Factors Affecting Online Bibliographic Retrieval: A Conceptual Framework for Research

Raya Fidel* and Dagobert Soergel
College of Library and Information Services, University of Maryland, College Park, MD 20742

This article presents a conceptual framework for the organization of factors (independent variables) affecting online bibliographic retrieval; the variables were collected from major sources. The first part describes the various roles that variables play in a research study. The second part gives the conceptual framework for the factors with examples of individual variables for illustration. We consider the following elements of the total retrieval situation: the setting , the user, the request, the database, the search system, the searcher, the search process, and the search outcome. For each of these elements (excluding search outcome) a detailed list of variables is given in the Appendix. The variables are organized in a table according to themes that are applicable across elements.

Introduction

Online retrieval studies—through surveys, observations, and experiments—have been carried out for the last ten years; Fenichel [1] gives a comprehensive review of these studies. She concludes that research into online bibliographic retrieval is still in the formative stage and that "...after this initial period of mostly exploratory research, it seems that the most could be gained from controlled experimentation." However, it is not clear whether the factors influencing the search process and outcome are understood well enough to be effectively studied in experiments. Be that as it may, it is useful to identify and organize the large variety of factors suggested by exploratory research so far. This will provide a framework for integrating the results of previous studies and for guiding future investigators in their choice of research problems and variables so that their studies might more easily form a cumulative body of knowldege.

In this article we will discuss briefly the roles variables play in studies of online searching. We will then present a framework for variables that may be important for online searching, giving examples of individual variables for illustration. In the Appendix we present a structured list of variables, based primarily on the following reports of major studies containing substantial lists of variables: the field study done by Carmon in the course of developing a model of the user interface [2]; the survey of users performed by Wanger, Cuadra, and Fishburn [3]; the quarterly report of the Individualized Instruction for Data Access (IIDA) project carried out by the Drexel University, School of Library and Information Science and the Franklin Institute Research Laboratories [4,5]; the reports concerning the IIDA Project by Meadow [6,7]; Fenichel's dissertation [8]; and the National Library of Medicine sponsored study by Wanger, McDonald, and Berger about the online search process [9]. The contributions of many other individual papers are considered using these sources, rather than examining them directly. In order to provide a fluid description, no reference to the specific sources is made in the text itself.

The Role of Variables in Studies of Online Searching

In experimental, observational, or survey studies the researcher tries to understand the relationships between the variables involved. Particularly in experiments, the researcher looks for changes in the *dependent variables* that occur as a consequence of change in the *independent variables*. For example, one may want to test the relationships between speed (the dependent variable) and the familiarity with the database searched (the independent variable). Assuming that other factors are either equal or of no consequence, one can set up an experiment to measure the speed with which a certain request is searched in one

*To whom all correspondence should be addressed. Present address: School of Librarianship, 133 Suzzallo Library, University of Washington, Seattle, WA 98195.

Received February 16, 1982; revised June 8, 1982; accepted August 9, 1982

database by searchers with different degrees of familiarity with the database. The results of such an experiment show the way in which speed of searching varies with the degree of familiarity with the database searched.

The dependent variables in online retrieval studies usually relate to the search process and/or the search outcome; for example, speed of searching and number of databases searched are variables characterizing the search process. Researchers are interested in identifying the effects of selected independent variables on elements of the search process and its outcome; they choose to investigate certain independent variables because they believe that these variables have some effect on the search process and/or its outcome. One may not bother to test the relationship between, say, speed of searching and the level of education of the searcher if it seems unlikely that the speed of searching is affected by the fact that the searcher acquired a higher academic degree in some field.

Search outcome variables are always dependent variables in studies of online searching and are only touched upon in this article. Search process variables may be used as independent variables, control variables (as when searchers are asked to search just one database), or as dependent variables. For example, Cooper [10] measured the dependence of search outcome variables on search process variables (e.g., the number of descriptors used). Oldroyd and Citroen [11] investigated the dependence of search process variables (such as the number of databases searched) on the type of request, without measuring search outcome variables. Fenichel [12] analyzed dependence of both search process variables and search outcome variables on the level of experience of the searcher, but also suggested dependence of search outcome variables on search process variables. Although there is obviously a relationship between the search process and its results, we know very little about the nature of the relationship and further research is needed (for such studies, search process variables are the independent variables).

With these examples in mind, a typology of the roles variables play in a study of online searching is presented in Figure 1.

Much would be gained if researchers would clearly state what the dependent variables in a study are, what variables might affect the dependent variables, and what role each of them play in the study. The list of variables given in the Appendix can serve as a checklist to aid researchers in deciding which variables to consider.

VARIABLES THAT MAY HAVE AN EFFECT ON THE RESULTS
- Independent variables: variables the effect of which is to be studied
  - Variables for which predetermined values are fixed
    - *The values are fixed by manipulation*
    - *The values are fixed by selection from a population*
  - Variables for which any occurring value is measured
- Control variables: variables the effect of which is to be neutralized
  - Variables that are held constant
    - *Values held constant by manipulation*
    - *Values held constant by selection from a population*
  - Variables that are measured and the effect of which is neutralized through statistical analysis
  - Variables whose values are randomized over the sample being studied
- Variables not considered at all

DEPENDENT VARIABLES: VARIABLES DESCRIBING THE RESULTS BEING STUDIED

FIG. 1. Typology of roles variables can play in a study.

The Variables

Variables of significance in online bibliographic retrieval refer to eight elements of the retrieval process and the interaction among them: (1) the setting, (2) the user, (3) the request, (4) the database, (5) the search system, (6) the searcher, (7) the search process, and (8) the search outcome. This list is based on the user interface model by Penniman [19,20] and the classification of variables in Fenichel's dissertation [8].

When we study a specific search we need to know the specific attributes (or values of variables) that characterize each of these eight elements. More importantly, we need to know the attributes that characterize specific combinations of elements, e.g., the match between cost restrictions imposed by the setting and the cost of searching a database, or the familiarity of a given searcher with a given database. Such combinations may express the degree to which values of a variable match with another variable or interactions between elements. To give a more complex example, we may want to know that the organization imposes strict cost restrictions (a Setting variable). When we examine a search performed by a certain member of an organization, the searcher's perception of these restrictions and whether his personal tendencies match such restrictions may affect the search process (two variables relating to the combination Searcher-Setting). If we want to analyze the specific request searched, we need to know whether the cost restrictions were considered by the user when submitting the request (a variable characterizing the combination Request-Setting). Combination variables are listed in this article only as examples, but many more can be generated by one of the following mechanisms:

- A match of two (or more) variables pertaining to different elements. For example, the match between the topic of request and the subject area of the database searched; or the match between the topic of the request and subject background of the searcher. Often the degree of match is more important than the individual values from which it is derived.
- Perception of A by B. For example, the perception of the subject knowledge of the searcher (a Searcher variable) by the user results in a User

variable (which may influence how the user states his query).

- Policies with respect to A. For example, the number of databases searched for a request (a Search Process variable) could be the subject to organizational policy thus giving rise to a Setting variable.
- Typical cases. For example, the database searched first for a specific request is a Search Process variable; the database usually searched first is a Searcher variable.
- Attitudes of A toward B, for example, the attitude of the searcher towards the search system.
- Familiarity of A with B, for example, the familiarity of the user with the database.

There are still other ways in which additional meaningful variables may be generated. For example, searching aids may be used in different search phases generating combinations which may be treated as separate variables. Thus, the list of variables presented in this article is not closed; rather, it could be considered a generator of variables.

We now come to the discussion of the individual elements of our framework.

A. The Setting

The type of parent organization is usually described according to one or more of five facets: the organization's orientation (research, education, etc.); whether it is a for-profit or not-for-profit organization; organization affiliations (governmental, commercial, etc.); the subject area with which the organization is concerned; and the organization's mission (to provide services, products, etc.).

The parent organization may also have general policies for online searching which influence the search process. Of particular importance are the charging policies and sources of funding for searching and the *general guidelines* for online searching procedures. These guidelines (which are not always explicitly stated) may address issues such as: Can users contact the online searcher directly and immediately when an information need is recognized, or do they have to follow certain procedures which may delay the delivery of the request or prevent them from having direct contact with the searcher? Is the user accessible to the searcher for further inquiries? These issues partially determine the distribution of responsibility between searcher and user which is imposed by the setting. In the one extreme, users may be required (or advised) to follow the search process and fully direct the outcome. At the other extreme, when users are not approachable, the searcher has to make all the decisions on his own. When the search is executed, is it possible for the user to be present at the terminal?

In addition to these general guidelines, an organization may make decisions relating to more specific policies that relate to the search process itself. The searching unit may, for example, put limits on the number of descriptors to be used. When organizational policy is examined, most of the factors in the Search Process section give rise to Setting variables.

Apart from the policies of the institution, the actual *status of online searching* within the organization should be described. Some elements of this aspect are: the position of the searching unit within the structure of the organization (part of a parent institution or an independent unit), management's attitude towards online searching, and the length of time online searching has been provided for a given group of users.

For the online searcher, the *group of users* forms an important part of the setting. It can be characterized by the orientation (research, education, etc.) of the group (which may or may not coincide with the orientation of the organization), the turnover in the group of users, and the degree of homogeneity.

B. The User

The user is a person who has an information need and who initiates a search. Whether the user decides to submit his request to a searcher or prefers to conduct an online search himself, a description of this person is important for a better understanding of request. General user studies are abundant in the information science literature (for a review, see, for example, Crawford [13]). It is beyond the scope of this discussion to enumerate all the user characteristics suggested in this literature.

The online literature does not specifically deal with user attributes. Online searching is affected by any user characteristic which is considered when analyzing information retrieval processes in general. For example, user characteristics such as cost behavior and education, whether perceived by the searcher or not, may have an effect on the user-searcher interaction and, in turn, may influence the search process. However, most online experiments control user variables (to be more precise—they eliminate users by providing the searchers-subjects with prepared query statements), and some of the surveys do not distinguish between the end-user and the intermediary by labeling "user" any person who performs the search online. The user's prior experience with computer-based information retrieval and the user's attitude toward computer-based retrieval are suggested as factors that should be taken into consideration.

C. The Request

The term "request" is construed here in its broadest meaning. It includes any specific attribute a searcher may take into consideration before and during the search, and any specific characteristic that may aid in a better understanding of the information need. A request is submitted by a user and is processed into an image by the searcher. The distinction between the user's statement of the request and the searcher's image of the request is very important. The

query is that part of the request that can be expressed by a formulation.

A request is submitted by a person and the description of this person forms an essential part of the request. The "objective" profile of the user is described in the section of the User variables. Perhaps even more important are the searcher's perceptions of the *user characteristics*; his image of the user will affect the search process and outcome. Thus, when perceived by the searcher, every User variable becomes a Request variable.

Three *characteristics of requests* have been identified and tested. Oldroyd and Citroen [11] tested two requests which varied greatly in their degree of specificity and in the amount of relevant documents that were supposed to result. Wanger, McDonald, and Berger [9] submitted three requests, which varied in their difficulty to be searched, to approximately 200 searchers (six sets of these requests each were used in the study). Searchers-judges identified the degree of difficulty of each request by ranking them on a five-point scale *a priori*.

The degree to which the query could be translated into a Boolean expression and the extent to which the concepts used could be translated into the system vocabulary should also be taken into consideration. Finally, the information given by the user when submitting the request is important; does this information include a written statement of the query, the names of authors of particular interest, precision-recall requirements, and preferred databases to search?

D. The Database

Database attributes include: coverage; frequency of updating; availability of thesauri, dictionaries, cross-reference listings, and scope notes for terms; cost; and type of vocabulary.

The *vocabulary* may be controlled, with or without hierarchical structure, or uncontrolled. When the vocabulary is controlled, the index terms may be formulated in natural word order, in inverted word order (important when using word adjacency in free-text searching), or the terms may be formulated to consist of a heading/subheading structure. Two possibilities exist for indexing: either index terms are specifically assigned by human indexers or a computer program, or the terms supplied by the author and/or abstractor in the title, abstract, or full text are used.

Many of these attributes depend on the specific form of the database, manual versus online, and if online, the specific search system. For example, free-text searching in addition to a controlled vocabulary is usually available only online; limiting searches by language, year, document type, etc., and weighted terms are other online features.

E. The Search System

General aspects of the service provided by database vendors should be taken into consideration before discussing the specific attributes of the search system. The first factor of this type is the charging procedure and the cost of using the system. The quality of the service is also considered according to factors such as: schedule of availability, response time, reliability, and the number of databases provided.

Systems may differ in the number and the nature of the *searching aids* they provide in hard copy, microform, or online. These searching aids may include: a users' manual (with various degrees of comprehensiveness), a toll-free number offered by the supplier, and listings of database postings (for presearch assistance).

Various *search support* capabilities may be provided by the search system for use during the terminal session. For example, the system may allow the searcher to do the following: obtain explanation of system features online, display a history of search strategy, and be notified if the system is down.

Searching capabilities influence the structure of search statements and of dialogues. The system may allow the searcher to: enter several terms in the same search statement, search all or part of the fields of the unit record, and search character strings sequentially or serially in any field.

Systems may allow different *output formats* and procedures. Searchers may be able to specify the output format according to features such as: their own print format, sorting of output by designated category (e.g., author, year), and having a search strategy entered online run later in batch mode.

F. The Searcher

Characteristics of searchers are widely described in the literature. They range from personality attributes to detailed analyses of online experience. Of the attributes that are easily defined, only a few have been proven to have any effect on online bibliographic retrieval. Of all searcher characteristics, *cost-consciousness* is regarded by most researchers as a major factor affecting online searching. Cost-consciousness may be induced by the setting when cost restrictions are imposed on searching. In discussing searchers' characteristics, this attribute refers to two issues. First, within the same setting, different searchers may follow the restrictions more or less rigorously, according to their personalities. Second, searchers who work for a certain period of time under rigid cost restrictions may build cost limitations into their searching styles. They may strive to perform cost-effective searches even when they are asked to place emphasis on other output characteristics. Fenichel [8], for example, discovered that subjects who were most cost-conscious (i.e., searchers who were expected to feel pressure to keep costs low) performed the shortest and most cost-effective searches (i.e., least time per relevant reference retrieved). These searchers were most satisfied, although their searches were the simplest (e.g., interacted least, used the smallest variety of commands).

Other searcher characteristics related primarily to personality traits, cognitive factors, demographic variables,

and education (e.g., academic degree, major field of study). Certain *personality traits* are mentioned in the literature only as a list of characteristics an online searcher *should* possess. Thus, an "average" searcher should have attributes such as: a logical (analytical) mind, flexibility and adaptability, and imagination.

Searcher's characteristics which relate directly to online searching are training and experience; these are most frequently mentioned in the literature. *Training* variables include the type of initial training (formal versus informal) and the duration of training; formal education in information science seems to be particularly important. Descriptions of searcher *experience* may refer to experience within the same setting or to total experience. Experience is operationally measured in different ways such as the average number of searches per month or the total number of searches a searcher has performed during his searching career. Each experience measure should be further divided according to specific databases and systems and to the combination of both. Other areas of experience include: experience with and the frequency of reference searching; and experience with hard copy equivalents of the databases. Related to experience is the fact that most searchers specialize (to different degrees) according to search system or database, and a large part of them prefer particular systems or databases. Searching preferred or familiar systems/databases may be different from searching other systems or databases.

The *attitude* of the searcher towards online searching seems to plan an important role in his searching behavior. The general attitude of the searcher may be determined by factors such as his perception of the utility or value of online systems (for the information providing unit and/or the user), interest in and enthusiasm towards online searching, and a sense of professionalism.

The concept of *searching style* is just beginning to emerge. Searching style is understood to be something which characterizes the way a certain searcher performs his searches. As explained later, most of the Search process variables, when looked at as typical procedures, can constitute elements of searching style. In a study performed by Fidel [14], two distinct styles of online searching were identified: the conceptualist searcher, defined as a searcher who bases the planning of the search and the interaction during the search on conceptual analysis; and the operationalist searchers who uses the various features provided by the search system to facilitate interactions during the search. The search process and its outcome are affected by style-related factors and their manifestations in individual searches. These include: the role that the searcher perceives himself to fulfill (e.g., providing the technical know-how, being creative and independent in analyzing the request); the amount of information about the request the searcher perceives himself to need in order to conduct a search; the general preference for free-text versus controlled vocabulary searching; and the willingness of the searcher to abandon a path when it proves to be not useful.

G. The Search Process

The term "search" as used here includes any activity performed by searchers in order to provide users with retrieved information to satisfy their needs. An online search includes at least one session at the terminal, but may include several sessions. This section discusses first the activities outside terminal time, and then the session at the terminal.

Most of the surveys and experiments considered the online search as a single-session event. Therefore, there is a clear line between "preterminal session" activities and "post-terminal session" activities. This distinction also indicates that the order in which these activities takes place is important. Searching is a process of accumulating information and the order in which the information is acquired by the searcher is likely to determine many elements in the search process. Therefore, the order in which the activities outside terminal time are performed should be carefully examined.

As mentioned before, the variables described in this section can be used to characterize procedures *typical* of a searcher giving rise to Searcher variables.

The first factor that should be described is the *interaction with the user*. Interaction may occur at the beginning of the search, during or at the end or it, or in any combination of these modes. When there is an interaction during the search, it is important to mention whether the user is present at the terminal. The main part of the interaction with the user may take place at the reference interview. (It is beyond the scope of this discussion to describe elements of the interview itself.) Generally speaking, the interaction process should be recorded concentrating on factors such as: the nature of the transformation of the request (e.g., from a broad one to a narrow one); topics for which explanations were provided by the searcher (e.g., search procedure, logic, output format); who selects the terms, logic, and databases; and the duration of the reference interview. It may be useful to mention here that Carmon [2] concluded that the user-searcher interface is a nondeterministic and highly adaptive process. It may be difficult, therefore, to develop predetermined guidelines for analyzing the interaction during the reference interview.

The amount and duration of *prelogon preparation* should be examined with regard to two aspects: (1) selection of access points, and (2) logic formulation. It is also important to note which resources were consulted during the preparation. Resources such as own subject expertise, another searcher, and vocabulary tools are mentioned in the literature. The *postlogoff activity* performed before the final results are provided to users should also be recorded; these include evaluation and packaging of results. The searcher may also conduct a postsearch interview with the user.

The search usually begins with an initial plan called *"search strategy."* It is assumed that in planning the search, the searcher decides what system(s) and database(s) to

search and develops the initial search formulation. Once the search strategy is planned, the searcher proceeds to the terminal session. One general factor mentioned in the literature which may have an effect on these decisions is the status of the search at the terminal with regard to other searching modes, namely, the ratio between online and manual searching for a specific search.

The searcher may have different reasons for choosing a certain *system/database* on which to search first. Search system/database attributes (which are described in previous sections) are believed to play a major role in the decision about these features. However, the literature suggests several factors that are not related directly to the characteristics of the search system/database, such as: a searcher may decide to search a certain database because that database was requested by the user, or a searcher may decide to search only one database or several.

Search system selection and database selection are highly interdependent. However, the literature suggests some independent criteria for search system selection. For example, the searcher may use personal preference criteria and use the search system with which he feels most comfortable.

The degree to which the request is formulated before logging on may differ from one search to another. The *query formulation* consists of terms to search and the logic used to combine them. Query formulation has not been surveyed or experimentally investigated. Therefore, no systematic analysis of terms and logic selected is provided by the literature.

Some characteristics of the query formulation are described in the literature. For example, a searcher may formulate the query initially on a trial basis to learn more about suitable descriptors rather than attempting to formulate an optimal and final query right away. The nature of the strategy planning process may be affected by the factors mentioned in previous sections. It seems to us, however, that three important Searcher-Request variables that were ignored by the literature play an important role in the query formulation process. These are: the degree to which the searcher can predict the nature of the resulting set of references, the amount of citations the searcher expectes to retrieve, and the degree to which the expected amount of retrieval determines the nature of the first query formulation.

Of special interest in the area of online retrieval is the *man-machine dialogue*. Qualitative analysis of actual dialogues is scarce in the online literature, however. In most studies of this type, the investigators analyze search protocols. Obviously, this may result in a limited analysis because protocols do not always provide sufficient information about the problem-solving behavior of the searcher during the dialogue. Therefore, investigations of actual dialogues (whether real or experimental) mainly involve counting. Researchers have measured variables such as: the number of databases searched, the total number of commands used in the dialogue, the number of logical combinations, the number of errors made and their types (for example, Sewell [15] identified 48 categories of errors), connect time, and speed (total number of commands divided by connect time).

The search (of one database in one terminal session) is sometimes divided into units, each consisting of a sequence of commands that represents a typical cycle in a search. Meadow [17] defines this unit (or "cycle") by the following sequence of DIALOG commands: BEGIN, EXPAND/SELECT, COMBINE, TYPE, PRINT. Studies in which such a unit was defined have measured factors such as the number of units in the search, or the elements of the bibliographic record displayed in the first unit.

In the NLM-sponsored study [9], the investigators define seven "search formulation styles" relating to steps taken in the search (e.g., the Specific-First Approach, the Progressively Narrowing Approach). The investigators hypothesize that these formulations have an effect on search time and on overall system-use efficiency. It should be pointed out, however, that "style" may be an unfortunate choice of terminology here. Whereas results indicate that "some searchers do not necessarily adopt a single style" (p. IV-39), style is usually construed to characterize a searcher.

Analyzing search protocols, and considering their measurable attributes as indicators of other attributes, one may also answer questions such as: how much time did the searcher spend on thinking and evaluation, did he replan the search strategy completely, did he browse (i.e., display sets of retrieved references and examine them).

Variables relating to the search process as a problem-solving process are described by Bates [16,17]. She identifies 17 idea tactics (i.e., "tactics to help generate new ideas or solutions to problems in information searching") and 29 information search tactics. A qualitative examination of a search process may record the tactics used by the searcher.

Manifestations of the problem-solving process (which is not easily observable) can be recorded. Such manifestations have a direct effect on the search outcome. Our experience in observing online searchers [14] assisted us in identifying attributes which characterize the interaction process such as: the feedback element to which the searcher reacts (e.g., number of postings, displayed controlled vocabulary terms), the point at the terminal session marking the first attempt in the process of constructing the final set, and the degree to which the searcher confines himself to the boundaries set by the explicit request.

Lastly, when the dialogue is terminated, the searcher puts together the final answer set. The format of the citations may be determined by the searcher's preference of output features. The searcher must decide on the ratio between online and offline printing of the results; factors he may take into consideration include: cost-related factors (e.g., cost per offline citation or page, cost of computer time), user-related factors (e.g., urgency of need and typical turnaround time from online supplier), and service-related factors (e.g., service goals, staff time).

H. The Search Outcome

Search outcome variables (which are always the dependent variables) refer usually to the "quality" of the retrieved set. The issue of what constitutes a "good" answer-set is not resolved as of yet, but various measures relating to different attributes of the search outcome have been suggested (e.g., expected search length [18]). It is beyond the scope of this article to discuss these measures in detail. Although search outcome is most commonly measured by precision and recall and/or by unit cost (the cost per relevant citation retrieved), these measures may not be the most suitable and investigators usually use them with some reservations. All measures involve judgment of relevance or utility which present thorny issues. User satisfaction has also been suggested as a measure [21]. At this point we should strive to discover outcome measures that relate directly to requirements arising out of problem situations of specific users. While such measures may not constitute universal indicators of quality, they may support the discovery of those.

Concluding Comment

In this article we systematically collect and arrange variables affecting online retrieval. Two problems in carrying out such a project are apparent. First, it is almost impossible to create an exhaustive list of variables affecting online retrieval since this process involves human elements, and almost any factor that affects human and organizational behavior may as a consequence affect online bibliographic retrieval. Second, with a few exceptions such as cost-consciousness, individual variables taken alone seem to have little influence; exploration of combination variables, such as searcher-request variables and of even larger and more complex patterns, hold more promise for understanding the search process and its outcome.

TABLE 2. Outline of search process variables.

- Overall description of search activities and their sequence
- The searcher's image and sources of information
  - The searcher's image of the request
  - Interaction with the user
  - The sources consulted during the search (perhaps differentiated by search phase)
- Functions in the search process
  - Query formulation
  - Selection of search system(s) and database(s)
  - Interaction with the search system(s) and database(s)
  - Termination of the search and final set
  - Postlogoff activities
- Search modification
  - Degree of intended flexibility of the search strategy
  - Degree of actual search modification
- The number of errors made

TABLE 1. Outline of factors influencing the search process and variables characterizing individual components.

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
Orientation, purpose	The nature of the organization, etc. The nature of the user group	Orientation	Purpose	Intended audience		Orientation
Personality characteristics	Organizational climate	Personality characteristics				Personality characteristics
Education and training	Education of management	Education and training				Education and training
Experience	Status of online searching in the organization	Experience				Experience
Attitudes	The management's attitudes	Attitudes				Attitudes
Guidelines and styles	General guidelines	Information-seeking style				Searching style
Subject and other requirements	Subject of the organization	Subject background	Subject and other requirements Information provided by the user	Coverage		Subject background
Structure				Structure	Services provided Search and display capabilities	
Complexity		Ability to handle complexity	Complexity	Complexity of database structure	Complexity of search system interface	Ability to handle complexity
Cost factors	Cost factors	Cost factors	Difficulty Cost factors	Cost factors	Cost factors	Cost factors

TABLE 3. Factors influencing the search process and variables characterizing individual components.

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
Orientation, purpose	*Nature of the organization* • Orientation (research, education, etc.) • Affiliation (governmental, commercial, etc.) • For profit/not-for-profit *Nature of the user group* • Turnover • Homogeneity • Orientation (research, development, education, etc.)	Orientation	Purpose	Intended audience		Orientation
Personality Characteristics	*Organizational climate* Actual intra-searcher communication	*Personality characteristics*				*Personality characteristics*
Demographic variables Personality traits • Intelligence • Cognitive style, etc.		• Demographic variables • Personality traits •• Intelligence •• Cognitive style, (etc., as a searcher				• Demographic variables • Personality traits •• Intelligence •• Cognitive style •• A logical (analytical) mind •• Major field of study (influence) on analytical ability) •• Imagination •• Flexibility •• Memory •• Communication skills •• Adaptability •• Persistence •• Curiosity •• Desire to learn •• Patience •• Self-confidence
Education and training	*Education of management*	*Education and training* (similar to searcher)				*Education and training* • Highest degree • Amount of education in information science and cognate fields • Training history • Each event in the training history is characterized by the following •• Type of education ••• Degree education ••• Continuing education •• Topics covered (e.g., general concepts, search systems, data bases) •• Initial vs. advanced •• Who does the training

TABLE 3. (*Continued from previous page*)

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
						•• Duration • Summary statistics of training (e.g., the number of data bases on which the searcher was initially trained)
Experience	*Status of online searching in the organization* • Experience of the organization with online searching • Location of the searching unit in the organization (relation to the library • Number of searchers in the unit • Job titles of the searchers • Systems to which the organization has access • Physical space in which searching is performed	*Experience* • Each variable should be further subdivided into: online; other • Experience with doing *own* • Experience with searches done for her/him • Average number of searches done for this user				*Experience* • Each variable should be further divided into: online; other • Experience with reference work • Range of subject usually searches • Total experience • Experience within the same setting •• In searching •• In other jobs • Total number of requested answered • Average frequency of reference searching • Current activity level (the amount of searching in a *recent* period of time) • Number of files in which the searcher specializes • Experience in supervising reference librarians as searchers • Percentage of work time spent on reference related activities • Experience with computers
Attitudes	*The management's attitudes toward online searching (subdivided as searcher)*	*Attitudes towards online searching*				*Attitudes towards online searching* • Perception of the utility or value (for the user, the organization) of online systems • Confidence in the completeness of online search • Perception of the quality of online searching in comparison to manual or batch searches • Interest in online searching

TABLE 3. (*Continued from previous page*)

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
						• Enthusiasm towards online searching • Perception of the permanence of automation • Sense of professionalism • Interest in new features of systems and data bases • Reading newsletters, manuals to acquire information about systems and databases
• Interest in new developments	Support for searcher to keep up with databases and search systems	• Interest in new services				
Guidelines and styles	*General guidelines regarding the procedures of online searching* • Procedures for submission of requests • User searcher communication procedure •• Can the user contact the searcher directly •• Can the user contact the searcher immediately •• Can the user be present at the terminal •• User's availability for searcher inquiries • Distribution of responsibility between user and searcher • General policy with respect to online versus manual searching • Scheduling procedures • Policy for intrasearcher communication • Guidelines for postsearch evaluation • Quality of documentation maintained about databases	*User's information seeking style*				*Searching style* • Overall orientation (conceptualist, operationalist) • Own perceived role (provide the technical know-how, be creative and independent in understanding the request, how much would the searcher like the user to be involved) • Willingness to independently resolve ambiguities • Perception of own general way of searching (e.g., from general to specific) • Adherence to general procedures in specific searches (the degree to which the searches are alike) • The amount of information about the request needed for a search, as perceived by the searcher • Consideration of new features of systems and data bases in searching • The amount of information about a database needed for a search, as perceived by the searcher • Nature of the answer set submitted to the user as perceived by the searcher (e.g.,

TABLE 3. (*Continued from previous page*)

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
						the specific answer, a broad set, a final product, an interim product, entry to the literature) • Criteria used for evaluating the result (e.g., precision, recall, specificity) • Criteria for deciding about the preferred databases and systems • General preference for free text vs. controlled vocabulary searching • General rules used to secure quality (e.g., "when more than two concepts are combined, the set for each concept should be enlarged") • Willingness to abandon a path when it proves not to be useful • Typical search-related behavior (see search process characteristics)
Subject and other requirements • Subject • Time aspects •• Retrospective coverage •• Currency	• Subject area • Document access capability	• Subject background	*Subject and other requirements* • Subject area • Formal specifications (e.g., journals to be included, language preference, type of documents, level) • Time aspects of information need •• Number of years required •• Currency required • Requirements for the final set •• Degree of specificity •• Precision-recall requirements •• Number of documents needed •• Number of documents expected by the user • Degree of urgency *Information provided by the user* • Form of submittal (is there a	*Coverage* • Subject coverage • Subject characteristics of indexing • (Data elements included) • Type of documents included • Level of documents included • Time aspects of coverage •• Number of years covered •• Currency (from publication to inclusion) ••• Frequency of updating • Quality of data		• Subject background

TABLE 3. (*Continued from previous page*)

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
		written query? is it in natural language? is it a Boolean expression) • Is the purpose of the request stated? • Is the subject area of the query mentioned? • Terms suggested by the user •• Free terms and synonyms •• Index terms • Authors of particular interest • Known relevant documents • The preferred databases to search		*Services provided by the vendor* • Quality •• Response time •• Reliability (i.e., down time) • Receive notice at log-on time of databases currently unavailable • Be notified if system is down • Schedule of availability • Log-on procedures • Control the length or form of system messages • Have a search strategy entered online run later in batch mode (and then, get results either online or by mail) •• SDI service • Enter comments to the vendor online		
			Structure of the database • Data elements included • Index language and thesaurus •• Type of index language •• Degree of precombination •• Hierarchy •• Form of terms • Indexing rules and parameters •• Type of indexing (manual, machine, use of titles/abstracts/text)	*Searching and display capabilities* • Vocabulary search and display play capabilities •• Online access to thesauri and indexes •• Display a list of terms occurring in relevant documents • Searching capabilities •• Truncate terms •• Specify data fields in which the term is to be searched		

TABLE 3. *(Continued from previous page)*

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
				•• Exhaustivity of indexing •• Specificity of indexing	•• Be able to search all the fields of the unit record or part of them •• Search character strings sequentially in any field •• Use a number for a vocabulary term in lieu of entering the term itself •• Enter natural language query •• Use Boolean operators •• Inclusive searching capability (exploding) •• Limiting capabilities •• Use relational operators •• Use word proximity operators •• Enter several terms in the same search statement •• Enter nested expressions •• Enter several instructions to the system at one time •• Display history of search strategy •• Store searches that can be run again at a later time •• Incorporate previous searches, by number, in new searches • Citation display capabilities •• Request standard print format •• Request predefined print format •• Specify own print format •• Specify sorting of output by designated category (e.g., author, year) •• Specify offline printing of search results •• Receive citation display in upper and lower case	

TABLE 3. (*Continued from previous page*)

0 General Heading	1 Setting	2 User	3 Request	4 Database	5 Search system	6 Searcher
Complexity		*Ability to handle complexity*	*Complexity* *Difficulty* • Degree to which the query can be translated into a Boolean expression • Match between the search requirements and the structure of the database	*Complexity of the database structure*	*Complexity of the search system interface*	*Ability to handle complexity*
Cost factors • Costs • Possibly detail cost element • Cost behavior	*Cost factors* • Source of funding for searching (organization, user, grant money, etc.)	*Cost factors* • Cost behavior •• Innate cost consciousness	*Cost factors* • Resource requirements • Actual cost • Importance of cost, resource availability	*Cost factors* • Cost	*Cost factors* • Cost • Charging procedures	*Cost factors* • Cost behavior •• Innate cost consciousness

TABLE 4. Factors influencing the search process. Variables characterizing pairs of components.

	Setting	User	Request	Database	Search system	Searcher
Setting		Preferences with regard to search procedures (including submission) Role in deciding the searching mode Attitudes towards charging policies	The degree to which the request is typical for the setting	Match between group and intended audience of database Match of database coverage with document access		Identification with organizational goals Perception of the status of online searching in the organization Attitudes towards procedures imposed by the setting Role in deciding the searching mode
User			Relation to the user's main area of interest	Match between level of documents covered in the database and level of user		Match with user's cognitive style Perception of the experience in information use Previous experience with this user Preferred ways to communicate with this user
Request	Setting policy with regard to the types of requests to be searched			Match between the subject of the request and the database vocabulary Match between types of documents covered in the database and types of documents requested	Match between capabilities and request requirements	Subject knowledge of the searcher in the subject matter of the request Personality traits needed for request The familiarity of the searcher with the literature in the subject area of the request Did the searcher search the same topic earlier

TABLE 4. (*Continued from previous page*)

	Setting	User	Request	Database	Search system	Searcher
Database	Match between intended audience of this database and user group Match of document access with the coverage of this database	Familiarity with the database. Understanding the database structure Match between level of documents covered in this database and level of use	Match between types of documents covered in the database and types of documents requested		Databases provided Search features available for this database (e.g., data elements searchable, "EXPLODE")	Training history with relation to database Familiarity with the database (online, printed) Understanding the database structure Match of database structure with searching style
Search system		Familiarity with search system	Match between requirements and search system	Search features available Number of years covered Frequency updating Currency (from database production to inclusion Cost		Training history with relation to this system Familiarity with this system Attitude to this system Complexity match with this system
Searcher	Management attitudes towards the searcher	Match with searcher's cognitive style Perception of searcher's ability to understand the query Previous experience with this searcher	Match between the searching style of the searcher and the type of request	Match of database structure with searching style	Complexity match with this searcher	

TABLE 5. Search process variables.

Overall description of search activities and their sequence

Levels of search activity
- Levels of search activity can be defined from the most comprehensive to elemental components as follows:
- The entire search
- A prelogon-terminal session cycle
- Prelogon preparation
- Terminal session
- Database session (that part of a terminal session dealing with one data base)
- Units within a database session
- Elemental component (e.g., one command)

For each search activity the following factors can be considered
- Number of occurrences in a higher level search activity
  - Number of types (e.g., number of different commands used)
  - Number of tokens per type or class of types (e.g., number of PRINT commands used)
  - Total number of tokens (e.g., total number of commands issued)
- For each occurrence of an activity (token)
  - Amount/intensity/complexity
  - Duration
  - Speed (e.g., number of commands per unit time)

Some sample combinations are included among the variables listed below.

The searcher's image of the request

The amount of information the searcher has about the request

The relative role in the formation of the image of
- Information supplied by the user, and
- The searcher's own ideas confirmed by the user

The degree to which the searcher understands the request

The searcher's perception of how well he understands the request

The searcher's recognition of the separate ideas which are implied by the query

The searcher's perception of search difficulty

The degree to which the searcher can predict what he is going to find

The expected amount of retrieval (for the whole search or parts of it)

Service-related factors (service goals, staff time) as components of the searcher's image

Interaction with the user

(See also Information provided by the user under Request)

Interaction with the user by step in the search process
- Interaction with the user at the beginning of the search - reference interview
- Interaction with the user at the terminal
  - Is the user present at the terminal?
- Interaction with the user between terminal sessions
- Interaction with the user after the search - postsearch interview

The total interaction, or each step, can be characterized by
- The length of the interaction
- Knowledge the searcher draws on for providing information to the user
- Aids used by user and/or searcher during the interaction (e.g., a thesaurus) (see Sources consulted during the search)
- The degree to which the user is prepared for the interaction
- The degree to which the searcher is prepared for the interaction

- Substance of the interaction
  - Topics for which explanations were provided by the user
  - Topics for which explanations were provided by the searcher (e.g., search procedure, logic, output format)
  - Weakness of the user's original query perceived by the searcher
  - Elaboration of the query by the user
    - The nature of the transformation of the query (e.g., from broad to narrow)
  - The degree of the user's and the searcher's participation in the search process
  - The nature of the user's and the searcher's participation in the search process
    - Examples are the user's and the searcher's contribution to:
      - identifying candidate terms
      - developing candidate logic formuations
      - identifying candidate databases
      - initiation of strategy modification
      - selection of terms
      - selection of logic
      - selection of databases
      - decision to revise strategy
      - decision to terminate the search

The sources consulted during the search (perhaps differentiated by search phase)

Values for the following variables could be simply used/not used or they could measure the use intensity and/or the importance of each source on a scale. In a detailed description the information obtained from each source could be given; also, the point in the search at which the source was used.

- Vocabulary tools
  - Printed vocabulary (hard copy or microform)
  - Online vocabulary file
- Printed version of the database, including its index
- General reference tools
- Textbooks
- Another searcher
- Professional with subject expertise
- Own subject expertise
- Own experience in doing searchers
- User manuals
  - Printed user manual
  - Online user manual

Query formulation

- *NT* the number of elements in the query formulation
- Conceptual query formulation
  - The number of concepts identified
  - Conceptual logic formulation
- Database specific query formulation
  - Selection of terms
    - Narrow or broad representation of each concept by a set of terms used in the file (selection of only the most relevant terms versus selection of a large number of terms for several concepts)
  - Database specific logic formulation
- The degree to which the expected amount of retrieval (see Image) determines the nature of the query formulation (at the beginning or throughout the search)
- The finality of the prelogon query formulation (perhaps differentiate for conceptual query formulation and database specific formulation)
  - The degree to which the query is formulated before logging on
  - The degree to which the first query formulation corresponds to the query statement (as opposed to being exploratory or representing a subsearch)
  - The extent to which the terms in the initial query formulation are thought of as final (as opposed to merely exploratory)

Selection of search system(s) and database(s)

- The status of the online search with regard to other searching modes (manual, batch)
- What is more important (selected first), search system or database?
- Selection of search system
  - System characteristics considered. Importance of each (RT level of use of system features):
    - Search capabilities
    - Access to databases
    - Personal preference
  - Selection of database
    - Database characteristics considered. Importance of each
    - Number of databases searcher decides to use
    - Selection of database to search first

Interaction with the search system(s) and database(s)

- The actual number of search systems used
- The actual number of databases searched
- The following variables should be measured for each database unless databases are searched interdependently
  - Connect time
  - Speed (e.g., number of commands per unit time)
  - The time spent on thinking
    - The time spent on evaluating system response
    - The time spent on planning the next step
    - These thinking time variables are hard to measure but may relate to the following observable times
    - The time spent on input
    - The waiting time for system response (including printing out records)
  - The number of elements in the query formulation
    - The number of search statements
    - The number of terms
      - The number of controlled vocabulary terms
      - The number of free text terms
      - The total number of terms
  - The number of logic combinations
- The level of use of selected system features
  - The level of use of different commands
    - The number of command types
    - The number of command tokens per command type or class of types
    - The total number of command tokens
  - The use of "advanced" techniques (subdivided as commands)
- The use of different moves/tactics
  - The number of operational moves
    - The number of types of operational moves
    - The number of tokens per type
    - The total number of operational moves (tokens)
  - The number of conceptual moves
    - The number of types of conceptual moves
    - The number of tokens per type
    - The total number of conceptual moves (tokens)
- Use of feedback
  - The feedback element to which the searcher reacts (e.g., number of postings, displayed citations)
  - Amount of browsing
    - The number of sets from which citations are displayed before the final answer is decided upon (or the number of print commands)
    - The total number of citations displayed before the final answer is decided upon
  - The elements of the bibliographic record displayed for feedback

TABLE 5. Continued.

- Evaluation of retrieval results and retrieved citations (perhaps differentiated by search phase)
  - Procedures for evaluating the retrieval results
  - Retrieval attributes taken into consideration when evaluating retrieval results (e.g., size, recall, precision)
  - Criteria for evaluating individual citations (e.g., subject, availability)
- See also narrower term degree of actual search modification

Units in the search of one database (see text for definition). The variables listed above can be measured for each unit. All the variables given under search activity above can be given for each unit.

Termination of the search and final set

The point in the terminal session and which the first attempt to construct the final set or a component subset is made

The nature of the final set (one final set versus several subsets)

The criteria used by the searcher to decide upon terminating the search

The element of the bibliographic record displayed when printing the final set

Decision about the ratio between online and offline printing of results
- Cost-related factors
- User-related factors

Postlogoff activities

Evaluation of the results

(Postlogoff interaction with the user)

Degree of intended flexibility of the search strategy (applies to all functions in the search)

The degree to which database selection is left open
- Sample values:
- Decision to search all preselected databases
- Decision to consider preselected databases not yet searched after results obtained so far are examined
- Decision to consider additional candidate databases after results are examined

The degree to which the query formulation is planned to be adaptive (see also Finality of the query formulation)
- Are interactions incorporated into the initial formulation, i.e., does the searcher formulate a search to display first a trial set (may be same variable as above)
- Did the searcher plan alternative query formulations before logging on

Degree of actual search modification

The number of actual search modifications

Modification of search system or database selection

Modification of the query
- Degree of actual reformulation of the query
- The nature of the actual reformulation of the query (broadened, narrowed, other changes)

Degree of adherence to a fixed search plan

Errors made (consider type and number)

Appendix: List of Variables Important in Online Searching

This Appendix lists the variables we have collected in a systematic framework. Our emphasis has been on the structure of this framework rather than on the exhaustive collection of variables. After giving outlines in Tables 1 and 2, we list first the factors influencing the search process: variables characterizing the components setting, user, request, database, search system, and searcher. Table 3 gives the variables characterizing the individual components in a form that shows themes in common to two or more components. Table 4 gives examples of variables that pertain to combinations of components, such as request-database or database-searcher. The examples have been selected to illustrate various ways in which variables characterizing such combinations can be generated; we did not include obvious variables such as the match between the subject area of the request and the subject background of the searcher. Table 5 gives the search process variables.

References

1. Fenichel, C. H. "The process of searching online bibliographic databases: A review of research." *Library Research*, 2(2):107-127; 1980.
2. Carmon, J. L. *Model of the User Interface for a Multidisciplinary Bibliographic Information Network*. Athens, GA: University of Georgia, Office of Computing Activities; 1975.
3. Wanger, J.; Cuadra, C. A.; Fishburn, M. *Impact of On-Line Retrieval Services: A Survey of Users, 1974-75*. Santa Monica, CA: System Development Corporation; 1976, 237 pp.
4. Drexel University, School of Library and Information Science; Franklin Institute Research Laboratories. *Individualized Instruction for Data Access (IIDA). Quarterly Report No. 2.* Philadelphia: Drexel University; 1978, 49 pp.
5. Drexel University, School of Library and Information Science; Franklin Institute Research Laboratories. *Individualized Instruction for Data Access (IIDA). Quarterly Report No. 4.* Philadelphia: Office of Research Administration, Drexel University; 1979, 20 pp.
6. Meadow, C. T.; Hewett, T. T.; Rafsmider, D. J.; Toliver, D. E.; Epstein, B.; Edelmann, J. V.; Maher, A. *Individualized Instruction for Data Access (IIDA); Final Design Report*. Philadelphia: Graduate School of Library Science, Drexel University; 1977.
7. Meadow, C. T. "The computer as a search intermediary." *Online*, 3(3):54-59; 1979.
8. Fenichel, C. H. "Online information retrieval: Identification of measures that discriminate among users with different levels and types of experience." Ph.D. dissertation, Drexel University, 1979, 257 pp.
9. Wanger, J.; McDonald, D.; Berger, M. C. *Evaluation of the Online Search Process: A Final Report*. Santa Monica, CA: Cuadra Associates; 1980, 1V. (PB81-132565).
10. Cooper, M. D. "Input-Output Relationships in On-Line Bibliographic Searching." *Journal of the American Society for Information Science*. 28(3):153-60; 1977.
11. Oldroyd, B. K.; Citroen, C. L. "Study of strategies used in online searching." *On-Line Review*, 1(4):295-310; 1977.
12. Fenichel, C. H. "Online Searching: Measures That Discriminate Among Users with Different Types of Experience." *Journal of the American Society for Information Science*, 32(1):23-32; 1981.
13. Crawford, S. "Information needs and uses." In: *Annual Review of Information Science and Technology*, M. E. Williams, Ed. White Plains, NY: Knowledge Industry Publications; 1978, Vol. 13, pp. 61-81.
14. Fidel, R. "Constructing a pattern model of man-machine interaction in online literature searching." Ph.D. dissertation, The University of Maryland, 1982. A preliminary account is given in Fidel, R. "Operationalist and conceptualist searchers: A case-

study-based pattern model of online searching styles." In: *Proceedings of the American Society for Information Science, 44th Annual Meeting, Washington, DC*, L. F. Lunin, M. Henderson, and H. Wooster, Eds. White Plains, NY: Knowledge Industry Publications; 1981, Vol. 18, pp. 68-70.
15. Sewell, W. Private communication, 1981.
16. Bates, M. J. "Information Search Tactics." *Journal of the American Society for Information Science*, 30(4):205-214; 1979.
17. Bates, M. J. "Idea Tactics." *Journal of the American Society for American Science*, 30(5):280-289; 1979.
18. Cooper, W. S. "Expected search length: A single measure of retrieval effectiveness based on the weak ordering of retrieval systems." *American Documentation*, 19(1):30-41; 1968.
19. Penniman, D. W. "Networked data bases—A view from the middle." In: *Proceedings of the 13th Meeting of the Geoscience Information Society, Geoscience Information: Publication-Processing-Management, 1978*, J. Bichteler, Ed. Boulder, CO: Geoscience Information Society; 1979, Vol. 9, pp. 39-54.
20. Penniman, D. W. "Domesticating the computer terminal." In: *Proceedings, Distributed Computing—COMPCON Fall '80 IEEE Computer Society 21st International Conference, Washington, DC, 1980.* NY: Institute of Electrical and Electronic Engineers, Inc.; 1980, 141-146.
21. Cooper, W. S. "On Selecting a Measure of Retrieval Effectiveness. Part II. Implementation of the Philosophy." *Journal of the American Society for Information Science, 24(6):413-424; 1973.*