Subject searching in the OPAC of a special library: problems and issues

M.S. Sridhar

The author

M.S. Sridhar is Head, Library and Documentation Division, ISRO Satellite Centre, Bangalore, India.

Keyword

Online catalogues, Information retrieval

Abstract

This paper draws on data from a comparative study of use of the online public access catalogue (OPAC) and the card catalogue of the ISRO Satellite Centre (ISAC) library, and examines the steady decline in the use of subject searching by end-users and the associated problems and issues. It presents data to highlight the negligible use of Boolean operators and combination searches, variations in descriptors assigned to books of the same class numbers, and too many records tagged to very broad descriptors. The article concludes that moving from a traditional card catalogue to a modern OPAC has not made subject searching more attractive or effective.

Electronic access

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1065-075X.htm

Introduction

A "subject approach" to knowledge has long been a concern of librarians, and is considered by them to be a major approach (access method) of users. In the print card catalogue days, both the classified part of the catalogue and the subject catalogue (based on assigned standard descriptors) were suppose to help users have subject access to the resources of a library. Access by classification number was believed to be more common in Europe and India than elsewhere. Unfortunately, very few card catalogue studies have been conducted by libraries in the past to judge the relative importance of the subject approach, success or failure of searches, and user behaviour towards subject access.

Print card catalogues had plenty of cross references to help users, even if they were not aware of standard descriptors. Today the same are lacking in online public access catalogues (OPACs) unless online thesaurus help or subject authority control tools are made available. Print card catalogues were primarily meant for pre-coordinated searching, whereas OPACs enable post-coordinated searches using Boolean operators and other combination searches. Further, OPACs can also execute vague and free text queries wherever keyword-in-context (KWIC) indexes are provided, which is a great boon to users who are normally unaware of descriptors selected from a thesaurus or subject heading list. Despite limitations, users do prefer free-text searching. There are efforts underway as well to create intelligent natural-language front ends which use subject headings and thesauri for searching the OPAC.

Some past studies

Although studies on use of card catalogues are rare, studies on the use of the OPAC are abundant, and there exist many reviews of OPAC studies (Larson, 1991; Hildreth, 1985; (O'Brien, 1994). In a 30-month transaction analysis of patron searches in the OPAC of Ohio State University, Norden and Lawrence (1981) found that the use of subject search commands increased rapidly (quoted from Hildreth, 1985, pp. 262-3). Unfortunately, OPACs are generally criticised as being more difficult to use and less serviceable

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · pp. 183-191
 · ISSN 1065-075X
DOI 10.1108/10650750410564691

The author wishes to express his gratitude to Dr P. S. Goel, Director, ISRO Satellite Centre, for his kind encouragement and permission to publish the paper.

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

than card catalogues, and are used more often to find known items rather than for seeking information or solving information-based problems (Borgman, 1996, p. 494). There has been, however, a significant and consistent decline in controlled vocabulary-based subject searching over the years in favour of title keyword searching (Larson, 1991). The relative use of subject indexes in OPACs has varied widely, from 10 per cent to 62 per cent in different studies. In the famous Council of Library Resources (CLR) project (jointly sponsored by OCLC and RLG), subject searching was found to constitute the majority (59 per cent)(Matthews *et al.*, 1983). Larson (1991), however, found a gradual decline in the use of the subject index as the size of the database increased. The frequency of title keyword searching exceeded that of subject searching over a similar time period. As mentioned before, however, many title keyword searches are for known items.

O'Brien (1994) says that "subject access is the most problematic area of online catalogues." It often leads either to failure or the retrieval of too many references. Most OPAC studies have identified the need to tackle related issues like free text search, field-directed searching, training users, adjunct thesaurus help, limiting devices with a "filtering" effect, relevance feedback, and ranking of retrieved references to reduce search failures, as most users cannot be expected to put in extra effort into subject searching. It was felt that even computer literacy does not ensure that users will do better subject searches.

The OPAC is a "black box" to users, and they know very little about what happens inside the system. For example, a tradeoff between precision and recall is rarely known to end-users. This may also be true of many library professionals who deal with the OPAC on a routine basis. OPACs are diverse in features as well as size. The functional layers mediating access to OPAC are user interface, DBMS interface, DBMS, and database (with indexes). Conceptually they can be grouped into those that deal with users (user interface) and those that deal with the storage and retrieval of bibliographic data. Most enhancements to subject searching are through enhancements to the database and DBMS layers. The database layer can be seen as some portion of the contents of the card catalogue in machine-readable form. The other layers provide procedures to facilitate the process of delivering information to users (Larson, 1991, pp. 190-91).

Boolean logic appears to be one of the most difficult aspects of information retrieval and is not "common sense" for most people (Borgman, 1986, p. 390). Users tend to perform simple searches using only the basic features. Even scientists and engineers who have expertise in logic for other applications, often use "AND" and "OR" in their linguistic sense (Borgman, 1986, 1996). Combination search or use of Boolean operators can greatly help users to reduce recall and increase precision so as to obtain a browseable number of hits. Users' information needs range from highly specific to very general, but they make only simple searches. Most users search with a single term that defeats the purpose of combination searches and Boolean logic:

> [. . .] Users rarely ventured beyond a minimal set of system features. The majority of searches were simple, specifying only one field or data type to be searched . . . [and] advanced search features were rarely used . . . (Borgman, 1986, pp. 389-90).

Most studies monitoring transactions have reported significant frequencies of input unidentifiable by the system, aborted sessions, and searches with no matches. Not only "usage rates are low enough that many online catalog users probably remain permanent novices" (Borgman, 1986, p. 390), but also users tended to perform simple searches using only the basic features and did not utilise index terms or headings unless "forced" to do so (Borgman, 1986, pp. 389-90). Like the use of other services of the library and interactions within the library, the use of the OPAC is also skewed with a few people using it frequently and most using very little. Further, most end-users search the OPAC only occasionally and do not access the system on a regular basis, tending to learn only enough to do simple searches quickly and to regard further instruction as unnecessary and more extensive expertise as a burden (Yuan, 1997, p. 218). In OPAC use studies, no-match subject searches ranged from 35 per cent to 57 per cent (Borgman, 1986, pp. 389-90). Dickson (1984) found that 37 per cent of all title searches and 23 per cent of all author searches resulted in no matches. She determined that 39.5 per cent of the no-match title searches and 51.3 per cent of the no-match author searches were for records that existed in the database and not due to user error in searching (Borgman, 1986).

Hirst (1998), through a questionnaire survey of users with different levels of information technology (IT) expertise about use of hypertext interfaces to library and information science (LIS), found that OPAC searches were mainly conducted for specific items and that most were successful. It may be noted that most specific item searches are not subject searches. Interestingly, novice users tend to achieve higher success rates than expert users:

> Identifying search terms for the subject catalog is hardest of all, since people often do not recognize that the subject entries are drawn from a controlled

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

> list or thesaurus that is separately searchable itself. Instead, they enter the catalog using the free-text keywords they know best, often on a trial and error basis (Borgman, 1996, p. 497).

One way to guide end-users is to enable them to choose a subject term from the classification number of the book. There is a great need to link subject headings to classification numbers.

Drabenstott and Weller (1994) summarised the problems of subject access based on the findings of the past studies:

- one-third of subject queries fail to produce results;
- large retrievals (high recall) discourage users from scanning the results;
- few instances of successful matching of query with controlled vocabulary are one-word queries; and
- users are discouraged by subject access and look for alternative approaches.

They feel that system designs enhancing subject headings, developing menu-based interfaces, and extending online catalogue functionality to other databases met the demands of library staff and not necessarily that of users for search functionality. Hence the problems found in the earliest OPACs still continue to exist as far as users are concerned (see Appendix 1). Drabenstott and Weller (1994) have proposed "search trees" for subject searching after running an experimental online catalogue called ASTUTE.

Search failures are usually due to misspelling, lack of knowledge of thesauri, "false drops," lack of user understanding of Boolean operators, lack of cross references, lack of online thesauri, and lack of training. There is a significant positive correlation between the failure rate and the percentage of subject searching (Larson, 1991), and a negative correlation between failure rate and time. The longer the processing time/rate, the greater the chances of a user abandoning the search. Subject searches often lead to unusually high recall and create the problem of information overload. The average number of records retrieved are very high (about 77) and users look at only a few (about nine) (Larson, 1991). Often, users prefer browsing the shelf rather than browsing through subject access. Larson (1991, p. 20) concludes that "the desire to do topical searching has not diminished, but that the penalties incurred by the user in the process of using the subject index have led to the decline in use". Even the CLR survey found that topical searching was more prevalent among those who are less experienced with the library and its catalogues (Matthews *et al.*, 1983).

O'Brien (1994, p. 223) says that the "... subject indexing of monographs is both superficial and inadequate" as they lack in-depth treatment particularly for composite books and conference volumes, as well as tables of contents (TOC), blurbs, etc. Larson (1991) suggests remedies to the problems of subject searching which are grouped under the database, search processing and retrieval algorithms, and the user interface. Database-related remedies included expanding records by adding words from TOC, indexes and blurbs of books, enhancing records with terms from classification schemes, and increasing the number of descriptors per book (see Appendix 2). The second group of remedies included limiting search results by Boolean intersections using additional terms or dates, partial matching and stemming of keywords, relevance ranking of outputs, automatic mapping from input search terms to controlled vocabulary terms through thesaurus look-up, automatic spelling correction or phonetic matching of terms, and relevance feed back. The last group of remedies related to user interface included "browsability" of existing subject headings using online thesaurus and classification assignments. In the context of digital libraries, however, Lesk (2003) questions the very need for traditional subject classification and indexing (which are usually meant for a possible future query) when the actual query itself can be searched on demand in seconds. Multistage searching and display, saving searches, set buildings, etc., are considered as not required any more in Web and future digital libraries. The new bibliographic model, The Functional Requirements for Bibliographic Records (FRBR), hopes for subject access by way of suggesting retrieval of groups of related documents. This program is supposed to generate sets of records that can be grouped for display as single works irrespective of their manifestations.

In a transaction log analysis of OCLC online catalogs, Tolle (1983) worked out correlation coefficients for transaction search patterns and found that the probability of going from "begin" state to "browse" is the highest (0.643) and going from one "error" to another "error" state is the next highest (0.598). In other words, once an error was made, the next transaction/command was an error in 59.8 per cent of times. Hardly less than 9 per cent moved from an error state to ending the session. This speaks of user frustration and waste of effort.

The degree of variability of subject searching in an OPAC at a university library revealed that subject searching varied from 22 per cent to 74 per cent over the hours of the day, from 17 per cent to 64 per cent over the days of the week, and from 12 per cent to 70 per cent over weeks of the semester (Kaske, 1988a, b). A more recent questionnaire-based survey (Oduwole *et al.*, 2002) of the use of the OPAC by 286 users at a Nigerian University

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

found that the OPAC was used mostly for self search rather than delegated search with "author" as major (59 per cent) access point, followed by "subject" (30.8 per cent), and the large majority were very satisfied (75 per cent) with the OPAC. A more extensive review of past OPAC use studies is neither feasible nor desirable. Some of the important findings and suggestions, however, are summarised in the adjacent boxes.

Background of the study

The Indian Space Research Organisation (ISRO) Satellite Centre (ISAC) is one of the major research and development centres of the ISRO. It is the lead centre for satellite technology. The primary objective of ISAC is to develop and operationalise indigenous satellites, and use space technology for socio-economic development of the country. It was established in 1972 at Bangalore as the Indian Scientific Satellite Project (ISSP) to build the nation's first satellite "Aryabhata". Today, ISAC has grown into a premier centre for research and development in satellite technology. Over the years, ISAC has planned and executed several satellite missions of ISRO. The missions represent a broad spectrum of satellite technology. Beginning with scientific and application experiments, these have culminated in operational space systems. Having successfully developed and deployed state-of-the-art operational satellites for communication, meteorology and remote sensing, the centre is poised to produce advanced versions of satellites like microwave remote-sensing satellites and direct broadcasting satellites in the near future. The ISAC library is the main source of information support to all projects, missions and activities of ISAC. It provides information and allied support services to over 2,500 users consisting of managers, technocrats, engineers, scientists and other supporting technical and non technical staff in the centre.

An observation-based study of use of the card catalogue of the ISAC library carried out during 1985 (Sridhar, 1986) revealed that the classified catalogue was not used, the report number catalogue (which is like the classified catalogue except for reports) was negligibly used, the author and title catalogues were moderately used, and the subject catalogue was heavily used. The library was automated during the early 1980s and the OPAC was made available on local area network (LAN) from 1991. A study (Sridhar, 2004) of the use of the OPAC based on observation, interaction with users, and recording by professional staff at the site was done for over 80 hours (the equivalent of ten working days or two working weeks) with due representation for all times of the day and all days of the week during July-August 2002, with the speculation that the use of the OPAC must increase substantially compared to the use of the card catalogue. Unfortunately, the new software did not provide for collecting transaction log data of the OPAC. At the time of study, the OPAC had over 200,000 records. Surprisingly, when the results of this study was compared with that of the earlier (Sridhar, 1986) card catalogue study, it was found that the volume of usage of the OPAC was not even as much as what it was in case of the card catalogue. It was concluded from this study that, despite the fact that the card catalogue was no match to the searchability of the OPAC, the quantitative comparison of the usage of the OPAC (without making it available as a web OPAC on any network) with that of the card catalogue did not favour the OPAC. Yet adequate weighting for qualitative factors like user convenience, time saved, and the effect of other variables like availability of Internet resources needed to be kept in mind. Further, the practice, attitude and behaviour of users were equally important when we examined tools of new technology. The data from this study were extracted to investigate further issues and problems relating to the decline in subject searching.

Data and discussion

The kinds of access that users prefer while searching the OPAC as well as the card catalogue are very interesting, and are very revealing for the development of search tools and techniques. The OPAC of the ISAC library was developed as a result of local cataloguing, without outsourcing or importing data. The OPAC does not include URLs, but has many e-resources like TOCs, blurbs, etc., tagged on to records. The search features of the OPAC are not comparable with that of internet search engines. The study (Sridhar, 2004), based on critical incident observation, was primarily aimed at examining the approaches of users of the ISAC library while searching the OPAC. Table I and Figure 1 present the statistics related to different approaches adopted by users for searching and querying the system. Also juxtaposed in the table and the corresponding figure are the data from the previous study of the use of the card catalogue in 1985 (Sridhar, 1986). The data reveal that the title approach was adopted by a maximum of 38.3 per cent while using the OPAC, as against a maximum of 54.2 per cent adopting the subject (descriptors) approach in the card catalogue case. It may be noted that additional search features/approaches like KWIC

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

Table I Access point/approach

Type of search/access	OPAC (2002) No.	OPAC (2002) %	Card catalogue (1985) No.	Card catalogue (1985) %
Title	150	38.3	8	8.3
Author	105	26.8	34	35.4
Subject (keyword)	90	23	52	54.2
Class no./report no.	5	1.3	2	2.1
Place of publication	1	0.2	NA	NA
Publisher	1	0.2	NA	NA
Word in title (KWIC)	30	7.7	NA	NA
Combination search	10	2.5	NA	NA
Total	392	100	96	100

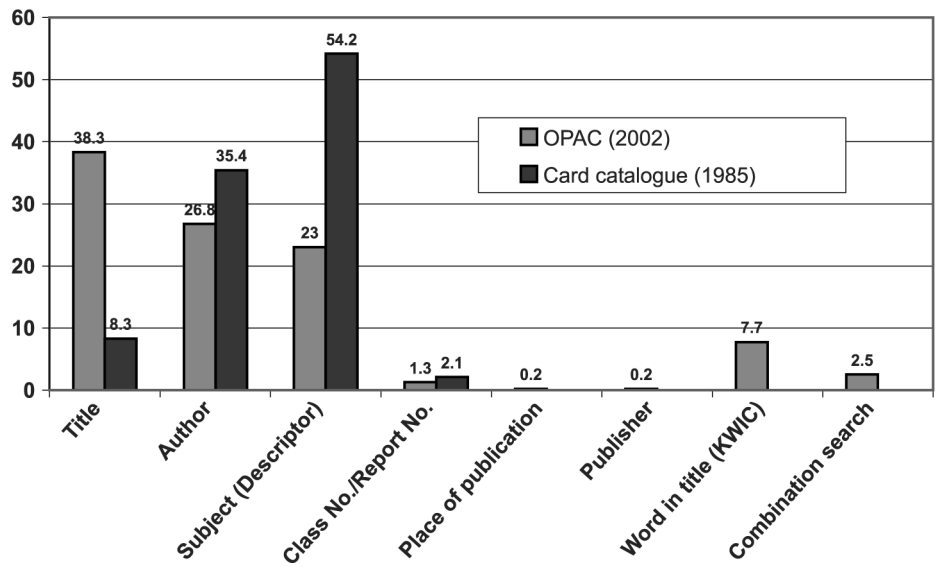
Note: NA = not applicable

and combination searches were obviously not found in the card catalogue, and hence the magnitude of subject searching on the OPAC can be assumed to be 33 per cent after adding the KWIC and combination searches. Even then the subject searches have substantially dropped from the time of card catalogue to the OPAC. This is in conformity with the findings of the past studies. As noted earlier, there is generally a small but significant decline in controlled subject searching in favour of keyword (free text) searching (Larson, 1991). Even the percentage of searches on the author catalogue, however, has dropped from 35.4 per cent in the card catalogue to 26.8 per cent in the OPAC. This also conforms to the findings of Norden and Lawrence (1981) that title searches were the most frequent, and the ratio of title to author searches was higher than that in the card catalogue. Assuming that specific item searches were mostly non-subject searches, Hirst (1998) through a questionnaire survey of users with different levels of IT expertise, found that OPAC searches were mainly conducted for specific items and that most were successful. In the present study, hardly 2.5 per cent of searches were "combination" searches. These kinds of advanced searches are done by very few end-users, and this trend also conforms to the findings of most of earlier studies of OPAC.

Concentrating only on major approaches, namely title, author and subject and merging data relating to KWIC searches into that of subject searches, the same data extracted from Table I are presented in Table II. This table, together with Figure 1 relating to title, author and subject searches, presents an interesting comparison of findings of the use of the OPAC with that of card catalogue. First, searches by title have substantially increased, from a mere 8.3 per cent in the card catalogue to 38.3 per cent in the OPAC. Second, subject searches have dropped substantially, from 54.2 per cent in the case of the card catalogue to a mere 30.7 per cent (including KWIC searches) in the OPAC. Author approach searching, however, has marginally dropped (from 35.4 per cent to 26.8 per cent) from the card catalogue to the OPAC. On the contrary, subject searching constituted the majority (59 per cent) in the CLR project (Matthews *et al.*, 1983).

A follow-up observation of 51 subject searches made by end-users is depicted in Table III. The data revealed that nearly half of them met with failure. The remainder of the searches (52.9 per cent) were considered reasonably successful searches in obtaining the desired

Figure 1 Access point/approach OPAC and card catalogue



OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

Table II Percentage of searches by author, title and subject

Access point/ type of catalogue	OPAC (2002)	Card catalogue (1985)
Title	38.3	8.3
Author	26.8	35.4
Subject (+ KWIC)	30.7	54.2

results. Out of those who met with failure, nearly one-quarter (23.5 per cent) abandoned the search. Another one-quarter (23.5 per cent) changed the search strategy. Out of those who changed the search strategy, 13.7 per cent changed keywords.

It is very clear from the results of the earlier studies that the majority of end-users face problems with subject searching, and one of the major issues is the concern with the selection of appropriate standardised descriptors by end-users. It is also clear that users do not prefer searching by classification numbers, and there is no online help regarding thesauri and/or classification scheme. There is a need to create links between the classification numbers and the corresponding descriptors. Further, too many or too few bibliographic records are linked to a subject heading, and also too few subject headings are assigned per record (an average of two per record). Hence there is a need to review the frequency of assigned subject headings. Table IV depicts an analysis of descriptors in 920 books from 20 sample class numbers at the ISAC library. Surprisingly, the average number of descriptors assigned per book varied from as low as 1.39 in the case of safety-related books, to a moderate 2.75 in the case of solar physics, and system-related books with an overall average a meager 1.89. Further, the standard deviation (SD) was also very low. In other words, contrary to the recommendation that more descriptors should to be assigned to the books, the collection continues to be indexed with fewer keywords than desired as decided by the indexers. The sample books had 599 unique keywords assigned 1,737 times. The last column of Table IV depicts the ratio of the number of unique descriptors to the number of books in each subject (class number). As this ratio increased, the precision of single descriptor search increased. Unfortunately, except in a few cases, this ratio was

Table III Success or failure of subject searches

	No.	%
Successes	27	52.9
Failures		
Abandoned	12	23.5
Changed keywords	7	13.7
Changed search	5	9.8
Sub total	24	47.1
Total	51	

Table IV Sample class numbers checked for descriptors

Sl. no.	Class. no.	Subject	No. of books	No. of descriptors	Average	SD	No. of unique descriptors	Ratio of unique descriptors to books
1	523.9	The sun, solar physics	24	66	2.75	1.27	39	1.63
2	528.8	Remote sensing	79	156	1.97	1.21	50	0.63
3	537.8	Electromagnetism, electromagnetic field, electrodynamics, Maxwell theory	52	86	1.65	1.26	28	0.54
4	539.216.2	Films, thin films	24	35	1.46	1.00	14	0.58
5	614.8	Accidents prevention, protection safety	28	39	1.39	0.88	21	0.75
6	621.3.049.7	Printed circuits and the like	10	21	2.10	0.89	11	1.10
7	621.314.5	Conversion of AC to DC and vice versa, convertors, invertors	17	41	2.41	1.18	20	1.18
8	621.38.049.771.14	Microprocessors	243	409	1.68	1.10	68	0.28
9	621.37.04	Microelectronics	32	71	2.22	1.11	28	0.88
10	621.381.542	Image analysis	30	71	2.37	1.57	38	1.27
11	621.383.5	Barrier layer photocells, photovoltaic cells, photodiodes, phototransistors	15	23	1.53	0.65	6	0.40
12	621.396.67	Aerial systems	79	150	1.90	1.86	55	0.70
13	621.390.96	Radar	58	107	1.84	1.92	55	0.95
14	629.7.036.5	Rocket propulsion	8	18	2.25	1.44	13	1.63
15	629.73	Aeronautical engineering	25	48	1.92	1.27	28	1.12
16	629.785	Space probes	21	52	2.48	1.01	21	1.00
17	658.562	Quality control	65	110	1.69	0.58	22	0.34
18	681.3.02	Design, construction layout of DP systems	77	177	2.30	1.01	62	0.81
19	681.3.06vhd	VHDL (computers)	15	25	1.67	0.36	9	0.60
20	681.351	Computer networks	18	32	1.78	0.40	11	0.61
Total			920	1,737	1.89		599	0.65

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

Table V Ranked list of descriptors in books

Rank	Descriptor	No. of books
1	Computer programming	1,127
2	Computer networks	748
3	Signal processing	745
4	Software engineering	610
5	Integrated circuits	577
6	Programming languages	534
7	Control theory	520
8	Neural nets	515
9	Artificial intelligence	510
10	Communication networks	470
11	Aerospace engineering	468
12	Computers	457
13	Micro processors	433
14	Remote sensing	426
15	Data processing	418
15	Image processing	418
17	Robotics	404
17	Telecommunication	404
19	Data base management systems	368
19	Dictionaries	368
20	Indexes (documentation)	349
21	Computer aided design	343
22	Astronomy	341
23	India	340
24	Circuits	329
24	Electrical engineering	329
25	Very large-scale integration	328
26	Communication	324
26	Micro computer	309
27	Computer programs	303
28	Multimedia	299
29	Management	285
30	Antennas	281
31	Libraries	279
31	Wireless communication	279
33	Computer graphics	271
34	Industries	266
35	Quality control	264
36	Expert systems	263
36	Operating systems (computers)	260
37	Architecture (computers)	259
38	Internets	257
39	Heat transfer	248
40	Bibliographies	247
41	Manufacturing	246
42	C (programming language)	245
43	Robots	238
44	Medical science	234
45	Algorithms	220
45	Electronic equipment	220
47	Lasers	217
47	Reliability	217
48	C + + (programming language)	213
49	Databases	207
49	Micro waves	211
50	Propulsion	206
51	Object-oriented programming	205
Total (57 descriptors)		20,452

Table VI Descriptors with least number of records/ books (frequency distribution of least assigned descriptors)

	No. of descriptors	No. of books (records)
	1,539	1
	658	2
	367	3
	281	4
	196	5
	159	6
	122	7
	93	8
	70	9
	75	10
Sub-total	3,560	
	936	>10
	13,004	(unused)0
Total	18,100	

not even 1.0 for most subjects, and the highest was just 1.63.

An analysis of the frequency of occurrence of descriptors in the entire OPAC revealed further surprises. A total of 57 of the most frequently-occurring descriptors appearing 20,452 times in the OPAC are listed in Table V. Most of the descriptors in Table V are very broad topics and, unless they are used in post-coordinated or Boolean searches or in combination with other fields, they will result in too many records and become unmanageable for end-users.

A frequency table of descriptors used for indexing books is depicted in Table VI. The ISAC Library uses the NASA Thesaurus for indexing books by way of assigning up to six postable terms from out of 18,100 postable terms listed in the NASA Thesaurus (which has a total of 41,300 entries). It is clear from the table that only 936 descriptors (postable terms) are assigned to more than ten books in a collection of about 39,000 books. In other words, as many as 13,600 descriptors out of 18,100 possible terms (i.e. 75.2 per cent) in the NASA Thesaurus remain unused. It is shocking that professional indexers have made minimal use of a specialised thesaurus meant for space science and technology while indexing books in a space science and technology library.

While most of the frequently-used terms (as shown in Table V) cause problems of precision, the unused and least-used descriptors (as shown in Table VI) cause problems of recall. As descriptors are drawn from a specialised thesaurus meant for space science and technology, such infrequent (including non-use) as well as the most frequent use of a small set of descriptors appears to be the result of variation in indexing level over time and also from indexer to indexer. The problem may not be uncommon in libraries, but the quality of indexing is a serious concern, if subject searching is

OCLC Systems & Services: International Digital Library Perspectives
Volume 20 · Number 4 · 2004 · 183-191

to be promoted among users. Past research has repeatedly found that failure to match the system's subject vocabulary, recall of subject browsing displays with thesaurus lookup facility, and lack of or inadequate cross-references are serious drawbacks for subject searching. Obviously, many have suggested allowing the use of natural language, providing additional search aids like search trees, pointing users from non-standard keywords to descriptors, and development of ontological and other automatic categorisation techniques.

Conclusion

The search process in online catalogues has more or less remained the same as in print card catalogues, with increased access points and varieties of search features but with increased complexity added to the process. End-users are not only expected to have technical searching skills, but also conceptual and semantic knowledge related to the query in the case of subject searching in order to articulate the query. Both indexing quality and under-usage of subject access in the OPAC cause serious concern about the so-called "subject approach" to knowledge. Inconsistent indexing quality has made subject searching more difficult and ineffective. In addition to end-user training and other suggested remedies, a totally new look into interactive searching like drag and drop text from hits, "more like this" features, online thesauri lookup with classification links, partial matching, and automatic AND searching are required. Among many suggestions made to software vendors should be the provision for indicating the intended use/application, or level of book such as beginner, general, specialised, and advanced, which often helps to improve the precision of the search and utility of hits in the OPAC. Lastly, adding intelligent components like quality additions based on community rating/use statistics and automatic updating are necessary.

References

Borgman, C.L. (1986), "Why are online catalogs hard to use? Lessons learned from information retrieval studies", *Journal of the American Society for Information Science*, Vol. 37 No. 6, pp. 387-400.

Borgman, C.L. (1996), "Why are online catalogs still hard to use?", *Journal of the American Society for Information Science*, Vol. 47 No. 7, pp. 493-503.

Dickson, J. (1984), "An analysis of user errors in searching on online catalog", *Cataloging and Classification Quarterly*, Vol. 4 No. 3, pp. 19-38.

Drabenstott, H.K.M. and Weller, M. (1994), "Testing a new design for subject searching in online catalogues", *Library Hi Tech*, Vol. 12 No. 1, pp. 67-76.

Hildreth, C.R. (1985), "Online public access catalogs", in Williams, M.E. (Ed.), *Annual Review of Information Science and Technology (ARIST)*, Vol. 20, Knowledge Industrial, New York, NY, pp. 232-85.

Hirst, S.J. (1998), *Hyperlib Deliverable 1.2: In-depth Survey of OPAC Usage. Part of Hyperlib Electronic Document Store*, University of Antwerp, Antwerp and University of Loughborough, Loughborough, available at: www.lia.ua.ac.be/MAN/P12/root.html

Kaske, N.K. (1988a), "The variability and intensity over time of subject searching in an online public access catalog", *Information Technology & Libraries*, Vol. 7 No. 3, pp. 273-87.

Kaske, N.K. (1988b), "The comparative study of subject searching in an OPAC among branch libraries of a university library system", *Information Technology & Libraries*, Vol. 7 No. 4, pp. 359-72.

Larson, R.R. (1991), "Between Scylla and Charybdis: subject searching in the online catalog", in Godden, I.P. (Ed.), *Advances in Librarianship*, Vol. 15, San Diego Academic Press, San Diego, CA, pp. 175-236.

Lesk, M. (2003), "Collecting for a digital library: size does matter", *Information Management and Technology*, Vol. 36 4, October-December, pp. 184-7.

Matthews, J.R., Lawrence, G.S. and Ferguson, D.K. (1983), "Using online catalogs: a nationwide survey. A report of a study sponsored by the Council on Library Resources", in Matthews, J.R. (Ed.), *The Impact of Online Catalogs*, Neal-Schuman, New York, NY.

Norden, D.J. and Lawrence, G.H. (1981), "Public terminal use in an online catalog: some preliminary results", *College & Research Libraries*, Vol. 42 No. 2, July, pp. 308-16.

O'Brien, A. (1994), "Online catalogs: enhancements and developments", in Williams, M.E. (Ed.), *Anuual Review of Information Science and Technology*, Vol. 29, Learned Information, Medford, NJ, pp. 219-42.

Oduwole, A.A. *et al*.et al., (2002), "On-line public access catalogue (OPAC) use in Nigerian academic libraries: a case study from the university of agriculture, Abeokuta", *Library Herald*, Vol. 40 No. 1, March, pp. 20-7.

Sridhar, M.S. (1986), "Pattern of card catalogue consultation in a special library", *IASLIC Bulletin*, Vol. 31 No. 1, March, pp. 9-16.

Sridhar, M.S. (2004), "OPAC vs card catalogue: a comparative study of user behaviour", *The Electronic Library*, Vol. 24 No. 2, pp. 175-83.

Tolle, J.E. (1983), *Current Utilization of Online Catalogs: Transaction Log Analysis. Final Report to the Council of Library Resources*, OCLC research report number OCLC/0PR/RR-83/2, Vol. I, OCLC, Dublin, OH.

Yuan, W. (1997), "End-user searching behavior in information retrieval: a longitudinal study", *Journal of the American Society for Information Science*, Vol. 48 No. 3, pp. 218-34.

Further reading

Alzofon, S.R. and Van Pulis, N. (1984), "Patterns of searching and success rates in an online public access catalog", *College & Research Libraries. March*, Vol. 45 No. 2, March, pp. 110-5.

Banks, J. (2000), "Are transactions log useful? A ten-year study", *Journal of Southern Academic & Special Librarianship*, Vol. 1 No. 3, February.

Drone, J.M. (1984), "A use study of the card catalogue in the University of Illinois Music Library", *Library Resources and Technical Services*, Vol. 28 No. 3, July/September, p. 253.

Ferguson, D. *et al.* (1982), "The CLR public online catalog study: an overview", *Information Technology & Libraries*, Vol. 1 No. 2, June, pp. 84-97.

Gorman, M. and Hotsinpiller, J. (1979), "ISBD: Aid or barrier to understanding?", *College and Research Libraries*, Vol. 40 No. 6, November, p. 251.

Gouke, M.N. and Pease, S. (1982), "Title searches in an online catalog and a card catalog: a comparative study of patron success in two libraries", *Journal of Academic Librarianship*, Vol. 8 No. 3, July, pp. 137-43.

McCarthy, C. (1975), "Colonial cataloguing", *New Library World*, Vol. 76 No. 897, March, pp. 55-6.

Matthews, J. (n.d.), "Value of library services: a presentation to the Joint Chapter Professional Development Day", available at: ww.sla.org/chapter/cpnw/postings/0401Mathews.ppt

Williams, M.E. (Ed.) (1994), "Online catalogs: enhancements and developments", *Annual Review of Information Science and Technology*, Vol. 29, Learned Information, Medford, NJ, pp. 219-42.

Sridhar, M.S. (1985), "A case study of lent-out use of a special library", *Library Science with a Slant to Documentation*, Vol. 22 No. 1, March, pp. 19-34.

Sridhar, M.S. (2002), *Library Use and User Research: With Twenty Case Studies*, Concept Publishing, New Delhi.

Wilson, P. (1983), "The catalog as access mechanism: background and concepts", *Library Resources and Technical Services*, Vol. 27 No. 1, January/March, pp. 6-16.

Appendix 1. Important findings

- Too many failed searches or no records retrieved.
- Retrieve unmanageably large number of records.
- Navigational frustrations.
- Failure to match the system's subject vocabulary.
- Recall of subject browsing displays.
- Users' search requests/ terms are too broad or too narrow.
- Inadequate/lack of cross references.
- No Boolean queries formulated.
- Lack of user perseverance.
- Users make a variety of errors when entering a search request.
- Too few subject headings per bibliographic record (average bibliographic record contains less than two subject headings).
- Too many or too few bibliographic records linked to a subject heading.

Appendix 2. Suggestions for improvements

- Records enhanced with table of contents (TOC), publisher summary (blurbs), index terms, chapter summary, MARC records etc.
- Allow use of natural language search.
- Provide additional search aids/ assistance tools like search trees.
- Use information visualisation software and user interface.
- Provide front-end database.
- Provide access to classification information.
- Point the user from keywords to subject headings.
- Clean up bibliographic records.
- Utilize authority control.
- Review frequency of assigned subject headings.