

# 서지데이터베이스의 품질관리

— K관의 MARC레코드 분석을 중심으로 —

김 지 훈\*

## <목 차>

- |                                 |                           |
|---------------------------------|---------------------------|
| I. 서 론                          | 2. 목록 기술상의 에러             |
| 1. 연구의 목적                       | 3. 주제관련 에러                |
| 2. 연구방법 및 제한점                   | 4. 편집상 에러                 |
| II. 서지데이터베이스 품질관리               | 5. 초록의 에러                 |
| 1. 품질의 개념                       | IV. 서지 데이터베이스 품질<br>관리 방법 |
| 2. 품질관리(Quality Control)        | 1. 수동방법                   |
| 3. 데이터베이스 품질                    | 2. 자동방법                   |
| 4. 서지데이터 베이스 품질관리               | V. 결 론                    |
| III. K관의 서지레코드 분석 및 에<br>러형태 조사 | 참고문헌                      |
| 1. 원문헌의 에러                      |                           |

## 1. 서 론

### 1. 연구의 목적

사회가 모든 분야에서 고도로 발전해 감에 따라, 그에 수반되어 생산되는 정보도 다양해짐과 동시에 엄청나게 증가하고 있다. 이러한 정보는 급속한 사회변화속에서 적시에 이용가능하도록 데이터베이스로 구축되어지고 있으며, 그와 병행하여 다양

\* 국토개발연구원 연구원

한 관련기술이 개발되고 있다. 나아가 이는 정보산업이라는 새로운 영역을 만들어 왔으며 앞으로 더욱 진보할 분야로 주목받고 있다.

이제까지 데이터베이스에 관한 기술은 주로 데이터베이스의 구축과 데이터처리기술, 서비스에 중점을 두어 왔으며, 데이터베이스 그 자체에 대한 평가는 상대적으로 미미하였다. 그러나 최근에는 데이터베이스의 품질에 대한 관심이 높아졌는데, 그러한 경향은 성능이 좋은 컴퓨터의 보급과 데이터베이스기술의 발달에 더불어 정보통신기술의 진보로 인해 데이터베이스를 이용할 기회가 많아지고, 또한 온라인 또는 CD-ROM으로 구축된 데이터베이스가 수적으로나 양적으로 늘어나고, 사회적으로 데이터베이스의 품질을 강조하는 경향이 증가함으로써 나타났다. 그 한 예로 오잘라(Ojala)는 사회적으로 품질의 중요성과 그에 대한 관심이 증가하는 증거로서 ABI/INFORM 데이터베이스의 디스크립터와 자연어에서 '품질(quality)'이란 단어의 사용빈도를 1975년부터 1991년까지 조사하였는데, 그 수적인 면에서의 증가뿐만 아니라 다른단어중 그 사용빈도도 상당히 증가함을 보여주었다.<sup>1)</sup>

데이터베이스의 품질평가는 크게 두가지부문으로 나눌 수 있다. 첫째는 거시적인 평가로서 데이터베이스가 주제영역에 대해 얼마나 많은 데이터를 보유하고 있으며, 수록범위가 어느 정도이며, 보유한 데이터의 주제의 망라성이 어느 정도인가에 대한 것이고, 둘째는 미시적인 평가로서 데이터베이스내의 레코드의 평가로서 레코드의 포맷, 데이터의 에러 등 데이터 그 자체에 대한 것이다.

본 연구는 미시적인 측면에서 데이터베이스의 품질관리에 대한 연구와 품질향상을 위한 가능한 방법에 대한 연구로서, 완전하고 정확한 서지데이터베이스를 구현하기 위해 서지데이터베이스의 MARC레코드를 분석함과 동시에 품질관리방안을 모색하고자 한다.

---

1) Marydee Ojala, "Quality Online and Online Quality," *Online*, Vol.16, No.1 (January 1992), pp. 73-75.

## 2. 연구방법 및 제한점

최근까지 대부분 도서관에서는 전통적인 카드목록의 보조수단이나 대체수단으로서 이용하기 위해 기본적인 서지정보만을 입력하였으며, 소급데이터를 컴퓨터를 통해 입력하는데 원자료를 대상으로 하기보다는 목록카드의 내용에만 국한하여 입력하였으며, 시간적인 제약으로 인해 입력하는 데이터도 기술목록규칙에 의해 요구되어지는 모든 사항을 입력하는 것이 아니라 단지 최소한 요구되는 필드만 포함하였으며, 비MARC레코드를 MARC레코드로 변환한 후 수정이 제대로 이루어지지 않았으며, 데이터를 입력하는 목록자가 MARC내용을 충분히 숙지하지 못한 경우가 많았다. 이러한 이유로 인해 구축된 데이터베이스의 품질이나 이용자의 신뢰성에 대해서는 그다지 관심을 두지 않았다.

이 연구는 먼저 데이터베이스 품질과 품질관리에 대해서 논하고 서지데이터베이스의 품질에서 고려할 사항, 품질개선방법 등을 다루었으며, 서지레코드에서 나타나는 에러를 K관의 MARC레코드의 에러검출 및 수정사례를 통하여 조사하고, 앞으로 고품질의 서지레코드를 유지하기 위한 방안을 고찰하였다.

K관의 경우, 비MARC 레코드에서 KORMARC레코드로 변환시키면서 기존의 레코드구조에서 각 필드사항을 MARC 포맷과 상응하는 부분은 그에 맞는 태그를 부여하였으며, 많이 나타나는 내용은 디폴트값으로 미리 MARC레코드에 입력되도록 하였다. 그리고 나머지 필드내용은 레코드구조를 MARC포맷으로 변환후 수작업을 통해 레코드를 수정하였다. 또한 레코드의 수정과 함께 부가적인 기입은 기존의 카드목록을 근거로 하여 기본적인 서지내용만으로 추가 및 수정을 하였으며, 작업자의 MARC에 대한 전반적인 지식의 부족으로 작업자마다 입력사항이 틀리는 경우가 비일비재하여 서지레코드의 품질이 상당히 불량하였다. 그리하여 1차적인 수정작업이 끝난 다음에도 많은 레코드에서 부분적인 에러가 그대로 방기된 상태였고, 데이터베이스 이용자들도 많은 불만을 가지고 있는 실정이다.

한편 서지레코드의 에러를 분석하고 형태별로 에러를 수치화 하여 데이터베이스의 에러율과 완전성을 조사하여 품질개선의 기준을 삼을 필요도 있으나, 앞서 지적한 것처럼 K관의 경우, MARC형태로 데이터를 구축한 시기가 얼마되지 않고, 2차 수정작업이 아직 끝나지 않은 관계로 인해 이 연구에서는 별로 의미가 없으리라 여겨져 제외하였다.

## II. 서지데이터베이스 품질관리

### 1. 품질의 개념

품질은 주로 상품학에 사용되어온 것으로 생산물에 대한 평가로서 많이 다루어졌으며, 대체적으로 품질에 대해 다음과 같은 논의가 있다.

품질의 척도는 명세, 일치, 신뢰성, 비용, 인도를 각각 반영한다. 첫째, 명세는 고객이 요구하는 것이 무엇인가를 의미하고 둘째, 일치는 명세서와 일치하는가를 나타내고 셋째, 신뢰성은 고객의 요구를 계속적으로 만족시키는가를 나타내며, 넷째, 비용은 금전적 측면의 가치이며, 다섯째, 인도는 정시에 제품을 인도하는 것이다.<sup>2)</sup>

영국표준연구소(British Standards Institution)에서는 품질을 ① 상대적 품질, ② 절대적 품질, ③ 목적에 부합하는 품질로 구분하였다. 이들 각 개념을 데이터베이스에 적용해 보았을 때, 상대적 품질은 그 품질에 근거하여 유사한 생산물과의 등급을 매기는 것이고, 절대적 품질은 절대적의미에서 생산물의 품질을 평가하기 위한 양적인 측정으로서 데이터베이스에서 레코드당 평균 에러수나 중복레코드의

2) 大韓, 서울商工會議所, 新경영기법 : 한국에서 유행인가 새로운 물건인가?, 서울, 大韓, 서울商工會議所, 1994, pp.56-57.

퍼센테이지를 나타내는 것이고, 목적에 부합하는 품질은 특정한 요구에 만족하는 것과 관련있는 생산물이나 서비스를 평가하는 것으로, 목적에 부합한다는 것은 예를 들어 혼란된 중개자에 의해 이용되어지도록 설계된 데이터베이스는 최종이용자에게 적합하지 않다는 것을 의미한다.<sup>3)</sup>

올레이센(Olaisen)은 인지적인 전거품질요인(어떻게 정보가 이용자에게 인식되는가)과 기술적인 이용자편의품질요인(이용자가 제공받는 것)을 정보품질의 철학적 측면과 서비스품질의 경영이론적측면을 고려하여 구별하였다.<sup>4)</sup> 전자는 확실성, 신뢰성, 적합성, 시간경과에 따른 의미, 타당성, 인지값 등의 특성을 고려한 것이고, 후자는 형식, 실제값이나 새로운것, 접근성, 적시성, 속도, 유연성, 완전성, 본질적인 가능성, 선별성, 브라우징능력 등의 특성에 대한 것을 말한다. 이러한 경험적인 연구에 의해 그는 정보품질모델을 다음의 4가지로 구별하였다.

- ① 인지적 품질(cognitive quality)
- ② 설계품질(design quality)
- ③ 생산품질(product quality)
- ④ 인수품질(delivery quality)

또한 마찬드(Marchand)는 정보품질에 대한 분석을 위해 8가지 틀을 다음과 같이 정의했다.<sup>5)</sup>

- ① 실제값(actual value) : 정보생산이나 서비스는 이용자를 위한 것이므로, 그

3) British Standards Institution, *Glossary of Terms Used in Quality Assurance(Including Reliability and Maintainability Terms)*, London, British Standards Institution, 1979, p.26 (Edward T. O'Neill and Diane Vizine-Goetz, "Quality Control in Online Database." *ARIST*, Vol.23(1988), p.127에서 재인용)

4) J. Olaisen, "Information Quality Factors and the Cognitive Authority of Electronic Information," In I. Wormell, ed., *Information Quality: Definitions and Dimensions(Proceedings of a NORDINFO Service, royal School of Librarianship, Copenhagen, 1989)*, London, Taylor Graham, 1990, pp.91-121. (Evelyn H. Daniel, "Quality Control of Documents," *Library Trends*, Vol.41, No.3.4(Spring 1993), pp.656-657에서 재인용)

5) D. Marchand, "Managing Information Quality," In I. Wormell, ed., *op. cit.* pp.7-17. (Evelyn H. Daniel, *op. cit.*, p.656.에서 재인용)

값의 속성은 정보자체 뿐만 아니라 이용자의 개별기준에도 적용

- ② 정보생산이나 서비스의 특징 : 정확성과 포괄성
- ③ 정보생산이나 서비스의 확실성
- ④ 정보의 의미변화 : 정보는 수명주기(life cycle)가 있으며 환경에 따라 변함
- ⑤ 정보의 관련성 : 정보가 이용자의 기준이나 표준을 따르는 정도
- ⑥ 정보의 타당성 : 정보가 어떻게 수집되고, 분석되며, 누가 전달하고, 어떻게 결과가 나타나는가 하는 것과 관련있음
- ⑦ 정보의 미학 : 정보가 나타나고, 전달되고, 패키지화되는 방법과 관련있는 주관적인 속성
- ⑧ 정보의 인지값 : 간접적인 비교측정을 통해 다른 부분의 부족을 보충하기 위해 사용함

이상의 논의를 종합해 볼때 품질은 생산물의 속성에 대한 것으로, 품질을 향상시키는 것은 생산물의 속성을 최선의 상태로 만드는 것을 말하며, 그렇게 하기 위해서 이용자의 요구에 근거하여 표준(명세)을 설정하고 그에 따르는 것이 가장 중요한 사항이라 하겠다.

## 2. 품질관리(Quality Control)

품질관리는 각종의 과학적 원리를 응용하여 제품품질의 유지 및 향상을 기하는 것을 의미하는데, 이는 품질표준의 설정, 품질의 검사, 그리고 보정활동으로 구성된다. 첫째 품질표준은 소비자의 요구를 품질에 반영하고 제품이 소비자를 만족시키는데 비용과 효과를 판단하는 것이며, 둘째 검사는 품질에 관한 측정치를 시계열적으로 상한과 하한의 관리한계선에 의해서 나타낸 것이고, 셋째 보정활동은 품질 보고를 바탕으로 불량원인의 발견과 그 시정조치를 말한다.<sup>6)</sup>

6) 東亞出版社, 東亞原色世界大百科事典, 서울, 동아출판사, 1989.

이러한 품질관리는 가장 시장성이 높은 제품을 가장 경제적으로 생산하기 위한 일련의 체계적 조치를 포함하여 TQC(Total Quality Control)로 발전하였으며, 나아가 고객의 요구조건에 최대한 만족함과 동시에 품질관리를 통해 기업의 수익을 높이는 데 초점을 두는 것으로, 요즘 한창 관심을 끌고 있는 기업경영기법의 하나인 TQM(Total Quality Management)으로 발전하였다.

따라서 품질관리는 이용자의 요구와 기대에 부합하는 만족할 만한 품질의 생산물과 서비스를 생산하는데 필요한 기술, 활동, 경영철학을 포함한다. 즉 품질관리를 하는 것은 가장 경제적이고 유용하며, 항상 소비자에게 만족하는 품질의 생산물을 개발하고, 설계하고, 생산하고 서비스하는 것이다. 여기서 이용자지향은 가장 중요하다.

이와같이 품질관리는 산업 전반에 걸쳐 아주 넓게 중요한 부문으로 인식되고 있으며, 정보산업부문에 국한시키더라도 매우 넓다. 정보산업의 생산물에 대한 품질은 하드웨어, 처리 및 검색소프트웨어, 통신, 도큐멘테이션, 이용자지원, 데이터베이스 품질을 평가함으로써 영향을 받는다. 이중 데이터베이스 품질은 데이터 그 자체의 품질과 예러발견과 수정으로 데이터품질을 강화하거나, 데이터베이스의 품질을 개선시키기 위한 방법론에 제한하였다.

지금까지 정보관련산업은 품질에 대한 만족할 만한 인정을 받지 못하였다. 1987년 미국에서는 회사의 동기부여를 위해 Malcolm Baldrige National Quality Award란 품질과 관련된 상이 제정되었다. 이 상의 기준은 리더십, 정보이용 및 분석, 전략적인 품질계획, 인적자원이용, 생산물과 서비스의 품질보증, 품질결과, 이용자만족도에 대한 것이었는데, 현재까지 정보서비스관련기관은 인간자원이용면에서 노동자 재편성이 너무 빈번하였고, 품질보증면에서 품질측정방법이 부족하였기 때문에 이 상을 수여하지 못하였다.<sup>7)</sup>

7) Evelyn H. Daniel, *op. cit.*, p.651.

한편, 품질보증(Quality Assurance)은 품질관리와 동일하게 사용되나, 원칙적으로 품질관리보다 넓은 개념으로서 소프트웨어 개발의 중요한 측면이 되었다. 그러나 데이터처리와 관련하여 품질보증은 대개 소프트웨어의 품질과 무결성을 언급하며, 일반적으로 데이터품질에 대해서는 사용되지 않았다. 정보무결성(Information Integrity)은 품질과 밀접하게 관련있는 개념이나, 데이터나 정보에 중점을 두는 것으로서 정확하고 적절한 시간에 정보의 수집, 처리, 저장, 배포를 포함하며, 이와 밀접하게 관련있는 개념은 전거성(authoritiveness), 즉 정보의 인지된 정확성이다.<sup>8)</sup>

### 3. 데이터베이스 품질

데이터베이스품질에 대한 연구는 1960년대부터 시작되었다. 초기연구는 주로 절차어검출과 수정, 자체체크데이터의 적용에 관한 것으로서 대개 특별한 에러를 검출하고 수정하는데 많은 기법들이 만들어졌다. 그러한 연구는 주로 서지데이터베이스를 다루었으며, 수치나 전문데이터베이스에 대한 품질은 최근에 중요하게 대두되고 있다.

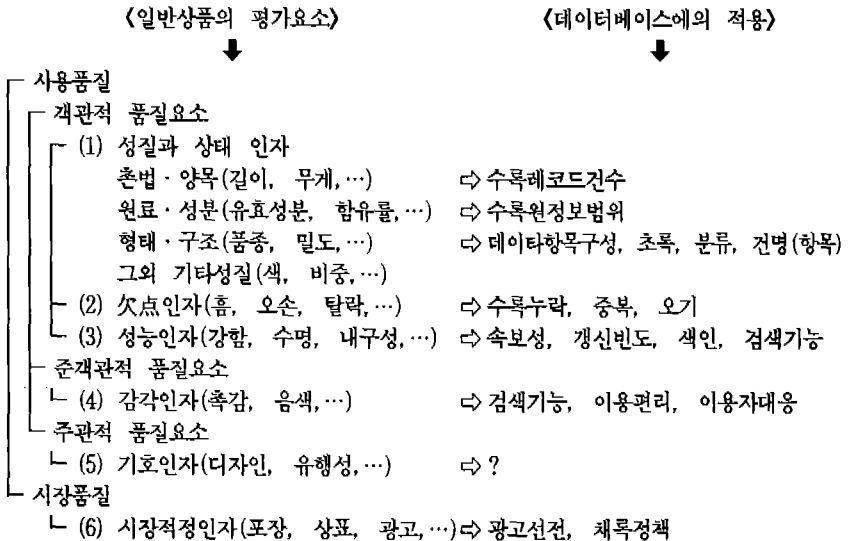
수치 및 전문데이터베이스의 에러는 예상치 못한 엄청난 손실을 초래할 수 있는데, 예를 들어 수치데이터가 원래 ₩5,000,000이 ₩5,000으로 오기된 것을 이용자가 그대로 사용한 경우 이용자는 잘못된 데이터의 사용으로 또다른 결과를 초래할 수 있다. 이러한 데이터의 에러에 대한 책임은 일차적으로 정보제공자에게 있다. 정보제공자는 이용자에게 제공된 정보가 정확해야할 책임이 있다. 일반적으로 정보제공자는 데이터의 정확성, 완전성, 데이터베이스가 목적에 부합하는가에 대한 책임이 있어야 한다. 따라서 정보제공자는 잘못된 정보에 대한 책임을 져야하고, 이용자는 법적인 보호를 받아야 할 의무가 있다.

8) Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, p.127.



이제껏 많은 연구자들이 데이터 자체에 대한 연구보다 컴퓨터시스템의 도입, 검색기법, 전통적인 업무에 최신 기술의 도입, 네트워크 등에 대한 연구가 활발하였으나, 이제 데이터베이스 품질은 많은 대형데이터베이스의 출현과 높은 이용, 일반적으로 사회에서의 품질에 대한 관심고조로 인해 점차 강조되고있다. 특히 1988년 「ARIST」에서 서지데이터베이스에서 데이터 자체의 품질, 에러검출 및 수정에 의한 데이터품질강화와 컴퓨터데이터베이스의 품질개선방법에 관한 전반적인 논의가 있었으며,<sup>9)</sup> 1994년 「情報の科學と技術」에서는 ‘데이터베이스의 품질관리’ 라는 제목하에 데이터베이스에의 품질관리 및 대응방안을 다루어 더욱 관심이 고조되고 있다.

根岸正光은 데이터베이스의 품질을 논하면서 먼저 일반상품론의 품질평가요소를 데이터베이스에 <그림. 1>과 같이 적용시켰다.<sup>10)</sup>



<그림 1> 일반상품의 품질평가요소체계와 데이터베이스의 품질요소

9) Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, pp.125-156.

10) 根岸正光, “アータベースの運用管理,” 情報の科學と技術, Vol.44, No.4(1994), p.155.

데이터베이스의 품질은 데이터베이스 생산자, 온라인 호스트시스템(시스템운영자), 탐색자(탐색전문가)에 의해 각각 다른 요소로서 나타난다.<sup>11)</sup> 데이터베이스생산자는 올바른 사실, 정확한 인용, 적절한 색인, 원문헌의 내용을 정확하게 반영하는 초록 등을 잘 구비해 생산해야 하며, 온라인 호스트시스템은 적당한 소프트웨어, 문서화된 명령어 및 다양한 도움말 기능, 정확하고 이해할만한 요금, 이용자훈련을 제공해야 한다. 또한 탐색자는 품질을 인지하고 평가하는 자로서 잘 훈련되고, 데이터베이스생산자와 호스트의 한계와 장점을 잘 파악하고, CD-ROM이나 전문가 또는 대체정보소스를 잘 숙지하고, 이용자질문에 답할 모든 가능한 최신정보에 민감해야 한다. 이 연구에서는 주로 데이터베이스 생산자에 의한 품질요소에만 국한하였다.

#### 4. 서지데이터베이스 품질관리

서지데이터베이스에서 품질을 요구하게 된 근본적인 배경은 입력된 데이터의 에러이며, 이러한 에러를 검출하고 수정하여 데이터의 품질을 강화하는데 노력을 기울여 왔다.

보이소나스(Boissonnas)는 법률자료에 대한 LC레코드와 OCLC회원입력레코드를 비교하여 OCLC레코드가 LC보다 두배이상의 수정이 요구된다고 밝히면서, 전자서지레코드(수정을 필요치 않는 레코드)의 품질은 AACR, ISBD포맷, LC 실행 지침(practice)에 준용하여 작성한 것이라고 정의하고, 전자레코드로 인정할 수 있는 데이터로서 LC레코드는 29%, OCLC 회원입력레코드는 1.3%밖에 안된다고 지적하였다. 또한 비슷한 에러연구를 한 리안스(Ryans)는 OCLC회원입력레코드

11) Marydee Ojala, "A Model for Quality Business Searching," *Online*, Vol.17, No.3 (March 1993), pp.78-79.

에서 올바른 데이터는 60%정도 나타난다고 하였다.<sup>12)</sup> 이 두 연구결과의 차이는 전거레코드를 정의하는 기준의 차이에 의해 기인된 것으로 보이며, 퍼센테이지의 수치는 사실상 큰 의미가 없는 것으로 보여진다.

그와 더불어 인터너(Intner)는 OCLC와 RLIN의 각각 215개의 같은 목록레코드를 비교하여 에러와 차이점을 조사했는데, 모두 1067개의 다양한 형태의 에러가 나타났으며, 한 레코드당 평균 2.48개의 에러가 있음을 보여주었다. 그의 조사중 흥미있는 부문은 두쌍의 목록레코드에서 서로 누락된 것이 269건이 나타났다는 것이다.<sup>13)</sup>

한편 대부분의 서지데이터베이스에 관한 연구는 영어권문자에 국한하였으나, 중국어레코드에 대한 품질관리와 그 개선방안에 대한 연구는 쟁(Zeng)이 처음이었다.<sup>14)</sup> 그는 OCLC의 CJK자료중 중국어레코드를 품질분석하여 여러가지 에러형태 별로 에러율을 조사하고 그 개선방안으로서 품질관리를 위한 규칙기반데이터검증시스템(rule-based data validation system)을 개발하였다.

서지데이터베이스 품질관리는 세가지 측면에서 고려될 수 있다.

첫째, 데이터베이스 내에 포함되는 문헌의 선택과 선택된 문헌의 처리

둘째, 데이터베이스의 최신성과 적합성

셋째, 이용자요구에 대해 데이터베이스로 부터 자료를 선택하는 문제

이와함께 데이터베이스 생산자에 의한 가장 일반적인 품질관리방법은 조사과정과 결합된 에러수정 소프트웨어나 자동관리시스템(automated control system)을 이

12) Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, pp.128-129.

13) Shelia S. Intner, "Quality in Bibliographic Databases: An Analysis of Member-Contributed Cataloging in OCLC and RLIN." In McCabe, Gerard B. and Kreissman, Bernard. ed., *Advances in Library Administration and Organization: A Research Annual*. Greenwich, Connecticut, JAI Press, Inc., 1989, pp.1-24.

14) Lei Zeng, *An Evaluation of the Quality of Chinese-Language Records in the OCLC OCLC Database and A Study of a Rule-based Data Validation System for Online Chinese Cataloging*(Ph.D. dissertation), Pittsburgh, PA, University of Pittsburgh, 1992.

용하는 것이다. 다음은 두개의 대표적 서지데이터베이스에서 적용하고 있는 품질관리방법이다.

OCLC의 경우, 중심기관은 품질관리부서를 신설하고, 목록작업을 통제하고 표준에 따르도록 유도하는 지역위원회와 목록전문가그룹을 개선하고, MARC 편목교육을 하는 워크숍을 개최하고, 모범이 되는 우수도서관을 지정하는 등 많은 프로그램을 개발하여 목록품질개선에 노력을 하고 있다. OCLC의 경우 일반공공도서관이나 소규모도서관은 그들의 주이용자가 정확하고 광범위한 목록레코드를 요구하지 않고, 경비절감의 차원에서 목록자를 줄일 목적으로 네트워크에 가입하여 대표적인 목록을 이용하는 경우가 많으므로 품질표준의 역할로서도 이러한 데이터베이스의 품질관리는 더욱 중요하다 하겠다.

연구중심의 RLIN은 OCLC보다 더 높은 품질을 관심을 두어 표준에 아주 집착하여 레코드의 완전성과 정확성도 높은 편이었다. 그 이유는 연구도서관그룹을 구성하는 회원도서관은 단지 학술적인 연구자료제공에만 관심을 가졌으며, 그러한 요구에 부응하기 위해 완전하고 정확한 목록레코드가 중요하였기 때문이다. 또한 이러한 연구기관들은 1차연구자료, 외국어자료, 기술보고서 등 원목록(original cataloging)을 요구하는 상대적으로 많은 특수한 자료를 수집하였으므로 많은 목록전문가가 있는 목록부서를 유지하였다.

### Ⅲ. K관의 서지레코드 분석 및 에러형태조사

#### 1. 원문헌의 에러

데이터베이스의 에러중 하나는 원문헌의 에러로서 잘못된 데이터나 오래된 데이터는 원문헌의 잘못된 인쇄나 참조로 나타난다. 이것은 고객에게는 물론 데이터베

이스 제작자에게도 어쩔 수 없는 문제로 이러한 실수는 저자나 참조자뿐만 아니라 잘못된 잡지명으로도 나타난다.

대부분의 데이터베이스 제작자는 이런 경우 수정을 가하지 않는 경우가 많으며, 특히 공식적으로 출판된 자료는 비록 에러가 발견되더라도 그대로 데이터베이스에 입력한다.

K관의 경우 이러한 에러는 거의 없었으며, 만약 이러한 에러가 나타난다면 이용자 검색시 영향을 미칠 수 있으므로 그대로 데이터베이스에 입력하지 않고 빠르게 수정한 상태로 입력을 수행하며, 다만 특별히 참조되어야 할 것은 주기사항에 기재 하도록 하였다.

## 2. 목록기술상의 에러

목록기술상의 에러는 목록카드나 레코드의 항목을 표현하는데 직접적인 영향을 줌과 동시에 탐색시 쉽게 드러나는데, 포맷에러와 내용에러로 구분할 수 있다. 포맷에러는 태그, 지시자, 하위필드코드, 구두점, 띄어쓰기 등의 부적절한 사용에서 나타나며, 내용에러는 필드와 하위필드의 누락, 필드와 속성간의 비일관성, 잘못된 내용 등에서 나타난다.

목록기술상의 에러는 전문적인 지식을 요구하는 기술목록규칙과 MARC에 대한 이해와 적용을 잘못하여 나타나는 것이 대부분이다. 에러중 1/3이 기술목록규칙의 에러라고 지적한 인터너(Intner)는 그의 표본서지레코드 중 OCLC의 기술표목에서 나타난 84개의 에러중 26개가 검색에 영향을 주고, RLIN의 기술표목의 94개 에러중 29개가 검색에 영향을 주는 것으로 조사하여,<sup>15)</sup> 이러한 에러가 검색에 많은 영향이 미침을 보여주었다.

15) Shelia S. Intner, *op. cit.*, pp.1-24.

포맷에러는 레코드의 모든 기술영역에서 나타나며, 목록기술상 많이 나타나는 에러로 알려져 있다. 특히 구뿃점에러와 띄어쓰기에러는 가장 많이 나타나는 것으로 여러 연구자의 연구결과 나타났다.

태그에러는 태그번호를 잘못사용한 경우로서 검색시 영향을 주는 것으로 나타났다. 인트너(Intner)는 태그에러의 대다수는 서명필드(245)와 제어필드(0xx), 총서명필드(4xx)에서 나타났음을 보여주었다.<sup>16)</sup> 이런 필드가 다른 필드보다 더 중요한 이유는 첫째, 245필드와 440필드는 탐색자에게 서명, 부서명, 총서명과 같은 중요한 데이터를 포함하며, 또한 245필드는 서명부출기입을 만드는 데 사용되기 때문이며, 둘째 제어필드는 온라인시스템에서 데이터관리에 아주 중요하기 때문에 에러가 나타나지 않도록 각별히 신경써야 한다.

다음은 K관의 MARC레코드로 변환후 수정작업 전후로 발생한 각 필드의 에러형태를 사례별로 나타낸 것이다.

#### ① 부호화정보필드에서의 에러

MARC레코드로 변환후 대부분 목록카드를 통해 수정을 하였기 때문에 누락되는 정보가 많다. K관의 경우 자료구성상 연구보고서가 많고, 최근의 연구보고서는 일반 단행본과 구별하기 힘들기 때문에 내용형식을 선정하는데 기준을 두어야 했다. 그리고 레코드 변환후 발행국명이나 언어부호를 새로 입력해야 했다.

#### ② 기본표목 및 부출표목에서의 에러

기본기입과 부출기입(기술표목, 주제명표목)에서의 에러는 탐색자가 데이터베이스에 접근하는데 심각한 영향을 주기 때문에 다른 기술필드에서의 에러보다 더 중요하다. 이는 MARC에서 1xx, 4xx, 7xx에서 나타난다.

K관의 경우 레코드변환시 모든 저자사항은 100으로 하고, 두번째 이하의 저자사

16) *Ibid.*, p. 17.

항은 700으로 하였기 때문에 단체명, 회의명에 대한 구분이 없었으므로 많은 수정이 요구되었다. 그리고 저자의 성과 이름의 배열순서도 수정하였으며, 단체명의 경우 상위단체명이 있을 시 기본요소에 추가시키는 것도 포함되었다.

보기1) 저자명의 배열순서

\$a Karl E. Taeuber → \$a Taeuber, Karl E.

보기2) 상위기관이 있는 단체명

\$a 내무부 → \$a 한국. \$b 내무부

③ 서명저자사항에서의 예러

관청, 관사의 처리, 각 하위필드의 기입시 구두점의 오기 및 누락이 많았다.

④ 발행사항에서의 예러

발행지불명, 발행자불명, 혹은 다른 소스를 통한 정보의 각괄호처리가 누락되어 있었으며, 발행자가 완전한 단어가 있는데도 약어로 된 것이 많아 아주 일반적인 것을 제외하고는 완전한 단어로 수정하였다. 또한 두개 이상의 발행사항에 대한 처리도 하였다.

보기1) 발행지불명, 발행자불명

\$c 1986 → \$a [발행지불명] : \$b [발행자불명], \$c 1986

보기2) 다른 소스를 통한 정보

\$a 서울 : \$b 민음사, \$c 1988

\$a [서울] : \$b 민음사, \$c 1988

보기3) 약 어

\$a Honolulu : \$b EWC, \$c 1992

→ \$a Honolulu : \$b East West Center, \$c 1992

## ⑤ 총서사항에서의 에러

총서사항은 부출이 되므로 검색효율에 영향을 미치지 때문에 연구보고서의 경우 '연구보고:2' 또는 '기본보고 IV'와 같이 특정한 의미가 없는 총서명을 총서로 인정하는가에 대한 기준설정이 먼저 요구된다.

K관의 경우 의미에 관계없이 모든 내용을 총서사항으로 인정하여 입력하였다.

## ⑥ 주기사항에서의 에러

주기사항에서의 대부분의 에러는 지금까지 별로 중요치 않게 다루어졌으며, MARC형식에 비추어 볼 때 이에 없거나 필요한 주기사항이 누락된 경우가 많았다. 더욱이 주기사항은 검색에 영향을 거의 미치지 않기 때문에(원저자명, 원서명주기는 부출되어 검색시 접근점이 됨) 다른 필드보다 특히 고려되지 않았다. 그러나 전문적인 자료를 이용하는 이용자들일수록 자료에 대한 더욱 상세하고 구체적인 내용을 요구하는 경향이 많아 앞으로 집중적으로 보강되어야 할 부분이다.

K관의 경우 자료의 일반적인 주기, 학위논문주기, 서지내력주기, 내용주기, 초록주기, 사용제한주기, 원서명, 원저자명에 대한 주기를 목록규칙에 의거하여 입력하는 것을 원칙으로 하여 작업하였다.

## ⑦ 로컬표목에서의 에러

로컬필드는 자관의 사정에 따라 기술할 필요가 있는 사항을 전개하기 위해 사용하는 필드로, K관의 경우 검색시 모든 언어로의 접근점을 제공하기 위해 로컬표목을 다음과 같이 확장하여 수정하였다.

- 보기1) 100 \$a 大竹英雄  
 245 \$a ……/\$b 大竹英雄  
 900 \$a 오오다케 히데오  
 900 \$a おおだけひでお



### 3. 주제관련에러

주제관련에러는 주제명표목과 분류번호와 관련있는 것으로, 주제명표목을 부여하는 데는 문헌내용에 대해 정확성을 기해야 하나, 전문적인 훈련에 관계없이 목록자마다 다르게 나타날 수도 있는 부분이다. 주제명표목에서의 실질적인 에러는 문헌내용과 아주 다르거나 잘못된 용어를 사용한 경우나 주제명표목의 누락 등이며, 분류번호에러는 아주 다른 분류번호를 사용한 경우나 분류번호의 누락 등으로 나타난다. 그러나 엄밀히 말해 이 부분의 누락은 서지입력표준의 위반은 아니다. 왜냐하면, 이 부분은 목록자의 주관적인 경향이 많이 나타나는 것으로 어떠한 기준이나 규칙이 없기 때문이다.

### 4. 편집상 에러

편집상에러는 레코드 입력시 단순한 실수나 착오로 나타날 경우가 많은 부분으로서 철자, 중복, 대문자에 대한 에러이다. 저자필드, 서명필드에 나타나는 철자에러는 검색에 많은 영향을 끼쳐 중요하게 고려되어야 하는 반면 초록필드에서의 철자에러는 검색에 영향을 주지 않는다. 대문자의 에러는 꼭 대문자가 입력되어야 하는 부분에 소문자로 입력되거나, 소문자가 있어야 할 부분에 대문자가 입력된 것을 말하나 검색시에는 별로 영향을 받지 않는다.

철자에러는 데이터베이스에서 가장 나타나기 쉬운 것으로서, 특히 데이터베이스의 이용자에 의해 쉽게 인식된다. 보우네(Bourne)는 11개의 데이터베이스의 색인 용어중 철자에러를 조사하여, 에러가 0.5%(BIOSIS)에서 23%(ABI/INFORM)까지 나타나는 것을 보여주었다.<sup>17)</sup> 다머로(Damerou)는 이러한 철자에러는 누락

17) Charles P. Bourne, "Frequency and Impact of Spelling Errors in Bibliographic Data Bases," *Information Processing & Management*, Vol.13, No.1(1977), pp.1-12.

(omission), 삽입(insertion), 대체(substitution), 전위(transposition) 등 보통 4가지 형태로 나누고, 이러한 형태의 에러가 철자에러의 80%에 이른다고 조사하였다.<sup>18)</sup> 여기서 누락은 한 단어에서 한 문자를 빠트린 것이고(great가 gret로 된 경우), 삽입은 한 단어에서 한 문자를 더 추가한 것이고(great가 greaet로 된 경우), 대체는 한문자가 다른 문자로 바뀐 것이고(great가 gteat로 된 경우), 전위는 두개의 인접문자가 서로 뒤바뀐 것(great가 graet로 된 경우)을 의미한다.

또한 철자에러중 미톤(Mitton)은 잘못철자된 단어가 또다른 단어로 될 수 있는 ‘실제단어(real word)’에러와, 한 단어가 두단어로 분할되거나 두단어가 한 단어로 된 ‘단어분할(word division)’에러를 조사하였다.<sup>19)</sup> 예를 들어 실제단어에러는 ‘form’이 잘못철자되어 ‘from’으로 또다른 단어로 오해될 수 있는 것이고, 단어분할에러는 ‘in side’ 또는 ‘inside’로 서로 상반되어 사용되는 경우이다.

이러한 철자에러는 자동철자체크시스템에 의해 도움을 받을 수 있으며, 특히 데스크톱터, 잡지명, 회사명 등과 같은 필드에서는 일관성을 유지하기 위해 통제어휘 사전을 이용한 자동통제기능을 가지는 것이 일반적으로 이용되고 있다. 아울러 검색시 서명의 우측절단이 빈번하게 사용되므로 단어 중에서도 첫단어의 에러는 두번째나 그 이후의 에러보다 더욱 중요하다.

중복에러는 서지데이터베이스에서 같은 서지항목을 가진 두개 이상의 서지레코드의 발생으로 나타난다. 이러한 중복레코드는 색인성능을 줄이고, 저장공간과 유지에 비용이 들며, 효과적인 탐색을 방해하므로 품질관리에서 또한 중요한 요소이다. 중복에러는 공동목록에서 더욱 두드러지게 나타나며, 중복에러수정은 기존의 서지데이터베이스에서 수정과 신규레코드의 입력시 기존레코드와의 비교를 통해 이루어

18) Fred J. Damerau, "A Technique for Computer Detection and Correction of Spelling Errors," *Communications of the ACM*, Vol.7, No.3(March 1964), pp.171-176. (Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, p.133에서 재인용)

19) Roger Mitton, "Spelling Checkers, Spelling Correctors and the Misspellings of Poor Spellers," *Information Processing & Management*, Vol.23, No.5(1987), pp.495-505.

질 수 있다. 공동목록에서의 중복에러원인은 목록규칙, 서지입력표준, 출판기술의 변화로 인해 주로 나타나는 것으로 알려졌다. 기존서지레코드의 중복수정은 酒井清彦이 NACSIS-CAT(학술정보센타목록시스템)에서 중복레코드의 처리방법에 대해 상세히 소개하였으며,<sup>20)</sup> 신규입력시 중복체크는 중복입력을 배제하기 위해 레코드의 비교대상으로서는 서명, 저자, 출판년도, 페이지수, ISBN, LCCN 등의 선행 비교를 통해 이루어지며 자동적으로 체크된다.

## 5. 초록의 에러

초록은 어떤 특정한 지식기록(연구결과물)의 필수적인 내용만을 요약한 것으로 해당 문헌의 대용물 또는 안내수단으로 사용할 수 있으며, 그 문헌을 읽을 지에 대한 판단요소가 된다.

초록에 구성하는데 기본적인 사항은 다음과 같다.<sup>21)</sup>

- ① 목적, 범위
- ② 방법론
- ③ 결과 또는 성과
- ④ 결론
- ⑤ 어떤 다른 유효하거나 주변적인 정보

초록품질을 판단하는 가장 유용한 기준과 방법은 보코와 베니어(Borko and Bernier)가 만든 다음의 리스트들이다.<sup>22)</sup>

- ① 품질의 전반적인 등급(인간의 판단에 근거)
- ② ANSI나 다른 표준에 근거한 범위

20) 酒井清彦, "オンライン総合目録データベースの重複排除", 情報の管理と技術, Vol.44, No.4 (1994), pp.183-189.

21) Carol Tenopir, "Quality of Abstracts," *Online*, Vol.17, No.4(May 1994), p.48.

22) *Ibid.*, p.50.

- ③ 중요한 정보의 포함과 주요하지 않은 정보의 배제
- ④ 에러의 발생빈도
- ⑤ 형태의 일관성과 읽기쉬움
- ⑥ 관련성의 예측
- ⑦ 원문헌의 대용물로서의 능력
- ⑧ 색인용어의 소스로서의 타당성

물론 이러한 기준은 분명하게 측정하기가 어렵다. 이는 모든 기준들이 주관적이기 때문이며, 아무리 유능한 초록작성사일 경우라도 똑같은 초록이 나오기 어려우며, 포함될 정보의 범위를 선정하는데도 차이가 있을 것이다.

K관의 경우 자체발간 연구보고서의 경우 초록을 수록하고 있으며, 초록작성은 연구보고서를 만든 저자에게 다음과 같은 「초록작성요령」을 따라 작성하도록 권유하고 있다.

- ① 문헌분석후 주요 주제들을 선별, 발췌하여 정확한 말로 문장을 구성하고, 간결하고 명확하게 표현하며, 중복되지 않게 하며, 글자수는 약 250단어 내외로 할 것.
- ② 첫문장은 초록평가에 중요한 기점이므로 논제에 대해 정확히 쓸것
- ③ 상식적인 내용은 제외
- ④ 독자의 관심사항 고려
- ⑤ 주관적 해석이나 비평 제외
- ⑥ 구조는 통일되고 논리적으로 전개
- ⑦ 가급적 원문의 대용물이 되도록 할 것

그러나, 위의 초록작성요령의 제시에도 불구하고 초록의 내용을 전반적으로 평가해보면 초록작성가로서의 저자가 가장 최선의 초록작성가라고는 할 수 없는 것으로 나타났다. 구체적으로 산술적인 평가한 것은 아니지만 위의 기준을 근거로 초록을 분석해 보면, 연구자들은 먼저 초록의 개념을 정확히 파악하고 있지 못하며 입수된

위의 사항이 누락되었을 뿐만 아니라, 필요없는 사항이 많이 들어간 관계로 초록이 연구보고서의 개요나 서론과 같은 수준으로 많이 나타났다.

#### IV. 서지데이터베이스 품질관리 방법

##### 1. 수동방법

품질관리방법은 사람이 직접 하는 것과 기계에 의한 방법이 있다. 자동방법은 수동방법에서의 불충분한 품질관리를 보충하는데 큰 역할을 한다. 오네일(O'Neill)은 수동적인 품질관리방법의 구성요소로서 교육(education), 검토(review), 이용자의 협조(user assistance)를 들었다.<sup>23)</sup> 첫째, 서지유틸리티를 위한 이용자 교육과 훈련은 데이터베이스에 있는 데이터의 품질을 관리하기 위한 가장 기본적인 방법이다. 최소한, 이용자는 데이터요소와 포맷의 기술과 입력표준과 절차에 대한 분명하고 적당한 정보를 가지고 있어야 한다. 또한 교육훈련과 밀접하게 관련된 것은 도큐멘테이션(documentation)이다. 데이터품질이 낮은 것을 분석해보면 종종 도큐멘테이션의 불충분하거나 잘못된 경우가 있다. 효과적으로 데이터베이스를 지원하기 위해 도큐멘테이션은 최신이며, 분명하고 완전해야 한다. 둘째, 자동탐색 절차가 있건없건간에, 교정은 데이터가 데이터베이스에 넣기 전에 많은 에러를 제거할 기회를 제공한다. 셋째, 포괄적인 데이터베이스 품질노력은 데이터베이스 이용자가 에러를 보고하고 더 나아가 OCLC의 경우와 같이 직접 레코드를 부가하거나 수정하는 것이다.

23) Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, p. 130.

## 2. 자동방법

### 1) 자체체크데이터(Self-Checking Data)의 이용

자체체크데이터의 이용은 다른 방법보다 상대적으로 간단하다. 이 방법은 에러가 어디 있는지는 알 수 없으나, 주로 여분의 문자를 데이터에 부가하여 에러의 유무를 검출하는데 이용하는 것으로서, 통신프로토콜의 에러검출기술과 비슷한 방법으로 사용한다. 통신상의 에러검출방식중 패리티비트(parity bit)는 각 단어의 끝에 붙여 사용하는 것인데, 그 예로서 ASCII전송에 있어서 각 7비트 ASCII문자에 하나의 패리티비트를 붙여 패리티비트값은 단어의 갯수가 짝수이거나 홀수가 되도록 선 정하여 전송되도록 한 것이다.

자체체크데이터의 첫번째 서지적이용의 하나는 1960년대 CODEN 코드의 체크 문자였다.<sup>24)</sup> 체크문자는 또한 ISBN, ISSN에 성공적으로 사용되었다. ISBN의 경우 10자리 숫자중 첫 9자리는 책의 고유번호이고, 마지막자리는 ISBN의 검증에 관한 자동체크를 제공한다. ISBN의 경우 각 문헌에 유일하게 있는 숫자를 확인하게 위해 10개의 숫자가 있는데, 마지막 숫자는 앞의 숫자들이 올바르게 되었는가를 확인하는 것으로 앞의 패리티비트와 같은 개념으로 사용되었다.

### 2) 자동데이터검증(Automated Data Validation)

자동검증(즉 데이터요소의 에러체크와 수정)은 데이터베이스품질을 보장하는 효과적이고 바람직한 기법이라고 알려지고 있다.

소여와 다비스(Sawyer & Davis)는 그 예로서 MARC 레코드를 이용하여 내용 지시자의 허용값, 레코드내의 MARC태그와 필드내의 서브필드의 타당한 존재패턴, 다른필드(고정장과 가변장 필드간의 날짜) 간에 일치성을 포함하여 자동적으로

24) *Ibid.*, pp. 131-132.

검출될 수 있는 에러종류를 포괄적으로 논의하고, 많은 경우에서 옳은 값이나 디폴트값이 레코드나 외부테이블값(external tables of values)에 근거해 자동적으로 제공될 수 있음을 보여주었다.<sup>25)</sup>

자동데이터검증방법중 최근에 많이 연구되고 있는 분야는 전문가시스템의 적용이다. 이것은 데이터가 어떤 요구를 만족하는지 결정하고, 검증테스트를 실패한 데이터를 위한 부가와 수정을 제시하기 위해 데이터가 입력될 때 몇가지 지능을 데이터 검토단계에 적용시키는 것이다.

최근 도서관업무영역에 전문가시스템의 적용에 대한 연구가 많았는데, 특히 편목 부문은 목록규칙과 MARC포맷에서 상당한 표준화가 이루어졌기 때문에 전문가시스템의 적용에 더욱 적절한 영역으로 인식되고 있다. 이러한 시스템은 인터페이스를 통해 편목과정에서 도움을 주는 것과, 더욱 진보된 형태로서 스캐너를 이용하여 형식과 문자를 인식한 다음 타이틀페이지를 해석하여 자동으로 편목하도록 하는 방법이 있다. 특히 후자의 경우, 표제지의 자동인식기능이 완벽하게 이루어지고, 완벽한 자동색인과 그에 연계된 전자통제시스템의 진보가 있으면 인간목록자의 역할의 거의 대부분을 대체할 수 있을 것으로 여겨진다.

### 3) 지원시스템이용

데이터품질관리를 위한 지원시스템의 이용은 다양하게 적용될 수 있는데, 이것은 앞서 논의한 전문가시스템에 지능적인 업무를 가능케 하는 기반시스템이기도 하다. 예를 들어 일본과학기술정보센터는 JICST파일의 품질을 유지하고 향상시키기 위해 데이터베이스 작성과정과 관련해 각 과정에서 품질관리지원시스템을 활용하고 있

25) Jeanne C. Sawyer and Jinnie Y. Davis, "Automated Error Detection in Library Systems," In Michael Gorman, ed., *Crossroads: Proceedings of the Library and Information Technology Association 1st National Conference; 1983 September 17-21; MD.* Chicago, IL, American Library Association, 1984, pp.213-217. (Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, p.132에서 재인용)

며, 특히 서지데이터베이스 작성시 초록/색인지원시스템, 공정관리시스템, 서지데이터베이스작성시스템, 디스플레이조정시스템 등이 지원되고 있다. 여기에 덧붙여 지원사전시스템으로서는 일본어자동화키워드시스템, 용어관리시스템, JICST분류표작성시스템, 디소러스변환데이터베이스, 일본어화합물사전작성시스템 등이 지원되어 사용된다.<sup>26)</sup> 또한 더 나은 품질유지향상을 위해 이용자의 지적사항을 즉각 수용할 수 있도록 하며, 오지수정시스템을 운영하고 각종지원사전을 유지관리하며, 기존의 데이터베이스를 재수정하여 최신의 품질을 유지하도록 노력하고 있다.

지원시스템중 사전의 구축은 지원시스템을 보다 효율적으로 이용할 수 있는 가장 중요한 기반이라 하겠다.

#### 4) 전거통제

전거통제는 전거업무과정과 각 파일의 일관성을 유지하기 위해 전거데이터의 이용을 포함하는 것으로서 지원시스템의 한 형태로 볼 수 있으며, 서지데이터베이스의 품질을 개선하는데 도움을 준다. 예를 들어 전거파일은 표목에서 철자와 인쇄상의 에러를 수정하는데 특수한 목적의 사전으로서 사용될 수 있다. 철자와 인쇄상의 에러를 검출하고 수정하는 모든 일반적인 방법은 전거파일에 적용할 수 있으며, 일반사전이나 시소러스 보다 더 나은 결과를 가져올 수 있다.

특히 전거통제는 전거데이터의 MARC포맷개발과 LC의 기계가독전거레코드외 배포로 자동전거통제시스템의 개발을 고무시켰으며, 이와 함께 전거통제시스템 자체에서도 자동편집기능, 전거파일의 자동발생 및 유지기능, 철자에러의 자동검출 및 수정을 포함해서 연구되고 있다. 전거통제시스템의 이용사례는 다음과 같다.<sup>27)</sup>

OCLC는 'LC저자명 및 주제명 전거파일'을 독립적으로 사용하고 있으며 온라인

26) 高野文雄, "JICSTファイルの品質管理支援システム", 情報の管理と技術, Vol. 44, No. 4 (1994), pp. 168-175.

27) Edward T. O'Neill and Diane Vizine-Goetz, *op. cit.*, pp.141-144.



으로 탐색을 제공한다. OCLC시스템에 서지레코드가 넣어질 때 표목이 전자파일에 대해 자동적으로 체크되지 않으나, 저자명표목의 배치변환으로 데이터베이스의 일관성을 유지한다. 이 파일은 OLC(Online Union Catalog)의 서지레코드로 연결되지 않았으나, 많은 회원도서관들이 레코드의 일관성을 유지하기 위해 사용하고 있다.

WLN은 통합된 서지파일과 전자파일을 유지하며 서지레코드에서 표목의 자동은 라인검증기능을 제공한다. WLN의 전자통제시스템은 MARC태그, 하위필드코드, 지시자의 자동체크와 모든 표목의 자동검증을 포함한다. 이 시스템은 서지레코드와 전자파일이 연결되었기 때문에 전자파일이 변화가 있을때 쉽게 서지레코드가 갱신된다.

## V. 결 론

데이터베이스산업은 혁신적인 기술발전과 시간과 장소를 가리지 않도록 한 통신 기술의 발달, 대량의 정보를 저장할 수 있는 매체의 개발, 많은 관심을 가진 사회 구성원의 지속적인 관심과 개발로 인해 놀랄만한 발전을 이루어 왔다. 그러나 이제 데이터베이스시스템의 외형적인 성장과 관련기술의 진보뿐만 아니라 동시에 데이터베이스의 품질관리에 대한 연구도 수행되어야 할 것이다. 데이터베이스의 품질관리란 데이터의 품질을 분석개선하여 이용자의 요구에 더욱 부응하는 것이다.

지금까지 서지데이터베이스의 경우 다른 수치 및 전문데이터베이스에 비해 어려가 발생하더라도 사회적인 책임성을 가지는데 인색하였는지 모르나, 전반적으로 많은 서지데이터베이스의 이용증가에 비취 볼 때 결코 간과할 수 없는 부분이라 하겠다. 특히 국내의 경우 MARC레코드로의 전환이 90년대에 와서 비로서 일반화 되어 비록 MARC레코드로 구성되었다하나 많은 레코드들이 품질이 낮음을 부인하

지 않을 수 없다. 아울러, 서지데이터베이스의 이용이 급성장함과 동시에 이용자의 수준도 상당히 높아진 만큼 그 요구를 충분히 만족시키지 못하는 경우도 나타났다. 따라서 이제는 서지데이터베이스의 품질관리에 가능한 모든 방법을 이용하여야 할 것이다.

K관의 서지레코드의 분석과 서지데이터베이스의 품질관리에 관한 전반적인 고찰을 통해 다음과 같은 사항들이 알 수 있다.

① 서지데이터베이스의 레코드는 목록기술관련, 주제관련, 편집과 관련한 다양한 종류의 에러를 가지고 있다.

② 이러한 에러는 검색효율과 밀접한 관련을 가지고 있다.

③ 데이터베이스의 품질관리의 척도는 이용자의 만족도에 근거한다.

그리고, 지금까지의 분석결과 다음과 같은 사항이 앞으로 서지데이터베이스의 구축에 고려되어야 한다.

① 목록지는 정확하고 높은 품질의 편목교육을 받아야 하며, 편목시 목록규칙적 용을 잘 준수해야 한다.

② 각종 사전의 구축과 상황에 맞는 지원시스템을 개발하여 서지데이터베이스 시스템과 연계해서 이용해야 한다.

③ 적절한 도구를 사용하여 지속적인 품질관리가 이루어지도록 해야 한다.

④ 궁극적으로 능동적인 데이터베이스 품질관리시스템이 제공되어야 한다. 이 시스템은 데이터 자체에 대한 전반적인 체크뿐만 아니라 데이터내용에 대한 전문가 수준의 검토를 수행하고, 자동적인 에러체크기능을 갖추고 있어야 한다.

參 考 文 獻

- 大韓. 서울商工會議所. 신경영기법, 한국에서 유행인가 새로운 물결인가?. 서울 :  
大韓. 서울商工會議所, 1994.
- 東亞出版社. 東亞原色世界大百科事典. 서울 : 동아출판사, 1989.
- 高野文雄. "JICSTフルイルの品質管理支援システムネ.". 情報の管理と技術.  
Vol. 44, No. 4 (1994). pp. 168-175.
- 根岸正光. "データベースの運用管理." 情報の科學と技術. Vol. 44, No. 4  
(1994). p. 155-161.
- 三輪眞木子. "データベースの品質評價 : サーチャーノ批評的檢索  
テクニク.". 情報の科學と技術. Vol. 44, No. 4 (1994). pp. 196-203.
- 酒井清彦. "オンライン總合目錄データベースの重複排除.". 情報の管理と技  
術. Vol. 44, No. 4 (1994). pp. 183-189.
- Bland, Robert. "Quality Control in a Shared Online Catalog Database :  
The Lambda Experience." *Technical Services Quarterly*. Vol. 4, No. 2  
(Winter 1986). pp. 43-58.
- Bourne, Charles P. "Frequency and Impact of Spelling Errors in  
Bibliographic Data Bases." *Information Processing & Management*. Vol. 13,  
No. 1 (1977). pp. 1-12.
- Daniel, Evelyn H. "Quality Control of Documents." *Library Trends*. Vol. 41,  
No. 4 (Spring 1993). pp. 644-664.
- Intner, Shelia S. "Quality in Bibliographic Databases : An Analysis of  
Member-Contributed Cataloging in OCLC and RLIN." In McCabe,  
Gerard B. and Kreissman, Bernard, ed., *Advances in Library*

- Administration and Organization : A Research Annual*. Greenwich, Connecticut : JAI Press, Inc., 1989. pp.1-24.
- Mintz, Anne P. "Quality Control and the Zen of Database Production." *Online* Vol.14, No.6 (November 1990). pp.15-23.
- Mitton, Roger. "Spelling checkers, Spelling Correctors and the Misspellings of Poor Spellers." *Information Processing & Management*. Vol.23, No.5(1987). pp.495-505.
- Ojala, Marydee. "Quality Online and Online Quality." *Online*. Vol.16, No.1 (January 1992). pp.73-75.
- . "A Model for Quality Business Searching." *Online*. Vol.17, No.3 (March 1993). pp.78-79.
- O'Neill, Edward T. and Vazine-Goetz, Diane, "Quality Control in Online Database." *ARIST*. Vol.23(1988). pp.125-156.
- Tenopir, Carol. "Quality of Abstracts." *Online*. Vol.17, No. 4 (May 1993). pp.44-55.
- Zeng, Lei. "Quality Control of Chinese-Language Records Using a Rule-Based Data Validation System. Part 1 : An Evaluation of the Quality of Chinese-Language Records in the OCLC OCLU Database." *Cataloging & Classification Quarterly*. Vol.16, No.4(1993). pp.25-66.

## A Quality control of bibliographic database

Ji-hoon Kim\*

According to database and information technology development, many interests of database quality control have being increase. The purpose of database quality control is improvement quality of data itself as well as database system to satisfy user's need.

As this paper was especially written about quality control of bibliographic database, to embody complete bibliographic database, it was invested numerous errors and its case by analyzing MARC records.

In addition, it was presented that high degree's cataloging education, introduction of supporting systems, and development of intelligent quality control system for quality improvement.

---

\* Research Fellow, Korea Research Institute For Human Settlements