USER MISCONCEPTIONS OF INFORMATION RETRIEVAL SYSTEMS

by

Hsinchun Chen
and
Vasant Dhar

Information Systems Department
624 Tisch Hall, 40 West 4th Street
New York University
New York, NY 10003

March 1988

Center for Research on Information Systems
Information Systems Area
Graduate School of Business Administration
New York University

Working Paper Series

CRIS #173
GBA #88-20

User Misconceptions of Information Retrieval Systems

Hsinchun Chen and Vasant Dhar
Department of Information Systems
New York University
New York, N. Y. 10003

Abstract

In this paper, we report results of an investigation where thirty subjects were observed performing subject-based search in an online catalog system. The observations have revealed a range of misconceptions users have when performing subject-based search. We have developed a taxonomy that characterizes these misconceptions and hypotheses about the causes of the misconceptions. Directions for improving search performance are also suggested.

Track: SCIENCE

Topic Area: COGNITIVE MODELING

Table of Contents

1. Introduction

While archival information sources such as libraries are becoming more computerized, access to such information is often difficult because of the indeterminism involved in the process by which documents are indexed, and the latitude users have in choosing terms to express a query.

Most online catalog systems offer capabilities for access using *known items* such as author, title, and call number, and *non known items* such as access based on subject area. While known-item search is easy to support, subject-based search can be difficult, often requiring the assistance of a reference librarian. For these types of queries, the problem of finding relevant documents can be difficult for two reasons:

1. it requires knowledge about the classification scheme (such as the Library of Congress Subject Headings or other indexing scheme) used in the retrieval system, and

2. the query itself may not be well defined because the user is not clear about the subject for which answers are being sought.

Our research goals are twofold. The first goal is to understand how reference librarians help users with subject search. The second goal is to understand the problems users have using computer-based retrieval systems for subject search. In a previous paper [5], we described a model of librarian-user interaction that showed the strategies employed by librarians to structure subject-based queries. Since then, we have focused primarily on the second goal. In this paper, we report results of an investigation where thirty subjects were observed performing subject-based search. The observations have revealed a range of misconceptions users have when performing subject-based search. Our goals in this paper are to

1. propose a taxonomy that characterizes these misconceptions,

2. develop hypotheses of the causes of misconceptions, and

3. provide directions for alleviating the problems we have observed in subject-based search.

We are not currently in a position to propose how the misconceptions might be detected by a computer-based system during the course of a dialog. This question is the focus of ongoing research.

2. Subject Search

The importance and difficulties of subject-based search have been well documented in the literature. In one card catalog study, subject access was found to constitute about 40 percent of catalog use [16]. Another study on online catalogs showed that subject access constituted between one-third and a one-half of all searches [14]. Even though keyword matching capabilities have been incorporated in many online catalog systems, the need for additional subject-related search capabilities has headed the list of desired improvements in several studies [17] [13] [15].

Several approaches have been proposed to alleviate difficulties involved with subject-based access. The National Library of Medicine's CITE public access online catalog offers natural language query input, automatic medical subject headings display, closest match search strategy, ranked document output, and the use of dynamic end user feedback for search refinement [7]. Other directions include weighted computations of relatedness between query and fiction records [18], more extensive linkages between fields in different records for browsing and navigating a database [19], and the application of the "hypertext" concept to catalogs, that is, breaking the linearity of the traditional file structure and providing links in a variety of different directions in records [11]. In addition, "intelligent front ends" [30] [22] [10] [9] and "expert systems" [26] [27] [29] [24] have been built in several restricted domains.

While the above approaches are recognized as being important improvements to subject search, there are surprisingly few theoretical or empirical analyses about the real causes underlying the problems with subject based access. More recently however, Bates [3] has proposed a framework based on general systems theory for understanding the problems associated with subject search. The framework highlights the role of uncertainty and indeterminism in the process of both indexing and searching. According to this framework, the problems associated with subject search can be understood more clearly by considering the indeterminism involved in the process of classification and in the terms searchers use in attempting to formulate a query, as well as those features of classification schemes that lead to reduced redundancy in cataloging. Specifically, the factors that make subject search problematic are as follows:

Uncertainty: The indexing of most catalogers is partly indeterminate and probabilistic. Evidence suggests that different indexers, well trained in an indexing scheme might assign indices for a given document differently. It has also been observed that an indexer might use different terms for the same document at different times [12] [28]. An even higher degree of uncertainty has been observed in users' searches. One study revealed that on average, the probability of any

two people using the same term to describe an object ranged from 7 to 18 percent [8]. In summary, the evidence suggests that there is considerable latitude involved in: i) the classification of a document into a particular category, and ii) the term a searcher might use to describe a subject area.

Matching: Due to the uncertainty in indexing and searching, generating an exact match between the user's term and that of the indexer becomes difficult. Bates [3] argues that for a successful match, the searcher must somehow generate as much "variety" (in the cybernetic sense, as defined by [1]) in the search as is produced by the indexers in their indexing. The variety produced by an indexer can also be viewed as redundancy in the sense that it consists of partially overlapping meanings applied to a document. To increase the chances of a successful match, there should be a number of labels for each document. This requires preserving the redundancy (generated by the indexer) associated with each document. In practice, however, catalog systems discourage redundancy for the following reasons:

1. Whole document indexing: The cataloger working according to the Library of Congress or some other scheme is trained to index the whole document, not parts or concepts within it.

2. Uniform Heading: The principle of uniform heading holds that for any description there is to be one and only one heading reflecting that description.

3. Specific Entry: Each book is to be entered under a category (heading) which is specific to the content of the book, neither broader or narrower in scope than the scope of the book's contents.

4. Limited cross-reference structure: cross references are frequently an afterthought to "augment" the basic catalog organization [2].

In summary, these tendencies to reduce redundant access points, decreasing the likelihood that a user will generate the right term for retrieval.

Rigidity: While indexers use the rule of specificity for indexing, users tend to approach a search by specifying broader terms first. There might be several reasons for this. One hypothesis is that users often do not have "queries", but what Belkin calls an "anomalous state of knowledge" [4]. Users often expect to refine (narrow down) this anomalous state into a query, *through* an interactive process. However, the organization of a catalog or a system does not always facilitate this type of query refinement. In contrast, reference librarians appear to be particularly adept at this function. Another reason for the "broader terms first" phenomenon that we described in [5] is

that a user might be unsure about the librarian's knowledge about the subject, and might consider it more prudent to start with the general terms first. While reference librarians are cognizant of this phenomenon and press the user towards focusing the query, the user has no such direct support from a computer system.

3. Research Design

The NYU online catalog system, Bobcat, lists over 600,000 catalog records including all new materials purchased after 1973 and many older items previously listed in the card catalog; journals are not listed under Bobcat. The system provides seven search options, namely, title search, author search, combination of author and title search, subject search, number search, keyword search, and Boolean search. These options are available in most online catalog systems. Figure 3-1 shows the initial screen of the Bobcat system.

```
170 BOBST LIBRARY              - GEAC LIBRARY SYSTEM - ALL   *CHOOSE SEARCH
 What type of search do you wish to do?
  1.  TIL - Title, journal title, series title, etc.
  2.  AUT - Author, illustrator, editor, organization, etc.
  3.  A-T - Combination of author and title.
  4.  SUB - Subject heading assigned by library.
  5.  NUM - Call number, ISBN, ISSN, etc.
  6.  KEY - One word taked from a title(TILK), author(AUTK) or subject(SUBK).
  7.  BOL - Boolean search on title, author, and subject.

Enter number or code, then press CARRIAGE RETURN
```

Figure 3-1: The Search Options in Bobcat

Thirty business school students ranging from Ph.D. candidates to freshmen participated in the study. These subjects were asked to perform a search for documents within a subject area of their interest. In general, the most frequently chosen option was 4 above, followed by 6. But there were also subjects using known-item search options to perform a subject-based search. That is, they used the title (or portion of the title), author, or call number to find documents within certain subject area. Subjects were asked to write down briefly, what they were looking for. Subjects were also asked to think aloud during the interaction. This protocol was tape-recorded, and the interaction between the user and the system was logged. Most of the interactions lasted between 5 and 40 minutes, with the median of about 15 minutes. After the inter-

action, they were asked a few follow-up questions pertaining to the search process and the problems encountered during search. Suggestions were also elicited about how the system might be improved.

4. A Typology of Misconceptions

We define a misconception as one where a user performs an erroneous action, uses erroneous terms, or goes about achieving a goal using an erroneous or suboptimal procedure. Under this definition, a lack of knowledge about something (ignorance) is also treated as misconception. While this definition stretches slightly the true meaning of misconception, we adopt it for simplicity.

The logs (and to a small extent the protocols) revealed between one and seven misconceptions per user. Some of these misconceptions precluded users from finding relevant material while others prolonged unproductive search. Three broad categories of misconceptions were identified. The first category includes misconceptions about the subject area itself. The second category includes misconceptions about the classification scheme. The third type of misconception is about the system's capabilities. In the remainder of this section, we describe these misconceptions, presenting examples for clarification. We do not provide tables of statistics for each type of misconceptions since these would increase significantly the length of this paper. These statistics can be obtained from the authors.

4.1. Subject Area Misconceptions

A lack of expertise in the subject area leads to three related problems: not choosing appropriate terms to initiate a search, not having good a priori estimate about how much material there might be for a subject, and not expressing the query at an appropriate level of specificity.

4.1.1. Inappropriate Terms

The LCSH in the system consists of "official" terms. Users on the other hand, use terms they feel best express the "semantic content" of their problem. However, even though the user may be "close" in some sense to the official term, it may not yield any relevant material. To illustrate, one subject looking for books on "measure theory" used the term "measurement" instead. Only after more than 10 minutes browsing did he realize that the screenfuls of citations (which had to do with all kinds of measurement such as pollution measurement, pollen count measurement, etc) had nothing to do with measure theory. In another case, a subject thought "information retrieval

systems" was synonymous to "database management systems", which is not the case from the system's standpoint. Users detected such errors only after a significant amount of browsing.

4.1.2. No Estimate About the Volume of Relevant Work

Problems can also arise when a user does not have a good estimate about how much material exists in a subject area. A common misconception among users, particular Ph.D. students, appears to be that their subject of interest is too specific for there to be books that are directly relevant. If an initial search attempt is unsuccessful, this bias tends to confirm the user's feeling that no relevant material exists. For example, on finding only one citation corresponding to the subject term "career", one subject thought that that was all the material to be expected -- not realizing the fact that there were over 50 citations listed under similar headings such as: "occupations", "professions" and "vocational guidance". Similarly, a subject was convinced that there was no book dealing with the "Contadora peace plan" because this topic is too specific and recent. Actually there were several books in the library that discuss the "Contadora peace plan" (which can be determined simply by using the title search option!).

4.1.3. Expressing the Query at Inappropriate Level of Specificity

A common tendency is one of not expressing the query at the appropriate level of specificity. The use of a "broader entry first" strategy has been observed in other studies [2] [4]. In this investigation, over 70% of our subjects used terms that were more general than they should have been. For example, one subject checked every citation under "statistics" when she actually needed something on "statistical power". In the other example, the subject browsed under "Nicaragua" and "Latin America" instead of the "Contadora peace plan". On the other hand, a minority of subjects used terms that were too specific, e.g. "Dempster-Shafer theory" and "software reusability" when they were really looking for literature on uncertainty and systems/software maintenance respectively. Such requests resulted in no matches. On the other hand, the broader terms first strategy did result in a match, although because of the specific entry principle in LCSH, they often matched documents which were at the wrong level of specificity [3]. One reason for why most queries tend to be expressed too generally appears to be that described in section 2, namely, that users often begin a search with an "anomalous state of knowledge". This tendency is probably reinforced by situations where prior experience with using specific terms results in no matches.

4.2. Classification Scheme Misconceptions

The backbone of the subject search option (SUB and SUBK) is the Library of Congress Subject Headings (LCSH) classification scheme. There were three types of misconceptions about this classification scheme: misinterpreting the terms used in subject headings, not realizing the indexing principles of LCSH, and simply ignoring the existence of LCSH.

4.2.1. Misinterpretation of Subject Headings

In contrast to situations where a user uses incorrect terms -- such as "measurement theory" instead of "measure theory", there are times when the user may in fact use correct terms, which nevertheless yield no matches. Reactions like the following were very common:

"Human factors, no match? This is impossible."

"There should be something under organizational theory.
That is a standard area within organization."

This type of problem stems from lack of knowledge about the classification scheme, no cross-referencing facility within the system, and the incapability of the system to infer synonymous terms.

Finally there was some confusion between title and subject heading. Some users believed there should be a one-to-one correspondence between the title and the subject area. As one user remarked:

"Invariably, the title is a reflection of its content.
If someone has written something which has major contribution
towards, let say, project management, there should be
project management in the title. So if I use TIL or TILK, it
works just as well as SUB."

This search strategy reflects a lack of knowledge about the classification scheme.

4.2.2. Not Realizing the LCSH Indexing Principles

A second problem was the lack of knowledge about the indexing principles of LCSH. There are three principles which were violated repeatedly: the specific entry principle, the whole document indexing principle, and the principle of subdivision. Based on the specific entry principle, subject headings assigned to a document are as specific as possible. However, users tend to think that a document classified under a certain heading should also be classified under a broader heading. The following remark is indicative of this type of misconception:

"If there is something comes out of corporate planning,

it should come out of planning too."

However, according to the specific entry principle, books classified under corporate planning would not be classified under planning.

Secondly, in order to reduce redundancy, LCSH indexers are trained to use a term that indexes a whole book, not a portion of it. This is referred to as the whole document indexing principle. Again, users without a knowledge of the LCSH violated this principle. The following remark illustrates this type of misconception:

"Some books talked about several sub-topics in different chapters. Perhaps I can search from these sub-topics. The system should then suggest these books."

However, since a book is indexed for its entire content, a broader term can be assigned to a book than any of the topics it covers. For example, a book which covers queuing theory, linear programming, and inventory theory is likely to be assigned a subject heading like "operations research", and is therefore not accessible by its specific topics.

Lastly, most users were unaware of the subdivisions in LCSH. Standard subdivisions within subjects headings are features like *topical* (which limit a concept term to a sub-topic), *period* (time), and *local* (like geographic area). Without this knowledge, users tend to explore combinations of terms which for the most part are unproductive. For example, one subject spent about 10 minutes searching for information on the "Contadora peace plan" under the subdivisions of "Nicaragua" and "Latin America". Clearly, the "Contadora peace plan" is highly unlikely to be a standard subdivision given how specific it is (for such queries it makes more sense to use a keyword search based on title -- this would provide materials consisting of documents whose title includes the keyword in any position).

4.2.3. Not Consulting the LCSH

Only one out of the thirty subjects asked for or consulted the two volume LCSH handbooks. Faced with difficulty generating system-recognizable terms, even experienced users did not consult it. If the user looking for information about "human factors" had looked at the LCSH handbook using the term "human factors", he would have found "human engineering" (an official LCSH term) via cross referencing.

We posit that the unwillingness of users to consult the two rather large looking LCSH handbooks stems from the infrequent usage of the system which might discourage the time investment

needed to learn it. It is probably worthwhile to try to incorporate the LCSH knowledge into an online system.

4.3. System Misconceptions

The last category of misconceptions are about the system itself. Twenty-five subjects exhibited erroneous perceptions about the system. System misconceptions are of two types: system's messages, and the system's capabilities.

4.3.1. System Messages

Some messages were misinterpreted by users due to a lack of precision in the system's language. For example, on typing "South Africa, sport" at the level where the system expected the input of a search option resulted the error message: "Your selection not recognized by the system." This message was interpreted by the user as stating that no document found under "South Africa, sport". In general, such misinterpretations arose because of the overally general content of the system's message.

Secondly, users confused the meaning of options. For example, some users were confused between the command PREVIOUS SCREEN which brought back the screen that was displayed previously, and BAC, which scrolled backward in the list of citations or headings. In other cases, users had difficulty distinguishing between the command IND which displayed the list of subject headings and the command CIT which displayed the list of matched documents.

Users also tended to ignore vital information on the screen. In many situations, users were not aware of subject headings printed on the screen which were actually relevant to the query. These subject headings were either displayed along with the titles of the documents (see boldfaced parts of figure 4-1) or as part of a detailed description of a book (see boldfaced parts of figure 4-2). We posit that this occurred because too much incidental information was presented on the screen, causing the user to overlook the important cues.

4.3.2. System Capabilities

There were three misconceptions about the system's capabilities. The first was the lack of clear understanding of the capability of each of the seven search options. Confusion was observed among the SUB, SUBK, TIL, and TILK options. In reality, SUB can only search for the subject headings if the user's terms happen to be in the leftmost position of an official LCSH subject heading. For example, "economic" will match headings like "economic development" but not

```
 170 BOBST LIBRARY              - GEAC LIBRARY SYSTEM -  ALL *AUTHOR SEARCH

Your author: grishman, ralph                          matches   2 citations
    Matches: Grishman, Ralph

Ref# Title                                  Subject                          Date
   1. Analyzing language in restricte>  Sublanguage -- Data processing>  1986
   2. Computational linguistics : an >  Linguistics -- Data processing  >  1986
```

Figure 4-1: The Screen Display of Matched Titles

```
 170 BOBST LIBRARY              - GEAC LIBRARY SYSTEM -  ALL *AUTHOR SEARCH

AUTHOR Logsdon, Tom, 1937-                                  citation  17 of  43
TITLE The robot revolution / by Tom Logsdon.
IMPRINT New York : Simon and Schuster, c1984.
PHYSICAL FEATURES 207 p. : ill. ; 22 cm.
NOTES Includes index. * Bibliography: p. 195-196
SUBJECTS   Robotics. * Robots, Industrial.
LC CARD  85001275
ISBN  0671467050 (pbk.) : * 0671507117
RLIN ID no. : 84-B29297
```

Figure 4-2: The Screen Display of Full Citation Information

headings like "international economic relations". SUBK on the other hand takes only one word but matches all headings which have that word appearing in them, regardless of position. For example, "business planning" can be matched by using "planing" as the keyword in SUBK. The same rules apply to TIL and the TILK. Several subjects who actually used these options interchangeably or did not realize the potential fruitfulness of using one option over others. This misconception reflects a limited knowledge of the system's functionality and a low frequency of use. There is no tutorial that might make clear the differences in capabilities of the search options.

A second major problem related to the system's capabilities was a lack of understanding of the system's match/search method. The system finds documents by matching alphabetically the terms supplied by the user. Six subjects exhibited lack of understanding of this process. For example, one subject typed in "South Africa, sport" after he had already browsed unsuccessfully all headings that had matched "South Africa". In another case, a subject typed in "salt

substitute" which matched alphabetically close headings starting from "salt -- social aspects". She didn't realize there were subject headings before the first screen of the list of headings displayed in front of her that also had to do with other aspects of salt. Typing BAC (to move backward in the list of matched headings), would have shown "salt -- physiological effect", which she was looking for. Finally, the use of operators such as "AND" and "OR" in non-Boolean search seems to suggest that users expect certain set operations to be performed automatically by the system. An example of this type of misconception is apparent in the following subject search (SUB) query: "Dempster-Shafer theory and expert systems" where the user expected the "AND" to be interpreted by the system to perform set intersection.

The last type of problem resulted from a lack of knowledge about "levels" of menus in the system. Basically the system's menus are at several levels as shown in figure 4-3. Users had problems remembering their position during their interaction. For example, instead of using CIT to return to scanning a citation list, many users typed in IND which incorrectly brought them back to the list of subject headings (one level higher than CIT), from where they again had to begin looking for the individual citations within headings. For example, a subject intially matched 107 citations using the term "game theory". After reviewing the detailed citation of a book in this list, instead of returning to the original position in the citation list (popping up one level), he went into the outermost system loop -- restarting the search by choosing SUB, "game theory", and moving forward in the citation list. This process was repeated 11 times. Apart from limited knowledge about how to traverse the system menus, this type of wasteful search is probably the consequence of limited short term memory. In the absence of an indication of the level of the dialog from a system, the user often tends to go back to the top level unnecessarily.

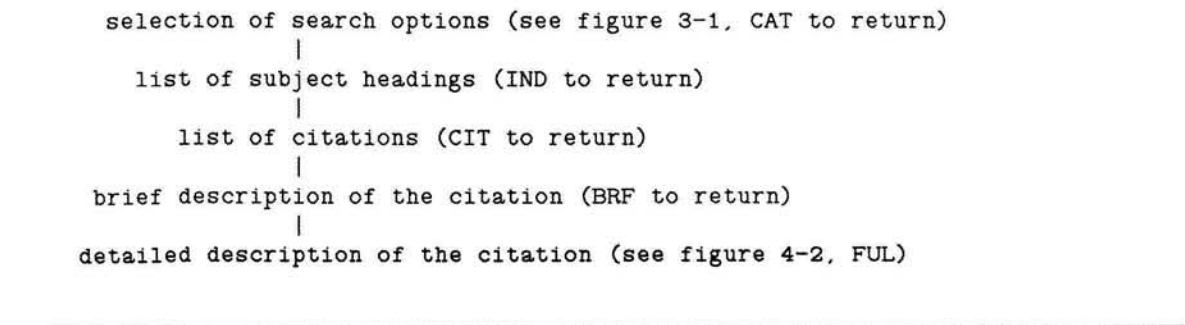

Figure 4-3: The Levels of Menus

Figure 4-4 summaries the hierarchy of misconceptions described above and a table of causes of

misconceptions. Connections from the causes to misconceptions reflect our hypotheses for the misconceptions described in this section. We have not drawn physically the actual linkages since there are too many such connections, which would congest the figure. Instead, the causes for each misconception are indicated in the square bracket listed with each misconception. The numbers in the square bracket correspond to causes in the lower part of the figure.

5. Improving System Performance

Based on our findings, it is clear that subject-based search needs to be better approached by users, and also better supported by computer-based systems. Specifically, there appear to be at least three ways of enhancing search success: via appropriate search strategies, query refinement mechanisms, and a user modeling component within the system.

5.1. Search Strategies

We propose several heuristics which should help avoid many of the obstacles we observed. These heuristics fall into two categories:

Browsing Rules:

1. While browsing, Look for the subject headings of each citation.

2. Browse both forward and backward in the list of headings when a subject heading produces matches. Relevant subject headings may be in the previous or subsequent screens.

3. If the number of matched subject headings or citations is too large, and little or no relevance is observed after browsing a screenful, make the search terms more specific.

Search Options Rules:

1. For subject areas that can be described using one word, use keyword searches (TILK and SUBK); otherwise use the standard title or subject search options (TIL and SUB). This increases the likelihood of a match since a keyword can match in any position.

2. If possible, begin by performing a known-item search first. Use author or title search (AUT, AUTK, TIL, and TILK) to identify one or a few citation and examine subject heading information from the detailed citation. More citations can be obtained by performing subject search (SUB) using the derived subject headings.

3. If the documents matched are too few, try keyword search options (TILK, AUTK, and SUBK) to broaden the search space.

A Taxonomy of Misconceptions:

1. **Subject Area Misconceptions**
   - 1.1. Misinterpretation of the terms [1.1]
   - 1.2. Inappropriate level of specificity [1.1/1.5]
   - 1.3. Not realizing the amount of relevant works [1.1]

2. **Classification Scheme Misconceptions**
   - o Misinterpretation of subject headings
     - 2.1. Treating title as subject [1.2]
     - 2.2. Using unofficial terms [1.2/2.5/2.6]
   - o Not realizing the LCSH indexing principles
     - 2.3. specific entry [1.2]
     - 2.4. whole document indexing [1.2]
     - 2.5. subdivisions [1.2]
   - o Not consulting LCSH [1.2/1.6]

3. **System Misconceptions**
   - o system message/display
     - 3.1. misinterpretation of the system message [2.1]
     - 3.2. confusion about the system options [2.1/2.2]
     - 3.3. missing vital information on the screen [1.7/2.4]
   - o system capabilities
     - 3.4. not understanding top-level search options [1.3/1.6/2.2]
     - 3.5. not understanding the match method of the system [1.3]
     - 3.6. not understanding the levels of menus [1.3/1.4/2.3]

Causes of Misconceptions:

1. **User-Attributed Causes**
   - o knowledge based characteristics
     - 1.1. subject knowledge inadequacy
     - 1.2. low classification scheme knowledge
     - 1.3. low knowledge of system functionality
   - o general human characteristics
     - 1.4. limited short term memory
     - 1.5. prior experience in using the system
     - 1.6. indifference due to low frequency of usage
     - 1.7. carelessness
2. **System-Attributed Causes**
   - o poor general system features
     - 2.1. ambiguous messages
     - 2.2. poor help screen
     - 2.3. non-transparent structure
     - 2.4. non-highlighting vital information
   - o problematic classification scheme
     - 2.5. no cross-referencing of terms
     - 2.6. no semantics in matching

Figure 4-4: Users' Misconceptions and the Causes

These rules should enable users to make the most of the capabilities provided by existing systems.

5.2. Query Refinement

A good example of the translation from an "anomalous state of knowledge" [4] to a precise statement of the problem is exhibited in figure 5-1 where the initial and final problem statements are boldfaced. The intermediate terms cropped up in the course of this unsuccessful interaction. The user progressed from the vague statement of how people "make errors" to that of "human factors in the routine work of data entry". During the interaction, the system's responses and messages provided no interactive support to the user.

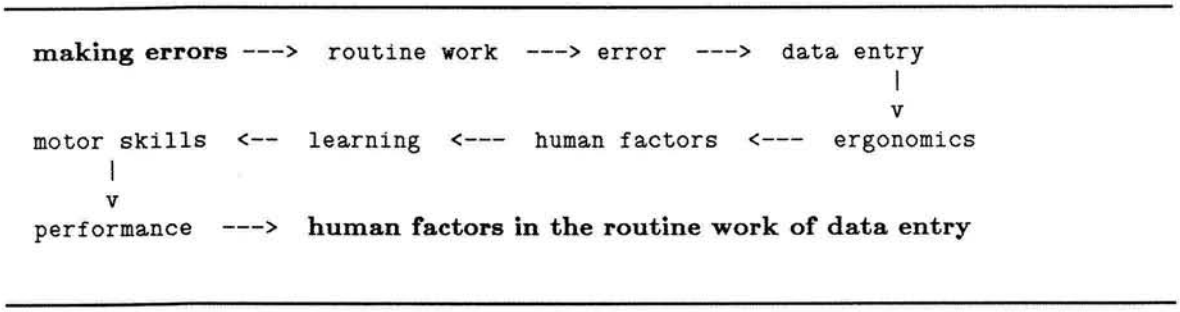

Figure 5-1: Example of Query Refinement

In contrast to the above interaction, we observed reference librarians play a very active role in what we call the term translation process. With the case above, for example, a librarian would actually force a complete specific description of the problem statement, and initiate a terms translation process (probably by consulting the LCSH) with the keywords "human" and "ergonomics". The details of that model are described in [5]. As a next step in this research, we intend to incorporate the librarian's expertise of query refinement into the retrieval system.

5.3. User Modeling

In previous research [5] we identified an important component of the librarian/user consultation model as one where on the basis of a cues, the librarian stereotypes the user. Specifically, hypotheses about a user's information needs are based on the purpose of search and the user's level of education. The advantages of stereotypical user modeling in question answer systems has been demonstrated in [20] [21].

The taxonomy of misconceptions we have described in this paper, provides a sound basis for building an intelligent user modeling component within an information retrieval system. Information about the user and his query can be obtained through a few questions and from an

analysis of the discourse. The system should be able to use such information to create an individual, implicitly-inferred model for a user [21]. The user model we have in mind would be similar to the student model constructed in Intelligent Computer-Aided Instruction (ICAI) systems [6] [31] [25].

The output of the user-modeling of the system could be used in two ways. It could be used to plan an appropriate search sequence, that is, suggest what search options to use and how. As we described earlier in this section, known-item searches can be used to anchor and/or determine terms that can be used to perform a focused subject search. The output of the user-modeling component could also be fed into a terms translation process which would require knowledge about the subject area integrated with the classification scheme into an "extended thesaurus" [23] [24]. Armed with these types of knowledge a retrieval system could become a responsive and intelligent information specialist capable of providing interactive support to a variety of users.

References

1. Ashby, W. Ross. *An Introduction to cybernetics.* Methuen, London, 1973.

2. Bates, Marcia J. "Systems meets user: problems in matching subject search terms". *Information Processing & Management 13,* 6 (June 1977), 367-368.

3. Bates, Marcia J. "Subject access in online catalog: a design model". *Journal of the American Society of Information Science 37,* 6 (November 1986), 357-376.

4. Belkin, N.J. et al. "Ask for information retrieval: Part I. background and theory". *Journal of Documentation 38,* 2 (June 1982), 61-71.

5. Chen, Hsinchun & Dhar, Vasant. Reducing indeterminism in consultation: an cogonitive model of user/librarian interaction. Proceedings of the National Conference on Artificial Intelligence, 1987.

6. Dede, Christopher. "A review and synthesis of recent research in intelligent computer-aided instruction". *International Journal of Man-Machine Studies 24,* (1986).

7. Doszkocs, Tamas E. "CITE NLM: natural-language searching in an online catalog". *Information Technology and Libraries 2,* 4 (December 1983), 364-380.

8. Furnas, George W. et al. Statistical semantics: how can a computer use what people name things to guess what things people mean when they name things. Proceedings of the Human Factors in Computer Systems Conference, Gaithersburg, MD. New York: Association for Computing Machinery, March, 1982.

9. Harris, L. R. "User oriented data base query with the ROBOT natural language query system". *International Journal of Man Machine Studies 9,* (1977), 697-713.

10. Hendrix, G. G. et al. "Developing a natural language interface to complex data". *ACM Transctions on Database Systems 3,* (1978), 105-147.

11. Hjerppe, R. Project HYPERCATalog: visions and preliminary conceptions of an extended and enhanced catalog. Proceedings of IRFIS, 6th, Frascati, Italy, September, 1985, pp. 15-18.

12. Jacoby, J.; Slamecka, V.. *Indexer consistency under minimal conditions.* Documentation, Inc., Bethesda, MD, 1962.

13. Kaske, Neal K.; Sanders, Nancy P. "On-line subject access: the human side of the problem". *RQ 20,* 1 (Fall 1980), 52-58.

14. Kranich, Nancy et al. *A study of user success with an online catalog.* New York University, 1983.

15. Larson, Ray R.; Graham, Vicki. "Monitoring and evaluating MELVYL". *Information Technology and Libraries 2,* 1 (March 1983), 93-104.

16. Markey, Karen. *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs.* , Dublin, OH: OCLC, 1984.

17. Matthews, Joseph R. et al. *Using online catalogs: a nationwide survey.* Neal-Schuman, New York, 1983.

18. Morehead, David R. et al. "The value of information and computer-aided information seeking: problem formulation and application of fiction retrieval". *Information Processing & Management 20,* 5/6 (1984), 583-601.

19. Noerr, Peter L.; Bivins Noerr; Kathleen T. "Browse and navigate: an advance in database access method". *Information Processing & Management 21,* 3 (1985), 205-213.

20. Rich, Elaine. "User modeling via steretypes". *Cognitive Science 3* (1979).

21. Rich, Elaine. "Users are individuals: individualizing user models". *International Journal of Man-Machine Studies 18* (1983).

22. Sager, N.. *Natural Language Information Processing: A Computer Grammar of English and Its Applications.* Addison-Wesley, Reading, MA, 1981.

23. Shoval, Peretz. *An expert consultation system for a retrieval data-base with a semantic-network of concepts.* U. of Pittsburgh, Ph.D. thesis, 1981.

24. Shoval, Peretz. "Principles, procedures and rules in an expert system for information retrieval". *Information Processing & Management 21,* 6 (1985), 475-487.

25. Sleeman, D. H. & Brown, J. S.. *Introduction: intelligent tutoring systems.* Intelligent Tutoring Systems, Academic Press, 1982.

26. Smith, Dana & Steve Hutton. "Back at 8:00 AM -- Microcomputer library reference Support Programs". *Collegiate Microcomputer 2* (November 1984).

27. Smith,Karen. "Robot at the reference desk?". *College & Research Libraries 47* (September 1986).

28. Stevens, Mary Elizabeth. *Automatic Indexing: A State-of-the-art Report.* U.S. Government Printing Office, Washington, DC, 65.

29. Urbanek, Val. "INFORM: library information at your fingertips". *Information Technology and Libraries 1* (December 1982).

30. Walker, Donald. "The organization and use of information: Contributions of Information Science, Computational Lingustics and Artificial Intelligence". *Journal of the American Society for Information Science* , (September 1981), 347-363.

31. Woolf, B. & McDonald, D. D. "Building a computer tutor: design issues". *Computer 17,* 9 (1984), 61-73.