

CD-ROM 데이터베이스의 품질평가 및 개선방안에 관한 연구*

A Study on Quality Evaluation & Improvement of CD-ROM Databases

이 응 봉(Eung-Bong Lee)**

목 차

1. 서 언	4. 2 평가 대상 CD-ROM 데이터베이스의 품질평가 측정 요소 및 방법
1. 1 연구의 목적	4. 3 평가 대상 CD-ROM 데이터베이스의 품질평가의 결과분석
1. 2 연구의 방법	4. 3. 1 정확성에 대한 평가
2. 관련 연구	4. 3. 2 일관성에 대한 평가
3. CD-ROM 데이터베이스의 특성	4. 3. 3 완전성에 대한 평가
3. 1 인쇄매체와의 차이점	4. 3. 4 검색의 편의성에 대한 평가
3. 2 온라인 데이터베이스와의 차이점	4. 3. 5 검색기능의 다양성에 대한 평가
3. 3 CD-ROM 데이터베이스의 향후 전망	5. 결론 및 제언
4. CD-ROM 데이터베이스의 품질평가 측정 요소 및 결과 분석	
4. 1 평가 대상 CD-ROM 데이터베이스	

초 록

본 연구는 국립중앙도서관의 “한국문헌목록정보”와 국회도서관의 “국회도서관문헌정보” CD-ROM 데이터베이스에 대하여 데이터 자체의 품질평가 요소인 정확성, 일관성, 완전성의 세 가지와 데이터베이스 서비스의 품질평가 요소인 검색의 편의성, 검색기능의 다양성의 두 가지를 설정하여 품질을 측정하였다. 이와 함께 품질 측정결과를 비교 분석하고, 문제점에 대한 개선방안과 정책적인 제언을 제시하였다.

ABSTRACTS

This study aims to evaluate the quality of CD-ROM databases: Korean MARC on Disc and Korean National Bibliographies on CD-ROM. The five criteria(accuracy, completeness, consistency on the quality of data itself and ease of use, variety of retrieval function on the quality of database services) are used to measure their quality. The purpose of this study was to diagnose the quality of two CD-ROM databases mentioned above and to analyses the measured result of quality comparatively and to find out the problems of them and also to provide possible suggestions for their improvements.

* 이 논문은 1997년도 학술진흥재단의 신진교수과제 연구비에 의하여 지원되었음

** 충남대학교 사회과학대학 문헌정보학과 조교수
접수일자 1999년 10월 22일

1. 서 언

1. 1 연구의 목적

정보기술의 발달에 따른 컴퓨터의 도입과 네트워크화의 진전에 의한 정보화는 산업·경제·학술·연구활동 뿐만 아니라 일반가정에서의 개인생활 수준까지 광범위하게 파급되고 있다. 정보화의 급속한 진전과 정보의 가치에 대한 인식의 향상에 의하여 폭발적으로 증가하는 정보를 적시에 효율적으로 이용하기 위한 데이터베이스의 구축은 가속화되고 있으며 그 이용률 또한 점증하고 있다.

온라인 데이터베이스를 이용한 정보검색은 최신의 정보를 가장 빠르게 입수할 수 있는 반면에 데이터베이스 사용료 및 관련 통신비용 때문에 장시간 이용시 부담이 많이 발생된다. 최근에는 퍼스널 컴퓨터에 CD-ROM 드라이브가 장착되어 보급되는 것이 일반화되어 있고, 대부분의 도서관이나 정보센터에서는 LAN을 활용한 CD Network System을 구축하여 CD-ROM 데이터베이스를 다수의 이용자가 동시에 접근하여 정보검색을 하는 것을 가능하게 하고 있다. 이와 같이 정보전달 수단으로서 정보기술을 이용한 온라인 데이터베이스의 이용과 더불어 CD-ROM 데이터베이스는 그 기능을 충실하게 수행하고 있으며, 수요 또한 계속 증가하고 있다.

그러나, 데이터베이스의 품질과 관련된 문제는 어제 오늘의 일이 아니라 데이터베이스가 출현하여 이용되기 시작한 이래 계속해서 제기되어 온 고질적인 문제이다. 인터넷과 같은 통신망의 확산과 상용·공공의 온라인 및

CD-ROM 데이터베이스의 개발·보급이 급속화되면서 데이터베이스의 사용이 보편화되고 대중화되고 있는 최근에 이르러 이 문제는 더욱 더 심각해지게 되었다.

데이터베이스의 사용자층이 종래에는 특수계층인 전문가에 한정되었지만, 현재에는 그 범위가 일반대중인 초보자에게까지 확대되었다. 전문가들은 데이터베이스에 수록되어 있는 불량 데이터를 예방하거나 발견하는 나름대로의 경험과 법칙을 가지고 있고, 컴퓨터 화면상에 나타난 데이터의 입력절차나 방법을 알고 있기 때문에 그 질적인 측면이나 정확도에 대하여 한번쯤은 생각을 하게 되지만, 초보자들은 화면상의 데이터를 그것이 명백한 철자의 오류가 아닌 한 거의 믿어버리는 경향이 있기 때문에 불량 데이터의 위험에 속수무책이 되고 마는 실정이다. 따라서 쉽게 믿어버리는 만큼 그것으로 인한 피해의 빈도와 심도는 당연히 커질 수 밖에 없다.

지금까지 우리 나라의 데이터베이스 육성지원 정책은 “규모의 확대와 양적인 증가 및 사용자층의 확산”에 치중하여 왔다. 따라서 기 구축된 데이터베이스의 품질에 관한 연구는 미약한 실정이다. 앞으로의 데이터베이스 관련 정책이 “데이터베이스의 품질 향상을 통한 질적인 발전과 사용자층의 만족 제고”에 초점을 맞추어야 한다는 것을 염두에 두면, 현재 국가도서관에서 구축하여 운영하고 있는 온라인 및 CD-ROM 데이터베이스의 정확한 품질 측정 및 개선방안에 관련된 다양한 연구는 그 이용대상이 전국민을 대상으로 하고 있고, 파급효과 등을 고려할 때 필요불가결한 것이며 시급한 과제가 아닐 수 없다.

본 연구의 결과는 향후 CD-ROM 데이터베이스 시스템의 설계자들에게 고품질의 CD-ROM 데이터베이스를 구축하는데 도움을 주고, CD-ROM 데이터베이스의 이용자들에게는 효율적인 정보탐색을 위한 하나의 지침으로서 활용될 수 있을 것이다.

1. 2 연구의 방법

지금까지 수행되어 온 데이터베이스 관련 연구는 대부분 데이터베이스 시스템 구축과 관련된 제반 기술, 데이터의 처리, 탐색서비스 및 이용자에 관하여 치중되었고, 데이터베이스 그 자체의 품질평가에 관한 것은 상대적으로 미약하다. 최근 들어, 데이터베이스의 품질에 관한 관심과 연구는 미국 및 영국 등에서 매우 활발하게 이루어지고 있는데, 서지 데이터베이스에 수록된 데이터의 철자에러 검출과 수정기법 등 뿐만 아니라 디렉터리/수치/전문/멀티미디어 데이터베이스에 대한 전반적인 품질관리 및 평가를 다루고 있다.

한편, 온라인 데이터베이스와 CD-ROM 데이터베이스는 시스템 환경이나 활용면 등 여러 가지 부문에서 명백한 차이가 있기 때문에 온라인 데이터베이스의 평가기준을 CD-ROM 데이터베이스에 그대로 적용하는 것은 무리가 따를 수 밖에 없다. 국내에서 수행된 관련 연구 또한 대부분 CD-ROM 데이터베이스의 선정기준 등과 관련된 것이 대부분인 실정이다.

따라서 본 연고에서는 앞으로 정보기술의 진전과 더불어 보급·이용이 급증할 것으로 예상되는 CD-ROM 데이터베이스의 품질 평

가기준 및 품질 측정요소를 제시함과 동시에, 특히 데이터 축적건수, 이용자 측면 및 파급효과 등을 고려하여 우리 나라의 납본도서관이자 국가도서관의 중추적인 역할을 수행하고 있는 국립중앙도서관과 국회도서관에서 구축하여 이용자서비스를 제공하고 있는 CD-ROM 데이터베이스인 “한국문헌목록정보”와 “국회도서관문헌정보”를 대상으로 하여 본 연구에서 제시한 품질 평가기준 및 품질 측정요소를 근거로 하여 품질을 평가하여 문제점을 도출하고 개선 방안을 제시하고자 한다.

2. 관련 연구

데이터베이스 산업이 규모면 및 질적인 면에서 우리 나라에 비하여 한 단계 높은 수준에 있는 미국이나 유럽 등의 경우에는 데이터베이스의 품질에 관하여 오래 전부터 관심을 기울여 왔으며, 데이터베이스의 품질을 향상하기 위한 노력 또한 활발하게 이루어지고 있다. 대부분의 데이터베이스 관련 저널에는 신규 제작된 데이터베이스를 소개하는 경우, 그들 나름대로의 기준을 적용하여 해당 데이터베이스의 품질 평가에 관한 내용을 수록하고 있는 것이 최근의 동향이다.

데이터베이스의 품질에 관한 논의는 National Online Meeting, Online/CD-ROM '90 및 International Online Information Meeting 등 국제회의에서 1990년의 슬로건으로 “데이터베이스의 품질”을 채택하여 선언하면서 이에 대한 관심과 연구를 전세계적으로 확산시켰다고 할 수 있다.

우선 주요국의 연구동향을 살펴보면, 미국의 SCOUG(Southern California Online User's Group)에서는 1990년에 개최된 제 4차 회의에서 데이터베이스 품질평가 기준으로 일관성, 수록범위, 적시성, 정확성, 접근성 및 사용의 편리성, 통합성, 출력결과, 도큐멘테이션, 이용자 지원 및 교육, 그리고 비용대 가치 등의 10가지를 제시하였고(Basch, 1990), 서지, 본문 및 디렉토리 형태의 데이터베이스의 출력물이나 검색결과를 중심으로 데이터베이스의 품질정도를 나타내는 "Quality Score" 방법을 개발하여 탐색자 관점에서 데이터의 품질을 평가할 수 있는 다양한 요소를 제시하고 있다(Granick, 1991). 영국에서는 LA(Library Association)와 UKOLUG(UK Online User Group)의 재정지원 하에 CIQM(The Center for Information Quality Management)을 설립하여 데이터베이스의 품질 개선 및 측정 기준을 개발하고 있다(Lester, 1994). 핀란드의 피니스 프로젝트(The Finnish Society for Information Services)에서는 서지 데이터베이스에 국한하여 평가기준을 제시하고 있는데 그 기준은 데이터베이스의 통신과 시스템간의 접속여부, 검색어와 검색지원 도구, 데이터의 내용에 대한 품질, 정보검색시 제공되는 도움말 기능 및 검색비용 등의 다섯 가지 항목이다. 일본에서는 정보과학기술협회가 주관이 되어 데이터베이스의 품질관리에 관한 연구를 계속적으로 진행하고 있다. 국내에서는 1996년에 한국데이터베이스진흥센터에서 데이터베이스 수록데이터의 품질기준과 데이터베이스 서비스의 품질기준을 제시하였는데 전자로는 정확

성, 완전성, 최신성, 일관성을, 후자로는 검색성, 사용의 용이성, 사용자 지원성을 제시하고 있다.

다음으로 테노피(Tenopir, 1987)는 데이터베이스의 품질관리에 영향을 가장 크게 미치는 요인으로 범위, 구조, 접근성, 정확성, 일관성의 다섯 가지를 제시하고 있다. AT&T Bell 연구소에서는 데이터베이스의 품질 평가 기준으로 정확성, 최신성, 완전성 및 일관성의 네 가지를 제시하고 있다. 오네일 등(O'Neill et al., 1988)은 데이터베이스의 검색결과에 품질에 영향을 미치는 요인으로 데이터베이스에 수록된 데이터 자체의 품질요소와 검색기능에 관련된 요소의 두 가지로 구분하면서 수동적 품질관리의 요소로 교육, 리뷰, 이용자의 협조의 세 가지를 능동적인 관리방법으로 자가점검 데이터의 이용, 자동 데이터 검증 및 전자통제의 세 가지를 제안하고 있다.

한편 CD-ROM 데이터베이스의 품질평가에 관해서 살펴보면, 자스코(Jasco, 1992)는 소프트웨어, 데이터웨어, 데이터베이스 및 하드웨어의 네 가지를 언급하고 있다.

3. CD-ROM 데이터베이스의 특성

CD-ROM 데이터베이스는 온라인 데이터베이스에 대하여 데이터베이스를 CD-ROM화한 것을 말하며, 이것은 PC를 사용한 정보 검색기법에 익숙한 이용자가 통신회선의 사용료를 의식하지 않고 그 부담에서 벗어나 각자의 위치에서 자유로운 검색이 가능하다. 이 장에서는 기존의 인쇄매체 및 온라인 데이터베

이스와 뚜렷이 구별되는 주요 차이점에 관하여 간략하게 살펴보기로 한다.

3. 1 인쇄매체와의 차이점

인쇄매체와 CD-ROM 데이터베이스는 일단 생산이 되면 매체 자체에서는 수록정보의 수정이나 갱신은 불가능하지만, 랜덤 액세스(random access)가 가능하다는 공통점을 지니고 있다. 인쇄매체 자료는 이용자의 친숙성에 장점을 가지는 반면, CD-ROM 데이터베이스는 최신의 정보기술을 응용한 CD-ROM이라는 전자매체를 이용하여 이용자가 원하는 정보에 대한 신속한 검색을 가능하게 한다. 인쇄매체와 CD-ROM이라는 전자매체의 전망에 관해서는 수많은 논쟁이 이루어지고 있지만, 서지 및 색인정보 CD-ROM 데이터베이스는 정보의 접근성, 신속성 및 보편성 측면에서 이미 인쇄매체보다는 우위를 점하고 있다는 것은 명약관화한 사실이며, 최근에는 학술저널, 학위논문, 뉴스 등을 대상으로 하여 텍스트 뿐만 아니라 화상, 동영상, 음성 데이터 등을 포함한 멀티미디어 형태의 전문 데이터베이스(full-text database) 구축으로 그 범위를 확장하고 있고, 기존의 인쇄매체를 중복 수록한 것이 아닌 CD-ROM에만 고유하게 수록한 데이터베이스도 증가하는 추세에 있다.

CD-ROM 데이터베이스는 기존의 인쇄매체와 비교하여 다음과 같은 뚜렷한 특성을 가지고 있다. 첫째, 물리적 접근에 의존하는 인쇄매체에 비하여 CD-ROM 데이터베이스는 전자적인 접근(키워드 및 주제검색, 불 연산자

사용, 개념조합 등)이 가능하므로 보다 신속하게 정보의 정확한 검색이 가능하다. 둘째, CD-ROM 데이터베이스는 네트워크 구축(CD Network)을 통하여 다수의 이용자로 하여금 동일한 데이터베이스에 동시접근을 가능하게 한다. 셋째, 인쇄매체의 증가는 그 속성상 공간문제를 야기하는데 반하여, CD-ROM 데이터베이스는 대용량의 저장매체이므로 이와 같은 공간문제에서 해방될 수 있다. 넷째, 정보자원의 보존이라는 측면에서 CD-ROM 데이터베이스는 인쇄매체에 비하여 내구성이 우수하고 정보기술의 발전 속도를 감안하면, 반영구적이라 할 수 있다. 다섯째, CD-ROM 데이터베이스는 인쇄매체에 비하여 타 저장매체로의 변환이 가능하고, 다운로드가 용이하며, 자유로운 편집이 가능하므로 이용자의 요구를 충족시킬 수 있다. 여섯째, 정보생산의 시차적인 관점에 보면, 초창기에는 CD-ROM 데이터베이스가 단순히 인쇄매체의 자료를 재수록하여 이용하게 하는 수단에 불과하였지만, 현재에는 인쇄매체의 자료와 동시에 생산되거나, CD-ROM 데이터베이스 자체만으로 간행되는 경우가 허다하기 때문에, 인쇄매체와의 시차가 발생할 여지는 거의 없다고 할 수 있다.

3. 2 온라인 데이터베이스와의 차이점

온라인 데이터베이스는 도서관 전산화 작업의 일환으로 시작된 목록시스템의 데이터베이스가 그 효시가 된다고 할 수 있다. 정보기술 특히, 컴퓨터기술의 급속한 발전으로 대용량의 하드디스크가 등장하고 가격 또한 저렴해

집에 따라 온라인 데이터베이스는 초기의 텍스트 정보만을 수록한 서지 데이터베이스의 형태를 벗어나 최근에는 전문(full-text) 및 멀티미디어 정보를 수록한 데이터베이스로 급속하게 발전하고 있다.

단순하게 성능만을 비교한다면, CD-ROM 데이터베이스는 온라인 데이터베이스에 비하여 그 효율성이 낮다고 할 수 있지만, 검색 성능이 뛰어난 CD-ROM 데이터베이스가 기존의 인쇄매체를 완전히 대체하지 못하는 것과 마찬가지로 온라인 데이터베이스에 비하여 그 성능에 있어 뒤떨어지는 CD-ROM 데이터베이스는 나름대로의 특징적인 정보저장 및 전달매체로서의 역할을 수행하고 있다. CD-ROM 데이터베이스는 온라인 데이터베이스와 비교하여 다음과 같은 차이점을 지니고 있다. 첫째, 온라인 데이터베이스는 이용자 접근에 아무런 제한이 없지만, CD-ROM 데이터베이스는 네트워크 장비의 드라이버 수에 따라 이용자를 제한할 수 밖에 없는 한계점을 가지고 있다. 그러나 최근 네트워크 기술(juke box)의 발전으로 특정 기관내에서의 이용에서는 이와 같은 이용자 접근의 제한이라는 어느 정도의 한계를 극복하는 것은 가능하다. 둘째, 온라인 데이터베이스는 수록된 데이터에 대한 자유로운 수정과 갱신이 가능하지만, CD-ROM 데이터베이스는 재생전용이므로 자체적으로 수록된 데이터의 수정 및 갱신이 불가능하다. 셋째, 온라인 데이터베이스 특히, 상용의 경우에는 고가의 사용료(접속료, 전송료 및 DB 이용료)를 지불해야 하므로 이용자가 사용시간에 제한을 느끼는데 반하여, CD-ROM 데이터베이스는 시간적인 제약이

없기 때문에 경제적이 정보검색이 가능하다. 넷째, CD-ROM 데이터베이스는 온라인 데이터베이스와는 달리 출판 및 배포의 형태를 취하기 때문에 수정 및 갱신과 동시에 서비스의 실시가 불가능하다는 최신성의 문제점을 지니고 있다. 다섯째, CD-ROM 데이터베이스는 온라인 데이터베이스와 마찬가지로 다양한 접근점을 제공하기 때문에 이를 활용하여 수록된 정보를 신속하고 다양하게 검색하는 것이 가능하다.

3. 3 CD-ROM 데이터베이스의 향후 전망

오늘날, 정보기술의 급속한 발전에 따라 새롭게 등장하고 있는 Optical Disc, Disc Array, DVD-ROM, 대용량의 Hard/Floppy Disc 등과 같은 정보저장 및 전달매체는 향후 CD-ROM 데이터베이스의 존속여부에 큰 영향을 미칠 것으로 보인다. 위에서 살펴본 바와 같이 CD-ROM 데이터베이스는 기존의 인쇄매체보다는 비교적 정보저장 및 전달매체로서 우수한 여러 가지 성능을 가지고 있지만, 온라인 데이터베이스에 비해서는 여러 가지 면에서 많은 단점을 지니고 있다. 그럼에도 불구하고, CD-ROM 데이터베이스는 여전히 앞으로도 계속 정보저장 및 전달매체로서의 기능과 역할을 수행할 것으로 생각된다. 그 주요 이유는 다음과 같다. 첫째, CD-ROM 데이터베이스는 정보 저장매체로서의 보편성을 확보하고 있다는 점이다. '80년대 중반부터 비약적으로 증가하기 시작한 CD-ROM 데이터베이스는 이미 각 정보센터, 도서관은 물론 가정에까지 널리 보급되어 광범

위하게 사용되고 있다. 둘째, CD-ROM은 이미 국제표준(ISO 9660)이 제정되어 있다는 점이다. CD-ROM 규격에 관한 국제표준의 확립은 CD-ROM 데이터베이스를 운용하는 주변기기의 표준화에도 동시에 중요한 영향을 미치며, 반대로 표준화된 주변기기는 CD-ROM 데이터베이스의 생산 또한 활성화하는 계기를 제공한다. 셋째, CD-ROM 데이터베이스는 대량복제가 용이하기 때문에 저렴한 가격으로 각 정보센터, 도서관, 기관 및 가정까지 손쉽게 배포가 가능하다는 점이다. 따라서, CD-ROM 데이터베이스는 특히, 정보유통의 측면에서 온라인 데이터베이스에 비해 저렴한 가격으로 정보를 공급하는 것이 가능하므로 앞으로도 계속해서 발전할 가능성이 있다. 넷째, CD-ROM 데이터베이스와 온라인 데이터베이스는 상호 보완적인 형태로 발전할 수 있다는 점이다. 양자는 가격경쟁력의 측면과 최신성의 측면에서 각각 장점을 가지고 있으므로 최신성을 요하는 정보는 온라인 데이터베이스로, 일정기간이 경과한 정보는 CD-ROM에 저장하여 이용할 수 있을 것이다.

4. CD-ROM 데이터베이스의 품질평가 측정 요소 및 결과 분석

4. 1 평가 대상 CD-ROM 데이터베이스

평가의 대상으로 선정한 CD-ROM 데이터베이스는 데이터 축적건수, 이용자 측면 및 파급효과 등을 고려하여 우리 나라의 납본도서관이자 국가도서관으로서의 중추적인 역할을

수행하고 있는 국립중앙도서관과 국회도서관에서 구축하여 이용자 서비스를 제공하고 있는 CD-ROM 데이터베이스인 “한국문헌목록정보(보판, 1998)”와 “국회도서관문헌정보(N.5, 1998)”를 대상으로 한다(〈표 1〉 참조).

“한국문헌목록정보”는 국립중앙도서관에서 수집한 1945년 이후 발간된 국내문헌을 대상으로 구축한 일반도서, 학위논문 등 약 70여만건의 도서목록 데이터베이스를 CD-ROM으로 제작한 것이다. 1994년 시제품을 생산한 후 1995년 가을판(1회)을 발행하였고, 1996년 이후 현재까지 년 2회 계속적으로 발행하고 있다.

“국회도서관문헌정보”는 국회도서관에 입수되는 국내 정기간행물 중에서 의정활동에 필요하거나 학술적 가치가 있는 기사를 수록한 정기간행물기사색인과 우리 나라의 전국 대학원 및 특수대학원에서 수여된 박사 및 석사학위 취득자의 논문목록, 인문사회과학분야의 박사학위논문 목록 및 국회도서관에 입수되는 단행본 목록 데이터베이스를 CD-ROM으로 제작한 것으로서 정기간행물 기사색인 데이터베이스는 1977년 1월부터 현재까지 국회도서관에서 발간된 정기간행물기사색인의 데이터베이스를 수록하고 있고, 주요 시사주간지의 기사색인은 1995년 1월 이후 자료를 대상으로 하고 있으며, 현재 데이터 축적건수는 5,000여종 85만여건에 이르고 있다. 그리고 한국 석사/박사 학위논문 데이터베이스는 우리 나라 각 대학·대학교 및 대학원(국방대학원, 한국과학기술원 및 한국정신문화연구원 포함)에서 수여된 석사/박사 학위취득자 총 47여만명의 학위논문을 수록하고 있으며, 수록범위는

〈표 1〉 평가대상 CD-ROM 데이터베이스의 주요 자원 비교

특성	구분	한국문헌목록정보	국회도서관문헌정보	비고
대상		한국문헌목록정보 볼판(Spring), 1998	국회도서관문헌정보 N.5 1998	
제작년도		1998. 04. 27	1998. 07. 01	
갱신주기		년 2회	년 2회(N.7부터 연4회 발행)	
데이터 축적 건수		일반도서 337,544건 학위논문 375,099건 등 70만 여건	정기간행물 기사색인 : 4,983종 837,067건 한국 박사/석사 학위논문 : 총466,979건 단행본목록 : 302,334건 UN·의회자료 : 14,309건	
H/W환경		IBM 호환 PC 386SX이상 640 KB의 메인 메모리(그래픽 모드 사용시 최소 1MB 필요) VGA 또는 HERCULES 디스플레이 어댑터	IBM호환 PC 586 이상 16M RAM(32M 권장) 4배속 이상 CD-ROM 드라이브(6배속 이상 권장)	
운영체제		MS-DOS 5.0 이상	한글 Windows 95 이상	
주요 기능		자료찾기 검색결과와 출력 (간략정보, 상세정보, KORMARC Data) 목록정보 편집 및 저장	자료검색/저장/인쇄 (간략정보, 상세정보, MARC Data)	
분류표		KDCP(1983년 이전 자료) KDC(1984년 이후 자료) DDC(1984년 이후 자료)	DDC 19판 (1991년 이전 자료는 DCC 16판 사용) 동양관계 세분전개표(국회도서관 편)	
수록범위		국립중앙도서관에서 수집한 1945년 이후 발간된 국내 일반도서 및 석·박사 학위논문 70여만 건의 목록정보	정기간행물기사색인 : 1997년 1월부터 1998년 3월까지 국회도서관에서 발간된 정기간행물기사색인 DB 주요시사주간지의 기사색인 : 1995년 1월부터 1998년 3월까지 한국박사/석사 학위논문 : 1945년부터 1998년 2월 졸업자까지 단행본 : 동양서(1982년~1998년 5월까지 등록된 자료), 서양서(1985년~1998년 5월까지 등록된 자료) UN자료 : 1998년 현재 국회도서관 소장목록	
목록기술		KORMARC(한국문헌자동화목록)	한국목록기술규칙 제3판(KCR 3)	
제작기관		서울시스템	대한민국 국회도서관	

1945년 이후 졸업자를 대상으로 하고 있고, 명예박사학위 수여자 목록은 1973학년도 이후부터 수록하고 있다.

4. 2 평가 대상 CD-ROM 데이터베이스의 품질평가 측정 요소 및 방법

제2장의 관련연구에서 살펴 본 바와 같이 데

데이터베이스의 품질평가는 크게 수록된 데이터 자체의 품질과 데이터베이스 서비스의 품질로 구분할 수 있는데, 본 고에서는 데이터 자체의 품질평가 요소로는 ① 정확성, ② 일관성, ③ 완전성의 세 가지를, 그리고 데이터베이스 서비스의 품질평가 요소로 ① 검색의 편의성, ② 검색기능의 다양성의 두 가지를 설정하여 전체 다섯 가지의 측정 요소를 사용하기로 한다.

연구대상으로 선정된 CD-ROM 데이터베이스들의 검색실험은 1999년 1월부터 4월까지 본 연구자와 연구보조원 2명(대학원생)에 의하여 1차로 수행되었고, 1차 실험검색에서 나타난 문제점을 보완하여 1999년 7월 한달 동안 2차 실험탐색을 실시하여 본 연구에서 설정한 측정요소에 따라 평가를 종합하였다.

4. 3 평가 대상 CD-ROM 데이터베이스의 품질평가의 결과분석

4. 3. 1 정확성에 대한 평가 데이터베이스를 이용하는 탐색자의 입장에

서 보면, 데이터베이스의 구축과정에 있어서 색인어의 입력시 오자 또는 일관성이 결여된 표기방법은 불완전한 검색의 원인이 되기 때문에 항상 이점을 염두에 두어 확인을 하면서 검색을 수행해야 한다. 특히, 외래어의 경우 지역이나 시대에 따라 하나의 용어가 서로 다른 방법으로 표기되어 사용되기도 한다.

본 연구에서는 평가 대상 CD-ROM 데이터베이스의 수록 데이터의 정확성을 측정하기 위하여 동일한 의미를 가지고 있지만 다양하게 표기되고 있는 외래어들 중 “네트워크, 데이터베이스, 디지털, 윈도우즈”의 네 가지 용어를 대상으로 하여 주제어(키워드)와 서명(단어 완전일치 및 우측절단) 필드에 국한하여 표기방법에 따른 검색결과를 비교·분석하였다.

“한국문헌목록정보” CD-ROM 데이터베이스의 서명단어 필드는 우측절단이 디폴트(default)로 지정되어 있는 경우이며, 서명 필드는 서명 중의 수록단어와 검색어가 일치하는 경우에만 검색되는 경우이다. <표 2>에서

<표 2> 한국문헌목록정보의 정확성 평가 내용

DB명	검색용어	필드별 검색결과			필드별 OR 연산자 적용건수			비고
		서명단어	서명	주제어	서명단어	서명	주제어	
한국문헌 목록정보	네트워	4	0	2	164 299 (1건 차이)	19 2 39 39 4	97 112 39	완성형 한글을 지원하지 않는 문자가 있다. (Ex. 획)
	네트워크	79	4	16				
	데이터베이스	96	9	72				
	데이터베이스	68	10	25				
	디지털	159	2	74				
	디지털	141	0	38				
	윈도	367	0	39				
	윈도우	340	1	39				
	윈도우즈	132	3	31				

〈표 3〉 국회도서관문헌정보의 정확성 평가 내용

DB명	검색용어	필드별 검색결과			필드별 OR 연산자 적용건수			비 고
		서명1	서명2	키워드	서명1	서명2	키워드	
국회도서관 문헌정보	네트웍	3	2	3	63	50	176	완성형 한글 을 지원하지 않는 문자가 있다. (Ex. ㅿ)
	네트워크	60	48	173				
	데이터베이스	75	62	231	140 (1건 차이)	120 (1건 차이)	326 (6건 차이)	
	데이터베이스	66	59	101				
	디지털	33	26	39	70	48	119	
	디지털	37	22	80				
	윈도	141	13	161	조합불능	113 (2건 차이)	조합불능	
	윈도우	125	63	157				
	윈도우즈	56	39	72				

나타난 바와 같이 검색어가 “디지털, 디지털”인 경우에 있어서 “서명단어 필드”의 경우 디지털(159건), 디지털(141건)의 OR 연산자 적용건수가 299건으로 1건의 차이가 발생하는데 그 이유는, “디지털”이라는 검색어로 검색을 했는데도 불구하고 “디지털”로 표기된 문헌이 한 건 중복 검색되었기 때문이다. 그리고 “윈도(367건), 윈도우(340건), 윈도우즈(132건)”의 OR 연산자 적용건수가 367건으로 나타난 것은 앞에서 설명한 바와 같이 서명단어 필드의 경우 우측절단이 디폴트로 되어 있기 때문이다. 그런데, 주제어 검색의 경우에는 특이한 사항을 발견할 수 있다. “주제어 필드”의 경우, “윈도(39건), 윈도우(39건), 윈도우즈(31건)”의 OR 연산자 적용건수가 39건으로 나타났는데, 그 이유는 영문인 “windows”란 주제어의 경우, 한글로 주제어를 부출할 때 “윈도, 윈도우, 윈도우즈”의 3가지 모두를 해야 함에도 불구하고, 입력시 8건의 문서 레코드의 경우에 있어서 “윈도우즈”란 주제어의 부출이 누락되어 있기 때문이다.

“국회도서관문헌정보” CD-ROM 데이터베이스의 경우에는 서명필드의 검색시 디폴트로 지정되어 있는 것은 우측절단을 의미(서명1 필드)하는 것이며, 디폴트를 해제할 경우에는 서명 중의 수록단어와 검색어가 일치(서명2 필드)하는 경우에만 검색되는 경우이다. 〈표 3〉에서 나타난 바와 같이 검색어가 “데이터베이스 데이터베이스”인 경우에 있어서 OR 연산자 적용건수에서 “서명1 필드”와 “서명2 필드”에서 각각 1건, 그리고 “키워드 필드”에서는 6건의 차이가 발생하는데 그 이유는 “서명1 필드”와 “서명2 필드”의 경우에 있어서는 동일자료가 중복 검색된 결과로 판명되었다. “키워드 필드”에서 발생한 6건의 차이는 6건의 레코드 공히 키워드로 “데이터베이스”와 “데이터베이스”를 동시에 부출하였기 때문인 것으로 밝혀졌다. 이와 같은 결과는 “국회도서관문헌정보”는 정확성의 문제가 매우 심각한 수준에 이르고 있다는 것을 나타내는 것으로 시급하게 수정이 되어야 할 사항이다.

그리고, “서명2 필드”에서는 윈도(13), 윈도

우(63), 윈도우즈(39)의 OR 연산자 적용건수가 113건으로 2건의 차이가 나타났다 “윈도우”와 “윈도우즈”를 탐색어로 하여 검색한 결과 2건의 레코드가 동시에 양자에서 검색되었기 때문이다. 해당 레코드 2건을 확인해 본 결과 1건은 번역서로 초판번역 서명과 재판의 서명이 달라진 경우였고¹⁾ 또 1건의 레코드는 서명/저자사항과 주기사항에 “윈도우즈”와 “윈도우”란 용어가 동시에 수록되어 있었기 때문이다²⁾. 그러나 “서명1 필드”와 “키워드 필드”의 OR 연산자 적용 건수는 확인이 불가능하였다. 왜냐하면 두 필드 공히 OR 조합이 가능한 탐색어의 수가 2개 이하로 한정되어 있었기 때문이다.

4. 3. 2 일관성에 대한 평가

일관성의 평가는 주제에 대한 일관성 유지를 위한 상호참조의 기능, 약어처리, 대소문자 구별 및 띄어쓰기 등의 4가지 항목을 설정하여 조사하였다(표 4 및 표 5 참조).

우선, 상호참조의 기능은 평가대상인 “한국

문헌목록정보”와 “국회도서관문헌정보” 두 곳 모두 지원을 하지 않고 있었고 두 번째, 약어 처리의 경우에 있어서는 DB라는 약어를 대상으로 하여 조사하였는데, 전자는 “서명단어 필드”에서만 약어인 DB가 12건이 검색되었을 뿐, “서명 필드”와 “주제어 필드”에서 검색결과가 한 건도 나타나지 않았다. 후자는 “서명1 필드”에서는 database가 98건, DB는 586건으로 약어인 DB가 오히려 488건이나 많이 검색된 것에 비해, “키워드 필드”에서는 database가 140건, DB는 68건으로 database가 “서명1 필드”와는 반대로 72건이 더 많이 검색되었다. 이와 같은 결과는 일관성의 측면에서 보면, 매우 큰 문제점을 내포한다고 할 수 있다. 그리고 database와 DB의 “OR 연산자 적용 건수”를 살펴보면, 전자의 경우는 “서명단어 필드”에서 3건의 차이를 보이고 있고, 후자는 “서명1 필드”에서 5건, “키워드 필드”에서 3건의 차이를 보이고 있는데, 원인은 전자와 후자 모두 동일자료의 중복검색에 기인하는 것으로 나타났다. 세 번째 대소문자 구별은 양자가 공

1) 해당 레코드의 상세정보는 다음과 같다

청구기호 : 001.6425 C167t2

서명/저자 : (초보자를 위한)윈도우 95 프로그래밍 21일 완성 / Charles Calvert 저: 김영훈 역

발행사항 : 서울 : 인포 북, 1996

형태사항 : 816 p. : 삽도 : 26cm + 디스켓 1장

주기사항 : 초판원서명 : Teach yourself windows programing in 21 days 초판번역서명 : (초보자를 위한)윈도우즈 21일 완성

원서명 : Teach yourself windows programing in 21 days -2nd ed.

키워드 : 윈도우:95:windows:95

2) 해당 레코드의 상세정보는 다음과 같다.

청구기호 : 001.6425 ㅅ323ㅎ

서명/저자 : 한글 윈도우즈 3.1 쉽고 빠르게 / 손동일 저

발행사항 : 서울 : PCBOOK, 1995

형태사항 : 247 p. : 삽도 : 26cm

주기사항 : 판권기서명 : 한글 윈도우 3.1 쉽게 빠르게

ISBN : 8981930031

키워드 : 한글윈도우즈3.1

〈표 4〉 한국문헌목록정보의 일관성 평가 내용

DB명	상호 참조	약어처리				대소문자 구별				띄어쓰기				
		검색용어	서명 단어	서명	주제어	검색용어	서명 단어	서명	주제어	검색용어	서명 단어	서명	주제어	
한국문헌 목록정보	없음	DB	12	0	0					cd rom	17	0	0	
										cd-rom	17	0	0	
							internet	96	12	0	cdrom	17	0	0
		database	88	1	0					data base	88	1	0	
		DB or database	97 (3)	1		INTERNET	96	12	0	database	88	1	0	
									on line	11	0	0		
									on-line	11	0	0		
									online	11	0	0		

〈표 5〉 국회도서관문헌정보의 일관성 평가 내용

DB명	상호 참조	약어처리				대소문자 구별				띄어쓰기				
		검색용어	서명 단어	서명	주제어	검색용어	서명 단어	서명	주제어	검색용어	서명 단어	서명	주제어	
국회도서관 문헌정보	없음	DB	586	0	68					cd rom	51	0	91	
										cd-rom	0	0	0	
							internet	94	93	147	cdrom	17	3	0
		database	98	72	140					data base	98	21	38	
		DB or database	679 (5)	72	205 (3)	INTERNET	94	93	147	database	33	72	140	
										on line	0	0	13	
										on-line	0	0	0	
								online	60	60	84			

〈표 6〉 국회도서관문헌정보의 일관성(띄어쓰기) 평가 내용

DB명	띄어쓰기							비 고
	검색용어	서명1	서명2	키워드	OR 연산자 적용건수			
					서명1	서명2	키워드	
국회도서관 문헌정보	cd rom	51	0	91				cdrom or cd rom cdrom or (cd rom) database or data base database or (data base) online or on line online or (on line)
	cd-rom	0	0	0	51	3	91	
	cdrom	17	3	0	(17)	(3)	(0)	
	data base	98	21	38	33	27	38	
	database	33	72	140	(98)	(72)	(140)	
	on line	0	0	13	29	4	13	
	on-line	0	0	0	(60)	(60)	(84)	
online	60	60	84					

히 소문자와 대문자를 구별하지 않고 동일하게 취급하는 것으로 나타났다. 마지막 네 번째 항목인 띄어쓰기 부분에 있어서는 일반적으로 혼용되고 있는 cd rom(cd-rom, cdrom), database(data base), online(on line, on-line)이라는 용어를 대상으로 하여 조사하였는데, “한국문헌목록정보”는 database(data base)를 제외하고는 모두 “서명단어 필드”에서만 검색결과를 확인할 수 있었으며, 검색결과 건 수 또한 띄어쓰기에 상관없이 동일한 값을 나타내주고 있다. 따라서 “한국문헌목록정보”에서는 띄어쓰기에 상관없이 동일한 검색결과를 제공한다고 볼 수 있다. 그리고 “서명 필드”에서 검색된 database(1건), data base(1건)이 역시 동일한 자료인 것으로 판명되었다.

“국회도서관문헌정보”는 띄어쓰기 부문에 있어서 조사대상으로 선정된 용어가 “서명1 필드”, “서명2 필드” 및 “키워드 필드”에서 검색결과를 각각 나타내고 있다. 그러나 “국회도서관문헌정보”는 띄어쓰기에 있어서 심각한 다수의 문제점을 가지고 있는 것으로 나타났다(표 5, 표 6 참조). 각 필드별 “OR 연산자 적용 건 수”를 살펴보면, 일관성이 전혀 없는 것을 알 수 있다. 첫 번째 문제점은 “서명단어 필드”를 제외한 “서명1 필드”와 “키워드 필드” 다시 말하면, 디폴트가 우측절단으로 되어 있는 필드에서는 2개 이상의 연산자를 사용할 경우 검색결과를 경우에 따라 제공하기도 하고 그렇지 않은 경우가 있는 등 시스템이 매우 불안정하다는 점이다. 두 번째 문제점은 2개 이상의 연산자가 사용 가능한 “서명단어 필드”에서도 일반적인 검색시스템에서 적용하

고 있는 연산자의 검색 우선순위(NOT→AND→OR)가 적용되지 않고, 먼저 사용된 연산자가 먼저 조합되는 다시 말하면, 좌측에서 우측순으로 연산자의 검색순위가 정해져 있다는 점이다. 이 점에 관해서는 일러두기에서 전혀 언급이 되지 않고 있다. 그리고 세 번째 문제점은 물론, 일러두기에 언급은 되어 있지만 연산자를 사용하지 않은 각 단어사이의 공간(space)을 “AND”로 인식하기 때문에 일반적으로 대부분의 시스템이 채용하고 있는 각 단어사이의 공간(space)을 “OR”로 인식하는 것에 익숙한 이용자는 상당한 주의를 필요로 하는 것으로 나타났다. 이와 같은 문제점으로 인하여 <표 6>에 나타난 바와 같이 조사 대상으로 선정된 검색어의 “OR 연산자 적용 건 수”는 일관성이 전혀 없는 것으로 나타났다. 마지막 네 번째의 문제점은 연산자의 조합이 초기화면에서는 불가능하다는 점이다. 일단 하나의 검색어를 입력하고 검색결과를 확인한 후부터 연산자의 조합이 가능한데, 이 점 역시 일러두기에 전혀 언급이 되고 있지 않은 사항이다.

4. 3. 3 완전성에 대한 평가

완전성에 대한 평가 항목으로는 데이터베이스의 구성에 있어서 불완전성에 대한 사전안내와 제한점의 언급 여부, 각 레코드의 필드 구성, 접근점이 되는 필드와 가능한 제한탐색의 범위 등을 설정하여 조사하였다(표 6 참조).

“한국문헌목록정보”의 경우에는 1983년 이전 자료와 이후 자료에 대한 검색방법의 차이를 보이고 있는데, 큰 차이점은 주제어 검색의 가능 여부와 채택된 분류표의 차이이다. 따라

서 이용자는 소급자료의 탐색에 있어서 주제어와 분류기호를 접근점으로 하는 검색의 경우에는 세심한 주의를 기울여야 한다. 그리고 필드구성은 “한국문헌목록정보”가 10개의 필드, “국회도서관문헌정보”는 8개의 필드로 구성되어 있다. 전자의 “자료구분 필드”와 후자의 “본문언어 필드” 이외에는 특징적인 차이점을 보이지 않고 있다. 그리고 탐색의 접근점이 되고 제한탐색이 가능한 필드는 “한국문헌목록정보”와 “국회도서관문헌정보” 공히 8개로 구성되어 있는데 앞에서 언급한 필드구성과 마찬가지로 특성을 보이고 있다. 이와 함께 이용자에게 검색된 레코드의 적절성 여부에 대한 판단의 근거가 되는 초록필드는 양자 공히 제공하지 않고 있다.

4. 3. 4 검색의 편의성에 대한 평가

검색의 편의성이란 이용자가 데이터베이스에 접속하고, 관련정보를 검색하며, 검색된 결과를 제공받는 S/W 인터페이스 및 유틸리티 기능의 편리성, 친절성, 이해가능성, 적합성 및 다양성 등을 망라하는 개념이라 할 수 있다. 검색의 편의성에 대한 평가기준으로는 <표 8>과 같이 10가지의 항목을 설정하였다. 물론 “한국문헌목록정보”는 DOS 버전인데 반하여, “국회도서관문헌정보”는 Window 버전으로 검색화면의 인터페이스 기능에는 차이가 있지만, 이용자의 측면에서 객관적인 조사를 행하였다. 정보검색에 있어서 가장 기본이 되는 검색대상자료의 선정, 검색결과의 디스플레이, MARC 데이터의 표시, 단축기의 활용, 검색결과와 저장 등은 “한국문헌목록정보”와 “국

<표 7> “한국문헌목록정보”와 “국회도서관문헌정보”의 완전성 평가 내용

DB명	불완전성/제한점의 언급	필드구성	접근점 필드/제한탐색	기 타
한국문헌 목록정보	<ul style="list-style-type: none"> • 주제어 검색은 1983년 이후 자료만 가능 • 1983년 이전 자료 : KDCP • 1984년 이후 자료 : KDC, DDC • ISBN 검색시 하이픈(-)없이 입력 · 발행년도는 다른 검색 요소와 조합하여 탐색 	자료구분 서명 서명단어 저자명 발행자 주제어 분류구분 분류기호 ISBN 발행년도	서명 저자명 발행자 주제어 ISBN 자료구분 분류기호 발행년도	초록 필드 없음
국회도서관 문헌정보	<ul style="list-style-type: none"> • 서명단어 검색은 서명단어가 서명 띄어쓰기에서 구분되어 있어야 함 • 키워드 검색은 키워드 색인어로 추출되어 있을 때만 가능 • ISBN은 하이픈(-)없이 입력 함 	검색범위 (해당년도, 본문언어) 서명 저자명 키워드 발행자 분류기호 ISBN	서명 저자명 키워드 발행자 분류기호 ISBN 해당년도 본문언어	초록 필드 없음

〈표 8〉 “한국문헌목록정보”와 “국회도서관문헌정보”의 검색의 편의성 평가 내용

평가 항목	한국문헌목록정보	국회도서관문헌정보	비 고
• 검색화면의 인터페이스	DOS	Window	
• 검색대상자료의 선정기능	Yes	Yes	
• 검색결과외 디스플레이 다양성	Yes	Yes	
• 색인의 열람기능	No	Yes	
• MARC Data 표시	Yes	Yes	
• 단축키	Yes	Yes	
• Mouse 이용	No	Yes	
• 언어별 검색	No	Yes	
• 검색결과내 찾기	No	Yes	
• 검색결과외 저장	Yes	Yes	

〈표 9〉 “한국문헌목록정보”와 “국회도서관문헌정보”의 검색기능의 다양성 평가 내용

항 목	한국문헌목록정보	국회도서관문헌정보	비 고
• 키워드(단어)검색	빈칸 채우기	빈칸 채우기	
• 데이터 필드(항목) 검색	Yes	Yes	
• 조건검색			“국회도서관문헌정보”는 조건검색에서 불연산자의 사용 개수에 제한이 있음(3개 이하). 그리고 우선순위지정 검색에 있어서도 첫 번째 검색어는 지정이 불가능하다.
- 불연산자	and, or, not	and, or	
- 인접연산자	No	No	
• 제한검색	Yes	Yes	
• 절단검색			
- 전방일치	No	No	
- 중간일치	No	No	
- 후방일치	Yes	Yes	
• 검색식 저장 검색	No	No	
• 조합검색	Yes	Yes	
• 우선순위지정 검색	Yes	Yes/No	

“국회도서관문헌정보” 양자가 제공하고 있다 그러나 전자의 경우, 검색에 있어 매우 중요한 색인의 열람기능이 제공되지 않고, 일반적으로 고도 보편적으로 이용이 되고 있는 마우스의 사용이 불가능하며, 특정언어를 지정하여 검색을 할 수 있는 언어별 검색기능, 그리고 일차로 탐색한 결과에서 이차적인 탐색을 수행할 수 있는 검색결과내에서의 탐색이 불가능한 것으로 밝혀졌다. 그리고 양자가 공히 제공하고 MARC data 표시에 있어서는 “한국문

헌목록정보”는 KORMARC 데이터의 편집 및 저장의 경우, MARC형태로의 변환이 용이한 반면, “국회도서관문헌정보”는 View만 가능하고 Text로만 저장이 되는 것으로 나타났다.

4. 3. 5 검색기능의 다양성에 대한 평가

검색기능의 다양성에 대한 평가기준은 〈표 9〉에 나타난 바와 같이 키워드(단어)검색, 데이터 필드(항목) 검색, 조건검색, 제한검색,

절단검색, 검색식 저장 검색, 조합검색, 우선순
 위지정 검색 등의 8가지 항목을 설정하였다.
 “한국문헌목록정보”와 “국회도서관문헌정보”
 는 조건검색(불연산자)과 우선순위지정 검색
 의 두 가지 항목 이외에는 별다른 차이를 보
 이지 않은 것으로 나타났다. 조건검색의 경우,
 전자는 통상의 세 가지 불연산자(and, or,
 not)가 검색어의 조합에 사용되는 반면에 후
 자는 두 가지(and, or) 불연산자의 사용이 가
 능함으로써 검색효율의 측면에서 보면, 정확
 률이 높은 검색을 원하는 경우에는 큰 문제점
 을 지니고 있다. 그리고 후자의 경우, 우선순
 위지정 검색에 있어서 첫 번째 검색의 경우에
 는 지정(예를 들면, ()의 사용)이 불가능하고,
 두 번째 이후부터의 검색어에 지정이 가능하
 기 때문에 이용자의 입장에서 보면 검색에 있
 어서 상당한 혼란을 일으키는 문제점을 가지
 고 있는 것으로 나타났다. 이와 함께 전자와
 후자 모두 조건검색의 경우, DIALOG과 같
 은 상용의 정보검색시스템에서 제공하는 인접
 연산자의 기능(with, adj, near 등)은 제공하
 지 않고 있고, 절단검색 역시 전방일치와 중간
 일치의 기능을 제공하지 않은 등 검색기능의
 다양성이라는 측면에서 보면 매우 초보적이고
 기본적인 기능만을 제공하고 있는 것으로 나
 타났다.

5. 결론 및 제언

본 연구에서는 “한국문헌목록정보”와 “국회
 도서관문헌정보” CD-ROM 데이터베이스의
 품질을 평가하기 위하여 다섯 가지의 측정기

준을 설정하여 조사하였다. 측정요소는 데이
 터 자체의 품질평가 요소인 ① 정확성, ② 일
 관성, ③ 완전성의 세 가지와 데이터베이스 서
 비스의 품질평가 요소인 ① 검색의 편의성, ②
 검색기능의 다양성의 두 가지를 설정하였다.
 본 연구의 결과를 요약하면 다음과 같다.

데이터 자체의 품질평가의 기준으로 설정한
 정확성, 일관성, 완전성에 있어서 “한국문헌목
 록정보”와 “국회도서관문헌정보” 양자 모두
 전반적으로 심각한 문제점을 지니고 있는 것
 으로 밝혀졌다. 정확성 평가의 경우, “한국문
 헌목록정보”와 “국회도서관문헌정보” 양자 모
 두 외래어 표기의 부정확성, 주제어 부출시의
 통일성 결여, 번역서인 경우의 원서명과의 불
 일치 문제 등으로 인하여 필드별 “OR 연산자
 적용건 수”에서 차이를 보이고 있는데, 특히
 후자에 있어서 이와 같은 현상이 많이 발생하
 는 것으로 나타났다.

일관성 평가의 경우, “한국문헌목록정보”와
 “국회도서관문헌정보” 양자 모두 상호참조의
 기능을 제공하지 않고 있으며, 약어처리 또한
 공통의 문제점을 지니고 있으며, 이 역시 특
 히, 띄어쓰기의 경우에는 후자가 심각한 문제
 점을 가지고 있는 것으로 나타났다.

완전성의 평가의 경우, “한국문헌목록정보”
 와 “국회도서관문헌정보” 양자 모두 불완전성
 이나 제한점에 대한 언급이 부족하였고, 발생
 가능한 오류처리에 대한 사전 언급은 없는 것
 으로 나타났다.

DB 서비스의 품질평가 기준으로 설정한
 검색의 편의성 및 검색기능의 다양성에 대해
 서는 “한국문헌목록정보”와 “국회도서관문헌
 정보” 양자 모두 만족할 만한 수준을 보이지

않고 있는 것으로 나타났다. 전자의 경우, 검색의 편의성 측면에서 색인의 열람기능, 언어별 검색, 검색결과내 찾기 등의 기본적인 기능이 제공되지 않고 있으며, 후자의 경우에는 검색기능의 다양성 측면에서 불연산자의 사용과 우선순위지정 검색에 문제점이 있는 것으로 나타났다, 양자 모두 조건검색의 경우, DIALOG, OVID와 같은 상용의 정보검색시스템에서 제공하는 인접연산자의 기능과 절단 검색에 있어서의 전방일치, 중간일치의 기능을 제공하지 않은 등 기본적인 검색기능만을 제공하고 있는 것으로 나타났다.

이와 같이 우리 나라 국가도서관의 역할을 수행하고 있는 국립중앙도서관의 “한국문헌목록정보”와 국회도서관의 “국회도서관문헌정보” CD-ROM 데이터베이스는 축적된 레코드의 양적인 면에 비하여 품질적인 측면에서는 아직까지 보완되고 수정되어야 할 요소가 많은 것으로 나타났다. 따라서 향후, 데이터베이스의 품질 향상을 통한 이용자의 만족도를 제고시키기 위해서는 아래와 같은 사항들이 고려될 수 있을 것이다.

첫째, 데이터베이스 관련 정책의 시급한 전환이 필요하다. 즉, “규모의 확대와 양적인 증가 및 사용자층의 확산”에서 “데이터베이스의

품질향상을 통한 질적인 발전과 사용자층의 만족도를 높이는 정책”에 초점을 맞추어야 할 것이다.

둘째, 데이터베이스 수록정보의 일관성을 유지하기 위한 전거통제의 기능을 수행할 수 있는 시소러스의 개발에 대한 지속적이고도 활발한 투자가 있어야 할 것이다.

셋째, 데이터베이스의 품질을 평가할 수 있는 공식적인 기준을 마련하여 기존에 생산된 데이터베이스의 품질수준을 평가하여 높은 등급을 받은 데이터베이스에 대해서는 인센티브를 부여할 수 있는 정책적인 배려가 있어야 하며, 신규 제작되는 데이터베이스에 대해서는 일정 수준 이상의 품질요건을 갖추어야만 유통될 수 있는 공식적인 제도를 마련해야 할 것이다.

넷째, 데이터베이스의 품질을 객관적으로 평가하여 품질등급을 공인할 수 있는 국가적 차원의 법제도적인 뒷받침이 있는 공식적인 기구의 설치/운영이 시급하다. 이 기구에서는 기존에 구축된 데이터베이스의 품질을 평가하여 높은 등급을 부여받은 데이터베이스에 대해서는 인센티브를 부여함으로써, 신규로 제작 유통되는 데이터베이스의 품질향상을 도모할 수 있을 것이다.

참 고 문 헌

유사라. 1999. 데이터베이스 정보 품질 평가의 메타분석. 정보관리학회지 16(1): 157-174
이응봉. 1996. 데이터베이스 품질에 관한 비평

적 평가. 『국회도서관보』 33(4): 13-28
이제환. 1996. “이용자의 관점에서 본 KRISTAL System과 DB의 품질”. KOSTI Proceedings, 1: 240-267.

- 한국데이터베이스진흥센터. 1996. 『데이터베이스 표준화연구보고서』 서울: 한국데이터베이스진흥센터.
- _____. 1998. 『1998 데이터베이스 백서』 서울: 한국데이터베이스진흥센터.
- 한국과학기술연구원/연구개발정보센터. 1996. 『DB 선정 및 품질평가 기준에 관한 연구』 과천: 과학기술처
- Basch, R. 1990. "Measuring the Quality of the Data: report on the Fourth Annual ACOUG retreat." *Database Searcher* 6(8): 18-23.
- Bosch, V. M. 1995. "CDROM User Interface Evaluation: The Appropriateness of GUIs." *Online & CDROM Review* 19(5): 255-270.
- Brodie, M. 1980. "Data Quality in Information Systems." *Information and Management* 13(1): 1-12.
- Daniel, E. 1993. "Quality Control of Documents." *Library Trends*, 41(4): 64-664
- Granick, Lois 1991. "Assuring the Quality of Information Dissemination: Responsibilities of Database Producers", *Information Services and Use*, 11:117-136.
- Harry, V. & Oppenheim, C. 1993. "Evaluations of Electronic Databases: Criteria for Testing CD-ROM Products. Part 1 & 2." *Online & CD-ROM Review*, 17(4): 211-222 & 17(6): 339-368.
- Jasco, P. 1992. *CD-ROM Software, Dareware and Hardware: Evaluation, Selection, and Installation*. Englewood, Colorado: Libraries Unlimited, Inc.
- O'Neill, E. T. & Vinzine-Goetz. 1988. "Quality Control in Online Database." *Annual Review of Information Science and Technology*, 23: 125-156.
- Rowley J. & Slack, F. 1997. "The Evaluation of Interface Design on CDROMs." *Online & CD-ROM Review* 21(1): 3-14.
- Tenopir, C. 1987. "Online Database: Quality Control." *Library Journal* 112(3): 34-51
- Valas, G. 1994. "Comparison of Some Widespread CDROM Information Retrieval Software Packages." *Online & CDROM Review* 18(4): 211-227.