Properties-Based Retrieval and User Decision States: User Control and Behavior Modeling

Gerald Benoit
Graduate School of Library & Information Science, Simmons College, 300 The Fenway, Boston, MA 02115-5898. E-mail: benoit@simmons.edu

As retrieval set size in information retrieval (IR) becomes larger, users may need greater interactive opportunities to determine for themselves potential relevance of the resources offered by a given collection. A parts-of-document approach, coupled with an interactive graphic interface and control panel, permits end users to tailor the information seeking (IS) session. Applying the model described by the author in a previous paper in this journal, this paper explores two issues: whether a group of information seekers in the same research domain will want to use this type of IR interaction, and whether such interaction is more successful than relevancy ranked lists, based on the general vector model. In addition, the paper proposes the use of gradient space as a means of capturing end users’ cognitive states—decision-making points—during a parts-of-document-based IR session. It concludes that, for a group of biomedical researchers, a parts-of-document approach is preferred for certain IR situations and that gradient space provides designers of systems with empirical evidence suited for systems analysis.

Introduction

Information seekers may be best served when information retrieval (IR) systems are comprehensible to them and end user’s online relevance judgments are comparable to those that would have been made if the seeker were evaluating familiar paper documents. Designers of IR applications however can build systems only to the limits of what is available to them. For example, they may be restricted by the full-text document collection, parsing algorithms, and models of interactive opportunities. This paper describes a model of information storage and retrieval intended to support tailoring of the information system’s behavior and support end user heuristic-based relevancy, or decision-making techniques, which also reveals to the designer more about user cognition during information seeking (IS). It proposes a “parts-of-document,” or finer-grained parsing, approach and a graphic user interface (GUI) that responds to the user’s input as cause for updating retrieval set membership and screen display. The results are expressed as behavior constructs in a gradient space, suggesting how individuals differ when given the same input opportunities.

Background

This project is motivated by the tendency to view human information seeking behavior in black-box terms, using techniques developed in control engineering to investigate behavioral responses to external stimuli. Naturally there are technical limits to be considered: the document collection size, storage space, retrieval algorithms, response time, etc. And because of this, IR design for large production systems may focus on limiting user input to terms, small sets of operators, linguistic manipulation (e.g., truncation, soundex, and term mapping), and refining a search to a subset of the initial retrieval set. This has the effect of promoting the designer’s model to the steering authority.

Tailoring Interaction for Individuals within a Class of Users

User studies that build solely upon this perhaps see a limited view of user interaction and cognition, stimulation being restricted to reading what surrogate information is provided, e.g., the URL, first line of the document, or parts of the bibliographic record. While useful, such studies may unwittingly accentuate systems that are suited to users with a “low need for cognition” (Cacioppo, Petty, Feinstein, & Jarvis, 1996) and support the “scientist-knows best” position (cf. Shneiderman & Maes, 1997). A model that permits an expanded form of interactivity, which gives the user greater control over the document collection’s retrieval behavior, might support better end user cognition and decision making. Such a model requires greater opportunities for the users to “steer” the IS session. For the designer heuristics...
enhance the explanatory power of user behavior (Nielsen, 1994a, 1994b).

**Heuristics**

As an alternative to the algorithmic approach to designing system behavior and to studying users, particularly individual decisions, some computer science and information science researchers turn to end-user heuristics—the following of general recommendations or personal decision making rules, akin to playing a chess game (Tversky & Kahneman, 1982; StatSoft, 2002). It is believed that a heuristic approach in building systems aids the analysis of “the relevance of the task from the user’s point of view and is important because it helps to identify the knowledge he needs” (Kovacevic, Diligenti, Gori, & Milutinovic 2002, p. 250). With this in mind, designers require a shift in understanding IS behavior and a different data model. Incorporating interactive opportunities along with a finer-grained structure of the documents within the IR system may facilitate the user’s success by encouraging trial-and-error experience without seeking to validate any particular hypothesis of the designer.

Such finer-grained parsing of document structure can be conceived of as identifying particular “properties” of the collection, beyond the usual bibliographic access points, which become the cognitive objects used by seekers. Broadly stated, by permitting the user to experiment with these objects the information seeking sessions emphasize the individual’s decomposition of the larger information seeking task into subtasks (Pohl, Kobsa, & Kutter, 2002) and perhaps the idea of IS design moves closer to task-based modeling (Bomosdor & Szwillus, 1999). As a result, it is possible to capture individual user decisions (subtasks) during information seeking based on what the IR system presents as well as their supposed “deviation” from the class of user behaviors on which the design is based (Priebeau & Vanderdonckt, 2002). As a consequence, IS focuses on the continuity and connectedness of sequences of events that take place in situations bounded by what the document collection offers and user decisions, and whose outcomes are thereby mutually influenced by their common localization. In this way, it is reminiscent of social networks, and can be mapped physically to represent spatial diffusion, such as Lexis-Becker demographic diagrams and “mean information field” (Hägerstrand, 1967).

**IR as a Bounded Space**

The heuristic elements contextualize the IS for the individual and form a bounded space, founded upon the usual quadruple of an IR session (logical views for the documents in the collection; a set of logical views for the user query; a framework for modeling document representations, queries, and their relationships; and a ranking function) (Baeza-Yates & Ribeiro-Neto, 1999, p. 23). The document collection contributes terms or other representations of concepts, along with “sub concepts” and their interrelations. The mutability of these relationships within the contained space is affected by the level of representational granularity of the document’s terms (e.g., the discretization (Benoit, 2002a)), sequence of previous events, and human perception at a given point in time.

We can abstract the repertoire of inputs and individual responses: (1) terms and interactivity that users draw on in order to realize their IR sessions and (2) constraints of (a) capability (limits of physical interactive opportunities), (b) coupling (defining where, when, and how long individuals can interact), and (c) authority or “steering.” This suggests that the system can distinguish between a repertoire of possible terms-decision paths and a concrete configuration of trajectories realized within these structural templates. In other words, information seeking sessions consist of a sequence of inputs and outputs, with multiple outcomes in a bounded space of concept representations, interactive opportunities, and a governing “authority” of the entire IR event based on the designer’s system model.

IR could be expressed, in its simplest form, as an input \((X_i)\), an output at a later time \((Y_{n_+})\) and a transfer function \(S\) that links the two, such that \(Y_{n_+} = S(X_i)\). However, one might expand this model to refer to continuous time with input/outputs occurring (or observed to occur) at discrete time scales, providing evidence of a user decision state change. Such models are used to examine the likely behavior of systems and their effects on an IR environment; otherwise the transfer function \(S\) is derived inductively without any knowledge of the causal mechanisms involved.

Capturing evidence in the form of individual user decision sequences would represent terms and interactivity opportunities as input and user decision-making during the IR session as output, the actual realized path of decisions as the transfer function. \(S\) would instead represent individual choices about terms, properties and their relationships over time, \(Y_{n_+} = S(X_i \in \{X_1, X_2, X_3, \ldots, X_n\})\). Adapting Churchland and Sejnowski (1989) and Hägerstrand (1984), this can be described as a gradient descent bounded by term/property space (Fig. 1).

**Users and Retrieval Sets**

As an example in an IR session, a user enters a query term, and, depending on the particular retrieval algorithm, the result set is populated with documents of likely utility or relevance to the user. The interface controls further user investigation, at times offering interaction, such as “more like these” buttons or an offer to display the full record with controlled vocabulary identifiers. Alternatively, IR could explicate result sets consisting of documents based on a finer level of granularity that might be useful to users and comprehensible in light of their own needs, knowledge, etc. For instance, a researcher may want a document because a particularly narrowly defined concept exists in the “methods” section, and a broader related concept is found in the “conclusions” section. A user with such an information need
FIG. 1. Gradient descent as a term/property space.

would not be served by most online systems because input opportunities are not supported and because the ranking algorithms [e.g., just the vector space model or tf-idf (Baeza-Yates & Ribiero Neto, 1999; Ponte & Croft, 1998)], create retrieval sets based on a rationale outside the user’s heuristic. Retrieval sets based on concepts and document properties that match the user’s heuristics more closely tailor the IR session to the individual’s cognition.

But to support the greater interactive opportunities, the interface requires more user controls and a means for displaying the retrieval set with greater dimensionality. Visualization techniques are popular response for collapsing large volumes of data and in suggesting relationships among the data without specific user input. It has been argued elsewhere (Barry, 1994; Benoît, 2002b) that providing end users with a graphic display of the retrieval set, increased control over the level of linguistic granularity, and interactive opportunities is a useful approach to IR design.

It stands to reason, too, that as the information seeker learns more about a topic or the potential resources in a collection—in other words there are reused inputs in IS session—then he or she may not make similar decisions compared to others in a group. This evinces the user decision-making (cognition) at different stages, based on the inputs.

It is theoretically fruitful to investigate this model by establishing the acceptability of interactions that are more cognitively engaging to examine how data generated by this model provide evidence of individual cognition during IR. The next section of this paper discusses the above themes from related work. Subsequently the ideas are tested with a group of information seekers interested in increased interactive opportunities, who need varying levels of granularity in the collection representation to match better their understandings of their information needs.

Related Work

There is a growing body of research that recognizes the impact that the system designers’ worldview has on IR system behavior and the display of data. These preferences promote the interests of the designers over an interest in providing greater end user controls (Benoît, 1998). There is corollary research in how designer perspectives are exerted over the user, or how intentional or accidental users are steered away from useful data, through the interface (e.g., Shneiderman & Maes, 1997). There is also investigation into the influence of the data themselves, as a time-dependent function, on the seeker’s cognitive states. Veerasamy and Belkin (1996, p. 85) noted that more is needed for the user than just a ranked list: “More explanation may be necessary in situations where the top retrieved documents are all clearly relevant but when the user needs to modify the query in order to get better results, understanding the causal relationship between query and document ranking could be useful in effective query reformulation.” Squires (2001, §3.1) likewise notes the influences on information seeking behavior and end user interpretation: “Clearly, many factors can influence the way users look for and use resources. Convenience, influence of outside or societal factors . . . will affect behavior and decision-making, and also influence user perception about the quality and usefulness of resources.” Because of this, Squires suggests guarding users from an instinctive personal favorite resource or behavior by permitting them to supply the IR system with their values, or “to enable the users to specify the importance of sources according to a variety of criteria” (§4). Other work (e.g., Chen, 2000) suggests that instead of building systems for a class of users, instead build to support individual decision-making. Naturally, the first step in building an IR system entails a needs analysis that includes a set of users or tasks, as no IR system is built to an individual’s specific interests.

Some research (e.g., Shaw, 1995; De Bliek, 1994) suggests that information seekers use also their own background knowledge of the concepts and literature of a field to propose utility for retrieved documents. Shaw writes “Language, source of publication, author, and length of work with longer materials preferred, as well as previous exposure to a citation all influenced relevance assessments” (1995, p. 327). Schamber (1994) also emphasizes the need to build on the users’ criteria in relevance judgments. Others (Cool, Belkin, & Kantor, 1993; Su, 1994) actually number specific behaviors that the researchers believe influence decision-making.

In addition to providing a different view of user behavior, system designers turn to visualization as a response to the issues of retrieval set size and interactive causality [see Chen (1999) and Keim (1999) for exhaustive reviews of visualization; Pearl (2002) on causality]). Typically, an interactive visualization system displays icons on a screen, relying on proximity to suggest degrees of similarity or lines drawn on the screen to specify relationship among icons, based on an underlying lexical attraction model of language (Yuret, 1999).

Some visualization systems support displaying additional information about the document (such as tool-tips
showing traditional bibliographic access points, e.g., author, title, and keywords). However, these approaches stop short of supporting the user’s ability to investigate the documents’ contents, examine their relationships, and answer contradictions the user sees in the retrieval set. Moreover, these visualizations do not help the user when facing a large number of retrieval items or linguistic concerns, such as polysemy or statistical relationships (Berger & Lafferty 1999; Baeza-Yates & Ribiero-Neto 1999). As Veerasamy and Belkin (1996, pp. 85–86) observe:

A possible means for addressing problems . . . is to display to the user something about the document which relates [the documents] directly to characteristics of the query, and which relates them to one-another. . . . Graphical display of the characteristics of retrieved documents (visualizations) which are relevant to their retrieval and ranking is one obvious approach to this problem. A further problem in IR systems in general has to do with the multi-stage nature of presentation of results. The initially-presented surrogates are meant to provide a concise picture of what a document is about. Based on these surrogates, the users may request more detailed information about particular documents which look promising (or for which the surrogate information is equivocal). In some cases, this might be the ‘full’ bibliographic information about the item, in others an abstract, an in many systems now, the full text of the item . . .

Similarly, this relates the user to the documents as an influence upon relevance. Froehlich (1994, p. 128) states succinctly that: “‘Relevance’ as used in relevance judgments by end users cannot and never will be defined in a clear and precise sense.” He concludes that a finite variety of criteria would be useful and that probable utility of resources is likely “based on classes of users and contexts of information seeking” (p. 329).

The dilemma is that a purely class-based user model employed by many designers may erase the significant differences between an individual seeker’s needs and the influence that different levels of domain knowledge may play (Lif & Sandblad 1996). Veerasamy & Heikes (1997, p. 236) write:

The overall concern of all components of an IR system is to present the user with as much relevant information as possible. While there has been a lot of work on effective algorithms for retrieving and ranking relevant documents, not much attention has been paid to the study of the effectiveness of user interface components of IR systems. Apart from retrieval mechanisms, interactive IR systems must also be concerned with the design of appropriate display mechanisms that present the retrieved information in the “best possible manner.”

Clearly designing and evaluating interfaces that support end user heuristics is difficult and no universal approach has emerged. The question of “best possible manner” remains open.

Properties-Based Retrieval and Search Controls Interface for Property and Term Weighting

In Benoit (2002b) an interactive retrieval system was described that broke down the document collection into identifiable sections, using xml tags, termed document “properties.” Every document has a form of properties structuring it. Most documents, for example, have identifiable sections, indicated by phrases, such as “Introduction,” “Results,” or “Conclusions.” Structured documents, such as a database table or xml-tagged document, also provide identifiable regions of a document that can be associated with terms extracted from that part of the document.

The properties-based application can individualize the information-seeking context including the link between the user’s knowledge and information resources. For instance, a biologist will understand better the nature of biological literature, such as where the most “useful” part of research results appear in a document or the preferred phrases of the field. Providing controls to the end user to identify the parts of the document and the type of terms that are useful allows the information seeker to control the behavior of the search engine in ways that are significant to him or her. The information retrieval system’s index must parse and maintain both term frequency data and part-of-document property information.

Besides identifying the parts of the document collection to search, the IR system must permit input from the user to identify a term and the property associated with it and indicate to the system the degree of importance of those term/property pairs. The example biologist might want to find the term “Saccharomyces cerevisiae” (Brewer’s yeast) within the abstract, the gene name “orf7” in the conclusions, and “regression” in the methods section. The IR system will return and rank resources based on these term frequencies. However, the ranking algorithm actually may obscure items of interest to the individual researcher’s needs. For instance, “S. cerevisiae” and “regression” would be very common terms so the user can reduce the influence of these terms in the search. Likewise the user knows orf7 is not likely to appear and so can increase the weight of this term to force any document(s) with the term to rise in the relevance ranking.

This “parts-of-document” orientation is reminiscent of Burkowski’s nonoverlapping list approach (1992a, 1992b), Navarro and Baeza-Yates’ work on proximal nodes (1995, 1997), and Baeza-Yates and Gomnet’s (1996) searching for regular expression on a trie. Documents so parsed provide identifiable regions that can be associated with terms extracted from that part of the document, essentially associating a “concept region” of the document with specific concepts expressed as terms within this region. Compare how people typically propose utility of a paper document based on some concept or fact within an identifiable part of the document, e.g., the “discussion” section may have good quotes, another document’s “methodology” section may look promising (Dillon, 1991; Dillon & Schaap, 1996), so a
computer application can be designed likewise to individualize the information seeking context, which is vastly more flexible than bibliographic-based keywords. As an alternative to detailing specific behaviors of classes of users and evaluating the interface to those specifications, some research acknowledges that individual users bring to information seeking interactions considerable, if passive, knowledge of the subject area. They also favor a personal searching style, and their own rationale, or heuristics, for proposing relevance of documents so that, even as members of an identifiable class of users, individual cognition and needs can be identified. An IR system, as Chen (2000, p. 529) notes, "should be adaptable to suit individuals and tasks at various levels."

To engage in a sustained, computer-mediated form of argumentation, the interface should situate the user with a familiar display, such as the common hierarchical list. The IR system should also support a novel, interactive interface, such as a graphic visualization of the retrieval set, suggestive of Kohonen et al. feature maps (2000). Lee, Park, and Choi (2001), Stapley and Benoit (2000), Lin (1992, 1997; Lin, Hassel, Song, & Doszkocs, 1995), and others suggest that graphic nodes demonstrating similarity between query and documents, as well as interdocument similarity, improve end user satisfaction with the search process and yield more useful retrieval results. As the user explores the set and makes tentative decisions, the interface can be updated to show the user the consequences of the emerging line of inquiry.

Finally, user inputs via a control panel can be stored in a transaction log and the user selected terms within the parts of the document. It is then possible to create a gradient space for each IR session and see how a group of information seekers will make different decisions given the same inputs, leaving a trajectory of individual cognitive behavior to create an image of group and individual information seeking behavior.

Research

A subject population was identified that was theoretically more likely to maintain a sustained interaction with a novel IR system. Twenty graduate student researchers in biological sciences were recruited to interact with the retrieval system. Scientific researchers are likely to want greater cognitive engagement (Cacioppo et al., 1996), will formulate queries with complex attributions (Petty & Jarvis, 1996), and devote attention to an ongoing cognitive task (Osberg, 1987). As information seekers, they are more likely to be curious about and be engaged with the system (Olson, Camp, & Fuller 1984), and they have a desire to maximize the information gain from the interaction (Sorrentino, Bobocel, Gitta, Olson, & Hewitt, 1988). Finally, as a group, they are likely not to be anxious when dealing with unknowns.

The programming mechanics of the system being used is detailed in Benoit (2002a). The underlying retrieval engine used generates a term/property matrix, calculating term weights both within-collection and by-property. Two sets of source documents were parsed—300 documents from Dialog and an xml-tagged journal Priorities for Health. The documents are in English, French, German, and Dutch, preserving the "parts-of-document" (or "properties") information for each term. The Dialog files' properties are file name, title, author, abstract, method, research, conclusion, discussion, future research, journal name, volume, issue, pagination, publication date, language, super taxa, diseases, organisms, biosystematic, major concept, chemicals, methods, miscellaneous, and alternative index. The Priorities' tags are title, author, editor, reference, abstract, science, institution, background, location, text, conclusions, disease, and [personal] name.

Users were given a 20-minute training session, explaining the concepts of queries, retrieval sets, relevancy ranking, and how full-text retrieval algorithms operate. Two types of queries were demonstrated. First a query term alone, without specific property, and then a combination query was shown, consisting of a query term + property. The resulting hierarchical lists and graphic interface were demonstrated, as were the "control panel" and how moving controls affected the retrieval set membership. For instance, users may seek a query in "all" properties, and secondary queries associated with specific parts of the document. The initial retrieval set is ranked according to tf * idf and displayed in a hierarchical list with the usual bibliographic elements, presented in an interactive graphic display. Users then employ slide bars to adjust the weight that the term/property combination exert on the retrieval set membership and ranking. As a result of the user's decision-making, the display is updated, demonstrating the cause-and-effect of user feedback on the set. Following Shaw (1995, pp. 331ff), these bioscience users were asked to perform their own searches (e.g., "is marijuana use physically addictive?") in the scientific document collection until they achieved what they felt was a successful outcome.

Measuring the effectiveness of retrieval is difficult (Tague-Sutcliffe, 1996; Ellis, 1996) and the small sample size rules out otherwise appropriate confirmatory factor analysis or structural equation modeling. Therefore, following Shaw (1995) and Treu (1994, p. 133) a combination of think-aloud and a 25-question survey was employed to determine the individual user's satisfaction measure was determined by four constructs: facultative correspondence (FC), depth penetration (DP), supportive context (SC), and system adaptiveness (SA). Quoting Treu (1994, p. 133):

- Facultative correspondence (FC): the mapping of task representation and actions, as implemented behind the interface, into user-visible interface and user-expressible language constructs in a many conductive (or close) to the expectations of the user's mental model and to what the user wants to accomplish
- Depth penetration or distributedness problem (DP): orderly transitioning between or among levels of detail, within an interaction state, from one such state to another,
and through the surrounding (distributed) environmental regions
- Supportive context (SC): providing specific forms or layers of help to the user; for both task-specific and task-peripheral purposes
- System adaptiveness (SA): compensating for the user’s current state of learning, knowledge, interest, experience, and capabilities.

A function based on these constructs generates a value for the individual or group, \( RC = f(FC, DP, SC, SA) \), suitable for generating views of group IS behavior and revealing which constructs and combinations of constructs most affect an individual’s experience.

Two hypotheses are put forward: whether properties-based information seeking with interactive controls is “more helpful” than full-text retrieval with ranked results using the general vector space model, and whether the user feels more cognitively engaged with the properties-based, graphic user interface. In addition, consideration is given to gradient space as a tool for researchers.

Results

The results of the user interaction are based on qualitative and quantitative analysis.

Quantitative: The small sample size rules out otherwise appropriate confirmatory factor analysis or structural equation modeling, but descriptive statistics and reliability analysis are useful nevertheless. An independent samples t-test was performed on both hypotheses. In the first case, the \( t_{19} \) is statistically significant so that users’ judgment that properties-based retrieval is, to this set of subjects, more useful than a relevancy ranked list based on the general vector model. The user group believes also that their cognitive engagement was increased when using the properties-based approach: their ability to help themselves determine how to manipulate the retrieval set, overall sense of satisfaction, confidence in the IR system itself, and confidence in their own relevancy judgments were all improved with this system.

A measure of overall satisfaction (“RC,” sense of cognitive engagement, overall system utility, and overall satisfaction with the experience [Treu, 1994]) was determined. The constructs (FC, DP, SC, SA) all correlate with each other at the .05 level, with the greatest associations in descending order between SC and SA, FC and DP, and FC and SC. Each correlated to each of the outcome variables (“sense of cognitive engagement,” “system utility,” and “overall satisfaction”). There was a positive correlation between SA and FC, negative between FC and SC, and FC and DP. This is not unexpected: Users have a greater sense of the system adapting to them if the correspondence between their thoughts and expectations of interactivity are close; and the closer the controls and behavior of the system are to the user’s mental model, the less cognitive depth penetration.

A reliability analysis was performed to test the survey itself using Cronbach’s alpha (\( \alpha = .9825 \)). This is a very high value, which, in this case, is not unanticipated since the average interitem covariance of the factors was so consistent. Cronbach’s \( \alpha \) for each (FC = .9401; DP = .8751; SC = .8837; SA = .9554) indicated, however, that questions concerning hierarchical lists, the utility of seeing the entire source document, and status box data (which were entirely correlated \( r^2 = 1 \)) were not good indicators.

Qualitative: Collapsing user comments to the most frequent observations, users on the whole enjoyed the IR experience. Users mentioned that they did not find it easy to intuit the rules guiding the hierarchical list’s behavior, primarily because they could not see immediately the relationship between the hierarchy and graphic display. Users also noted that their relevancy decisions were based on evidence in the “status display” (the area that describes the search terms, system ranking, etc.) but they used color and movement in the graphic display as the foundation for relevancy. This observation maps to the statistical data: the greatest variance was between the survey questions that addressed this topic.

Users also consistently assumed that the collection was domain-specific: that is, that the documents represented a research-oriented collection of scientific literature suited to their interests. When searching with a specific task, most users initiated the search with a broad search term and a document-structure-oriented property (e.g., traditional library access points and identifiable regions of “introduction,” “conclusion,” and “methods”) because, as one subject said “that’s what I’m accustomed to doing.” When reminded about the other available properties, most users still formed their original query with a broad search term and property (e.g., “addictive + all”) and then specifically chose a concept oriented property (e.g., “… looking for terms ‘addiction’ or ‘marijuana’ in the ‘science’ property”).

The subjects felt that they understood the precision of terms and could project context for isolated terms because of the linguistic specificity of the domain. Some mentioned that they felt comfortable with term/property pairing for searching, but appreciated the greater scope of properties in the demonstration application. It is interesting to observe, too, that users who felt that the query term was absolutely unambiguous or represented a specific scientific concept refined the search by entering the same term but within a specific part-of-document and manipulation of the weight bar to see the resulting update in relevancy ranking display on the graphic panel. Searchers who were in a well-defined stage (especially the literature review stage) would search for terms within “title” and “conclusion” property pairs, or “abstract” and “conclusion” pairs.

Although the document collection used was small, the specificity of the collection meant a large number of items were retrieved in the initial search. Because of the size of the set, subjects were glad to be able to refine the set. Verbatim comments below are shown in italics.
I like not having to work with a lot of hits in a list, and the ability to refine the set is great. I was able to isolate particular documents ... the change of colors on screen and the change of position was great ...

However, domain-specificity seems to be an important assumption of the user:

... but I wouldn't like having to pull out all the biomedical records from a bigger, all-purpose collection. The nodes are useful, but it is still hard to remember what documents were where.

Other searchers said they used very specific terms and properties depending on the stage they were in their research (literature review, project design, analysis, discussion, conclusions). Some subjects said they liked the ability of searching parts-of-document that matched their sense of stage:

When I'm writing a paper the research method is pretty much determined by the project. But I want to know who else is using a particular method both to see what journals are publishing that type of method and to see how other researchers express their use of the method.

All users consistently mentioned the use of the graphic display (a) to get a sense of the entire collection, (b) to confirm relevancy decisions based on the hierarchical list, and (c) that they prefer the hierarchical list for small retrieval sets, but that as they became more familiar with this type of searching would likely switch to the graphic.

As a document gradient space: The data generated by the system's capture of user inputs (terms and property pairs) can be mapped in the gradient space to visualize quickly the divergence of individual decision making. Figure 2 demonstrates.

For each user of the system, the output $Y_{t+n}$ of the interaction is the result of a set of $X$ inputs over time $t$. This is similar to relevance feedback systems; permitting the user to update the query or refine the retrieval set by additional terms. Unlike that approach, the term/property pair creates a conditional probability of each user's individual choices, given the document property ($dp$). In other words, each decision state for the user is $Y_{t} = (X_{t}) \mid dp$. The sequence of individual user inputs, then, is $Y_{t+n} = S(X_{t} \in \{x_{1}, x_{2}, x_{3}, \ldots, x_{n}\}) \mid dp$. For each user's output sequence $U_{y}$, the difference between $U_{y_{1}}, \ldots, U_{y_{n}}$ offers a view into individual decision-making differences (cognitive states), given the same input opportunities.

Conclusions

The results suggest that if the vector model's ranking were sufficient for user needs, there would be little difference between the system's relevance rankings and the user's final decision state, and between users. Of the sample, every participant manipulated the set to alter retrieval set membership before he/she was satisfied. In addition, since users individualize information seeking, there should be differences in the Euclidean distances and set membership. This may mean that a general vector model ranked list is not sufficient, at least not for domain-specific searching by knowledgeable users.
As the qualitative and quantitative data imply, there are no dominant trends of searching in this small sample. Such trends, if confirmed by a larger random sample, would demonstrate that when given the chance, knowledgeable domain-specific users would tailor their information seeking behavior to match better their knowledge level, phase of information need, and understanding of the concepts and documents within a given online IR resource. This supports task-based modeling over user-group modeling in system design. Interfaces, then, could be integrated into existing IR systems and well adapted to XML or semantically-tagged document retrieval. Greater flexibility in document parsing may better serve the heuristics of end users.

The quantitative data suggest that users' own sense of cognitive engagement is increased by the use of the GUI controls as a means to perform their thoughts (e.g., "I want more of this, . ..") and the novelty of seeing the IR system's ranking data and dynamic cause-and-effect of the graphic display. They are better able to answer their own questions by manipulation of the controls. As the data imply, however, the additional features of the interface (status panel, dynamic nodes, and color changes) did not entirely match the user's mental model of the system, which is not unreasonable considering they have never used this type of interface before. There was no statistically significant covariance among the constructs for "system useful" and "overall satisfaction" measures, although FC and DP remain strongly correlated, SA and SC negatively so. This suggests that users' mental models are challenged and engaged more by attempting to analyze the workings of the system itself than the actual use of the controls. Some variance may be accounted for the novelty of the interface which no subject had previously used (the Macintosh's "Aqua" GUI) and users' admitted eagerness to try something new.

Turning to individual participants, that each user chose to emphasize different properties suggests that a single presentation of retrieval results is not sufficient and that individual interactive styles need to be addressed in information retrieval design. Discussion with the participants indicated that the users were at different stages in their research and that, as they grow in their knowledge of their work (cf. Kuhntau, 1991), they might prefer different properties during retrieval. For example, one participant said that, during the literature review stage of a project, she prefers to search by author and title. When considering which research methodology would be appropriate, she would like searching only the "research property" to find statistical methods.

**Gradient Space**

The use of a gradient space to represent graphically the user's decision-making has several implications. By casting information seeking as a distance-decay pattern on a spatial grid, designers of various types of systems (e.g., guided information retrieval and educational software) see the diffusion of user information seeking behavior within a known set of inputs. What Hägerstrand (1967) labels a "mean information field" can be shaped on the known properties of the document collection: properties on one axis, major shared concept representations on another, and user decision-state changes on the third. Aware of this, designers may opt for more flexible interactive opportunities in the system—giving freer range to user cognition during IS, or when appropriate, they can maintain the "steering" component mentioned above—to keep the individual user's behaviors within a preferred set of cognitive responses. Similarly, it can be used to help designers identify when there is potential for a break in the system model of user behavior. Equally, the model can expose unanticipated phenomena outside the original design.

It is interesting, too, to propose the application of a gradient space model for evaluation of system effectiveness. Recalling Shannon and Weaver's (1949) original formula, one can construct a statistic \( H \) that represents the maximum information gain for the individual seeker and a group of seekers:

\[
H = \sum_{x} \log \left( \frac{1}{p} \right) = \sum - x_{i} \log x_{i} \tag{cf. Marchand, 1972; Thomas, 1981). \]

The individual \( x_{i} \) are the probabilities of \( N \) possible outcomes of the stochastic IR. As \( H \) approaches 0, one of the \( x_{i} \) approaches a near certainty (unity). If all the \( x_{i} \) are approximately equal at 1/\( N \), then \( H \) approaches a maximum given by \( \log H \). This can be used to define an index \( R \) as a measure of redundancy. Stated differently, having a known number of properties, with a known set of input terms, it is possible to predict the maximum information gain for a group using a given collection, or an individual's maximum information gain from the collection. At one extreme of a continuity of user behavior, this suggests that individual decision-making indicates ambiguity because a highly concentrated spatial pattern (\( H = 0, R = 1 \)) evinces one kind of order in the user behavior, perhaps comparable to data generated by latent semantic analysis. At the other extreme a nonuniform, scattered spatial pattern (\( H = H_{\text{max}}, R = 0 \)) suggests highly individualized behavior and maximum information gain.

**Future Research**

Data generated by a small sample are not a sufficient basis for adopting these suggestions. Given that the project intends to study the individual's heuristics during IS, the next phase of this project is to study the disaggregate choices in a larger domain-specific environment. Furthermore, it would be instructive to study, for each subject, the sequence of choices over time, particularly the causal relationship of concept- and document-structure-oriented properties affect end user decision-making. Users stated that the novelty of the interface held their interest, but prolonging the information seeking session beyond the novelty is a challenge. Nielsen (2002) recommends some of the design considerations already incorporated in the test IR system, such as user control and freedom, aesthetic and minimalist design, and flexibility of use. In addition, other suggestions, such as increased error recovery dialogs, different methods.
of matching the system’s and the user’s “real worlds” (through icons and familiar language), and system status visiblity, may engage users longer. The project will analyze the results of user choice sequences by a multinomial logistic regression controlling for time and subject. The dependent variable would be selected document properties and the generalized odds would be estimated. The purpose of the analysis would be to order document properties from the most preferred (most often used) to the least preferred and establish a basis for a more flexible data model for properties-based tailoring of information seeking contexts.

References


496 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY—April 2004


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.