

INFORMATION-SEEKING BEHAVIOR OF CATALOG USERS

R. TAGLIACCOZZO and M. KOCHEN

Mental Health Research Institute, The University of Michigan, Ann Arbor, Michigan

Summary—This study is based on the data from a survey of catalog use at three university libraries and one public library. Both “known-item” searches and “subject” searches were analyzed. The characteristics of the user population were examined and methodological problems of the survey were discussed. A relation was found between the academic rank of the catalog users and type of search that they carry out. Some of the factors influencing the success or failure of the search were analyzed, and the meaning of “success” for the two types of search was discussed. The study investigated trends in search strategies as well as degree of perseverance of catalog searchers. Implications for the design of modern information retrieval systems were pointed out.

IT SEEMS likely that future generations of scientists, scholars, technologists, and students will make use of partially automated information and education systems. When this happens, new skills, attitudes and habits will emerge in response to the change in technology. Though many would agree that the manipulative acts required to use automated information retrieval systems may be different from those required to use traditional information systems, almost everybody seems to accept the assumption that the same basic cognitive processes are involved in both cases. Is this assumption justified? Do we know so well, and can we describe so precisely, the information-seeking behavior of people, that we can assert that this behavior will remain basically the same, even when the information systems have changed? Or, if we do admit that there will be a change in behavior, can we predict in which directions it will occur? These general questions are part of the context of the present study.

The future users of automated information retrieval systems will be people who, for at least a part of their lives, have had recourse to libraries whenever the need to gain information not available in their immediate surroundings arose. (We will not consider here the hypothetical case of individuals born into a culture in which libraries, as we know them, no longer exist.) Any innovative information technology [1] will have to deal with the information-seeking behavior of these individuals. Studies of the *actual* use of existing libraries or other forms of traditional information systems, while helping us to improve the information systems of the present, will lay the foundations for the automated systems of the future.

The study that we are reporting concerns that aspect of information-seeking behavior which has to do with searching through a directory. It arose out of our concern with principles of directory design and was part of a project for learning to design a “growing encyclopedia system” [2]. Instead of creating an experimental directory, we chose, as a source of data, the library catalog—a ready-made, universally used directory. Other forms of directory, such as a subject index, a telephone book, a guide, an encyclopedia, could as well have served the purpose. Three of the libraries where our study was carried out are part of the University of Michigan Library System. The fourth library is the Ann Arbor

Public Library. The University of Michigan Library System ranks with the libraries of Harvard, Yale, and the University of Illinois as one of the largest academic libraries in the United States. With its 23 divisional libraries and a collection of four million volumes, it serves a student body of 36,000. The Ann Arbor Public Library is a medium-size library: it has about 155,000 volumes and serves a population of 100,000. As of June 30, 1969 the holdings of the three university libraries that we surveyed (General Library, Undergraduate Library, and Medical Library) were respectively 1,684,605 vols., 155,986 vols., and 175,339 vols.

The study dealt with various aspects of library catalog searching and was designed to reveal some of the characteristics of human information-seeking behavior in a "natural" environment. We were mainly interested in the college and research library; the public library was added for comparison purposes. In this paper we report on the general results of the study and discuss some of the trends observed. Details of our research can be found in Refs. 3 and 4. For recent studies by other researchers on library catalog use, see Refs. 5 through 11.

THE COLLECTION OF DATA

Data on the behavior of catalog users were collected by a method which combined interviewing and observation. We followed the users to the catalog, asked them questions before and during their searches, and watched them while they were searching the catalog. This method seemed preferable to other systems such as written questionnaires, simple interviewing, and diary keeping [10, 12], because it elicits "on the spot" information, it does not rely on the user's recollection and interpretation of his behavior, and, by separating the role of the observer (the interviewer) from that of the actor (the user), it attains a higher degree of accuracy.

Undoubtedly the observer has some influence on the behavior of the catalog user [13]. Although our interviewers were instructed to avoid pressuring anyone into participating in the survey, and to make themselves as unobtrusive as possible when observing the searcher's performance at the catalog, we cannot exclude the possibility that in some cases the natural development of a search might have been affected by the searcher's knowledge that he was being observed. Some searchers were perhaps more accurate and persistent than they habitually were, just to impress the interviewer. On the other hand, some searchers may have cut the search shorter than they would under normal circumstances, or may have suppressed responses which they feared would appear "silly". But even if the interview situation in some cases induced deviations from the typical behavior of the user, these were minor and of such a nature as not to alter the basic pattern of the search, i.e. the user's choice of entries and his strategy of searching.

The extent and accuracy of information elicited by the interviewer was satisfactory in the case of "known-item" searches. These are the searches for a particular item (book or other document) which is known to exist (although not necessarily in the library where the search takes place) and on which the searcher has such information as author and title. The collection of data did not present, for these searches, any great difficulty. In the case of "subject" searches, the user's behavior was not as simple to analyze and to record. These searches involve more complex forms of behavior than known-item searches. The searching process includes locating one or more entries in the catalog, screening the cards filed under those entries and selecting the relevant ones, consulting subheadings or cross-references. In addition, all sorts of personal devices and strategies are employed by the

users to reach their goals. At times the interviewers had difficulty in following all of what was going on, especially in the case of complex searches and of searchers not inclined to verbalize the details of their actions. While analyzing the data, we realized that occasionally there were gaps in the interview record, but in most cases we were able to reconstruct the missing steps.

The interviews were carried out between February 1968 and April 1969. Altogether we conducted 2681 interviews: 887 at the General Library,* 659 at the Undergraduate Library, 618 at the Medical Library and 517 at the Ann Arbor Public Library. Each interview covered a single catalog search. We planned to have a fixed number of subject searches (about two hundred) in the four samples. Since the ratio of subject searches to known-item searches turned out to be different at each of the four catalogs, we ended up with interview samples of different size.

THE CATALOG USER

As there are no prerequisites for the use of a library catalog, except a minimum level of literacy, the users of catalogs may form a very heterogeneous population. In a university library system, however, a certain degree of uniformity can be expected, at least in such characteristics as age, occupation, education, interests and information requirements. Among the many variables which characterize a population of users, we chose the following as being relevant to our study: occupational status, academic rank, sex, field of study, and frequency of catalog use.

Since the University of Michigan Library System serves not only college students, but also graduate students, teaching staff and research staff, the two first variables, status and rank, are important in defining the user population. The sex composition of a population of catalog users is, in itself, of no great interest, but may point to other differences which in turn are related to sex (e.g. subject of study or membership in an occupation having sex bias). Field of study, as indicative of the preferences and interests of the users, is an important variable and can be used to separate subgroups of students with interest in the same topics. After reviewing our data, we found that our fifth variable—frequency of catalog use—had doubtful validity as an indicator of proficiency in the use of the catalog. A better index may result from the combination of frequency of use with previous experience and perhaps with *attitude* towards learning a new skill. We may remark at this point that the level of skill demonstrated even by senior scientists is generally rather low, rarely going beyond the rules of alphabetization. It seems that years of familiarity with library catalogs have not induced a great sophistication in their use, which suggests that factors having to do with attitude, rather than skill, should be taken into consideration.

Our study shows that very similar populations use the catalog of the General Library and the Undergraduate Library. In both, the users are primarily university students, about equally divided between graduates and undergraduates, with a prevalence of students from the humanities and the social sciences. At the Medical Library, faculty and graduate students are found in higher proportions than in the other two libraries and they are primarily in the field of the biological sciences. At the Public Library, the two largest groups of catalog users are formed by high school students and by housewives. Women predominate in this library, and the catalog is used less often than at the university libraries.

* The catalog of the General Library includes catalog cards for the collection housed in the General Library and for the divisional library collections.

These data suggest questions which may be of some interest to those involved in planning or operating libraries. For instance: Why is the catalog at the Undergraduate Library not used primarily by undergraduates? If the Undergraduate Library was planned to serve an undergraduate population, does it fill this role? Is the inclusion in the General Library catalog of duplicate cards from the catalogs of all the technical and scientific divisional libraries (which have their own catalogs) justified in view of the fact that almost 85 per cent of the users of the General Library catalog are from the social sciences and the humanities? These are questions for the librarian to answer. Even if, from the point of view of the librarian, they are not the most important questions, they point out that studies like this one are required to gain much needed knowledge about the interaction of the user and the information retrieval system.*

TYPE OF CATALOG SEARCHES

Data on the ratio of known-item searches to subject searches performed on catalogs or other information systems are of great interest to information scientists, and it is easy to see why. The processes involved in the two types of searches are to a large extent dissimilar; consequently, the searching task requires two different sets of responses from both the user and the catalog. In nearly 90 per cent of the cases of known-item searches, all that is needed for the retrieval operation is an alphabetical listing of authors and titles on the part of the catalog, and some knowledge of filing rules on the part of the user. But the kind of cognitive behavior which is needed to catalog a book on the basis of its subject content, or to decide which subject heading will lead to that book, is much more complex and involves a fundamentally different mechanism.

The distribution of different types of catalog searches in the four libraries is shown in Table 1. The category "Other" includes searches for *any* book by an author, searches for a *shelf range*, searches for bibliographic information and other miscellaneous categories.

TABLE 1. TYPES OF CATALOG SEARCHES PERFORMED AT THREE UNIVERSITY OF MICHIGAN LIBRARIES AND THE ANN ARBOR PUBLIC LIBRARY

Search type	General Library		Undergraduate Library		Medical Library		Ann Arbor Public Library	
	No.	%	No.	%	No.	%	No.	%
Known-item search	636	71.7	448	68.0	405	65.5	256	49.5
Subject search	197	22.2	174	26.4	200	32.4	215	41.6
Other	54	6.1	37	5.6	13	2.1	46	8.9
Total	887		659		618		517	

We have investigated the relationship between the type of search and some of the characteristics of the users, namely their occupational status and academic rank. For the University of Michigan, the figures are as follows: at the General Library, 16.0 per cent of the graduate students who use the catalog do subject searches, as compared to

* The third question suggests that librarians re-examine the need for a centralized catalog which duplicates the records of each branch library catalog. The argument that such centralized catalogs are no longer cost-effective is further enhanced by the availability of MARC tapes [14] which would allow librarians to create duplicate records when needed or to obtain needed data directly from computer searches.

31.2 per cent undergraduate students; at the Undergraduate Library, 18.2 per cent of the graduate students and 33.1 per cent of the undergraduates do subject searches; and at the Medical Library the relative proportions are 26.9 per cent and 46.5 per cent. These figures are averages over a period of one semester or longer. Within the semester, fluctuations occur which are presumably related to exams, term papers, and breaks between terms. When the three library samples are pooled, and the academic rank of the undergraduates is taken into account, we have the following figures:

TABLE 2. STATUS AND ACADEMIC RANK OF USERS DOING SUBJECT SEARCHES

	All searches		Subject searches	
	No.	No.	No.	%
Undergraduate students:				
(a) Underclassmen	269	114		42.4
(b) Upperclassmen	596	189		31.7
Graduate students	930	184		19.8
Faculty	223	28		12.6

Upperclass college students do relatively fewer subject searches than underclassmen, graduate students do less than college upperclassmen, and faculty less than graduate students. The difference in subject search percentage between upperclassmen and underclassmen, and between graduate students and upperclassmen is significant at the 0.01 level of confidence; between faculty and graduate students the difference is significant at the 0.05 level of confidence.

We have tried to interpret these results in terms of the user's competence and specialization. The graduate student or the faculty member who is going to the library usually knows which books he wants to borrow or consult. He therefore uses the catalog primarily to locate particular items. The undergraduate student, on the other hand, is more likely to have little or no knowledge of the literature of a given field, and more often will rely on the catalog as a guide to identifying and locating books relevant to the subject of his search. Also, the pressure to keep up with the current literature differentiates the information-seeking behavior of the graduate student or the faculty member from that of the undergraduate. For the former, the periodical literature plays a larger role than the non-periodical; therefore, when he needs to make a subject search, he uses indexes, abstracts, journals, and other means for retrieving articles, in preference to card catalogs.

SUCCESS AND FAILURE IN CATALOG SEARCHING

To evaluate a catalog as an instance of a directory, we need performance measures. But we also need a clearer conceptualization of its functions. The primary function of the catalog is to point or to refer. A secondary function is to teach, i.e. to update the user's internal map of the structure of knowledge in the library holdings, insofar as the catalog embodies such structure. Though our main long-range interest is in directories as educational tools, our concern in this paper is entirely with the pointing function of the catalog.

Catalog and user form an interactive system: search failures may be due to deficiencies in the structure or functioning of the catalog as well as to inadequate performance on the part of the catalog user. We say that the catalog has succeeded in directing the user to an item (or items) that he needs, when the user notes the identification number of the item with the apparent intention of using it in searching the stacks or asking at the loaning desk. This is our operational definition of success. The user's success in using the catalog does not necessarily mean that the items he is directed to will meet his information needs.

When a catalog user wants to find a particular item in the catalog (known-item search), the goal of the search is clearly defined as that of locating a catalog card containing the description and identification number of that item. Ambiguity arises only in the case of searchers who, lacking adequate information, believe that they have located the reference that they were seeking, when in fact they have found something else. In such cases, it is hard to decide whether the search was successful. The searcher may believe that he has accomplished what he set out to do, but judged from an objective point of view (the matching of the *intended* reference with the corresponding catalog card), his search was a failure. This, however, happens rarely. Usually, the successful searcher is being directed to the document he was seeking.

In the case of subject searches, the meaning of *success* and *failure* is less clear. A subject search includes two distinct phases: in the first, the searcher attempts to match one or more "query" terms to catalog entries; in the second, he selects, out of the set of catalog cards filed under those entries, items which he thinks will meet his needs. Only if both of these actions have a positive outcome can the search be considered successful. The task of judging whether the first phase of the search was successful does not present serious difficulties; but an analogous judgment is not always possible in the second phase of the search. Here, our operational definition of success in terms of taking down the call number of books does not seem adequate. The apparent success may turn out to be completely illusory if the book, in spite of a promising title, does not contain the desired information. In other cases, although the title proves to be informative, the searcher may have misjudged the relevance of the title to the topic of his search. Furthermore, the judgment of relevance itself is subject to fluctuations, according to times and circumstances. The searcher may be influenced, for instance, by the size of the stack of cards that he is scanning. When a large number of books is indexed under a subject heading, he may be likely to adopt a higher standard of selection; when his choice is limited to a smaller number of cards, he may be inclined to relax his standard and accept books which, under other circumstances, he would have judged to be of dubious relevance to his query. In conclusion, the fact that the searcher has picked up one or more references from the catalog does not necessarily mean that he has obtained the information that he wanted.

On the other hand, the searcher's failure to select any item from the stack of cards filed under the subject heading of his choice is not a sure indication that the catalog did not have anything to offer. A number of reasons may have induced the searcher to reject relevant items: perhaps he had already read the books which were listed in the catalog or had decided to buy a copy of some of them, or the items were old and he wanted only recent material, or some of the relevant books were written in a language unfamiliar to him. In these cases the failure of the search should be ascribed to a higher threshold of acceptance rather than to the inability of the catalog to meet the user's needs.

In the following sections, we will discuss our results separately for the known-item searches and for the subject searches.

SUCCESS AND FAILURE IN KNOWN-ITEM SEARCHES

On the average, 28.1 per cent of the users failed to locate the desired item in the catalog. The figures concerning the distribution of success and failure in the four catalogs are shown in Table 3.

TABLE 3. RESULT OF CATALOG SEARCHES

Libraries	Known-item searches			
	Success		Failure	
	No.	% of all searches	No.	% of all searches
University of Michigan				
General Library	515	81.0	121	19.0
Undergraduate Library	298	66.5	150	33.5
Medical Library	287	70.9	118	29.1
Ann Arbor Public Library	155	60.5	101	39.5
Total	1255		490	

In a good number of cases the failure could not be ascribed to the searcher's faulty performance or to catalog deficiencies, but was simply due to the fact that the wanted book was not part of the library collection. If from the total we exclude this type of failure (we called it a *collection failure*), what is left are the cases in which the catalog card of the desired item was in the catalog but the user failed to find it (*user's failure*). Table 4 shows all the failures in the first column, and the user's failures in the second.

TABLE 4. SEARCH FAILURE AND USER'S FAILURE

Libraries	Failure of search		Failure to find existing entry	
	No.	% of all searches	No.	% of all searches
University of Michigan				
General Library	121	19.0	82	13.7
Undergraduate Library	150	33.5	27	8.3
Medical Library	118	29.1	76	20.9
Ann Arbor Public Library	101	39.5	12	7.2

Since the skills required to consult an alphabetically ordered list of names or titles are well within the abilities of a college population, we would expect that the number of failures due to searcher's errors and carelessness (assuming that the number of misfiled or missing cards is negligible) would be very low. But this is not what happened. The percentage of failures to find the catalog card—when the catalog card was in the catalog—is, especially in the case of the General Library and the Medical Library, rather substantial.

In about 90 per cent of the cases, either the author or the title of a book was used to search the catalog. But author and title information, as the searcher has it, is not always right. In order to test the relation between degree of completeness and correctness of the user's information about the author, and success of the search, we divided the items of our sample into four groups. To the first group belong searches in which the searcher had perfect information about the author of the item, i.e. the author's name was both correct and complete. In the last group the author's name was badly misshapen. In the two

intermediate groups the author's name had different degrees of imperfection. We have labeled the four groups "high precision", "moderate precision", "low precision", and "grossly defective". We did the same for the titles.* This analysis was made on the three University Library samples taken together, excluding items lacking personal author or having multiple authors ($N = 864$). Table 5 indicates the percentage of failures for the four author groups and the four title groups.

TABLE 5. FAILURE OF SEARCH IN RELATION TO DEGREE OF CORRECTNESS AND COMPLETENESS OF AUTHOR OR TITLE INFORMATION

Author	No. of searches	Failure of search (%)	Title	No. of searches	Failure of search (%)
Group I (high precision)	459	4.1	Group i (high precision)	670	5.8
Group II (moderate precision)	292	6.8	Group ii (moderate precision)	63	14.3
Group III (low precision)	67	14.9	Group iii (low precision)	93	10.7
Group IV (grossly defective)	46	28.3	Group iv (grossly defective)	38	10.5
	—			—	
	864			864	

It is apparent that a deterioration in author information, as represented by our groups II, III, and IV in this order, goes together with an increase in percentage of failures. Among adjacent groups, however, only the difference between Group II and Group III is statistically significant ($p < 0.05$). The differences in percentages of failure between non-adjacent groups (Group I and Group III, or Group I and Group IV, or Group II and Group IV) are all significant at the 0.01 level of confidence. In conclusion, the chances of success in the search for a given item are affected by the extent of the searcher's knowledge of the author's name. Considerable differences in user's information are needed, however, to produce significantly different results.

Information about the author seems to be a better predictor of success than information about the title. No consistent trend can in fact be noticed when one compares the percentage of search failures in the various groups of title information (from Group i to Group iv). With the exception of the pair Group i/Group ii, either there are no differences or the differences are not significant. Searches with titles in Group ii present a significantly higher rate of failure than searches with titles in Group i. Paradoxically, worse titles, such as those included in Group iv, do not increase the chances of failure.

On the basis of these data it seems legitimate to conclude that correctness and completeness of title play a less important role in the retrieval of known items than correctness and completeness of author's name. This may be due to the fact that titles are used less frequently than names to gain access to the catalog [3]. A defective title, i.e. a title with missing or wrong words, or with words in the wrong order, may be very difficult to match to a catalog entry, but it can still be useful in identifying the right entry, after the author's name has been located in the catalog.

* For details about the categorization of author and title information, see Ref. 3.

The joint influence of author and title correctness and completeness on search results is shown in Table 6. Only 13 failures (3.3 per cent) occurred among the 388 searches where the searcher had a complete, correct author's name (Group I) and a complete,

TABLE 6. FAILURE OF SEARCH IN RELATION TO AUTHOR AND TITLE INFORMATION

		Author	
		Group I (high precision)	Group II (moderate precision)
Title	Group i (high precision)	13/388 3.3%	10/208 4.8%
	Group ii (moderate precision)	1/23 4.3%	5/30 16.7%

correct title (Group i). But when both the author and title were defective, the percentage of failures went up to 16.7 per cent. (The difference is significant at the 0.01 level of confidence.) By itself, neither the decrease of information about the author (from Group I to Group II) nor the decrease in information about the title (from Group i to Group ii) was sufficient to alter significantly the chances of success of the search. Together, an imperfect author and an imperfect title produced a five-fold increase in the percentage of failures.

SUCCESS AND FAILURE IN SUBJECT SEARCHES

The first condition for the success of a subject search is that the catalog must have a subject entry which matches the query term generated by the user and that the user must find the entry and convince himself of its suitability. If he does, he has gained appropriate access to the catalog, and in this respect the search can be deemed successful.

We have to distinguish between the cases in which the query term is identical to the catalog entry (*exact match*) and those in which the one is an orthographic, morphemic or semantic variant of the other (*partial match*). In any partial match, at least the respective first words of the query term and of the catalog entry will correspond exactly or differ only slightly (as when "infections" was matched to "infection"). Because of the nature of the matching process, the entry must be very near the expected alphabetical position of the query term. In addition to morphemic variations, semantic variations may occur. These are of two kinds: (a) the query term and the entry term have the same number of words, but some words of the former do not match the corresponding words in the latter (as when "social pathology" was matched with "social medicine"); (b) the query term and the catalog entry differ in the number of words they contain, but the shorter of the two coincides exactly with a part of the longer (as when "sheet metal" was matched with "sheet metal work"). We called the first type of semantic variation *substitution variation* and the second *expansion variation*. These basic classes of variation may also appear in combinations, so that the complete scheme of partial matches is as follows:

1. Morphemic variation alone (e.g. infections → infection).
2. Substitution variation alone (e.g. social pathology → social medicine).
3. Morphemic + substitution variation (e.g. mental retardation → mentally handicapped).
4. Expansion variation alone (e.g. sheet metal → sheet metal work).

5. Morphemic + expansion variation (herbs → herb gardening).
6. Substitution + expansion variation (French theater → French literature–20th century –history and criticism).
7. Morphemic + substitution + expansion variation (Election laws → Elections–U.S.–statistics).

While some catalog users generated, in the course of a search, only one query term, others shifted from one query term to another, either because they were unsuccessful in matching the previous query terms to any catalog entry, or because they were not satisfied with the information gained through their previous access points, or because they wanted additional information. Table 7 shows the distribution of searches when the number of query terms used by the searcher is considered.

TABLE 7. NUMBER OF QUERY TERMS GENERATED BY SEARCHERS

No. of query terms per search	General Library		Undergraduate Library		Medical Library		Public Library	
	Searches		Searches		Searches		Searches	
	No.	%	No.	%	No.	%	No.	%
1	126	66.0	125	72.3	131	66.8	170	79.1
2	41	21.5	28	16.2	42	21.4	37	17.2
3	14	7.3	12	6.9	17	8.7	4	1.9
4	8	4.2	3	1.7	3	1.5	3	1.3
5	1	0.5	1	0.6	2	1.0	1	0.5
6 or more	1	0.5	4	2.3	1	0.5	—	—
	191		173		196		215	

Each query term generated by the user initiated a *trial*, i.e. a new attempt to gain access to the catalog by matching that query term to a catalog entry. All together there were 293 trials at the General Library ($126 \times 1 + 41 \times 2 + 14 \times 3 \dots$), 259 at the Undergraduate Library, 298 at the Medical Library and 273 at the Public Library. The percentages of exact and partial matches over the total number of trials are shown in Table 8.

On the average, then, about 75 per cent of the trials resulted in a successful matching (either "exact" or "partial") of the query term to a catalog entry. We can conclude that there was a relatively high degree of correspondence between the users' terminology and the subject catalog terminology, suggesting that there is a "naturalness" about the catalog's lexicon which comes to the aid of most users most of the time, although the times when it has failed may be more vivid in the mind of the frustrated user. It seems that the conventions and artifices adopted by catalogers to describe complex and multifaceted subject matter are less confusing to users than one might assume; whether this is due to the essential psychological soundness of the catalogers' decisions, or to the users having been educated to interact with catalogs and similar indexes remains to be seen.

We now turn from the discussion of the matching process at the level of the trial, to an analysis of the same process at the level of the search. While the former was useful in

TABLE 8. THE MATCHING OF QUERY TERMS TO CATALOG ENTRIES

Type of match	General Library N = 293	Undergraduate Library N = 259	Medical Library N = 298	Public Library N = 273
Exact match	66.6 %	55.6 %	57.7 %	50.0 %
Partial match	16.0 %	18.2 %	15.4 %	21.9 %
Total	82.6 %	73.8 %	73.1 %	71.9 %

providing the basis for a comparison between the user's query language and the indexing language of the system, the latter helps us to evaluate the degree of success of the users in gaining access to the catalog.

Table 9 shows the distribution of subject searches on the basis of number of matched query terms, ranging from no matches up to four matches per query. Subtracting from the total the searches which resulted in zero matches, we find that in 96 per cent of the General Library searches, 91 per cent of the Undergraduate Library searches, 89 per cent of the Medical Library searches and 86 per cent of the Public Library searches, at least one of the terms generated by the user found an exact or partial match in a catalog entry.

TABLE 9. NUMBER OF EXACT AND/OR PARTIAL MATCHES PER SEARCH

No. of matches per search	General Library		Undergraduate Library		Medical Library		Public Library	
	No.	%	No.	%	No.	%	No.	%
0	8	4.2	15	8.7	21	10.7	31	14.4
1	139	72.8	135	78.0	141	71.9	169	78.8
2	33	17.3	15	8.7	28	14.3	14	6.5
3	8	4.2	6	3.5	5	2.6	1	0.5
4 or more	3	1.6	2	1.2	1	0.5	—	—
Total	191		173		196		215	

The searches in which no query term was matched to catalog entries—zero matches—evidently resulted in failure. What about the searches in which one or more matches occurred? Finding a point of entry into the catalog is a necessary but not sufficient condition for the success of the search. Some of the searchers who succeeded in matching one or more query terms to catalog entries failed to select any book among those listed in the part of the catalog to which they gained access. This, according to our operational definition of success, was a search failure. The percentage of unsuccessful searches is generally higher than the percentage of searches with zero matches, as one can see in the table below.

TABLE 10. ZERO-MATCH SEARCHES AND UNSUCCESSFUL SEARCHES

Libraries	Zero-match searches	Unsuccessful searches
University of Michigan		
General Library	4.2 %	15.7 %
Undergraduate Library	8.7 %	19.1 %
Medical Library	10.7 %	14.8 %
Ann Arbor Public Library	14.4 %	14.4 %

Success or failure of the search results from the interplay of two basically distinct sets of variables. The first set, which is related to the matching of the user's query term to a catalog entry, may include such things as the skill and resourcefulness of the user, his persistence, his familiarity with catalog rules; it may also include the availability of subject headings relevant to the topic of his query and the availability of cross-references connecting related subject headings in the area of the search. The second set of variables has to do with the user's acceptance of the retrieved material and therefore concerns the user's expectations and attitudes, his satisfaction with the retrieved information, his knowledge of the topic and his ability to discriminate between what is relevant and what is not. This is borne out by the data of Table 10. At the General Library, only 4.2 per cent of the searchers failed to find any suitable match for their query terms; however, 15.7 per cent of the searches were unsuccessful, which means that 11.5 per cent of the searchers succeeded in matching at least one of their query terms to a catalog entry, but, for one reason or another, were not satisfied with the documents that they found listed under that entry. For the Public Library users, by contrast, the number of unsuccessful searches equaled the number of failures to gain access to the catalog, suggesting that when these users succeeded in matching their query terms to catalog entries, they always found some book which satisfied their needs.

STYLE AND STRATEGIES IN CATALOG SEARCHING

Different catalog users may pursue the task of locating a particular item or a particular subject in different ways. It would be hard to determine, with the data that we have, how different users respond to the same search problem, since only in a few cases was the same search performed by more than one individual. In the course of the search, however, catalog users are confronted with similar options concerning the steps that they are going to take. It is here—i.e. in the choice of courses of action—that we are going to look for similarities or differences in the behavior of the users.

If we define "strategy" as the planned use of available repertoires of responses in order to achieve a given goal, we can talk of strategy in catalog searches, although at times the user's behavior appeared to be more the result of trial and error than of accurate planning. By "style" we mean a consistent way of behaving which is characteristic of an individual or a group. Style in regard to catalog searching appears to be the expression of the searcher's inclination to act in a certain way under given circumstances. We are not completely sure that the distinction between strategy and style, in the case of catalog searching, is always valid. We feel that questions of strategy apply particularly to known-item searches, where the goal is well defined and the steps to be followed clearly marked, while questions of style seem to be more pertinent to "subject" searches, in which the searching situation is somewhat fluid and less apt to be defined in terms of discrete steps or an algorithm.

The user who is looking for a particular item in the catalog has to choose a point of access to the catalog in order to locate the item. If he has the complete reference of the item, at least three choices are open to him: (a) the author (or editor); (b) the title; (c) a subject heading. If his information is incomplete, his options are more limited. This part of the study is described in detail elsewhere [3]. Here we want to report only the final results and conclusions. In their first attempt, nearly 2/3 of the users chose the author as access point to the catalog, while less than 1/3 chose the title (see Table 11). In the second and third attempts, author and title are about equal as access point choices. The subject

TABLE 11. SEQUENCE OF CHOICE OF ACCESS POINTS

	1st Trial	2nd Trial	3rd Trial
	(N = 1718)	(N = 295)	(N = 62)
Author (or editor)	65.5 %	34.5 %	19.3 %
Title	28.9 %	37.3 %	17.7 %
Subject heading	5.6 %	28.1 %	62.9 %

heading approach, which in the first trial had been used in only a small percentage of cases, became the preferred method of searching the catalog in the third trial.

The figures above include searches in which the user could choose and searches in which the user could not choose between author and title. (For example, the user's options were limited if his item did not have an author, as in the case of a journal, or when his reference was incomplete.) Of the users who *did* have a choice between author and title, 85.2 per cent started their search with the author. There seems to be a marked preference, therefore, for the author over title as first access point. The study showed that even when the user had extremely poor or totally wrong information about the author, he still chose it in preference to his *perfect* title. Some speculations about possible reasons for this seemingly dysfunctional behavior are reported in Ref. 3.

Another interesting observation we made concerns what we called the "perseverant behavior" of the user, i.e. the number of repeated attempts to locate the item by the user who met with failure. We defined the "perseverance index" of a given population of searchers as the ratio of the number of searchers who continued the search to the number who failed. This index was, for the three University Libraries taken together, 0.459 after the first trial, 0.368 after the second trial, and 0.178 after the third trial. It was rather surprising to discover that almost half the users who failed in their first attempt gave up the search. The reasons which may have determined this early discouragement are discussed in Ref. 3.

The data on subject searches are more difficult to analyze than those on known-item searches. Some of the difficulties lie in the fact that the searcher has a much larger repertoire of available responses and that his goal is less well defined. Furthermore, the extent of the effort that the user spends in searching the catalog, and therefore the length of the search, is much more variable than in the known-item searches. Some catalog users give up the search when their first attempt to locate a suitable subject entry fails, while other users do not stop even after success, but go on happily collecting information under a variety of subject headings.

In Table 7, we have shown the distribution of searches on the basis of number of query terms generated by the user. From the number of query terms per search we can also infer how many *trials* there were in each search, since each new query term marks the beginning of a new trial. The majority of users generated only one or two query terms per search. The percentage of searches containing multiple query terms (3 or more) is very similar at the three University Libraries (12.5 per cent, 11.5 per cent, 11.7 per cent), but is much lower at the Public Library (3.7 per cent).

The matching of query terms to catalog entries is not the only source of suitable subject headings for the user. After the user has gained access to the catalog, he may discover subject headings that are as relevant to the topic of his search or more relevant than the initial entry. These are subject headings which happen to be in proximity with the subject

entry, or which are connected to it by cross-references. In these cases, the clues for extending the search to additional subject headings have been generated by the catalog, rather than by the user; they have been, so to speak, *suggested* by the catalog.

We will call a "pass" the act of exploiting the alphabetical proximity of conceptually related entries or the cross-references which link them, to arrive at further usable subject headings. The number of passes which follows each matched query term provides us with a rough measure of the extent to which the user utilizes the associational structure of the catalog. In Table 12, we tabulated the distribution of number of passes following any query term/subject heading match. We excluded from the count those catalog-generated subject headings which could be considered "form" subject headings, i.e. those sub-headings which have the function of singling out sets of documents which present differences in type of publication or type of presentation but, conceptually, can be considered to belong to the same main heading. (The sub-heading consists of the main heading followed by such an expression as: "Addresses, essays, lectures, etc.", or "Congresses", or "Bibliography".)

TABLE 12. NUMBER OF PASSES PER QUERY TERM

No. of passes per query term	General Library		Undergraduate Library		Medical Library		Public Library	
	Query terms		Query terms		Query terms		Query terms	
	No.	%	No.	%	No.	%	No.	%
0	172	58.7	155	59.8	186	62.2	158	57.9
1	78	26.6	63	24.3	62	20.7	66	24.2
2	22	7.5	21	8.1	28	9.4	32	11.7
3	15	5.2	11	4.3	10	3.3	9	3.3
4	2	0.7	7	2.7	9	3.0	3	1.1
5	2	0.7	2	0.8	2	0.7	2	0.7
6 or more	2	0.6	—	—	2	0.6	3	1.1
Total	293		259		299		273	

On the average, in about 40 per cent of the cases, the searchers made use of catalog-generated subject headings as a basis for continuing the search. It may be concluded, therefore, that more than half of the users doing subject searches are satisfied with self-generated query terms and do not feel the need to follow any catalog "leads". It is also possible that some users fail to try catalog-generated terms because they lack confidence in the catalog's associational structure or do not understand its use, or perhaps because the catalog is inadequate in the area of their search.

Do searchers differ in the extent of their reliance on self-generated query terms or on catalog-suggested subject headings? Could we define a "self-reliant" style of searching as opposed to a "catalog-oriented" style? We certainly have found that some users are inclined to generate a large number of entry terms and to ignore the catalog as a source of suitable related subject headings, while other users generate just one entry term, and then painstakingly follow all possible links connecting that entry to other subject headings of the catalog. On the basis of our data, it would be very difficult to say if this difference in behavior is due to user differences, or if it is the result of the demands of different search problems, or of the different catalog resources in a particular area, or some combination

of these factors. Only with a group of searchers all engaged in the same search problem could one get some insight on the roles of these variables. But this would require a different setup from that in which our study was conducted. To explore the problem under controlled experimental conditions, we have recently started an investigation of the behavior of a group of college students who are given a task which simulates a catalog subject search. The results of the experiment are in the process of being analyzed.

FROM THE CATALOG TO THE STACKS

Although what happens after the catalog searches are completed is not of direct concern to this study, we thought that it would be interesting to obtain some estimate of how successful the catalog users were in finding the books to which the catalog had directed them. A follow-up questionnaire, containing questions on the outcome of the search at the stacks, was given to the users who had successfully completed a subject search. The questionnaire was limited to a maximum of four book searches. Ninety-nine questionnaires were completed and returned from the Medical Library users, 81 from the Undergraduate Library users. The following table illustrates the results.

TABLE 13. SEARCH FOR BOOKS IN THE STACKS OF THE MEDICAL LIBRARY AND THE UNDERGRADUATE LIBRARY

Medical Library					Undergraduate Library					
Number of books found	searched				Number of books found	searched				
	4	3	2	1		4	3	2	1	
4	9	4	6	18	4	6	9	10	11	
3	6	4	6	16	3	10	5	4	11	
2	12	4	6	22	2	9	5	10	11	
1	11	3	10	24	1	4	—	4	11	
0	1	1	4	16	0	1	—	2	10	
	39	12	20	28	99	30	14	16	21	81

About 16 per cent of the users, then (16 at the Medical Library and 13 at the Undergraduate Library), did not find any of the books they were looking for. Thirty-seven searchers at the Medical Library (37.6 per cent) and 36 at the Undergraduate Library (46.6 per cent) found all the books, up to a maximum of four, to which the catalog had directed them (complete success). The remaining searchers had partial success in locating the books on the shelves.

A catalog search which leads the searcher to books that he reads or consults should be considered more successful than a search leading him to books that he discards as irrelevant or not appropriate to his needs. Our follow-up questionnaire was designed with

the intention of gaining information on this point. Potential usefulness of a book was established on the basis of the searcher's answer to either of two questions: "Did you check out the book?" and "Did you take notes from it?" Forty-three of the 83 searchers who were able to locate one or more books on the shelves of the Medical Library either checked out or took notes from all the books that they had located; 18 searchers checked out or took notes from none of them; the remaining 22 searchers were in between. At the Undergraduate Library, 28 out of 68 users utilized all the books that they had located; 21 utilized none of them; and 19 were in between.*

To obtain the total number of failures at the stacks, the 18 Medical Library searchers and the 21 Undergraduate Library searchers who found the books on the shelves but did not utilize them are to be added respectively to the 16 and the 13 users who failed to find any of the books that they were searching for (see Table 13). The failures encountered at the stacks are therefore 35 per cent of the total at the Medical Library, and 42 per cent at the Undergraduate Library. If we combine these failures with the failures in catalog searching (14.8 per cent at the Medical Library and 19.1 per cent at the Undergraduate Library), we reach an estimate of almost 50 per cent unproductive searches at the Medical Library and 60 per cent at the Undergraduate Library.

In conclusion, on the basis of our sample of subject searches and the follow-up questionnaires, we can state that at the libraries under study, over half of the searchers who were using the catalog to reach potentially useful documents met with failure somewhere along the way.

To brighten this somewhat gloomy picture, we have the fact that a good number of searchers, although unable to locate on the shelves the books to which the catalog had directed them, or unable to utilize the books they had located, found other books which apparently had some potential usefulness. To the question "Did you examine any adjacent books whose call numbers you had not written down?" 80 of the 99 Medical Library users and 62 of the 81 Undergraduate Library users answered in the affirmative. These totals include searchers who had met with failure in the "directed" search as well as searchers who had met with success. Twenty-six of the Medical Library users and 17 of the Undergraduate Library users also said that they had checked out one or more of the books that they had found by browsing (we do not have information, unfortunately, on how many searchers utilized these books at the library by taking notes). We may conclude, then, that browsing in the stacks within the limited range determined by similarity of topic is a widespread phenomenon [15]. In our sample, it occurred in 80.8 per cent of the subject searches at the Medical Library, and in 76.5 per cent of the subject searches at the Undergraduate Library.† Of the searchers who browsed, 39.5 per cent at the Medical Library and 27.4 per cent at the Undergraduate Library checked out books that they had found by browsing.

* We cannot exclude the possibility that some of the searchers who neither checked out nor took notes from a book, utilized it by reading it at the library. Our estimate of failures at the stacks may thus be a little higher than it should be.

† It is possible that browsing has contributed to produce a number of completely successful multiple searches which is consistently higher than what would be expected on the basis of failure/success ratio in single searches (see Table 13). Searches for two books at the Undergraduate Library, for instance, where such ratio was approximately 50/50, could hardly have resulted in 10 complete successes, 4 partial successes and 2 failures (instead of the expected 4-8-4) unless some bias was present. We suspect that in some cases the searchers did not discriminate between books for which they had catalog call numbers and books which they picked up by browsing.

CONCLUSIONS

The insight gained from the analysis of library catalog use can provide valuable guidelines for building catalogs, indexes, and other types of directories for the retrieval of information from large stores of recorded knowledge. Our survey of the use of three University of Michigan library catalogs and one public library catalog clarifies some important points about the user's interaction with the catalog, his strategies and his style of searching. We summarize and briefly discuss some of the results:

The catalogs of University libraries are used primarily for known-item searches and secondarily for subject searches. In the three University libraries of our study, the ratio was, on the average, three to one. There is a relation between the academic rank of the catalog users and the type of catalog searches that they perform: college upperclassmen perform relatively fewer subject searches than college underclassmen, graduate students fewer than college upperclassmen, and faculty fewer than graduate students. The reverse applies to known-item searches. This suggests that if a library serves a predominantly undergraduate student population, its catalog's function is, to a large extent, *educational*, in the sense that we discussed on page 367: it updates the user's internal map of the information on a given topic available in the library. But if the catalog users are graduate students or faculty members, the *pointing* (or referring) function of the catalog assumes a central role. The ability to predict, on the basis of the composition of a catalog user population, the frequency of occurrence of a certain type of catalog search may improve catalog design; it may also be helpful in determining the structure of other information retrieval systems.

The completeness and correctness of the user's information about the author and title of the book he is seeking is important for the success of the search, even more in the case of the author than in the case of the title. This, of course, can be explained by the fact that most users prefer to use the author name to retrieve books, even when they have better information on the title than on the author. There is some rationale for this behavior: with few exceptions, books always have a catalog entry for their authors (or editors), but not always a catalog entry for their titles. (We estimated that at the university catalogs used in our study, about 30 per cent of the books published after 1930 were missing the title added entry.) And, of course, while incompleteness of the author name is easily perceived by the user, incorrectness often goes unrecognized. (Misspellings often produce different but plausible family names.) The searcher, therefore, may not realize that his title information is better than his information about the author. Furthermore, many catalog users are unwilling to learn any of the catalog conventions and filing rules, and so feel safer in searching the catalog with the author's name, whose use does not require any other knowledge than that of the alphabet.

If, as some studies suggest [5, 8, 16], the efficient use of library catalogs requires that a larger role be assigned to title searching, perhaps we should contemplate a re-education program for catalog users and for librarians. On the other hand, if the title entry is found to be a less efficient means of retrieval than the author entry, greater emphasis should be placed on obtaining exact author references. The feasibility of computer-aided author verification should be investigated and studies on correction codes for proper names should be encouraged [17, 18].

The subject approach is seldom used in the first attempt to locate a specific item in the catalog. Its use increases after one unsuccessful trial and becomes predominant (about 63 per cent of the cases) after two unsuccessful trials. The use of a subject heading to retrieve a specific document represents an overlap between known-item searches and subject

searches. A catalog or directory, designed to retrieve specific documents only, could certainly be built more cheaply if one were to forgo the demanding and costly process of content analysis and surrogation needed for subject indexing; but in so doing one would abandon a powerful instrument for the retrieval of items with imperfect author and title.

The perseverance of the searchers who look for a specific document does not seem to be very high: in the population that we studied, more than half the searchers who failed in their first attempt to locate a book gave up the search. The reasons for this early discouragement should be investigated more deeply. Studies concerning the motivation of catalog users, their resourcefulness, their skill and their knowledge may have implications for estimating the need for sophisticated query-negotiation devices in future information retrieval systems [19].

About 75 per cent of the query terms used by the searchers who did subject searches matched a catalog entry, either exactly or partially. The majority of searchers in our sample generated one or two query terms in the course of their search, but some of the searchers produced as many as six or seven query terms. In over 90 per cent of the subject searches performed at the University of Michigan library catalogs, the searcher obtained at least one successful match and thus gained access to the catalog. A shift from one query term to another is not necessarily the consequence of failure in the previous trial; it often indicates the searcher's intention to obtain additional references. The principles regulating the number and type of query terms generated by the catalog users are at present under investigation in a simulated catalog search situation.

Of the searchers who met with failure in their first trial, a larger percentage persisted (about 70 per cent) than was the case among searchers for known items. In about 40 per cent of the cases, leads provided by the catalog helped the searcher. These leads included subheadings and main subject headings which happened to be in proximity to the subject entry or were connected to it by cross-references.

Acknowledgements—This work was supported in part by the National Science Foundation GN716.

We are indebted to Dr. F. H. Wagman, Director of the University of Michigan Libraries, and to Mr. H. R. Chance, Director of the Ann Arbor Public Library, for allowing us to carry out our survey of the catalog users. We thank the following persons for their assistance in the collection and analysis of the data, and for their contributions to the design of the study and the discussion of the findings: Kathie Coblenz, Nancy Green, Peter Headly, Betty Linschied, Lawrence Rosenberg, Susan Ryan, Dorothy Semmel, Arvo Tars, and William Van Loo.

REFERENCES

- [1] F. G. KILGOUR: *A. Rev. Inf. Sci. Tech.*, 1969, 4, 305-337.
- [2] M. KOCHEN: *Integrative Mechanisms in Literature Growth*. Final Report to the National Science Foundation, GN716, Mental Health Research Institute, The University of Michigan, Ann Arbor, January, 1970.
- [3] R. TAGLIACCOZZO, L. ROSENBERG and M. KOCHEN: *J. Docum.* To be published (September, 1970).
- [4] R. TAGLIACCOZZO *et al.*: Patterns of Searching in Library Catalogs. In: M. KOCHEN, *Integrative Mechanisms in Literature Growth*, Vol. 2, Part IV. Final Report to the National Science Foundation, GN716, Mental Health Research Institute, The University of Michigan, Ann Arbor, January, 1970.
- [5] F. H. AYRES, J. GERMAN, N. LOUKES and R. H. SEARLE: *J. Docum.*, 1968, 24, 266-272.
- [6] M. BLACKBURN: *The Information about Books Known by Users of the Catalog who are Looking for a Particular Work*. Unpublished Master's Dissertation, Graduate Library School, University of Chicago (1968).
- [7] W. S. COOPER: *J. Am. Soc. Inf. Sci.*, 1970, 21, 112-127.
- [8] W. A. HINKLEY: *On Searching Catalogs and Indexes with Inexact Title Information*. Unpublished Master's Dissertation, Graduate Library School, University of Chicago (1968).
- [9] B. A. LIPETZ: *User Requirements in Identifying Desired Works in a Large Library*. Final Report to the U.S. Department of Health, Education, and Welfare, Grant No. SAR/OEG-1-71071140-4227. Yale University Library, New Haven, Connecticut, June, 1970.

- [10] E. MONTAGUE: Card Catalog Use Studies, 1949–1965. Unpublished Master's Dissertation, Graduate Library School, University of Chicago (1967).
- [11] The University of Chicago, Graduate Library School. *Requirements Study for Future Catalogs*. Report No. 2 to the National Science Foundation, March, 1968.
- [12] H. MENZEL: *A. Rev. Inf. Sci. Tech.* 1966, **1**, 41–69.
- [13] R. ROSENTHAL: *Experimenter Effects in Behavioral Research*. Appleton, New York (1966).
- [14] H. D. AVRAM: *Drexel Libr. Q.*, 1968, **4**, 279–309.
- [15] H. H. FUSSLER and J. L. SIMON: *Patterns in the Use of Books in Large Research Libraries*, Chapter VII. The University of Chicago Library, 1961.
- [16] J. Z. NITECKI: *Coll. Res. Libr.*, 1968, **29**, 431–436.
- [17] C. N. ALBERGA: *Commun ACM*, 1967, **10**, 302–313.
- [18] J. L. DOLBY: An Algorithm for Noisy Matches in Catalog Searching. In: *A Study of the Organization and Search of Bibliographic Holdings Records in On-line Computer Systems: Phase I*. J. L. CUNNINGHAM, W. D. SCHIEBER, and R. M. SCHOFNER, Institute of Library Research, University of California, Berkeley, March, 1969, pp. 119–136.
- [19] R. S. TAYLOR: *Coll. Res. Libr.*, 1968, **29**, 178–194.