AN ANALYSIS OF QUESTIONS AND ANSWERS IN LIBRARIES

by

Caroline E. Hieber

A THESIS
Presented to the Graduate Faculty of Lehigh University in Candidacy for the Degree of Master of Science in Information Sciences

Lehigh University 1966
Center for the Information Sciences
Lehigh University

STUDIES IN THE MAN-SYSTEM INTERFACE
IN LIBRARIES

Report No. 1

An Analysis of Questions and Answers in Libraries

by

Caroline E. Hieber

The work reported here was partially supported by grant from the Air Force Office of Aerospace Research, AF-AFOSR-724-66, and by the Office of Education, Cooperative Research Project No. S-235.

June 1966
AN ANALYSIS OF QUESTIONS AND ANSWERS IN LIBRARIES

by

Caroline E. Hieber

ABSTRACT

Traditional methods of analyzing questions and answers as they occur in reference libraries are discussed and criticized. Methods of examining questions, the question-answering process, and answers to questions are evaluated. A pragmatic scheme is suggested which classifies answers by their formats, dividing them into exact-reproduction type, fill-in-the-blank type, short descriptive type, information-about type, and list-of-references type. This classification is applied to a collection of actual inquiries. Recommendations for further testing of the suggested classification are given, with a discussion of its value in present and future libraries.
This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

January 4, 1966
(date)

[Signature]
Professor in Charge

[Signature]
Head of the Department
Work on this thesis was partially supported by grant from the Air Force Office of Aerospace Research AF-AFOSR-724-66.
CONTENTS

Introduction...............................................................2
Part I. Use Studies.........................................................4
Part II. Some Alternative Methods.................................14
Part III. A Suggested Classification.................................29
Part IV. Recommendations and Implications.......................37
Bibliography..............................................................43
Vita.....................................................................49
Traditional methods of analyzing questions and answers as they occur in reference libraries are discussed and criticized. Methods of examining questions, the question-answering process, and answers to questions are evaluated. A pragmatic scheme is suggested which classifies answers by their formats, dividing them into exact-reproduction type, fill-in-the-blank type, short descriptive type, information-about type, and list-of-references type. This classification is applied to a collection of actual inquiries. Recommendations for further testing of the suggested classification are given, with a discussion of its value in present and future libraries.
AN ANALYSIS OF QUESTIONS AND ANSWERS IN LIBRARIES

INTRODUCTION

Librarians, at least for the last half-century, have been concerned about library users' needs and the users' opinions about the filling of these needs. As members of a profession which considers itself service-oriented, librarians have attempted by various techniques of examination and evaluation to assess the services offered by libraries, and then, in light of the results, to improve these services. The acquisition of a well-rounded and useful book collection was the purpose of many of the early studies; later ones were conducted with an eye to improving training of personnel and arrangement of facilities, establishing policies such as telephone reference service, and developing better means of record-keeping.

In no area of librarianship has there been greater difficulty in analyzing services rendered than in the field of reference work. The problem is inherent in the non-routine nature of the work, and has been compounded by a general lack of knowledge about the question-and-answer process itself.

Libraries now are bearing increasing responsibilities for making information available to a growing number of serious inquirers, and new methods of doing this are being developed and implemented. Reappraisals of reference
service and better knowledge of the types of answers desired by seekers of information are therefore needed. To this end, traditional techniques of evaluation of reference work should be analyzed and untried ones considered.
PART I
USE STUDIES

Most attempts to evaluate library reference work have taken the form of use studies, performed either by counting and classifying questions, or by investigating the ways in which library users, generally scientists, seek information.

Among the simplest types of use studies are reference or citation counts. Footnotes and bibliographies of journal articles are examined in order to discover which journals are most frequently referred to. In this way librarians can ensure that the most-cited journals are in their collections. The disadvantage is, of course, that the citations do not necessarily reflect what the articles' authors knew at the time of writing, but rather what was available (and indexed).

Acquisitions policies have also been reviewed by studying card catalog use. Patrons using library card catalogs are watched or interviewed to learn which materials they are seeking and how successful they are in their

searches. Some of these studies have as their purpose the improvement of descriptive cataloging or the rearrangement of the catalog itself. Examples of card catalog use studies are the investigations of Jackson\(^2\) and of Brooks and Kilgour.\(^3\)

The first analyses of actual reference questions were done in public libraries, a result not only of these libraries' being the first to offer reference service but also of the social changes brought by the Depression era.\(^4\) The results of most of the studies conducted during the 1930's and 1940's are summarized in Bernard Berelson's *The Library's Public*.\(^5\) While the various studies differed in approach and in emphasis, most of them classified users by occupation and questions by subject. Reference questions asked in public libraries during periods ranging in length from one day to 20 years were categorized, usually according to the Dewey Decimal system. It was generally found that the majority of users could be classed as professional workers and students, and that their queries fell mainly

---


into the 300's, 600's, and 900's of the Dewey Decimal classification.\(^6\)

The results of these investigations seem to have little relevance to present-day reference work. Classification of questions into broad subject areas was useful in formulating acquisitions and training policies, but, with the ever-increasing number of subjects of inquiry, it is unrealistic to use categories set up nearly a century ago. The Dewey Decimal classification is inadequate for collections of large public libraries, let alone research libraries, and there is no reason to suppose that it is any more satisfactory for research questions.

An exception to classification by subject was the system devised by Alexander.\(^7\) He grouped questions into the following seven types:

1. Fact type
   
   (a) Meaning type
   
   (b) Numerical or statistical type
   
   (c) Historical type
   
   (d) Exact wording type
   
   (e) Proper name type

2. How to do type

\(^6\)Examples of these studies are listed in the bibliography.

3. Trends type
4. Supporting evidence type
5. "All about" type
6. Evaluation of reference type
7. Duplication of previous work type.

Alexander's seven types were later expanded to 13 by Dorothy E. Cole, who was unable to fit all of the inquiries she was analyzing into Alexander's categories. Cole found that fact type questions accounted for 55% of her total; general information type (her substitution for Alexander's "all about" type), for 20%; how to do type for 10%; and supporting evidence type for 8%.

A different approach was followed by Chait, whose purpose was to learn which sources were most frequently used to find answers to questions asked in branches of the Brooklyn Public Library. Chait felt that an analysis of sources could form a basis for acquisitions policies, aid in training reference personnel, and help in arriving at a convenient and useful shelf arrangement.

Florence Van Hoesen also analyzed sources as well as questions in her doctoral dissertation, an attempt to

---


9 William Chait, "Books and Other Sources Used in Adult Reference in Branches of the Brooklyn Public Library" (Unpublished, 1938).
outline a new method of teaching reference work. Both the Chait and Van Hoesen studies, incidentally, found that questions in the fields of social science, science, and biography predominated. Van Hoesen also classified her questions by type. Requests for specific facts, general information, and "how-to-do-it" information were the most frequent.

Since the publication of Berelson's book in 1949, there has been little organized effort on the part of public libraries to assess their reference services. This is probably partly because of the definitive nature of The Library's Public and partly because of the increased work-load of reference librarians in the 1950's and 1960's. Many reference librarians tally the questions they receive, often breaking them down by the amount of time it takes to answer them, a factor that depends as much on the quality of the librarian and the collection as on the question. The totals appear in statistical reports but the subtotals are seldom evaluated.

The general conclusion of librarians seems to be that quantitative measurements of reference work have little value. Louis Shores writes:

Classification of questions by type, or by inquirer, or by sources consulted, or by purpose,

all present formidable obstacles both as to
validity and reliability."\textsuperscript{11}

A recent exception has been the work of Margaret
Slater, who has analyzed questionnaires administered to
users of industrial, government, university, and learned
society libraries in Great Britain.\textsuperscript{12} She classified her
data by user employment and type of use of libraries.
Among her findings was that most requests for information
were for simple factual data or for a description of an
object, process, or method.

The examinations of public libraries were, for the
most part, conducted from within. The recent concern with
scientists and how they use information has led to broader-
based analyses of "information-seeking behavior." These
studies include the library as only one of many possible
sources of knowledge for a specialized group of people.\textsuperscript{13}

It is difficult to draw any conclusions from these in-
vestigations. One problem is pointed out in the DOD User
Needs Study, which found that the studies generally
"suffered from the weakness of being based on
the opinion of the user. The problem is that

\begin{footnotes}
\item[11] Louis Shores, "Measure of Reference," \textit{Southeastern
\item[12] Margaret Slater, \textit{Technical Libraries: Users and
Their Demands, a Classification of User Groups and User
\item[13] Several studies of this type are listed in the bib-
liography. The classic example is perhaps D B. Hertz
and A. H. Rubenstein, \textit{Team Research} (Boston, 1953).
\end{footnotes}
the user may not have an adequate basis for rendering a reliable opinion as to the types or characteristics of information services since he may not have had any direct experience with them.

Another difficulty is that of inconstant variables. Groups of users of various numbers and subject specialties and in different settings have been asked a wide variety of questions. The resulting answers may be helpful in the particular organization studied, but one cannot easily extrapolate from them to the total universe of serious information-seekers.

The findings on which most investigators - social scientists, scientists, and librarians - agree\textsuperscript{15} are that:

1. Pure and applied scientists, technicians, management personnel, and other non-technical personnel have different aims in seeking information.

2. Pure scientists are likely to seek information within their own field of endeavor; applied


\textsuperscript{15}These points are drawn from the studies of Herner and Menzel in the United States, and those of Cole; Mote and Angel; Sabel, Terry, and Moss; Martyn; and Slater in Great Britain. A discussion and evaluation of the results of all of these investigations may be found in C.S. Hanson, "Research on User's Needs; Where Is It Getting Us?", \textit{Aslib Proceedings}, 16 (1964), pp. 64-74.
scientists and non-technical personnel need
information from peripheral fields.\textsuperscript{16}

3. Pure scientists are less likely than applied
scientists to ask for help from a librarian.

4. Scientists are more likely to use libraries
to perform exhaustive searches than to keep
abreast of current developments in their fields.

5. Scientists use informal communication net-
works to get information, often more than they
use libraries.

6. Journal articles are the favored printed
sources of information.

7. Library users will more often ask for speci-
fic material by its author or title than by its
subject.

8. When a reference question is asked, the query
is usually either for a single fact or for a
description of a process, method or technique.

Most of these tentative conclusions could have been
reached by a little intelligent observation and hard

\textsuperscript{16} Relevant to points 1 and 2 is Derek J. De Solla
Price's article, "Is Technology Historically Independent
of Science? A Study in Statistical Historiography,"
Technology and Culture, 6 (1965), pp. 553-568. One of his
conclusions is that the "cumulations of the literatures of
science and technology serve separate ends."
thinking on the part of someone working in a research library. Unfortunately, this was not formally done by anyone, and use studies have proliferated. Mortimer Taube, use studies' most devastating critic, asks:

"Why do use studies continue to accumulate in the face of a general recognition that they have so far contributed little to the purpose for which they are primarily intended? The answer to this question is to be found in the fact that use studies have been criticized in terms of their methodology....So much of social activity seems amenable to polls, surveys, questionnaires, etc., that it seems reasonable to apply similar techniques to the problems of scientific information." 17

The reason for use studies' lack of success, says Taube, is that professional services, unlike consumer services, cannot be measured in terms of consumer response, but possess "criteria of evaluation which are independent of consumer response." 18 He concludes that "use studies have no value as direct guides to the design of information systems, any more than consumer acceptance or rejection is a guide to the Salk vaccine." 19

Most critics, such as Taube, who have attacked use studies have offered no positive substitutes for them.

18 Ibid., p. 8.
19 Ibid., p. 11.
One who holds out some hope is R. D. Rogers who, although he calls statistics "an occupational hazard," believes that "precise, detailed, and expert analysis of small samples of circulation and reference work might produce valid conclusions of significance."20

---

Rogers feels that the proper consumers of significant conclusions about reference work are appropriating bodies and patrons. The hope for eventual question-answering via man-machine interaction, with the concurrent need to translate questions into linguistic forms computers can handle, is another pragmatic justification for the study of questions and answers. Furthermore, the question-and-answer process is interesting enough to provide a rationale for such a study for its own sake.

The remainder of this paper is a report of an examination of a set of actual reference queries in light of several theories about questions and answers. The subjects of this analysis are 272 inquiries received at the reference desk of the Lehigh University Library during a two-year period, from November 1963 through October 1965.

The Lehigh University library serves a faculty of 400 and a student body of 4500, as well as many visitors from industry and other academic institutions. While the emphasis of the curriculum has traditionally been on science and engineering, the university has Colleges of Arts and Science and Business, a graduate school, and a large graduate program in education, so that the library is used for reference and research in many fields.
The inquiries studied were of two types: (1) those considered interesting or difficult by experienced reference librarians, and (2) those recorded by inexperienced librarians, who were instructed to note all questions they were asked. While the 272 questions are only a fraction of the total asked during the two years, it was felt that they were sufficiently representative to permit fruitful study. Since some of the questions were recorded because they were unusual, while others are run-of-the-mill, the collection should be varied enough to include most types of inquiries. Directional questions and requests for specific library holdings or for instruction in the use of library materials are not considered here.

It should be noted that the questions studied are in their final form, as agreed upon by the inquirer and the librarian. The "negotiations" leading up to this agreement are themselves deserving of considerable study. Seldom is a question as it is first asked the real question to which the inquirer wants an answer. Every reference librarian is familiar with such queries as "Have you anything on...?" or "Will you give me all of the information about...?" which turn out, after extensive interviewing, to be expressions of needs for very specific pieces of information. The patron who asks for "books on science" when he wants to know how to dissect a frog is typical of library users who
believe that the only way to get anything from a library
is in large, general packages. Probably among the reasons
for this behavior are libraries' lack of effective advertising
of their services, and previous poor service the
questioner has received.

In addition to narrowing the field of inquiry as far
as the questioner will allow, the librarian must ascertain
such facts as the educational level and interest of the in-
quirer; the purpose of his question, if this will be help-
ful in answering it; and the format in which he expects or
prefers the answer to be presented.

After the final form of the question is decided upon,
still further refinement may be required. For example, a
patron might request material on the diplomatic history of
the United States. This inquiry is "negotiated" until the
librarian learns that the question is "Has there ever been
an ambassador or envoy from the U. S. to the Vatican?" Al-
though this seems to be a perfectly sensible query, the
simple yes-no alternative presented has a further implica-
tion: if yes, "Who was it?" It is necessary therefore to
follow the line of thought suggested by each question to its
logical conclusion.

Analysis of the Question

The most obvious, and most often attempted, method of
examining the question-answer process is the study of the
question itself. A question has been defined as a reveal-
tion of an inadequacy in the inquirer which must be met by a readiness to help in the inquirer. This definition will hardly help a reference librarian, who is presumably already motivated to answer questions.

The traditional approach to studying questions has been to classify them and, as previously mentioned, the classification has usually been by subject. The rationale for this is that a knowledge of subjects commonly asked for is helpful in building collections and in training staff. A more theoretical justification is Hiz's claim that "knowledge can be classified according to what questions it answers." This may not have been what Dewey and other classificationists had in mind when they devised their schemes, but if all the questions that have been and could be asked in libraries were classified by their subjects, the resulting array would indeed be a classification of knowledge.

In real life, however, there is always the unexpected, for the human mind, especially in the aggregate, ranges widely in its interests. The queries recorded at the Lehigh University Library, for example, included such diverse

---


items as requests for a criticism of *Walden* and for a list of the standard available sizes of cold rolled steel rods. Even in a more specialized library, the range of subjects cannot be predicted. Herner found in a study of questions received at atomic energy research organizations that 18% of the inquiries concerned non-technical subjects, and this finding has been borne out by others.

Another approach is to classify a question by the number of concepts it contains, and their relationships to one another. This, however, may be unnecessary for practical purposes for, as Herner, one of the foremost students of information-seeking behavior and of retrieval systems, concludes,

"Questions produced by the groups investigated [workers in a pharmaceutical laboratory] were considerably less complicated than the questions that a majority of retrieval systems are designed to handle. They generally covered fewer concepts, and required far simpler logical manipulations than would seem likely from most discussions of storage and retrieval systems in the documentation literature."  

Although this may be a result of the inquirers' previous experience with information retrieval systems, it may, on the other hand, reflect the way in which questions are

---


24 Hanson, op. cit., p. 67.

asked. Whichever is the case, the method is undesirable because of the difficulty of determining for a given question what the concepts are, and their relationship to one another. While it is probably possible to analyze such a question as "How many paroled prisoners imprisoned for murder have committed murder during parole?" in terms of its concepts, it is a somewhat time-consuming task to perform this operation on many questions.

Cole's revision of Alexander's classification of types of questions is perhaps more useful. One of its advantages is that it can be adapted to specific situations, as was done by Petrella in a study of inquiries in a business library. He changed several of the categories under "fact type" so that they fit the questions most often asked. Difficulties arise, however, in assigning questions to their proper categories. For example, the "supporting evidence type" may overlap any of the "fact types", because the inquirer may be using the facts as supporting evidence. The "how to do type" may turn out to be of the "all about type", while the "duplication of previous work type" could include any of the other categories. A requirement of any question

26. All questions discussed in Parts II and III are actual inquiries received at Lehigh, with the exception of the George Washington example.

classification is that it may be applied without agonizing decision-making.

Another description of questions has been done by Beinap, who discusses them in terms of the alternatives they present. He classifies questions as "whether questions" (having a finite set of alternatives), "which questions" (having a condition to be met as well as a set of names or terms acceptable as answers), and "why and how" questions" (ambiguous alternatives). 28

The "whether" type could include all queries of the "do you have" variety, which, although not discussed as such in this report, form a large proportion of any library's requests. Other "yes or no" questions when analyzed, however, usually resolve themselves into "which" questions. "Was Beethoven born on December 16?" is really asking "When (on which date) was Beethoven born?" The "which" category essentially includes "whether", "who", "when", "what", and "where," leaving out only "why" and "how". The "why" and "how" class is composed of demands for descriptions and how-to-do-it material, and all generalized requests for information about a subject. Thus boiled down to only two categories, such a classification is of little use.

Answering the Question

Some hint of the complexity of a question may be gained from a study of the steps required to find the answer. Such an investigation is included in Goldhor's examination of reference service at the Evansville, Indiana, Public Library.29 A fact question which can be answered from the World Almanac might be considered simpler than the inquiry which involves searching a periodical index and then proceeding to the periodical itself. The latter process, however, is quite straightforward, while answering a simple fact question in one operation may require considerable skill. Furthermore, an investigation of search strategies is not really an investigation of questions and answers.

Several studies have been conducted to learn the sources in which answers to reference questions are found,30 and one such investigation was carried out briefly at Lehigh. The method was found to have one advantage, that of ease of use. An indication of specific sources requires no decisions about categories, and is easy to perform retrospectively. However, such a list as:

---


30 The analyses of P. F. Cole and Van Hoesen show the types of sources. Lists of specific sources, arranged by frequency of use, are given by Chait, Chase, D. E. Cole, Ludewig, and Petrella.
grows "cold", and nothing of the nature of the question or its answer is remembered. Only the list of most commonly used materials remains, and this list is probably largely a function of the reach of the librarian. As in the case of the amount of time taken to find the answer, the source used depends on the skill of the librarian and the scope of the collection as much as on the question or the answer.

The Inquirer's Need

Some theoreticians believe that the best way of categorizing use of library materials is by the need of the inquirer, or his purpose in asking a question. Voigt groups these purposes into three approaches - the current, the everyday, and the exhaustive. He believes that the approach involved is reflected, not only in the inquirer's question, but also in the materials which will best answer that question.

"Printed information used in the current approach normally appears in only one location. Information wanted in the everyday approach usually can be found in a number of sources. Where data is concerned, the source easiest to find will generally be used, while for explanations or methods, other considerations - the clearest explanations or the best method - also enter in..."

---

The material ultimately used in the exhaustive approach is most often the original research article.\textsuperscript{32}

Menzel has reclassified the approaches discussed by Voigt into five: current awareness, reference, exhaustive search, brushing up, and stimulation to alter the boundaries of one's area of attention.\textsuperscript{33} He says that "the definition of information...must be in terms of characteristics of the messages transmitted or in terms of the relation of these messages to the receiving scientist's information-seeking behavior."\textsuperscript{34}

Since in real life it is often not feasible, or even polite, to ask the inquirer why he needs the answer to his question, unless a knowledge of how he will use the information will be of definite help in answering the question, practical investigations must be in terms of message characteristics. A reversal of Voigt's theory about approach determining source may be in order. Perhaps a knowledge of what answered the inquirer's question will reveal something of his approach.

\textsuperscript{32}Voigt, op. cit., pp. 185-186.


\textsuperscript{34}Ibid., p. 8.
Analysis of the Answer

The most promising method of discussing the question-answer process, therefore, may be in terms of the answer. An examination of the answer should be just as useful, if not more so, than an examination of the question. Knowing the question may involve the problem of the inquirer's not stating his need precisely, or even of his not knowing his need at the time he asked his question. On the other hand, if the answer fulfills the conditions finally set by the inquirer, it is in some way a statement of his need. It must be stressed that the inquirer must be satisfied, at least consciously, by the answer he receives, in order for a study of answers to have any meaning.

According to Hamblin, "knowing what counts as an answer is equivalent to knowing the question."\(^{35}\) This is not strictly true, for "George Washington was the first president of the United States," may answer more than one question. The question may have been, "Who was George Washington?"; "Which president of the United States was George Washington?"; "What office was George Washington the first to hold in the United States government?"; or "Of what country was George Washington the first president?" However, "knowing what counts as an answer" is equivalent to

knowing what satisfies the questioner, and if he receives the answer he needed, his question is no longer important.

It is essential that all of the alternatives presented by the question be matched by the answer. If the query is "Was George Washington the first president of the United States?" then the answer "yes" is sufficient, for the alternatives are simply "yes" or "no". If the query is "Who was the first president of the United States?" the answer "George Washington" satisfies the condition "first president of the United States," and is the only member of the set of names acceptable as answers. In the case of "Who were some presidents of the United States?", the condition changes to "presidents of the United States" and the set of acceptable names increases to 35.

The size of the set of acceptable terms in the answer is referred to by MacKay as the "logon-content" or "logical dimensionality" of the question. The logon-content of "Who was the first president of the United States?" is one; that of "Who were some presidents of the United States?" is the total of all combinations of any two or more of the 35 names. Answers of a logon-content of one are usually specific facts or figures. Answers of larger logon-contents may be lists, as in the case of the presidents, or descriptive

---

36 MacKay, op. cit., p. 474.
material which may be worded in a number of different ways.

The obvious procedure in studying the question-answer process from the standpoint of the answer would be to state the answer as given. This quickly becomes unsatisfactory. Although the statement, "Hitler became Chancellor of Germany in 1933", answers the question, "When did Hitler become Chancellor of Germany?", recording the answer to a request for proofreader's marks is hardly practical.

David Kronick, in an article describing reference requests in a medical library, classifies answers according to whether they are "direct" or "indirect." "Direct answers" are those in which the librarian responds, "It says here..."; "indirect answers" are those in which the response is "Read this and find out." Because of the frequency of biographical and directory inquiries and requests for verification of references, Kronick makes subdivisions for them in his "direct answer" category. With only two (or four, when "direct answers" are subdivided) categories, and these categories clearly defined, this is perhaps the simplest classification of all. Because of the findings of

---

such investigators as Hanson\textsuperscript{38} and Herner\textsuperscript{39} that pure scientists in particular prefer to do their own searching, this type of division is valid in discussions of information-seeking behavior. The drawback is that little is revealed about the answer, other than whether the librarian or the questioner read the material containing it.

Kronick's classification may be expanded, or adapted, to one in which the answer is categorized by its type, or format. Mary Francillon, in an examination of a year's reference questions in a research library, asked users about the forms in which they preferred their answers. Lists of titles were favored over copies of primary source material by 80\% of the respondents. There was also a strong preference for abstracts.\textsuperscript{40}

Probably most information-seekers as they enter a library have some unconscious expectation of the package in which the information will be presented. They may have in mind a paragraph of descriptive text accompanied by a diagram, an uttered word (\textit{is the answer}), or even "that big red book I used last time." The format of the response to a request will become more important as more strides are

\textsuperscript{38}Hanson, \textit{op. cit.}, p. 72.


taken toward making man-machine interaction a reality in question-answering systems. J. C. R. Licklider, in discussing precognitive systems in Libraries of the Future, writes:

"we need to substitute for the book a device that will make it easy to transmit information without transporting material, and that will not only present information to people but also process it for them, following procedures they specify, apply, monitor, and, if necessary, revise and reapply."\(^4\)

He mentions, as does the description of Project Intrex,\(^4\) optical screens, cathode-ray tube displays, light pens, keyboards and hard-copy displays, and telephone communication as means by which this may be accomplished\(^4\).

A knowledge of the most common ways in which answers are currently presented, and accepted, especially in scientific libraries, will be helpful in designing question-answering systems, as well as in analyzing present practices.

A possible classification of answers by their formats is given in Part III.


The classification proposed here includes five types: exact reproduction type, fill-in-the-blank type, descriptive type, information-about type, and list-of-references type. (An outline of these types and their subdivisions, with examples, follows an analysis of the five categories.)

The exact reproduction type includes both pictorial and verbal materials, which must be seen by the inquirer. Pictorial materials are pictures, maps, diagrams, and charts. Verbal materials in this category are copies of historical documents, laws, speeches, poems, musical scores, etc. In cases where this form of answer is provided, the questioner needs the verbatim text. Often these materials, since they are not listed as individual documents in whatever passes for a card catalog, require long and arduous searches, using indexes and hunches, that take them out of the simple "do you have this book" category.

The fill-in-the-blank type is comprised of three categories: (1) short answers, in which names, terms, or figures are presented orally or in writing; (2) lists of names or terms; and (3) tables of statistics, scientific numerical data, or other figures. This type is characterized by the fact that there is only one term, or one set of terms, which meets the condition or conditions imposed.
by the question. The question has generally been of Belnap's "which" type, including "who", "when", "where", and sometimes "what". Specific citations or verifications of references are in this category, since in these cases only one answer is acceptable. Where lists or tables are given as the answer, the whole list may have been needed, or only one item of data. (E.g., the patron who received a list of the Marines who were in the picture of the flag-raising at Iwo Jima may have wanted all of the names, or he may have been interested only in the name of Ira Hayes. In either case, the list was sufficient answer.)

The descriptive type is here divided for convenience by length. "Short" answers contain a sentence to a paragraph; "long" ones run up to a few pages. In any event, the material is read by both librarian and inquirer, who agree that it will answer the question. The words of a descriptive answer are only one of many possible collections of words which comprise the ambiguous alternatives, theoretically infinite in number, presented by the question. A definition, for example, may be worded in many different ways, each of which is correct. The question has been of the "who" or "how" type, and occasionally "what", as in "what is".

The information-about type answers the familiar "information on a subject" questions, or "subject requests." The inquirer is presented with one or more documents and
told that here he will find what he wants to know about his subject.

The list-of-references type is another form of answer to the same question, but the seeking out of the "information about" is left to the inquirer.

It can be seen that these five types form a continuum of precision. At one end is the "exact reproduction" type, where the librarian knows that what is presented is just what the inquirer asked for, and this is verified by the inquirer. At the other end is the "list-of-references type", where the questioner leaves with a bibliography or a scribbled set of call numbers in his hand, and the librarian is not at all sure that the information provided will in fact lead the inquirer to what he seeks.

The first two types ("exact reproduction" and "fill-in-the blank") are likely to correspond most closely to whatever package the inquirer had in mind before beginning his search. They also have smaller logon-contents, for there are only a given number of pictures of the Trojan horse, only one Gettysburg Address, and only a few terms or sets of terms which will fill a given blank. The other three types, being less specific, have high logon-contents, and may not be at all the type of package the questioner was expecting.
An outline of the classification scheme is given below, with examples from the Lehigh collection of questions and answers.

A. Exact reproduction type

1. Pictorial
   
a. Pictures
   Example: A picture of the Phi Beta Kappa key.

   b. Maps
   Example: A map of Allentown, Pa. (Question: location of a specific street).

   c. Diagrams
   Example: A flow chart. (Question: I need to draw a flow chart and don't know what they look like.) It is true that an article on flowcharting might also have been useful, but in this case the diagram was sufficient.

2. Verbal
   Example: A copy of the Congressional resolution making Winston Churchill an honorary citizen of the U.S.

   Example: A copy of "Paul Revere's Ride."

B. Fill-in-the-blank type

1. Short answer
   Example: Francis Thompson (Question: Who
wrote "The Hound of Heaven?")
Example: 34 Beacon Street, Boston 6
(Question: What is the address of Little, Brown & Co?)

2. List
   Example: A list of the members of the United States Congress from Pennsylvania in 1811-12.

3. Tabulations
   Example: Periodic table (Question: What is the atomic weight of mercury?)

C. Descriptive type
   1. Short
      Example: "A monopsonist is one who is a single buyer for a product or service of many sellers."

   2. Long
      Example: Two and a half pages about George Bernard Shaw in Twentieth Century Authors.
      Example: A page in The Chemical Engineer's Handbook (Question: How does a pitot tube work?)

D. Information-about type
   Example: Several books and a lengthy article in the Catholic Encyclopedia on the Roman Catholic mass.
E. List-of-references type

Example: Pages in several issues of the Reader's Guide listing articles on Afghanistan.

Hanson's remarks about information-seeking behavior have some relevance to this outline. In discussing all "acts of library use," he asserts that:

"Something approaching one fifth of the demand is for a figure, or a single, simple fact... between a quarter and a third of the total demand is for a description of an object, a process, a method or procedure." 44

The "fill-in-the-blank type" answer is analogous to Hanson's "demand for a figure or single simple fact;" the "descriptive" and "information-about" types together to his "demand for a description of an object, a process, a method or procedure."

A tabulation of the five types of answers as they were presented at Lehigh is shown in Figure 1.

The figures in the table are presented to show that each category contains enough items to give it some validity, although the percentages may not present an accurate picture. As was stated, 272 inquiries are a fraction of the total received. While it was felt that all types of questions and answers were included in the collection, it is likely that the proportion of each type to the total of the

44Hanson, op. cit., p. 66.
<table>
<thead>
<tr>
<th>Type of Answer</th>
<th>Number</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact reproduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures</td>
<td>8</td>
<td>2.9</td>
</tr>
<tr>
<td>Maps</td>
<td>7</td>
<td>2.8</td>
</tr>
<tr>
<td>Diagrams</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>6.2</td>
</tr>
<tr>
<td>Verbal</td>
<td>16</td>
<td>5.9</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>12.1</td>
</tr>
<tr>
<td>Fill-in-the-blank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short answer</td>
<td>101</td>
<td>37.9</td>
</tr>
<tr>
<td>List</td>
<td>25</td>
<td>9.2</td>
</tr>
<tr>
<td>Tabulation</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
<td>47.8</td>
</tr>
<tr>
<td>Descriptive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>25</td>
<td>9.2</td>
</tr>
<tr>
<td>Long</td>
<td>28</td>
<td>10.3</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>19.5</td>
</tr>
<tr>
<td>Information about</td>
<td>42</td>
<td>15.5</td>
</tr>
<tr>
<td>List-of-references</td>
<td>14</td>
<td>5.1</td>
</tr>
<tr>
<td>Total</td>
<td>272</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**Figure 1**

collection is not the same proportion of that type to the total number of questions asked.

The large number of short, specific answers is probably a reflection of the nature of the collection. Users of an academic library, because they are accustomed to performing their own searches for general information about a subject, are inclined to approach librarians only for specific data. Also, librarians are more likely to record a specific fact that was difficult to find than an unusual subject request. Probably the proportion of
"information-about" and "list-of-references" answers was much higher than is shown here. In light of Francillon's findings about the number of users who needed primary source material, the figures for "exact reproduction type" are probably fairly realistic. Some specialized libraries may have more requests, and answers, involving pictorial material. An art library is an obvious example, or a library serving chemists who need to see models of molecular structures.

Francillon, loc. cit.
PART IV
RECOMMENDATIONS AND IMPLICATIONS

No claim can be made that the analysis of the question-answer process just described is better than other methods, without further testing of it. The collection of questions and answers examined was a small one, and much of the data were recorded before the classification system was developed. Because the boundaries of each class are sufficiently clear to require a minimum of difficult decision-making, testing should be relatively easy. A form such as the one shown in Figure 2 is all that would be required.

<table>
<thead>
<tr>
<th>Check form of answer supplied</th>
<th>Date ______</th>
</tr>
</thead>
<tbody>
<tr>
<td>___  Pictorial</td>
<td>___ Short description</td>
</tr>
<tr>
<td>___  Map</td>
<td>___ Long description</td>
</tr>
<tr>
<td>___  Diagram</td>
<td>___ Information about</td>
</tr>
<tr>
<td>___  Fact</td>
<td>___ List of references</td>
</tr>
<tr>
<td>___  List</td>
<td>___ Table</td>
</tr>
</tbody>
</table>

Was the inquirer satisfied?

| ___ Yes  | Partially    | ___ No  | ___ Unknown |

Figure 2

As can be seen from this suggested form, information about the answer could be combined with other data, such as user satisfaction. If desired, space could be provided
for a statement of the question itself, type of user, the name of the source used, reasons for user dissatisfaction, etc.

There is some justification for carrying out such a test. As well as being simply performed, its results are reasonably plain and easily quantified, and can be cross-tabulated with other variables.

If Voigt's assertion that the user's approach is reflected in the materials that answer his question is accepted, an analysis of these materials, combined with data about the user, may be of assistance to those who wish to study information-seeking behavior. Investigating reference service in this manner is more objective than administering questionnaires to ascertain how many journals a person reads or which colleagues he asks for help.

Knowledge of format preferences should be of particular interest in the design of and experimentation with question-answering systems involving man-machine interaction. Lists of references and exact reproduction-type material, if predictable, such as molecular structures or maps, are formats which can be stored in memory and displayed on optical screens or cathode-ray tube devices, as described in Libraries of the Future and the Project Intrex report. Interaction via alphanumeric keyboards in conjunction with hard-copy devices can produce from storage information in the form of lists of references, predictable
short answers, and lists. Telephone connections between user and machine may ultimately make feasible the communication of short fill-in-the-blank type or descriptive type information that is at present often transmitted by voice. An analysis of desired configurations of answers combined with a study of system capabilities should aid in designing the best possible system for a given situation.

If a library still insists on classifying questions by type or subject, such a classification can be related to a classification of answer formats. There could be some interesting implications, for example, in a finding that people who ask questions about coal mining usually want diagrams, or that astrophysicists prefer short, specific answers. Results could lead to new arrangements of reference collections, or the development of different search strategies for different clienteles or questions.

The happiest arrangement, of course, would be for the librarian to learn at the outset of a search how the inquirer wants his answer delivered. Having, for instance, the hypothetical knowledge that people interested in coal mining usually want diagrams, the librarian could ask, "Do you want this in diagrammatic or in some other form?" Or, after a librarian has established what the question is and has learned something, either by observation or by asking, about the inquirer, knowing that the answer is likely to be best presented and best received as... (in the hypothetical
a diagram will induce the librarian to look first in sources that provide diagrams.

Search strategies approached from this point of view lend themselves to programmed instruction. To the knowledge of this writer, only one programmed text has been developed for library use, an instruction in use of the card catalog at West Virginia University. A programmed lesson about search strategies and desired answers might look somewhat as shown in Figures 3-5.

Choose one of the following subjects you would like to know more about and turn to the page indicated.

The year Hitler became Chancellor of Germany. Turn to page 4.

Atomic weights of some of the elements. Turn to page 97.

What the moon looks like. Turn to page 32.

The names of George Bernard Shaw's plays. Turn to page 45.

How a pitot tube works. Turn to page 81.

The student decides he wants to know how a pitot tube works and turns to page 81.

In what form do you prefer the answer?

List of references. Turn to page 43.

All the information about. Turn to page 101.

Diagram. Turn to page 76.

Picture. Turn to page 68.

Short definition. Turn to page 52.

One-or-two-page description. Turn to page 60.

The student selects a one-or-two-page description and turns to page 60.

The sort of information you are looking for may be most conveniently found in a handbook or technical encyclopedia. For a list of chemical handbooks in the library, search the card catalog under the heading:

"Chemistry - Tables, etc."

Some suggested handbooks in the reference collection are the following:

Lange. Handbook of Chemistry.


If the student had desired information about other subjects, or in other formats, he would have found similar discussions and suggested reference books on the other pages.

A programmed textbook of this type could be used, as is the West Virginia manual, in assisting college and university students in learning about the library. It might also have some value in training reference personnel in small libraries whose employees have not had formal education in library science.

The classification of answers described in this thesis is but one small step toward the development of new techniques for analyzing and evaluating reference services in libraries. Other methods, and refinements and emendations of this one, are doubtless needed. The system proposed here is put forward because of its simplicity and relative straightforwardness, the ease with which it may be used in combination with other data, and its applicability in "libraries of the future" as well as libraries of the present.
BIBLIOGRAPHY


Caroline Hieber was born in Sayre, Pennsylvania, on November 27, 1929, the daughter of John T. and Jean Urquhart Hieber. She attended elementary and high school in Towanda, Pennsylvania, and graduated from the Towanda High School in 1946. She received her A.B. degree from Miami University, Oxford, Ohio, in 1950, having been elected to Phi Beta Kappa in 1949. After a few years as an editorial assistant in a publishing house, she returned to school for a B.S. in library science, received in 1954 from the University of North Carolina. She worked in several libraries until September 1964, when she enrolled at Lehigh University as a candidate for the degree of Master of Science in the information sciences.