OCLC Search Key Usage Patterns in a Large Research Library

Kunj B. RASTOGI: OCLC; and Ichiko T. MORITA: Ohio State University, Columbus.

Many libraries use the OCLC Online Union Catalog and Shared Cataloging Subsystem to perform various library functions, such as acquisitions and cataloging of library materials. As an initial part of the operations, users must search and retrieve a bibliographic record for the desired item from the large OCLC database. Various types of derived search keys are available for retrieval. This study of actual search keys entered by users of the OCLC online system was conducted to determine the types of search keys users prefer for performing various library operations and to find out whether the preferred search keys are effective.

INTRODUCTION

In the last decade, many information systems have been developed that use search keys to retrieve bibliographic records from large databases. The OCLC Online Union Catalog and Shared Cataloging Subsystem in particular is one of the larger of these systems.[1-6] There are currently more than 7 million bibliographic records in the OCLC database.

The OCLC online system uses search keys to access various index files that locate bibliographic records in the database. Index files are maintained for name/title, personal author, corporate author, CODEN, ISBN, and LCCN indexes. The first four of the above index files contain search keys that are derived from information (e.g., author, title) present in the piece or citation. Search keys in these four indexes are in general not unique, because the derived key could be the same for different bibliographic records. The last three indexes (CODEN, ISBN, and LCCN) contain search keys or identifiers that are unique in general. A user enters a search key consisting of characters (letters, numbers, symbols, commas, hyphens) formatted according to specific rules that identify to the system which index file to search. For example, to search the name/title index, the user enters a search key consisting of the first four characters of the author's last name and the first four characters of

the first nonarticle word of the title of the work, separated by a comma. To search the title index, the user enters a search key consisting of the first three characters of the first nonarticle word in the title, the first two characters of the second word, the first two characters of the third word, and the first character of the fourth word, each separated by a comma.[7]

The system compares the user-entered search key with the search keys contained in that index file. This comparison results in one of three possible cases:

1. Only one index file search key matches the user-entered search key.
2. More than one index file search key matches the user-entered search key.
3. No index file search key matches the user-entered search key.

In the first case, the system retrieves the unique bibliographic record corresponding to the search key and displays it on the user's terminal screen. In the second case, the system retrieves all records that correspond to the search key, prepares truncated entries (consisting of author, title, imprint data, etc.) for those records, and displays the truncated entries on the user's terminal screen. The user then selects the truncated entry that corresponds to the desired record and requests the system to display the full record for that item. In the third case, the system responds with the reply that a record matching the user-entered search key was not present (a "not found" response) in the index.

In the OCLC online system, 2,500 member libraries using 3,800 terminals search the OCLC database to perform various library functions such as acquisitions, monograph cataloging, and serials cataloging. Users can choose to enter any type of search key from the various types of search keys permitted by the system. Users' preferences to enter a particular type of search key will depend in part upon the kind of information they have about the item to be searched and the type of library function they wish to perform. If users receive a "not found" response after entering a particular type of search key, they may then try a different type of search key that they consider next best.

The purpose of this study was to determine what types of search keys are preferred to perform various library functions and whether the preferred search keys are effective. The study also investigated what type of search key is used next when particular types of search keys are unable to retrieve the desired record to determine if there are any discernible search patterns.

MATERIALS AND METHODS

For conducting this study, data were needed on the pattern of search-key use in OCLC member libraries. Further, the data had to include the actual time of day when work was performed for a particular library

function on a specific terminal. This requirement would permit iden-
tification in the Online System Use Data collected by OCLC of search
keys entered to perform specific library functions. Ideally, a library with
several OCLC terminals, each used exclusively for only one library
function, was desired. The Ohio State University (OSU) Library met
this requirement. The OSU Library has eleven terminals: two of the
eleven terminals are used exclusively for performing acquisition func-
tions, seven are used for monographic cataloging, and one terminal each
is used for serials cataloging and public use. The terminal assigned for
serials cataloging is used for monograph cataloging after 5 p.m. Library
staff at OSU use all the terminals exclusively, except for the public-use
terminal. This public-use terminal can be used by anyone, including
faculty, students, and library staff.

Two full days' transactions for each of the OSU terminals were
obtained from the OCLC Online System Use Statistics (OLSUS) file.
During the online operation, the system writes a record on the OLSUS
file for each message entered by the user. This record includes the in-
stitution number, a number identifying the terminal from which the
message came, the time of the transaction, and the first nonblank six-
teen characters of the message. If the user-entered message is a search
key, the system response is either a "not found" response or a "found"
response.

With the "found" response, the system displays the bibliographic rec-
ord (if unique) or displays a truncated entry screen. However, a "found"
response does not necessarily mean that the truncated entry screen in-
cludes information about the bibliographic record the user was actually
seeking.

For the study, a program was written to scan the records in the
OLSUS file for two full days in October 1978. The program extracted all
the records for messages that came from the eleven OSU terminals and
wrote the records on two tapes—one for each day's activity. These tapes
were sorted first by the terminal number and then within each terminal
number by the time of transaction. Each sorted tape was fed to another
program that printed, for each terminal, the actual messages in chrono-
logical order and the associated system response.

From this printout, it was possible manually to go through the com-
plete sequence of messages entered to search a single bibliographic
item. The printout for an entire day's activity for each terminal was thus
divided into sections, each section containing all transactions that were
performed to search for a single item. For each section, the type of
search key first entered and the system response was noted. In case of a
"not found" response, the type of search key next entered (if the search
process was continued for the item) also was noted. The results were
combined for all the terminals used to perform a specific library function
(e.g., acquisitions) and for the two days.

RESULTS AND DISCUSSION

Table 1 and figure 1 show the different types of search keys used as the first choice to perform various library functions. Note that at the time of data collection for this study, the Interlibrary Loan Subsystem was not operational.

Table 1. Different Types of Searches for Various Applications

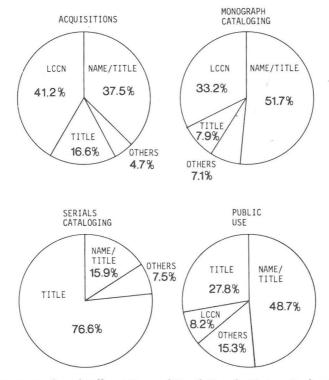| Type of Search	Acquisitions		Monograph Cataloging		Serials Cataloging		Public Use	
	Items Searched	% of Total	Items Searched	% of Total	Items Searched	% of Total	Items Searched	% of Total
Name/Title	111	37.5	313	51.7	15	15.9	77	48.7
Title	49	16.6	48	7.9	72	76.6	44	27.8
Personal Author	0	0.0	9	1.5	0	0.0	16	10.1
LCCN	122	41.2	201	33.2	1	1.1	13	8.2
ISBN	14	4.7	34	5.6	1	1.1	3	1.9
ISSN	0	0.0	0	0.0	5	5.3	3	1.9
CODEN	0	0.0	0	0.0	0	0.0	2	1.3
Total	296	100.0	605	100.0	94	100.0	158	100.0

Fig. 1. Number of Different Types of Search Keys for Various Applications.

During the two-day period, a total of 605 items were searched for monograph cataloging, 296 items were searched for acquisitions operations, and 94 items were searched for serials cataloging. A total of 158 items were searched on the public-use terminal. Most types of search keys were used to some extent. The use of ISBN and ISSN search keys was quite limited for all types of library functions. The CODEN search key was used only twice, and both times through the public-use terminal. The corporate author search key was not used at all. The use of the personal-author search key was much smaller than expected. This was probably because at the time of the study the system did not permit use of personal author keys during peak hours (9 a.m. to 5 p.m.) of online system operation.

For the acquisitions function, the LCCN search key was used most often, followed by the name/title key. These two types of keys together were used for about 80 percent of the acquisitions items searched. For the monograph cataloging function, the most frequently used search key was the name/title key. This key was entered for about 52 percent of items searched. The next most frequently used key for monograph cataloging was the LCCN key, used for about 33 percent of the items searched. For the serials cataloging function, the title key was used most often, for more than 75 percent of the items searched. Searches performed through the public-use terminal included all types of search keys. The name/title key was used most frequently, followed by the title key.

Before performing an actual search, a user must choose, from among the various types of search keys available in the OCLC system, the particular search key to use. If the search key used for a first try (primary choice of search key) results in a "not found" response from the system, a second key may be entered (secondary choice of search key). This sequence may continue through many search-key choices until the user retrieves the desired record ("found" response) or decides to abandon the search at some point upon obtaining a "not found" response. For this study, the investigation was confined to only primary and secondary choices of search keys. The results of the "found" responses for the primary choice of key and for the secondary search key entered after receiving the first "not-found" response are presented in tables 2 through 5.

For the acquisitions function (table 2), the most frequently used primary search key was the LCCN key, which retrieved the desired record about 89 percent of the time. When the LCCN key could not retrieve the record, the user chose mostly the name/title key as his/her secondary choice or abandoned the search. The next most frequently used primary search key was the name/title key, which retrieved the desired record about 51 percent of the time. When the name/title key was unsuccessful, the users entered as their secondary search key a title key

Table 2. Number of Primary and Secondary Choices of Search Keys for Acquisitions

Type of Search Key Used First	Items Searched	Found Responses	% of Found Responses	Not-found Responses	Name/ Title	Title	Personal Author	LCCN	ISBN	Search Discontinued after the First Not-found Response
					Types of Search Key Used after the First Not-found Response					
Name/Title	111	57	51.3%	54	17 (31.5%)	22 (40.7%)	0 (0.0%)	1 (1.9%)	0 (0.0%)	14 (25.9%)
Title	49	17	34.7%	32	6 (18.8%)	11 (34.4%)	0 (0.0%)	2 (6.2%)	1 (3.1%)	12 (37.5%)
Personal Author	0	—	—	—	—	—	—	—	—	—
LCCN	122	109	89.3%	13	5 (38.4%)	1 (7.7%)	0 (0.0%)	2 (15.4%)	1 (7.7%)	4 (30.8%)
ISBN	14	1	7.1%	13	8 (61.5%)	3 (23.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (15.4%)
ISSN	0	—	—	—	—	—	—	—	—	—
CODEN	0	—	—	—	—	—	—	—	—	—
Total	296	184	62.2%	112	36 (32.1%)	37 (33.0%)	0 (0.0%)	5 (4.5%)	2 (1.8%)	32 (28.6%)

Note: To calculate the percentage given in parentheses, the number of "Types of Search Key Used after the First Not-found Response" was divided by the number of "Not-found Responses."

about 41 percent of the time, or a different name/title key about 31 percent of the time. Approximately 26 percent of the time they abandoned the search. It seems that acquisitions users mostly try the LCCN key first if available (the LCCN is not present in all the records) and the name/title key first if the LCCN is not available. Thus, users adopted the right approach since the LCCN key has the highest hit rate. Furthermore, the LCCN key is more efficient than other keys because it results, on the average, in a fewer number of replies.

For the monograph cataloging function (table 3), the name/title key was used most often as the primary search key, resulting in retrieval of the desired record about 57 percent of the time. When the name/title key could not retrieve the record, the users next attempted a title key (52 percent of the time) or a different name/title (21 percent of the time). About 23 percent of the time they discontinued the search. The LCCN key was the second most frequently used primary search key and successfully retrieved the record about 79 percent of the time. When the LCCN key was unsuccessful, the users tried the name/title key (58 percent of the time) as their secondary choice or abandoned the search. Unlike the search-key usage pattern for acquisitions, the use of the LCCN key for monograph cataloging was lower than use of the name/ title key, although here also the hit rate was highest for the LCCN key. The reason the LCCN use was lower is that Ohio State University, being a research institution, processes a large number of items from var-

Table 3. *Number of Primary and Secondary Choices of Search Keys for Monograph Cataloging*

Type of Search Key Used First	Items Searched	Found Responses	% of Found Responses	Not-found Responses	Types of Search Key Used after the First Not-found Response					Search Discontinued after the First Not-found Response
					Name/ Title	Title	Personal Author	LCCN	ISBN	
Name/Title	313	180	57.5%	133	28 (21.1%)	69 (51.9%)	1 (0.7%)	4 (3.0%)	1 (0.7%)	30 (22.6%)
Title	48	24	50.0%	24	9 (37.5%)	2 (8.3%)	1 (4.2%)	3 (12.5%)	2 (8.3%)	7 (29.2%)
Personal Author	9	3	33.3%	6	4 (66.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (16.7%)	1 (16.7%)
LCCN	201	158	78.6%	43	25 (58.1%)	4 (9.3%)	0 (0.0%)	2 (4.7%)	1 (2.3%)	11 (25.6%)
ISBN	34	3	8.8%	31	20 (64.5%)	4 (12.9%)	1 (3.2%)	1 (3.2%)	3 (9.7%)	2 (6.5%)
ISSN	0	—	—	—	—	—	—	—	—	—
CODEN	0	—	—	—	—	—	—	—	—	—
Total	605	368	60.8%	237	86 (36.3%)	79 (33.3%)	3 (1.3%)	10 (4.2%)	8 (3.4%)	51 (21.5%)

Note: To calculate the percentage given in parentheses, the number of "Types of Search Key Used after the First Not-found Response" was divided by the number of "Not-found Responses."

ious sources other than regular acquisitions channels, and many of these sources do not have LCCN information.

For the serials cataloging function (table 4), the title key was the first primary choice and retrieved the desired records 44 percent of the time. If this key failed to retrieve the desired records, the users entered as their secondary key a different title key 55 percent of the time and a name/title key 17 percent of the time. Approximately 23 percent of the time, users decided to discontinue the search. Although for serials cataloging the title key was used most frequently, its hit rate was less than 45 percent. On the other hand, the ISSN key was used very little, but its hit rate was as high as 80 percent. The use of the ISSN key is likely to increase in the future, however, because the United States Postal Service now requires the ISSN to be present on serials.[8] Therefore, the ISSN will be more readily available to the user.

Among the searches performed through the public-use terminal (table 5), the most frequently used primary search key was the name/title key, which resulted in a successful search about 29 percent of the time. When patrons encountered a "not found" response, they tried as their secondary choice a different name/title key 29 percent of the time, or a title key 29 percent of the time. They abandoned the search 38 percent of the time. As mentioned earlier, the public-use terminal can be used by anyone, including faculty and students. The hit rate for name/title

Table 4. Number of Primary and Secondary Choices of Search Keys for Serials Cataloging

| Type of Search Key Used First | Items Searched | Found Responses | % of Found Responses | Not-found Responses | Types of Search Key Used after the First Not-found Response | | | | | Search Discontinued after the First Not-found Response |
					Name/Title	Title	Personal Author	LCCN	ISBN	
Name/Title	15	3	20.0%	12	6 (50.0%)	4 (33.3%)	1 (8.3%)	0 (0.0%)	0 (0.0%)	1 (8.3%)
Title	72	32	44.4%	40	7 (17.5%)	22 (55.0%)	2 (5.0%)	0 (0.0%)	0 (0.0%)	9 (22.5%)
Personal Author	0	—	—	—	—	—	—	—	—	—
LCCN	1	0	0.0%	1	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
ISBN	1	0	0.0%	1	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
ISSN	5	4	80.0%	1	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
CODEN	0	—	—	—	—	—	—	—	—	—
Total	94	39	41.5%	55	14 (25.5%)	28 (50.9%)	3 (5.4%)	0 (0.0%)	0 (0.0%)	10 (18.2%)

Note: To calculate the percentage given in parentheses, the number of "Types of Search Key Used after the First Not-found Response" was divided by the number of "Not-found Responses."

Table 5. Number of Primary and Secondary Choices of Search Keys for Public Use

| Type of Search Key Used First | Items Searched | Found Responses | % of Found Responses | Not-found Responses | Types of Search Key Used after the First Not-found Response | | | | | Search Discontinued after the First Not-found Response |
					Name/Title	Title	Personal Author	LCCN	ISBN	
Name/Title	77	22	28.6%	55	16 (29.1%)	16 (29.1%)	0 (0.0%)	2 (3.6%)	0 (0.0%)	21 (38.2%)
Title	44	20	45.4%	24	11 (45.8%)	9 (37.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (16.7%)
Personal Author	16	5	31.3%	11	0 (0.0%)	0 (0.0%)	3 (27.3%)	0 (0.0%)	0 (0.0%)	8 (72.7%)
LCCN	13	5	38.5%	8	2 (25.0%)	2 (25.0%)	0 (0.0%)	1 (12.5%)	1 (12.5%)	2 (25.0%)
ISBN	3	2	66.7%	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)
ISSN	3	1	33.3%	2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
CODEN	2	0	0.0%	2	0 (0.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)
Total	158	55	34.8%	103	29 (28.2%)	38 (27.2%)	3 (2.9%)	3 (2.9%)	2 (1.9%)	38 (36.9%)

Note: To calculate the percentage given in parentheses, the number of "Types of Search Key Used after the First Not-found Response" was divided by the number of "Not-found Responses."

key at this terminal was rather low. From this study, it is not possible to say whether this was due to patrons' lack of knowledge in key construction or lack of sufficient information needed for the construction of the key.

SUMMARY AND CONCLUSIONS

Among various types of search keys available to the users, the name/title, LCCN, and title search keys were entered most frequently. The use of personal-author, ISBN, ISSN, and CODEN search keys was very limited for all library functions. Corporate-author search keys were not used at all.

For the acquisitions function, system users most frequently entered the LCCN key, followed by the name/title key. For monograph cataloging, the users entered the name/title key most frequently, followed by the LCCN key. For serials cataloging, the use of the title key was the most common. Persons using public-use terminals entered mostly name/title and title search keys.

For acquisitions and monograph cataloging functions, the LCCN key was most successful in retrieving the desired records. The next most successful key was the name/title key. For both of these functions, when the name/title key failed to retrieve the record, users next tried the title key most of the time.

For serials cataloging, the title key was used most frequently but was not very successful in retrieving serial records. On the other hand, the ISSN key was the most successful but it was used very little.

Individual identifiers such as LCCN, ISSN, ISBN, and CODEN are very efficient search keys because they retrieve, on the average, far fewer numbers of replies than other types of search keys. With the exception of LCCN, the individual indentifiers were used only to a small extent. From this study, it is not possible to answer questions such as: Why weren't individual identifiers' search keys not used more often? Did a searcher use a name/title key even when the LCCN was available? To answer such questions, data will have to be collected concerning what kind of information is available to the searcher when constructing the search keys.

REFERENCES

1. F. G. Kilgour, P. L. Long, and E. B. Leiderman, "Retrieval of Bibliographic Entries from a Name-Title Catalog by Use of Truncated Search Keys," *Proceedings of the American Society for Information Science* 7:79–82 (1970).
2. F. G. Kilgour and others, "Title-only Entries Retrieved by Truncated Search Keys,"

Journal of Library Automation 4:207–10 (Dec. 1971).
3. P. L. Long and F. G. Kilgour, "A Truncated Search Key Title Index," *Journal of Library Automation* 5:17–20 (March 1972).
4. A. L. Landgraf and F. G. Kilgour, "Catalog Records Retrieved by Personal Author Using Derived Search Keys," *Journal of Library Automation* 6:103–8 (June 1973).
5. A. L. Landgraf, K. B. Rastogi, and P. L. Long, "Corporate Author Entry Records Retrieved by Use of Derived Truncated Search Keys," *Journal of Library Automation* 6:151–61 (Sept. 1973).
6. J. D. Smith and J. E. Rush, "The Relationship between Author Names and Author Entries in a Large On-Line Union Catalog as Retrieved Using Truncated Keys," *Journal of the American Society for Information Science* 28, no.2:115–20 (March 1977).
7. OCLC, Inc., *Searching the On-Line Union Catalog* (Columbus, Ohio: OCLC, Inc., 1979).
8. *Library of Congress Information Bulletin* 37:35 (1 Sept. 1978).

Kunj B. Rastogi is a research scientist at OCLC. **Ichiko Morita** is assistant professor at the Ohio State University Libraries.