

과학기술분야 서지 DB의 품질관리 및 평가 방안: KORDIC의 KRISTAL DB를 중심으로

Methods for Quality Control and Evaluation in the Scientific and Technical Bibliographic Databases

이 제 환(Jae-Whoan Lee)

목 차

- | | |
|---------------------------|-----------------------------|
| 1. 연구를 시작하며 | 3. KRISTAL DB의 품질에 대한 평가 |
| 2. 이용자가 보는 KRISTAL System | 3. 1 DB의 품질평가를 위한 기준 |
| 2. 1 시스템에 대한 인지도 | 3. 2 KRISTAL DB의 품질 (이용자관점) |
| 2. 2 시스템이 제공하는 정보의 내용 | 3. 3 KRISTAL DB의 품질 (검색실험) |
| 2. 3 시스템이 제공하는 정보서비스 | 4. KRISTAL DB의 품질개선을 위한 제언 |

초 록

이 연구는 국내 과학기술계에서 구축하고 있는 서지 DB의 품질 문제를 다루고 있다. 구체적으로, 과학기술분야의 중추적인 정보유통기관인 KORDIC이 1990년대 초부터 구축해 오고 있는 KRISTAL DB 중에서 핵심이라 할 수 있는 UNION DB와 SATURN DB를 선정하여 그 품질을 평가하고 개선방안을 제시한다. 특히, UNION DB와 SATURN DB의 품질 평가를 위한 접근 방식에서, 기존의 시스템개발자 중심의 논리를 배격하고 이용자 중심의 평가 기준을 제시함으로써 '정보시스템의 설계 및 평가'에 있어서 이용자의 중요성을 강조한다.

ABSTRACT

This study discusses the quality issue of large scientific and technical (S&T) bibliographic databases in South Korea. In details, this study develops the criteria to evaluate the quality of S&T bibliographic databases, evaluates the quality of the selected two databases - UNION DB and SATURN DB of the KORDIC, and finally, suggests both organizational and technical methods for the quality improvement of such bibliographic databases.

* 이 논문은 1996년도 학술진흥재단의 자유공모과제 연구비에 의하여 지원되었음.

** 부산대학교 문헌정보학과 조교수

접수일자 1997년 8월 14일

1. 연구를 시작하며

1.1 연구배경

과학기술 정보시스템, 특히, 온라인 정보시스템의 구축에 있어서 핵심이 되는 것은, 구축된 시스템을 통해 유통될 데이터의 집적체, 즉 DB의 구축이다. 과학기술 분야의 연구특성상, 이 분야의 DB는 주제면에서 전문적이고 수록자료의 범위가 포괄적이며 그 내용면에서 최신성이 요구되는 등, DB에 포함되는 레코드의 량과 질이 우수할 것을 필요충분조건으로 요구한다. 현재 국내에서는, KINITY (산업기술정보원)을 중심으로한 산업기술정보시스템이 구축되어 관련 기관이나 인력들에게 정보서비스를 제공하고 있고, 다른 한편에서는, KORDIC (연구개발정보센터)가 중심이 되어 과학기술분야의 연구자들을 이용대상으로 하는 과학기술정보시스템의 구축을 위해 노력하고 있다. 특히 후자인 KORDIC의 경우, 자체적으로 축적한 기술적 노하우를 바탕으로 온라인 정보시스템인 KRISTAL 시스템을 개발하여 운영하고 있다. KRISTAL 시스템은 KORDIC이 독자적으로 개발한 검색엔진인 KRISTAL 소프트웨어와 KIST를 비롯한 정부출연연구기관들과의 협력하에 구축중인 과학기술분야 종합 데이터베이스인 KRISTAL DB로 구성된다. KRISTAL 시스템은 1996년 8월 현재, 그 검색엔진의 성능개선 작업과 종합 DB의 구축작업이 일정단계

에 이르면서 과기처산하 연구기관들에 소속된 연구인력들 뿐만 아니라 학계와 민간연구소에 속해 있는 연구자들에게도 개방되어, KORDIC이 국가단위의 과학기술 정보유통시스템으로서의 역할을 수행하는데 있어 주요 도구가 되고 있다.

그러나 과학기술정보의 조직, 축적, 탐색, 그리고 검색을 위한 도구로서의 KRISTAL 시스템의 효율성은, 이 시스템의 주 이용자그룹으로 설정된 과학기술분야 정부출연연구소의 연구원들에 의해서 조차 의문시되고 있다. KRISTAL 시스템의 효율성과 관련한 문제점은, 그 운영시스템인 KRISTAL 소프트웨어의 개발과정과 종합 DB인 KRISTAL DB의 구축과정에서 이미 내포되어 있었던 것으로 보인다. 즉, KRISTAL 소프트웨어의 개발과정에서 봉사대상으로 삼은 이용자그룹에 대한 선행연구가 결여되어, 이용자중심이라기 보다는 시스템의 설계자 혹은 개발자 중심의 검색엔진이 개발된 것이 아닌가 하는 의문이 들며, 또한, 시스템의 핵심이 되는 주요 DB의 구축에 있어서도 KORDIC은 유통을 위한 기술적 인프라만을 제공할 뿐 참여 연구기관들이 할당된 주제분야의 자료에 대한 서지레코드를 제작하여 통합 DB에 입력하는 분산시스템의 형태를 띠다 보니, 구축 중에 있는 통합 DB의 품질관리가 주요한 문제로 부각되고 있다.

특히 후자와 관련하여, KRISTAL DB 중에서 핵심이 되는 UNION DB와 SATURN DB의 구축에 참여하고 있는

대부분의 연구기관들이 입력 데이터의 표준성을 결여한 채 각 기관이 보유하고 있는 자료들에 대한 서지정보를 質보다는 量위주로 가공하고, 그것도 빈약한 전문인력에 의존하여 단시일내에 서지레코드를 제작하여 입력하는 작업을 DB 구축 초기부터 지속적으로 추진해 오고 있다. 그 결과는, 입력 데이터의 표준성과 완전성의 결여, 그리고 DB에 포함된 서지레코드의 중복률이 높게 나타나는 등, KRISTAL DB의 전체적인 품질관리 및 유지와 관련하여 심각한 문제점으로 나타나고 있다. DB에 수록되는 데이터의 품질관리 및 유지를 위한 대책이 미흡한 상태에서 아무리 방대한 양의 자료를 DB로 구축하였다 하더라도, 실제로 관련분야의 연구자들에게 얼마나 유용한 정보소스로 이용될는지 의문이다.

DB의 개발은 DB를 궁극적으로 이용하게 될 사용자그룹의 정보요구 및 이용행태에 대한 포괄적이고 체계적인 시스템연구의 결과에 의존해야 한다. 특히, 데이터의 품질을 검사하고, 품질관리 및 유지를 위한 평가방안을 마련하는 등, 공급하는 데이터의 質면에서 사용자들의 정보요구를 만족시키려는 노력을 기울여 한다면, DB의 구축에 들인 노력과 비용은 이용자의 외면속에 회수되기 어려울 것이기 때문이다. 一 例로, 세계 최대의 문헌정보망인 OCLC의 창립자 Kilgour는 일찍이 'OCLC 데이터베이스의 품질저하는 OCLC 데이터베이스의 관리자들에 물론 이용자들에게 있어 최대의 적'으로 등장

할 것임을 예언한 바 있다 (Boissonnas, 1979). 歷史는 Kilgour의 예언이 옳았음을 보여준다. 그의 예언대로, OCLC는 초기에 양적 측면에서의 DB의 성장에 몰두한 나머지 DB에 포함된 데이터의 질적 측면에 대한 고려를 게을리 하였고, 그 결과, DB의 양적 방대함에도 불구하고 근래에 이르기까지 전문 이용자들로부터 '쓰레기 정보의 저장소'라는 오명을 받아 왔다 (Davis, 1989). DB의 품질문제를 다룬 대부분의 선행연구들은, DB에 포함된 데이터의 품질이 열악할 경우 예산과 노력의 낭비는 물론이고 그 정보시스템을 이용하는 연구자들의 연구생산성이 저하될 수 있음을 경고하고 있다 (Redman, 1992 & 1995). 실제로, 북미의 4대 문헌정보망인 OCLC, RLIN, WLN, 그리고 UTLAS의 경우, 중앙 DB에 수록된 수준 미달 혹은 중복 레코드들로 인한 품질저하는, 이들 정보시스템의 관리자뿐 아니라 이용자들에게 심각한 고민거리로 남아 있다 (Mandel, 1986; Davis, 1989:).

1. 2 연구목적

이와 같은 배경下에, 이 연구의 목적은 국가단위의 과학기술정보의 유통을 촉진하기 위한 목적에서 설립된 KORDIC이 자체적으로 개발하여 운영중인 KRISTAL 시스템을 대상으로 하여, 그 핵심이 되는 KRISTAL DB의 품질을 분석하는 것을 통하여 국가단위의 과학기술 정보시스템

의 구축 및 관리를 위해 실무자들이 참고할 수 있는 종합적인 제언을 하고자 하는데 있다.

구체적으로, 이 연구는: (1) KRISTAL 시스템의 핵심이 되는 KRISTAL DB의 품질을 평가하기 위한 기준을 개발하고; (2) 개발한 평가기준에 의거, KRISTAL DB의 품질을 분석하여; (3) KRISTAL DB와 DB를 이용한 정보서비스의 품질을 개선하기 위한 방안을 제시하는데 그 초점이 주어진다. 특히, KRISTAL DB의 품질평가를 위한 접근방식에서 기존의 시스템개발자 중심의 논리를 배격하고 이용자 중심의 평가기준을 마련함으로써, 궁극적으로는 과학기술분야 정보시스템의 질적 발전을 도모하고 시스템의 이용효율을 증대시키고자 하는데 이 연구의 목적이 있다.

1. 3 연구방법

이 연구를 위해 필요한 데이터의 수집은 다음 방법에 의거하였다. 먼저, KORDIC이 정보서비스의 주 대상으로 삼고 있는 정부출연연구소소속 연구원들과의 집중 인터뷰 (focus interview)를 실시하였다. 구체적으로, KIST를 비롯한 12

개 정부출연연구소 소속 연구원을 대상으로 하되, 각 연구소의 정보관리부서 (기술정보실, 정보자료실 등)에서 추천한 4-5명의 연구원을 표본으로 선정하여, 모두 50명의 연구원들을 대상으로 인터뷰가 진행되었다.¹⁾ 인터뷰는 이 연구를 위해 특별히 제작된 인터뷰 스케줄을 이용하여 60-90분 가량 진행되었으며, 1차 인터뷰의 내용에 대한 분석과정에서 데이터의 보충이 필요하다고 판단된 11명을 대상으로 2차 인터뷰가 실시되었다.

한편, KRISTAL DB의 품질 평가를 위해 필요한 데이터는 다음 두 가지 측면의 분석을 염두에 두고 수집되었다. 첫째, 이용자의 경험에 바탕한 종합적 평가를 위해서는 앞서 언급한 인터뷰 과정에서 수집된 데이터가 이용되었고; 둘째, 본 연구팀이 개발한 품질 평가기준에 의한 KRISTAL DB의 품질 측정을 위한 데이터는, 본 연구팀의 직접적인 검색실험을 통하여 수집되었다. 검색실험의 대상은 KRISTAL DB를 구성하는 여러 專門 DB들 중 이용자그룹에 의한 인지도가 비교적 높은 UNION DB (과학기술 단행본 DB)와 SATURN DB (국내제작 과학기술 문헌 DB)를 선정하였으며, UNION DB에서 100개의 레코드를, SATURN

1) Focus interview에 참여한 50명의 연구원중에, KAIST 석/박사과정에 재학중인 6명을 제외한 11개 연구소소속 연구원 44명의 배경 (즉, 연령, 직급, 최종학위, 연구경력, 연구 분야 등)에 대한 데이터는 다음과 같다: (a) 연령: 평균연령이 37.4세로 30대가 26명, 40 대가 18명; (b) 직급: 책임연구원급이 14명, 선임연구원급이 30명; (c) 최종학위: 박사학 위 소지자가 35명, 석사 학위 소지자가 9명; (d) 연구경력: 현 연구소에서의 연구경력은 평균 11.2년; (e) 연구분야: 기초연구분야가 12명, 응용연구 분야가 32명; (f) 소속연구소: 한국과학기술원, 한국과학기술연구원, 한국기계연구원 (대전), 한국에너지기술연구소, 한 국전기 연구소, 한국원자력연구소, 한국해양연구소, 한국화학연구소, 생명공학연구소, 한국자원연구소, 한국항공우주연구소, 한국표준과학연구원.

DB에서 600개의 레코드 (12개 SATURN DB 구축 참여 연구기관별로 50개씩)를 출력하여 분석하였다.

2. 이용자가 보는 KRISTAL 시스템

인터뷰를 통해 수집한 이용자그룹의 정보추구행태에 관한 데이터를 기초로 하여 KRISTAL 시스템의 현재 기능을 분석해 본 결과, 여러 문제점이 있는 것으로 파악되었다. 특히, 시스템에 대한 인지도, 시스템이 제공하는 정보의 내용, 그리고 시스템이 제공하는 서비스의 효용성과 관련한 문제가 가장 심각한 것으로 나타났다.

2.1 시스템에 대한 인지도

가장 먼저 부각되는 문제점은 KRISTAL 시스템에 대한 이용자그룹의 인지도가 지극히 낮다는 사실이었다. 인터뷰에 참여한 12개 정부출연연구소 소속 50명의 연구원들 사이에서 KRISTAL 시스템에 대한 인지도는 매우 낮았다. "KRISTAL 시스템에 대해서 전혀 들어본 적이 없었다"는 연구원이 모두 15명으로 전체의 30%를 차지하였는데, 이 사실은 인터뷰를 위한 대상의 선정이 해당 연구기관의 정보관련부서의 추천에 의해 이루어졌다는 점을 고려할 때 상당히 의외의 결과로 판단되었다. 더욱이, 이들 50명이 KRISTAL DB의 품질평가를 위해 각

기관에서 선정한 모니터 요원이라는 점을 고려할 때, 전체 50명 중 15명에게서 "본 인터뷰를 앞둔 최근 3-4일 전까지는 KRISTAL 시스템에 대해서 들어본 적이 없었다"는 솔직한 고백은 충격적이었다.

이에, 전체 연구원들을 대상으로 한 인지도 조사가 다시 필요하다는 본 연구팀의 판단에 따라, 12개 연구소 소속 연구원들 중에서 무작위로 추출한 50명을 대상으로 전화인터뷰가 행해졌다. 그 결과는, 예상했던대로, 50명중에 단 6명만이 "KRISTAL 시스템에 대해서 잘 알고 있다"는 응답을 함으로써, KORDIC이 정보서비스를 제공하기 위해 주 이용자그룹으로 선정했던 연구원들 사이에서 조차 KRISTAL 시스템에 대한 인지도는 지극히 낮은 것으로 나타났다.

KRISTAL 시스템의 실질적 이용도는 더욱 낮았다. "KRISTAL 시스템에 대해서 알고 있다"고 응답한 35명중에서, "지난 1달 동안 전혀 사용해 본 적이 없다"고 응답한 연구원이 6명이나 되었다. 다시, "사용해 본 적이 있다"고 응답한 29명을 대상으로 "지난 1달 동안 평균 몇 회 정도 이용하였느냐"는 질의에 대해서는, 평균 1.6회 정도 이용한다는 통계가 나와, 평균 20일에 한 번 꼴로 KRISTAL 시스템에 접속하는 것으로 나타났다. 이 통계는 KRISTAL 시스템이 정부출연연구소 소속 연구원들에게 있어 주 정보시스템으로서의 역할을 거의 수행하지 못하고 있음을 암시한다. 특히, 인터뷰 대상 연구원들 사이에 DIALOG나 STN Inter-

national 같은 해외 DB 벤더나 KINITI 같은 국내 DB 벤더에 대한 인지도와 이용도가 비교적 높게 나타났던 점을 고려할 때, KRISTAL 시스템에 대한 인지도와 이용도가 지극히 낮다는 사실은 KORDIC이 과학기술분야의 중심적인 정보유통기관으로서의 역할을 수행하기 위해서 시급히 극복해야할 문제점으로 지적된다.

무엇보다도 현재와 같이 낮은 인지도와 이용률은 시스템자체의 존립근거를 약하게 만들 가능성이 있다. 정보시스템의 기능 중에 public relations 기능의 중요성은 아무리 강조하여도 지나침이 없다. 시스템의 존재를 알리고 그 기능을 홍보하는 과정을 통해 보다 많은 고정 사용자그룹이 생겨나고, 다시, 이들로부터 시스템 개선을 위한 조언이 피드백 형식으로 지속적으로 유입될 때, 시스템은 존재 이유를 입증하는 단계를 넘어 확대 발전을 위한 근거를 얻게 된다. KRISTAL 시스템의 경우, 시스템의 기능과 서비스에 대한 체계적인 홍보활동을 위해 KORDIC이 시급히 개선책을 마련할 필요가 있다. 특히, 조직구조상의 개편을 통하여 PR 관련 부서나 전담인력을 배치하고, 이 부서나 인력이 KRISTAL 시스템에 대한 홍보자료의 발간, 시스템 소개를 위한 다양한 회의 참석, 그리고 이용자교육을 위한 단위 연구기관의 방문 등, PR 관련 업무를 전담하게 하는 것이 바람직하다. 나아가, KORDIC 기관차원에서는, 기관소식지의 발간, 전문저널 (trade or scholarly journal)의 발간, 전문교육과정의 설치, 혹은 학술회의의 개최

등을 통하여, 기관자체에 대한 홍보에 나설 필요성이 절실히 보인다. 현단계에서 기관과 시스템에 대한 홍보가 체계적으로 이루어지지 않는다면, 국가의 주요 정책이 조령모개식으로 바뀌는 우리의 풍토에서 KORDIC의 존립근거마저 흔들릴지 모른다.

2. 2 시스템이 제공하는 정보의 내용

다음으로 문제가 되는 것이 정보시스템으로서 KRISTAL 시스템이 제공하는 정보의 내용이다. KRISTAL 시스템이 제공하고 있는 정보는, 그 내용의 포괄성이나 적합성과 관련하여 이용자들로부터 그리 좋은 평을 받지 못하고 있는 것으로 나타났다. 이는 역설적으로 KRISTAL 시스템이 이용자들의 요구에 적합한 정보를 제공하는데 실패하고 있음을 의미한다.

KRISTAL 시스템이 대상으로 하고 있는 이용자그룹은 매우 다양한 소그룹으로 이루어져 있다. 이들은, 소속연구기관에 따라, 관심 연구영역에 따라, 연구기관에서의 지위와 역할에 따라, 수행중인 프로젝트의 성격에 따라, 그리고 프로젝트에서의 역할에 따라, 매우 다양한 정보요구를 표출하고 있다. 이들의 다양한 정보요구에 대처하는데 KRISTAL 시스템이 현재 제공하고 있는 정보서비스는 적합한가? 이 질의에 대한 대답은 부정적이며, 이는 앞서 논의하였듯이 KRISTAL 시스템의 실제 이용률이 지극히 낮다는 사실로 입증된다. KRISTAL 시스템의 이용률이 낮은

이유는 여러 가지로 설명될 수 있다. 그러나 그 근본 이유는, 봉사대상으로 설정한 이용자그룹이 절실히 필요로 하는 정보가 무엇인지에 대한 총체적인 분석이 선행되지 않은 채 시스템구축이 이루어진데서 찾는 것이 타당해 보인다. 다시 말해, 시장조사도 없이 기계를 들여놓고 제품부터 만들어 놓은 제조공장의 상황과 유사하다. 이처럼, 이용자들로부터의 외면은 예견되어진 결과일지도 모른다. 본 연구에 참여한 연구원들과의 인터뷰에서도 나타났듯이, 연구자들이 가장 필요로 하는 서비스 중의 하나는 ‘관련 연구분야에서 현재 진행 중에 있는 연구들에 대한 체계적인 정보의 제공’이다. 그렇다면 과연 KRISTAL 시스템은 연구자들의 이와 같은 정보요구를 충족시켜주고 있는가?

연구원들은 말한다: “정보에 대한 학제적인 접근, 공동연구를 위한 인력 정보, 자료의 소재지 및 입수방법에 대한 정보, 문헌에 대한 간접 정보보다는 심도있게 분석 요약된 내용이 자신들의 연구활동을 위해 절실히 필요하다.” 연구원들의 정보요구는 이어진다: “관심분야에서 생산되고 있는 주요 연구물들을 총체적으로 파악할 수 있는 이차자료가 필요하다” 혹은 “국내외에서 개최되는 관련 회의에 대한 체계적인 정보가 필요하다.” 이상과 같은 이용자들의 정보요구 중에 과연 몇 가지나 현재의 KRISTAL 시스템은 충족시켜주고 있는가? KRISTAL 시스템을 개발하고 관리하는 인력들이 관심을 갖고 고민해야만 하는 질문이다.

2. 3 시스템이 제공하는 정보서비스의 효용성

또 하나의 문제점은, 이용자들이 궁극적으로 요구하는 것은 ‘실질적인 데이터나 그러한 데이터를 포함하고 있는 정보자료(원문)의 입수’라는 사실과 관련한다. 정보자료의 원문 혹은 분석/편집/추출된 데이터의 제공 기능을 KRISTAL 시스템은 갖추고 있는가? 후자의 기능은 고단계 정보기술의 발전과 전문인력의 확보라는 문제와 맞물려있어 논외로 치더라도, 원문의 제공기능은 개선의 여지를 남겨두고 있다.

가령, 본 연구를 위한 인터뷰에 응했던 연구원들은 비용에 대한 부담을 감수하면서까지 국내외의 데이터뱅크나 원문서비스기관의 원문제공서비스를 이용하고 있는 것으로 조사되었다. 이 문제는, 소속 연구기관의 정보자료실이 소장하고 있는 장서를 이용하여 원문을 입수하는 비율이 28%에 머물고 있는 반면, 상업적 외부기관을 이용하여 원문을 입수하는 비율이 30%에 이르고 있는 현실과 맞물려, KRISTAL 시스템이 과학기술분야의 연구자들에 의해 활발히 이용되는 정보시스템으로서 거듭나기 위해 앞으로 지향해야 할 방향을 제시한다. 뿐만 아니라, 대부분의 연구자들은 ‘필요하지만 입수하지 못한 정보자료’의 경우, ‘안타깝지만 시간적 비용적 이유로 인해 포기’하는 행태를 보이고 있어, 결과적으로, 관련 정보의 입수의 어려움이 이들에 의해 생산되는 연구물의 질적 저하를 유발하는 원인이 되고

있는 것으로 나타난다.

이와 같은 상황은, 과학기술분야의 종합 정보시스템으로서 KRISTAL 시스템이 무엇을 우선적으로 추구해야 할지를 명확히 보여준다. 가령, (a) 이용자의 정보요구를 충족시키는데 필요한 적절한 정보자료를 자체적으로 개발 수집하거나 관련 연구기관에서 이루어지고 있는 개발 수집 활동을 효율적으로 지원하는 일; (b) 더불어, 수집된 정보자료에 대한 체계적인 분석 및 조직작업을 감독하고 지원하는 일 (이 작업은 특히 후에 논의될 KRISTAL DB의 품질에 직접 영향을 미친다); (c) 나아가, 원문을 수록한 全文 DB를 자체적으로 개발하거나 개발을 지원하는 것을 통하여 관련 분야의 연구자들이 궁극적으로 원하는 것을 제공하는 일 등은, KRISTAL 시스템이 그 존재 가치를 대내외적으로 밝히고 지속적인 발전을 위한 외부로부터의 투자를 유치하기 위해서 우선적으로 추구해야 할 작업들이 될 것이다.

2. 4 종합적 견해

이용자그룹과의 인터뷰의 결과는 그들이 어떤 정보시스템을 선호하는지를 명확히 보여준다. 그 결과는 KRISTAL 시스템이 시스템 측면의 개선을 위해 추구해야 할 기준을 제시하고 있는 바, 여기서 인용하면서 구체적으로 어떠한 점에서 개선이 필요한 지에 대해 논해보기로 한다.

먼저, 연구원들의 정보탐색방법의 선택

에 있어 영향을 미치는 중요한 요소로는: 얻을 수 있는 데이터의 질 (quality - accuracy & reliability of data), 얻을 수 있는 데이터의 포괄성 (coverage of data), 탐색 도구에 대한 접근성 (accessibility of search tool), 탐색방법의 편이성 (easiness in the use of search method), 얻을 수 있는 데이터의 최신성 (currentness of data), 탐색하는데 드는 비용 (cost in money) 등이 제시되었다. 한편, 정보입수방법의 선택에 영향을 미치는 요소들로는: 원문 혹은 실질적인 데이터의 입수까지 걸리는 시간 (time), 입수 절차의 간편함 (easiness), 입수하는데 드는 비용 (cost) 등이 거론되었다.

이상의 요소들을 시스템 관점에서 재분류하면, (a) 시스템을 이용하는데 드는 비용, (b) 시스템에 대한 접근 및 이용의 편이성, (c) 검색엔진의 능력, 그리고 (d) 시스템이 제공하는 정보의 품질 등으로 요약되는데, 이 네 가지 기준이 이용자들이 정보시스템을 선택하는 데 있어서 주요한 변수가 된다. 특히, 이 중에서, 비용적 측면을 제외한 나머지 세 요소와 관련한 문제들은, KRISTAL 시스템이 과학기술분야의 종합정보시스템으로서의 역할을 수행하기 위해서는 개선의 여지가 많은 부분으로 지적된다. 이에, 다음 장에서는 이들 세 요소와 관련한 문제점들을 KRISTAL DB의 품질을 분석하는 과정을 통해 구체적으로 논의하고, 나아가, 품질 개선을 위한 구체적인 방안을 제시하고자 한다.

3. KRISTAL DB의 품질에 대한 평가

3.1 DB의 품질평가를 위한 기준

DB의 품질은 크게 세 가지 관점에서 평가될 수 있다: 상대적 품질 (relative quality), 절대적 품질 (absolute quality), 그리고 목적 부합 측면에서의 품질 (quality as fitness for purpose). 먼저, <상대적 품질>의 개념은 여러 DB의 품질을 서로 비교하기 위한 목적에서 주로 사용되며, 다음 <절대적 품질>의 개념은 개별 DB의 품질을 절대적인 기준에 의거하여 평가할 때 주로 사용된다. 이 두 개념의 경우, DB에 포함되어 있는 데이터의 정확성 (accuracy), 일관성 (consistency), 신뢰도 (reliability), 권위도 (authoritativeness)와 같은 단위 레코드의 완전성이나 DB내에 중복 수록된 레코드의 비율 등이 품질평가를 위한 주요 기준이 된다. 그러나 세 번째 관점, <목적 부합 측면에서의 품질>의 개념은 이들과는 다르다. 이 세 번째 개념에서 DB의 품질은 해당 DB가 이용자의 특수한 정보요구를 어느 정도 만족시키고 있는가, 즉 해당 DB에 대한 이용자의 만족도 (혹은 정보요구의 충족도)를 기준으로 평가되는데, 구체적으로 DB가 검색결과로 제공하는 '해결책 (answer)'의 품질이 평가를 위한 주요 기준이 된다 (O'Neill & Vizine-Goetz, 1988).

이 분야에서 발표된 연구문헌들을 분석해 보면, 대부분의 연구가 DB에 포함된 레코드의 완전성이나 중복 수록된 레코드의 비율을 DB의 품질평가를 위한 주요 기준으로 삼아 왔음이 드러난다. 즉, 많은 연구가 <절대적 품질>을 측정하고, 절대적 품질의 개선을 위한 알고리즘의 개발에 치중해 온 것으로 나타난다 (Wang et al, 1995). 이와 같은 연구의 결과로, DB를 구성하는 수준미달 혹은 중복 레코드의 비율을 줄이기 위한 여러 방안들이 개발되었으며, 특히, 데이터 요소에 있어서의 자동에러탐지 및 수정과 관련하여서는 대폭적인 기술 진전이 이루어져 왔다. 그러나, 서지레코드를 구성하는 데이터 요소들에서 발생하는 에러, 특히, 철자상의 에러에 대한 탐지와 수정 기능의 개선만으로 DB의 전체적인 품질이 향상되기를 기대하는 것에는 무리가 있다 (Strong et al, 1994; Wang et al, 1994).

정보시스템의 개발과 시스템의 핵심이 되는 DB의 구축은, 관련 정보의 원활한 유통을 통하여 정보의 잠재적 이용자들의 정보요구를 충족시켜 줌으로써, 궁극적으로, 이들의 생산성을 향상시키는 데 그 기본 목적이 있다 (Melone, 1990). 따라서, 구축된 혹은 구축중인 DB의 평가에 있어 <절대적 품질>의 개념 못지 않게 중요한 것이 <목적 부합 측면에서의 품질>의 개념이다. 즉, DB에 포함된 자료의 적합성이나 이용자 요구의 충족도 등과 같은 이용자 측면에서의 품질평가를 위한 연구의 중요성은 아무리 강조하여도 지나침이 없

다. 그럼에도 불구하고, 국내의 경우 이 분야에 대한 연구의 진척이 활발해 보이지 않는다. 1990년대로 접어들면서 중앙정부의 '정보화' 정책에 힘입어 여러 분야에서 다양한 DB가 구축 중에 있으나, 양적 팽창에 몰두한 나머지 질적 평가와 관리를 위한 기초 연구가 거의 이루어지지 않고 있는 것이 현재의 실정이다.

정보검색을 위한 소프트웨어의 개발이 일정 수준에 이른다 할지라도, 검색대상이 되는 자료의 집합체인 DB의 '내용면'에서의 품질이 열악하다면, 정보시스템 전체의 효율성과 활용도가 저하될 것은 자명한 사실이다. 정보선진국이라 할 수 있는 미국과 일본 그리고 유럽 여러 나라의 사례를 보면, DB의 구축을 위해서는 포함될 데이터나 정보자료의 선정에서부터 레코드의 구조설계에 이르기까지 잠재적 이용자그룹에 대한 종합적인 평가가 반드시 기초자료로서 반영이 되고 있다. 이러한 점을 고려할 때, 국내에서 개발 중에 있는 과학기술 관련 DB의 품질관리 및 유지를 위해서는, 앞서의 <절대적 품질>과 <목적부합 측면에서의 품질>을 동시에 커버하는 종합적이고 체계적인 평가 기준의 마련이 필요하다.

이와 같은 배경하에, 이 연구에서는 이용자의 관점에서 DB의 품질을 객관적으로 평가하기 위한 기준을 다음 세 가지 항목에 초점을 두고 개발하였다. 첫 번째 항목은 DB에 포함되어 있는 정보자료의 품질을 평가하기 위해; 두 번째 항목은 DB에 포함되어 있는 단위 레코드의 품질

을 평가하기 위해; 세 번째 항목은 DB의 검색과 관련하여 제공되는 서비스의 품질을 평가하기 위해.

3. 1. 1 DB에 포함되어 있는 정보자료의 품질에 대한 평가 기준

이 항목은 하나의 DB가 서비스 대상으로 설정한 이용자들의 정보요구에 대처하기 위해서 관련있는 정보자료를 '어느 정도 포함하고 있는나'를 평가하고자 하는 목적에서 설정되었다. 따라서 주요 분석대상은 DB의 구축을 위해 수집/분석/입력된 정보자료의 품질이 되며, 이때의 정보자료의 품질은: 포괄성 (comprehensiveness), 적합성 (appropriateness), 그리고 최신성 (currentness in time)에 의해서 평가된다. 가령, 하나의 DB가 물리학자들을 위해 구축되었다면, 그 DB의 구축을 위해 수집/분석/입력된 1차자료 (e.g., 저널, 회의록 등)가 種의 측면에서 어느 정도 해당 분야의 주요 정보소스들을 포함하고 있으며, 이와 같은 자료들은 내용면에서 어느 정도 물리학자들의 정보요구에 적합하며, 그리고 이들 자료의 최신성은 어느 정도 되는지를 측정하는 것이 해당 DB의 품질을 평가하는 주요 기준이 된다. 이때 측정방법은 일정 수의 자료를 DB로부터 검색하여 언급된 3가지의 평가기준에 따른 통계적 분석을 행하는 방법 (양적 측정)과, 일정 수의 이용자들을 직접 검색실험에 참여하게 하여 각 평가기준에 대한 그들의

종합적인 견해를 묻는 방법 (질적 측정)이 각기 독립적으로 혹은 병행해서 쓰일 수 있다.

3. 1. 2 DB에 포함되어 있는 레코드의 품질에 대한 평가 기준

이 항목은 하나의 DB에 포함되어 있는 단위 레코드들의 완성도를 측정하는 것을 통하여 해당 DB의 품질을 평가하고자 하는 목적에서 설정되었다. 따라서, 주요 분석 대상은 DB를 구성하는 레코드들이 되며, 이때의 레코드의 품질은: 그 구조의 일관성, 데이터 필드의 적합성, 수록 데이터의 정확성, 데이터의 갱신주기, 그리고 레코드사이의 중복률에 의해서 평가된다. 측정방법은 DB로 부터 검색된 일정 수의 레코드를 대상으로 언급된 5가지의 평가 기준에 의거하여 통계적 분석을 행하는 방법이 보편적이며, 일정 수의 모니터 요원들을 선정하여 일정 기간 검색실험에 참가하게 한 후 그들의 의견을 구하는 질적 측정 방법도 때로는 병행할 수 있다.

3. 1. 3 DB의 검색과 관련하여 제공되는 서비스의 품질에 대한 평가 기준

이 항목은 언뜻 볼 때, DB 자체의 품질을 평가하는 것과는 관련이 없는 것처럼 보인다. 그러나, 하나의 DB가 관련 데이터의 축적작업의 결과일 뿐만 아니라, 궁극적으로는 관련 데이터의 검색을 위한 도

구라는 점을 고려하면, DB 품질에 대한 종합적 평가작업에는 반드시 이 항목이 포함되어야 한다. 이 항목의 주안점은 DB의 검색과 관련하여 제공되는 여러 가지 정보서비스의 품질을 평가하는데 주어진다. 구체적으로, DB에 대한 접속방법과 접속과정의 용이성, 이용자 지원방법 (documentation, training, HELP 등)의 유무와 방법의 효율성, 인터페이스의 적절성, 검색방법의 효율성 등이 이 항목을 평가하는 주요 기준이 된다. 측정방법은 질적 측정을 통한 방법이 효율적인데, 가령, 최종이용자들을 일정 기간 검색실험에 참가시킨 후 그들의 종합적인 견해를 묻는 방법이나, 또는, 주의 깊게 선정된 검색요원들을 이용하여 전문적 견해를 구하는 방법이 바람직하다. 검색엔진의 효율성과 관련하여서는 검색된 자료의 관련도를 재현율이나 정확률 등의 기준을 활용하여 통계적 측정을 하기도 하나 본 연구에서는 배제된다.

3. 2 이용자의 관점에서 본 KRISTAL DB의 품질

각 항목별로 설정된 평가 기준에 의거하여, KRISTAL DB 중에서 비교적 인지도가 높은 UNION DB와 SATURN DB를 대상으로 품질평가를 실시하였다. 평가작업은 두 가지 방법을 이용하였는데, 본 연구를 위해 모니터 요원으로 선정된 이용자들의 종합적 견해를 묻는 방법과 본 연구팀의 직접적인 검색실험을 통해 데이

터를 수집 분석하는 방법이 병행되었다.

먼저, 인터뷰에 참여한 50명의 연구원들 중에서 KRISTAL 시스템을 사용해 본 경험이 있는 35명을 대상으로, KRISTAL 시스템과 DB에 대한 종합적인 의견을 구하였다. 인터뷰의 방식은 미리 설정된 세 가지 평가항목에 대해 연구원들이 평소 경험한 바를 자유로이 이야기하는 방식으로 진행되었으며, 수집된 데이터는 통계적 분석보다는 연구원들이 공통적으로 지적한 문제점들을 요약하여 기술하는 방식으로 처리하였다. 특히, 연구원들이 지적한 문제점의 심각성을 검증하기 위하여, 본 연구팀이 검색실험 도중에 발견했던 문제점들을 보충하여 기술하는 방식으로 KRISTAL DB에 대한 평가결과를 요약하였다.

3. 2. 1 DB에 포함되어 있는 정보자료의 품질에 대한 평가

KRISTAL DB에 포함되어 있는 정보자료의 품질에 대한 연구원들의 평가는 한마디로 '문제가 많다'라는 말로 압축된다. 특히, 주 분석대상이 되었던 SATURN DB의 경우, 포함되어 있는 정보자료의 포괄성, 적합성, 그리고 최신성 모두에 대해 연구원들의 평가는 매우 부정적인 것으로 나타났다.

구체적으로, "SATURN DB를 구성하는 정보자료가 자신의 연구분야에서 생산된 정보자료를 어느 정도 포함하고 있는나"는 포괄성을 묻는 질의에 대해, 대다수 연

구원들의 대답은 "거의 포함하지 못한다"는 것이었다 (5단위 Likert Scale을 활용하여 '모두 포함한다'를 5로, '전혀 포함하지 못한다'를 1로 설정한 뒤 포괄성을 측정하였을 때, 평균이 1.6으로 집계되었다).

다음, "SATURN DB에 포함되어 있는 정보자료가 현재 수행 중인 연구에 어느 정도 도움이 되느냐"는 적합성을 묻는 질의에 대해, 대다수의 연구원들은 "별로 적합하지 않다"라는 반응을 보였다 (5 단위 Likert scale에서 평균 1.7로 집계되었다).

마지막으로, "자신의 연구분야에서 최근에 생산된 연구물에 대한 정보를 SATURN DB로 부터 어느 정도 얻을 수 있는나"는 최신성을 묻는 질의에 대해, 연구원들의 반응은 '전혀 얻지 못한다'가 다수를 차지해, 이 항목 측정을 위한 세 가지 평가 기준 중에 가장 부정적인 것으로 나타났다 (5단위 Likert scale에서 평균 1.2로 집계되었다).

이처럼, 정보자료의 품질을 측정하기 위해서 설정된 세 가지 평가기준 모두에 있어 연구원들의 반응은 극히 부정적이었다. 여기서, "과학기술정보의 효율적 유통을 위해서 KRISTAL 시스템의 존재는 매우 중요하다"고 역설한 한 연구원의 진솔한 의견을, KRISTAL DB를 개발 운영하고 있는 관계자들에게 유용한 충고가 될 듯 하여 그대로 인용한다:

"... KRISTAL DB에 들어가 보고는

이것을 계속 이용한다는 것은 시간낭비라는 생각이 들었다. 우선 나에게 필요한 자료가 너무 없었다. 예를 들어, 과기처 연구보고서를 찾아보았는데 너무 오래된 자료였고 빠져있는 자료가 너무 많았다. 특히, SATURN DB의 경우, 내용면에서 적합한 자료가 거의 없어 신뢰성이 없어 보였다. 게다가, 최근 자료는 거의 찾아볼 수 없었다... 현재 수행 중인 project와 관련하여 중요한 탐색어를 선정하여 검색해보았더니 30여건의 레코드가 검색되었다. 그러나, 내가 필요로 하는 자료는 그 중에 단 한 건에 불과했다... 나는 주로 CAS를 이용한다. CAS를 보완 혹은 보충해 주는 자료가 이 DB에 포함되어 있다면 모를까, 현 상태에서는 이용의 필요를 못 느낀다... <과학기술인력정보>의 경우 나에게 는 유용하다. 그러나, 내용의 update가 제대로 안되어 있다. 직급이 틀린 경우는 예사이고, 심지어 연구소를 떠난 사람도 그대로 수록하고 있는 경우를 여러 번 보았다... DB의 구축이 시간적으로 오래되지 않은 점을 고려해도 너무 문제가 많다..." (S연구소소속 선임연구원)

3. 2. 2 DB에 포함되어 있는 레코드의 품질에 대한 평가

앞서 우리는 KRISTAL DB를 이용해 본 경험이 있는 35명의 연구원들 조차도 평균 이용빈도는 20일에 한 번 꼴로 매우 낮은 이용률을 보이고 있음을 이야기한 바 있다. 이와 같이 빈약한 이용경험에 근

거하여 DB에 포함되어 있는 단위 레코드의 품질을 평해 달라는 것에는 무리가 있었다. 그러나 이들 중 10여명은 SATURN DB에 포함되어 있는 서지레코드의 품질에 대해 평소 자신들이 느꼈던 문제점을 비교적 상세히 지적하였는데, DB에 포함되어 있는 단위 레코드의 품질 평가를 위해 본 연구팀이 설정한 5개의 평가기준에 따라 이들의 견해를 요약한다. (보충 분석은 검색실험을 통한 평가 부분에서 이루어진다).

먼저, SATURN DB에 포함되어 있는 레코드의 구조적 일관성 (표준화 정도)과 관련하여 제기된 가장 큰 문제점은, 레코드의 제작 입력기관에 따라 레코드의 구조가 다양하여 일관성이 없다는 점이었다. 특히, 레코드구조의 일관성 결여는 심지어 동일 연구기관에서 제작된 레코드의 경우도 입력년도에 따라 상이하다는 지적도 있었다.

다음, 단위 레코드를 구성하는 데이터필드의 적합성에 대해서도 여러 연구원들의 지적이 있었다. 가령, 레코드의 입력일자는 이용자의 입장에서는 불필요한 데이터인데도 '이용자보기' 화면에 포함되어 있었고, 소재정보나 원문수록처, 그리고 무엇보다도 주제명의 경우 반드시 필요한 데이터필드임에도 제작기관 (입력기관)에 따라 빠져 있는 경우가 많다는 지적이었다.

수록 데이터의 정확성은 레코드의 품질과 관련하여 가장 심각한 문제로 지적되었다. 가장 많은 지적이 있었던 부분은 철

자나 띄어쓰기의 오류였고, 데이터필드가 아예 비어있는 경우도 많다는 지적이 있었다. 잘못된 필드에 데이터가 수록되어 있는 경우와, 특히 내용의 불완전성과 표기방법의 통일성 결여에 대한 지적은 심각하게 제기되었다.

한편, 레코드의 갱신주기와 관련하여서도 연구원들의 지적이 있었는데, 가장 많은 지적은 잘못 입력된 레코드에 대한 수정작업이 거의 이루어지지 않고 있다는 점이었다. 더불어, 새로운 레코드의 입력 시기가 일정하지 않아 (현재 구축 중에 있는 DB라는 점을 고려하더라도) 새로운 자료를 탐색하고자 시간을 내어 접속하였다가 실망하는 경우가 빈번하다는 불만도 있었다.

마지막으로, 레코드의 중복률 또한 문제점으로 지적되었다. 동일 정보자료에 대한 여러 개의 서지레코드가 발견되는 경우를 경험한 연구원들이 많이 있었는데, 한 연구원의 경우 4건의 레코드를 출력하였는데 그 중에 3개가 동일한 자료에 대한 서지레코드인 경우도 있었음을 토로하였다.

3. 2. 3 DB의 검색과 관련하여 제공되는 서비스의 품질에 대한 평가

KRISTAL DB의 품질은 이 DB를 이용하여 관련 정보를 검색하는 과정에서 더

욱 부정적인 요소들에 직면하게 된다. KRISTAL DB를 이용해 본 경험이 있는 연구원들은 이구동성으로, 접속방법, 이용안내, 인터페이스, 그리고 검색방법 등, DB의 검색을 위해 제공되는 서비스의 품질이 지나치게 열악하다는 의견을 피력하였다.

구체적으로, 접속방법 및 과정과 관련하여 지적된 문제점으로는 “접속자체가 매우 힘들다”는 의견에 더하여 “접속하는 시간이 너무 오래 걸린다”는 불만이 많았으며, “검색도중 다운이 잘 되어 이용을 포기했다”는 연구원도 있었다. 본 연구팀의 검색실험 도중에도 검색한 양이 많을 경우 목차를 보여 주다가 혹은 레코드를 검색하다가도 다운이 잘 되어 다시 접속해야만 하는 경우가 적지 않았다. LAN이 구축되어 있는 연구소의 경우 “소속 연구소의 정보시스템을 통해 접속할 수 있으면 좋겠다”는 의견도 다수 표출되었다. 대다수 연구원들이 Netscape를 이용하여 Web상에서 접속하고 있었고, KRISTAL 소프트웨어 전용검색프로그램을 다운받아 이용하는 경우도 소수 있었다.²⁾

이용자에 대한 안내 및 지원방법에 대해서도 부정적인 견해가 많았다. 가령, Web상에서 열람용 DB를 선택하는 화면의 경우, 6개의 분류항목이 정확히 어떤 내용의 DB들로 구성되어 있는지, 각 DB

2) KRISTAL II 전용검색프로그램을 이용하고 있는 연구원은 세 명이었는데, 그들은 '전용 검색프로그램'을 통한 접속이 Web상에서의 접속 보다 시간적으로 빠르다는 편을 인식 하면서도, 이 프로그램에 대한 상세한 안내가 없을 뿐 아니라 홈페이지에서 이 프로그램을 다운받으자 할 때도 시간이 너무 오래 걸리기 때문에, 연구원들이 많이 이용하지 않는 것 같다는 의견을 피력하기도 하였다.

의 내용은 무엇인지 등에 대한 이용안내가 전혀 없었다. 구체적으로, '과학기술단행본도서'라는 명칭의 DB의 경우 DB구축에 참여한 기관만이 나열되어 있을 뿐이었고, '국내제작과학기술문헌'이라는 명칭의 DB의 경우도 그룹 10까지만 제작기관이 나열되어 있을 뿐이었다.

검색방법과 관련하여서도, 일반검색과 복합검색의 구체적인 차이는 무엇인지, 초보자의 경우 어떤 방식의 검색방법이 검색결과와 관련하여 효율적일 수 있는지 등에 대한 이용안내가 없었다. 특히 연산자의 사용과 관련하여, 검색전문가들로 구성된 본 연구팀의 입장에서도 그 기능이 무엇인지 애매한 기호들이 많았다 (가령, 저자명 중, AU, COH, PNH 등이 어떻게 구분되는지에 대한 설명이 없다 - 거의 동일한 기능을 가지는 연산자임에도).

KRISTAL 소프트웨어 전용검색프로그램의 경우도 이용안내는 그 내용이 무척 부실하였다. 도움말 (HELP) 기능이 있긴 하지만, 아주 간략한 설명만이 기술되어 초보자의 경우 별로 도움이 되지 않는다는 지적이었다. <이용자안내>라는 제목의 인쇄본형태의 자료 역시 그 내용이 지나치게 간략하여 초보자들에게는 거의 도움이 되지 않는다는 지적이 많았다. 이와 관련하여, 온라인상에서 DB 이용과 관련하여 연습할 수 있는 기능이 있었으면 좋겠다는 의견도 다수 표출되었다.

인터페이스에 대한 불만도 많았다. 무엇보다도, DB 열람용 선택 화면에 대한 지적이 많았는데, 각 분류항목과 DB의 명칭

이 모호하여 그 내용을 파악하기 어렵다는 점과 DB의 분류방법이 애매하다는 점이 지적되었다. 지금의 방식보다는 분류항목 화면을 만들어 (각 분류항목의 내용에 대한 간단한 설명과 함께) 이용자들이 자신의 정보요구에 적합한 분류항목을 선택하게 하고, 다음 화면에서 각 분류항목내의 DB를 선택하게 하는 것이 보다 이용자에게 친숙하게 다가갈 것 같다는 의견이 제시되었다.

Web 상에서 검색문 입력화면의 경우도, 일반 검색과 복합 검색으로 나누어 놓는 것이 오히려 혼동스럽다며, 차라리 일반이용자용 화면과 전문검색사용 화면으로 분리하고 간단한 안내문을 곁들여 놓는 것이, 이용자 입장에서 보면 보다 편리할 수도 있다는 의견도 있었다. KRISTAL 소프트웨어 전용검색프로그램의 경우도, '초보자'가 사용하기에는 오히려 어려워서 익숙해지려면 많은 시행착오를 되풀이해야 하며 이 때문에 탐색을 중지하는 경우가 많았음을 토로하였다. 검색엔진의 효율성 문제는 가장 심각하게 거론되었다.

검색엔진과 관련한 문제점은 본 연구팀의 검색실험과정에서도 빈번하게 발생하였는데, 가장 큰 문제는 앞서 언급되었던 검색방식의 문제에 더하여 색인어의 선정과 관련이 있었다. 다양한 색인필드가 제공됨으로써 검색을 위한 접근점이 다양해지는 것은 바람직하지만, 문제는 각 색인필드에 포함되어 있는 색인어의 선정기준이 모호하며 일관성이 결여되어 있는 경우가 자주 발견된다는 데 있다. 가령, '규

격'과 관련된 색인어 사이에 '규제'와 같이 유사 주제의 색인어로는 부적합한 단어들도 포함되어 있었는데, 이는 색인시스템에서 서로 관련된 색인어를 체계적으로 보여주지 못하고 있음을 의미한다. 색인어의 띄어쓰기 처리도 불분명하였는데, '돌연변이'로 검색하였을 때는 전혀 검색이 안되다가 '돌연 변이'로 검색을 하니 많은 자료가 출력되는 등 색인어의 통제와 관련하여 많은 문제가 있음이 노출되었다(유사한 예로는, "인공 지능", '유전 공학' 등이 있었다).

이외에 검색엔진과 관련한 문제점으로는, 연산자를 이용한 검색의 경우 검색에 성공하는 경우보다는 검색 자체가 안되는 경우가 많았고, 단위 필드명 검색(가령, 주제어 검색)의 경우에도 출력된 결과물을 보고 다시 같은 주제어를 입력했는데도 동일한 문헌이 검색되지 않는 경우가 발생하였다. 또한, 반드시 필요한 필드에 의한 검색은 불가능하고 전혀 불필요한 필드가 검색필드로 지정되어 있었다. 가령, 공저자의 경우, 저자필드에 이름들이 입력되어 있어 저자명검색만으로도 자료가 검색되는데, 공저자검색필드가 따로 필요한 이유가 분명하지 않았다. 더구나 SATURN DB의 경우, 소장기관별(혹은, 레코드입력 기관별) 검색의 중요성에도 불구하고 소장기관에 의한 검색이 불가능하였다.

3. 3 검색실험을 통해 본 KRISTAL DB의

품질

본 연구팀에 의한 검색실험은, UNION DB와 SATURN DB에 포함되어 있는 레코드의 품질을 측정하는데 일차적인 목적을 두었고, 나아가, 최종이용자인 연구원들이 지적하였던 검색엔진의 문제점을 구체적으로 검증해 보고자 하는 목적에서 시행되었다. 연구책임자를 비롯하여 모두 4명의 연구원들이 검색실험에 참여하였는데, 다음은 그 결과에 대한 요약이다.

3. 3. 1 UNION DB의 품질에 대한 평가

UNION DB에서 서지레코드는 다음의 15개 데이터필드를 그 기본 구조로 하고 있다: 제어번호, 자료형태, LCCN, ISBN, 언어코드, LC청구기호, DDC청구기호, 저자, 서명, 출판사항, 형태사항, 주제명, 소장기관, 레코드타입, 그리고 소장기관의 영문코드. 그러나 40개가 넘는 기관에서 레코드의 입력작업을 하다보니 입력기관에 따라 구조가 다른 레코드가 출력되는 등, 레코드 구조의 일관성과 관련하여 문제점이 발견되었다. 가령, 기본 데이터필드에는 <내용주기>나 <초록> 등의 필드가 없음에도 불구하고, 일부 기관에서(KIST 등)에서 입력한 레코드의 경우, 기본 구조에는 없는 <초록>이나 <내용주기> 필드를 포함하고 있는 경우가 다수 있었다. 특히, 국방과학연구소에서 입력한 레코드들은 레코드의 구조와 데이터필드의 순서에 있어서 다른 기관들에서 입력

한 것들과 커다란 차이를 보이고 있었다. 가령, <기본기입개인명>, <부출기입개인명>, <서명저자표시사항>, <일반 주기>와 같이 데이터필드가 상세하게 나누어지거나 추가되어 있었고, <자료형태> 필드의 경우 제일 끝에 위치하는 등 데이터필드의 순서도 다른 기관과 달랐다.³⁾

데이터 필드의 적합성과 관련하여서도 여러 문제점이 발견되었다. UNION DB와 같은 통합 DB는, DB구축에 참여하고 있는 기관들이 공동 편목 (shared cataloging)과 자원 공유 (resource sharing)를 통하여 과학기술 정보자료의 가공과 수집비용을 절감하고 나아가 과학기술정보의 유통을 효율화 하자는데 그 존재가치가 있다. 이 점을 고려할 때, UNION DB의 레코드에 있어서 빠져서는 안될 데이터필드가 '서지데이터의 입력행위'와 '원자료의 입수행위'와 관련된 부분이다.

구체적으로, 레코드의 <입력기관>과 <입력 혹은 갱신날짜>, 그리고 원자료의 <소장기관>과 관련한 데이터필드는 반드시 필요하다. UNION DB의 경우, '입력행위'와 관련한 데이터필드의 보완은 시급해 보인다 (이 문제는 특히, 통합 DB에서 레코드의 품질유지와 관련하여 매우 중요하다). '원자료의 입수'와 관련하여서도, <소장기관> 필드에 더하여 <소장기관의 영문코드> 필드가 분리되어 있는 이유가 명확해 보이지 않았다. 이외에 <저자> 필

드의 경우도 문제가 있었는데, <저자> 필드에 수록된 데이터의 내용을 보면 개인저자명과 단체저자 혹은 연구기관으로 보이는 단체명이 동시에 수록되어 있는 경우와 그렇지 않은 경우가 혼재하고 있는 등, 데이터필드의 기준 혹은 정의가 명확하지 않았다.

수록 데이터의 정확성은 UNION DB의 품질과 관련하여 가장 심각한 문제로 부각된다. 무엇보다도, 기본 데이터필드의 값이 완전하게 채워진 레코드는 거의 발견되지 않았으며 각 데이터필드내에 수록된 데이터도 그 표기방법의 통일성과 내용의 완성도에 있어서 큰 구멍이 있었다. 먼저, 기본 데이터필드로 설정된 15개의 필드 중에서, 특히, <소장기관>, <주제명>, <레코드타입> 등의 필드는 데이터가 누락되어 있는 경우가 빈번하였다. <소장기관>은 DB 구축에 참여하고 있는 일부 대학과 연구소의 경우 대부분 누락되어 있었다. <주제명>은 영문자료의 경우 비교적 충실했으나 국내자료나 그외 언어로 된 자료의 경우 데이터 자체가 누락되어 있는 경우가 많았으며, 주제명이 수록된 경우에도 지나치게 일반적인 주제가 선정되어 주제에 의한 자료탐색에 큰 도움이 되지 않았다. <레코드 타입>은 데이터가 누락되어 있는 경우가 더 많았으며, 해당 기호가 무엇을 의미하는지를 (가령, 'a'가 무엇을 의미하는지) 파악하는 것이 힘들었다.

표기방법 역시 통일되지 않아, 입력기관

3) 국방과학연구소의 경우는 특히, 입력시기에 따라 레코드의 구조가 서로 다른 것으로 조사되었다.

에 따라 다양한 형태로 표기되어 있었다. 가령, 저자 필드의 경우, 일부 레코드는 한글 이름과 영문 이름을 동시에 표기하고 있거나, 일부 레코드는 저자관련사항을 수록하고 있는 경우도 있었다. 언어코드의 경우도 일본어를 표시하는 JPN과 JA가 병행해서 쓰이는 등 표기가 통일되어 있지 않았고, 출판사항의 경우도 표기방법이 다양하게 나타나고 있었다 (가령, [대전]과 '대전'). 각 데이터필드의 내용의 완성도도 큰 문제였는데, 가령, 출판사항의 경우 출판장소나 출판사와 같은 기본 값이 누락되어 있는 레코드가 많았다. 특히, '문자지원' 문제에서 비롯된 것으로 보이는 한글과 영문 이외의 문자에 대한 '□' 처리가 많아 데이터의 불완전성을 심각하게 노출하고 있었다.⁴⁾

언어처리 또한 문제였다. 가령, 일본어 자료의 경우, 저자명을 한글로 읽어서 표기하고 원어는 표기하지 않거나, 주제어를 영문으로 표기하는 등 일관성이 없었다. 표기방법과 내용의 완성도는 검색작업시 검색의 효율에 지대한 영향을 미치는 점을 고려할 때 시급한 시정을 요하고 있다. 오자나 탈자 등의 철자상의 잘못은 다른 문제들에 비해 그 에러율 (error rate)이 비교적 양호하였으나, 검색점이 되는 주요 필드 (가령, 저자나 서명, 심지어, ISBN)에서 발생하는 오타가 3% 정도 발견되어

UNION DB의 품질과 관련하여 문제의 심각성을 더해주고 있었다. 레코드의 갱신 주기는 참여기관별 입력일자나 갱신일자가 표기되어 있지 않아 전체적으로 파악하기는 어려웠으나, 1995년 이후에 출판된 자료의 경우 전혀 검색되지 않아 (출판년도별 hit rate: 1995년 이후는 0%), 레코드의 최신성 유지와 관련하여 의문의 여지가 있었다. 더불어, 데이터에 오류가 있는 레코드에 대한 수정작업은 거의 이루어지지 않고 있는 것으로 보였는데, 2년여 전에 입력된 레코드의 오타가 그대로 방치되어 있었다.

레코드의 중복률에 있어서는 UNION DB의 경우는 그 정도가 심각하지 않아 여기서는 논외로 한다.

3. 3. 2 SATURN DB의 품질에 대한 평가

검색실험을 통해 수집된 데이터에 대한 1차 분석의 결과, 본 연구팀이 내린 결론은 "SATURN DB에 수록된 서지레코드의 품질은 매우 열악하여 시급한 개선방안이 제시되지 않을 경우, OCLC의 창시자 Kilgour의 표현대로, <쓰레기 데이터의 저장소>라는 오명과 함께 이용자들로 부터 외면당하는 DB가 될지도 모른다"는 것이었다. 먼저, 레코드의 구조와 관련한

4) 불어나 독일어의 독특한 언어형식이 표현되지 못하는 문제 뿐만 아니라, 화학식이나 기호도 표현되지 못하고 '□' 처리되는 사례가 많았다.

5) 적게는 9개의 데이터필드를 기본 구조로 하는 연구기관 부터 많게는 16개의 데이터필드를 기본 구조로 하는 연구기관에 이르기까지, 연구기관별로 서지레코드를 구성하는 데이터필드의 수가 다양하였다.

문제점으로는, 앞서 '이용자에 의한 평가' 부분에서도 언급하였듯이, SATURN DB의 구축에 참여하고 있는 12개 연구기관들이 서로 상이한 레코드 구조를 가지고 있었다.⁵⁾ 레코드의 구조적 일관성 결여 문제는 생각보다 심각하여 심지어 동일 연구기관에서 제작한 레코드의 경우에도 입력일시에 따라 서로 상이한 구조를 가지고 있었다. 물론, 자료의 유형에 따라 (저널논문, 보고서, 학위논문, 회의록논문 등) 레코드의 구조가 다른 점은 논외로 하더라도, 동일 유형의 자료에 있어서도 주요 데이터필드 (e.g., 주제코드, 식별자, 초록 등)가 일정기간 존재하다가 없어지거나 다른 필드로 대체되고 다시 나타났다가 없어지는 등, DB를 관리하고 있는 기관이나 DB를 구축하고 있는 기관이나 레코드구조의 일관성은 전혀 고려하고 있지 않은 듯 보였다.

레코드를 구성하는 데이터필드의 적합성은, (a) SATURN DB의 목적이 과학기술정보의 공유를 통한 연구생산성의 향상에 있으며, (b) 검색의 효율성이 데이터필드의 존재유무와 밀접한 연관이 있음을 고려할 때 매우 중요한 이슈가 된다. 가령, 전자의 목적을 위해서는 <소장기관에 대한 데이터>가, 후자와 관련하여서는 <자료의 유형에 대한 데이터>가 주요 데이터필드로서 반드시 포함되어야 한다. SATURN DB의 경우, 전자와 관련하여서는 <소장정보>라는 독립된 데이터필드가 존재하고 있어 큰 문제는 없어 보이나 <소장정보>에 의한 레코드의 검색이 실질

적으로 불가능하다는 점은 중요한 결점으로 보인다.

다음, SATURN DB의 경우처럼 여러 유형의 자료들이 하나의 DB에 포함되어 있을 경우 <자료의 유형>에 의한 검색은 매우 중요하다. 그러나 대부분의 연구기관에서 입력한 레코드에 <자료유형> 필드는 누락되어 있었다. 검색결과와 출력형태 지정양식에는 <자료유형> 필드가 존재하는 것으로 보아 KRISTAL DB의 기본 레코드구조에는 기본 필드로 설정되어 있었던 것이 분명해 보이나, 실질적으로 레코드를 제작/입력하는 연구기관들이 이 필드를 소홀히 취급하고 있는 듯한 인상을 받았다. 현재처럼 자료의 유형에 따라 레코드의 구조가 상이하게 검색되는 상황에서, <자료유형>에 따른 검색이 실질적으로 불가능하다는 점은 데이터필드의 적합성과 관련하여 또 하나의 문제점으로 지적된다.

인터뷰에 참여했던 연구원들도 지적하였듯이, 수록 데이터의 정확성 문제는 SATURN DB의 품질을 조악하게 만드는 주범이었다. 구체적으로, 아예 데이터필드가 비어있는 경우가 많았으며 (주제명, 원문수록처, 심지어 저자 필드까지), 특히, 기본 데이터 값이 누락되어 있는 경우도 많았다 (e.g., 저자 필드에 저자의 성 혹은 이름만 입력되어 있는 경우 등). 잘못된 필드에 잘못된 데이터가 수록된 경우도 있었으며 (e.g., 저자 필드에 제목이 입력된 경우), 철자나 띄어쓰기의 잘못은 이루어야 할 수 없이 많았다. 특히, 검색점이 되는 저자, 제목, 주제명, 심지어, 보고서번

호 등에도 많은 철자상의 잘못이 발견되었다.

데이터의 내용과 표기방법에도 문제가 있었다. 가령, 주제명의 경우 연구기관에 따라 지나치게 일반적인 주제명을 기입하여 검색효율에 오히려 부정적 영향을 미치는 경우가 빈번하였다 (e.g., '단백질', '세포동물', '규격', '표준' 등과 같이 제목에 나오는 한 단어를 뽑아 그대로 주제명으로 삼는 경우가 많았다). 저자의 경우도 서로 식별이 불가능할 정도로 약어 형태로 기록되거나 한/영문 이름이 혼재하여 표기되기도 하였고, 단체명의 경우도 동일기관의 명칭이 다양하게 표기되는 등 (e.g., 한국과학기술대 - 한국과학기술대학 - 과기대 - 한국과기대) 데이터의 내용적 부실에 더하여 표기방법에도 통일된 원칙이 없었다. 또한, UNION DB의 경우에서도 지적하였듯이, 화학식 같은 각종 공식이나 영어 일어 이외의 외국어의 경우 특수문자는 지원이 되지 않는 듯, '□'로 처리하는 경우가 많았다 (e.g., 10□SO-6□TPa).

한편, 레코드의 갱신주기는 일정하지 않은 것으로 나타났다. SATURN DB의 구축에 참여하고 있는 연구기관들의 기관단위의 정보시스템에는 수록되어 있는 서지레코드들이 SATURN DB에는 누락되어 있는 경우가 많았는데, 이는 각 연구기관들이 일정한 주기에 따라 레코드를 입력하기보다는 목표로 한 일정량의 레코드를 작성한 후 일괄적으로 입력하기 때문에 발생하는 문제인 것 같았다 (이 문제는

입력일자별로 레코드를 출력해 볼 때 명확히 드러난다). 이 문제는 SATURN DB의 최신성 유지와 관련하여 개선방안이 요구된다.

레코드의 갱신과 관련한 또 하나의 문제는, 잘못 입력된 데이터에 대한 수정작업이 거의 이루어지지 않고 있다는 점이다. 본 연구실험은 1996년 8월과 10-11월에 걸쳐 두 차례 진행되었는데, 앞서 실험에서 발견된 여러 오류들이 전혀 수정되지 않은 채 2차실험에서 그대로 발견되는 경우가 대부분이었다. 특히 몇몇 연구기관의 경우는 기관명칭이 바뀌면서 제어번호나 소장정보에 수록된 내용도 바뀌어야 하나, 과거의 명칭에 대한 약자를 그대로 쓰고 있었다.

마지막으로, 레코드의 중복률 또한 심각하여 동일한 자료에 대한 서지레코드가 중복되어 검색되는 경우가 빈번하였다. 중복 레코드의 유형은 크게 두 가지로 나누어지는데, 동일 정보자료를 다른 연구기관들이 따로 처리하여 입력한 경우와 동일 연구기관에서 중복 입력한 경우가 있었다. 중복 레코드의 가장 전형적인 例로는, 저자와 제목은 동일한데 그 외의 데이터(주제코드, 초록, 주제명 등)가 다르게 기입된 것이 있었으며, 심지어 모든 데이터 필드의 내용이 동일한 레코드들도 종종 발견되었다.

4. KRISTAL DB의 품질개선을 위한 제언

과학기술분야의 종합적인 정보유통시스템으로서 현재 KRISTAL 시스템이 지니고 있는 문제점은 크게 세 가지로 요약된다. 그 첫째는 시스템자체에 대한 인지도 및 이용도와 관련된 문제이고, 그 둘째는 시스템의 핵심이 되는 DB의 품질 문제이며, 마지막은 DB의 검색과 관련하여 제공되는 서비스의 품질 문제이다. 특히, 두 번째 항목인 DB의 품질 문제는 이 연구의 중심을 형성한다. 이 연구에서는 KRISTAL DB의 품질을 측정하기 위한 기준을 이용자의 관점을 중시하여 설정한 후, 그 기준에 의거 KRISTAL DB의 핵심이 되는 UNION DB와 SATURN DB의 품질을 구체적으로 측정해 보았다. 품질에 대한 측정의 결과는 한마디로 '문제가 많다'로 요약된다. 구축대상 DB의 선정에서부터, 구축대상 정보자료의 선정, 구축인력의 전문성, 실질적인 서지레코드의 작성 및 입력과정, 그리고 구축후의 품질 검증과정 등, DB의 구축과 관련된 전 과정 (process)에서 문제점이 발견되었다.

그렇다면, 이와 같은 문제점은 어디에서 비롯되었을까? 아마도 근본적인 원인은 KRISTAL DB의 구축에 앞서 KRISTAL 시스템의 역할 및 기능에 대한 장기적이고 종합적인 계획 (master plan)이 제대로 마련되어 있었는지에 대한 의문에서부터 찾아내는 것이 빠를 듯 싶다. 장기종합계획이 제대로 마련되어 있었

다면, 구축 DB의 선정기준, 구축과정에 대한 감독, 그리고, 구축후의 품질검증 등 전 과정에 걸친 관리가 초기부터 정례화 되어 지금과 같이 DB의 품질이 열악한 상태로 추락하지는 않았을 것이다. 물론, '양적 확충'을 우선시하는 대한민국의 가치판단기준이 DB구축에 있어서도 예외일 수는 없었을 것이며, 따라서, KORDIC이 KRISTAL DB의 양적 확충에 몰두해 온 것을 비난할 사람도 정책적 근거도 이 땅에는 존재하지 않을지 모른다. 오히려, 100만 건에 이르는 방대한 양의 레코드를 불과 4-5년의 짧은 기간동안 DB화하는데 성공한 KORDIC의 사례는 이 분야에 관심을 가진 모든 기관이나 인력들에게 '참고할 만한 사례'로 기억될지도 모른다.

그러나 이제는 우리의 시선을 '量' 보다는 '質'로 옮길 때가 되었다. 이 정보시대에 KRISTAL DB도 이점에서는 예외가 되어서는 안된다. 아니 오히려, 質 위주의 새로운 가치를 만들어 가는데 선도해야 할 의무가 과학기술분야의 중심적인 정보유통기관인 KORDIC에게 주어져 있다. 이제 여기서, 어떻게 하면 KRISTAL DB의 품질을 지금부터라도 개선할 수 있을는지, 그래서, 이용자에게 '없어서는 안될' DB로 거듭날 수 있을는지, 그 방안에 대해서 궁리해 보자.

본 연구팀이 제안하는 첫 번째 방안은 KRISTAL DB를 구성하는 단위 DB들의 선정과정과 관련된다. 본 연구를 위해 행해진 인터뷰에서 여러 연구원들이 지적하였듯이, 우선 지나치게 많은 단위 DB의

수에 비해 KRISTAL DB만의 특성을 보여줄 수 있는 단위 DB가 별로 없다는데 문제가 있다. KRISTAL DB를 통해서만 포괄적 탐색이 가능한 그런 내용의 단위 DB가 선정되어 구축되어야 한다. 그래야만, 이용자들의 관심은 증대된다. 특히, 집중 분석대상이 되었던 SATURN DB의 경우, 국내에서 생산된 과학기술문헌을 분석하여 서지정보를 수록한다는 초기의 취지와는 달리, 구축에 참여하고 있는 연구기관별로 제작된 서지레코드의 내용을 보면 해외에서 생산된 문헌에 대한 서지정보가 주종을 이루고 있음이 드러난다. 해외생산 문헌에 대한 서지정보를 구하고자 하는 연구원들에게 SATURN DB이외의 대안이 없다면 모르되, 수록 데이터의 내용이 보다 포괄적이고 그들의 정보요구에 보다 적합하며 최신성에 있어서도 앞서는 많은 주제별 서지 DB가 존재하고 있는 상황에서, 그 누가 SATURN DB를 이용하려 들겠는가? 이는 KRISTAL DB, 특히, SATURN DB의 구축을 책임지고 있는 관계자들이 고민해야할 첫 번째 문제이다.

이 문제를 해결하기 위한 방안은, 이용자들의 '정보추구행태'에 대한 연구를 DB 선정에 앞서 반드시 시행하는 것을 통하여 찾아질 수 있다. "어떤 DB를 구축할 것인가"를 결정하기에 앞서, 그 DB가 봉사대상으로 설정하고 있는 잠재적 사용자 그룹의 정보요구에 대한 포괄적이고 체계적인 조사가 반드시 필요하다. 사용자그룹의 정보요구에 대한 분석의 결과는, 어떤

DB가 우선적으로 제작되어야 하며, 그 DB에 포함될 정보자료는 어떤 기준에 의해 선정되어야 하며, 그 DB의 레코드는 구체적으로 어떤 구조를 가져야 하며, 그리고, 검색을 위한 데이터필드는 어떻게 선정해야 하는지 등에 대한 구체적인 해답을 제공하기 때문이다. 특히, 분산체제로 구축된 통합 DB인 SATURN DB의 내용을 여러 분야를 추가하여 확충하고자 한다면, 관계자들은 새로운 주제분야의 선정에 앞서 반드시 '정보요구조사'를 실시할 것을 권하고 싶다.

문제해결을 위한 두 번째 제언은 KRISTAL DB를 구성하는 단위 DB들의 구축과정을 개선하라는 것이다. 앞서 우리는 UNION DB와 SATURN DB를 사례로 하여 이들 DB에 포함되어 있는 레코드의 품질을 평가하였다. 품질평가를 위한 기준으로 설정된 5가지 항목(레코드 구조의 일관성, 데이터필드의 적합성, 수록 데이터의 정확성, 레코드와 데이터의 갱신 주기, 그리고 레코드의 중복률) 모두에 있어 두 DB의 품질은 결합 투성이었던 것으로 드러났다. 특히 SATURN DB의 경우, 문제는 매우 심각한 것으로 조사된 바 있다. 구축과정의 개선이 절실하다. 앞서 이야기하였듯이 우선은 양 보다는 질 위주의 구축전략이 필요하며, 이를 위해 현재의 분산구축과정은 보완되어야 한다. 연구기관마다 전문인력의 부족을 임시직원으로 보충하며, 정보자료에 대한 전문적인 분석을 통한 서지레코드의 제작이 아닌 어디선가 제작된 레코드의 단순한 다운로

드를 통한 입력작업을 반복하고 그 구축량을 과시하는 현재의 체제로는, 국가예산의 낭비이외의 결과를 기대할 수 없다.

UNION이나 SATURN DB와 같은 통합 DB의 경우 특히, 구축과정에 대한 대대적인 수술이 시급하다. 이 문제의 개선을 위한 방안은 두 가지로 압축된다. 현재의 분산구축과정을 중앙기관에 의한 통합구축과정으로 대체하는 것이 그 첫 번째이고, 철저한 감독 및 관리체제의 구축이 그 두 번째이다. 전자보다는 후자의 현실성이 높는데, 이를 위해서는 DB구축기관에 대한 DB관리기관 (여기서는 KORDIC)의 철저한 감독과 구축레코드에 대한 평가작업을 우선 정례화하고 평가결과에 의거하여 예산지원 등을 차별화하는 등, DB구축기관이 입력 데이터의 질적 향상을 위해 스스로 노력하도록 유도하는 정책적 고려가 필요하다. 더불어, DB구축기관이 입력한 레코드의 품질을 DB관리기관에서 검증하여 수정할 수 있도록 조직적이고 기술적인 기능을 KORDIC은 갖추어야 한다. 그렇게 될 때, 다소 늦은 감은 있지만 KRISTAL DB가 'junk DB'라는 오명을 벗을 수 있을 것이다.

마지막 제언은 KRISTAL DB의 검색과 관련한 서비스의 질을 개선하여야 한다는 것이다. 연구원들과의 인터뷰와 본 연구팀에 의한 검색실험의 결과는, 이 항목의 품질평가를 위해 설정된 4가지 기준 (접속 방법 및 과정의 용이성, 이용자에 대한 안내 및 지원방법, 인터페이스, 그리고 검색

엔진의 효율성) 모두에 있어 시급한 개선이 필요함을 보여주고 있다. 本文에서 지적된 여러 가지 구체적인 문제들을 개선하기 위한 방안은 KRISTAL DB를 관리하고 검색을 위한 소프트웨어를 개발하여 운영하고 있는 KORDIC의 조직개편에서 우선적으로 찾아야 한다. 검색과 관련한 서비스를 전담하는 새로운 부서가 필요해 보인다. 새로운 부서의 전문인력들은, 이용자들의 불만과 충고를 능동적으로 파악하고 잠재하는 요구사항을 적극적으로 찾아내어 그 결과를 KRISTAL 소프트웨어 개발팀에 지속적으로 제공함으로써 시스템개선에 협조하고, 더불어, 이용자에 대한 교육 실시, 교육과 홍보에 필요한 각종 도구를 제작 배포하는 등 KRISTAL DB의 이용과 관련한 다양한 종류의 서비스를 전담하여야 한다. 그래서, 이용자와 시스템간에 게이트키퍼 (gatekeeper)로서의 역할과 기능을 수행해야 한다.

결론적으로, 이제는 質의 개선 내지는 확보에 모두가 관심을 기울여야 할 때이다. 質이 담보되지 않는 한 量의 확충은 그 의미가 반감된다. 아니 오히려 시스템의 존재 자체를 위협하는 惡材가 된다. 여러 곳에서, 여러 기관에 의해서 혹은 여러 기관의 지원하에, 여러 종류와 유형의 DB가 만들어지고 있다. DB의 제작은 데이터의 보관을 위한 관리적 의미도 물론 있지만 데이터를 보다 효율적으로 활용하고자 하는데에 보다 큰 목적이 있다. 활발히 이용되는 DB를 만들어야 한다. 활발히 이용되는 DB를 만들려면 소비자가 무엇을 원

하는지 시장조사는 필수적이다. 소비자가 원하는 것은 무엇일까. 정보의 소비자는 무엇을 원할까? 그들이 원하는 것은 결코 많은 양의 정보가 아니라 우수한 품질의 정보이다. 그들의 요구에 적합하고 시의적절하며 그 내용을 신뢰할 수 있는 그런 良質의 정보라면, 소비자는 양이 적고 많음을 그리 심각하게 여기지 않는다. 이것이 다양한 종류의 DB 제작에 관계하고 있는 이들에게 이 연구가 주는 忠告이자 結語이다.

Acknowledgement

이 연구결과의 일부는 1996년 11월 25-26일 한국정보과학회와 KORDIC이 주최한 제1회 Korea Science and Technology Information Workshop에서 발표되었고, 거기서 제시된 여러 제언들은 KORDIC의 경영진에 의해 적극적으로 수용되었으며, 이후 KORDIC의 KRISTAL 시스템은 품질과 기능의 대폭적인 혁신을 진행중에 있다.

참고문헌

- 과학기술처. 1993. 과학기술정보 유통사업 추진계획. 서울: 과학기술처.
- 남영호 외. 1994. 2000년대를 향한 국가 과학기술정보 유통시스템 구상. 서울: STEPI. 연구개발정보센터.
1994. 과학기술정보 유통체계 강화 방안. 대전: 연구개발정보센터.
- Arnold, S. 1992. "Information Manufacturing: the Road to Database Quality." *Database* 15(5) : 32-39.
- Basch, R. 1990. "Measuring the Quality of Data: report on the Fourth Annual Southern California Online User Group (SCOUG) Retreat." *Database Searcher* 6(8): 18-23.
- Boissonnas, C. 1979. "The Quality of OCLC Bibliographic Records: the Cornell Law Library Experience." *Law Library Journal* 72(1): 80-85.
- Bourne, C. 1977. "Frequency and Impact of Spelling Errors in Bibliographic Databases." *Information Processing and Management* 13(1): 1-12.
- Davis, C. 1989. "Results of a Survey on Record Quality in the OCLC Database." *Technical Services Quarterly* 7(2): 43-53.

- Daniel, E. 1993. "Quality Control of Documents." *Library Trends* 41(4): 644-664.
- Decoster, B. 1984. A Comparative Study of OCLC and WLN in Twenty-nine Pacific Northwest Academic Libraries. Unpublished Ph.D. Dissertation, University of Texas at Austin.
- Dolan, D. 1992. "Quality Control at System Level." *Online* 16(2): 30-35.
- Hafter, R. 1986. *Academic Librarians and Cataloging Networks: Visibility, Quality Control, and Professional Status*. New York: Greenwood Press. Ch.VI. Quality Control Standards.
- Harry, V. and 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.
- Jones, B. and Kanster, A. 1983. "Duplicate Records in the Bibliographic Utilities: A Historical Review of the Printing versus Edition Problem." *Library Resources and Technical Services* 27(2): 211-220.
- Lundeberg, M., Goldkuhl, G., and Nilsson, A. 1981. *Information Systems Development: a Systematic Approach*. Englewood Cliffs, NJ: Prentice-Hall.
- Mandel, C. and Rhee, S. 1986. "Shared Cataloging: Some Remaining Issues." *Cataloging and Classification Quarterly* 7 (Winter 1986): 29-38.
- Melone, N. 1990. "A Theoretical Assessment of the User-satisfaction Construct Information Systems Research." *Management Science* 36(1) 598-613.
- Norton, N. 1981. "Dirty Data: A Call for Quality Control." *Online* 5(1): 40-41.
- Oglebay Institute on Quality Control. 1983. *Proceedings on Quality Control*. Pittsburgh: OCLC.
- O' Neill, E. and Vizine-Goetz, D. 1988. "Quality Control in Online Databases." *ARIST* 23: 125-156.
- Redman, T. 1992. *Data Quality: Management and Technology*. New York: Bantam Books.
- _____. 1995. "Improve Data Quality

- for Competitive Advantage.”
Sloan Management Review
36(2): 99-109.
- Schoenung, J. 1981. The Quality of the Member Input Monograph Records in the OCLC Online Union-Catalog. Unpublished Ph.D. Dissertation, Drexel University.
- Strong, D., Lee, Y.W. and Wang, R.Y. 1994. Beyond Accuracy: How Organizations are Redefining Data Quality. Cambridge, Mass.: Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management.
- Sule, G. 1986. “Problems of Duplicate Records, Standards and Quality Control.” In: A. Helal, ed. Future of Online Catalogues. Essen, West Germany: Essen University Library. pp. 145-167.
- Tague-Sutcliffe, Jean. 1995. Measuring Information: an Information Service Perspectives. San Diego, CA: Academic Press.
- Tenopir, C. 1993. “Quality of Abstracts.” Online 17(4): 44-55.
- Tillett, B. 1989. Authority Control in the Online Environment. New York, NY: The Haworth Press.
- Wang, R.Y., Storey, V. and Firth, C. 1995. “A Framework for Analysis of Data Quality Research.” IEEE Transactions on Knowledge and Data Engineering 7(4): 623-640.
- Wang, R.Y., Strong, D. and Guarascio, L. 1994. Beyond Accuracy: What Data Quality Means to Data Consumers. Cambridge, Mass.: Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management.