

---

# Formulation Rules for Posing Good Subject Questions: Empowerment for the End-User

DONNA RUBENS

---

## ABSTRACT

LACKING AN UNDERSTANDING of bibliographic organization and information retrieval systems, end-users have difficulty expressing an information need. By starting with the end-user's intuitive understanding of knowledge creation and the institutional structure, end-users can be taught a technique for analyzing their questions and translating them into information system terms. Based on a model developed to describe how professional searchers think, the question formulation technique employs five operations that transform a question into a description of the characteristics of potential answer-providing sources.

## INTRODUCTION

This article provides rules for formulating subject questions that can be taught to end-users. The method is based on a model developed over the six-year period from 1982 to 1988. The model, called "thinking like a searcher," was developed originally to explain how the professional searcher uses mental associations to develop a search strategy (Rubens, 1989). In this article, the results are applied to end-user searching (manual and online).

The question formulation process proposed here challenges the traditional relationship between the end-user and the professional intermediary by advocating a more active role for the end-user in the reference transaction. Rather than relying on the professional intermediary to draw out from the end-user what is really wanted

Donna Rubens, Bibliographic Instruction, Institute of Technology Libraries, University of Minnesota, 108 Walter Library Building, 117 Pleasant Street S.E., Minneapolis, MN 55455

LIBRARY TRENDS, Vol. 39, No. 3, Winter 1991, pp. 271-98

© 1991 The Board of Trustees, University of Illinois

and to translate the question into information system terms, the end-user learns how to analyze his or her need for information and to specify the characteristics of potential answer-providing sources. The process is one that can be taught because it is based on what the end-user already knows about information sources.

The rationale for this project grows from the observation that many information seekers have difficulty asking subject questions. As Ingwersen (1984) noted, present IR (information retrieval) systems, whether printed or computerized, are based on the "best match" principle:

"Best match" implies the assumptions that users are able to specify the information required—that the information need expressions are functionally equivalent to document texts; i.e., equivalent to information. Optimistically, it implies that user queries exactly mirror the underlying problem situation, that users may describe very well what they might not know about and that the applied search terms are always valid.... (p. 86)

As Ingwersen implied, the difficulties that end-users face stem from their lack of understanding of the structure of the information environment (terms used in a particular way in this article are defined in the appendix). They have no model of information flow, no knowledge of the bibliographic chain, no knowledge of subject relationships or of knowledge creation. This study suggests that end-users who hold an accurate and detailed view of the information environment are able to pose a question that takes into account the variables in that environment.

As opposed to the usual halting attempts to get help at the reference desk, asking a question and getting an answer require consumer behavior. When seeking a product, the consumer asks for certain features using appropriate language. When buying a computer, the consumer might specify a 32-bit memory board upgradable to 13 MB; a 20 MHz, 383 processor; two serial ports; and the like. Similarly, when seeking an answer to a question, the end-user might get a better answer if able to specify the features desired in terms of format, currency, intellectual level, and subject.

Information specialists have an extensive vocabulary for describing information sources. They use this knowledge to negotiate questions and to develop search strategies. For example, during the reference process, it is useful to distinguish research journals from trade magazines and consumer magazines, or to distinguish encyclopedias from handbooks and monographs, or children's books from adult fiction, and fiction from nonfiction. Professionals recognize thousands of subject fields and topics using technical and nontechnical terms derived from all sectors of society including academia and the popular press. In addition, they use an extensive

vocabulary for describing the characteristics or attributes of each type of source. Periodicals tend to be current or contemporaneous media. Children's books tend to have illustrations, larger print, and easy vocabulary. Handbooks tend to have tables and numbers describing properties or formulas.

However, this knowledge about classes and attributes of information sources is not systematically applied to the reference process or to bibliographic instruction. There is such a thing as "thinking like a searcher" (Huston, 1988), but there are no accepted models that are used for teaching reference or library use. This article describes one such model and applies the findings to teaching students how to formulate subject questions.

## METHODOLOGY

The model here is based on ideas from the library and information science literature; the author's own experience working at the reference desk at a special, public, and academic library; and on an analysis of published case studies of reference.

### *Literature Review*

The literature review for this study covered the years 1950 to early 1990 and included the topics of search strategy, the reference process, education for librarianship, bibliographic instruction, and expert systems for reference. The literature was scanned for studies of cognitive behavior. What does it mean to think like a searcher? What are the important variables in creating a search strategy? How does the question formulation influence the search strategy? How does the question negotiation influence the search strategy?

Although there was scant attention given to the cognitive aspects of these subjects, some ideas and studies proved useful. Benson and Mahoney's (1975) outline of query parameters made clear that a question has many dimensions beyond the topic itself (p. 318). The requestor typically makes only some of these parameters explicit when first posing the question. As Taylor (1968) discussed in his landmark paper, each of the other relevant parameters are identified through a question negotiation process. Benson and Mahoney also contributed the useful image of bridge building to describe the process of closing the gap between the query and the information system (p. 317). This image motivated this author's search for that bridge and resulted in the idea that "attributes" of information sources are the cognitive link between query and answers.

Neill (1975) argued that questions should be analyzed at a high level of abstraction (p. 313). Although a search strategy is ultimately expressed in character strings of specific terms, the initial stages of

query analysis require conceptualizations of the topic that explore the meaning behind the question as posed. This process allows for more flexibility in positioning the question in one or more subject domains and facilitates the creative approach to searching that Bates (1979a, 1979b) explored in her two-part series on search tactics.

It is generally agreed that knowledge of the system is necessary before the system can be interrogated and that end-users typically lack such knowledge. Ingwersen (1984) stressed that "the searcher must possess sufficient 'IR (information retrieval) knowledge'" (p. 471). He divided searchers into four categories depending on subject knowledge and system knowledge. The "end-user" and the "layman" are defined as groups having scarce or no system knowledge and therefore are dependent on a professional intermediary (p. 473). Current theories of bibliographic instruction suggest that users can learn the structure of information systems if they have a conceptual understanding of the search environment. Borgman (1982) showed that use of mental models improves end-user searching.

The literature on knowledge creation and information flow contributed terms for categorizing information sources and mapping the information environment. For example, the Doyle-Grimes Model of the Bibliographic Chain defines the relationships between primary, secondary, and tertiary sources as a function of knowledge creation and information flow (Doyle & Grimes, 1976, p. 3). Kerezstesi (1982) showed the importance of understanding information flow for bibliographic instruction and presented his analysis of the evolution of research information in the domain of academia (pp. 15, 17, 19).

Swift, Winn, and Bramer (1979) went beyond the narrow domain of academia to explore the origins of information from a sociological viewpoint. The authors articulated the relationship between the institutional structure of society and the structure of the information environment. They pointed out that "knowledge is created as human beings interact....Interactions of people give rise to the institutional order." The authors concluded that it is useful in designing information systems to understand "how society works" (p. 218).

For example, governmental bodies produce official or public documents, and this is reflected in the institution we call depository libraries. Similarly, community groups and associations produce practical, brief reports on various subjects in the form of pamphlets, and this is reflected in vertical file collections. Or, to express it another way, depository libraries and vertical file collections are acknowledged divisions of the information environment as we know it. Professional understanding of these collections stems from an understanding of knowledge creation, which implies a knowledge of social structure. What kind of information do we associate with Congress, or the

courts, or travel writers, or publishers, or industrial labs? How are these documents reflected in bibliographic organization? How can knowledge of information flow and knowledge creation be used in the reference process?

Studies of end-user searching tend to conclude that users have trouble selecting the appropriate subject domain during a search. For example, Allen (1990) found that only one patron in five used the most appropriate database available and one in five used the least appropriate database (p. 69). This suggests that users have trouble classifying questions in subject terms and matching these terms to the selected information system.

In a pilot study on course-integrated instruction funded by the National Endowment for the Humanities and the Council on Library Resources, Tiero and Lee (1983) concluded that there is a need to teach generic sources rather than titles (p. 291). Their observations reinforce the idea that information seekers can be more flexible if they think in categorical terms.

The advent of end-user searching has forced some rethinking of the role of the intermediary and with it some re-examination of how far the end-user can go before the intermediary must intervene. Most models of the search process assume that interaction between the end-user and the intermediary is required to develop the interface language that bridges the gap between query and answer-providing sources. However, those researchers and educators whose instruction is motivated by the social ideology that information is power emphasize that users know more about information seeking and the information environment than they are given credit for. The bridge to an understanding of bibliographic organization is the user's intuitive knowledge of community information and social structure:

Teaching about information access, then, can best be achieved by first emphasizing the familiar (students' experiential and topical information) and then linking that to the new (librarians' bibliographic knowledge). To maximize communication requires our rethinking our approach to both students and faculty. (Huston, 1983, p. 186)

### *Analysis of Case Studies of Reference*

For analysis of the reference process, fifteen published case studies of reference by British librarian Denis Grogan (1967, 1972) were examined. Grogan's two-volume work is unique in that each case contains descriptions of his reasoning and decision-making steps (see Figure 1). Each case starts with a description of the question as Grogan understood it. The running commentary of what he thought as he proceeded allows us to see his mind at work. It might have been valuable to have a transcript of the reference transaction, but from the way Grogan analyzed the question it can be concluded that a

good strategy depends on an analysis that translates the question into system terms. Underlying this analysis is a mental map of the information environment that supports the process of thinking like a searcher.

---

**"What turned out to be a very long hunt for the date of the first use of the SOS distress signal was initiated by a national newspaper wishing to confirm the impression given in a film shown on television that it was by the sinking Titanic. Several approaches immediately suggested themselves to the librarian: the DATE, 14th/15th April 1912, was easily ascertained from the nearest ENCYCLOPEDIA, thus opening up access to the NEWSPAPER and PERIODICAL press of the day. Since a HUGE DISASTER was almost certain to have led to an OFFICIAL ENQUIRY, a 'FORM' APPROACH through the PARLIAMENTARY PAPERS would be another possibility; it was known that the film was based on Walter Lord's BOOK *A Night To Remember* (1956) and an approach that way seemed hopeful; even the SUBJECT APPROACH revealed several promising aspects--RADIO, SHIPPING, DISASTERS, etc...."**

Figure 1. Example of Grogan's reasoning and use of interface language (capitalized). From Grogan, D. (1972). *More case studies in reference work*. Hamden, CT: Linnet Books, pp. 251-55

---

### THINKING-LIKE-A-SEARCHER MODEL

The following is a description of the thinking-like-a-searcher model that outlines the searcher's knowledge base and how that knowledge is put to use during the course of a search. The section following the presentation of the model is the application of the model to question formulation and a discussion of requirements for teaching question formulation as part of a bibliographic instruction program.

#### *A Good Answer Starts With a Well Thought Out Question*

The literature of question negotiation, notably Taylor's (1968) paper, makes clear that a question as originally posed by the patron typically contains insufficient information to proceed with a search. The searcher must discuss the question as asked until certain critical information is made explicit. Benson and Maloney (1975) identified nine query parameters (p. 318). Figure 2 is based on their list. Each of these parameters helps channel the question to the appropriate source in or out of the library.

## TOPIC PARAMETERS

1. Subjects and subject relationships
2. Time frame (current/historical)
3. Geographic scope

## PROJECT PARAMETERS

4. Purpose (expected use for the information)
5. Scope (amount of material desired)
6. Depth, breadth, technical level
7. Language comprehension (foreign language fluency)
8. Literacy level
9. Media
10. Deadline

Figure 2. Query parameters used in EXPAND and SEGMENT operations. Based on Benson, J., & Maloney, R. K. (1975). Principles of searching. *RQ*, 14(4), 318.

---

### *A Well Thought Out Question Starts With an Understanding of Documents and Bibliographic Organization*

Searchers understand that the question must be translated into the language of the information system in order for an answer to be forthcoming. This requirement is often described as bridging the gap between question and answer (Benson & Maloney, 1975, p. 318). The job of the information seeker, then, is to analyze and classify the question in terms that can relate to documents and bibliographic organization.

### *Understanding Documents and Bibliographic Organization is Complemented by Understanding the Origin of Information*

Information arises as a result of people at work in different settings. Because people have different interests and work within organizations and institutions with different responsibilities, focus, or jurisdiction, the information that arises is about many different subjects. Some information comes about voluntarily. For example, writers pick a subject they are interested in for the purpose of writing a book or magazine article. Researchers choose a topic and, according to traditions of scientific communication, they report their findings in a journal article, conference paper, or technical report. Some

information arises because of reporting or communication requirements, such as government statistics and Congressional testimony. Dissertations are written because of academic requirements for the doctoral degree. Tradition, economics, technology, and market factors dictate that information about many different subjects gets published, packaged, and distributed in certain predictable formats.

The library world recognizes these characteristic differences and accommodates them through arrangement and classification. Academic and special libraries collect and organize reports of scientific, technical, and academic research. These libraries typically segregate this literature into one or more journal collections. Public libraries specialize in consumer, trade, and popular literatures, which get into many separate collections depending on format, content, and audience. Fiction collections reflect the publishing patterns of trade publishers and journal collections reflect the publishing patterns of professional and learned societies and commercial academic publishers. In other words, bibliographic organization (libraries, collections, subject departments, subject headings, and classification) reflects the evolution of information within the social structure.

To the extent that information professionals understand the jurisdiction of government, or the habits of trade publishers, or the activities of citizen groups, they will understand what is in library collections. And if the information is not in their collections, they will anticipate that it exists and be able to guess where it might be located.

### INFORMATION PROFESSIONALS HAVE A TWO-PART MENTAL MAP OF THE INFORMATION ENVIRONMENT

The searcher mentally organizes the information environment into two complementary categories of answer providing sources—one that reflects the institutional structure of society, and one that reflects bibliographic organization (see Figure 3). The institutional structure is the set of institutions, organizations, and professions from which knowledge arises, hence the term *prebibliographic terrain*. The bibliographic structure is the set of form and subject divisions that characterize collections in each type of library, hence the term *bibliographic terrain*.

On the prebibliographic terrain, the searcher encodes information about key institutions and professions, notably academia; government

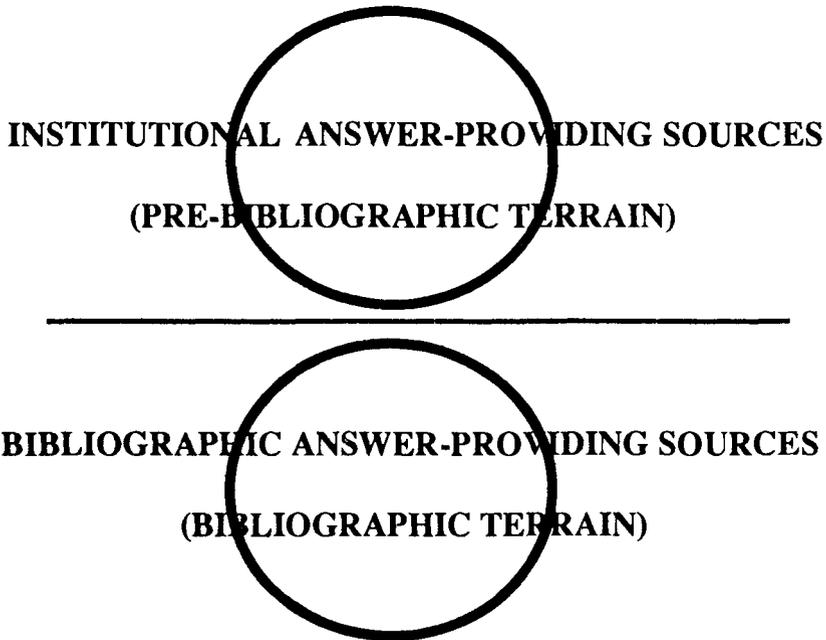


Figure 3. The Searcher's Two-Part, Dual-Context Knowledge Base. The searcher encodes knowledge of the structure of the information environment, including knowledge of the major divisions of the institutional structure through which knowledge arises, and knowledge of the major divisions of the bibliographic organization.

---

agencies; Congress and the courts; industrial, university, and government laboratories; religious groups; trade and professional associations; corporations; small presses; commercial and scientific publishers; news organizations; and free-lance writers, reporters, and consultants. For each institution or category of institutions, there is an associated set of terms, or "attributes," by which the searcher characterizes each source in terms of jurisdiction, purpose, or subject domain. For example, the attributes of government include "public," "official," "regulatory," "taxes," "statistics," "health," "welfare," "education," "environment," and so forth (see Figure 4). The number, scope, and choice of terms will depend on the searcher's familiarity with the institution and the professional demands dictated by the library setting in which he/she works. The choice of terms may be both idiosyncratic, based on what the searcher learns through experience, general education, and the news media; and standardized based on professional education and experience with subject headings and thesauri.

DIVISIONS	PURPOSE, SCOPE	JURISDICTION
ACADEMIA	Research, scholarship	All disciplines, fields of knowledge
GOVERNMENT AGENCIES	Official, public inquiry, investigation, regulation	Taxes, welfare, educ., health, disasters, environment, trade
CONGRESS	Official, public inquiry, oversight, investigation, law-making	(Same as Gov't)
COURTS	Official, public ruling, investigations	Murder, theft, white collar crime, abortion, discrimination
SCIENTIFIC LABS	Basic, applied research, testing	Sci-tech advances all disciplines, industries
BUS., MKETING RESEARCH;	Surveys, scanning of business, markets, politics, economy	All industries
TRADE/PROF. ASSOCIATIONS	Statistics, trends, surveys	All disciplines, industries
CORPORATIONS	Statistics, trends, surveys of products, services, markets	All industries
TRADE PUBLISHERS	Current, historical popular, general, useful info; fiction	Travel, crime, money, self-help, romance, consumerism, reference
SMALL PRESSES	Esoteric, specialized, literary, scholarly info., opinion, lit.	Politics, social issues, controversies
MEDIA	News, entertainment	Current internat'l, nat'l, local events, notices
RELIGIOUS GROUPS	Opinion, interpretation	Religion, welfare, social issues

Figure 4. The Searcher's Knowledge Base: Prebibliographic Divisions of the Information Environment. The searcher characterizes institutional sources in terms of jurisdiction, subject domain, purpose, and scope of activities. The searcher classifies the query in terms of these attributes in order to translate the query into system terms. Attributes are part of the interface language that bridges the gap between the query and search paths. Only selected attributes of major divisions of the prebibliographic terrain are shown for purposes of illustration. A comprehensive associative network with all divisions and a rich vocabulary of attributes has yet to be created and graphically displayed.

Figure 5 summarizes the most important divisions of the bibliographic terrain along with associated attributes. Attributes characterize bibliographic divisions by subject, content, genre, audience, format, and currency. Functionally, each division of the information environment is a generic set of answer-providing sources or search path. Benson and Mahoney (1975) called these classes "macrosystems" to distinguish them from special tools or "microsystems" (p. 317).

Because of the relationship between knowledge creation and bibliographic organization, these two main divisions of the information environment can be conceived as complementary contexts for encoding information about information sources, hence the label "two-part, dual context knowledge base."

The prebibliographic terrain may, in fact, be the mental model held by any information seeker who is not trained in library science. Given people's experiential relationship to social structure, and their secondary relationship to bibliographic organization, people may be more attuned to institutions and organizations than to libraries and documents. As Hunt (1989) notes in a discussion of cognitive theories of classification, novices use different classification schemas than experts, "simply because they have less sophisticated theories about how the field is organized" (p. 622). The reliance on a mental model of institutional sources may persist for end-users who use it exclusively. However, the analysis of case studies of reference and this author's experience suggest that professional intermediaries use the prebibliographic terrain as a backup. There are situations when a search path is more easily identified by thinking in terms of the origins of the information.

For example, the patron asks a science librarian about getting information about DNA fingerprinting. In the course of the question negotiation and the search, the librarian might ask, "Who would be involved with that issue?" or, "Who's likely to be engaged in that topic?" Maybe the librarian's mind even scans the prebibliographic terrain for likely sources, forming a picture of institutional activity and people at work. Lawyers, expert witnesses, geneticists, and regulatory agencies are considered. These thoughts evoke ideas about likely search paths. This helps in querying the patron further to determine which aspect of the subject is most important in order to make the search more relevant.

Perhaps the librarian learns that the patron is interested mainly in the legality of forensic evidence and decides that the patron is better off starting the search at the law library. If the patron wants a comprehensive search, the librarian is reminded to search *Legal Research Index*, the NTIS database, and the *Monthly Catalog* as well

<b>DIVISIONS</b>	<b>CONTENT, USE</b>	<b>SUBJECT</b>
<b>ACADEMIC LIB.</b>	Technical, scholarly	All disciplines, fields
<b>SPECIAL LIB.</b>	Technical, applied	All industries, fields
<b>PUBLIC LIB.</b>	Practical, popular, trade	Consumer, car repair, travel, finances
<b>READY REFERENCE</b>	Available, "wonder," "first facts," stats, overview, current	People, companies, events, assn., terms products, words
<b>HANDBOOKS</b>	How-to, tables, facts	Science, consumer
<b>DICTIONARIES</b>	Meanings, definitions	Lang., all subjects
<b>ENCYCLOPEDIAS</b>	Historical, background	All fields, disciplines
<b>ALMANACS</b>	Historical, facts, stats	Politics, economics, geog., biog., events
<b>DIRECTORIES</b>	Addresses, names	All fields, industries
<b>MAPS</b>	Graphic, overview	Places, features, routes
<b>GOV'T DOCUMENTS</b>	Public, official, legal	Trade, social welfare, envir., war, disasters
<b>BOOKS</b>	Established facts	All, fiction and non
<b>JOURNALS</b>	Current, research, technical	All disciplines, fields
<b>MAGAZINES</b>	Current, popular, pictures	Hobbies, household, news, social issues
<b>TECH REPORTS</b>	Current, sponsored research, technical	Advancing, hot fields
<b>PAMPHLETS</b>	Inside, facts, how-to	Consumer, hobbies
<b>NEWSPAPERS</b>	Current, local, news, photos	Crimes, politics, war
<b>NEWSLETTERS</b>	Up-to-the-minute, inside, opinion	Hobbies, politics, industries, social
<b>CHILDREN'S LIT.</b>	Large-print, illustrated, easy	Families, pets, science
<b>BIOGRAPHIES</b>	People facts, views	Famous, prominent

Figure 5. The Searcher's Knowledge Base: Bibliographic Divisions of the Information Environment. The searcher characterizes libraries, collections, and formats in terms of subject, content, use, availability. The searcher classifies the query in terms of these attributes in order to translate the query into system terms. Only selected attributes of major divisions of the bibliographic terrain are shown for purposes of illustration. A comprehensive associative network with all divisions and a rich vocabulary of attributes has yet to be created and graphically displayed.

as *Biotechnology Abstracts*. In other words, thinking prebibliographically evokes options for positioning the question in one or more subject contexts and formats. Because the prebibliographic terrain and the bibliographic terrain are complementary, it requires only a small mental shift to reposition the search paths from institutional sources to bibliographic sources.

Every question does not require this train of thought. It depends on experience with the subject, the library setting, and the requestor's true need. At the opposite end of the scale is the known item search in which the requestor asks for a specific answer source by title and the searcher does not have to reason out a search strategy. In between these extremes are questions that require only bibliographic thinking. The searcher analyzes the question in terms of the attributes of libraries or documents, a process which directly links the question to one or more document classes, without reference to institutional sources.

Theoretically the map of the information environment is the same for all searchers. Operationally it will differ depending on the library setting, clientele, expectations, reference policy, and the particular question being asked. In actuality, it will differ by skill level, education, training, and experience. The less knowledge of bibliographic organization, the more the searcher will rely on prebibliographic thinking to evoke ideas for search paths. On the other hand, the more experienced searcher will also rely on prebibliographic thinking for difficult questions and to avoid mental ruts.

#### SEARCH PATH SELECTION PRECEDES SELECTION OF A SPECIFIC ANSWER SOURCE

As the DNA example illustrates, selection of a search path often precedes selection of a specific answer source. The searcher chooses law, biotechnology, or government before selecting *Legal Resource Index*, *Biotechnology Abstracts*, or the *Monthly Catalog*. If the question is about tuna fishermen and the killing of dolphins, the searcher can choose one or more subject domains—e.g., public affairs, science, environment, government. Each is a different search path. The choice also may be driven by format given the characteristic relationship between content and format. Technical reports, newspaper stories, and films are very different search paths.

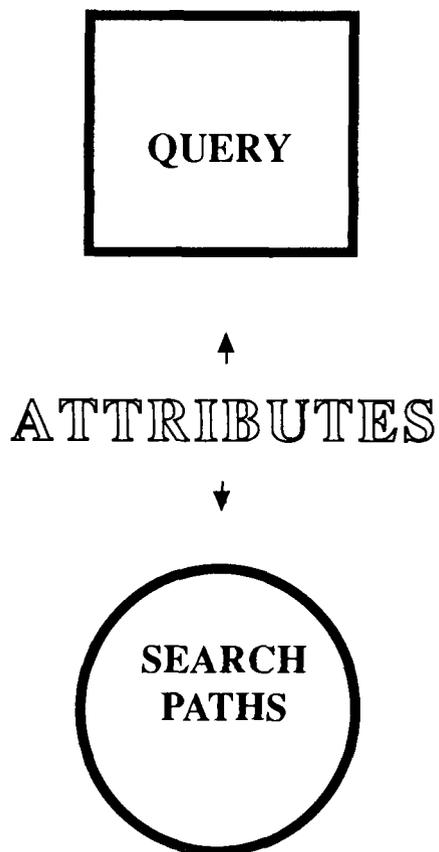


Figure 6. Attributes are the Link Between the Query and the Information System. The searcher analyzes the query in terms of the characteristics of answer-providing sources. In so doing, the query is translated into system terms.

---

### COGNITIVELY, ATTRIBUTES LINK THE QUESTION AND SEARCH PATHS

Attributes are the interface language that translates the question into system terms. Questions must be formulated and further analyzed with these attributes in mind. This is required in order to make the mental leap between questions and answer sources. Attributes link questions and answers. They are the mental bridge (see Figure 6).

### THE MORE ATTRIBUTES FOUND, THE MORE FLEXIBLE AND CREATIVE THE SEARCH STRATEGY

In one of Grogan's (1972) case studies, the user wanted to know the local authority for "Great Cheverell." Grogan classified the question as "geographic," an attribute of gazetteers, atlases, and maps. However, this search path hit a dead end. Great Cheverell did not seem to be a place name. He re-analyzed the question and reclassified Great Cheverell as an "administrative unit," an attribute of administrative area maps prepared by government agencies. This search path was fruitful. Ultimately he checked Ordnance Survey county administrative area maps where he found the answer.

The second example is a question that came over the telephone to the science reference desk. A caller asks "What is the correct spelling for the mineral meehanite and what is it made from?" "Mineral" is an attribute of "geology," which is a specialized subject. "Correct spelling" is an attribute of dictionaries. "Made from" is conceptualized as "formula" or "properties," which are conceived as "factual data." These are all attributes of reference materials. Putting it all together, the searcher now knows that she is looking for a subject dictionary in geology. Knowing the arrangement of her reference collection and the classification for geology, she goes to the shelf and retrieves the appropriate source. The example is simple because we make this translation almost automatically. Cognitively, however, it is a fascinating walk through the associative map of the information environment.

### APPLICATION OF THE MODEL TO QUESTION FORMULATION

Question formulation is a process whose goal is to position the question along one or more search paths. A good subject question is expressed in interface language, which orients the question outward toward the information system rather than inward toward the topic of the question. Interface language uses the vocabulary of attributes and search paths to specify the direction of the search. Although the requirements for a good subject question are the same for all queries, the amount of cognitive work that the user must do depends

on the subject of the question, its complexity, the original wording of the question as it comes into the user's mind, and the outcome of the initial search for information.

The utilization of the question formulation method makes the end-user an active participant in the reference process and thus a better consumer of what the library has to offer. To the extent that the end-user understands the divisions of the information environment and the attributes of answer-providing sources, the more independent the user will be. Knowing the rules for question formulation empowers the user (Huston & Perry, 1987).

Figures 7 and 8 summarize the operations that seem to be required for clearly and completely expressing a need for information. Figure 7 shows that the method is not always a straight path from "representation" to "specification." Depending on the complexity of the question or the outcome of the initial search, there may be need to re-analyze the query until new search paths open.

The end-user starts with a topic-oriented "representation" of the information need and concludes with interface language that re-orientes the question to the information environment. At that stage the intermediary can complete the translation into system terms and define a search strategy directed to sources in or out of the library—i.e., to specific libraries, collections, tools, or outside agencies.

### *Rules for Question Formulation*

To "represent" is to start the process by internally voicing the main topic of the information need. The initial "representation" is oriented inward to the subject, not outward to the system that will satisfy the information need. For example, the user wonders about the first use of the SOS signal, or the effect of infant bonding on adolescent development, or the design requirements for a smart robot, or the colleges that offer a major in evolutionary biology.

To "expand" is to shift the focus of attention beyond the narrow confines of the topic. An expand defines *the project* (not just the subject of the question) in terms of topic, deadline, kind and amount of information, and so forth. Expansion of the topic also involves a "segmentation" of the topic into all relevant concepts, similar to preparing a database search. "First use," "SOS signal," "infant bonding," and "adolescent development," "design requirement," "smart robot," "college majors," and "evolutionary biology" are the concepts from the examples above. The following is an illustration of the "expand" operation for the question about infant bonding:

#### *Topic Parameters*

Subject  
Time frame  
Geographic scope

Bonding/adolescent development  
Current  
United States

*Project Parameters*

Purpose	Undergraduate term paper
Amount of information	Ten sources
Depth, breadth, technical level	Basic, background
Language comprehension	English
Literacy level	College
Media	Print
Deadline	Next week

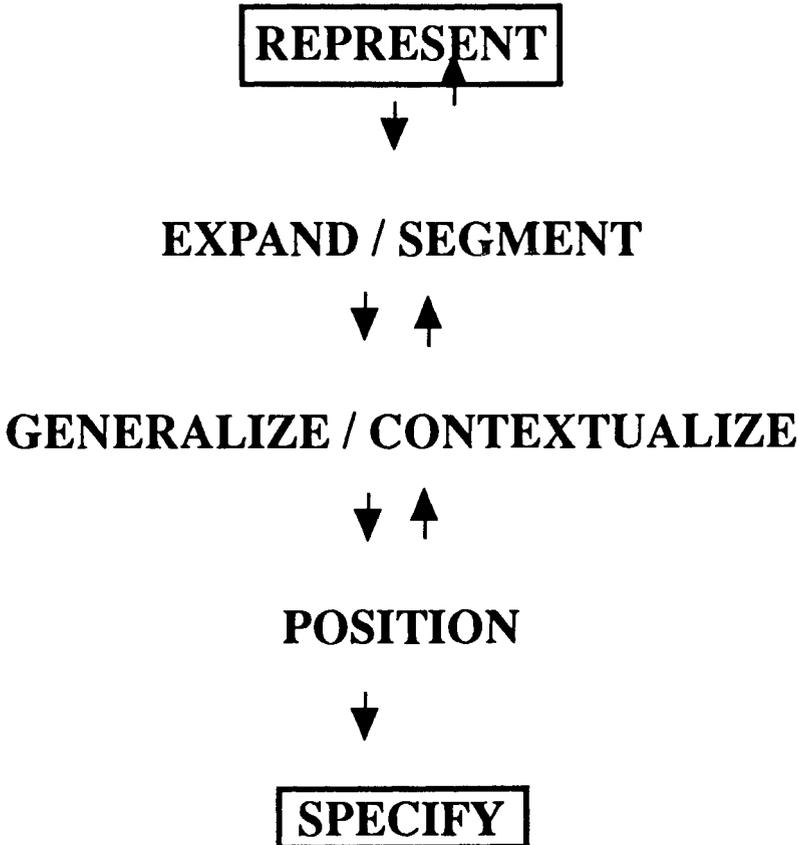


Figure 7. Operations for Formulating Good Subject Questions. The initial REPRESENTATION is topic-oriented. The final SPECIFICATION is system-oriented and describes the characteristics of potential answer-providing sources. The re-orientation from REPRESENTATION to SPECIFICATION depends on the end-user's knowledge of the divisions of the information environment and their attributes.

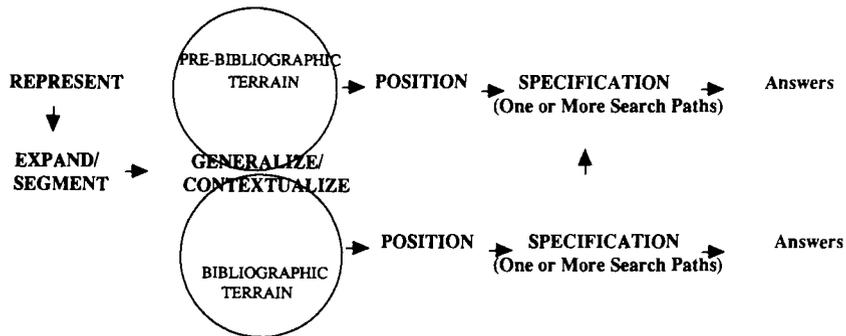


Figure 8. Formulating Good Subject Questions Using Knowledge of Institutional and Bibliographic Answer-Providing Sources. The initial REPRESENTATION of the question is EXPANDED to provide a complete description of the project, including topic, deadline, amount of information, and intended use. The results of the EXPAND are then analyzed in order to conceive of the query in broader, more abstract terms and to place the topic in context (GENERALIZATION/CONTEXTUALIZATION). The analysis, which relies on an understanding of the divisions of the information environment, defines the query in terms of the characteristics of answer-providing sources. These characteristics, or attributes, of sources are the interface language that translates the query into the language of the information environment. This language is used to SPECIFY potential search paths and the characteristics of ideal answer-providing sources.

To generalize/contextualize is to interpret the *project* (topic and other parameters) at a higher level of abstraction or to identify equivalent, related, or broader terms that evoke for the user ideas about likely answer-providing sources. Using the preceding example, the user can say that he is looking for some "basic," "background," "current," "available," "printed" materials about a "psychology" topic. These attributes are a more abstract, information system-oriented way of saying that this undergraduate is writing a brief term paper on adolescent development and infant bonding and that the paper is due next week.

The purpose is not to find subject headings or indexing terms. That is a later step outside the scope of these operations. Rather, the goal is to conceive of the query in as many different ways as

necessary to get ideas for likely search paths—e.g., for the example above, psychology texts and magazines. By looking at the subject of the query conceptually, these operations will help guide the user to appropriate domains and formats.

These interrelated operations typically involve normal everyday classification and categorization of ideas, terms, and concepts. The operations rely on experience, general education, and attunement to the culture. The outcome is also heavily influenced by the user's cognitive flexibility, which is required in order to avoid "hardening of the categories" (Neill, 1975, p. 314).

Specifically, to "generalize" is to broaden the categories for each concept involved in order to conceive of the query in more general terms. An airplane accident is a transportation accident. Holsteins are cows. Cars are vehicles. Peonies are flowers, and robins are birds. Wilson (1968) called these "analytic associations." Analytic associations are a matter of definition and tend to imply a hierarchical relationship.

"Contextualizations" help define context. Broadly defined here, they are a form of conceptualization that places subjects in a certain light. Contextualizations define subjects and link one subject to other subjects, events, ideas, and subject domains that are related as a matter of historical, biographical, and social fact. Donald Trump is a billionaire, a real estate developer, a tycoon, an empire builder. "Infant bonding" is a psychological variable. Adolescent development is a popular topic for the media, clinicians, and social scientists. A question about Holsteins may be a question about cows in the context of biology or farming or agribusiness or all of the above. A question about sauna baths is a question about machinery or consumerism or fads or health or Swedish popular culture. A question about diamonds can be a question about economics or minerals or jewelry or wedding customs. A question about knots can relate to cowboys, religion, science, fishing, sailing, and execution by hanging. A plane crash is a technological failure, a civil disaster, and perhaps, an international incident.

The purpose of these abstractions is to evoke ideas about sources of information. Perhaps the concepts "international incident" or "public disaster" are more evocative than "plane crash" for identifying different search paths.

Each person has a unique semantic network of associations. Certain terms are more or less likely to evoke ideas about answer-providing sources and will vary according to the user's need, purpose, experience, and frame of reference.

To "position" is to use the output of "generalizing" or "contextualizing" to orient the question to one or more subject

domains and formats. This operation uses either bibliographic or prebibliographic thinking, depending on the user's familiarity with the information system.

Using prebibliographic thinking, the user thinks in terms of the origins of information within the social structure. In this case, "positioning" is a journalistic approach to the question. The ability to carry out the operation relies on the user's knowledge about how society works rather than on knowledge of the information environment.

When users learn to "draw from their own experiences with social organization" they will figure out where to find information about many subjects (Huston, 1983, p. 186). By approaching the question like a journalist rather than like a librarian, the user will figure out that news reporters cover plane crashes, the federal government must launch an official inquiry, and insurance companies will certainly develop expertise because it is their business to cover the airlines and passengers (Rubens, 1982, p. 14). Maybe the user forgets government sources when the subject is conceptualized as "plane crash." However, "contextualizing" the subject as a question about a "public disaster" or a "regulated industry" may activate the associations that lead to the anticipation that government is a likely source of information.

In approaching the question this way, the user will use a "probe" technique that is typically used by professional searchers in practice but underreported in the literature. Huston (1983) mentioned the following "probing" questions to illustrate the journalistic or sociological approach to question analysis: "What group of people/kind of profession would have thought about this topic? How would they have presented the information—a film, report, book?" (p. 186).

It does not require library training to make these associations. These associations come about through what is learned about and experienced in the world. The ability to think abstractly like this is a function of general education, upbringing, and life experience (Rubens, 1982, p. 3). As Huston (1983) says:

With such a sociological approach, information is not cataloged in terms of bibliographic organization but in terms of the sectors of society and human enterprises from which information about different subjects, packaged in different formats, arises. Such an awareness is developed as we are encultured in society.... (p. 186)

If the user "positions" using bibliographic thinking instead, the output will be expressed in terms of literatures or collections rather than institutions or professions—i.e., government documents rather than government agencies; medical literature rather than hospitals; research reports rather than laboratories or scientists; newspaper collections rather than news organizations or reporters; fiction

collection rather than trade publishers or creative writers. However, regardless of the cognitive terrain used to position the query, the outcome is a description of an answer-providing source (see Figure 8).

To "specify" is to express, in one or more declarative sentences, the attributes of the ideal answer-providing source(s). A "specification" is the voiced query using interface language.

Using the example about adolescent development, the initial "representation" is something like "I'd like to know more about infant bonding and adolescent development." The final "specification" gives the context for the question, the subject, and the characteristics of answer-providing sources. For example, the user might say the following (interface language is printed in all capitals):

"I'm writing a term paper on infant bonding and adolescent development. I need about 10 BASIC sources that are CURRENT and NOT TECHNICAL. I'll need material available here because the paper is due next week. I know the topic is pretty popular right now so I figure that research is being reported in PSYCHOLOGY MAGAZINES, PARENTING MAGAZINES, and NEWSPAPER articles. If I have to use the PSYCHOLOGY LITERATURE, it should be a TEXTBOOK I can understand. I want to stick with the PSYCHOLOGY LITERATURE because I'm interested in INDIVIDUALS, and not GROUPS of people, and I'm interested mostly in my OWN CULTURE."

If selected materials are not on the shelf, the user theoretically will have alternatives in mind, such as using related literature in "sociology" and "anthropology" and "education."

In this "specification," the user names attributes such as "basic," "current," "not technical," "can understand," "available," "individuals not groups of people," "my own culture," "popular topic," "psychology," all of which result from his "expansion," "generalization," and "contextualization." The attributes of current, basic, easy to read, popular, and readily available point to the popular press and general magazines and texts. Although not expressed in the "specification," the user has classified the topic as belonging to the field of psychology because "bonding" is a psychological variable and "development" is a focus of interest for psychologists.

In this case, the thinking seems to be bibliographic, resulting in the "specification" of literatures and formats rather than institutions or professions. However, depending on experience and the motivation for the project (perhaps the user is in therapy), the user may have first thought prebibliographically in terms of psychologists, writers, and reporters. If so, it is probably a small mental leap for most college students from "psychologists" to

“psychology journals” and from “writers” and “reporters” to “magazines” and “newspapers.” The route to a bibliographic “specification” is sometimes by way of a nonbibliographic “specification” (see Figure 8).

Here is a complete example, based on a reference transaction at the author’s library, to illustrate the question formulation method. In this example, the initial question is straightforward. The “representation” is close to a “specification.” However, a printed source was not available. The student had to find the answer indirectly by networking through the academic community. Prompted by the librarian, the student approached the question journalistically and came up with a search strategy using prebibliographic thinking. Representation:

“I’m trying to find colleges that offer a major in evolutionary biology.”

*Expansion/Segmentation:*

*Topic Parameters*

Subjects	College major/evolutionary biology
Time frame	Current
Geographic scope	United States

*Project Parameters*

Purpose	Admissions
Amount of material	One list
Depth, breadth, technical level	Description/ranks
Language comprehension	NA
Literacy level	NA
Media	Print
Deadline	Three weeks

*Analysis/contextualization:*

Selective, evaluative, current list of colleges in the biological or natural sciences.

*Position:*

*Bibliographic thinking:* directories or guides in the science collection.

*Specification 1:*

I want a current list of colleges that have majors in evolutionary biology. I want to contact some programs so I need the address of the admissions office and a phone number. I want to know the best programs so I need some reviews or evaluations or ratings. Maybe there’s a selected guide like I’ve seen for some other science programs.

*Re-analysis and contextualization:*

People who are prominent in the field or considered to be authorities.

*Position:*

Prebibliographic thinking using “probe”: If people are prominent in the field or considered to be authorities, where might their name be seen? What role do they play? Who would know about good programs?

EDITORS and EDITORIAL BOARD of KEY JOURNALS  
in the field, PROFESSIONAL ASSOCIATIONS,  
AUTHORS of current books and review articles.

*Specification 2:*

I'm looking for a list of the key journals and books in the field of evolutionary biology. I want to see which schools these people are with. I may also need the names of large professional associations.

## DISCUSSION

Will students appreciate the value of learning the question formulation model and techniques? Yes, if introduced in consumer terms. The question formulation model teaches the attributes of answer-providing sources. Students, like other consumers, understand the importance of knowing product characteristics before making a purchasing decision. In this age of empowerment and autonomy, students may welcome an opportunity to participate actively in the reference transaction.

Can the question formulation model be taught? Yes. The literature on teaching students by building on what they know (Huston, 1983) suggests that students understand a great deal about social structure and the attributes of information sources, and that they learn to relate that knowledge to bibliographic organization. Much of this knowledge is intuitive based on observation and experience; some aspects must be taught or brought to conscious awareness. Students can learn the attributes of answer-providing sources if told what they are and why they are important to information seeking. They can be taught to pose questions in terms of interface language if they are given practice in analyzing and contextualizing questions and given proof that a “specification” is more evocative and personally empowering than a “representation.”

Is the thinking-like-a-searcher model sound? Based as it is on soft methodological techniques, the model and its application to question formulation need to be field tested using model curricula. It will be a challenge to create curriculum material. A comprehensive map of the information environment has yet to be created and will

require the combined efforts of cognitive scientists, knowledge engineers, and information specialists. However, major divisions of the information environment and their attributes, as suggested in Figures 4 and 5, are readily identifiable. Librarians work everyday with subject classifications, authority lists, and reference tools that can be used as a basis for teaching the institutional structure, subject fields, academic disciplines, literature types, and formats. Some of these tools are Library of Congress subject headings, Library of Congress and Dewey Decimal classifications, Superintendent of Documents classification, *Encyclopedia of Associations*, and DIALOG's list of subject categories. Using creative thinking techniques such as brainstorming and visualizations, hands-on experience with different literatures and literature types, and practice analyzing questions in terms of attributes of answer-providing sources, students can begin to speak the language of knowledge creation and bibliographic organization. In so doing, end-users become active partners in the information-seeking process.

#### ACKNOWLEDGMENT

Nancy J. Rohde and Robin Crickman, my faculty advisors at the University of Minnesota School of Library and Information Science, provided invaluable support during the initial phase of the research leading to my master's thesis. I am indebted to Mary Huston for seeking out my work, assisting me in the revision and publication of the original model, and inviting me to apply the results to bibliographic instruction.

## APPENDIX

(Terms from this glossary used to define other glossary terms are italicized)

**Attributes.** Characteristics of institutional and bibliographic sources. "Current" is an *attribute* of newspapers, newsletters, magazines, and journals. "Public" and "official" are *attributes* of government sources.

**Bibliographic Terrain.** The part of the *knowledge base* having to do with bibliographic organization. Complemented by the *prebibliographic terrain*.

**Contextualize.** To conceptualize the question in terms of the context to which it belongs. A question about Holsteins can be a question about veterinary medicine, farming, food.

**Expand.** To shift the focus of the question beyond the topic. Defines the project, not just the topic, by considering deadline, purpose, use of the material, language requirement, amount of material needed, and other *question parameters*.

**Generalize.** To interpret the question at a higher level of abstraction. A question about Holsteins is a question about cows or animals.

**Information Environment.** All sources of information, including institutional sources and document sources. The *information environment* extends beyond the local collection to include all potential answer-providing sources, be they persons, groups, or documents.

**Interface Language.** Vocabulary consists of *search paths* and *attributes*. Used to specify the characteristics of selected answer-providing sources and the direction of the search.

**Knowledge Base.** What the professional intermediary knows about the *information environment*. Elaborate map of the divisions of the *information environment* and their *attributes*. A two-part, dual context representation of the *information environment*, encompassing knowledge of institutional sources from which information arises (*prebibliographic terrain*), as well as knowledge of libraries, collections, and classes of documents (*bibliographic terrain*).

**Position.** To orient the question to one or more subject domains or formats. To determine appropriate *search paths*.

**Prebibliographic Terrain.** The part of the *knowledge base* having to do with institutional (primary) sources. Complemented by the *bibliographic terrain*.

**Probe.** To approach the question journalistically or sociologically. To ask "Who is interested/writes about this topic?"

**Question Parameters.** Topic and Project-oriented facets of the question. Topic facets include subject, time frame, and geographic scope. Project facets include purpose, amount of material desired; depth, breadth, technical level; language comprehension, literacy level, media, deadline (see *Expand*).

**Represent or Representation.** The initial internal expression of an information need. Tends to be topic oriented.

**Search Path.** The direction that the search will go. Expressed in terms of classes of documents, collections, libraries, institutions. "Newspapers" is a *search path* as is "trade associations."

**Segment.** To divide the topic of the question into its individual concepts as is done to prepare for a database search.

**Specify or Specification.** To express, in one or more declarative sentences, the *attributes* of appropriate answer-providing sources and the direction of the search. Final result of question formulation method (compare *represent*).

**Thinking Bibliographically.** Analyzing a question in terms of the *attributes* of libraries, collections, documents.

**Thinking Prebibliographically.** Analyzing a question in terms of the *attributes* of institutional sources.

## REFERENCES

- Allen, G. (1990). Database selection by patrons using CD-ROM. *College & Research Libraries*, 51(1), 69-75.
- Bates, M. J. (1979a). Information search tactics. *Journal of the American Society for Information Science*, 30(4), 205-214.
- Bates, M. J. (1979b). Idea tactics. *Journal of the American Society for Information Science*, 30(5), 280-289.
- Benson, J., & Maloney, R. K. (1975). Principles of searching. *RQ*, 14(4), 316-320.
- Borgman, C. L. (1982). Mental models: Ways of looking at a system. *ASIS Bulletin*, 9(2), 38-39.
- Doyle, J. M., & Grimes, G. H. (1976). *Reference resources: A systematic approach*. Metuchen, NJ: Scarecrow.
- Grogan, D. J. (1967). *Case studies in reference work*. London: Clive Bingley.
- Grogan, D. J. (1972). *More case studies in reference work*. London: Clive Bingley.
- Hunt, E. (1989). Cognitive science: Definition, status, and questions. *Annual Review of Psychology*, 40, 603-629.
- Huston, M. M. (1983). Rethinking our approach for research instruction. *Research Strategies*, 1(4), 185-187.
- Huston, M. M. (1988). *A theoretical framework and an instructional model for educating end users of online bibliographic information retrieval systems*. Unpublished doctoral dissertation, Union Graduate School, Cincinnati, Ohio.
- Huston, M. M., & Perry, S. L. (1987). Information instruction: Considerations for empowerment. *Research Strategies*, 5(2), 70-77.
- Ingwersen, P. (1984a). A cognitive view of three selected online search facilities. *Online Review*, 8(5), 465-492.
- Ingwersen, P. (1984b). Psychological aspects of information retrieval. *Social Science Information Studies*, 4(2&3), 83-95.
- Kereztesi, M. (1982). The science of bibliography: Theoretical implications for bibliographic instruction. In C. Oberman & K. Strauch (Eds.), *Theories of bibliographic instruction: Designs for teaching* (pp. 1-26). New York: Bowker.
- Neill, S. D. (1975). Problem solving and the reference process. *RQ*, 14(4), 310-315.
- Rubens, D. (1982). *Cognitive strategy and the search process: A four-fold, dual context memory representation*. Unpublished master's thesis, University of Minnesota School of Library and Information Science, Minneapolis.
- Rubens, D. (1989). Thinking like a searcher. In J. C. Fennell (Ed.), *Building on the first century* (Proceedings of the fifth national conference of the Association of College and Research Libraries) (pp. 337-42). Chicago, IL: ALA.
- Swift, D. F.; Winn, V. A.; & Bramer, D. A. (1979). A sociological approach to the design of information systems. *Journal of the American Society for Information Science*, 30(4), 215-223.
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29, 178-194.
- Tiero, M. J., & Lee, J. H. (1983). Developing and evaluating library research skills in education: A model for course-integrated bibliographic instruction. *RQ*, 22(3), 284-291.
- Wilson, P. (1968). *Two kinds of power: An essay on bibliographic control*. Berkeley, CA: University of California Press.

## ADDITIONAL REFERENCES

- Cohen, P. R., & Kjeldsen. (1987). Information retrieval by constrained spreading activation in semantic networks. *Information Processing & Management*, 23(4), 255-268.
- Foskett, A. C. (1977). *The subject approach to information* (3d ed.). London: Clive Bingley.
- Gardiner, G. L. (1969). The empirical study of reference. *College & Research Libraries*, 30(2), 130-155.
- Jahoda, G., & Braunagel, J. S. (1980). *The librarian and reference queries: A systematic approach*. New York: Academic Press.

- Ingwersen P. (1982). Search procedures in the library analyzed from the cognitive point of view. *Journal of Documentation*, 38(3), 165-191.
- Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology*, 30, 363-396.
- Thompson, F. B. (1973). The dynamics of information. *The Key Reporter*, 38(1), 2-4, back cover.
- Walker, D. E. (1981). The organization and use of information: Contributions of information science, computational linguistics and artificial intelligence. *Journal of the American Society for Information Science*, 32(5), 347-363.
- Watters, C. R., & Shepherd, M. A. (1987). A logic basis for information retrieval. *Information Processing & Management*, 23(5), 433-445.