THE DESIGN OF BROWSING AND BERRYPICKING TECHNIQUES

FOR THE ONLINE SEARCH INTERFACE

by

Marcia J. Bates

Graduate School of Library and Information Science

University of California at Los Angeles

Los Angeles, CA 90024-1520

Copyright © 1989

by Marcia J. Bates

ABSTRACT

First, a new model of searching in online and other information systems, called "berrypicking," is discussed. This model, it is argued, is much closer to the real behavior of information searchers than the traditional model of information retrieval is, and, consequently, will guide our thinking better in the design of effective interfaces. Second, the research literature of manual information seeking behavior is drawn on for suggestions of capabilities that users might like to have in online systems. Third, based on the new model and the research on information seeking, suggestions are made for how new search capabilities could be incorporated into the design of search interfaces. Particular attention is given to the nature and types of browsing that can be facilitated.

I. INTRODUCTION

As more and more different types of databases are brought online, the universe of information available to search online is beginning to to resemble the vast array of sources available in manual print environments. From an original emphasis on bibliographic databases, which are the online equivalent of abstracting and indexing (A & I) services, databanks have expanded to the full text of journals and other documents, as well as directories, encyclopedias, and other reference sources traditionally available in libraries. Soon there really will be something approaching whole libraries accessible by computer.

As more types of resource are brought online, however, the searcher has a more complex search environment to consider, both in terms of types of sources to use and search techniques to employ with these sources [1,2]. We need to expand our understanding of these resources and the search techniques to use in them, and at the same time, expand our

view of how the search interface should be designed to assist searchers with their new, complex tasks.

In this article the following is done:

•A new model of searching in online and other information systems, called "berrypicking," is discussed. This model, it is argued, is much closer to the real behavior of information searchers than the traditional model of information retrieval is, and, consequently, will guide our thinking better in the design of effective interfaces.

•The research literature of manual information seeking behavior is drawn on for suggestions of capabilities that users might like to have in online systems.

•Based on the new model and the research on information seeking, suggestions are made for how these capabilities could be incorporated into the design of search interfaces. Particular attention is given to the nature and types of browsing that can be facilitated.

II. A "BERRYPICKING" MODEL OF INFORMATION RETRIEVAL

The classic model of information retrieval (IR) used in information science research for over twenty-five years can be characterized as follows (Compare Robertson [3], especially p. 129):
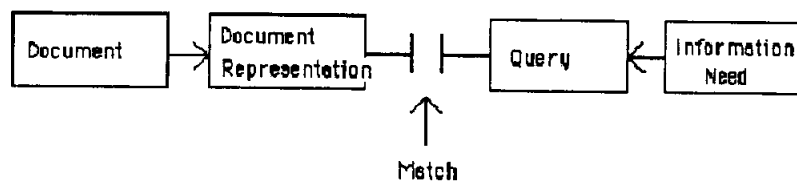

FIGURE 1: The Classic Information Retrieval Model

This model has been very productive and has promoted our understanding of information retrieval in many ways. However, as Kuhn [4] noted, major models that are as central to a field as this one is, eventually begin to show inadequacies as testing leads to greater and greater understanding of the processes being studied. The limitations of the original model's representation of the phenomenon of interest become more and more evident.

It is only fitting, then, that in recent years the above classic model has come under attack in various ways [5-8]. Oddy [9] and Belkin et al. [10] have asked why it is necessary for the searcher to find a way to represent the information need in a query understandable by the system. Why cannot the system make it possible for the searcher to express the need directly

as they would ordinarily, instead of in an artificial query representation for the system's consumption?

At the other end of the model, that of document representation, powerful developments in computing make possible free text and full text searching so that the traditional document representation (controlled vocabulary) takes on a different role and, for some purposes, is less important in much information retrieval practice.

Here I want to challenge the model as a whole--to the effect that it represents some searches, but not all, perhaps not even the majority, and that with respect to those it does represent, it frequently does so inadequately. As a formal model for testing, it has taught us much; as a realistic representation of actual searches, it has many limitations. As a consequence, as long as this model dominates information science thinking, it will limit our creativity in developing IR systems that really meet user needs and preferences.

The model I am about to propose differs from the traditional one in four areas:

1. Nature of the query.

2. Nature of the overall search process.

3. Range of search techniques used.

4. Information "domain" or territory where the search is conducted.

The first two areas will be dealt with in this section and the second two in the next section.

Let us return for a closer look at the classic model. Fundamental to it is the idea of a single query presented by the user, matched to the database contents, yielding a single output set. One of Gerard Salton's [11] contributions to research in this area was the idea of iterative feedback to improve output. He developed a system that would modify the query formulation based on user feedback to the first preliminary output set. The formulation would be successively improved through the use of feedback on user document preferences until recall and precision were optimized.

But Salton's iterative feedback is still well within the original classic model as presented in Figure 1--because the presumption is that the information need leading to the query is the same, unchanged, throughout, no matter what the user might learn from the documents in the preliminary retrieved set. In fact, if a user in a Salton experiment were to change the query after seeing some documents, it would be "unfair," a violation of the basic design of the experiment. The point of the feedback is to improve the representation of a static need, not to

provide information that enables a change in the need itself.

So throughout the process of information retrieval evaluation under the the classic model, the query is treated as a single unitary, one-time conception of the problem. Though this assumption is useful for simplifying IR system research, real-life searches frequently do not work this way.

In real-life searches in manual sources, end users may begin with just one feature of a broader topic, or just one relevant reference, and move through a variety of sources. Each new piece of information they encounter gives them new ideas and directions to follow and, consequently, a new conception of the query. At each stage they are not just modifying the search terms used in order to get a better match for a single query. Rather the query itself (as well as the search terms used) is continually shifting, in part or whole. This type of search is here called an *evolving search.*

Furthermore, at each stage, with each different conception of the query, the user may identify useful information and references. In other words, the query is satisfied not by a single final retrieved set, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search. A bit-at-a-time retrieval of this sort is here called *berrypicking.* This term is used by analogy to picking huckleberries or blueberries in the forest. The berries are scattered on the bushes; they do not come in bunches. One must pick them one at a time. One could do berrypicking of information without the search need itself changing (evolving), but in this article the attention is given to searches that combine both of these features.

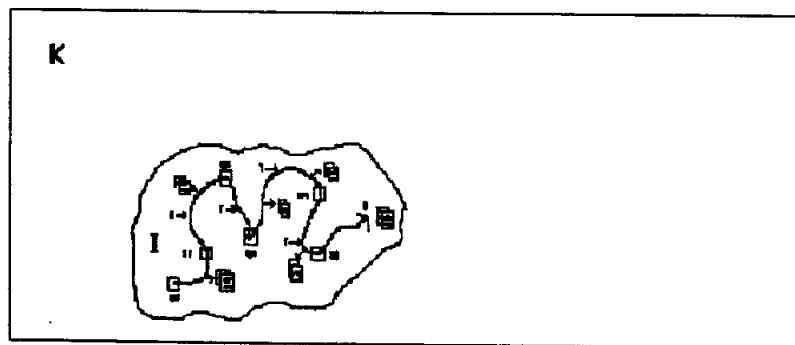Figure 2 represents a berrypicking, evolving search.
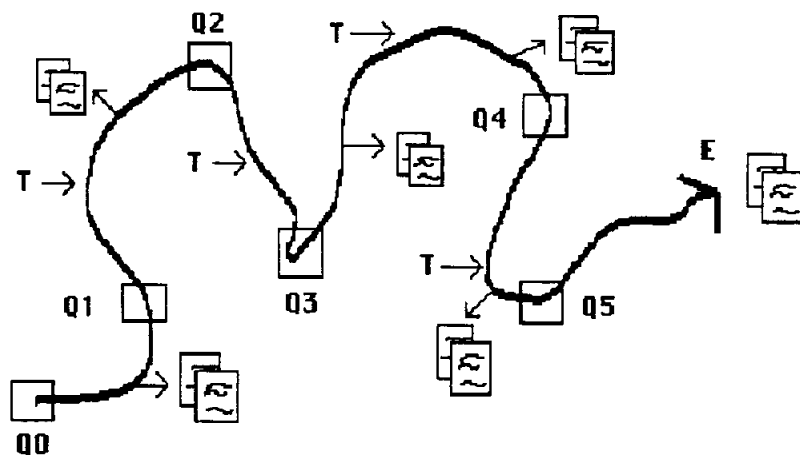
Q = query variation
T = thought
E = exit
= documents, information

FIGURE 2: A Berrypicking, Evolving Search

In Figure 3 we see the size of the picture shrunk in order to show the context within which the search takes place.

K = universe of knowledge
I = universe of interest

FIGURE 3: Content of Berrypicking Search

The focus of the classic model in Figure 1 is the match between the document and query representations. The focus of the model in Figures 2 and 3 is the sequence of searcher behaviors. The continuity represented by the line of the arrow is the continuity of a single human being moving through many actions toward a general goal of a satisfactory completion

of research related to an information need. The changes in direction of the arrow illustrate the changes of an evolving search as the individual follows up various leads and shifts in thinking. The diagram also shows documents and information being produced from the search at many points along the way.

In the case of a straightforward single-match search of the classic sort, we can think of the arrow as being very short and straight, with a single query and a single information output set. Thus, we can see that this model differs from the classic one in the first two respects mentioned above: 1) The *nature of the query* is an evolving one, rather than single and unchanging, and 2) the *nature of the search process* is such that it follows a berrypicking pattern, instead of leading to a single best retrieved set.

There is ample evidence of the popularity of searches of the evolving/berrypicking sort. Reviews of research by Line [12], Hogeweg-de Haart [13], Stone [14], and Stoan [15] attest to the popularity of this approach in a variety of environments, particularly in the social sciences and humanities. A recent landmark study by Ellis [16] on social scientists supports and amplifies the results of earlier studies. Kuhlthau's work [17] with high school students suggests that there is a great deal of exploratory searching that goes on, both before and after a topic for a paper is selected. While the research reviewed here refers largely to the academic environment, I would suggest that many searches by people in many contexts other than academic can also be better characterized by the berrypicking/evolving model than by the classic IR model. The sources consulted may differ, but the process is similar.

III. HOW AND WHERE USERS SEARCH FOR INFORMATION NOW

It was argued in the previous section that information seekers in manual environments use a berrypicking/evolving search mode. In this section we will examine in more detail some of the search techniques used and information sources consulted by users in manual environments.

We might be tempted to say that the path taken in Figures 2 and 3 is simply a series of mini-matches of the classic sort. That is, that at each point where searchers identify documents of interest, they are making a match as represented in Figure 1, and that Figure 2 is simply a representation of searching at a higher level of generality. To make that assumption, however, would be to misrepresent what is being proposed here. Figure 2 is different in essential character, not just in level of generality. Specifically, in a real search there are many different ways people encounter information of interest to them. We will discuss several of them below. Only one of those ways is the kind represented by the classic model.

Users employ a number of strategies. With the help particularly of Stoan [15] and Ellis [16], I will describe just six of them, which are widely used:

• **Footnote chasing** (or "backward chaining" [16]). This technique involves following up footnotes found in books and articles of interest, and therefore moving backward in successive leaps through reference lists. Note that with this technique, as with other citation methods, the searcher avoids the problem of subject description altogether. This method is extremely popular with researchers in the social sciences and humanities. See, for example, Stenstrom and McBride [18].

• **Citation searching** (or "forward chaining" [16]). One begins with a citation, finds out who cites it by looking it up in a citation index, and thus leaps forward.

• **Journal run**. Once, by whatever means, one identifies a central journal in an area, one then locates the run of volumes of the journal and searches straight through relevant volume years. Such a technique, by definition, guarantees complete recall within that journal, and, if the journal is central enough to the searcher's interests, this technique also has tolerably good precision. In effect, this approach exploits Bradford's Law: the core journals in a subject area are going to have very high rates of relevant materials in that area.

• **Area scanning**. Browsing the materials that are physically collocated with materials located earlier in a search is a widely used and effective technique. Studies dating all the way to the 1940's confirm the popularity of the technique in catalog use. Frarey [19], in reviewing three of those early studies, found that use of the subject catalog is divided about equally between selecting books on a subject on the one hand, and finding the shelf location of a category in the classification in order to make book selections in the stacks on the other hand. The latter is, of course, the sort of area scanning described here. Recent work by Hancock [20] again confirms the importance of this approach.

• **Subject searches in bibliographies and abstracting and indexing (A & I) services**. Many bibliographies and most A & I services are arranged by subject. Both classified arrangements and subject indexes are popular. These forms of subject description (classifications and indexing languages) constitute the most common forms of "document representation" that are familiar from the classic model of information retrieval discussed earlier.

• **Author searching**. We customarily think of searching by author as an approach that contrasts with searching by subject. In the literature of catalog use research, "known-item" searches are frequently contrasted with "subject" searches, for example. But author searching can be an effective part of subject searching as well, when a searcher uses an author name to

see if the author has done any other work on the same topic [16].

Until now most of the emphasis in online databanks and other automated IR systems--theoretical, experimental, and operational--has been on use of just one of the above techniques, namely, searching abstracting and indexing services. It is assumed that to do an automated information search, one is searching on a bibliographic database, a list of references with or without abstracts, that is just like an abstracting and indexing service, except that it is online. In experiments, the "document representations" in the classic IR model may involve very sophisticated methods, but most come down to some form of representation of the contents of documents that is usually much shorter, and different from, the documents themselves. In short, most IR research, until a recent flurry of interest in full text databases, has been research on databases of document surrogates.

Real searches, by contrast, use all the above techniques and more, in endless variation. It is part of the nature of berrypicking that people adapt the strategy to the particular need at the moment; as the need shifts in part or whole, the strategy often shifts as well--at least for effective searchers. So, to return to an earlier point, the berrypicking model does not represent a number of mini-matches of the classic sort, i.e., between search term and A & I service (database) term. Rather, the evolving/berrypicking search also involves the third and fourth features mentioned earlier: 3) the search techniques change throughout, and 4) the sources searched change in both form and content.

We have generally assumed in library/information science that the fifth technique in the list above, the A & I search, is clearly superior to the others. That is an important reason for the primacy given given to the bibliographic search in our research and practice. However, Stenstrom and McBride [18] found, when they asked social science faculty where they got the references for journal articles they used, that over 87 percent of them said they got the references from abstracting journals only occasionally, rarely, or never (p. 429). They relied far more heavily on footnote chasing: 69 percent (p. 429). Both Stoan [15] and Ellis [16] provide evidence and are very persuasive on the power and effectiveness of these other techniques for academic researchers and students at the very least.

Some of the other search techniques described above are possible on some systems--see, for example, Palay & Fox [21], Croft & Thompson [22], Cove & Walsh [23], Noerr & Noerr [24]. See also Hildreth's masterly review of intelligent interfaces for bibliographic retrieval systems [25]. Nowhere, to my knowledge, however, are all of these techniques easily applied by a searcher within a single system.

A model containing a unified perspective, incorporating the full range of searcher behaviors in the information seeking process, may make it easier to design many more such

features for information retrieval systems. Ellis [16] has presented the results of his own research on social scientists and, on that basis, argues for the implementation of most of the above techniques, as well as others not discussed here. The particular mix of different capabilities that should ultimately be made available is a question deserving much more attention in the future.

Citation searching is also available, of course, in online systems in the Institute for Scientific Information databases. This searching method is now widely accepted in library/information science as another valuable database approach. Not all readers may be aware, however, of how hard Eugene Garfield had to work in the 1960's and 1970's to persuade librarians of the value of citation searching. I vividly recall observing an otherwise very capable reference instructor telling a class in the late 1960's that a citation index was a waste of money, that it was just a vanity publication for professors--its only value being for them to look up and see who was citing their own work. My point here is that we have yet to fully accept all six of these techniques as valid, effective approaches to information. Even citation indexing, now widely used, was not received easily into the thinking of library/information science.

From the standpoint of general effectiveness in searching, it is clear, on reflection, however, that, other things being equal, the searcher with the widest range of search strategies available is the searcher with the greatest retrieval power. We in information science feel that information searchers should take more advantage of A & I services in online or manual form. We, in our turn, should recognize that these other techniques used so commonly by researchers must have some real value for them, and that there may be times when they are preferable (see Stoan [15]). With each of the six retrieval techniques described above, it is possible to think of instances when that technique is clearly superior to the others as a route to the desired information.

I would argue on two grounds that these techniques should all be available in at least some future automated IR systems, and that our model of information retrieval should include berrypicking through use of these and other techniques:

1. The more different strategies searchers can use an information store, the more retrieval effectiveness and efficiency is possible.

2. There are many experienced searchers who use these techniques already--in a berrypicking mode--with great satisfaction. These approaches represent well established patterns that are handed down from scholars to their students and which work well for them in many cases. If we want to meet users' needs, we should enable them to search in familiar ways that are effective for them.

To summarize the argument to this point, this model of searching differs from the traditional one not only in that it reflects evolving, berrypicking searches, but also searches in a much wider variety of sources, and using a much wider variety of search techniques than has been typically represented in information retrieval models to date. With this broader picture of information retrieval in mind, many new design possibilities open up. In the next section, some of those possibilities will be examined, with particular attention to the role of browsing in the broader search process.

IV. SEARCH CAPABILITIES FOR A BERRYPICKING SEARCH INTERFACE

Browsing. The view of searching as frequently being an evolving/berrypicking process, and one which uses a variety of types of information sources and search techniques, changes our sense of what browsing capabilities should be like in online systems, and how the database and the search interface should be designed.

Concepts of browsing in IR systems are becoming more and more sophisticated. See Noerr & Noerr [24], Wade & Willett [26], Cove & Walsh [23], Hildreth [27], Bawden [28], Ingwersen & Wormell [29]. But there is still a lingering tendency in information science to see browsing *in contrast* to directed searching, to see it as a casual, don't-know-what-I-want behavior that one engages in separately from "regular" searching.

However, as Ellis notes [16], browsing is an important part of standard information searching; he calls it "semi-directed or semi-structured searching" when used this way. He recommends that browsing of a variety of types of information, e.g., contents pages, lists of cited works, subject terms, should be made available in automated systems. He further argues that since the user is doing the browsing, and we therefore do not have to design a cognitive model of user browsing into the system, that providing browsing features should be relatively simple.

Relatively simpler perhaps, but making effective provision for browsing capabilities involves its own complexities. The techniques above combine browsing and conventional use of the information access apparatus in a variety of *specific configurations*. With all of the six techniques above, as well as with other features that might be designed for browsing, it will be desirable to set up combinations of features that incorporate browsing in different ways in each case.

The nature of browsing associated with each of the techniques listed above is examined

in more detail below. Key Design features recommended for automated IR systems will be stated for each technique.

So that there is no confusion, however, I want to emphasize that browsing and berrypicking are not the same behavior. There will be a great deal of discussion of browsing in the remainder of this article, but only because browsing has gotten less attention in our field that other kinds of searching. Berrypicking involves the use of a wide variety of techniques, some of which are very standard, and others which involve a considerable amount of browsing. One of the points emphasized in this model is precisely that people use a wide variety of techniques.

Each of the six techniques is discussed below, followed by some general points about database and interface design for berrypicking and browsing.

_____Footnote chasing:_ In footnote chasing one might want both to be able to browse through the article or book that generates the references as well as through the list of references--in fact, to move back and forth easily between the two parts of the document. The body of information browsed in footnote chasing has a coherence and meaning that clusters around the indiosyncratic purposes of the author of the article or book. Browsing in the footnotes or endnotes will be minimal if the searcher only looks up individual references found in the text, and sticks to them. Browsing of the references can be more extensive if the searcher scans the list, independently of an originating textual reference.

_____*Key design features:*_ User can get the following easily, preferably by direct manipulation, e.g., with mouse and pull-down menus: 1) Overview of document contents--chapter or section headings, 2) Full text of documents and references, 3) Ability to jump back and forth between text and references.

_____Citation searching:_ In citation searching, one might want either to browse the set of references that cite a given starter reference, or read any of the citing articles. No single human created this grouping of citations; rather they come together because they all happened to cite the originating reference; they may otherwise be quite unrelated. Such a collection of references is likely to be stimulating to creativity, as the citing articles may not be on the "same" topic in the conventional sense, yet nonetheless create a grouping that has at least one key thread of similarity that may go along unconventional lines. (See also Bawden [28].) Because of this unconventional grouping, the user might well want to expand the search indefinitely in any direction, that is, upon finding a citing article, learn which articles cite it, and so on.

_____*Key design features:*_ Users should have the ability to 1) Scan lists of citing references, 2)

Make simple single step jumps to a) full text of citing articles, b) full list of references in citing article, and 3) Make jumps in any direction ad infinitum, i.e., the user should not have to "return to go" and reenter a starting article for each jump in any direction.

<u>Journal run</u>: Looking through journals manually, the searcher flips through issues, scanning large chunks of the text of the articles, as well as the contents lists and abstracts. Here the grouping of articles is that subject area represented by the coverage of the journal. When the journal has a very broad subject coverage, such as that of *Science* or *American Psychologist,* it is unlikely to meet a searcher's need for information on a topic of the normal degree of specificity associated with a research project. To put it differently, browsing such general journals is probably useful more for general monitoring of the environment, rather than contributing to a well defined need.

In cases where the journal coverage is a more specific subject area, however, reviewing the contents lists or articles in that journal may be an excellent way to see quickly a large number of articles exactly in the heart of an area that interests one. The grouping of articles that results from their joint publication in a journal can be expected to be coherent and well thought out, since the focus of journals is generally well defined by editors for prospective authors.

<u>*Key design features:*</u> 1) Easy specification of journal title and starting date in a journal run search, 2) Easy jumps between contents lists and articles and back again, 3) Capability of requesting, if wanted, standard section headings in scholarly articles, such as "Methodology," or "Conclusions," so the searcher is shown these sections directly.

<u>Area scanning</u>: This technique is most commonly used with books arranged by a library classification scheme on the shelves of a library. With area scanning, one may either follow the exact arrangement of the classification scheme by reading linearly along the shelves, or alternatively, and, I suspect, more commonly, deliberately not follow that order. In practice, one of the most useful aspects of area scanning is that one can visually scan in a random manner over the shelves in a subject area of interest.

The effect of this latter method is to "jump the rails" of the classification scheme, to skip to other parts of the scheme that are near the starting point, without having to look at every single intervening book and category. This technique represents a deliberate breaking up of the conventional classified order, while enabling the searcher to remain in the same general initial subject area. Thus the search domain may consist of a variety of specific areas within one larger area.

Area scanning is the quintessential form of browsing in manual environments. As noted

earlier, the research shows that it has remained very popular over many years among users. It is reasonable to presume that it meets some real needs. More research into why this approach is popular is desirable. However, here are a couple of guesses: 1) The searcher is exposed to a variety of related areas, some of which, because of the jumping around, may be related in unexpected ways--thus producing serendipitous discoveries. 2) The searcher can look directly at the full text of the materials. By flipping through the pages and reading a passage here and there, the searcher gets a quick gestalt sense of the "feel" or character of the author and his or her approach. Whatever that feel is, it is almost never accessible through any classification or subject description.

Key design features: 1) A library's listing of its books on the shelves arranged by the order of the classification scheme is called a shelf list. Thus, for area scanning linearly along the shelves, a capability of browsing the shelf list can be provided. 2) For "jumping the rails" of the classification scheme, browsing at several levels of generality within the classification scheme itself can be provided, i.e., giving the searcher the option of browsing a list of the most general categories in the scheme, or a list of the general categories plus their subdivisions, and so on, down to the full detail of the scheme. 3) At any point, with either of the first two capabilities listed in this section, the searcher should be able to ask for "snapshots" of full text of books (more discussion later).

Subject searching in bibliographies and A & I services: In discussions of "browsing" in online databases, the term usually refers to reading short lists of alphabetically arranged subject terms or reading citations and their associated abstracts. But, in fact, in such activities, there is little sense of the random visual movement usually associated with browsing. Indeed, the lists of terms printed out are short, and the printing of citations is costly, so searchers often keep it to a minimum. When the cost of printing out abstracts falls, and/or CD-ROM database use become more widespread, true browsing may be easier to do.

It may help the discussion here if we compare the manual form of A & I services, and consider how they are used for browsing. We may be able to do more, of course, with the online form, but let us first see if the text lends itself to browsing in principle. A very common pattern in manual forms of A & I services is to arrange the abstracts by a classified order, and attach a subject index using more specific subject terms. When an online searcher searches by controlled vocabulary, or by free text on the titles and abstracts, all the entries associated with the more specific subject terms are brought together in one location, so they become easy to examine. In the manual form, usually only the abstract numbers are brought together in the index. So grouping entries by these specific terms is a useful function of online services, though the browsing potential is limited for the reasons given above.

Since the A & I services generally arrange the abstracts by a classified order, it is possible

in the manual form to browse through the abstracts in a classified section. This is generally impractical in online databases unless the search is also limited to certain dates or issues of the service, because the online database usually combines many years of the service in one, and each classification category therefore contains very large numbers of items (see Bates [30]). However, in a database in which cost per reference is not a factor, then some sort of browsing in the classified sections might be possible, particularly if brief forms of the reference were printed out, so many could be seen on the screen at once.

_____*Key design features:*_ The user should have the capability of 1) Rapid browsing of many references without cost, and/or ability to ask to see every nth reference in a large set (see further discussion in Bates [31], p. 21ff.), 2) Browsing the classification used in an A & I service, as well as abstracts within each classification, either all or every nth one.

_____Author searching: Author searching makes sense as a form of *subject* searching in that authors tend to write on similar things from one article or book to another. Thus, if one item pays off, maybe another by the same person will too. While bibliographies and catalogs have brought together in one place the references to an author's work since time immemorial, it would be a novel contribution of online systems if they made it possible to see grouped in one place the full text of an author's works. Library stacks do it for books, but there is currently no way to bring together other forms of publication, or to combine book texts with those other forms. When the day comes that full text online becomes very cheap, this grouping of an author's work in one place will be possible. The question in the meantime is, can we design the interface to make it easy to "flip through" the pages of the author's work?

_____*Key design features:* When author searching, the user should have the capability of calling up 1) Bibliographies of authors' works, 2) "Snapshots" of the text of works (see discussion later), and 3) Features that enable footnote chasing and citation searching.

Each of these approaches can be seen as a different way to identify and exploit particular regions in the total information store that are more likely than other regions to contain information of interest for the search at hand. To put it differently, these are different ways of identifying berry patches in the forest, and then going berrypicking within them.

Database and Interface Design. Suggestions for implementing specific design features have been made above. In this section some across-the-board proposals are made for the design of databases and interfaces for browsing and berrypicking:

• To reproduce the above search capabilities, databases will need to contain very large

bodies of full text, as well as different types of text (narrative, statistical, bibliographic references, etc.). At the same time the structure of the databases will need to be such that the searcher can move quickly from one form of information to another, in other words, not have to follow a complicated routine to withdraw from one database and enter another.

    • Several authors have pointed out the value of helping the user of a system develop a mental model or "metaphor" of the system to guide them [32-34].   Various models have been used in the design of interfaces for information systems, for example, Weyer [35] used the book, which approach was also supported by Elkerton & Williges [34] in their research, and Borgman [36] used the card catalog. In teaching students general information searching, Huston [37] has suggested using the model of community-based information networks as a basis for explaining the online literature reviewing process. Hannabuss [38], on the other hand, has argued for a view of information seeking as a form of conversation, especially with reference to the pattern of turn taking in conversation, and those parts of conversation that involve question asking and answering.

    Now that so many different types of information are going online, including much full text, a good place to start as a model of information searching for a berrypicking interface might be the physical library itself. It is the actual physical layout of a library that people are most familiar with, rather than the complex intellectual relationships we develop among catalog entries, books, periodical indexes, journals, etc. Creating a virtual physical layout on the screen may make it easier for the searcher to think of moving among familiar cateogries of resources in an information retrieval system, in the same manner in which they move among resources in the actual library. This may be particular useful at the beginning of a search, when the user could see a physical representation of an imaginary library on the screen. The searcher might then be reminded of whole classes of resource which they might otherwise forget.

    Many years ago, the psychologist George Miller [39] pointed out how very physical our memories are, and how easily we remember things by their physical location. Jones & Dumais [40] challenge the idea that spatial metaphors help information system users recall where something was filed. However, I am suggesting the idea primarily as an orientation device, a way to give users a familiar basis from which to move forward. (See also [41-44], and Hildreth's [25] discussion of the General Research Corporation's "Laserguide" CD-ROM online catalog, p. 90-94.)

    There are many complex issues involved in adapting such a model in an interface, which cannot be dealt with here. Suffice it to say that the transfer will not be simple, and may ultimately be modifed somewhat away from the more literal image of the library as testing proceeds and as users gain greater familarity with computer interfaces generally.

• Browsing in a manual environment is a physical activity, involving body or eye movements of a fairly random character. Thus to be effective in an online environment, a browsing capability should also allow for random movement, at least of the eyes. An aspect of browsing that has been commented upon is the juxtaposition, in time or space, of different ideas or documents that stimulate the thinking of the information searcher ([45], p. 53). To reproduce this in an online environment, it will be necessary to make rapid movement across large amounts of text possible.

The physical metaphor of the library that was suggested above may facilitate such searching particularly well. For example, if the interface can produce a picture on the screen that looks like the books on a shelf, the searcher can transfer a familiar experience to the automated system. If then, a mouse or similar device makes it possible to, in effect, move among the books, a familiar physical experience is reproduced and the searcher can take advantage of well-developed browsing skills. Until the full texts of books are online, the searcher may examine extensive subject information about the book, such as contents lists, index entries, and the like [46].

Once such a form of movement is possible online, it should be transferable to other kinds of information environments where such movement was more difficult in manual situations. For example, the searcher might move among categories of a classification scheme used in an A & I service, or follow up leads of related terms in a high-powered online thesaurus. (See also Bates [8].)

• As noted earlier, the value of flipping through the pages of a book may be due, at least in part, to being able to read passages of a writer's work to get a feel for his or her approach and determine whether it appeals. In large full text databases it will be desirable to be able to do this as well. It would be easy to program a command that would produce a series of randomly selected passages, or "snapshots," each two or three paragraphs in length. Such passages should be truly randomly selected--just as happens whe we flip through an article or book--because it is precisely what is not indexed that we want to sample.

Incidentally, in a recent study, based on a random sample drawn from three different types of libraries, I learned that both reference books and "regular" books use a surprisingly limited and robust set of patterns of organization within the book. These patterns have endured in very stable form over hundreds of years and in many Western cultures [47]. The overwhelming majority of contents lists, for example, are two pages or less in length. Thus, plans to use snapshots of text for browsing purposes, therefore, should not produce nasty surprises in terms of displaying segments of complex or unusual file structures. (I am speaking of the structure of the book as a whole, not of what may appear internally to a diagram or

illustration.)

• The searcher should be able, with a single command, to call for a search mode and screen that is set up for one of the six techniques above (or others). That is, it should not be necessary to issue a string of commands to get the information needed on screen to begin. *Each whole technique should* be built in as a package that the searcher can call upon when desired. Movement through screens should resemble movement through a real-life source using a given strategy (again the physical metaphor). For example, for the searcher doing a journal run, it should be possible to type in a journal title and year, preceded by some phrase such as "journal scan." The contents page of the first issue of that year then appears on the screen. The searcher can then by, say, highlighting a title, easily ask to see the article full text. Another command or highlight sends parts or all of the article to be printed. And so on.

• Hypertext approaches appear tailor-made for berrypicking searching [48]. Being able to jump instantly to full bibliographic citations from references in the text, for example, is a technique that hypertext handles well.

• Berrypicking frequently requires the capability of seeing substantial quantities of information on the screen at once. Screens used should be high definition for easy reading and scanning.

• The interface design should make it easy to highlight or otherwise flag information and references to be sent to a temporary store. Said store can then be printed out when the searcher is ready to leave off searching. The necessity otherwise either to write information down by hand or print out information in bits and pieces interspersed between search commands would be tiresome and would reduce search effectiveness.

V. CONCLUSIONS

As the sizes and variety of databases grow and the power of search interfaces increases, users will more and more expect to be able to search automated information stores in ways that are comfortable and familiar to them. We need first, to have a realistic model of how people go about looking for information now, and second, to find ways to devise databases and search interfaces that enable searchers to operate in ways that feel natural.

A model of searching called "berrypicking' has been proposed here, which, in contrast to the classic model of information retrieval, says that

•typical search queries are not static, but rather evolve

•searchers commonly gather information in bits and pieces instead of in one grand best

retrieved set

•searchers use a wide variety of search techniques which extend beyond those commonly associated with bibliographic databases

•searchers use a wide variety of sources other than bibliographic databases.

Drawing on the research of Ellis [16], Stoan [15], and others, a half dozen typical search techniques used in manual sources have been described (footnote chasing, citation searching, journal run, area scanning, A & I searches, author searches). The specific behaviors associated with these techniques, in particular, browsing behaviors, have been analyzed. Methods have been proposed for the implementation of these techniques in database design and search interface design in online systems.

In conclusion, as Rouse & Rouse [49] note, after an extensive survey of the literature of information seeking behavior:

> Because information needs change in time and depend on the
>
> particular information seeker, systems should be sufficiently
>
> flexible to allow the user to adapt the information seeking process
>
> to his own current needs. Examples of such flexibility include the
>
> design of interactive dialogues and aiding techniques that do not
>
> reflect rigid assumptions about the user's goals and style (p. 135).

REFERENCES

[1]   Martha E. Williams, 'Transparent Information Systems through Gateways, Front Ends, Intermediaries, and Interfaces', *Journal of the American Society for Information Science,* **37**, 4, 1986, pp. 204-214.

[2]   Donald T. Hawkins, Louise R. Levy, and K. Leon Montgomery, 'Knowledge Gateways: The Building Blocks', *Information Processing & Management,* **24**, 4, 1988, pp. 459-468.

[3]   S. E. Robertson, 'Theories and Models in Information Retrieval', *Journal of Documentation,* 33, 2, 1977, pp. 126-148.

[4]   Thomas S. Kuhn, *The Structure of Scientific Revolutions,* 2nd ed. enl., Chicago: University of Chicago Press, 1970.

[5]   D. Ellis, 'The Effectiveness of Information Retrieval Systems: The Need for Improved Explanatory Frameworks,' *Social Science Information Studies,* **4**, 4, 1984, pp. 261-272.

[6]   David Ellis, 'Theory and Explanation in Information Retrieval Research', *Journal of Information Science,* **8**, 1, 1984, pp. 25-38.

[7]   Marcia J. Bates, 'An Exploratory Paradigm for Online Information Retrieval', IN B.C. Brookes, ed., *Intelligent Information Systems for the Information Society.* Proceedings of the Sixth International Research forum in Information Science (IRFIS 6), Frascati, Italy, September 16-18, 1985, Amsterdam: North-Holland, 1986, pp. 91-99.

[8]   Marcia J. Bates, 'Subject Access in Online Catalogs: A Design Model,' *Journal of the American Society for Information Science,* **37**, 6, 1986, pp. 357-376.

[9]   R. N. Oddy, 'Information Retrieval through Man-Machine-Dialogue', *Journal of Documentation,* **33**, 1, 1977, pp. 1-14.

[10]   N.J. Belkin, R.N. Oddy, and H.M. Brooks. 'ASK for Information Retrieval: Part I: Background and Theory', *Journal of Documentation,* **38**, 2, 1982, pp. 61-71.

[11]   Gerard. Salton. *Automatic Information Organization and Retrieval.* New York: McGraw-Hill, 1968.

[12]   Maurice B. Line. 'Information Requirements in the Social Sciences',

       IN *Access to the Literature of the Social Sciences and Humanities.* Proceedings of the Conference on Access to Knowledge and Information in the Social Sciences and Humanities. Library Science Dept., Queens College, City University of New York. New York: Queens College Press, 1974, pp. 146-158.

[13]   H.P. Hogeweg-de-Haart, 'Characteristics of Social Science Information: A Selective Review of the Literature. Part II', *Social Science Information Studies,* **4**, 1, 1984, pp. 15-30.

[14]   Sue Stone, 'Humanities Scholars: Information Needs and Uses', *Journal of Documentation,* **38**, 4, 1982, pp. 292-312.

[15]   Stephen K. Stoan, 'Research and Library Skills: An Analysis and Interpretation', *College & Research Libraries,* **45**, 2, 1984, pp. 99-109.

[16]   David Ellis, 'A Behavioural Approach to Information Retrieval System Design', *Journal of Documentation, in* press.

[17]   Carol Collier Kuhlthau, 'Developing a Model of the Library Search Process: Cognitive and Affective Aspects', *RQ, 28*, 2, 1988, pp. 232-242.

[18]   Patricia Stenstrom, and Ruth B. McBride, 'Serial Use by Social Science Faculty: A Survey', *College & Research Libraries, 40*, 5, 1979, pp. 426-431.

[19]   C.J. Frarey, 'Studies of Use of the Subject Catalog: Summary and Evaluation', IN Maurice F. Tauber, ed., *The Subject Analysis of Library Materials,* New York: Columbia University, School of Library Service, 1953, pp. 147-166.

[20]   Micheline Hancock, 'Subject Searching Behavior at the Library Catalogue and at the Shelves: Implications for Online Interactive Catalogues', *Journal of Documentation, 43*, 4, 1987, pp. 303-321.

[21]   Andrew J. Palay and Mark S. Fox, 'Browsing through Databases', IN R.N. Oddy, et al., eds., *Information Retrieval Research,* London: Butterworths, 1981, pp. 310-324.

[22]   W.B. Croft and R.H. Thompson, 'I^{3}R: A New Approach to the Design of Document Retrieval Systems', *Journal of the American Society for Information Science, 38*, 6, 1987, pp. 389-404.

[23]   J.F. Cove and B.C. Walsh, 'Online Text Retrieval via Browsing', *Information Processing & Management, 24*, 1, 1988, pp. 31-37.

[24]   Peter L. Noerr and Kathleen T. Bivins Noerr. 'Browse and Navigate: An Advance in Database Access Methods', *Information Processing & Management, 21*, 3, 1985, pp. 205-213.

[25]   Charles R. Hildreth, *Intelligent Interfaces and Retrieval Methods for Subject Searching in Bibliographic Retrieval Systems,* Washington, DC: Library of Congress Cataloging Distribution Service, 1989. Advances in Library Information Technology #2.

[26]   Stephen J. Wade and Peter Willett, 'INSTRUCT: A Teaching Package for Experimental Methods in Information Retrieval. Part III. Browsing, Clustering and Query Expansion', *Program, 22*, 1, 1988, pp. 44-61.

[27]   Charles R. Hildreth, 'Online Browsing Support Capabilities', *Proceedings of the ASIS Annual Meeting, 19*, 1982, pp. 127-132.

[28]   David Bawden, 'Information Systems and the Stimulation of Creativity', *Journal of Information Science,* **12**, 5, 1986, pp. 203-216.

[29]   Peter Ingwersen and Irene Wormell, 'Improved Subject Access, Browsing and Scanning Mechanisms in Modern Online IR', *Proceedings of the 9th Annual International Conference on Research and Development in Information Retrieval; Pisa, Italy, September 8-10, 1986,* New York: Association for Computing Machinery, 1987, pp. 68-75.

[30]   Marcia J. Bates, 'The Fallacy of the Perfect 30-Item Online Search', *RQ,* **24**, 1, 1984, pp. 43-50.

[31]   Marcia J. Bates, 'Rigorous Systematic Bibliography', *RQ,* **16**, 1, 1976, pp. 7-26.

[32]   Donald A. Norman, *The Psychology of Everyday Things,* New York: Basic Books, 1988.

[33]   John M. Carroll and John C. Thomas, 'Metaphor and the Cognitive Representation of Computer Systems', *IEEE Transactions on Systems, Man, and Cybernetics,* **SMC-12**, 2, 1982, pp. 107-116.

[34]   Jay Elkerton and Robert C. Williges, 'Information Retrieval Strategies in a File-Search Environment', *Human Factors,* **26**, 2, 1984, pp. 171-184.

[35]   S.A. Weyer, 'The Design of a Dynamic Book for Information Search', *International Journal of Man-Machine Studies,* **17**, 1,1982, pp. 87-107.

[36]   Christine L. Borgman, 'The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype Online Catalog', *International Journal of Man-Machine Studies,* **24**, 1, 1986, pp. 47-64.

[37]   Mary M. Huston, 'Search Theory and Instruction for End Users of Online Bibliographic Information Retrieval Systems: A Literature Review', *Research Strategies,* **7**, 1, 1989, pp. 14-32.

[38]   Stuart Hannabuss, 'Dialogue and the Search for Information', *ASLIB Proceedings,* **41**, 3, 1989, pp. 85-98.

[39]   George A. Miller, 'Psychology and Information', *American Documentation,* **19**, 3, 1968, pp. 286-289.

[40]   William P. Jones and Susan T. Dumais, 'The Spatial Metaphor for User Interfaces:

Experimental Tests of Reference by Location versus Name', *ACM Transactions on Office Information Systems,* **4**, 1, 1986, pp. 42-63.

[41]   John L. Bennett, 'Spatial Concepts as an Organizing Principle for Interactive Bibligraphic Search', IN Donald E. Walker, ed., *Interactive Bibliographic Search: The User/Computer Interface,* Montvale, NJ: AFIPS Press, 1971, pp. 67-82.

[42]   Richard A. Bolt, *Spatial Data Management System,* Cambridge, MA: Massachusetts Institute of Technology Machine Architecture Group, 31 December 1978. NTIS #AD-777 878/0GA.

[43]   D.D. Woods, 'Visual Momentum: A Concept to Improve the Cognitive Coupling of Person and Computer', *International Journal of Man-Machine Studies,* **21**, 3, 1984, pp. 229-244.

[44]   Dee Michel, 'When Does It Make Sense to Use Graphic Representations in Interactive Bibliographic Retrieval Systems?', manuscript, Los Angeles: University of California Graduate School of Library and Information Science, 1986.

[45]   D.J. Foskett, *Pathways for Communication,* London: Bingley, 1983.

[46]   P. Atherton, *Books are for Use. Final Report of the Subject Access Project.* Syracuse, NY: Syracuse University School of Information Studies, 1978.

[47]   Marcia J. Bates, 'What is a Reference Book? A Theoretical and Empirical Analysis', *RQ,* **26**, 1, 1986, pp. 37-57.

[48]   Jeff Conklin, *A Survey of Hypertext,* Austin, TX: Microelectronics and Computer Technology Corporation, 3 December 1987. MCC Technical Report #STP-356-86, Rev. 2.

[49]   William B. Rouse and Sandra H. Rouse, 'Human Information Seeking and Design of Information Systems', *Information Processing & Management,* **20**, 1-2, 1984, pp. 129-138.