precision, and Reliability
- Currentness, Age, and Timeliness
- Completeness and Duplication
- Consistency and Integration

이 연구는 DB 데이터 품질 부분에 초점을 두

고 있으며, 각 품질기준에 대한 개념, 용어, 논리 등을 비교적 탄탄한 이론적 배경에 의해 설명하고 있다.

이상에서 보는 바와 같이 여러 연구에서 DB품질을 파악할 수 있는 기준들이 제시되고 있다. 데이터의 정확성 같은 일부 기준은 대부분의 연구에서 등장하고 있다. 어떤 기준은 여러 연구에서 서로 다른 의미로 사용되기도 하고, 상이한 용어로 표현되는 기준이 실상 동일한 의미를 지니는 경우도 있다.

특히 데이터 자체의 품질 부분보다 CD ROM의 기능과 특성을 나타내는 서비스 품질 부분 기준들이 더욱 통일이 되어 있지 않음을 볼 수 있다.

3. 본 연구의 DB 품질기준

본 연구의 목적 중 하나는 DB품질을 측정할 수 있는 기준을 개발하는 것이다. 이를 위해 국내외 수십편의 관련 분야 연구논문 및 보고서를 분석하여 1차 리스트(17개 품질기준)를 추출하였고, 1차 리스트를 놓고 DB분야 전문가들이 수차례에 걸쳐 토론하여 2차 리스트(10개 기준)를 작성하였다. 2차 리스트는 다시 여러 실무자 및 산업종사자에게 우편으로 배달되었으며, 그 과정에서 수집된 수정 제안을 반영하여 몇몇 기준들의 통폐합이 이루어졌으며 그 결과 표1에서 나타난 바와 같은 최종 리스트(7개 기준)가 만들어졌다. 7개 품질기준 중 4개는 DB 데이터 품질에, 3개 기준은 DB 서비스 품질에 속하고 있음을 알 수

〈표 1〉 본 연구에서 제안하는 DB 품질기준

구분	품질기준(criteria)	핵심 관심
DB 데이터 품질	정확성 (accuracy)	DB 데이터가 실제 값과 정확히 일치하는가?
	완전성 (completeness)	실세계의 중요한 객체들과 속성이 담겨있는가?
	일관성 (consistency)	둘 이상의 데이터가 불일치하지 않는가?
	현행성 (currentness)	가장 최근의 데이터로 갱신되었는가?
DB 서비스 품질	검색성능 (search performance)	검색이 얼마나 신속하고 정교하게 이루어지는가?
	사용용이성 (ease of use)	DB접근과 정보의 활용이 얼마나 쉽고 편리한가?
	사용자지원성 (customer support)	문서화, 훈련, 도움말 등 사용자 지원이 적합한가?

있다.

(1) 정확성

데이터를 “실세계에 존재하는 사물, 사건, 개념 등의 객체(objects)에 관한 기록”으로 정의하자면, 데이터 정확성은 다음과 같이 세가지 시기에 의해 분석될 수 있다.

- 데이터 구조의 정확성
- 데이터 값의 정확성
- 데이터 표현의 정확성

① 데이터 구조의 정확성. 데이터 구조를 통해 실세계 객체들의 존재와 상태의 골격이 나타난다. DB 데이터 구조가 정확하지 않으면 실세계 정보가 왜곡된다. 예컨대 학교DB에서 선생과 학생을 하나의 파일로 묶어 버리거나, 서로 다른 속성을 보유하는 정규직원과 일용직원을 하나의 그룹으로 묶는 경우이다. DB 데이터 구조는 사용자 정보요구 분석과 데이터 모델링 단계에서 주로 선

택되며 이 과정에서 사용자의 정보요구를 얼마나 정확하게 충족시킬 수 있는지, 실세계를 얼마나 정확하게 반영하고 있는지의 판단은 객관적으로 검증되기가 어려우므로 실제 품질 측정대상에서 제외되는 것이 보통이다.

② 데이터 값의 정확성. 데이터 값의 정확성이란 실세계에서 객체들이 가지고 있는 속성 값(value-1)과 DB내 데이터 값(value-2)이 서로 일치하는 정도를 의미한다. 특정 데이터의 정확도는 value-1과 value-2가 얼마나 차이가 나느냐에 따라 계량화될 수 있다[Fox et. al, 1992]. 예컨대 홍길동의 실제 몸무게(value-1)가 70kg인데 DB 데이터 값(value-2)이 71kg이라면 1/70 정도의 비정확도를 가지게 되는 것이다. 따라서 실제 값 70kg을 79kg으로 입력한 오류는 69kg으로 입력한 오류보다 10배의 비정확도를 가지게 된다. 가격, 재고량, 땅 넓이 등과 같이 오류 데이터로 인한 피해의 심각도가 큰 숫자 데이터의 경우에 이러한 정확도 개념이 적용될 수 있다.

그러나 현실적으로 특정 데이터의 정확도는 의미적 차원보다 구문적 차원에서 판단되는 것이 보통이다. 즉, 70kg을 79kg으로 나타낸 것은 숫자 1개가 정확하지 못한 것이고, 69kg은 숫자 2개 모두가 틀린 것이다. 이러한 논리에 의한다면 “홍길동”이라는 문자로 된 value-1과 “홍일동”, “송길동”, “송일동” 등 value-2 사이의 비정확도를 비교적 간단하게 나타낼 수 있다. 전체 DB의 정확도는 DB에 담겨있는 전체 레코드(또는 화면) 중에서 정확한 레코드(또는 화면)의 비율로 나타낼 수 있다. 예컨대 1000개 레코드 중에서 100개 레코드가 정확하지 못한 값을 일부라도 담고 있다면 $900/1000 = 90\%$ 정도의 정확도로 표현된다.

정확성 기준의 가장 큰 난점은 실세계에 존재하는 실제 값 value-1 자체를 구하기가 어렵다는 점이다(물론 value-2는 DB에 확보되어 있다). 실제 값이 둘 이상 존재하여 어느 것을 선택하기가 애매모호한 경우(예: 취미)나 시간의 흐름에 따라 실제 값이 지속적으로 변해가는 경우(예: 주식시세, 신생아의 몸무게)도 문제를 어렵게 만든다. 특히 DB내 대규모 데이터의 실제 값을 정확히 파악하는 것 자체가 시간과 비용 문제로 인해 현실적으로 불가능한 경우도 있다. 이런 경우 value-1 값은 측정자에 의해 추정되는 것이 불가피하며, 따라서 정확성 측정 결과를 100% 믿을 수가 없다.

데이터 값의 정확성은 원시데이터의 수집, 가공, 입력, 그리고 현행화 단계를 거치면서 점점 떨어지는 것이 보통이다. 전통적으로 입력 오류는 저품질 데이터의 가장 주요한 원인으로 인식되어 왔으나 최근에는 자동판독기술의 등장과 디스크에 저장된 상태의 원시데이터가 많아짐에 따라 그 심각성이 점차 줄어들고 있다. 오히려 사생활보호나 지적재산권 강화 추세에 따라 “정보 보유자가 정확한 정보를 제공하는가?” 또는 “어떻게 왜곡 없이 데이터를 수집하는가?”라는 점이 DB 정확성에 큰 영향을 미치는 문제로 부상하고 있다.

③ 데이터 표현의 정확성. 데이터 표현이란 실세계 객체의 존재나 상태가 데이터로 표현되는 양식이나 형태를 의미하며, 데이터 표현의 정확성이란 실제로 존재하는 원시데이터와 DB에 담긴 데이터가 의미상, 내용상, 표현상으로 서로 정확하게 일치하는 정도를 의미한다. DB에 따라 정도의 차이는 있지만 수집된 원시데이터를 가공 또는 전환하는 작업이 필요하다.

예컨대 원시데이터의 필요한 부분만을 발췌하

거나, 내용을 요약하거나, 스크린에 적합하게끔 재배열하거나, 여러 데이터를 한 곳으로 통합하거나, 텍스트로 된 정보를 화상화하는 등의 가공 및 전환 작업이 이루어진다. 이렇게 가공된 데이터가 얼마나 정확한지를 나타내는 것이 데이터 표현의 정확성이며, 다음과 같은 다소 정성적인 질문을 통해 분석될 수 있다.

- 요약 또는 발췌된 데이터에 담긴 내용과 의미가 원문(원시데이터)과 동일한가?
- 분류, 재배열, 또는 재구성된 데이터의 내용과 의미가 원시데이터와 동일한가?
- 원시 데이터와 그것으로부터 추출된 키워드(keyword)가 정확하게 대응하는가?
- 기타 형태로 가공된 데이터가 원시 데이터의 내용, 의미와 일치하는가?

한편 정확성과 유사한 용어로서 많은 사람들이 신뢰성이란 용어를 사용하고 있다. 이 용어는 "데이터의 전체적인 정확성"을 의미하기도 하고 [Hoare, 1975; Chapple, 1976], "정확할 가능성 [probability of correctness]"으로 이해되기도 하며[Fox et. al, 1992], 때로는 "소프트웨어의 견고성"으로 사용되는 경우도 있다[Citroen, 1989]. 그러나 이 용어는 상황에 따라 달리 이해될 여지도 많고 직관적으로도 그 범위가 다소 명료하지 못하므로 본 연구에서는 사용하지 않는다.

(2) 완전성

DB 데이터의 완전성은 위에서 살펴 본 정확성과 마찬가지로 다음 세가지 시각에 의해 접근될 수 있다.

- 데이터 구조의 완전성
- 데이터 값의 완전성

● 데이터 표현의 완전성

① 데이터 구조의 완전성. 데이터 구조를 통해 실세계의 정보구조가 나타나는 바 구조가 완전하지 못하면 중요한 객체가 나타나지 않거나, 객체에 관한 중요한 속성이 누락되는 현상이 발생한다. 예컨대 도서정보DB에서 한 사람의 저자만을 수록할 수 있게끔 설계가 되어 있어서 공동저자가 빠지거나, 기업정보DB에서 팩스번호 필드가 없다면 이는 실세계를 완전하게 반영하지 못하는 DB라 볼 수 있다. 데이터 구조의 완전성을 측정하기 위해서는 "무엇이 중요한 객체이며, 그 객체를 나타내기 위한 필요한 속성이 무엇인가?"라는 문제가 대답되어야 하나 무엇이 옳고 그른지를 판단하기 곤란한 경우가 많으므로 현실적으로 품질측정대상에서 제외되는 경우가 대부분이다.

② 데이터 값의 완전성. 데이터 값의 완전성은 실세계에 존재하는 객체들의 존재나 상태가 데이터에 완전하게 담겨있는 정도를 의미한다. 특정 데이터의 완전성은 누락된 레코드가 있는지 그리고 레코드의 필드 값이 비어있는지에 의해 측정될 수 있다. 즉, 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 측정 결과를 100% 믿을 수는 없다. 데이터 값의 완전성 문제는 원시데이터 수집 또는 입력단계에서 오류가 발생한 결과로 볼 수 있다.

한편 필드 값이 비어있었다함은 null 값이 있다는 말이다. 언뜻 생각하기에는 null 값이 들어있는 레코드를 불완전한 레코드로 판단할 수 있다. 그러나 문제가 이처럼 간단치 않은 이유는 바로

이 null 값이 상황에 따라 서로 다른 의미를 지니고 있다는 점이다. 표 2는 null 값의 여러가지 의미를 보여주고 있다. 이러한 null 값의 여러가지 의미는 DB 데이터의 완전성 측정에 어려움을 더해 주고 있다.

③ 데이터 표현의 완전성. 데이터 표현의 완전성이란 가공된 DB 데이터가 원시데이터에 담긴 정보를 완전하게 담고 있는지를 의미하며, 실제로 다음과 같은 질문을 통해 분석될 수 있다.

- 분류, 재배열, 또는 재구성된 데이터의 내용이 원시데이터의 일부를 누락하거나 그 의미를 왜곡하고 있지 않는가?
- 추출된 키워드는 해당 주제 분야의 주요 개념을 완전히 커버하는가?
- 기타 방법으로 가공된 데이터가 원시 데이터의 일부를 누락하고 있지 않는가?

(3) 일관성

DB 데이터의 일관성이란 정확성이나 완전성과는 달리 그 의미가 간결하게 정의되지 않는 개념이다. 좁은 의미에서 보자면 DB내 둘 이상의 데이터 값이 서로 상충되지 않고 일관된 상태를 이루는 정도로 정의되고[Date, 1983], 보다 넓은 의미에서 보자면 데이터가 지켜야 할 구조상, 값 표현상의 모든 제약조건을 위배하지 않고 충족시키는 정도로 정의된다[Elmasri & Navathe, 1989]. 본 연구에서는 품질의 실제적인 측정가능성을 고려하여 다음과 같이 일관성을 정의한다.

① 데이터 구조의 일관성. 데이터 구조의 일관성이란 객체의 존재나 상태를 나타내는 속성이 전

〈표 2〉 Null 값의 여러가지 의미

null 값의 의미	사 례	비 고
속성이 적용되지 않는다 (attribute not applicable)	홍길동은 삐삐를 가지고 있지 않다	완전성에 영향이 없다
속성 값을 모른다 (value unknown)	홍길동이 가지고 있는 삐삐의 번호를 모른다	완전성에 영향을 미친다
속성이 적용되는지를 모른다. (unknown applicability)	홍길동이 삐삐를 가지고 있는지 아닌지를 알 수 없다	완전성에 영향을 미치지 않는다 아닌지를 알 수 없다

체 DB를 통해 중복과 혼동없이 나타나고 있는 정도를 의미한다. 예컨대 생년월일, 나이 등과 같이 본질적으로 동일한 정보를 제공하는 필드가 둘 이상 중복한다든지 또는 어느 대학원생의 레코드가 학생파일과 조교파일에 중복적으로 존재하게끔 구조가 설계되는 것을 일관성 부족의 경우로 간주할 수 있다.

② 데이터 값의 일관성. 데이터 값의 일관성이란 두 데이터 값이 서로 상충없이 존재하는 정도를 의미한다. 예컨대 이중 국적을 가진 사람의 국적이 외무부에서는 미국으로, 내무부에서는 한국으로 나타난다든지(둘 다 정확하지 않는 것은 아니다), 어떤 사람이 가진 두 개의 여권번호가 서로 다른 파일에 존재한다든지, 남학생 파일에 여학생 레코드가 끼어 있다든지, 전화번호 필드에 팩스번호가 있는 경우를 일관성 부족의 사례로 들 수 있다.

③ 데이터 표현의 일관성. 데이터 표현의 일관성이란 동일한 종류의 데이터를 표현하는 방식의 일관성을 의미한다. 예컨대 두 권의 세계명작에 관한 데이터의 요약 정도가 특별한 이유없이 현저히 다르다든지(요약하는 이의 스타일에 따라), 한 클래스는 학번 순으로 정렬하고 다른 클래스는 이름의 가나다 순으로 정렬하는(특별한 이유없이) 경우를 일관성 부족의 경우로 들 수 있다. 때로는

데이터 표현의 일관성을 DB 데이터 뿐만 아니라 스크린 설계, 명령어, 메뉴, 색인 등의 체계성과 통일성을 포함하여 사용하는 경우도 있으나 [Granick, 1991] 본 연구에서는 이러한 부분을 DB 서비스 품질로 넘기고, 일관성은 DB 데이터에만 국한시킨다.

(4) 현행성

현행성이란 DB 데이터가 최신의 데이터로 갱신되고 있는 정도를 나타낸다. 특정 시점 t_1 에서는 정확했지만 시간이 흘러 시점 t_2 에서는 더 이상 정확하지 않게 된 데이터를 낙후된 데이터라 부른다. 그러므로 현행성은 앞에서 살펴본 정확성 기준에 포함될 수도 있다. 즉, 시간의 흐름에 따라 정확했던 데이터가 부정확한 데이터로 변질된다는 의미에서 낙후된 데이터를 정확하지 않는 데이터로 간주하는 입장도 있을 수 있다. 그러나 일반적으로 낙후된 데이터는 부정확한 데이터와 별도로 취급되고 있고, data warehouse 개념에서 보는 바와 같이 이력(historic) 데이터는 그 나름대로의 의미와 가치를 가지고 있으며, DB 관리에서 데이터 현행화는 특별한 중요성을 가지고 있으므로 본 연구에서는 현행성을 별도의 품질기준으로 취급한다.

DB 데이터의 현행성은 전체 데이터 중 얼마나 많은 데이터가 낙후되었는지를 통해 측정될 수 있다. 예컨대 전체 데이터의 40%가 최근 데이터로 갱신되었고 60%가 낙후되었다고 말하는 식이다.

이와는 달리 시간적으로 얼마나 최근에 갱신되었는지에 의해 표현되기도 한다. 예컨대 새로운 데이터가 발생하는 주기를 1 사이클 타임이라고 하면(예: 일간지는 1일, 월간지는 1개월) 현재 DB는 2 사이클 타임이 낙후되었다고 말하는

식이다.

DB의 종류에 따라 낙후된 사이클 타임이 별로 의미를 가지지 못하는 경우도 있고(예: 팔만 대장경DB, 조선왕조실록DB), 다른 어떤 품질기준 못지 않게 중요하게 취급되는 경우도 있다(예: 주식시세현황DB, 신착도서DB).

(5) 검색성능

검색성능은 앞에서 언급한 데이터 품질기준과는 달리 서비스 품질기준에 속한다. DB 서비스 품질이란 데이터가 사용자에게 서비스되는 과정에서의 매체, 수단, 방법, 그리고 공급자 영업정책 등의 바람직한 정도를 의미한다. 즉, 데이터 자체가 아니라 서비스의 신속함, 편리함, 친절함, 다양함을 망라하는 개념이다.

DB 서비스의 가장 중요한 부분을 차지하는 것이 검색 소프트웨어(또는 DBMS)이다. 소프트웨어의 메뉴 체계, 명령어 체계, 색인 종류, 검색 속도, 스크린 설계, 도움말 기능, 예러처리 기능 등이 사용자의 요구를 얼마나 충족시키느냐에 따라 서비스 품질이 주로 결정된다. 물론 DB 서비스 품질은 소프트웨어적 성능과 기능 외에도 제공매체(온라인 통신망 또는 CD ROM), 사용자 매뉴얼 등의 문서, DB업체의 사용자 훈련, 도움말 서비스, 각종 지원 메카니즘 및 정책에 의해 영향을 받는다.

검색성능이란 사용자가 DB에 접근하여 데이터를 검색하는 과정에서 제공되는 방법, 도구, 매체들의 적합 정도를 의미한다. 주로 검색 소프트웨어(또는 DBMS)와 온라인 통신망의 기능과 성능에 의해 결정되며, 일반적으로 다음과 같은 시각에 의해 분석될 수 있다.

- 데이터 검색 속도 (예: 평균 및 최악 상황의 반응속도)

- 메뉴에 의한 검색(menu-driven search)의 다양함과 편리함
- 검색 명령어의 표준화와 다양함
- 필드 검색의 제약(색인의 다양함)
(예: 모든 필드마다 검색이 가능한가? 둘 이상 필드의 조합에 의한 검색이 가능한가?)
- 색인(index)의 정교함 (예: 이름 끝자가 "자"인 여성을 검색할 수 있는가?)
- 논리 연산자(AND, OR, NOT)의 존재여부
- 키워드 및 시소러스 체계의 완전함과 정확함

(6) 사용용이성

사용용이성이란 사용자가 DB에 접근하고 정보를 검색하며, 산출물을 활용하는 과정에서 제공받는 소프트웨어 인터페이스 및 유틸리티 기능의 편리함과 적합한 정도를 의미한다. 이 개념은 정보기술의 성공여부를 결정적 영향을 미치는 변수로서 수많은 연구자들에 의해 연구되어 왔다. 본 연구에서는 사용용이성에 영향을 주는 요소들을 다음과 같이 다섯가지로 구분한다.

① 명령어의 사용용이성

- 일반화 또는 표준화된 명령어가 있다면 그것을 사용하는가?
- 명령어가 다양한가? (예: 화면전후이동, 항목전후이동, 메뉴상하이동, 화면재생, 연속 출력, 입력취소 등)
- Prompt가 일관성있게 제공되며 입력위치 식별이 용이한가?

② 메뉴의 사용용이성 (예: 메뉴 구조가 간단명료하고 이해하기 쉬운가?)

③ 화면 및 산출물의 사용용이성

- 화면에 나타나는 데이터를 쉽게 이해할 수 있는가?

- 화면당 출력량, 배열, 색상, 하이라이트 사용이 적절한가?
- 여러 화면들 사이에 일관성이 있는가?
- 화면상에 나타나는 타이틀과 안내 메시지가 유용한가?
- 산출물을 사용자가 원하는 매체와 형태로 전환하기 용이한가?
(예: 텍스트 형태로 특정 순서에 의해 프린터로 출력할 수 있는가?)

④ 용어의 사용용이성

- DB에서 나타나는 용어를 사용자가 이해하기 쉬운가?
- 용어를 일관성있게 사용하는가?

⑤ 도움말 및 에러 처리 기능의 사용용이성

- Help와 Exit 기능이 언제나 이용가능한가?
- 도움말 메시지가 간단명료하고 친절한가?
- 도움말 메시지가 사용자가 처한 상황에 맞는가?
- 에러 메시지가 사용자가 처한 상황과 대책을 간단명료하고 친절하게 설명하는가?

(7) 사용자 지원성

사용자지원성이란 DB개발자가 사용자에게 제공하는 교육훈련, 각종 문서, 도움 서비스 등 기술적/정책적/제도적 지원의 적합정도를 의미한다. 본 연구에서는 다음과 같은 요소로 구분한다.

① 제반 문서의 충실도, 간단명료성, 목차와 색인 구비여부

- 사용자지침서
- 운영지침서

② 도움 서비스의 충실도 및 친절도

- 온라인 튜토리얼

- 전화 도움라인
- ③ 사용자 훈련의 빈도 충실도
- ④ 마케팅 정책 및 제도
 - DB입체 과실로 인한 피해의 보상 및 배상제도
 - 사용자 구분에 따른 할인 또는 무료 정책

IV. 품질 측정문항의 개발

1. 품질 측정문항의 개발에서 고려된 사항

DB 품질에 관한 연구 프레임워크에서 기술된 바와 같이 품질측정은 측정의 목적, 주제, 방법, 도구, 절차상의 여러가지 문제를 안고 있다. 이러한 문제들 중 현실적으로 가장 많은 관심이 모아지고 논란의 대상이 될 수 있는 것이 바로 측정도구, 즉, 측정문항(instruments)이며, 본 연구의 목적은 DB의 품질측정에 채택될 수 있는 실제적인 문항을 개발하는 데 있다. 이 과정에서 고려된 사항은 다음과 같다.

(1) 측정비용과 시간의 최소화

어떠한 목적으로 측정이 필요하든지 간에 측정작업 자체에 많은 시간과 비용이 주어지기는 어렵다. 품질측정의 기본의의를 훼손하지 않는 범위 내에서 최소한의 시간과 비용이 요구될 수 있도록 측정작업이 고안될 필요가 있다. 특히 설문조사의 경우 보통 응답자가 모든 문항을 읽고 답할 수 있는 시간이 4시간을 넘지 않도록 하는 것이 중요하다. 사람의 집중도와 인내력이 4시간 이상 꾸준히 유지되리라 기대할 수 없기 때문이다.

(2) 측정결과의 계량화

DB 품질측정의 어려움은 일부 품질기준이 정

량적이 아니라 정성적이라는 점에서 우선 찾을 수 있다. 예컨대 사용용이성 또는 사용자지원성과 같은 품질기준은 수와 양에 의해 측정되는 것이 아니라 질과 깊이에 의해 결정되므로 계량적 방법에 의한 측정이 불가능하다. 더우기 정량적인 품질기준이라도 그 완벽한 측정을 위한 경비와 시간이 비현실적인 경우가 대부분이다. 예컨대 일관성을 철저히 파악하기 위해서는 모든 레코드의 크로스체크가 이루어져야 하는데 그 작업 자체에 드는 시간과 노력이 현실적으로 허용될 수 없다. 그러나 품질측정의 일반적인 목적을 달성하기 위해서는 가능한 한 측정결과를 표준화된 의미의 수치 또는 형태로 나타낼 수 있어야 한다. 흔히 DB품질의 측정결과를 "좋다, 보통이다, 나쁘다"라는 식으로 나타내지만 그 결과가 어떤 의미를 가지고 있는지 정확하지 않다. 예컨대 "좋다"라는 것이 절대적 의미에서 좋은 것인지 아니면 다른 DB와 비교하여 상대적 의미에서 좋은 것인지 알 수 없다.

본 연구에서는 DB품질의 측정결과를 0부터 100까지의 수치로 나타낼 수 있도록 문항의 척도를 조정하고 있다. 즉, 각 문항의 척도는 다섯단계로 구성되어 있으며, 각 단계마다 0, 25, 50, 75, 100이라는 점수가 부여되어 있다. 예컨대 다음은 데이터의 철자오류를 묻는 문항이다.

철자오류나 잘못된 데이터 값이 발견되는 레코드가 100 레코드당 평균 몇 개 정도 있습니까?

7개 이상	5 내지 6개	3 내지 4개	1 내지 2개	없음
0점	25점	50점	75점	100점

위에서 보는 것처럼 오류 레코드가 전혀 없으면 100점, 3 내지 4개가 있으면 50점, 7개 이상

있으면 0점이 된다. 다섯 단계의 척도는 현재 DB 기술과 서비스 현황을 고려하여 다음과 같은 의미에 의해 작성되었다.

100점	최고 수준, 완벽한 수준
75점	(100과 50의 중간)
50점	현실적으로 최소한 달성되어야 할 수준
25점	(50과 0의 중간)
0점	도저히 허용될 수 없는 불량 수준

물론 DB 종류에 따라 측정결과가 달리 해석되어야 할 경우도 있고, DB기술의 발전에 따라 앞으로 조금씩 갱신될 필요가 있다. 이러한 계량화를 통해 다음과 같은 효과도 기대할 수 있다.

- 특정 DB의 전년도 품질 75와 금년도 품질 80을 비교함으로써 동일 DB의 시간의 흐름에 따른 개선여부 및 정도 파악이 가능하다.
- “갑” DB의 품질 75와 “을” DB의 품질 80을 비교함으로써 서로 다른 DB들의 상대적 품질 비교가 가능하다.
- 모든 공공DB는 품질지수 50이상을 획득해야 한다는 등의 명료한 목표설정이 가능하다.

2. 본 연구의 품질 측정문항

본 연구에서는 앞에서 기술된 7개 품질기준별로 각기 문항 Pool을 개발하였다. 예컨대 정확성 Pool에는 8개의 문항, 완전성 Pool에는 10개의 문항이 Pool을 이루고 있다. 각 Pool내의 모든 문항을 반드시 사용할 필요는 없으며 측정하고자 하는 특정 DB의 특성과 평가 여건에 따라 적합한 문항을 선택하여 사용할 수 있다. 각 품질기준별로 개발된 문항은 다음과 같다.

(1) 정확성 측정문항

1-1: 철자오류나 잘못된 데이터 값이 발견되는 레코드가 100레코드당 평균 몇 개 정도 있습니까?

7개 이상	5 내지 6개	3 내지 4개	1 내지 2개	없음
0점	25점	50점	75점	100점

1-2: 한 화면에 나타난 데이터에서 철자가 틀린 부분이 평균 몇 군데 정도 발견됩니까? (5 화면 이상을 무작위로 검토한 후 대답해 주십시오)

3군데 이상	2군데 정도	1군데 정도		없음
0점	25점	50점	75점	100점

1-3: 요약 또는 발췌된 데이터가 있다면 그런 데이터의 문장구조와 표현방식이 어느 정도 정확합니까?

매우 어색함		비교적 정확함		아주 정확함
0점	25점	50점	75점	100점

1-4: 요약 또는 발췌된 데이터가 있다면 그런 데이터에 담긴 내용과 의미가 원시 데이터와 어느 정도 일치합니까?

전혀 다름		비교적 일치함		아주 일치함
0점	25점	50점	75점	100점

1-5: 재분류, 재배열, 또는 재구성된 데이터가 있다면 그런 데이터에 담긴 내용과 의미가 원시 데이터와 어느 정도 일치합니까?

전혀 다름		비교적 일치함		아주 일치함
0점	25점	50점	75점	100점

1-6: 키워드(keyword)와 해당 데이터가 정확하게 일치합니까?(예: 키워드를 통해서 찾고자 하는 데이터가 정확하게 찾아집니까?)

전혀 안 맞음		비교적 맞음		아주 잘 맞음
0점	25점	50점	75점	100점

1-7: 부정확한 데이터가 나타났다면 그러한 부정확한 데이터로 인해 어느 정도 실망했습니까? (또는 불만의 정도)

아주 불만임		어느정도 불만		개여치 않음
0점	25점	50점	75점	100점

1-8: 부정확한 데이터를 예방하거나, 발견하거나, 또는 교정하려는 노력이 어느 정도라고 생각합니까? (예: DB업체의 노력이나 소프트웨어적 기능)

전혀 없음		보통 정도임		아주 열심히
0점	25점	50점	75점	100점

(2) 완전성 측정문항

2-1: 어떤 필드가 비어있는 레코드가 10 레코드 당 평균 몇 개 정도 있습니까?
(실제로 필드 값이 존재하는데도 수집을 하

지 못한 경우만 해당됩니다)

7개 이상	5개지 6개	3개지 4개	1개지 2개	없음
0점	25점	50점	75점	100점

2-2: 한 화면에 나타난 데이터에서 필드가 비어 있는 곳이 평균 몇 곳 정도 발견됩니까?
(5 화면 이상을 검토한 후 대답해 주십시오)

3곳 이상	2곳 정도	1곳 정도		없음
0점	25점	50점	75점	100점

2-3: 전체 레코드에서 누락된 레코드의 비율이 어느 정도입니까?
(특히 인명, 지명, 기업명, 사물명 등의 리스트를 제공하는 DB의 경우에 해당됩니다)

4퍼센트 이상	3퍼센트	2퍼센트	1퍼센트	없음
0점	25점	50점	75점	100점

2-4: 정기간행물 DB의 경우 누락된 권/호는 평균 몇% 정도입니까?

15퍼센트 이상	10퍼센트	5퍼센트	3퍼센트	없다
0점	25점	50점	75점	100점

2-5: DB 데이터는 주제분야의 폭과 깊이를 어느 정도 완전하게 커버하고 있습니까?

아주 부실함		어느 정도		거의 완전함
0점	25점	50점	75점	100점

2-6: 실세계에서 중요하다고 인식되는 속성이 DB 내의 필드로서 완전하게 담겨 있습니까?
 (예: 누락된 속성이 많습니까? 중요한 속성이 누락되어 있습니까?)

누락이 큼		어느 정도		누락 없음
0점	25점	50점	75점	100점

2-7: 재분류, 재배열, 또는 재구성된 데이터가 있다면 그런 데이터에 담긴 내용과 의미가 원시 데이터를 어느 정도 완전하게 표현하고 있습니까?
 (예: 중요 내용이 누락 또는 상실되지 않았습니까?)

매우 불완전		비교적 완전함		아주 완전함
0점	25점	50점	75점	100점

2-8: 키워드를 통해서 찾고자 하는 데이터를 완전하게 찾을 수 있습니까?
 (예: 찾아지지 않는 데이터가 있습니까?)

매우 불완전		비교적 완전함		아주 완전함
0점	25점	50점	75점	100점

2-9: 데이터가 누락되고 완전치 못하다면 그러한 데이터로 인해 어느 정도 실망했습니까?
 (또는 불만의 정도)

아주 불만임		어느정도 불만		개의치 않음
0점	25점	50점	75점	100점

2-10: 불완전한 데이터를 예방하거나, 발견하거나, 또는 교정하려는 노력이 어느 정도라고 생각합니까?
 (예: DB업체의 노력이나 소프트웨어적 기능)

전혀 없음		보통 정도임		아주 열심히
0점	25점	50점	75점	100점

(3) 일관성 측정문항

3-1: 동일한 사실을 표현하는 값이 서로 틀리는 경우가 어느 정도 있습니까?
 (예: 동일인의 주소가 여러 곳에서 서로 다르게 기록된 경우)

가끔 있음		거의 없음		전혀 없음
0점	25점	50점	75점	100점

3-2: 한 파일 내에 불필요하게 중복되는 레코드가 어느 정도 있습니까?
 (예: 동일 레코드가 두번 이상 중복적으로 나타나는 경우)

가끔 있음		거의 없음		전혀 없음
0점	25점	50점	75점	100점

3-3: 특정 파일에 해당되지 않는 레코드가 엉뚱하게 끼어 있는 경우가 어느 정도 있습니까?
 (예: 신입생 DB에 졸업생 레코드가 들어 있는 경우)

가끔 있음		거의 없음		전혀 없음
0점	25점	50점	75점	100점

3-4: 한 속성에 해당되지 않는 값이 들어 있는 경우가 있습니까?

(예: 주소 필드에 전화번호가 있는 경우)

가끔 있음		거의 없음		전혀 없음
0점	25점	50점	75점	100점

3-5: 요약 또는 발췌된 데이터가 있다면 그러한 데이터가 요약 또는 발췌된 정도와 방식이 어느 정도 일관성을 지니고 있습니까?

각기 다름		비교적 일관적임		아주 일관적임
0점	25점	50점	75점	100점

3-6: 재분류, 재배열, 또는 재구성된 데이터가 있다면 그러한 데이터가 가공된 정도와 방식이 어느 정도 일관성을 지니고 있습니까?

각기 다름		비교적 일관적임		아주 일관적임
0점	25점	50점	75점	100점

3-7: 데이터가 불일치하고 일관성이 없다면 그러한 데이터로 인해 어느 정도 실망했습니까? (또는 불만의 정도)

아주 불만임		어느정도 불만		개의치 않음
0점	25점	50점	75점	100점

3-8: 일관성없는 데이터를 예방하거나, 발견하거나, 또는 교정하려는 노력이 어느 정도라고 생각합니까?

전혀 없음		보통 정도임		아주 열심히
0점	25점	50점	75점	100점

(4) 현행성 측정문항

4-1: 필드 값이 낙후되어(out-of-date) 있는 레코드가 100 레코드당 평균 몇 개 정도 있습니까?

16개 이상	15개 정도	10개 정도	5개 정도	없음
0점	25점	50점	75점	100점

4-2: 새로운 데이터가 발생하는 주기를 1 사이클이라 정의합니다(예: 주간지는 7일, 월간지는 30일). 새로운 데이터가 발생하는 시점과 그것이 DB에 갱신되는 시점은 평균 어느 정도의 시간 차이가 납니까?

5사이클 이상	4사이클 정도	3사이클 정도	2사이클 정도	1사이클 이내
0점	25점	50점	75점	100점

4-3: 새로운 데이터의 발생주기와 규모를 고려할 때 DB가 얼마나 자주 갱신 되는 편입니까?

전혀 갱신안됨		비교적 자주		아주 자주
0점	25점	50점	75점	100점

4-4: 데이터가 갱신되지 못하고 낙후되어 있다면 그러한 데이터로 인해 어느 정도 실망했습니까? (또는 불만의 정도)

아주 불만임		어느정도 불만		개의치 않음
0점	25점	50점	75점	100점

4-5: 오래 되어서 불필요한 데이터를 발견하고, 제거하거나 교정하려는 DB업자의 노력이

어느 정도입니까?

전혀 없음		보통 정도임		아주 열심임
0점	25점	50점	75점	100점

4-6: 정보의 발생, 수집, 그리고 갱신 시기가 명료하게 표현되고 있습니까?

전혀 아님		어느 정도		아주 상세함
0점	25점	50점	75점	100점

(5) 검색성능 측정문항

5-1: 메뉴를 이동하거나 기본 명령문의 반응이 나타나기까지 걸리는 시간은 평균 어느 정도입니까?

(주의: 각기 다른 5 케이스를 검토한 후 대 답해 주십시오)

7초 이상	7초 이내	5초 이내	3초 이내	1초 이내
0점	25점	50점	75점	100점

5-2: 데이터 검색에서 “비교적 간단한” 질의어를 입력한 후 그 결과가 나타나기까지 걸리는 시간은 평균 어느 정도입니까?

(주의: 각기 다른 5 케이스를 검토한 후 대 답해 주십시오)

8초 이상	8초 이내	6초 이내	4초 이내	2초 이내
0점	25점	50점	75점	100점

5-3: 데이터 검색에서 “아주 복잡한” 질의어를 입력한 후 그 결과가 나타나기까지 걸리는 시간은 평균 어느 정도입니까?

(주의: 각기 다른 5 케이스를 검토한 후 대 답해 주십시오)

20초 이상	20초 이내	10초 이내	5초 이내	3초 이내
0점	25점	50점	75점	100점

5-4: 메뉴를 통한 데이터 검색에 관해서입니다. 메뉴 구조의 체계성이나 메뉴에서 제공되는 검색경로의 다양성이 어느 정도 우수하고 정교합니까?

아주 빈약함		보통정도		매우 우수함
0점	25점	50점	75점	100점

5-5: 데이터 검색 명령어에 관해서입니다. 제공 되는 명령어가 여러 사용자의 필요성과 비 교해 볼 때 어느 정도 다양합니까?

아주 빈약함		보통정도		매우 다양함
0점	25점	50점	75점	100점

5-6: 단일 필드를 통한 검색에 관해서입니다. 검색이 필요한 필드마다 데이터 검색이 신속히 이루어질 수 있습니까? 기술적으로 말하자면, 중요한 필드마다 색인이 충분히 갖 추어져 있습니까?

아주 불충분함		비교적 충분		아주 충분함
0점	25점	50점	75점	100점

5-7: 단일 필드를 통한 검색에 관해서입니다. 중요한 필드의 데이터 검색이 얼마나 유연하고 탄력적으로 이루어질 수 있습니까? 즉,

중요한 필드에서 다양한 검색이 이루어질 수 있습니까?

(예: 이름 석자 중 끝 자가 "자"인 레코드를 구할 수 있는지의 여부)

전혀 다양하지 않음		비교적 다양함		아주 다양함
0점	25점	50점	75점	100점

5-8: 둘 이상 필드의 조합에 의한 데이터 검색에 관해서입니다. 중요한 몇가지 필드를 조합한 데이터 검색의 필요성에 비해 실제 제공되는 기능이 어느 정도 충분합니까?

아주 불충분함		비교적 충분		아주 충분함
0점	25점	50점	75점	100점

5-9: 논리곱(AND), 논리합(OR), 부정(NOT), 비교연산자(=, <, >, >=, <=) 등에 관해서입니다. DB 서비스에 의해 제공되는 이러한 연산자 기능이 어느 정도 충분합니까?

아주 불충분함		비교적 충분		아주 충분함
0점	25점	50점	75점	100점

5-10: 데이터 검색에 사용되는 키워드 및 시소러스(thesaurus)가 그 필요성에 비해 어느 정도 잘 준비되어 있습니까?

전혀 불충분함		비교적 충분		거의 완벽함
0점	25점	50점	75점	100점

(6) 사용용이성 측정문항

6-1: 메뉴에 관해서입니다. 메뉴 구조가 어느 정도 간단명료하고 이해하기 쉽습니까?

매우 어려움		비교적 명료함		아주 명료함
0점	25점	50점	75점	100점

6-2: 명령어에 관해서입니다. DB에 의해 제공되는 명령어가 어느 정도 간단명료하고 이해하기 쉽습니까?

매우 어려움		비교적 명료함		아주 명료함
0점	25점	50점	75점	100점

6-3: Prompt와 입력위치가 어느 정도 식별하기 쉽게 나타납니까?

혼란스러움		비교적 좋음		아주 보기 좋음
0점	25점	50점	75점	100점

6-4: 화면 설계에 관해서입니다. 레이아웃(layout)이 간단명료하고 이해하기 쉽습니까?

매우 복잡함		약간 명료함		아주 명료함
0점	25점	50점	75점	100점

6-5: 화면설계에 관해서입니다. 데이터 출력량, 배열, 색상, 하이라이트 사용이 적절합니까?

적절치 못함		비교적 적절함		아주 적절함
0점	25점	50점	75점	100점

6-6: 화면에 나타나는 데이터의 표현방식에 관해서입니다. 사용자가 쉽게 이해할 수 있게끔

간단명료하게 나타납니까?

매우 복잡함		비교적 명료함		아주 명료함
0점	25점	50점	75점	100점

6-7: 텍스트가 아닌 데이터에 관해서입니다. 그 표현형태가 그 내용 및 주제와 어느 정도 어울립니까?

(예: 수치화, 테이블화, 화상화, 영상화, 음성화)

전혀 맞지 않음		어느 정도		아주 어울림
0점	25점	50점	75점	100점

6-8: 검색된 정보가 사용자 요구에 따라 적절히 정렬, 배열, 등급화 됩니까?

전혀 아님		어느 정도		아주 적절함
0점	25점	50점	75점	100점

6-9: 검색된 정보를 어느 정도 쉽게 페이지 단위 또는 그림 단위로 프린트할 수 있습니까?

매우 어려움		어느 정도		아주 쉬움
0점	25점	50점	75점	100점

6-10: 검색된 정보를 어느 정도 쉽게 사용자가 지정하는 format나 응용소프트웨어(예: word-processor, DBMS, spread sheet)로 download할 수 있습니까?

매우 어려움		어느 정도		아주 쉬움
0점	25점	50점	75점	100점

6-11: 에러 발생시 나타나는 메시지가 사용자가 처한 상황과 대책을 명료하고 친절하게 제시합니까?

위협적임		알듯 말듯함		친절 상세함
0점	25점	50점	75점	100점

6-12: 검색과정에서 help와 exit 기능이 어느 곳에서나 지속적으로 제공되고 있습니까?

거의 없음		어느 정도		아주 완벽함
0점	25점	50점	75점	100점

6-13: 처음 DB에 접근하고자 하는 사용자가 어느 정도 복잡한 절차를 거쳐 DB에 접근할 수 있습니까? (logon 절차의 복잡도)

매우 복잡함		보통이다		아주 간단함
0점	25점	50점	75점	100점

(7) 사용자지원성 측정문항

7-1: 잘못된 정보서비스에 대한 보상 또는 배상 정책이 적절합니까?

전혀 없다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-2: 처음으로 사용하는 사용자에게 무료 또는 할인 정책이 적절합니까?

(예: 처음 얼마 동안의 무료 사용 기간 등)

전혀 없다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-3: 온라인 도움말 메뉴가 얼마나 잘 준비되어 있습니까?

전혀 없다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-4: 고객 서비스를 위한 별도의 전화나 전자메일이 제대로 준비되어 있습니까?

전혀 없다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-5: 온라인 도움말 요원이 상주하고 있습니까? 있다면 얼마나 친절합니까?

없거나 매우 불친절하다		어느 정도		아주 훌륭하다
0점	25점	50점	75점	100점

7-6: 사용자 질의에 대한 응답서비스의 신속/정확도가 어느 정도입니까?

형편없다		보통이다		아주 좋다
0점	25점	50점	75점	100점

7-7: 사용자가 품질이 좋지 않은 정보를 발견했을 경우 그것을 표시/통보하는 장치가 얼마나 잘 준비되어 있습니까?

전혀 없다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-8: 사용자 교육/훈련의 빈도, 개최장소, 참가비용과 교육내용이 어느 정도 적절합니까?

전혀 없다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-9: 인쇄물로 된 사용자 메뉴얼의 가격과 내용이 적합합니까?

전혀 없거나 너무 비싸다		어느 정도		아주 적절하다
0점	25점	50점	75점	100점

7-10: 사용자 메뉴얼이 초보 사용자에게 알기 쉽고 도움이 되도록 작성되어 있습니까?

너무 어렵다		보통이다		아주 쉽다
0점	25점	50점	75점	100점

7-11: 사용자 메뉴얼이 전문 사용자에게도 도움이 될만큼 정교하고 심층적으로 작성되어 있습니까?

전혀 도움이 안된다		어느 정도		아주 유용하다
0점	25점	50점	75점	100점

7-12: DB 사용에 필요한 시소러스(thesaurus)가 인쇄물로 잘 준비되어 있습니까?

전혀 없거나 너무 비싸다		어느 정도		아주 유용하다
0점	25점	50점	75점	100점

7-13: DB 제공자로부터 소식지, 탐색지원도구 및 기타 지원자료가 정기적으로 배부되고 있습니까?

전혀 없거나 너무 비싸다		어느 정도		아주 유용하다
0점	25점	50점	75점	100점

V. 맺는 말

본 연구에서 제안된 DB품질에 관한 연구 프레임워크는 앞으로 이 분야에 대한 연구의 필요성과 방향을 제시하고 있으며, 품질기준에 대한 본 연구의 정의는 그동안 경험적 수준에 머물던 DB 품질개념의 이론적 근거를 제시했다고 판단된다. 그리고 품질측정도구로서 개발된 문항과 척도는 상당한 논쟁의 여지를 안고 있음에도 불구하고 그 실용성과 체계성을 통해 본 연구의 가치를 입증하리라 믿는다. 한편 DB품질에 관한 주요한 문제들 중 품질관리와 품질인증 부분은 본 연구에서는 생

략되었다. 이 이슈들은 각기 그 자체로서 별도로 취급되어야 할 만큼의 폭넓고 깊이있는 내용을 지니고 있기 때문이다.

본 연구는 텍스트 DB의 품질에 초점을 맞추고 있으며 따라서 최근에 관심을 모으고 있는 멀티미디어 정보의 품질은 제외되어 있다. 그 이유로서는 첫째, 이러한 DB의 품질을 이해, 평가하기 위해서는 전통적 DB와는 전혀 색다른 패러다임이 요구되기 때문이고(예: 색상의 품질, 음향의 품질, 여러 미디어가 디자인된 상태의 품질 등), 둘째, 이러한 새로운 DB가 현재 실무적으로 활용 단계에 있다기 보다는 아직까지 연구 또는 개발단계에 머물러 있으며, 따라서 그 기술이나 개념이 일반화되지 않는 상태에 있으므로 좀 더 시간을 두고 지켜보는 것이 바람직하다고 판단되었기 때문이다. 끝으로 본 연구에서 DB 품질측정을 위해 개발된 문항 및 척도는 그 신뢰성과 타당성에 대한 통계적 검증이 이루어지지 않는 상태에 있음을 밝힌다.

〈참고문헌〉

- Armstrong, C.J., "CIQM Report on Database Quality. Quality : Deaf to Definition", *Database*, Vol. 17, No. 6, December 1994, pp. 45-48.
- Armstrong, C.J., "Database Quality : How Users Can Help - And Be Helped!", In *Computers in Libraries International '94. Proceedings of the Eighth Annual Computers In Libraries International Conference*, February 1994, pp. 105-109.
- Armstrong, C.J., "Quality : the Double Standard", *Inform 170*, December 1994, p6.
- Armstrong, C.J., "The Centre for Information Quality Management : 097-421 441 - a single 'phone number for all your woes!", *Library Technology News*, Issue 12, April 1994, pp. 3-5.
- Basch, R., "Measuring the Quality of Data : Report on the Fourth Annual SCoug Retreat", *Database Searcher*, Vol. 6, No. 8, October 1990, pp. 18-23.

- Basch, R., *Electronic Information Delivery : Ensuring Quality and Value*, 1995.
- Basch R., "Measuring the Quality of the Data : Report on the Fourth Annual SCOUG Retreat", *Database Searcher*, October 1990, pp. 18-23.
- Behesti, J., "Retrieval interfaces for bibliographic databases", *CD ROM Professional*, January 1991, pp. 50-53.
- Bendel, A., "Principles of reliability database", *Reliability databanks*, 1991, pp. 8-11.
- Brodie, M.L., "Data quality in information systems", *Information and Management*, Vol. 3, 1980, pp. 245-258.
- Burch, J.G., Strater, F.R. and Grudnitski, G., *Information systems : Theory and practice*, No. 3, 1983.
- Citroen, C., "Microcomputer software for information retrieval : How to make a well-founded choice", *Program*, Vol. 23, No. 2, 1989, pp. 141-150.
- Ciuffetti, P.D., "A plea for CD-ROM interchangability", *CD-ROM Librarian*, Vol. 6, No. 8, 1991, pp. 27-28.
- Codd, E.F., "Data models in databases management", *ACM SIGPLAN Notices*, Vol. 16, No. 1, 1981.
- Cox, J., *Online and CD-ROM Database Searching*, 1992, Meckler Publishing, London.
- Date, C.J., *An introduction to database systems*, Vol. 1, 1986.
- Davis, G.M. and Olson, M.H., *Management information systems : Conceptual foundations, structure, and development*, No. 2, 1985.
- Dolan, D.R., "Quality control at the system level", *Online*, Vol. 16, No. 2, March 1992, pp. 30-35.
- Doll, W.J. and Torkzadeh, G. "The Measurement of End User Computing Satisfaction" *MIS Quarterly*, June 1988, pp. 258-274.
- Dorn, P.H., "Business information in the eighties", *Business information systems*, Vol. 9, No. 7, 1981, pp. 245-260.
- Elmasri, R. and Navathe, S.B., *Fundamentals of database systems*, 1989.
- Fidel, R., *Database design for information retrieval*, 1987.
- Flavin, M., *Fundamental concepts for information modeling*, 1981.
- Foulds, M.S. and L.R. Foulds, "CD-ROM disk selection and evaluation", *Reference Services Review*, Vol. 18, No. 2, 1990, pp. 27-38.
- Fox, C., "The Notion of Data and Its Quality Dimensions", *Information Processing & Management*, Vol. 30, No. 1, January 1994, pp. 9-19.
- Fox, C., Gandel, P. and Frankes, B., "Foundational issues in knowledge-based information systems", *Canadian Journal of Information Science*, Vol. 13, No. 3, 1988, pp. 90-102.
- Fry, J.P. and Sibley, E.H., "Evolution of database management systems", *ACM Computing Surveys*, Vol. 8, No. 1, 1976, pp. 7-42.
- Goldberg, M., "CD-ROM periodical indexes : Bertter evaluation necessary", *The Indexer*, Vol. 18, No. 1, 1992, pp. 11-15.

- Granick, Lois., "Assuring the quality of information dissemination : responsibilities of database producers", *Information Services and Use*, Vol. 11, 1991, pp. 117-136.
- Green, M., "Setting standards for CD-ROM", *Refer*, Vol. 7, No. 3, 1991, pp. 15-19.
- Harry, V. and Oppenheim, C., "Evaluations of Electronic Databases, Part I : Criteria for Testing CD-ROM Products", *Online & CD-ROM Review*, Vol. 17, No. 4, 1993, pp. 211-222.
- Harry, V. and Oppenheim, C., "Evaluations of Electronic Database Part II : Testing CD-ROM Products", *Online & CD-ROM Review*, Vol. 17, No. 6, 1993, pp. 339-368.
- Herther, N., "A planning model for optical product evaluation", *Online*, Vol. 10, No. 5, 1986, pp. 128-130.
- Herther, N., "How to evaluate reference materials on CD-ROM", *Online*, Vol. 12, No. 2, 1988, pp. 106-108.
- Hoare, C.A.R., "Data reliability", *ACM SIGPLAN Notices*, Vol. 10, No. 6, 1975, pp. 528-533.
- Imielinski, T. and Lipski, W., "Incomplete information in relational databases", *Journal of the ACM*, Vol. 31, No. 4, 1984, pp. 761-791.
- Jardine, D.A. and Matzov, A., "Ontology and properties of time information systems", *Data and knowledge*, No. 2, 1988, pp. 173-188.
- Kent, W., "The realities of data : Basic properties of data reconsidered", *Database semantics*, No. 1, 1986, pp. 175-188.
- Kent, W., *Data and reality*, 1978.
- Kleijnen, J.P.C., *Computers and profits : Quantifying financial benefits of information*, 1980.
- Langlois, J., "CD-ROMs : Considerations before purchasing", *CD-ROM Librarian*, Vol. 5, No. 11, 1990, pp. 17-18.
- Large, J.A., "Evaluation of Online and CD-ROM Reference Sources", *Journal of Librarianship*, Vol. 21, No. 2, 1989, pp. 87-108.
- Laudon, K.C., "Data quality and due process in large interorganizational record systems", *Communications of the ACM*, Vol. 29, No. 1, 1986, pp. 4-18.
- Lee, R.M., "Logic, semantic, and data modeling : An ontology", *Data and knowledge*, No. 2, 1988, pp. 221-244.
- Liepins, G.E., and Uppuluri, V.R.R., *Data quality control : Theory and pragmatics*, 1990.
- Loebl, A.S., "Accuracy and relevance and the quality of data", *Data quality control : Theory and pragmatics*, 1990, pp. 105-144.
- Mealy, G., "Another look at data", *Proceedings AFIPS 1967 Fall Joint Computer Conference*, 1967, pp. 525-534.
- Medawar, K., "Quality in Library and Information Services : a Review", *Program*, July 1995.
- Miller, D.C., "Evaluating CD-ROMs : To buy or what to buy", *Database*, Vol. 10, No. 3, 1987, pp. 36-42.
- Morey, R.C., "Estimating and improving the quality of information in a MIS", *Communications of the ACM*, Vol. 25, No. 5, 1982, pp. 337-342.
- Nahl-Jakobovits, D. and Tenopir, C., "Databases online and on CD-ROM : how do they differ, let us count the

- ways", *Database*, Vol. 15, No. 1, February 1992, pp. 42-50.
- Nicholls, P., "A framework for evaluating CD-ROM retrieval systems", *Laserdisk Professional*, March 1990, pp. 41-46.
- Nieuwenhuysen, P., "Criteria for the evaluation of text storage and retrieval software", *The Electronic Library*, Vol. 6, No. 3, 1988, pp. 160-166.
- Ojala, M., "The dollar sign : Quality online and online quality", *Online*, Vol. 16, No. 1, January 1992, pp. 73-75.
- O'Neill, E.T., and Vizine-Goetz, D., "Quality control in online databases", *Annual review of information science and technology*, Vol. 23, 1988, pp. 125-156.
- Parnas, D.L., "The influence of software structure on reliability", *Current trends in programming methodology*, Vol. 1, 1977, pp. 111-119.
- Peters, C., "Databases on CD-ROM : comparative factors for purchase", *The Electronic Library*, Vol. 5, No. 3, 1987, pp. 154-160.
- Piternick, A.B., "Functions and capabilities of online searching systems : a checklist", *Online Review*, Vol. 13, No. 6, 1989, pp. 457-476.
- Quint, B., "The Database Connection : Full of Text", *Document Delivery World*, Vol. 9, No. 3, April/May 1993, pp. 34-36.
- Richards, T. and Robinson, C., "Evaluating CD-ROM Software : A Model", *CD-ROM Professional*, Vol. 6, No. 5, 1993, pp. 92-101.
- Rosen, L., "CD-ROM user interfaces : consistency or confusion?", *Database*, April 1990, pp. 101-103.
- Schwartz, C., "Evaluating CD-ROM Products : Yet Another Checklist", *CD-ROM Professional*, Vol. 6, No. 1, January 1993, pp. 87-91.
- Sy, K.J., and Robbin, A., "Federal statistical policies and programs : how good are the numbers?", *Annual review of information science and technology*, Vol. 25, 1990, pp. 3-54.
- Teskey, F.N., "User models and world models for data, information, and knowledge", *Information Processing & Management*, Vol. 25, No. 1, 1989, pp. 7-14.
- Travis-Nicholls, P., "A buyer's guide to CD-ROM selection : CD-ROM product directories and review tools", *CD-ROM Professional*, Vol. 3, No. 3, 1990, pp. 13-21.
- Van Camp, A.J., "User advisory groups and the online industry", *Online*, Vol. 16, No. 2, March 1992, pp. 40-45.
- Wilson, L., "Devil in the data", *Information week*, August 1992, pp. 48-54.
- Zink, S.D., "Towards more critical reviewing and analysis of CD-ROM User Software Interfaces", *CD-ROM Professional*, Vol. 4, No. 1, January 1991, pp. 16-24.