

Apr 16th, 12:00 AM

Systems Versus Users: An Analysis of the Art and/or Science of Search Negotiations

Kathleen T. Bivins
Linköping University

Kathleen T. Bivins, "Systems Versus Users: An Analysis of the Art and/or Science of Search Negotiations."
Proceedings of the IATUL Conferences. Paper 8.
<https://docs.lib.purdue.edu/iatul/1981/papers/8>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.
Please contact epubs@purdue.edu for additional information.

SYSTEMS VERSUS USERS: AN ANALYSIS OF THE ART AND/OR SCIENCE OF SEARCH NEGOTIATIONS

Kathleen T. Bivins

Guest Professor, Linköping University, Linköping,
Sweden

1. Introduction

Each time a user has a problem which necessitates using an information retrieval (IR) system of any kind, each time that happens, a negotiation of some kind must take place. This negotiation can be between the end-user and the system, between the end-user and a librarian/mediator, or between the librarian/mediator and the system. The process is complex and delicate, involving, variously, deductive and inductive logic, combinatorial thinking, knowledge of system structures, knowledge of knowledge structures, personal communicative abilities, guessing, sheer speculation, and, in the end, luck.

All this just to find a book in a library, to take a simple example? Yes and no. In a typical large research library, once a decision is made to add a document of some kind to the collection, the process of deciding where to place it within the context both of its subject content and the system structure must be made. And therein lies the first part of the problem. The matter of access by title, author, and/or other specific information unique to a particular document is not of concern here. What is of concern, of course, is the assignment of subject terms from an indexing language. Generally speaking, indexing languages can be assumed to include classification systems and the various forms of thesauri and simpler alphabetical subject heading systems. The person responsible for the indexing decision is working both from his/her framework, as well as from that of the indexing language. In addition, since it is normal practice for an indexer to be instructed to consider the user of the IR system, a situation exists whereby the indexer is attempting to figure out what a user would be apt to look for in regard to the document in question. Who knows? Or can really say? So, some knowledge (of the indexing language and of the knowledge structure of the document), some technical competence (in applying the "rules" of the indexing language) and some guesswork is involved. Now, the user must do a sort of reversal of this process. S/he must deal with his/her information need in all its vagueness and try to both figure out how the system works as well as how his/her need can be expressed in the terms of that system. The user is really the only one who can truly say what is his/her need, although this is frequently a very difficult process. Indeed, there is some validity in saying that the system itself may well influence that need, perhaps in a way that wouldn't have been the intent of the user if s/he had known of the influence! More often than not, this hapless user has little knowledge of the system, and finds it a rather mysterious, not to say arcane, procedure to delve into a formal IR system. In the situation where the user asks for the aid of an intermediary (mediator), we add yet another dimension, that of the person-person communication. At this stage, one might well wonder how indeed anyone manages to retrieve anything at all from an IR system! Of course, in many cases, users find information which presumably satisfies their needs, and when looking for specific information, that is not so difficult - provided that the IR system has the document in stock. So, what have we arrived at? A person on the one hand indexing from certain sorts of conditions, and

a user searching from other sorts of conditions, and often an intermediary' as a sort of buffer between the two. Having attempted to describe the parameters of the IR process, I shall proceed to discuss IR systems in general, followed by different search and research approaches, along with alternatives and concluding remarks.

2. IR systems

As a generalism, one might say that IR systems are system, rather than user, oriented. If one is discussing, for example, a card catalog for a large research library, it is possible to state that the catalog is constructed by librarians for use by librarians. It is well-known that card catalogs are complex structures which are based on an equally complex series of rules and elaborations thereof. It is also well-known that librarians have long since constructed a rather neat set of principles for the establishment of just such complex card catalogs, and therefore have provided a basis of justifications for these IR devices. Of course, the problem of use of these catalogs is a difficult one. Not only is a card catalog hard to use, one often wonders how much of the elaborately constructed information is actually pertinent. We do not know these answers, despite the fact that the problem is one which is quite frequently addressed. In the same mood, one might question the current machine-readable files of bibliographic references which have been constructed more or less in the same spirit. The wonders of online searching and the like have long since become a fixture of sorts in the information world, but again, the same cavils apply to these types of systems. At best, they provide a union catalog sort of approach to different subject areas, that union catalog being in machine-readable form. One might say that this could provide a speedier mechanism for collecting literature and for searching that same literature. While this is possible, it is more likely that the large machine-readable data bases which exist (many hundreds at this point) probably do not significantly improve service to the end-user.

The systems here generalized about are most certainly document-oriented, and even system-oriented, systems, rather than user-oriented, which may account for their relative failure. The most common means of searching them is none other than that of trial-and-error, certainly the most inefficient method of problem-solving, but at present about the only workable approach available to system users. The aim of the "trial-and-error" method is to produce relevant documents/information (almost invariably documents) with as few trials as possible and with a minimum of error. This is done primarily through manipulating the document surrogates, be they cards or representations in a machine-readable file. In order to attempt to improve performance of the system (and thereby providing the most relevant documents with the least irrelevant ones) many efforts have been made to manipulate (usually) subject terms. One can easily see that this is an effort to improve the system, with, it is to be hoped, benefits to the end-user of that system.

Why has this approach been taken in this field? The answer is at once simple and complex. Simple, because there has long been a number of assumptions made regarding IR systems, and complex, because it has seemed next to impossible to construct IR systems without making these kinds of assumptions. The assumptions made are that relevance (the primary measure for determining "success" of an IR system) is a function of both the document and the question, and that the question represents accurately the information need. Both these assumptions are integral to the vast majority of IR system construction, development, and manipulation. Of course, the validity of these assumptions, while frequently questioned, has never been wholly verified. In fact, the rather large body of literature which has grown up around the assumptions has for many years effectively removed attention from the actual user of the IR system in favor of the system itself and its contents, the

representations of documents. Thus, comments on the "art" and/or "science" of search negotiations must dwell heavily on this aspect of IR systems. If one makes a rather broad assumption that the work done in attempting to improve the effectiveness of IR systems, based on relevance, is "science", then one must also assume that the construction of search languages and the learning thereof by intermediaries (or end-users) has some element of "science" to it. The "art" may indeed be in the ingenuity of intermediaries to "guess" at the mechanics of the system to attempt to retrieve relevant information.

3. Search approaches

In this section, I shall focus on the various ways and means of search negotiating, be it within manual or automated IR systems. I shall by no means attempt a thorough review of the rather vast literature, but shall direct my comments to what seems to be a stream of effort over the years, with digressions on the popular methods of automatic term manipulation.

3.1 Traditional approaches

Two threads are apparent in the literature of the past twenty or so years. First, there are studies involving question negotiation, usually focussing on the "reference" process, and, second, there are information transfer studies which attempt to trace information flows in varied settings among different population groups.

Question negotiation has been the focus of a sizeable number of studies appearing in the traditional library literature (1-2). The majority of these studies focus on the role of the reference librarian as the facilitator of a particular IR system and, as such, I consider them too restrictive to be discussed in any further detail. Other studies, however, attempt to analyze the question negotiation process as a process in itself, and are not necessarily restricted to a particular institutional setting.

For example, Shera's model (3) introduces the importance of the personal characteristics of both the system user and the negotiator, and incorporates a feedback loop. Jahoda and Olson (4) and Rees and Saracevic (5) are particularly concerned with the formulation of the "information need", while Taylor (6) has analyzed the process itself in a more comprehensive manner. In Taylor's work, a potential user of an IR system is seen as an individual who has doubts which can be resolved by information. Formulating a question in regard to these doubts is considered the critical aspect of the search process. Taylor's analysis of question formulation takes the form of four levels: unarticulated need for information, conscious description of the need, formal statement of the need, and the question stated to the formal store. In addition, before the question is actually presented to the formal store, it undergoes five "filters" which enable the negotiator to pre-select relevant information. These filters involve subject determination, user objectives and motivation, user personal characteristics, relationship of inquiry to files, and acceptable answers. Taylor's model, while not invalid, suffers from the assumption that the user has an ability to perceive a gap in his/her knowledge space. Indeed, this assumption has indirectly led to a number of interesting studies which are discussed in another section.

Some relevant studies have been performed which are concerned with analysis of information needs and information-gathering processes. These are concerned with a more generalizable process which is irrespective of the user-librarian negotiation, and are thus grouped under the broader heading of information transfer studies. Harmon (7) has established a four-

stage model of "user information need" that parallels Taylor's model of the four levels of questions. He refers to five variables associated with user information need identified in a study conducted by Cuadra and Katter (8). These include the intended use of the information, differences in the user's background knowledge and "cognitive style", the influence of economic and temporal pressures, the limits placed on the information sought, and user reaction such as "surprise". Harmon also cites a study by Rees and Schultz (7) that suggests the importance of these variables in user information needs.

Harmon's model of user information need comes closer to representing the overall search process than does Taylor's model of the levels of questions. In Harmon's model, the user gathers information, places it in order, perceives the gaps in his/her knowledge and gathers the specific information to fill these gaps. Harmon speculates that the formation and articulation of questions asked differs with each of these four stages. As the user gathers more information and perceives the gaps in the order of his/her knowledge, s/he is better able to form specific questions expecting specific information in response. Harmon further speculates that information systems could be designed to enhance the assistance to the user gathering and ordering needed information during each of these stages.

Kennington (9), in an article on managerial decision-making, outlines four modes of scanning, four different levels of information search applicable to the needs of the searcher. These modes consist of undirected viewing, which is aimed at the user who has not defined an information need, and conditioned viewing, which has the searcher looking for more specific kinds of information. In the third mode, a limited and unstructured search is conducted, while in the fourth mode, a fully structured, fully defined search takes place in a formal system, for specific literature or information.

White (10), in her study of the research behavior of academic economists, identifies three stages of research which can be subdivided into seven information functions. In order of occurrence, these information functions are perception of idea, definition of problem, development of methodology, provision of data, suggestion of information source, analytical assistance, and practical assistance. While her study was not designed for the purpose of examining relevance judgments, the results indicate a general agreement with the findings of others. The economists were increasingly aware of their specific information needs as their projects progressed, and White notes that this acute awareness seemed to peak during the methodology phases of the research projects.

These studies have considerable value, but are limited in that they focus on specific population groups, institutional settings, or interactions with formal IR systems. Let us now have a look at the "non-user"-oriented work to see what is going on in that area.

3.2 Manipulation techniques

In recent years, there has been an increasing amount of research in the area of manipulation of the search language (i.e., the terms employed in the Boolean mode which "match" terms in the document description) of automated IR systems, in order to attempt to provide some more satisfactory results. In order to improve the effectiveness of retrieving documents from the store, a number of theories and models have been proposed, almost invariably based on a form of mathematical reasoning. Fuzzy set theory, utility theory, cluster models, vector space models, and probability theory have all been used to address the problem of IR. Applications in IR systems based on theories and models involve most frequently, term weighting, cluster files, automatic feedback processing, and ranking algorithms (11-13).

There has developed a controversy between the fuzzy set adherents and those favoring probability theory, the two theories which have been applied more often than others. Fuzzy set theory allows for degrees of membership in a set, an admittedly appealing notion in the fuzzy area of IR and relevance, but while it seems somewhat analogous to IR, its very imprecision seems to be a barrier to further understanding of IR systems. Probability theory, on the other hand, is significantly better developed in IR, and an increasing amount of work is being based on it. The central notion here is that it is possible to estimate the probability of the relevance of a given document. If one assumes this, then various activities can take place, such as using the theory to develop indexing and retrieval models based on word occurrences, creating document vectors (one here assumes that the pattern of index terms in relevant documents differs from that in non-relevant documents), and rank ordering according to probability of relevance.

One of the few applications on a large working database is dependent on probabilistic associations being used to find terms associated with yet other terms. These associations are based on term occurrences in a Boolean query retrieval set, and the frequency of terms in a database (14). Another approach uses automatic term weighting based on relevance feedback. Here, an online dialog permits users free term listing, following which system computations weight terms and rank retrieved citations. The user then enters relevance judgments on the ranked display and the system computes new term weights, thus modifying the query (15). These two methodologies are both operating on live systems, as opposed to the experimental work done in more-or-less laboratory environments. At the present time, the use of probability theory with weighting, ranking, and automatic or user-generated feedback, seems quite promising from the system point of view.

In addition, some work is being done on alternative methods for IR. The use of free text, as opposed to controlled vocabulary terms, the use of natural language in "sentence" form, and the use of dialogs are all areas of current interest. Free text searching, as currently practiced, has been employed in the models described above. Most of the uses of natural language in IR are in the form of attempts at dialog situations (16). Invariably, these dialogs are quite rudimentary and it is dubious as to whether they are beneficial for users or not. Other approaches involve different forms of display of the contents of files. One that is conceptually similar to IR thesaurus approaches is the use of document-to-document relationships (citations) coupled with subject terms in a graphic presentation (17). This work is quite interesting, and will soon be implemented in a database.

Now, having presented (albeit briefly) the "scientific" approach, we return to the "art" approach, which is now being rather interestingly investigated.

3.3 Cognitive methods

The one large issue which we have not yet addressed except in passing is that of the end-user. As stated previously, end-user oriented (or user-friendly) systems consist by far and large of dialog-based approaches, or attempts at freeing the end-user from Boolean commands and the like. The major drawback to this is the assumption that the system designer and those who provide the input make in regard to the user (to return to our initial comments). It is assumed that the design structure and input choices are appropriate to users (to repeat), however, very little is known about that user. If the computer or other IR system is the intermediary, then what is critical is the dialog between producers and users. This dialog is the means by which the capacity and capability of the system can be best utilized.

If one can discern something of the process by which users (whether end- or intermediate) search for information, then it may be possible to implement other than the current assumption-based systems, which cannot be considered as more than ad hoc representations. In discussing user-system interactions, what is really being addressed is the fact that very often (and to some degree, always) systems are being used in ways that do not agree with the assumptions on which they are based. It may be that some form of user-system dialog will be the most effective means of interaction. A dialog which attempts both to represent the assumptions of the system as well as the assumptions of the users may well be an appropriate step. Some work along these lines is currently underway in IR (18-22).

In my opinion, the most exciting contributions to the literature are in the area of cognitive processing of system users. There are a few studies here, which consist of both theoretical and experimental efforts aimed at an understanding of what it is that occurs during the search process. This cognitive approach has been taken by Ingwersen and Kaae (20), Mark Pejtersen (22), Belkin and Oddy (18) and Bivins (19-20) among others. Significant elements of these studies include the use of the "thinking aloud" technique (23) combined with various analyses of data so collected. It is most interesting that the "thinking aloud" technique is gaining both popularity and credibility among researchers in this field. While long-validated as a methodology, it of course suffers from the problem of great obtrusiveness. However, much data can be generated through the use of verbal protocols (another name for data generated through this means), the majority of which come from tape recordings, sometimes combined with video recordings.

Although Ingwersen et al. did not pioneer the method in this field (24), their work is certainly the most thoroughly performed and documented, and should serve as a model and a pattern to current and future researchers. These researchers, at the Danish Library School, were primarily concerned with attempting to find out something of patterns and the like in search method/routines of public librarians and/or students searching in public libraries, with an eye towards improved educational methods in the library school. Their careful work and extensive data analysis has led to more experiments and has shown promising leads for future studies.

Mark Pejtersen, also of the Danish Library School, using the same methodological technique, was able to develop a new kind of classification scheme for fiction in public libraries. Based on analyses of verbal protocols, she developed a scheme which involved types and characteristics of fiction works, and which was used by librarians in directing users to what they wanted. An unexpected side result of this was that many of the books which had not previously circulated were borrowed by readers! This is a small, but convincing, piece of evidence for developing user-oriented systems. Readers really borrowed books which were classed (in a multi-dimensional way) through a scheme developed from their (the readers') wants and needs.

Belkin and Oddy have been working on a design study for a system which would be based on a model other than the traditional one. They first postulate that IR system users have an anomalous state of knowledge (ASK), whereby they cannot specify exactly what they need, and therefore can't be expected to "match" that ASK to the specific terms required by an indexing language. Once again, using the "thinking aloud" technique, users of a particular library were queried as to their reasons for coming there. The recorded information was transcribed into verbal protocols, which were then subjected to rather rudimentary linguistic analyses, resulting in the development of what can be termed association networks revolving around the central concepts expressed in the protocols. Next, the subjects queried were asked to

verify the networks, which were accordingly adjusted. In order to attempt to construct a novel IR system, Belkin and Oddy performed a similar linguistic analysis on abstracts of certain articles, sent the resulting networks to the authors, and, again, made adjustments as necessary. They then had the beginning of what would be an IR system - document concepts expressed as networks of associations, and representations of verbal protocols expressed in a similar way. It is obvious, of course, that the proposed system would attempt to develop an ASK of the user (through word co-occurrence strengths and so forth) and then search its other networks for, presumably, relevant information. It will be most interesting to see whether, in fact, and how, if so, such a system would actually work.

Bivins has performed some experiments aimed at determining a level of indexing expected by users versus the actual level used in the system. This was done by presenting IR system users with a search query, and then asking them to "think aloud" while developing search terms they thought might be appropriate for the query. These verbal statements were subsequently transformed into lists which the same subjects used as points of departure for actual search on the same query in a real IR system. Subjects were again tape-recorded during the actual search. Results indicated an extremely high level of consistency in initial development of search terms, and the expected lack of agreement with terms actually used in the system. This is an interesting finding, and could have some impact on the development of user interfaces to IR systems.

4. Conclusions

Having at some length discussed some of the pertinent literature in the area of search negotiations (and no doubt neglecting other pertinent material!), I shall attempt to pull it together. The continuous stream of literature in this area, whether it be intuitive, system, "scientific", or otherwise-oriented, is a valuable indication of the concern for both investigating and attempting to develop, various kinds of interfaces between the IR system and its users. Have we made progress? And is there any sign of model and/or theory development? To the first question, I would say that yes, we have made some progress, but partially because of the lack of some kind of concerted model and theoretical development, our efforts, on the whole, remain rather diffuse. The encouraging fact of a number of researchers banding together regularly to discuss IR, and, in particular, user-oriented theories, as well as an increasing interest in methodological issues, gives one considerable hope that we may yet develop a serious body of literature and serious students. The steady increase of interest in "cognitive" research in IR is also extremely heartening. This is certainly a fruitful area to investigate now, as witness the developments in many other areas, some of which may surely bear relevance to ours. We are most assuredly still in the "art" stage, but there are signs that we may some day progress to a more ordered stage, perhaps even that of "science"! (26)

References

1. BREED, P.F. Analysis of reference procedures in a large university library. Master's thesis. Graduate Library School, University of Chicago, 1955.
2. COLE, D.E. An analysis of adult reference work in libraries. Master's thesis. Graduate Library School, University of Chicago, 1943.

3. SHERA, J.H. The challenging role of the reference librarian. In Knowing books and men, knowing computers, too. Littleton, Colorado, Libraries Unlimited, 1973.
4. JAHODA, G. and OLSEN, P.E. Analyzing the reference process. RQ, 12, 1977: pp. 148-156.
5. REES, A. and SARACEVIC, T. Conceptual analysis of questions in information retrieval systems. In Annual Meeting of the American Documentation Institute, part II. Washington, D.C., ADI, 1963.
6. TAYLOR, ROBERT S. Question negotiation and information seeking in libraries. College and Research Libraries 29, 1968: pp. 178-194.
7. HARMON, GLYNN. Information need transformation during inquiry: a reinterpretation of user relevance. In Proceedings of the American Society for Information Science, v.7, 33d Annual Meeting. Washington, D.C., ASIS, 1970, pp. 41-43.
8. CUADRA, C. and KATTER, R.V. Implications of relevance for library operations and training. Special Libraries, 59, 1968: pp. 503-507.
9. KENNINGTON, DON. Scanning the operational environment: the librarian's role. Journal of Librarianship, 9, 1977: pp. 261-269.
10. WHITE, MARILYN DOMAS. The communication behavior of academic economists in research phases. Library Quarterly, 45, 1968: pp. 178-194.
11. MCGILL, M.J. and HUITFELDT, J. Experimental techniques of information retrieval. In: Williams, M.E., ed. ARIST, v.14. New York, Knowledge Industry Publications, 1979.
12. RIJSBERGEN, C.J. VAN. Information retrieval. 2d ed. London, Butterworths, 1979.
13. ROBERTSON, S.E. and SPARCK JONES, K. Relevance weighting of search terms. JASIS, 27, 1976: pp. 129-146.
14. DOSZKOCZ, T. AID: an associative interactive dictionary for online searching. Online Review, 2 (2), 1978: pp. 163-173.
15. JAMIESON, S.H. and ODDY, R.N. Low-cost implementation of experimental retrieval techniques for on-line users. Proc. of 4th International Online Meeting. Oxford, Learned Information, 1980.
16. ODDY, R.N. Information retrieval through man-machine dialogue. Journal of Documentation, 33 (1), 1977: pp. 1-14.
17. SMALL, H. Co-citation context analysis and the structure of paradigms. Journal of Documentation, 36 (3), 1980: pp. 183-196.
18. BELKIN, N.J. and ODDY, R.N. Design study for an anomalous state of knowledge based information retrieval system. Final report on Grant S/1 SG 09, 2May78-15June79, to the British Library Research and Development Department, September 1979.
19. BIVINS, K.T. Users of information retrieval systems: a frame-theoretical approach. Proc. 3d International Research Forum in Information Science. Oslo, Statens Bibliotekskole, 1979.

20. BIVINS, K.T. Project Athena: are there basic levels of search? (In preparation)
21. INGWERSEN, P. and KAAE, S. User-librarian negotiations and information search procedures in public libraries: an analysis of verbal protocols. Final research report. Copenhagen, School of Librarianship, 1980.
22. WILSON, T.D., STREATFIELD, D.R., MULLINGS, C., LOWNDES SMITH, V. and PENDLETON, B. Information needs and information services in local authority social services departments (Project INISS). Final report to the British Library Research and Development Department on Stages 1 and 2, October 1975 through December 1977. 2 vols. 1978.
23. MARK-PEJTERSEN, ANNEISE. Design of a classification scheme for fiction. In: Harbo, O. and Kajberg, L., eds. Theory and applications of information research. London, Mansell, 1980: pp. 146-159.
24. VYGOTSKY, L.S. Thought and language. Cambridge, Mass., MIT Press, 1962.
25. CARLSON, G. Search strategy by reference librarians. Sherman Oaks, California, Advance Systems Division, Hughes Dynamics, 1964.
26. ERIKSSON, L. Personal conversation, 1981.