

1997

Improve OPAC Searching by Reducing Tagging Errors in MARC Records

Terry Ballard

Gill Library, College of New Rochelle, tballard@cnr.edu

Follow this and additional works at: <http://digitalcommons.cnr.edu/gill-publications>

 Part of the [Library and Information Science Commons](#)

Recommended Citation

Ballard, T. (1997). Improve OPAC searching by reducing tagging errors in MARC records. *Library software review*, 16(3), 152-155.

This Article is brought to you for free and open access by the Gill Library at Digital Commons @ CNR. It has been accepted for inclusion in Gill Library Publications by an authorized administrator of Digital Commons @ CNR. For more information, please contact lfazzino@cnr.edu.

Improve OPAC Searching by Reducing Tagging Errors in MARC Records

Terry Ballard and Anthony Grimaldi

One of the most common errors in cataloging library materials involves miscoding of the nonfiling indicator of title fields. This article notes the extent of the problem and its negative effect on searching in the library's on-line catalog and surveys how librarians have approached solutions to the problem. The article describes how the major library automation systems address this problem.

Statement of the Problem

A tour of on-line catalogs through the Internet will confirm that the problem of tagging errors in OPACs is widespread. These can be located by making a title browse search for the command "the." Books that are mistakenly indexed by the initial article will then display on the browse screens (see Figure 1).

Searching for some of these records by their correct titles causes the OPAC to mistakenly tell the searcher that the library does not own that title (see Figure 2).

This problem is a result of a bad value in the second indicator of the 245 field, which specifies the number of characters to ignore in indexing. If it is set to 0 when there is an initial article, the record has been mistagged (see Figure 3).

Differences Between Systems in the Use of Full Title Search

A study of searches performed at a university in New York using a PALS OPAC in the summer of 1995 examined 723,279 searches performed. Of these, 61,436, or 8.4 percent, of the searches were for exact title. A similar study on a PALS system was reported by Flaherty (1993). She found that 15 percent of searches in the Mankato State University system were for exact title. Given the relatively light usage of title browse searches in PALS, the initial assumption might be that mistagged 245 fields are not a large problem. However, one of these libraries produced a microfiche backup of their title index and found that the hundreds of mistagged 245 records were essentially lost records.

For a full account of search preferences in different systems, a good starting point is volume 42 of *Library Hi-Tech* (1993), which was largely dedicated to transaction log analysis. Zink's (1993) article in this volume measured search patterns in INNOPAC. Ballard's (1994) study of staff-only versus public terminal searching demonstrated that title searching accounts for nearly 50 percent of staff inquiries in that INNOPAC system.

Identifying the Problems

Telnetting to randomly selected OPACs employing NOTIS, PALS, INNOPAC, DRA, and DYNIX was used to gather

You searched for the TITLE: the

710 TITLES found, with 990 entries; TITLES 1-8 are:

1	The Art Of Cross Examination	1 entry
2	The Bulletin Of The Institute Of Social ...	1 entry
3	The Diminished Nation	1 entry
4	The East Asian Miracle	1 entry
5	The Fact Book	1 entry
6	The International Law On Trafficking.....	1 entry
7	The Jurisprudence Of Public Choice	1 entry
8	The Legal Resources Approach	1 entry

Figure 1. Example INNOPAC search for items beginning with "the."

You searched for the TITLE: jurisprudence of public choic
Your TITLE not found, Nearby TITLES are:

1	Jurisprudence Of Medicine	1 entry
2	Jurisprudence Of My Time	1 entry
3	Jurisprudence Of Orthodoxy	1 entry
4	Jurisprudence Of Prophylactic Rules	1 entry
Your entry Jurisprudence Of Public Choice would be here		
5	Jurisprudence Of The General Claims Comm ..	2 entries
6	Jurisprudence Of The Jewish Courts	2 entries

Please type the NUMBER of the item you want to see, OR

F > Go FORWARD

N > NEW Search

B > Go BACKWARD

A > ANOTHER Search by

TITLE

W > Same search as WORD search

+ > ADDITIONAL options

Choose one (1-8, F, B, W, N, A, P, D, T, J, M, +)

Figure 2. Errors in the title field can lead to incorrect search results.

samples of tagging errors in those systems. On-site visits were made to university libraries using these systems. A dozen academic sites around the country were contacted to augment the collection of data and probe for further information.

On-site visits provided anecdotal evidence regarding the efficacy of these systems from the point of view of catalogers and technical services librarians in academic libraries. NOTIS users seemed to be the least satisfied with their

OPACs; INNOPAC users seemed to be the most pleased. PALS and DRA users were somewhere in the middle, expressing neither strong dislike nor great passion for their systems. When tagging errors in the 245 field were pointed out to these librarians, it was found that the majority of catalogers were not enthusiastic about the prospects of systematically correcting their MARC records. The notion of cleaning up all of the mistagged titles beginning with the word "the," as a case in point, interested only two of the

```

You searched for the TITLE: the
100 1 Spodick, David H.,|d1927-
245 10 The pericardium :|ba comprehensive textbook /|cDavid H.
Spodick
260 New York :|bM. Dekker,|cc1997
263 9610
300 x, 464 p. :|bill. ; 26 cm
440 0 Fundamental and clinical cardiology ;|vv. 27
500 Includes index
650 12 Pericardium
650 22 Pericarditis
650 0 Pericardium|xDiseases

```

Figure 3. Example of incorrect coding in indicators for 245 field.

catalogers. Having as perfect a database as possible—with all of the troubles that such perfection entails—did not seem to be a top priority here.

The following questions were put to the technical librarians and/or catalogers in connection with the systems they were using—that is, DYNIX, NOTIS, PALS, INNOPAC, and DRA:

1. I am researching tagging errors/quality control in your on-line catalog system. How are errors detected when you do on-line cataloging? How do you systematically find errors?
2. How are errors corrected? What do you do in the way of quality control?
3. Can you give me a printout of corrections made?
4. What percentage of the searches are title, subject, term, and so on?
5. Can you think of any questions that I should have asked you? Is there anything that you would like to add?

The DYNIX, PALS, and DRA librarians all responded to the first question by saying that they detected tagging errors by chance. They did not deliberately seek to discover mistakes in the 245 field. The NOTIS librarian mentioned that his department was extremely shorthanded, so he felt that it was the responsibility of the public services librarians to point out mistakes for correction.

There was more variety in the answers to the questions about how errors are corrected and quality control. The DYNIX librarian claimed that DYNIX sometimes balks on maneuvers. For example, for a video recording of *The Merchant of Venice*, it was impossible to access the H-subfield

to indicate its media designation in the 245 field. Ergo, the librarian “tricked” the system by including information in the B-subfield as a “line parallel to the title.” Quality control was addressed by downloading records mostly from OCLC and ARLIN. DYNIX handles the on-line indications, and there is also a 949 field to accommodate this process.

The NOTIS librarian seemed to indicate that modifying records with NOTIS software is cumbersome. Tagging errors are unearthed by trial and error, and the backlog of errors (given the fact that this library imported vast quantities of MARC records from sundry other libraries) discouraged cleaning up the database.

The PALS librarian said that errors are corrected in the database as soon as they are brought to his attention. Otherwise, he does not go out of his way to find mistakes. Usually, he was sensitive to problems in the 740 field in which the indicator fields 01 and 41 are confused.

Two DRA sites were visited. In the first, at least one paralibrarian is assigned the task of screening fields in the MARC record—especially the 245 field—for errors. This activity is built into her job description. Moreover, the technical services librarian takes an aggressive position about keeping in touch with his database and ferreting out errors. For example, *Cataloging Services Bulletin* is checked each month to anticipate trends and head off problems.

The other DRA site, like the NOTIS site, inherited a large number of problem records. The librarian said that her staff exercises control over scanning MARC records but that they were sometimes reluctant to remove errors. “Why bother?” “Who cares?” This librarian, however, had a special fondness for media and, thus, took special interest in monitoring the 590 fields of the MARC records.

All of the above expressed a strong interest in making global changes in their MARC records, but the software technology does not seem to be quite there yet. All of these librarians were able to furnish printouts of both tagging errors in the 245 field and the subsequent corrections that they had implemented. Not one of these librarians, except for the INNOPAC and PALS librarians, was able to furnish percentage breakdowns of how their on-line catalogs were searched in terms of keyword, title, subject term, subject searching, and so on.

One DRA librarian allowed that he ran a verification against the authority record. Every new record is matched against the authority file. Originally, he tossed this operation out to a vendor, but this did not prove to be satisfactory. He stressed the importance, instead, in keeping in touch with one's own database, verifying heading by heading as it came in. Every series heading is verified too.

Conclusion

Tagging errors can hurt search performance, and they can lead to missing records if microfiche backups are run. The severity of the problem is a function of the number of mistagged records versus the likelihood that a user will search by title. Vendors should work to make these problems fixable in rapid update mode. None of the systems observed in this study had that capability, and no evidence was found that any system allows rapid updates of tagging errors;

hence, often, the decision to fix them slows down in the political process of finding staff time to fix individual problems.

References

- Ballard, Terry. 1994. Comparative searching styles of patrons and staff. *Library Resources and Technical Services* 38 (3): 293-305.
- Flaherty, Patricia. 1993. An investigation into Zero-hit searches: A translation log study. Speech given at the LITA Online Catalog Interest Group session at the American Library Association midwinter meeting, 24 January, Denver, CO.
- Zink, Steven, D. 1993. Monitoring user search success through transaction log analysis: The WolfPac example. *Library Hi-Tech* 42: 49-56.

About the Authors

Terry Ballard is the library automation coordinator at the New York University School of Law and an adjunct professor at the Palmer Library School at Long Island University. He is the author of INNOPAC: A Reference Guide to the System. He can be reached by e-mail at ballardt@turing.law.nyu.edu, and his web page is at <http://geocities.com/Athens/Delphi/3632>. Anthony Grimaldi is an architecture librarian at the New Jersey Institute of Technology, Newark.