

종합목록 데이터의 오류 유형에 관한 연구

- KERIS 종합목록의 학위논문 서지데이터를 중심으로 -

A Study on Error Data Types in the KERIS Union Catalog

- Focused on Dissertation Bibliographic Database -

조 순 영(Sun-Yeong Cho)

초 록

KERIS의 대학도서관 종합목록은 330개 대학도서관이 가입하여 570만 레코드를 보유하고 있는 국내 최대의 목록이다. 그러나 각 회원기관이 이미 구축한 DB를 짧은 기간 내 통합하면서 중복 및 오류 레코드가 많이 포함되어 있기 때문에 검색의 효율성이 떨어진다는 지적이 있다. 따라서 본 연구에서는 전체 자료의 10%를 차지하는 학위논문 데이터 1000건을 대상으로 오류 데이터의 유형을 분석함으로써 데이터의 품질을 측정하고 기계적으로 오류데이터를 색출할 수 있는 방안을 제시하였다. 분석 결과 오류데이터는 전체 표본 DB의 30%를 차지하였고 발생하는 주요 원인은 1)입력오류 2)MARC의 사용 오류 3)목록규칙의 적용 오류가 가장 큰 원인으로 나타났다.

ABSTRACTS

The KERIS Union Catalog is the largest bibliographic database in Korea. It has over 5.7 million bibliographic records and at present 330 university libraries are participating in shared cataloging services. The database, however, includes various errors and a large number of duplicate data because separate databases from many university libraries were merged without proper quality control in a short period. This study intends to find error data types by analyzing dissertation bibliographic data. The results show that error data are mainly caused by input errors, and the mistakes in using MARC formats and cataloging rules.

키워드: 종합목록, 오류데이터, 중복데이터

한국교육학술정보원(KERIS) 학술연구정보화실장(chosy@keris.or.kr)

논문접수일자 2002년 08월 25일

게재확정일자 2002년 11월 10일

참고문헌

『2001년도 종합목록서비스 운영현황보고서』. 서울: 한국교육학술정보원, 2001

국립중앙도서관 편. 1993. 『한국문헌자동화목록법형식: 단행본용』. 서울: 국립중앙도서관, 1993

김지훈. 서지데이터베이스의 품질관리-k관의 MARC 레코드 분석을 중심으로 『도서관학논집』, 제21집 (1994): 401-429

이지은. 1999. 『첨단학술정보센터 종합목록데이터베이스 품질관리에 관한 연구』. 석사학위논문. 숙명여자대학교 대학원. 문헌정보학과.

OCLC. 1990. Bibliographic input standards. 4th ed. Dublin, Ohio: OCLC.

Cousins, S. A. "Duplicate detection and record consolidation in large bibliographic databases: the COPAC database experience in Great Britain." Journal of Information Science" v. 24 no4 (1998) p. 231-40

Determining Duplicate Records: Kinetica Duplicates Guidelines for Monographs.

<<http://www.nla.gov.au/kinetica/detduplicates.html>>

OCLC. Bibliographic Formats and Standards: Chapter 4. When to Input a New Record.

<<http://www.oclc.org/oclc/bib/chap4.htm>>

_____. Bibliographic Formats and Standards: Chapter 5. Quality Assurance / OCLC

<<http://www.oclc.org/oclc/bib/chap5.htm>>

_____. OCLC Cataloging Service User Guide. 3 ed.

<<http://www.oclc.org/oclc/cataloging/guide/frameset.htm>>

O'Neill, E. A. "Duplicate detection". Annual Review of OCLC Research. (1988/89): 15-16, (1989/90): 13-14

____, Sally A. Rogers, and W. Michael Oskins. "Characteristics of duplicate records in OCLC's online union catalog. LRTS, 37(1): 59-67

Ridley, M. J. "An Expert system for quality control and duplicate detection in bibliographic databases" . Program 26(1) 1992: 1-18

Rittberger, M. Rittberger, W. "Measuring quality in the production of data-bases" . Journal of Information Science, 23(1): 25-37

K C I