

A Least Cost Searching Sequence

By GERALD J. LAZORICK and THOMAS L. MINDER

IN EVALUATING the procedures of the acquisitions department of the Pennsylvania State University library system, the cost of searching information to order a book was found to be \$.73. In addition, an order form had a rather high in-process time. An analysis of the flow of order forms through the department showed that a bottleneck was occurring with the searching procedure. As a result of this many order forms were being marked "rush," and routine items were faced with an even higher in-process time. The library was faced with adding searchers or with accepting the fact that there would almost always be a backlog of unprocessed orders. The latter would result in a high in-process flow time thus reducing service to patrons and requiring that needed course material be ordered well in advance of the time when it was actually needed.

The technical services librarian, the catalog librarian, an industrial engineer, and the librarian in charge of library systems research met and compared the amount of searching needed to identify adequately a publication to be purchased with the amount of searching needed for complete bibliographic information. At the time, it was decided that the library would adopt for a trial period an "adequate information" philosophy. The possibility that duplicate copies would occasionally be received was considered together with the increased probability of obtaining the wrong books. However, the fact that searching time would be considerably reduced thus increasing the number of possible searches in a given period of time was assumed to outweigh

Mr. Lazorick is Industrial Engineer, and Mr. Minder is Engineering Librarian, in Pennsylvania State University.

the disadvantages in accepting the philosophy.

The definition of adequate information was also discussed. It was decided that adequate information would include the author, title, publisher, and date of publication. It was also assumed that the latest edition would always be ordered unless otherwise specified.

With this in mind, a study of the searching procedure was made by the industrial engineer. Out-of-print items and foreign language items were not considered in the study. Out-of-print items required special processing by acquisitions personnel, and foreign items required searching in a large variety of bibliographies. In addition, the bulk of the books ordered by the acquisitions department were in the English language.

The searching sequence was considered to be an area for study. What was the optimum searching sequence? Realizing that one bibliographic source would contain more information than another and that the time for searching varied considerably with the source used, a method was needed to determine in what sequence bibliographic information should be sought. In other words, should a searcher first look in the LC catalog, and if adequate information was not found then continue the search in *Books In Print*, and again if not found, continue in the *Cumulative Book Index*, etc., or should some other sequence be used?

It was decided that *Books In Print*, *Publishers' Trade List Annual*, *LC proof slips*, *Cumulative Book Index*, and the LC catalog would be studied as the sources of bibliographic information. The two variables which had to be measured were the per cent of time in which adequate information was found in each of the sources and the average time to complete a search in each of the sources. The first of these, the percentage, is really the probability of finding adequate information in each of the sources.

THE STUDY

A random sample of twenty-five English language order slips were selected. A searcher was asked to take the twenty-five slips and check each of the five sources to see if the items could be identified with adequate information. The percentage of time adequate information was found was recorded for each bibliographic source. Another searcher was then given five random samples of ten order forms and was asked to search the samples in the five bibliographic sources while being timed separately for each source. The outstanding order file and the public catalog were not studied, since previous studies indicated that about 20 per cent of the orders received in acquisitions were for books on order or already in the collection. These sources would always be searched first. *Books In Print* and *Publishers Trade List Annual* were both included for control and comparison purposes. The results of the study are summarized in Table 1.

The second column in Table 1 shows the average time of a search in minutes. The third column shows the probability of finding the item searched. It should be noted that the probability of not finding adequate information after searching all five sources is .03. That is, we can expect that 3 per cent of the books searched will not be identified if we search only the five sources.

TABLE 1
SEARCHING SUMMARY

Source	Time to Search	Probability of a Successful Search
Library of Congress Catalog	3.24	.72
<i>Cumulative Book Index</i>	2.14	.60
<i>Books In Print</i>	.56	.40
<i>Publishers' Trade List Annual</i>	1.09	.36
Library of Congress Proofs	.94	.28

There are two commonly accepted philosophies in this type of sequencing situation. One person would argue that we should put the least time-consuming source first in the sequence, the second time-consuming source second, and the most time-consuming source last; whereas others would argue that the source with the best chance of a successful search should be put first, and the source with the least chance of a successful search last. However, neither of these philosophies will necessarily result in the optimum sequence, that is, the least time-consuming sequence. One could argue that all combinations should be evaluated. However, there would be 51 or 120 possible combinations. Fortunately, operations research and industrial engineering have solved this sequencing problem for an analogous industrial application. The optimum sequence in this instance can be determined by taking the ratio of the time consumed to the per cent of success and ordering the tests such that the ratios will be in increasing sequence. The ratio for each of the bibliographic sources is shown in Table 2.

TABLE 2
SEARCHING SEQUENCE RATIOS

Source	Ratio
Library of Congress Catalog	4.50
<i>Cumulative Book Index</i>	3.56
<i>Books In Print</i>	1.40
<i>Publishers' Trade List Annual</i>	3.03
Library of Congress Proofs	3.36

The optimum searching sequence together with the "least time first" sequence

TABLE 3
POSSIBLE AND OPTIMUM SEQUENCES

Least Time First	Most Success First	Optimum
BIP	LC Catalog	BIP
LC Proof	CBI	PTLA
PTLA	BIP	LC Proofs
CBI	PTLA	CBI
LC Catalog	LC Proof	LC Catalog

and the "most success first" sequence are shown in Table 3.

Based on this study, it was pointed out that adoption of the optimum searching sequence would result in 27 per cent more searches per year of English language publications. Certain exceptions to the procedure should be followed based on certain idiosyncrasies of the sources. For example, the publications listed by an obscure scientific society would rarely be listed in *Books In Print*. Therefore, one would be wasting his time searching in *Books In Print*.

Special considerations were also given to the LC proof file, a rather recent addition to the Pennsylvania State acquisitions department. The other sources are rather stable bibliographic sources which are not likely to change in the searching sequence in the years to come. Since the use of the LC proof file is still in its infancy at Pennsylvania State, it is expected to grow in size and usefulness. It was therefore recommended that the LC proof file be re-evaluated annually as a searching tool.

One weakness of the study should be noted. In order to make the times and percentages more accurate, a larger sample would have to have been taken. This was not done because neither the industrial engineer nor the searchers had the time to gather as much information as would have been necessary. Further data is being collected, and the searching pro-

cedure will continuously be re-evaluated.

The least-cost searching sequence method is derived from a methodology described by L. G. Mitten.¹ This method also has promise in other areas of library work. For example, the reference staff might establish a card file of searching sequences for certain general types of questions. The searching sequence preferred would be determined by the probability of searching success in specific sources and the searching time through these sources. The sequence might also be determined by the other variables such as the level of competence of the requester or the librarian.

To carry this one application a step further, one would expect its most fruitful area of use in the middle ground between the very special narrow topic and the extremely broad "encyclopedia" topic. These two extremes usually direct the searcher to the obvious sources. There is usually not a choice of six or eight different sources and sequences for this type of search.

Another possible application lies in the area of "state of the art" searches. The questions, "What is the probability of finding useful information in source X since date Y?" and "What is the expected redundancy between indexes V and W?" are tied in very closely with the economics of literature searching versus laboratory experimentation.

Least-cost searching sequence methodology might profitably be applied to the oft-stated proposition that it is cheaper to re-invent than to search the literature. It might also reduce the cost of searching.

¹ L. G. Mitten, "An Analytic Solution to the Least Cost Testing Sequence Problem," *Journal of Industrial Engineering*, XI (January-February 1960), 17.