THE EVALUATION OF END-USER ONLINE SEARCHING: A REVIEW OF LITERATURE

Lai Ting-Ming *

INTRODUCTION

I. Need for Research

Traditionally, owing to complicated search procedures and database structure, online information retrieval has been conducted primarily by intermediaries, called variously librarians, search analysts, or information specialists. Recently, the situation has begun to change: online searching is not likely to remain the exclusive province of information specialists again.

For instance, some information searchers have begun to do their own online searching. They are tired of making tedious appointments with intermediaries who spend most of the time discussing and interviewing, and of trying to make their needs understandable to the intermediary.\(^1\) Furthermore, if end-users choose not to be present during a search, they have to accept any biases in the intermediary’s interpretation of the query. This approach tends to distort the information requirement, and the users may have lost the opportunity to modify the initial question during the search.\(^2\)

Also, modern society is quite volatile, and this situation is reflected in the proliferation of literature. It is very difficult for the librarian or information specialist in such a society to provide effective in-depth support to the users across all disciplines. Richardson indicates the “the day of the Universal Renaissance Engineer clearly is over and, sadly, so is that of the Universal Renaissance Information Specialist.”\(^3\)

If technical complexity is leading us to a crisis of confidence

---

* 著者為美國威斯康辛大學麥迪遜校區圖書館與資訊科學研究所研究生。
in delegated searching, what are our alternatives? The answer is in the end-users themselves. The advent of new end-user products has made it possible for library users to conduct their own searches. In addition, some databases are being established to meet end-user needs for easy use and frequent searching from selected sources. Some systems are capable of downloading or allow leasing magnetic tapes of specific portions of files from database producers. These trends have made it possible for libraries to install user-friendly databases at the library and help users access the online searching easily and quickly.4

It seems that the era of end-user searching is coming. However, before any strong assertion of this tendency can be made, many problems still have to be understood in order to help end-users obtain optimal results. For example:

1. Can a system be designed for use by the naive user?
2. What are the characteristics and behaviors of end-users in online searching?
3. In what ways can the end-user's performance be improved?

Many studies have been conducted in the above areas since 1970. The purpose of this article is to review this research, to present what has been done and to discuss the techniques and methodologies used. It will also discuss the results of these studies, to present the characteristics of end-users in online searching. Finally, this article will indicate what studies should be conducted in the future in order to cope with the coming of the era of end-user online searching.

II. Definition of Terms

Command-driven system—-an online interactive computer system in which the searcher communicates with the system using a command language. Command can be entered when the system has displayed a prompt. Compare with “menu-driven.”

Delegated search—-one in which the person with the information
need delegates the responsibility for searching the database
to one or more information specialists or intermediaries.
Do-it-yourself online searching—searching that is conducted by
end-users themselves.
Downloading—the process of transferring part of the records
from a major databases to a storage medium connected to
the user's computer.
Front-end system—a system that is used between the user and
database to offer short searches on current topics to end-
users, for training and introducing online searching to end-
users.
Gateway system—microcomputer software that permits online
searches to be conducted by naive users who know little or
nothing about information retrieval.
Intermediary—the librarian or information specialist who con-
ducts searches for users.
Menu-driven—this typically means that when one first boots
the program one will be presented with a range of numbered
choices. Selection is made by keying in the appropriate
number and then being presented with a second menu with
other choices. The process continues until one reaches
whatever point was aimed at when the search was begun.
Compare with "command-driven."
Nondelegated search—the user goes directly to the database. In
this situation the user must conceptually analyze the in-
formation need and translate the analysis into the language
of the system.
Online searching—the process in which a human being uses a
computer terminal to interact with a search service, in an
attempt to satisfy an information need.
Precision ratio—number of relevant documents retrieved among
the total number of documents retrieved.
Recall ratio—number of relevant documents retrieved among the
total number of relevant documents in the database.
User-friendly system—a system with which untrained users can
interact easily. Essentially, user friendly is supposed to mean easy to use.

III. Scope of Literature

This review of end-user online searching will confine itself to research reports published since 1969, as opposed to opinion papers. Dealing only with untrained end-user interactions with online systems; the interactions of intermediaries with online systems will not be discussed in this review.

METHODODOLOGY

I. Sampling

Most end-user studies are limited due to sampling procedures. The sample size of much research within this area is rather small, not reaching the minimum of 30 for reference to the normal distribution. For example, the sample size in Haines was 15, Olson 16, Trzebiatowski 20, Richardson 20, Vollaro and Hawkins 22, Janke 25, Kirby and Miller 27. Among these, only Olson manipulated t-statistics to test the results.

In addition to the problem of sample size, the construction of the sampling frame is crucial. Some studies chose the sample from actual users, not potential users. This portion includes Olson, Vollaro and Hawkins, Janke (1983), Janke (1984), Horowitz, Broering. Furthermore, most of the resources did not select their samples randomly or did not explain their procedures explicitly. Therefore, those results should be doubted in some way.

II. Data Collection and Analysis

Some of the studies used questionnaires to collect data,
including Richardson, Janke (1983 and 1984), Hansen, Trzebiatowski, Vollaro and Hawkins, Harman. Some studies collected data from online search printouts, including, Olson, Kirby and Miller, Lancaster (AIM-TWX), Lancaster and Rapport and Penny, Barber and Barraclough and Gray. Haines used search records and monthly invoices, Sewell and Teitelbaum’s report used transaction log, interview and questionnaire. Horowitz used computer records and questionnaires together, while Broering used computer records only.

Lancaster’s study used online searching printouts to assess the relevancy of citations by end-user after the search. This approach has a major problem; if the saying “don’t judge a book by its cover” is true, then how can we expect the user to judge citation relevancy without reading the article? Therefore, to determine the precision ratio by the user in the search process has some problems to it.

Among these researchers, only Lancaster, Rapport and Penny used microevaluation to analyze the results and to offer recommendations. Janke had excellent questionnaires, and derived interesting information, but did not interpret it very well.

END USER’S PERFORMANCE AND SATISFACTION
AS THE SOURCES OF SYSTEM EVALUATION

The most standardized effectiveness studies were conducted by Lancaster, Rapport and Penny. In 1971, Lancaster implemented a microevaluation method to determine the effectiveness of AIM-TWX. In that same year, these three researchers used the same method to test EARS (Epilepsy Abstracts Retrieval System) in England. The search was conducted by medical practitioners and physicians respectively. After each search, participants were asked to indicate on their printouts the relevance of items retrieved, using the scale of major relevance, minor relevance, or no relevance. Then, the users were asked to evaluate the search
results, on a similar scale—major value, considerable value, minor value, or no value. In order to get comparable recall ratio, a parallel search was conducted by a specialist with ample experience and knowledge in using the system. Additional items retrieved in these parallel searches were submitted to the original searcher for her/his relevance assessments, using the same scale as before.

Under the AIM-TWX study, Lancaster found that the average precision ratio was 63.1%. The average unit cost in time was 4.5 minutes per relevant citation, and the average recall was 57.6%. Furthermore, 67% of the users “judged their searches of considerable or major value.” Lancaster indicated that, on the whole, the results were surprisingly good, that the majority had conducted productive searches, and that AIM-TWX appeared to meet a definite need. It was a good system and was easy to use.

EARS is a free text searching database of approximately 8,000 abstracts drawn from Epilepsy Abstracts. The physician users of EARS were free to operate the terminal, without the help of information specialists. In this study, 47 searches were completed and “almost 60% were judged to be of major or considerable value, and only 10% were judged of no value.” The average precision ratio was 64.8%. The overall average recall for the group was 57.4%. Lancaster, Rapport and Penny found that EARS can be used reasonably successfully by relatively inexperienced searchers. Improved training and the addition of some types of searching aids could raise the overall effectiveness. The authors indicated that

EARS appears to meet a definite need inasmuch as it provides a capability for conducting searches on topics that would be exceedingly difficult to handle either in the printed index of in a more general machine-based retrieval system such as MELDARS.

In 1973, Barber, Barraclough and Gray used Lancaster's methods to test the effectiveness of the MEDUSA system. This
system was designed to allow medical workers to formulate their own searches without using Medical Subject Headings. The results showed that users obtained 32% precision ratio and 59% recall ratio.\textsuperscript{12} According to this study, the user obtained poorer recall but better precision than the searchers. The authors did not indicate how effective the system was, but from the overall results, they concluded that physicians can formulate a search comparable in accuracy with that of a trained searcher.

In 1983, Janke used questionnaires to survey 25 users’ reaction to BRS/After Dark. He asked questions of relevancy of their search and ease of the system in order to learn about system performance. Thirty-five per cent of respondents indicated that BRS/After Dark either “not too difficult to use” or “easy to use”; 30% found it “extremely easy to use.”\textsuperscript{13}

In the relevancy dimension, 57.1% of respondents obtained over 60% of highly relevant citations. 66.6% of all respondents judged that fewer than 30% of all citations retrieved were not at all relevant to their research needs. If the number of printed citations is an indicator of relevancy, in this study, 54.5% of the participants printed from 26 to 50 references, 22.7% printed 51–59 citations and 4.5% of the respondents printed over 100 references.\textsuperscript{14}

In July and August 1983, Janke ran a follow-up survey of 50 actual end-users in Merisset Library, University of Ottawa about their reaction to the Online After Six. He used “relevant,” “cost,” “how long did it take you to run your search,” and “the using of the Online After Six,” to evaluate the Online After Six. In this study, 40% of respondents printed more than 26 references. Forty-four per cent of respondents replied that they obtained more than 60% of highly relevant citations. Forty-eight per cent indicated that fewer than 30% of the citations retrieved were not at all relevant. Only 36% stated it was “not too difficult to use,” while 44% said it was “easy to use” and 20% indicated it was “extremely easy to use.” Fifty-two per cent of the respondents spent 11–15 minutes to run their search, 20% spent
6–10 minutes and 8% only spent 0–5 minutes. The majority (52%) paid $6.00 to run the search.\textsuperscript{15}

Vollaro and Hawkins used seventeen patent attorneys in A T & T Laboratories to run their own online searches on the DIALOG system. The study found that patent attorneys were satisfied with the results of their online searches. However, the average search time were astonishing: two attorneys spent about two hours per search in their online searching; six had average search times in the 40–60 minute range. Only two attorneys ran searches that averaged 15 or fewer minutes.\textsuperscript{16}

Kirby and Miller examined BRS Colleague/Medical, a menu-driven information system, to test the success of untrained end-users in satisfying their own needs. The results showed that in 37 of 75 searches, the end-users found what they wanted. In 21 of 52 searches (40%) the end-user was as successful as the experienced searcher.\textsuperscript{17} However, after a follow-up search by two trained searchers, the results showed that of the 37 initially judged successful by the end-user, more than half (54%) were missing important articles compared to the follow-up search.\textsuperscript{18}

Harman’s study used questionnaires to learn 75 journalists’ opinions of the usage and value of Newsbank, the full-text database produced by Reuters. He assessed the frequency of use and usefulness to know if Newsbank was judged good or bad. Of the 33 respondents (44%), nobody used Newsbank ten times or more during an average week, only 9% used it between 5–10 times and 78% said they used it five times or less. Astonishingly, 42% replied that they never used it. Apparently, no one found it very useful and only 15% said it was useful. Only 24% of the journalists said they knew enough about it to use it, and 27% indicated that they did not understand Newsbank at all. Regarding the Newsbank searching aids, 42% of respondents indicated that they were not very useful and 33% said they were not at all useful.\textsuperscript{19} The results showed that the Reuters’ journalists seldom used Newsbank, that they found it was not useful and that they did not know the system very well. The user guidance was not
helpful to them. Accordingly, they would rather let the librarian run the search for them.²⁰

Horowitz and Bleich’s article in PaperChase presented the system success from the approaches of the total number of user searches and the average number of citations printed. PaperChase is an online system that permits the end-user to search the medical literature without any previous training. During the first year of deployment, 1,032 users conducted 8,459 searches and displayed 399,821 references. Among these references, 97,769 (24.6%) were printed. The average time of each search was 13 minutes. Among these users, 49% returned to perform five or more searches, and 13.9% returned to perform 20 or more searches.²¹

Not all users of PaperChase found what they were looking for. Horowitz and Bleich indicated that the proportion of unsatisfied users decreased after they had used it 20 times or more.²²

Broering also discussed miniMEDLINE SYSTEM from the perspective of the total number of searches and the average time per search. miniMEDLINE SYSTEM is a user-friendly online information system. The system is self-instructional and easy to learn. Broering indicated that over a two year period, over 28,000 searches had been conducted by end-users, he said that that “represents almost five times the number of MEDLINE searches conducted by the librarians in the period.”²³ The average time for each search was less than ten minutes in this study.

From the above studies, the results showed that if an online system is user-friendly or permits natural language searching, the end-user tends to get higher performance and satisfaction. AIM-TWX, EARS, MEDUSA, BRS/After Dark, Online After Six, BRS Colleague/Medical, PaperChase and miniMEDLINE SYSTEM either permit natural language searching or are user-friendly; as a result, they are capable of fulfilling the needs of end-users to some extent. Originally, DIALOG was not designed for end-users. This is why Vallaro and Hawkins’ study found that end-users had to spend one to two hours in their searching. Richardson’s study mentioned that even within the period of no fee
searching, the overall usage of DIALOG was surprisingly low.\textsuperscript{24} We cannot rule out that complicated searching protocol was the main obstacle. Lancaster indicated that if users have to spend excessive effort in exploiting a system, they may prefer not to use it.\textsuperscript{25} This does not mean that DIALOG is a poor system; the problem is that it is not a system for end-users. The system is for information specialists; only experienced searchers can fully explore DIALOG very well.

Harman’s study in Newshénk also confirmed this tendency: if a system is complicated, the end-users tend to be bewildered, especially, if they are busy like journalists. However, as the majority of journalists indicated that they did not know how to use the system, and librarians did not give them any training, an objective assessment on their part is unlikely. If they had been trained, perhaps the results would have been different. On the other hand, if they were so busy and even had no time to run the search, they probably would not have attended a training program. The judgments of journalists about Newshénk are highly subjective and of limited research value.

ARE END-USERS CAPABLE OF DOING ONLINE SEARCHING?

In Lancaster’s 1968 study of MEDLARS, trained searchers obtained only a 50.4\% precision ratio and a 57.7\% recall ratio.\textsuperscript{26} In AIM-TWX study, end-users achieved an average 63.1\% precision ratio and 57.6\% on the recall ratio. Lancaster considered

it extremely encouraging that a group of biomedical specialists, with a minimum of exposure to the system should be able to conduct searches to satisfy their own needs, using a controlled vocabulary, and achieve a precision in excess of 60 per cent.\textsuperscript{27}

He concluded that end-users should be able to conduct their own online searching.\textsuperscript{28} Vollaro and Hawkins’ study confirmed this
assertion, also concluding that "end-users can do successful searches, even with only a little training."\textsuperscript{29}

Lancaster found that end-users seldom modified their search statements, and that "frequently the searcher will stick to his original search strategy and will not be led to alternative approaches."\textsuperscript{30} In order to test this finding, Olson conducted research on the interactive capabilities of end-users on the MEDLINE system. He found end-users made an average of 5.7 search statement modifications and "it is concluded that non-librarian users can and do use the interactive capabilities of MEDLINE."\textsuperscript{31} Moreover, the study of Barber, Baraclough and Gray even showed that an end-user (physician) "can formulate a search which is comparable in performance with that of a trained search editor."\textsuperscript{32}

Janke drew the most optimistic conclusion. Owing to the success of BRS/After Dark pilot project, he concluded that a new era of online searching is coming, and soon any one will be able to run his or her own online search, whether (s)he come to the library to search on publicly available terminals or microcomputers, or runs searches at home, in the office or at work.\textsuperscript{33}

He said that the time is ripe for the implementation of after hours online service, both in libraries and at home, throughout North America. He even predicted (perhaps somewhat over-optimistically) that by 1986, "the electronic home library will very likely be fully established, in literally millions of homes and offices throughout the U.S. and Canada."\textsuperscript{34}

**PROBLEMS ENCOUNTERED BY END-USERS**

Most of the above research demonstrated the ability of end-users in online searching. However, these studies also indicated several problems that often occur in the process.
For instance, end-users tend to be effective and efficient when the searches are simple and unambiguous. Lancaster found that "users are most successful in relatively simple approaches. When they try more sophisticated techniques, they frequently go astray." Kirby and Miller's study confirmed this view: they found that "it is relatively easy for an untrained user to search a topic which is wholly expressed by a single descriptor or by a unique phrase consistently used in titles or abstracts." Hansen found that if the topic is simple, specific and clear, end-users can design searches without problem.

Moreover, end-users encounter obstacles in grasping Boolean logic. Janke's study indicated that end-users always want to zero in quickly on their topic, will often cut corners, generally stick to their Boolean operators and are unaware of the more sophisticated approaches to online searching.

Sewell and Teitelbaum found AND was the only logical operator used by more than half the end-user, and it was used 5.4 times per search. Trzebiatowski also confirmed this tendency: she found AND was the only Boolean operator used in all searches. Only five searchers used the OR operator, and the NOT operator was not used. She indicated that they had problems in understanding concept combinations.

Furthermore, Janke's study indicated that although 80% of the end-users are willing to execute their online searches, they still need the help of a professional intermediary in order to obtain satisfactory results. Harman's study found that the need for a sympathetic and knowledgeable specialist was widely requested. Janke's study indicated that reasons for the need of professional searchers were related to the lack of understanding of the computer and the problems of hardware or software malfunction. Sewell found that "the major problem is that the end-users do not search often enough to remember details." They need a professional specialist readily available to solve prob-
lems when help is needed.

HOW TO HELP THE END-USER IMPROVE PERFORMANCE

The most detailed analysis of end-user failure in online searching was conducted by Lancaster, Rapport and Penny in their EARS study. They found that end-user precision failures were caused mainly by problems with vocabulary (false coordination, incorrect term relationship) or search strategy (strategy not exhaustive, illogical Boolean construction). The major problems in recall failure were in search strategy (failure to cover all approaches to retrieval, search too specific).  

It seems that the vocabulary and search strategy problems were the most prevalent failures for end-users in online searching. How can we help end-users to improve performance?

Sewell and Teitelbaum's study indicated that there are three ways to help users: system improvement, training and user aids.  

Speaking to system improvement, in their EARS study, Lancaster, Rapport and Penny indicated that an end-user in a natural-language system needs some kind of "searching thesaurus" to help him or her construct comprehensive strategies and think of all approaches. Such a searching thesaurus would consist of all possible synonyms or near-synonyms and would display related terms for helping to decide whether the original terms are appropriate or not. They advocated that this "searching thesaurus" should "be stored in machine readable form and thus be available for examination and manipulation on-line."

Also, they suggested that an ideal online retrieval system should have its own built-in instructional capabilities to guide end-users in the construction of a comprehensive search strategy.

This idea of a "searching thesaurus" with "built-in instructional capabilities" in online searching is known as the "gateway," "interface" or "front-end" in the modern sense. Many of these systems have been developed and evaluated. For instance, CONIT
(developed by Marcus and Reintjes)\textsuperscript{51} and IIDA (developed by Meadow, Hewett and Aversa)\textsuperscript{52} are typical examples. Detailed reviews are included in Eisenberg.\textsuperscript{53}

Concerning training, Trzebiatowski suggested that it should include components that will teach end-users how to:

1. Analyze concepts in a search statement;
2. Gather appropriate terminology including synonymous terms, variant spellings, and correct use of truncation;
3. Create sets and use Boolean operators effectively to express the relationship between sets;
4. Select appropriate databases for the topic and recognize their scope and limitation;
5. Evaluate their search results to identify any errors, and to modify their search strategy.\textsuperscript{54}

Speaking to searching aids, Sewell and Teitelbaum's study found that an easy to use manual would increase effectiveness of end-users in online searching.\textsuperscript{55} Lancaster's study also suggested that "what is still lacking is a brief, clear, well illustrated (with example) booklet describing in simple terms how to use the system."\textsuperscript{56} Harman's study indicated that most users were bewildered by the manual: they recommend a total rewrite for the user guide.\textsuperscript{57}

**FUTURE STUDY**

As a new phenomenon, there is no doubt that end-user online searching will affect the library and information field profoundly. In order to adapt this new tendency, many studies have to be done in the future.

First, if end-users will run their online searching without the help of intermediaries, then what is the future of these intermediaries? Moreover, is this tendency going to affect librarian's education? If the fully established home computer is coming true, is end-user online searching going to decrease the usage of
libraries? If it is, then what will be the future of libraries?

Second, from Harman's study, it seems that the journalist of the Reuters company were not willing to do their online searching. Does this imply some special professions have significant different in their reaction to do-it-yourself online searching? And how many disciplines have these reaction and Why? How can we help them?

And last, most of the studies in this review tend to be small sample size in scale. In the future, a large scale study may have to be conducted in order to determine the real reactions of the end-user in user-friendly online searching.

REFERENCES


3. Ibid.


6. Ibid., p. 7.

7. Ibid., p. 11-2.

9. Ibid., p. 231.
10. Ibid., p. 235.
18. Ibid., p. 261.
20. Ibid., p. 41.
22. Ibid., p. 930.
24. Richardson, p. 46.


28. Ibid., p. 10.

29. Vollaro and Hawkins, p. 70.


32. Barber, Barraclough and Gray, p. 438.


34. Ibid., p. 18.


36. Kirby and Miller, p. 262.


39. Ibid., p. 20.


43. Harman, p. 42.
47. Sewell and Teitelbaum, p. 241.
49. Ibid., p. 243.
50. Ibid., p. 244.
54. Trzebiatowski, p. 449.
BIBLIOGRAPHY


. “Online searching: Measures that Discriminate among


———. "A Computer Intermediary for Interactive Database


Vollaro, Alice J. and Hawkins, Donald T. “End-User Searching in