
This study describes an analysis of content of the old and the current online literature search request forms used at the Library of the United States Environmental Protection Agency in Research Triangle Park. An analysis of terms from six years worth of online literature search request forms was undertaken (Study 1), and these results were compared to an analysis of terms from two months worth of searches that were submitted on a revised form designed to elicit more and better information from the patron (Study 2). While searches that used the revised form did not contain a greater overlap in terms provided by the patron with terms used in the final search strategy than searches that used the original form, the revised form did elicit more information from the patron at the onset of the search request. These results suggest the importance of input from both the patron and the librarian in the literature search process.

Headings:
Information Retrieval
Electronic Information Resources
Query Formulation
Search Term Selection
Elicitation of Information Needs
Patron Term Selection in Online Literature Searches at the EPA

by
Jillian E. Tanner

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina
April, 2006

Approved by:
Table of Contents

Introduction ......................................................................................................................... 2

Literature Review ................................................................................................................ 9

Methodology ...................................................................................................................... 27

Results ............................................................................................................................... 41

Discussion .......................................................................................................................... 52

Conclusions ....................................................................................................................... 64

References ......................................................................................................................... 66
Introduction

Information retrieval is an area of study in library and information science that addresses how people access information and how to improve the systems and mechanisms in place to aid people in their quest for information. It is also a fundamental process that is employed pervasively in all types of libraries, in all types of businesses, and in all situations in which information is sought. Information retrieval can be as simple as a person asking another person for directions or as complicated as a searcher using a series of complex databases and terms to perform a patent search, and it runs the gamut of everything in between. Some of the questions that this area of study asks pertains to what information people are attempting to get access to, what type or format they want it in, why they need or desire this information, why they need this information at this particular point in time, how they are going to find this information, and what they want to do with the information once they get access to it.

In the present day information landscape in which there are a myriad of channels through which to obtain information, such as libraries, schools and universities, colleagues, journals, newspapers, television, radio, the Internet, podcasting, and weblogs, people have a great amount of autonomy and choice in selecting how and when they retrieve information. The variety of sources and the emergence of new ways of accessing information are certainly interesting and exciting for library and information science professionals to learn to use and also to study. In studying these media and the methods by which people make use of them, information professionals learn new skills and
subsequently new ways of helping people to access these media in better and more efficient ways. Not only have librarians learned how to use these new media to help their patrons, they have been able to assist patrons who desire to learn more about these resources and how to use them effectively on their own. Especially in a world where information overload and information quality are perpetual concerns of those who provide and also those who access information, librarians are particularly adept at making sense of, and adding value to a fairly chaotic and overwhelming mass of resources and information.

Libraries and the librarians who run them are arguably still one of the best institutions for people to turn to when they need information of any kind. One function of a library is as a storehouse of information and a portal through which a plethora of information sources can be accessed, such as books, journals, newspapers, and documents, as well as maps, local information sources, and reference books and materials. Other sources that the library can act as a gateway to are the vast collection of online databases, directories, dictionaries, newspapers, merchants, and every other type of resource the imagination can think of that are available on the Internet. Some of the preceding resources are free and freely available to anyone who has access to them, while many others are subscribed to and paid for by the library.

One of the aforementioned resources that has been the subject of much research with respect to information retrieval has been the online database, a plethora of which are available today. Research has been devoted to online databases because of many factors: their depth of coverage, their quality of information, their cost, their variability in subject matter and the types of content they provide, their different interfaces and searching
functionality, their use by end-users, and their use by librarians to help end-users retrieve the information they need. Some of these online databases are useful for statistical purposes or medical purposes, legal purposes, and still others are useful for retrieving articles on many other different topics. Articles, diagrams, and statistics can be accessed in different ways according to the design of the database, via drop-down boxes, drilling down from a general topic to a more specific topic through navigating menus, searching functions using terms that a person enters, or a combination of several techniques.

One of the widely used methods for accessing information in an online database is the selecting and entering of search terms, which specify the subject of the information need, the scope of the information being sought, and the type or format of information that the requestor prefers, among other attributes. When prompted, a user must choose some words that accurately represent what the user would like to learn about. These words comprise the search syntax: in other words, these terms and phrases will be searched for in the database records. The system may or may not offer any help to the user in choosing the terms or hinting about the correct format that the terms should take when the user enters them into the search box or the fields provided. The user may not know how the database will use the terms that he or she inputs and consequently may not understand why certain results were retrieved and why certain results were not retrieved. This latter case may be a consideration only if the user is savvy enough to know that there may be other articles or resources that exist that were not returned by the database by his or her initial query. While this entire process of conceptualizing, selecting, and articulating words to represent a specific information need to the online database can be, and often is, performed by the user without much reflection, the process of user search
term selection has serious implications on the definition of the information need itself. The terms selected and used also have a significant impact on the retrieval actions of the online database system, as well as considerable bearing on the relevance and usefulness of the results to the user.

Since the process of good, productive term selection and use in certain online databases is often quite complicated and requires much specialized knowledge about database structure, indexing practices, and retrieval techniques, in many instances a librarian, often called a search intermediary or a searcher, may be consulted. This is the case in many special libraries, which cater to a specific clientele, often at the corporate or government level. The search intermediaries gather information from the patron who has an information need, and the search intermediary plans a strategy based on input from the patron and runs or performs the search in the online database. This searching procedure has many important components which are all vital to producing useable and meaningful results for the patron or user requesting the search.

**Purpose**

In order to better understand the processes of search term selection in a specialized library setting, an analysis of online literature search request forms will be performed. These online literature search request forms are used to document the initial information request submitted by the patron to the searching staff at the Library.

Research to answer the following question will be undertaken: How much overlap in search terms is there between a patron’s written initial information request and the final search strategy run by librarians at the Library of the United States Environmental
Protection Agency in Research Triangle Park, and does the revision of the online literature search request form result in a greater overlap?

The purpose of this research is to study the process of search term selection as it has occurred in the past, and how it occurs currently at this particular library in the effort to identify any trends present in the number and types of words patrons supply to the librarians, and if and how these words are used in the final search strategy. Also under study is if the introduction and use of the revised online literature search request form has had an impact on these same factors. The revised form aims to gather information in a way similar to that of the traditional reference interview. If patrons begin the process of refining their search topics at the very inception of the literature search process, as prompted by the questions on the revised searching form, then the subsequent reference interview that occurs could gather more and better information early in the process.

From the data collected from these two studies, the librarians at the EPA Library will be able to understand quantifiably how many terms patrons provide at the onset of the search, and what the characteristics of those terms are. These terms can be compared with the terms that appear in the final search strategy, which will provide insight into the sources of the search terms: whether certain terms were present at the onset, or were added during the reference interview or at some other point in the search process. The difference between the initial information request and the final search strategy as listed on the forms point to the librarian’s role as an elicitor of information, as a researcher, as a strategy deviser, and as a strategy reviser. If, for instance, a certain type of term is lacking in many initial requests that is found in the final strategy that is deemed important, then perhaps there is a way to address this shortcoming. If there is a
difference between the terms found on the original form versus the revised form, then
that difference can be analyzed and taken as evidence of the worth of the revised form.
Gathering more or better information at the onset of the literature search from the patron
may have implications for the quality of the search. This work, however, will be useful
to the librarians at the EPA regardless of the outcome of the comparisons between the
original and revised online literature search request form, because information will be
gathered on how each form has been used to perform literature searches, which is an
important component of the library’s services.

This analysis will also highlight the various procedures the librarians and patrons
undertake as part of the online search process. The revised search form may capture on
the search form itself some of the currently performed processes, like the reference
interview for example, since the revised form was based on reference interview
techniques that give the librarian the exact type of information he or she needs to develop
a search strategy for use in complex databases. Patrons may be able to provide more
information at the onset of the search, and librarians may be able to perform a more in-
depth reference interview if the patron has already begun to formulate a more specific
search topic by answering questions on the form. The end result of this research is
intended be two-fold: one, to educate the librarians with concrete figures about what the
composition of the terms used in a search is, with all its accompanying data on numbers,
Sources, and characteristics for searches that used either the original or the revised online
literature search request form; and two, to help the patrons define and formulate more
specific search topics using the recently implemented revised form. The latter end result
may have implications for providing patrons with a better search experience overall, in
that patrons may receive more defined and precise search results from less interchanges with the searchers, who themselves may be able to save time and resources. The goal of this research is to elevate the level of service for the patrons whose research the librarians support.

This research, if it is published, may also have an impact on the profession of library and information science. The more information that is gathered about the processes by which users supply search terms and the ways in which searchers utilize this input along with information from other sources, the more the profession as a whole can understand how these processes occur and how to improve the experiences of both the librarians and the users involved.
Literature Review

This literature review will discuss the variables involved in this research in relation to work that has already been done. The variables include terms that are given at the inception of an online literature search request, the addition and reexamination of terms during the course of the search, and the terms found in the final search strategy. First, there will be a discussion of queries and some larger issues that pertain to information retrieval in general. More specific elements of this research will be discussed next, such as implications for the number of search terms found in a search strategy, as well as the source of search terms and particular searching styles of intermediaries. Other factors that influence search strategy building are the indexing of databases to be used and the inclusion of synonyms in the search strategy, which are discussed next. Finally, the literature forming the basis for the revised online literature search request form will be discussed.

Queries

There are many elements that factor into how an information need request is understood, how the query is built from this request by the search intermediary, and how the query performs in the particular database being used for the search. These topics must be discussed before a deeper investigation of the qualities of the terms used in these queries can be examined. In one of the seminal articles of information seeking and information retrieval, Belkin (1980) explains "the problem of the effective and efficient
transfer of desired information between human generator and human user,” or a patron and a search intermediary (p. 133). Belkin’s theory of a user’s anomalous state of knowledge, or ASK, is a way to conceptualize the reasons behind the user’s information seeking behavior, or what drives a person to find out information. The process of information seeking starts when “a user decides to investigate or use some part of his or her state of knowledge or image, but on consideration, the user realizes that there is an anomaly in that state of knowledge with respect to the problem faced” (p. 135). Once the user recognizes this anomalous state of knowledge within himself or herself, the user will try to remedy this “perceived inadequacy” in his or her understanding (p. 136). The ASK model attempts to address “the user’s initial difficulty in specifying or even explicitly recognizing what is wrong, and especially in recognizing and specifying what is necessary to make things better” and also the concept of the “non-specifiability of information need” (p. 136, p. 137). The user’s difficulty in understanding and articulating to another person what the user does not know, but wants to know, and how well formed in the user’s mind this gap in knowledge is all factor into how well an information need will be dealt with by the user and by the intermediary involved. Belkin’s research relates to the current study because it examines the internal state of the user and once an information need arises in the user, how the user attempts to understand it and communicate it to a librarian. The face-to-face reference interview and revised online literature search request form are designed to facilitate this communication.

It has been suggested that one of the ways in which the intermediary understands an information request from a patron and subsequently builds a search strategy to fulfill it is by the intermediary traveling through various “discourses” (Iivonen, 1998). Iivonen
states “that the selection of search terms is more than a translation process between the client’s search request and IR system-specific search terms,” asserting that intermediaries perform more functions in a search than only converting natural language terms into those the system can understand (p. 313). For instance, intermediaries navigate through the discourses of, or thinking about the information need in the context of, controlled vocabulary, indexing, and their own search experience (p. 322). Keeping these and other discourses in mind while constructing a strategy helps the search intermediary ensure that the search will incorporate a deeper and more thorough understanding of the task from these different but mutually reinforcing perspectives.

Once a search strategy has been built, however, a searcher can not be entirely sure of its effectiveness until the search has been executed in the necessary databases and a review of the retrieved information has been completed. Cronen-Townsend, Zhou, and Croft (2002) state that: “Even what a user believes to be well-formulated queries may, in fact, perform poorly depending on the nature of the collection” (p. 299). The terms that the user supplies to the system is the only data it has to try to make sense of what the user wants to retrieve, and if this language is not specific enough, understandable to the system, or does not match that of the documents that contain the answer, the query will perform poorly in retrieving exactly what the searcher wants. The authors have developed a “clarity score” in order to measure the “degree of ambiguity of a query with respect to the collection of documents being searched,” which “is often closely related to query performance” (p. 299).

In order to address these potential issues in query performance, Belkin et al. (1993) suggest that “progressive combination of query formulations leads to
progressively improving retrieval performance” (p. 339). The study these authors carried out showed that combining techniques or representations of queries, such as those generated by different searchers, and also those that were produced by different sources such as a searcher and a retrieval system, improved the retrieval performance of the query. This combination of “multiple techniques or representations” works in theory because “their combination will address more aspects of the situation, and thus retrieve more relevant documents” (p. 339). From this logic, it follows that the more representations, or the more input that different people have and the different techniques which are used in building queries, which includes combining different sources of information, the better the query will be in retrieving documents that are deemed relevant to the requestor.

**Number of Search Terms**

The number of terms involved in a search is influenced by a number of factors: the subject matter of the topic, the breadth and depth of the information needed, and any topic-related or people-related factors. Haas (1997) found in her examination of terms in different disciplines that the hard sciences had more domain-specific terms than social sciences and humanities (p. 78). These domain terms were also more likely to be found in patterns in the text in the hard sciences as well. This finding has implications for the number of terms involved in a search because the hard sciences may tend to have a larger number of specific terms, and therefore may not require as many terms to convey certain ideas or concepts in a search as opposed to a search based in another subject area.

The number of terms involved in a search strategy also relates to the input of the people involved in the search, namely, the patron or user, and the searcher or
intermediary. Spink (1995) found that using a method of query expansion that involves the selection of additional terms from retrieved records, intermediaries selected more terms that retrieved relevant results in subsequent searches than the user did (p. 167). What this means is that the intermediary’s expertise in recognizing good terms to add to the search strategy (also called query expansion) to make the results more relevant will tend to cause the number of search terms that are productive to be greater than if the user himself selected terms. This data supports the intermediary’s role in adding terms to the user’s initial information request because of his or her experience in selecting and adding terms to the strategy to improve the search results.

Source of Search Terms and Searching Style

The source of the search terms in this research question has been categorized into patron and librarian generated, as supported by the search processes in practice. The initial information request will most often consist of terms solely from the patron, and the final search strategy will contain a mix of terms from the patron, from the searcher, from the indexing of the database, from retrieved articles, from thesauri, as well as from other possible sources. The collaboration between the patron and the searcher can be productive in yielding terms, just like other methods the searcher follows can add to the terms utilized in a search. Spink and Saracevic (1997) found that users were responsible for producing 61% of the terms used in their analysis of 40 questions. They point out that 39% of the search terms came from sources other than the user, in addition to 23% of the terms arising from the interaction between the searcher and the user, for a total of 62% of the total search terms that the authors attribute to the interaction process (p. 748). The reference interview and the subsequent information exchange between these two parties
have a large impact on the makeup of the search strategy. Spink and Saracevic also found that the source of terms has bearing on their ability to contribute to retrieving relevant results. In all searches, at least one term from the user’s written question statement was used in the final strategy, and those terms written as the information request, called the question statement, were found to be the most productive category of terms in yielding relevant results (p. 750).

Another consideration related to the source of search terms is how consistently a librarian or searcher will choose terms to represent the query posed by the patron or user. This is an important issue to consider because the level of agreement that different searchers have with each other and that searchers have with themselves at different points in time can have an effect on the number and types of terms they select for inclusion in the search. An analysis of the consistency with which librarians select terms to represent these various information needs has been done by Livonen (1995) and that study has produced some intriguing results that have specific implications for this study.

Livonen (1995) focused on studying search term selection consistency between searchers (intersearcher consistency) and also between a searcher and the same searcher two months later (intrasearcher consistency) (p. 173). Under study were these two types of consistency in the searcher’s selection both of search terms and of search concepts. A search term is defined as “a string of characters that is bounded by Boolean operators, or that starts or ends a query statement,” while a search concept is “the subject of the search, as units of information taken into account in a query statement and described with one or more search terms” (p. 174). While Livonen used an elaborate set of rules to define search terms and search concepts and used these exact definitions in her study, the
current study used simplified versions of her definitions as guidelines on how to classify terms. Ivonen found that intersearcher search term selection consistency based on search terms was fairly low, with a mean value of 31.2% (p. 180). This finding is on par with other such studies comparing term by term how consistently searchers select and use the same terms for search strategies, such as Saracevic (1984). Saracevic states that there is “considerable difference” in search term selection, but he also found that the group of searchers “agreed 78% on the concepts to be included in the search statement” (p. 229).

Ivonen (1995) finds similar results, in that intersearcher consistency on the basis of search concepts was “rather high” with a mean of 87.6% (p. 181). Ivonen’s two intersearcher values were further broken down by type of search environment, and that of “special service-oriented search environment” is the closest to the environment in which this study is being done. The intersearcher consistency based on search terms and search concepts (defined on the previous page) for this type of environment is 38.8% and 86.8% respectively (p. 181, p. 183). What this finding means is that even though searchers did not agree exactly on the words or forms of words with high consistency, they did agree on the concepts that should be included in the search strategy. The intrasearcher consistency follows the same patterns in that consistency in this environment is close to if not higher than the average consistency by search term and by search concept (p. 184-186). These figures show evidence that special library environments and their searchers are consistent in the terms and concepts they select to meet their customer’s needs.

Along with consistency, there are other issues surrounding an intermediary’s selection of search terms. A good overview of many of these issues is given in a series of three articles by Fidel entitled “Searchers’ Selection of Search Keys” (1991a, 1991b,
In “The Selection Routine,” Fidel outlines many of the facets a searcher takes into account when understanding the information need, including the searcher’s having to keep in mind the “semantic and pragmatic” aspects of the search: the topic of the user’s need, and the practical “use to which the information will be put” (1991a, p. 490). The process by which an intermediary goes about providing a service to the patron is stated succinctly: “Each concept requires a set of search keys for its representation. Thus, the searcher looks for search keys that will best capture the literature on the topic of each individual concept or of the concepts in combination, and at the same time retrieve an answer set that satisfies other request characteristics, such as recall, precision, or timeliness” (1991a, p. 491). Fidel discusses choices searchers make in using common terms, descriptors, and textwords in their strategies, and reasons why these choices were made. She continues this discussion about term selection decisions in the next article: “Controlled Vocabulary or Free-Text Searching” (1991b). Fidel points out the benefits of constructing and using controlled vocabularies, saying that “despite the expense and difficulties in the construction” of them, “they are built and used because they improve retrieval” (1991b, p. 501). While controlled vocabulary terms are quite useful in information retrieval, the author states that “there is increasing evidence that textword and descriptor searching actually complement one another” (1991b, p. 501). Indeed, many search strategies used in information retrieval do incorporate a mixture of free text and controlled vocabulary terms in order to address the immediate information problem of the user.

Fidel reports some interesting findings to lend support to her assertions. She states that “Searchers as a group did not show a preference for one type of search key:
When they had a choice, they selected descriptors and textwords with the same frequency” (1991b, p. 505). This discovery reinforces the statement mentioned above regarding textword and descriptor usage, since the end goal of the searcher is to retrieve relevant documents for the patron, and the searcher demonstrates that by including different types of words equally, they are achieving this goal. However, external conditions, like the type and number of databases being searched, do influence the types of words searchers include in their search strategies: Fidel found that “Searches which require several databases, and searchers who habitually search several databases for a request, are likely to use more textwords than searches or searchers using a single database” because of the differences in indexing between databases (1991b, p. 509). Also reported is the finding that “When searchers had options in the selection of search keys, their choice was most frequently (48% of the time) determined by the databases they were searching and least frequently (20%) by their habitual searching behavior” (1991b, p. 509). This result illuminates that the habits searchers develop change dynamically to suit the particular databases in which the search is being performed.

In the third article in the series “Searching Styles,” Fidel examines the various methods and techniques different searchers use when approaching an information problem for which they will do an online search. The two styles examined were the use of operational moves and conceptual moves. The former of these is a move or any search action made that does not change the meaning of the search strategy. These moves include reducing or enlarging the size of the retrieved set of documents, and “moves to simultaneously increase both precision and recall” (1991c, p. 517). Conceptual moves on the other hand, are meant to change the meaning of the search request, and they are
organized into the same categories as operational moves, i.e., increasing and decreasing the number of results, and moves affecting precision and recall. Fidel’s examination of moves that searchers make during a search speaks to the level of manipulation searchers use in altering the search strategy, whether by adding or removing terms, adding or removing databases, or restricting or enlarging the retrieved set of documents by publication year, publication type, or language (1991c, p. 519). For instance, Fidel finds that “interactive searchers” or those who use more moves than others, “are likely to use a larger number of search keys than searchers who are less interactive” (1991c, p. 519). Thus, the series of three articles by Fidel show that the different searching styles of searchers have implications for the conceptualization of the information problem at hand, the number and types of terms (free text or controlled vocabulary) they select, depending on their own preferences and the databases used in the search, and also the degree to which searchers manipulate the individual components of a search strategy in order to retrieve relevant and useful results.

*Indexing and Synonyms*

Also important in the area of the source of search terms is the availability and use of the indexing of the particular database or databases the search will be performed in. If there is good indexing available, meaning that the database vendor does a good job with indexing the articles contained in their products, intermediaries will use this knowledge to their advantage in retrieving relevant search results. Gomez, Lochbaum, and Landauer (1990) state “the probability that a system will return the desired object is a function of how often the user and the indexer refer to the same object by the same words” (p. 548). Following this logic, the authors also posit that there are issues with indexer consistency
because “each object in a database has many aspects and dimensions, and different people choose different words to describe the same features. Further, different people, or the same person on different occasions, will be interested in different aspects of the same object” (p. 548). This article as well as those that follow point to several facets of the practice of indexing, both as it relates to indexers as a group, and as indexing relates to the intermediaries and end-users who must make use of it.

One of the databases used very frequently at the searching site where the analysis is to be performed is MEDLINE. Funk and Reid (1983) studied MEDLINE records that had been indexed twice by mistake by indexers at the National Library of Medicine, which built and maintains the database. The study showed that the twice-indexed records agreed on the central concept of the article by including the same central concept subject headings in 61.1% of cases (p. 180). The authors also found that central concepts in the article were indexed with more consistency than non-central concepts. Because MEDLINE records are consistently indexed, as both the study found and search experience shows, searchers can rely on indexing in this database, knowing that every article will be indexed according to the MeSH schema and indexed consistently. For this reason, terms derived from consistent indexes are likely to appear in final search strategies.

A study on the indexing consistency of another database was done by Sievert and Andrews (1991). Information Science Abstracts or ISA was studied and compared with the MEDLINE study mentioned above. The authors point out a key difference between MEDLINE indexing practices and ISA indexing practices: namely that MEDLINE indexers can choose more terms to describe the subject of the document than those at
ISA, and MEDLINE indexers can also designate certain terms as central concepts, which ISA indexers cannot (p. 5). The indexing consistency of ISA as compared to the MEDLINE central concepts was lower for both the main headings and subheadings, but this is not a perfect comparison (p. 5). From these two studies, two main points arise that have bearing on this study: one, that databases have different indexing practices and guidelines which affect the number, type, and scope of indexing terms as well as how they will be applied to documents; and two, that the consistency with which databases are indexed according to their own rules can and does vary. Search intermediaries are responsible for being aware of and considering the various indexing practices when putting together a search strategy.

To further this point of the importance of intermediaries especially in the context of indexing and information retrieval, Anderson and Pérez-Carballo (2001) make an important point about the guidelines with which human indexers are provided today: “Most guidelines…focus mostly on the content and features of messages, texts, and documents, and less on potential uses by interested persons…and even less on the relevant characteristics of potential users and their information needs and information seeking behavior” (p. 250). Therefore, intermediaries can and do play an integral role in guiding users to the materials they need by using their skills in searching techniques and learning about indexing practices in order to utilize indexing effectively in a search session.

Another aspect of this research question is the presence of synonyms in both the initial information request and the final search strategy. Synonyms are very important in informal retrieval because of the myriad of words humans have to describe anything and
everything. Synonyms are needed because in order to retrieve a given article, one must use the words mentioned in the article or index terms assigned to the article after the fact. Therefore there is a need to be as verbose as possible on both the retrieval side and the article/database side so that a user or searcher can find documents that will address a need. Furnas et al. (1987) address this “vocabulary problem” that is inherent in the interplay between human language and machine retrieval. The authors state that “in information retrieval systems, the keywords that are assigned by indexers are often at odds with those tried by searchers. The seriousness of the problem is indicated by the need for professional intermediaries between users and systems” (p. 965). Furnas et al. argue that it is in the best interest of all people involved in information retrieval to come up with as many terms as possible to describe whatever the object, idea, or concept is, because there is rarely only one good term for most objects or ideas.

The addition of synonyms to a search strategy also relates to the searching experience of the person who adds them. Hsieh-Yee (1993) studied both search experience and subject knowledge and their effects on how novice and experienced searchers perform various functions associated with building a strategy and performing the search. When the participants searched for information about a topic that was outside of their subject knowledge base, there were several key differences between the groups observed by the author. The experienced searchers made more use of the thesaurus for term suggestion, included more term synonyms, and tested out more term combinations than the novice searchers (p. 169). Hsieh-Yee posits that novices had more trouble in synonym generation, so they adhered to the terms that were provided in the question, whereas experienced searchers included more synonyms and refined the search strategy
to a greater extent. The author states that “user’s reliance on nonthesaurus terms and their low use of the thesaurus when dealing with a topic they had little knowledge of suggest that end users use a limited number of sources for term selection and do not seem to appreciate the distinction between natural language and controlled vocabulary and limitations for online searches” (p. 169). Ruthven (2003) reports similar findings, in that “users cannot always identify useful semantic relationships for retrieval” (p. 219). Ruthven points out that users are sometimes not very skilled in understanding why concepts or terms are related to each other in the context of selecting terms for query expansion to retrieve additional relevant documents. Since users could not pinpoint useful terms for future retrieval, they “concentrated mainly on terms they viewed as safe; those that were semantically related to the topic description rather than the retrieved relevant documents,” which is similar to the finding in Hsieh-Yee (1993) that users stuck to the terms found in the topic question when they didn’t know what other terms to use (p. 219).

*Elicitation of Patron’s Needs at the EPA*

All of this literature pertains to issues surrounding the gathering of information from patrons and the subsequent online literature searches that librarians perform using this information. The method by which the EPA Library collects this information is through patrons filling out an online literature search request form and subsequent conversations. In order to add a second dimension to the analysis of the terms from the original online literature search request forms that the site has been using for at least the past six years, a comparison between this data and data gathered from the revised form will be performed. The current form was a good candidate for revision because much of it relates to administrative content and the section for the topic or question is more or less
a blank space to be filled in. The revised form drew on aspects of the literature that is described in this section in order to collect more and better information from those submitting online literature search requests. The revised form took cues from the elements of a face-to-face reference interview, which is a necessary step in most information requests made by patrons to librarians in order to clarify the topic and details of the request. Typical reference interview questions asked in the current study’s setting are aimed at finding information about what the patron already knows about the topic, what he or she needs to know, why the patron needs this information or to what ends will the resources found be put, and any limits that the patron would like to impose on the result set, for instance, date ranges and limiting to specific language.

Two of the more important aspects at the inception of a literature review request are the interaction that occurs as part of the reference interview, and the subsequent capturing of this information on the online literature search request form. Though separate, these processes are inextricably linked because the gathering of information would not be as effective if it was not recorded, and the recording of information could not happen without its being gathered. In this study, the form was revised in order to better incorporate and capture the information that occurs as part of a reference interview for two purposes: one, to require the patron to respond to a “reference interview on paper” by answering the questions on the revised form; and two, to make the subsequent face-to-face interview a more productive process, because some of the preliminary information will have already been gathered on the form. In this way, the revised form serves to make the reference interview an easier and better process because the questions
on the form will have already begun to elicit more information, encouraging the patron to think about and focus his or her information need.

As was just discussed, the interaction between the librarian and the patron requesting information is extremely important. Spink, Goodrum, Robins, and Wu (1996) studied the various elicitations that were made by either the user or the search intermediary in an information retrieval session. Of the total number of elicitations coded in the study, 64% were from the search intermediary, asking for information from the user, and the other 36% of elicitations were from users asking for information from the search intermediary, demonstrating the intermediary’s leading role in defining and refining the search goals and strategies (p. 122). Spink et al. categorized these elicitations by their purpose and found that both users and intermediaries were “concerned primarily with eliciting information from each other about selection of search terms and strategies” (p. 124). This research speaks to the importance of the interaction between the user and the intermediary and also to the capturing the important contributions each makes to the search process.

This study shoes that not only is it important that these interactive processes take place in order for the search intermediary to gain the clearest perspective on the information need of the client, but it is also important to document this process for several reasons. First of all, the act of writing is a clarification skill in and of itself, and this also serves as a meeting point between the user and the intermediary; agreeing upon the language and the meaning of the terms in the search is very necessary to accurately pinpoint the topic of the search request. Documenting this process also helps to serve as a reference for future follow-up requests and also as a general learning tool for training
new online searching staff. The form also serves as a record of work requested and completed, and it attests to the valuable services the library offers to its patrons.

The other important aspect that will be considered in this discussion is the recording of the patron’s information need. One technique for collecting information from patrons is called “polyrepresentation” developed by Ingwersen (1996), in which the information need is addressed from different angles in order to elicit the most useful elements about the need itself (p. 3). This method aims to obtain from the user information that the user may not necessarily provide or think to provide to the search intermediary for use in the online literature search. Polyrepresentation speaks to, among other aspects, the basic topic of the information need, “what is currently known about the unknown,” and also why the user desires this information (p. 18). With the addition of this information, it is possible that a searcher may have a much better grounding in the nuances of the specific information need of the patron.

Kelly, Dollu, and Fu (2005) presented users with a web form that contained a set of four questions designed to elicit more information from users for use in query expansion or to clarify ill-defined information requests. The authors believe that “users should be considered as useful sources of terms for query expansion, independent of any information from the system” used in retrieval (p. 458). The questions from the form used in the study provided the basis for the revision of the online literature search request form. This is because the questions used in the study are similar to questions already typically asked in the reference interview interaction between the searcher and the user. The form was also shown to elicit additional terms for query expansion, which is one of the goals of this project. One of the questions outperformed the others in eliciting terms
used for query expansion; that of “Describe what you already know about the topic” (p. 462). This question produced the “lengthiest responses” from users, typically about “background and contextual information” which aided in search strategy revision (p. 462).

Conclusion

This literature review has discussed some of the more important facets of this current study. Ample literature exists to address information retrieval issues, and only some of the more relevant articles have been presented here. Emphasis was given to variables and issues directly under study in this research, such as the subject of queries, the number and source of search terms found in search strategies, the particular search strategies of intermediaries, and the background information necessary to understand why the online literature search request form was revised in the way that it was. This overview of the literature available helps to ground the reader in the thinking behind the analysis of the term composition of searches that have been completed using the original form and searches that have been completed using the revised form, and the variety of issues that such an analysis must consider.
Methodology

All data emanated from and all research was done at the EPA Library in Research Triangle Park, North Carolina. The data for Study 1 of this research was retrieved from the online literature search request form used up until January 2006 at the Library and Study 2 data came from the revised online literature search request form. Currently the author is an intern in the Library at the EPA and she has completed the Online Searching Rotation. One of the job responsibilities of this Rotation was to suggest process improvements and ideas to improve the online searching experience for the librarians who do the online searching using Dialog databases, and for the patrons who request online literature searches. The author received permission from the Library Director April Errickson and the Assistant Library Director Susan Forbes to revise the paper form used to collect this online searching data as part of those job responsibilities. This revised form replaced the original form permanently as of January 9, 2006, as mandated by the aforementioned people. The revised form was not an experimental form as such; it has been adopted for permanent and exclusive use by the EPA Library. The term data for Study 2 was collected for the first two months that this form was in use. This study was approved by the UNC Behavioral IRB (Study # LIBS-2005-84).

Study 1: Collection and Analysis of Existing Data

Design, Methods, and Procedures. Approximately six years worth of searches that used the original online literature search request form was available for analysis for
Study 1. These forms are organized in folders by month and the twelve folders for one year are kept together. To perform stratified systematic sampling, these forms were first divided into groups by year, then by month, and then further organized by the searcher or searchers who handled the information request. For example, all searches done in a month by one searcher were placed together by order of the date the form was submitted. If there was more than one searcher who worked on a request, then these dual-searcher forms were given their own category and were sorted like the single searcher forms, by date of submission of the form. Figure 1 displays this form: “Original Online Literature Search Request Form.”

Figure 1. The Original Online Literature Search Request Form

<table>
<thead>
<tr>
<th>Cost</th>
<th>Date Run</th>
<th>Database(s)</th>
<th>Number of Records</th>
<th>Emailed</th>
<th>Mailed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Searcher
# Search Questions ______ x # Databases Searched ______ = Total Searches ______
It is necessary to explain what forms were included in the sample and which forms were not. The exclusion of some forms directly relate to the content of the question they have been used for. Exclusions were made as follows:

- If the form was used for a citation search, it was excluded. Citation searches do not involve term selection and revision in the same way that other ad hoc searches do. Citation searches consist of a citation or a group of citations that the patron had written on the form or communicated to a searcher in another parallel way (email, phone, etc.), and had asked for the number of articles that have cited that article or for the actual articles themselves which have cited that initial article. These searches are more straightforward and do not involve the interchange and the accompanying reference interview and topic clarification that other searches do, and so were not appropriate for this analysis.

- If a form contained a question or topic that could not be adequately addressed by the searching methods used, it was excluded. Some topics submitted on the online searching form are better answered by techniques other than online searching, for instance, an in-depth chemistry question may be better answered by a textbook than by articles published and indexed in the databases the Library makes use of. These searches that have not been adequately answered by the resources gathered as a result of an online database search were not included because there was a lack of confidence in the final search strategy. The strategy was not able to provide the patron with what he or she was looking for, and so the analysis of the search terms and strategy involved in the final search would not have given an accurate representation of a patron’s needs being addressed.
• Searches that did not conform to the usual parameters of an online search request that characteristically used Dialog files and were one-time requests for information were excluded. Searches in which PubMed, other Internet, or other resources were used were excluded, as this does not fit the usual requirements for an online literature search. Also excluded were search forms that were used solely to set up alerts, which are saved searches that are run automatically by Dialog and typically use broader search strategies than those that address a single information need. Another important facet of alerts is that the results are sent to patrons without being filtered by a librarian, so there is no opportunity for revision of the strategy by the librarian, or discussion between the librarian and the patron for each result set.

• Searches that the author was involved in were excluded from Study 1 as well. This was done in order to prevent bias towards the results due to familiarity with the question topic and the various processes involved in those searches. Since the author did not conduct searches in cases where the revised online literature search request form was used, this issue was also avoided in Study 2.

• If any of the term information, namely the initial information request terms or the final search strategy was missing, the form was excluded from the analysis.

After the forms were sorted according to the schema outlined above and the forms that meet the criteria for exclusion were removed, the forms that were included in the sample were gathered. The process of eliminating forms was carefully tracked and reasons for discarding forms were noted. The method for determining the sampling interval depended on the total number of forms in the population after the excluded forms
were removed. Once the size of the population was determined, an appropriate sampling interval was determined. The number of forms in the entire population was 543, and the number of usable forms (i.e., not meeting any of the criteria for exclusion) was 240. From these 240 forms, a sampling interval of every fifth form was taken, for a total sample size of 48 forms that used the original online literature search request form.

The only data that was analyzed on the forms was that relating to the terms relevant to the search. Personal or identifying data collected on these forms was not used in analysis. The included original form collected at the top such administrative information as name, phone number, mail drop, and department. The space on the form relevant to this study is the middle section, which states: “Please explain the subject of interest. Describe your request in sentences to ensure that we understand the relationships between the concepts in your search.” Patrons typically wrote a few words or sentences in this middle space, whereas the top and bottom of the form, used to track cost and work done on the search, are used more for administrative and tracking purposes.

**Study 2: Collection and Analysis of New Data using Revised Form**

**Design, Methods, and Procedures.** The included revised form, which permanently replaced the original form as of January 9, 2006, has been administered by the new Online Searching Rotation intern and the Assistant Director of the Library, who comprise the online searching staff at the EPA Library. These two librarians are the only people who carry out the online searching functions in the Library, and both have been using the revised form exclusively. This two month time frame was determined as reasonable in order to gather some new data under current time constraints, which are
affected by the author’s graduation in May. The author had no part in the administering of this new form because her tenure in the Online Searching Rotation had ended and she has moved into another rotation. She also removed herself from this process in order to eliminate bias in having an interest in patron’s filling out the form completely as she would like them to. The intent was to have the form be used in a natural setting that was identical to the setting in which the previous unrevised form was used. Figure 2 displays this form: “Revised Online Literature Search Request Form.”
Figure 2. The Revised Online Literature Search Request Form. The front of the form and the back of the form are below.

ONLINE LITERATURE SEARCH REQUEST FORM

Requestor's name: ___________________________ MD: _____ Phone: ______ Date received: ______ Needed By: ______ Date Completed: ______

Circle one: NRMRL/APPCD NCEA NERL NHEERL OARM OAQPS OE/EOTOP NHISRC When the results are ready, please Call Me or Email Me at: ________________

Time Period: Search from ______ to _______ ENGLISH ONLY? Yes ___ No ___

Would you like your results TAGGED for use in a bibliographic manager such as EndNote or ProCite? Yes ___ No ___

Please describe your request in sentences to ensure that we understand the relationships between the concepts in your search.
(Note: For author searches, please include first and middle initials. For chemical searches, please try to include CAS registry numbers.)

How many times have you searched for information about this topic in the past?

Never 1-2 times 3-4 times 5 or more times

What is the topic of your request? What do you already know about the topic that is relevant to the search?

How will you use this information in general? (Laboratory research, writing an EPA Document, etc.)

Please write any additional keywords that describe your topic, or anything else we should know.

January 2006

<table>
<thead>
<tr>
<th>Cost</th>
<th>Date Run</th>
<th>Database(s)</th>
<th>Number of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Searcher __________________

# Search Questions ______ x # Databases Searched ______ = Total Searches ______
The first two months of data generated by this revised form was included for analysis for Study 2. Again, this revised form has permanently replaced the original form due to administrative decisions unrelated to the existence of this proposed research. Due to the small population of data that was collected in just two months, all usable forms were included in the data analysis. The exclusion criteria for this study were the same as for Study 1.

After the forms were sorted according to the schema outlined in the above section and the forms that met the criteria for exclusion were removed, the rest of the forms were analyzed. The process of eliminating forms was carefully tracked and reasons for discarding forms was noted, just as in Study 1. The number of forms that comprised the total population was 12, due to a relatively slow two months in online searching. The number of forms that were usable (i.e., not meeting any of the criteria for exclusion) was 5, and since the data set was small for this study, all of the usable forms were included for analysis, forgoing the stratified systematic sampling technique.

The only data that was analyzed on the forms was that relating to the terms relevant to the search. Personal or identifying data collected on these forms was not used in analysis. The included revised form collected such administrative information at the top as name, phone number, mail drop, and department just as the original form did. The space on this form relevant to this study is everything on the front page below the dashed line beginning with “Please describe your request in sentences to ensure that we understand the relationships between the concepts in your search.” Other, more specific questions were asked on this form including “How many times have you searched for
information about this topic in the past?,” “What is the topic of your request?,” “What do you already know about the topic that is relevant to the search?,” “How will you use this information in general?,” and “Please write any additional keywords that describe your topic, or anything else we should know.” These questions were created using research by Kelly, Dollu, and Fu (2005) as a guide. The grid used to collect cost and progress data on the current form was moved to a second page, which appears on the back of the form when it is printed as double-sided. It was expected that in asking more specific questions and providing more space for patrons to write information on the revised form, that the topic information and thus the terms provided would be found to have a greater overlap with the final search strategy than the original form.

Data Collection and Analysis – Study 1

In order to track each of the forms examined in Study 1, a form number and a date were given to each form on the attached Excel spreadsheet. These were the only identifying features noted in the collection part of the research, merely to ensure that the same form was not analyzed twice and also to serve as a way to identify the form if term decisions needed to be reconsidered. A short explanation on which words were considered to be terms is necessary: though the definition of “term” is approached in many different ways in the literature, for this study, a term was considered to be any word that added meaning to a request. Words like “asthma” or “particulate” were considered terms, whereas prepositions and stop words were not considered terms and thus were not recorded as such. Words that aided in the description and clarification of the information problem at hand were considered to be terms in this analysis.
Terms for Study 1 were identified for analysis and were categorized as follows. Those terms that fell into the “Initial Information Request Terms” category were recorded. These terms included those terms written by the user on the front of the form in the space provided to detail the topic of the literature search request. If these terms were written by a librarian for a patron who called in a request, these terms were still counted as the initial information request, or IIR. These terms, and all in the subsequent categories, were totaled for each online literature search request form. Also notated was the source of these initial information request terms; if the term was written on the form by the patron or for the patron by a librarian, the source of the term was the patron. If it was clear that a librarian added words to the space on the form used for the initial information request, then the librarian was deemed as the source of the terms. The terms from the librarian could have originated from several sources, such as the reference interview that the librarian had with the patron, the librarian’s own perception and comprehension of the topic, or from the librarian’s knowledge of indexing practices of particular topics or databases.

The next category of terms analyzed were those that comprised the Final Search Strategy or FSS, which is printed out from the search buffer of the Dialog databases used at the completion of the search process and taped onto the back of each search form. These were the terms used to actually perform the search and serve as a good source of information with which to compare patron generated terms. Also notated were the final search strategy limits, or the various commands used by the search intermediary in order to manipulate the terms in the strategy. For example, Boolean operators used were transcribed, as well as commands to limit the results to articles with certain terms in the
title or descriptor fields. These limits were tracked to demonstrate the thought processes that go into devising and performing a search strategy, but were not used as part of the term analysis.

The next category of terms was the Term Overlap between the IIR (initial information request) and the FSS (final search strategy). These terms are the direct overlap of those terms found on the front of the form and those that were actually used to perform the search. This category consists of strictly overlapping terms – variations on terms were not placed in this category in the effort to show exactly how many terms found in the initial information request (the vast majority produced by patrons) were used in the final search strategy by searchers. The next category, Terms in FSS not IIR, lists all of the terms in the final search strategy that were not deemed to be part of the initial information request. These terms were added by the search intermediary to the final search strategy in order to retrieve relevant results. Of these terms in this category, a subset was listed in the Synonyms for IIR in FSS. This final category outlines of the terms that were added by the librarian to the search strategy, which and how many were synonyms of those terms found in the initial information request, most of which were provided by the patron. Synonymy was decided by the researcher and was rather strict: related concepts, of which many of the “Terms in FSS not IIR” were, were not counted as synonyms. This was done for two reasons: one, that the terms in fact were describing elements of the search that were not introduced by the patron, and two, that the distinction between the librarian’s adding of synonyms and adding of other terms could be preserved.
Data Collection and Analysis – Study II

Much of the same collection and analysis that was developed for Study 1 was used in Study 2. Each form was given a number and its date was recorded. All categories were the same for Study 2 except that the “Initial Information Request Terms” column in Study 1 was broken down into several columns to account for the various questions asked on the revised online searching form. Categories that were subparts of the IIR were as follows: Times searched on this topic, Topic Terms, Topic Related Terms, Use of Information, and Any additional keywords. These categories reflect those terms and information gathered in response to the questions listed in Section II. The Topic Terms category was intended to collect terms describing the information need of the patron. The Topic Related Terms category collected those terms that the patron provided that were related to the information request and reflected already known data about that specific information request. Responses to the question of what the patron will use this information for were recorded in the Use of Information category. The last category that was used exclusively with the revised online literature search request form was the “Any additional keywords” category that corresponded to the last question patrons were asked. This question and space was meant to allow patrons to write down any additional terms or other information that they thought it was important for the searchers to know. All other categories described above in Section III were included for analysis in Study 2.

A comparison of the statistics of each of these term categories across the two sets of form data was done to determine the extent to which the revision of the online literature search request form impacted the search terms present. While the sample size
for Study 1 was most likely sufficient to gather general trends about term selection, the small sample size in Study 2 may have ramifications in its generalizability and accurate depiction of the usefulness of the revised form.

Other considerations

Risks and measures to minimize risks. For the first part of the project (Study 1), there was no direct interaction with subjects, as existing data was analyzed for its content. The second part of the project (Study 2) used data that was created on the revised form during the form’s first two months of being used for online literature search requests in the Library. The expected risks to patrons were minimal, as the subjects are being asked to write some more specific information about a search they requesting for work purposes, and all participation in the literature search process was voluntary. This scenario encouraged collection of realistic data from searches that were actually requested because of some information need. A small amount of embarrassment of the subject may ensue if the subject could not answer the questions given, but these questions were not more invasive than the typical verbal reference interview questions that take place currently.

Confidentiality of the data. The only data that was included in the analysis of Study 1 was that of the terms relating to the search: any identifying characteristics pertaining to the subjects was not considered or included for any reason. The same applied to the revised forms that were collected for Study 2. Since the author did not include any data from online searches that she had performed, she was not able to identify the patron based on the content of the search. The same held true for the revised
forms, since they were administered by another online searcher, and so the author was not able to identify the patron based on the content of the search.

All search forms were kept in a locked office when the searching intern was not present, and so could not be accessed by unauthorized Library personnel. Substitute identifiers were not necessary for the data on the forms because the data from the patrons and from the librarians was being examined as a whole, and the analysis was not based on any other criteria except for those terms the patron provided and those terms that were provided by other sources, one of which may have been the librarian or searcher involved.
Results

Study 1

The sample of 48 forms was taken from the useable portion (240 forms) from the total corpus (543 forms) for Study 1. There were 303 forms that met the exclusion criteria outlined in the Methods section. Forms that had no final search strategy, citation searches, forms used solely for new alerts, search forms in which the initial information request of the user was unclear, and searches that were done using resources other than Dialog database files were excluded.

Results of the analysis of Study 1 data are presented in Table 1 and the data from two example forms is listed in Table 2. There were a total of 518 terms that were recorded as originating from the initial information request, or IIR. These terms are those that are written on the front of the form in the space provided in order to detail the topic of the request. For the most part, these terms are written or provided by the patron. The mean number of IIR terms per form was 10.79, with the smallest number of IIR terms being 4 and the largest being 24. The most commonly occurring number of terms in the IIR was 7 terms, occurring 9 out of 48 times.
Table 1. Results of Study 1

<table>
<thead>
<tr>
<th>Initial Information Request Terms</th>
<th>Source of terms from IIR (patron and librarian)</th>
<th>Final Search Strategy Terms</th>
<th>Term Overlap between IIR and FSS</th>
<th>Terms in FSS not IIR</th>
<th>Synonyms for IIR in FSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total IIR Terms: 516</td>
<td>Patron Terms: 488</td>
<td>Total FSS Terms: 709</td>
<td>No. forms patron-only terms: 41</td>
<td>Total FSS not IIR: 371</td>
<td>Total Syn for IIR in FSS: 201</td>
</tr>
<tr>
<td>Mean IIR Terms: 10.79</td>
<td>Librarian Terms: 30</td>
<td>Mean FSS Terms: 14.77</td>
<td>No. forms patron and lib. terms: 7</td>
<td>Mean FSS not IIR: 7.73</td>
<td>Mean Syn: 4.19</td>
</tr>
<tr>
<td>Lowest No. of IIR Terms: 4</td>
<td>No. Forms only Patron terms: 37</td>
<td>Lowest No. of FSS Terms: 4</td>
<td>Total Overlap: 338</td>
<td>Lowest FSS not IIR: 0</td>
<td>Synonyms/Total FSS not IIR: 54.18%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Library Overlap: 15</th>
<th>Highest No. Syn: 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Overlap: 7.04</td>
<td>Lowest No. of Overlapping terms: 3</td>
</tr>
<tr>
<td>Highest No. of Overlapping terms: 16</td>
<td></td>
</tr>
<tr>
<td>Overlap/FSS Terms: 47.67%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Examples of Data Collection from 2 forms from Study 1

<table>
<thead>
<tr>
<th>Form No.</th>
<th>Date Form Submitted</th>
<th>Initial Information Request Terms</th>
<th>Source of terms from IIR</th>
<th>Final Search Strategy Terms</th>
<th>Final Search Strategy limits</th>
<th>Term Overlap between IIR and FSS</th>
<th>Terms in FSS not IIR</th>
<th>Synonyms for IIR in FSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>6/8/2001</td>
<td>Aerocoil characterization</td>
<td>patron</td>
<td>aerosol? and or particles</td>
<td>i statistical low</td>
<td>data</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patron size?</td>
<td></td>
<td>aeroel size?</td>
<td>statistical low</td>
<td>data</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diameter</td>
<td>particle</td>
<td>particles electrical</td>
<td>measure?</td>
<td>measure?</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Statistical</td>
<td></td>
<td>statistical</td>
<td>measure?</td>
<td>measure?</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data</td>
<td></td>
<td>statistical</td>
<td>measure?</td>
<td>measure?</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reported</td>
<td></td>
<td>statistical</td>
<td>measure?</td>
<td>measure?</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PM</td>
<td></td>
<td>statistical</td>
<td>measure?</td>
<td>measure?</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Form No.</th>
<th>Date Form Submitted</th>
<th>Initial Information Request Terms</th>
<th>Source of terms from IIR</th>
<th>Final Search Strategy Terms</th>
<th>Final Search Strategy limits</th>
<th>Term Overlap between IIR and FSS</th>
<th>Terms in FSS not IIR</th>
<th>Synonyms for IIR in FSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>1/2/2000</td>
<td>Black</td>
<td>patron</td>
<td>black</td>
<td>i, (in)</td>
<td>data</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carbon</td>
<td>carbon</td>
<td>carbon</td>
<td>elemental</td>
<td>data</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Atmospheric</td>
<td>carbon</td>
<td>carbon</td>
<td>elemental</td>
<td>data</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emissions</td>
<td>inorganic</td>
<td>inorganic</td>
<td>elemental</td>
<td>data</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emissions</td>
<td>1980-2012</td>
<td>emissions</td>
<td>air</td>
<td>data</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emissions</td>
<td></td>
<td>emissions</td>
<td>air</td>
<td>data</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emissions</td>
<td></td>
<td>emissions</td>
<td>air</td>
<td>data</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emissions</td>
<td></td>
<td>emissions</td>
<td>air</td>
<td>data</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Of these 518 IIR terms, 488 were produced by the patron. This means that the source for 94% of the terms for the initial information request was the patron, which was not surprising, given the fact that the patron fills out the form at the onset of the literature search request. Of the 48 forms, 37, or 77% contained only terms given by the patron. Thirty terms written on the front of the form as the initial information request were attributed to the librarian, which accounts for just about 6% of the terms. Eleven of the 48 forms, or almost 23% of the sample contained terms that were produced by two sources: the patron and the librarian.

The total number of final search strategy (FSS) terms as found on the search strategy attached to the back of the form for all 48 forms was 709, with a mean number of 14.77 terms per form, almost 4 terms more than the average for the IIR. The lowest number of FSS terms matched that of the IIR terms at 4, while the highest number of FSS terms was 39, which was larger by 15 terms than the highest number of terms for IIR. The most commonly occurring number of final search strategy terms was 13, occurring 7 out of 48 times. The difference between the number of IIR terms and the FSS terms was not a simple subtraction of the two totals, as reviewed in the Discussion section.

The next category of results deals with the term overlap between the initial information request (IIR) and the final search strategy (FSS) for each literature search. The total overlap of terms between these two categories was 338, with 323 terms emanating from the patron and 15 terms coming from the librarian. In other words, of the total overlap, the source of 323 of these terms was the patron, and the remaining 15 terms were written on the search form by the librarian at some point during the search process.
The mean overlap in terms of the 48 forms was 7.04 terms, the lowest overlap being 3 terms and the highest overlap being 16 terms. The most commonly occurring number of overlapping terms was 6 terms, occurring on 8 of the 48 forms. There were 41 forms on which only the patron had terms overlapping and 7 forms on which the terms that overlapped were derived from both the patron and the librarian.

In terms of the percentage of overlap between the initial information request and the final search strategy, which roughly equates to how productive patrons (and in a small part, the librarians) were in generating terms that are actually used to run the search, the results are varied. The IIR terms overlapped with the FSS terms a mean of 47.67% of the time. This average, however, can be misleading. The overlap values range from the lowest overlap of terms, which was 13%, all the way up to 100% overlap. There were five instances of these “perfect overlaps”: that is, every term in the initial information request was contained exactly in the final search strategy. Even this result has a caveat: only three of these five “perfect overlap” searches were truly perfect in that all the terms the patron wrote were included in the final search strategy and no other terms were included in the FSS. The other two instances of “perfect overlap” used a portion of terms from the IIR in the FSS, and did not add other terms, but excluded some of the IIR terms given by the patron. So, out of the 48 searches examined, only three used all of and only those terms provided by the patron. The overlap of IIR and FSS terms was 338, and the total number of IIR terms was 518, the number of terms that were discarded from the IIR and not used in the FSS was 180.

The next category of terms under consideration was that of the terms in the final search strategy that were not found in the initial information request. In other words,
these were the terms that were added to the search strategy by the librarian, taking into account those terms already provided in the IIR. The total number of terms found in the FSS but not in the IIR was 371. These terms account for slightly over half (52.32%) of the terms included in the final search strategy, with the balance of the terms originating from the initial information request terms, as discussed above. The mean number of terms in this category was 7.73 per search, the lowest number of terms in the FSS but not in the IIR was 0, for those five “perfect overlaps,” and the highest number of terms was 33. The most frequently occurring number of terms was a tie, with six instances each of two and eight terms.

The last category was a subset of the preceding category to determine what portion of the terms added to the FSS by the librarian were synonyms for terms in the IIR. There were a total of 201 terms that were synonyms for terms in the IIR that were contained in the FSS. The mean number of synonyms was 4.19 per search, the lowest number of synonyms being 0 and the highest number being 15. The most commonly occurring number of synonyms was 2, occurring in 11 out of the 48 forms. From these last two categories, it was determined that 54.18% of the terms in the final search strategy that were not in the initial information request were synonyms for terms in the IIR. This finding shows the role of the librarian in adding synonyms to the search, but it also points to the finding that 46% of the terms that the librarians added to the final search strategy were not direct synonyms, but were related in subject matter to the topic being researched.
Study 2

Results of the analysis of Study 2 are presented in Table 3 and the data from an example form is given in Table 4. Of the 12 forms that were used for online searches during the first two month period in which the revised form was being used, 7 met the criteria for exclusion, including one search for which the patron used an original form. Other forms were excluded because they were citation searches, searches that were not finished or had no final search strategy, or they were searches that dealt exclusively with alerts.

Table 3. Results of Study 2

<table>
<thead>
<tr>
<th>Study 2: 5 forms</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Times searched on this topic</th>
<th>Topic Terms</th>
<th>Topic Related Terms</th>
<th>Use of Information</th>
<th>Any Additional Keywords</th>
<th>Source of terms from IIR (patron or librarian)</th>
<th>Final Search Strategy Terms</th>
<th>Term Overlap between IIR and FSS</th>
<th>Terms in FSS not IIR</th>
<th>Synonyms for IIR in FSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Responses</td>
<td>5 responses</td>
<td>3 responses</td>
<td>4 responses</td>
<td>1 response</td>
<td>70 (all) terms from patrons</td>
<td>Total FSS Terms: 57</td>
<td>26 overlapping terms</td>
<td>Total FSS not IIR: 31 terms</td>
<td>Total synonyms: 16</td>
</tr>
<tr>
<td>2 Nevers</td>
<td>58 topic terms</td>
<td>2 list articles already obtained</td>
<td>3 papers</td>
<td>12 additional terms given</td>
<td>Mean IIR Terms: 14</td>
<td>Mean FSS Terms: 11.4</td>
<td>Mean: 5.2 overlapping terms</td>
<td>Mean FSS not IIR: 6.2 terms</td>
<td>Mean synonyms: 2.2</td>
</tr>
<tr>
<td>Mean: 11.6</td>
<td>1 &quot;assume nothing&quot;</td>
<td>1 seminar session</td>
<td>Lowest No. FSS terms: 8</td>
<td>Lowest No. overlap: 3</td>
<td>Lowest No. FSS not IIR: 4</td>
<td>Synonyms/Total FSS not IIR: 51.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest No. topic terms: 6</td>
<td></td>
<td></td>
<td>Highest No. FSS terms: 18</td>
<td>Highest No. overlap: 8</td>
<td>Highest No. FSS not IIR: 12</td>
<td>Lowest No. Syn: 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest No. topic terms: 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Highest No. Syn: 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Example of Data Collection from one form in Study 2

<table>
<thead>
<tr>
<th>Form No.</th>
<th>Date Form Submitted</th>
<th>Times searched on this topic</th>
<th>Topic Terms</th>
<th>Use of Information</th>
<th>Topic Related Terms</th>
<th>Any Additional Keywords</th>
<th>Source of terms from IIR</th>
<th>Final Search Terms</th>
<th>Final Search Strategy limits</th>
<th>Term Overlap between IIR and FSS</th>
<th>Terms in FSS not IIR</th>
<th>Synonyms for IIR in FSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>2/2/2006</td>
<td>N/A</td>
<td>mathematical</td>
<td>N/A</td>
<td>preparing to write a paper on this topic</td>
<td>N/A</td>
<td>N/A</td>
<td>patron</td>
<td>steroidogenesis and or steroidogenesis estrogen biosynthesis progesterone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>models</td>
<td>patron</td>
<td>steroid</td>
<td>rti</td>
<td>steroid</td>
<td>progesterone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>steroidogenesis</td>
<td>patron</td>
<td>estrogen</td>
<td>org</td>
<td>synthesis</td>
<td>testosterone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>synthesis</td>
<td>patron</td>
<td>progesterone</td>
<td>orgy-1999</td>
<td>model</td>
<td>biosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>steroid</td>
<td>patron</td>
<td>testosterone</td>
<td>rd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>hormones</td>
<td>patron</td>
<td>biosynthesis</td>
<td>orgy-1979</td>
<td>model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|   |   |   |   |   |   |   |
| 6 | 6 | 9 | 4 | 4 | 1 |

Results for the remaining five forms are as follows. The Initial Information Request of Study 1 was broken down into 5 categories for Study 2 to reflect the 5 specific questions patrons were asked on the revised form. There were a total of two responses given out of five search forms for the first category of analysis for the revised search form: Times searched on this topic. This question gave patrons a scale of how many times they have searched for information on this topic, and the two responses were “Never.”

The next category was “Topic Terms,” which was the only one of the five IIR categories in which data was collected for each of the five forms under consideration. In other words, this was the only field on the search form that was filled out on every form that was being analyzed for Study 2. On those 5 revised search forms, there were a total of 58 Topic Terms, with an average of 11.6 Topic Terms per search. The lowest number of terms in this category was 6 and the highest number of terms in this category on a search was 16. Each value of Topic Terms occurred once across the five forms.
The category of Topic Related Terms elicited responses on three of the five forms and the responses were of two types. Two forms detailed what resources that patron already had and thus wouldn’t need, for example, this category on form 49 stated that the researcher already had parts 3 and 4 of a particular series, and was interested in obtaining the first two parts only, and form 51 had a similar comment. Form 53 indicated that the patron wanted the searcher to assume that she did not know anything about the topic whatsoever when approaching her search.

The Use of Information category was the most homogenous of the categories in terms of the four responses this question received. Three of the four patrons who responded to this question explained that they were writing a paper, and one patron said the information was for a seminar or learning session. The catch all Any Additional Keywords question was only answered by one patron, who supplied an additional 12 terms.

All terms given on the forms were supplied by the patrons themselves, so the category that describes the source of the IIR terms was very straightforward. There were a total of 70 terms on all 5 forms, which was the sum of the Topic Terms (58) and the Additional Keywords (12). Therefore, the average number of terms in the Initial Information Request was 14, with the lowest number of terms being 6 and the highest being 27. The number of terms in the Final Search Strategy category was 57, with an average of 11.4 FSS terms per form. The lowest number of FSS terms was 8, the highest was 18, and the most frequent was 8, occurring in three of the five search forms.

There were 26 terms that overlapped between the Initial Information Request and the Final Search Strategy. This equated to an average of 5.2 terms that overlapped, with
the lowest overlap being 3 and the highest overlap being 8 terms. Searchers discarded 44 of the 70 terms patrons provided them with as part of the five categories of the Initial Information Request. There were no “perfect overlaps” as there were in Study 1.

There were 31 terms across the five searches that were found in the FSS but not in the IIR, with an average of 6.2 terms per search. The lowest number of terms in this category was 4, the highest was 12, and the most frequent was 5, which occurred in three of the five searches. In other words, these terms in FSS but not in the IIR were the terms (31) that the searchers added to those they selected from the IIR terms (26) to comprise the terms that were used in the Final Search Strategy (57). Of the 31 terms that the searchers added, 16 or 51.6% were synonyms of terms in the Initial Information Request. There were an average of 3.2 synonyms added per search, the lowest number of synonyms being added was 1 and the highest being 6.

Comparison of Results between Study 1 and Study 2

In order to compare the results from the two studies, a brief explanation of the Initial Information Request category must be reiterated. In Study 1, the initial information request consisted of those terms that the patron wrote down in the one large space provided on the form that were elicited by the following sentences: “Please explain the subject of interest. Describe your request in sentences to ensure that we understand the relationships between the concepts in your search.” In Study 2, this one space and one question were broken down into five spaces and five questions in order to determine if more questions and better questions would elicit more information from the user. Thus, all the terms from these five spaces, which corresponded to five new categories in
Study 2 were considered to be part of the Initial Information Request category. Further discussion will be given to these five categories and the types of information extracted by each in the next section.

The number of terms that were part of the IIR in Study 1 was 518, translating to a mean number of IIR terms of 10.79 per search. The five categories that functioned as the total IIR produced a total of 70 terms, with the mean number of IIR terms being 14 per search. There was an increase in the average number of IIR terms provided by patrons on the search forms in Study 2 over Study 1. Even when comparing the one category of Topic Terms from Study 2, which most closely resembled the question used in Study 1, there was an increase in the average number of terms given by the patron: 11.6 terms in Study 2 versus 10.79 terms in Study 1.

All terms in the IIR in Study 2 were given by the patron, contrasted with Study 1, in which 94% of the terms were given by patrons, and the balance emanated from the librarian or librarians who worked on the search. The total number of Final Search Strategy terms in Study 1 was 709, with an average of 14.77 terms per search. The total number of FSS terms in Study 2 was 57, with an average of 11.4 terms per search. So, Study 2 searches had a lower average of total terms used for the final search strategy than Study 1 searches.

In the category of the overlap of the IIR and the FSS, Study 1 yielded a total of 338 overlapping terms, with a mean overlap of 7.04 terms per search. Study 2 produced 26 overlapping terms, with a mean overlap of 5.2 terms per search. This finding is important to emphasize: in Study 1, the overlapping terms accounted for 47.67% of the final search strategy terms. In other words, of the terms that were used as part of the final
search strategy, almost half came directly from the patrons. This percentage did not change much in Study 2: 45.61% of the final search strategy terms came from the patron’s initial information request. Gathering these statistics was the focus of this research project, though many other interesting and useful data also have been collected as a result of this research.

The next category also saw a slight decrease in Study 2: for the terms included in the final search strategy but not found in the initial information request, Study 1 searches had a total of 371 added terms, with a mean of 7.73 per search. Study 2 however, had a total of 31 added terms, with a mean of 6.2 terms per search. The last category, Synonyms for IIR in FSS, had a total of 201 synonyms, with an average of 4.19 synonyms for Study 1. This same category in Study 2 had a total of 16 synonyms, with an average of 3.2 synonyms per search. Of the terms librarians added in Study 1, 54.18% were synonyms for IIR terms, whereas in Study 2, 51.6% of the terms added were synonyms for IIR terms. This finding means that in both two studies, of the terms that the librarians added to the search strategies, about half were synonyms of terms that the patron provided, and about half were other related terms.
Discussion

Study 1

To reiterate, there were 518 terms in Study 1 that were counted as the initial information request from 48 searching forms. Of these 518 IIR terms, 488 were given by the patron and 30 were given by the librarian, accounting for 94% and 6% respectively of the total number of IIR terms given. Thirty seven search forms contained only terms produced by the patron, and 11 forms contained terms from both the patron and the librarian. The terms produced by the librarian were probably added while the librarian was conversing with the patron about the request and the librarian wrote down additional terms that were generated from this reference interview, speaking to the importance of interaction, as studied by Spink, Goodrum, Robins, and Wu (1996) and Spink and Saracevic (1997). These terms also could have come from the librarian’s experience with indexing terms or perhaps were terms that the librarian thought would be useful for other reasons.

The terms found on those 11 forms which contained terms from both the librarian and the patron document the fact that at sometime during the early stages of the literature search request, a librarian engaged in a follow-up interview with the patron. Some type of reference interview, whether it was formal or a series of follow-up questions and conversations after the form has been submitted, must have taken place because the terms were written down on the form by the librarian in the same space as the initial information request. This reference interview is vital to the search process because it
addresses the user’s anomalous state of knowledge as posited by Belkin (1980) and because it clarifies the language and meaning involved in the request, which was explored by Furnas et al. (1987). The reason why the existence of these terms written by the librarian on the form matters is because though the reference interview is a standard part of the online searching process and follow up conversations occur with very high frequency, the results of this reference interview are not consistently documented, as shown by librarian terms occurring only on 11 of 48 search forms. True, in the differences between the IIR terms and the FSS terms, one can discern that terms have been added and discarded; yet it is difficult to pinpoint the stage at which these decisions were made. Since these 30 terms were written directly on the front of the form by the librarian, it is plausible that these terms were added while conversing with the patron or were triggered in the librarian’s mind because of this conversation. Thus, unlike those terms that were added solely by the librarian in the final search strategy, these terms can be attributed to the interaction between the librarian and the patron. Terms produced by the librarian as a result of talking to the patron may not have been documented as such and may have led to an inflated number of final search strategy terms attributed to the librarian only. The point to take away here is that the interaction and clarification of search topics and goals is an important step in the searching process and was not as well-documented as it could have been on the forms used for Study 1.

This category of data is reminiscent of the study done by Spink and Saracevic (1997) in which the authors looked at the importance of interaction between the searcher and the user in generating terms that were the most productive terms in the resulting search strategy. One reason why these terms might be fewer in number than expected in
this study at this stage of the literature search process is that, as discussed before, this interaction process is not consistently documented, and so terms that were actually generated as part of the librarian-patron interaction may have been accounted for in the final search strategy as terms emanating from the librarian. In any event, the fact that the terms from any of the aforementioned sources (patron, librarian, and interaction) were documented and considered for inclusion in the final search strategy in order to address the patron’s information need is one of the more important functions of the online search form and the subsequent reference interview.

The next category under consideration is that of the number of Final Search Strategy terms found in the 48 searches examined. There were a total of 709 FSS terms in this category, which demonstrates that the search terms patrons provided were not enough in number or in content to retrieve the same amount or types of results that the librarians gathered as a result of the final search strategy. While this result is interesting, the more telling result concerning this category is its overlap with the terms found in the initial information request.

The total overlap of terms between the initial information request and the final search strategy was 338, with 323 terms emanating from the patron and 15 terms coming from the librarian. There were 41 forms on which only the patron had terms overlapping and 7 forms on which the terms that overlapped were derived from both the patron and the librarian. These results make sense, in that a large portion of the terms used in the final search strategy came from the source of the patron. The patron after all is the one who is requesting the literature search on a specific topic and it follows that he or she most likely has a fairly good idea of the terms that would aid in the retrieval of helpful
articles. While of course the literature posits that this is not always the case, due to the patron’s inability to articulate a need or that the topic of interest is not necessarily well formed or known about (Belkin, 1980), the patron is still treated as a valuable source of information on the topic he or she is inquiring about (Kelly, Dollu, and Fu, 2005).

In terms of the percentage of overlap between the initial information request and the final search strategy, which roughly equates to how productive patrons (and in a small part, the librarians) are in generating terms that are actually used to run the search, the results are varied. The IIR terms overlapped with the FSS terms a mean of 47.67% of the time. This average, however, as mentioned in the previous Results section can be misleading. The occurrence of the three “perfect overlaps” was discussed, but further discussion is warranted for this result. Out of the 48 searches examined, only three used all of and only those terms provided by the patron. In other words, if the patrons had the exact same extensive experience with databases and formulating search strategies as the searching librarians, which is one of their specialties in this government agency setting, in only three cases would the patrons have been able to use all and only their terms to retrieve the results that the librarians did. This result speaks to the importance of the services that librarians provide to their patrons by adding terms for use in searches, in addition to selecting databases, and constructing search queries in order to return relevant and useable results to the patrons who request literature searches. Thus, searchers are proficient in selecting terms that best gather the articles available that address a specific topic of interest, as mentioned in Fidel (1991a).

Since the overlap of IIR and FSS terms was 338, and the total number of IIR terms was 518, the number of terms that were discarded from the IIR and not used in the
FSS was 180. These terms were not included in the final search strategy by the librarians running these searches probably for a variety of reasons, including repetition of other better terms or they were terms that would not add anything to the search results. This is another facet of the query construction process in which the work of the librarian shows through. Following the same scenario as above, if the patron for instance, had access to and knowledge of databases and complex searching strategies, and actually had the time and resources available in order to do the searching themselves, the inclusion of these discarded terms may have returned less relevant or completely irrelevant articles. Judging the worth of the terms for inclusion in the search strategy is another way in which searchers aid the library patrons. This finding is reminiscent of the finding in Spink (1995) that intermediaries selected more terms that retrieved relevant results than the user did.

The terms that are found in the final search strategy but not in the initial information request can be called the “added terms” because these are terms that librarians added to the search, taking into account those terms provided by the patron, in order to formulated the final search strategy. The total number of terms found in the FSS but not in the IIR was 371. These terms account for slightly over half (52.32%) of the terms included in the final search strategy, the rest of which came from the initial information request category. This finding means that roughly half of the terms in the final search strategy came from the patron and half came from the librarian. Again, it is important to emphasize that the terms that are attributed to the librarian are those terms that the librarian himself or herself added to the search whether from his or her own faculties, knowledge of indexing practices in particular databases, or due to other factors.
The last category in Study 1 is the adding of synonyms by the librarian for terms that were found in the initial information request. It was determined that 54.18% of the terms in the final search strategy that are not in the initial information request were synonyms for terms in the IIR. What this means is that of the terms that the librarians use in the final search strategy, a portion came from the IIR, and a portion came from terms that the librarian added to the search. Of the portion of terms that came from the librarian, about half were synonyms of terms that the patron gave in the IIR, and about half were terms that were related to the search, but emanated from the librarian. Hsieh-Yee (1993) discusses the ability of experienced searchers to add more synonyms than inexperienced searchers, which is reflected in the findings of the current study. While adding synonyms is a very important function in order to address the variety of words to describe topics in various databases, it is not the only method by which librarians add terms to the FSS. Other terms that address related topics or those that approach the topic in a broader or narrower way are added by the librarian in order to encompass the topic in the best way possible to meet the patron’s information needs.

Study 2

The number of forms included for analysis for Study 2 was 5, this small population size being attributed to the short data collection time period constrained in the interest of the author’s research. The useable population size also may be attributed to variations in the number of searches that are submitted to the library, which may be impacted by a myriad of factors, such as fluctuations of research projects at different times of the year, vacation time taken by the researchers, and budget issues affecting researchers that may have restricted the amount of new research they can do. While the
entire population was small as well (12 forms), the proportion of forms selected for inclusion was not terribly different from that of Study 1. In Study 1, for example, 240 of the 543 forms were included for analysis, or roughly 44%. In Study 2, 5 of the 12 forms were included for analysis, or almost 42%. In Study 1, 303 of the 543 forms were excluded, comprising about 56% of the total forms; in Study 2, 7 of the 12 forms were excluded, which amounts to roughly 58%. So while there were only 5 forms available for analysis in Study 2, this number is not far off with respect to the total population size, using Study 1 as a baseline.

The five categories that encompassed the initial information request in Study 2 deserve some further discussion, which will touch on how these categories met expectations and how these categories were used in term analysis. To reiterate, the five categories that corresponded to questions on the revised searching form were: Times searched on this topic, Topic Terms, Topic Related Terms, Use of Information, and Any Additional Keywords. The author expected that most, if not all of the fields would be filled out by patrons submitting online literature search requests. As the Results section points out, however, this was not the case, as none of the five searching forms included had every question answered and every section filled out. The reasons for this vary from a particular request not needing to occupy every space (for instance, if all keywords were already written down, the last category of Any Additional Keywords would not need to be filled out) to patrons not understanding the questions, or possible lethargy on the part of the patron filling out the form.

Some of the responses that these questions elicited on the revised searching form were somewhat different than the author expected. The categories to which answers were
not surprising were the Times searched on this topic, Topic Terms, Use of Information, and Any Additional Keywords. The answers to these questions were straightforward, the two categories of Topic Terms and Any Additional Keywords being those used as the IIR source of terms, and the remaining two categories being more informational in purpose. The Times searched on this topic gave choices on a Likert scale from which patrons had to select, and the Use of Information category was intended to help librarians know what the end use of this research was to be.

Unlike the Times searched on this topic and the Use of Information categories, which were not expected to gather new terms to add to the search, the remaining category of Topic Related Terms was expected to be a great source for additional terms. This category on the revised form asked patrons “What do you already know about the topic that is relevant to the search?” which corresponded to the question of “Describe what you already know about the topic” in the study completed by Kelly, Dollu, and Fu (2005), which was a question that “resulted in the lengthiest responses from users” (p. 462).

Instead of eliciting more terms for query expansion, this category extracted responses from patrons that were more akin to the “informational” categories. Two patrons wrote, for example, which articles they had already found in their research, and another wrote that the librarian should assume that the patron herself knew nothing about the topic of the search in order to get a broad overview of the topic. So while this category did not gather more terms for possible inclusion in the final search strategy, it served as another contextual source of information for the librarian which also plays an important role in the relevancy and number of database results. Thus, the only two categories which were a source of IIR terms were Topic Terms and Any Additional Keywords. The results for
this category and those that follow will be discussed in comparison to the results for Study 1.

Comparison of Study 1 and Study 2

The revised form appeared to “work” and simultaneously “not work” according to the original research question: How much overlap in search terms was there between a patron’s written initial information request and the final search strategy run by librarians at the Library of the United States Environmental Protection Agency in Research Triangle Park, and did the revision of the online literature search request form result in a greater overlap? Some of the more interesting comparisons are detailed below.

In comparing the terms that were given as part of the Initial Information Request, the searches in Study 2 had a higher per search average: 14 terms, versus 10.79 terms in Study 1. Remember, the IIR in Study 1 contained both patron and librarian generated terms; removing the librarian generated terms from the equation yields an average of 10.16 patron-only given terms per initial information request. Another way to compare these results would be to compare the category of Topic Terms from Study 2, which most closely resembles the question used in Study 1, with the IIR from Study 1. Even when comparing this category in Study 2 to the IIR category in Study 1, there was an increase in the average number of terms given by the patron: 11.6 terms in Study 2 versus 10.79 terms in Study 1.

Calculating the term averages in any of these ways yields a lower average of terms in Study 1 than in Study 2. This means that the revised form gathered more terms than the original form about the search topic the patron was requesting. Whether this increase in terms is due solely to the revised form or due in part to the nature of the
requestors or the search, the increase still lends support to the worth of the revised
searching form in eliciting more information up front from the patron submitting an
online search request. The two categories of Topic Terms and Any Additional Keywords
elicited more information than the single blank space in the middle of the original
searching form. The layout of the form may also have played a role in gathering more
information because there were more questions asked and more space provided for the
patrons to expand upon their topics of interest. Perhaps the existence of pointed
questions and/or the availability of more space on the form on which to write about the
topic encouraged patrons to write more. It was demonstrated in Study 2 that patrons gave
a higher average number of terms as part of the initial information request using the
revised form, providing the librarian with more information at the inception of the
literature search request.

The averages for the rest of the categories for Study 2 were slightly lower than
their counterparts in Study 1, as can be seen in the Results section, by between about one
to three terms. One of the more consistent findings was from the last category; that of the
proportion of synonyms versus other terms added to the search strategy by the librarian.
Of the terms librarians added in Study 1, 54.18% were synonyms for IIR terms, whereas
in Study 2, 51.6% of the terms added were synonyms for IIR terms. The relatively small
amount of variability in this category between the two studies points to the librarian’s
role in adding terms to the search strategies, about half of which are synonyms for IIR
terms and about half of which are other terms that address the search topic.

The small differences in the statistics for these categories may be due to several
factors. The small population of data available for analysis for Study 2 most likely
impacted the results for each of the categories of terms. The time frame for data collection was much smaller for Study 2 (2 months) than for Study 1 (6 years), which may account for some of the difference between the two, both in terms of time elapsed and in the amount of data that could be collected. This drastic change in the amount of available data for analysis may have impacted the results for Study 2, as any “outliers” as it were, would influence the Study 2 dataset much more so than the Study 1 dataset, because of the much larger sample size. One factor that may have had an impact on the number of searches initiated was the reorganization and relocation of one of the major departments within the EPA-RTP location. During the time frame in which the revised form was put into use, one department was concerned with a large internal relocation of employees and this focus on other pressing matters may have contributed to the small number of searches submitted. Another factor in Study 2 may be that the particular searches requested during this two month time period tended to include more or less terms due to research fluctuations, patron experience or inexperience in the online searching process or due to other factors.

Another cause for variation in the data of Study 2 might have been the introduction of the new searching form, and the resulting adjustment period that patrons must go through to learn about and use the new form. The longer the revised searching form is used may impact the completeness with which patrons fill out the form and so the form’s usefulness may increase after a certain period of time. Chance also could have played a role on the variation in results between studies or within the studies themselves.

Overall, the revised form was more useful than the original form for gathering more terms from the patron at the time of the submission of the literature search request.
While the overlap in Study 2 was not more than the overlap in Study 1, which was one of the points of focus in the research question, the overlap and other categories analyzed were not tremendously different between the two studies.
Conclusions

The goal of this research was to determine how much overlap in search terms was there between a patron’s written initial information request and the final search strategy run by librarians at the Library of the United States Environmental Protection Agency in Research Triangle Park, and if the revision of the online literature search request form result in a greater overlap. An analysis of the terms contained on completed searches which used the original online literature search request form was performed in order to gather information about the process and also to obtain a baseline for how the online searching process had been done in the past. A revised online searching form was introduced and data from the first two months of its use was analyzed in the same way as the preceding body of data. Results were then compared between the two studies, showing that the revised online searching form gathered a higher average of terms from the patron at the initial information request stage of the literature search process than the original search request form. While there was not a greater overlap between the patron’s initial information request and the final search strategy for the revised searching form, more terms were elicited from the patron at the onset of the search and the other categories of terms analyzed were not that different between studies.

The reason for this analysis of terms was to provide a better understanding about the number and types of terms that are presented to the searchers by the patrons, who have the information need and are also ostensibly more grounded in the subject matter than those who are doing the searching. The patterns about terms that are written on the
online literature search request form and used for the final search strategy can offer the librarians information about what types of terms patrons have a history of offering, and which types of terms either need to be elicited from the patrons themselves, or gathered from other sources in order to devise a search strategy.

Further research on the long term efficacy of the revised online search request form needs to be done in order to provide the best searching form possible. In addition to testing the form’s ability to gather terms for a longer period of time, questions on the form could be revised or added to if need be. Using the revised form in other settings may improve the searching process elsewhere and would serve as an interesting comparison to its use in its current setting.

There is much support for this type of research question in the literature because information retrieval as it pertains to patron-searcher interaction and the queries formulated as a result of that interaction is one of the important facets of library work and providing services that will be useful to the patrons. The desired outcome of this study was to provide greater understanding about the entire process for the searchers themselves, and through this concrete knowledge, the searchers may be able to provide the patrons with a better searching experience that is more streamlined and more useful to the patrons whose work the library supports.
References


