THE CONCEPT OF INFORMATION SEEKING STRATEGIES AND ITS USE IN THE DESIGN OF INFORMATION RETRIEVAL SYSTEMS

N.J. Belkin & C. Cool

School of Communication, Information & Library Studies, Rutgers University
4 Huntington Street, New Brunswick, NJ 08903

Abstract. The support of effective interaction of the user with the other components of the system is a central problem for information retrieval. In this paper, we present a theory of such interactions taking place within a space of information seeking strategies, and discuss how such a concept might be used to design for effective interaction. In particular, we relate this view to ideas of case-based reasoning, and show how these two concepts can complement one another in the design of information retrieval systems.

1. Introduction

A central problem for information retrieval (IR) is designing for effective interaction. Although the traditional issues of representation and comparison techniques are still important, it is becoming increasingly clear that IR is an inherently interactive process, from a variety of points of view, and that supporting, and taking advantage of the interaction of the user with the other components of the IR system is crucial for effective IR system design (see, *e.g.* Belkin, Marchetti & Cool, 1993). Although there is a fairly long tradition of interest in this issue in IR (early concern with user interface design, e.g. Walker, 1971; the long-standing concern with, and research program in relevance feedback, *cf.* Salton and Buckley, 1990), only fairly recently has it been suggested that interaction is a central, and organizing process in IR (Oddy, 1977; Belkin, 1982; Belkin & Vickery, 1985; Croft & Thompson, 1987; Bates, 1989; Thiel, 1990; Ingwersen, 1992; are all examples of this point of view).

Some common features of this viewpoint are that: asking people to specify information 'needs' is in general unrealistic; people's conceptions of their information problems change through their interactions with the IR system; there are different kinds of information problems, for which different kinds of interactions might be appropriate; the rest of the system can better represent the person's information problem through interaction; and, entire information seeking episodes can consist of a variety of types of interaction.

A major problem with this interactive approach is being able to gain an understanding of the nature of the interaction itself, and then to move from such understanding to the specification of a system structure which supports and enhances it. This is the major theme of this paper.

A significant problem for case-based reasoning approaches to system design lies in knowing when a given stored case is relevant to a current situation. This problem can take several forms; most commonly, it is perceived as a problem of indexing the cases, and retrieving the relevant cases at appropriate times, for use in supporting the current situation.

The case-based approach has obvious intuitive relevance to the IR interaction problem, in particular because we can consider records of a person's interaction in an IR system to be a case of problem-solving activity. But how to move from this intuitive generalization, to a specific solution, is not terribly clear.

In this paper, we propose that the concept of a space of information seeking strategies, combined with a view of IR interaction as dialogue, can lead to possible solutions to both of the problems we have posed here: how to structure IR interaction effectively; and, how to use cases of IR interaction effectively.

2. Information Seeking Strategies

Elsewhere, (Belkin, Marchetti & Cool, 1993), we have proposed a model of information seeking behavior based upon the concept of a multidimensional space of information seeking strategies (ISSs). According to this model, the variety of behaviors people engage in while searching for information in some knowledge resource can be viewed as ISSs, and these ISSs can be construed as interactions between the user and the other components of the IR system. These strategies arise from characteristics of the person's *problematic situation* (Wersig, 1979); in particular, the user's state of knowledge and information seeking goals. For instance, a person may or may not be able to specify a desired item, or may or may not know

how to use an information resource effectively. A person may begin an information interaction with only a vague understanding of his or her information problem or need. Furthermore, a person's knowledge and goals may change over the course of a single information seeking episode. The specific values of the ISS taken at any one time are contingent upon these factors, related to the larger context of a person's information problem.

We suggest that any single information seeking interaction is a complex activity, which can be characterized according to its values on a relatively small set of factors, or dimensions. In our preliminary attempts to develop this model, we propose the following four dimensions of ISS, based upon our own observations and on the empirical findings of others, including Belkin, *et al.* (1990), Ellis (1989) and Hancock-Beaulieu (1990):

Method of Interaction
(scanning - searching)

Goal of Interaction
(learning - selecting)

Mode of Retrieval
(recognition - specification)

Resource Considered
(information - meta-information)

The first of these, method of interaction, can be understood in terms of the classic distinction between *searching* for a known item and looking around, or *scanning*, for something interesting among a collection of items. The goal of the interaction may be *learning* about some aspect of an item or resource, or *selecting* useful items for retrieval. Furthermore, looking for identified items can be characterized as retrieval by *specification*, while identifying relevant items through stimulated association can be characterized as retrieval by *recognition*. And interaction with *information* items themselves can be contrasted with interaction with *meta-information* resources that describe the structure and contents of information objects.

We suggest that any single ISS can be described according to its location along these four dimensions. In the simplest case, we can consider each of these dimensions as orthogonal, with dichotomous values, resulting in sixteen distinct ISSs. Figure 1 illustrates all of the possible ISSs that can be derived from our binary, four-dimensional model. At this point, it is unclear whether the dimensions suggested in our model can be treated as continuous rather than dichotomous variables, although it seems likely. If so, we have a multidimensional space of ISSs, in which certain regions might exemplify standard interaction strategies.

ISS	METHOD		GOAL		MODE		RESOURCE	
	Sc	S	L	S	R	S	I	M
1	x		x		x		x	
2	x		x		x			x
3	x		x			x	x	
4	x		x			x		x
5	x			x	x		x	
6	x			x	x			x
7	x			x		x	x	
8	x			x		x		x
9		x	x		x		x	
10		x	x		x			x
11		x	x			x	x	
12		x	x			x		x
13		x		x	x		x	
14		x		x	x			x
15		x		x		x	x	
16		x		x		x		x

Method: Sc = Scan; S = Search
Goal: L = Learn; S = Select
Mode: R = Recognize; S = Specify
Resource: I = Information; M = Meta-information

Figure 1. Information Seeking Strategies (after Belkin, Marchetti & Cool, 1993)

According to our matrix, ISS2 represents a situation in which a person needs to learn about characteristics of the knowledge resource before the information search can begin. This can also be understood as the ISS associated with an unformulated and unspecified information problem. A person in this situation might begin an information seeking episode by looking through a meta-information resource such as a classification scheme or thesaurus in order to learn about the organization of the knowledge resource.

In contrast, ISS15 is a prototypical example of a well understood information problem, in which the goal of the interaction is not to learn about the system, but to select items which can be specified by the user. After learning about the information system by consulting the classification scheme, the person represented above in ISS2 might choose some descriptors from that scheme as the specification of a topic of information items to be searched for. Similarly, ISS1 represents a person who scans the current periodical shelves in order to learn what journals exist on a particular topic of interest. ISS6 corresponds to a person scanning through a table of contents of a journal, with the

goal of selecting articles relevant to a particular topic.

We do not wish to make the claim that the dimensions suggested in our model are exhaustive, but, on the basis of our examples, they appear to be necessary if not sufficient, and represent at least a useful starting point for characterizing ISSs. The matrix can be used to illustrate several less usual ISSs. For example, while scanning is typically associated with retrieval by recognition, and searching is typically associated with retrieval by specification, the ISSs identified in our matrix demonstrate that these are not inevitable associations. Scanning occurs in conjunction with retrieval by specification in ISS4. A person may know precisely what she is looking for, but not where it is located. Searching occurs in conjunction with retrieval by recognition in the case of knowing what one wants to retrieve, but not being able to specify it (ISS13). The inability to specify identifying characteristics of the item means that retrieval by recognition is the only mode available for this user.

According to our conceptualization, information seeking behavior is characterized by movement from one strategy to another within the course of a single information seeking episode, as a person's problematic situation changes (*cf.* Bates, 1989). For example, a person may not be able to specify the title or author of a book she is looking for, but may remember its approximate shelf location. In order to recognize the item, this person might go to this location and scan the shelves (ISS5). While at the shelves, she may find some other book or item that seems relevant. This might prompt her to search for other similar items, but first she needs to know how to find these items. In order to learn how other items have been characterized in the information system, she may turn to a meta-information resource, such as a catalogue, to identify the subject terms under which the item has been indexed (ISS12). Having finished this step, this user would be able to continue the information seeking episode by specifying these terms in a search for other relevant items (ISS15).

Having ISSs described by, and located in, the kind of space we suggest gives us a means to describe movement from ISS to ISS, as well as to describe the individual ISSs, and potentially the means to understand such movement well enough to devise methods for supporting it in a principled fashion. From this point of view, we can consider ISSs as types of user interactions within the IR system, rather than as queries or demands put to that system.

3. The Concept of Prototypical Interactions

As well as characterizing ISSs according to the dimensions of the ISS space (effectively, locating them in that space), we also suggest that we can associate a prototypical interaction pattern with each region of the space, or each ISS. Such an interaction pattern, or script, would characterize the most usual, or most effective, or in some sense standard means by which the user and the system interact, in order to accomplish the ISS associated with that region of the ISS space. Such scripts, based, for instance, on, and abstracted from observations of people as they engage in information seeking, could be used as means for structuring human-computer interaction aimed at achieving the goal of that particular ISS. Of course, there could be more than one pattern associated with any one region, although the reverse is rather less likely. For simplicity's sake, in our examples, we consider the case of only one script per region.

In the previous section, we have seen examples of different kinds of ISSs, each associated with a general region of our example ISS space. These examples were static, in the sense that they described situations, rather than activities. Here, we give example scripts for human-computer interaction for some of these ISSs, in order to demonstrate the nature of such interaction patterns. These scripts are construed as patterns of moves in a two-party interaction; that is, a conversation. As such, these interactions are properly dialogues, and can be formally characterized as such (*cf.* Belkin, Cool, Stein & Thiel, 1993).

Any ISS interaction begins with a standard introductory section, in which the system informs the user of what classes of interactions it can support, the user chooses one, and the system informs the user of how that interaction will proceed. The various interaction choices (say, sixteen for the different ISSs in our simple model) will be described for the users according to the values of the dimensions, and especially according to the goals and knowledge that they assume of the user. The user having chosen, and learned about one such interaction support pattern, the specific interaction proper begins.

Consider ISS5, which is appropriate for situations in which the user wants to find some item that is known about (or known to exist), but which cannot be specified by searchable characteristics. However, the general 'location' of the item may be

known (by location, we mean some conceptual region, such as red items, or items on the third floor, or like some other items). The interaction commences with the system asking the user to identify the location of the item. The user is prompted to respond by suggesting the area that is known to be potentially relevant, which s/he does. The system then asks the user to indicate aspects of the item that would be useful for recognizing the right one (title, author, frontispiece, etc.), giving the user a choice of the characteristics available to the system. The user specifies those that would be useful, and the system displays the items in the specified location, focusing on the characteristics which the user has indicated as helpful for recognition. The user scans through these items and, in the simplest case, recognizes the desired item.

ISS12 is relevant to the situation in which a person would like to find items like a specified known item, but does not know how to characterize the desired items appropriately. This ISS is then in the region of searching, learning, specifying, meta-information After the user has chosen the appropriate interaction script, the system begins by asking the user to identify or specify the known item, which the user does. The system then asks the user what characteristics of this item she would like to learn about (perhaps as a choice from available ones for descriptive and searching purposes). The user chooses one or more such characteristics. The system finds the specified item, and displays the requested characteristic associated with that item (e.g. the terms used to index it). The user can then choose from the display as a basis for a subsequent search, or, ask the system to display things related to one or more characteristics selected from the display (e.g. browse the thesaural structure for a given term).

ISS15 responds to a quite different situation: one in which the user knows how to describe desired information items according to searchable characteristics, and desires to have such items retrieved. In this case, having entered the appropriate script, the system asks the user to specify the value(s) of the desired characteristics, from some general choice of characteristics. The user chooses one or more and specifies them, the system retrieves and displays items 'matching' the specification, and the user chooses those which are appropriate.

From these few examples, we can see that the pattern of interaction between system and user differs greatly between the different regions of the ISS space. These differences arise in terms of what each partner asks of the other, and what each offers the other, and in the general structure of the interaction. Notice that the script in each case is tailored to taking advantage of what the user knows, in order to help the user to accomplish the desired goal.

Of course, our examples are simplified in that they are successful at each point in which there is a possibility of non-success, each ISS script is independent and self-contained, and the user's problem, goal and knowledge are constant. These are clearly unrealistic simplifications. Relaxing these simplifying assumptions leads us to the concept of a single information seeking episode consisting of a sequence of several such scripts, as was described in section 2.

4. Information Seeking Episodes

A realistic concept of human information seeking behavior would respond to the observation that people, in such activities, routinely change from one ISS to another. This can be due to their goals, problems, and knowledge changing, through the course of the interaction, or because their overall information seeking plan for the particular episode required a sequence of ISSs to achieve the overall goal which led them to the activity. From the point of view of our model, this can be seen as the combination of scripts, into patterns of information seeking episode interactions. Such combinations might be pre-planned or pre-stored in a system, but they are much more likely to arise in response to the specifics of any particular episode. Below, we offer an example of an interaction which combines the three IIS scripts described in section 3, and then discuss how it might be possible to predict and support potential moves from one place in the space of ISSs to another.

Consider the case of a person whose goals and knowledge correspond to those associated with ISS5. This person would like to look at a book that she had seen last week, but cannot remember its author or title. She does remember that it was on the bottom shelf of some set of stacks on the third floor of the library. Being in a progressive institution, the library catalog allows her to engage in an ISS5 script with the specified location as region to be scanned. While looking over these items, the user sees an interesting item that she didn't know about before. This phenomenon is usually called serendipity, and is something that every information system tries to support. This item looks so interesting, that the user would like to see more like it. Unfortunately, the other items nearby, that she is scanning, don't appear to be very similar to it, at least in ways that are relevant

to this user. Now the user finds herself in the position of wanting to learn how to describe this item in a way that will help her to find others like it. This is the situation that ISS12 responds to. So the user selects the item of interest, and informs the system that she would like to engage in a new ISS. This might be accomplished, for instance, by selecting the appropriate ISS script from a menu of alternative scripts that the system keeps displayed. Once engaged in the ISS12 script, the user asks to see the subject headings associated with this book, and the headings related to them. She finds one of the terms in the display to be relevant to her interests, and so wishes to see what other documents indexed by this term look like. This brings her to ISS15, which supports her in searching to select items in an information resource according to some specified characteristic(s). In this case, ISS15 might be invoked by the user selecting the appropriate term from the display, and informing the system that this should be used as a search key, again from the menu of ISS scripts. The system, having been informed of the user's desires, performs the search as specified, and displays the other items indexed by the chosen term. Clearly, such movement from ISS to ISS could continue indefinitely, but we'll bring it to a halt here. For an example of different combination of scripts, see Belkin, *et al.* (1993).

Predicting, or constructing such sequences is clearly a problem for the design of a system based on these principles. One way to address this problem is to make each script in the system always available for selection by the user, at each interaction move. A slight variant of this approach has been followed in the interface design suggested by Belkin, Marchetti & Cool (1993). But with many available scripts, and with the many ways in which they could be implemented highly dependent upon the context of the interaction, this might become rather cumbersome.

An alternative approach is to base the script combinations on likely branching points, such as points in the interaction where the system might not be able to display the desired response because of problems with the input, or no match between what the user is able to specify and what is in the database. A typical example of this situation is when a user specifies a search term which is not used in the database. Such failure points can be anticipated by detailed analysis of the moves in each script, and potential remedial strategies (e.g. movement to other scripts) can be enumerated. In the example given, for instance, the user might be prompted to begin again with a different search characteristic, or it might be suggested that she consult a thesaurus in order to find different terms.

This leaves us still with the problems of supporting movement from ISS to ISS based upon the user's overall goal and plan in entering the system, or upon unanticipated changes in user's problem, goal and knowledge which arise during the course of the interaction. Some of these can again be anticipated, based upon analysis of what people are likely to do when presented with particular kinds of information, and what problems they are likely to face in attempting to achieve particular classes of goals. A general characterization of information seeking goals, and a related cognitive task analysis is one way to address these problems (*cf.* Belkin, Marchetti & Cool, 1993). Another is through empirical observation of instances of interaction patterns, or storing of such patterns for use in guiding similar new interactions. This, of course, suggests the use of case-based reasoning techniques for helping to structure and organize interaction sequences.

5. Strategies, Cases, Interaction

We have suggested that the concept of a space of ISSs can lead to the specification of a relatively small number of prototypical interaction sequences, or scripts, which can be used to guide effective user interaction in IR systems. Two problems with this approach are the specification of the prototype scripts, and the support of interactions which require multiple sequences of scripts. We suggest that using cases of real interactions is a possible way to address each of these problems.

These problems can be addressed in the case-based approach by collecting numbers of cases, analyzing them to determine to which region of the ISS space their different parts are relevant, and characterizing each such sequence as a series of dialogue moves. With sufficient data, or as data are gathered, general patterns for the different regions can be induced. That is, some empirical observation in typical information seeking situations is required in order to begin the original system design. The concept of ISSs provides a means for structuring such observation, and for analysis of the data which would lead to standard interaction sequences, or scripts. And once the system is operational, collecting and maintaining the cases of interaction could lead to learning about new interaction sequences, both within a single region of the ISS space, and among them.

Another means to address these problems is to make use of cases of information seeking episodes directly, to drive the current interaction. The problem here is how to decide which case (or part of a case) is relevant to the given situation. We suggest that the ISS approach provides a means to address this particular problem.

A single, real information seeking episode, a *case*, could be described according to the scripts whose patterns it follows. That is, it could be indexed according to these scripts, effectively, the region of the ISS space. Then, the case itself is a concatenation of scripts, or parts of scripts. This means that each bit of the case is indexed by its region or script. Users, on entering the system, can choose a region appropriate to their current goal and state of knowledge. This instantiates a script, and all of its associated cases (at least those cases which begin in this script). The details of this person's case; that is, their input, the knowledge they have, and other characteristics of the person, then provide search points for retrieving the cases most likely to be relevant from this subset of all of the cases. These cases, with their individual sequences with the ISS and among ISSs, can then serve as the basis for identifying potentially appropriate branching points and paths for this particular user, which can be offered to the user at the relevant times. As the interaction progresses, new cases are retrieved and displayed or made available, according to the choices made by the user in her information seeking path. Furthermore, the matching of specific details between the current case and the cases relevant at any point, such as search terms used, can lead to the system's proposing highly specific courses of action in any given situation. This general pattern thus suggests how individual cases can be indexed and used effectively in support of new cases of information seeking interaction. Note that many of the problems normally faced by the case-based approach are addressed in this paradigm by relying rather heavily on user interaction with the system, in particular in asking the user to make appropriate choices among suggested options.

6. Conclusions

The concept of a space of information seeking strategies seems a good candidate for a basis for designing IR systems which support effective and direct user interaction. It seems especially promising in that it moves away from the idea of developing queries which represent more-or-less static information needs, to a more integrated concept of supporting users in all of their information seeking behaviors. A major problem with this approach, identifying and structuring such interactions, can be addressed by application of aspects of the case-based reasoning approach. And some major problems with the potential application of the case-based reasoning approach to IR appear resolvable through embedding it in the highly interactive ISS structure. Thus, the appropriate combination of these two approaches seems to us to be quite likely to lead to the design of IR systems which effectively support individual information seeking. Belkin, et al.. (1993) and Tißen (1993) provide examples of how the detailed problems of this combined approach can be addressed, and of what the benefits of such an approach can be. The next step, evaluation of this approach, awaits.

7. References

Bates, M. J. (1989) The design of browsing and berrypicking techniques for the online search interface. *Online Review,* 13,5: 407-424.

Belkin, N.J. (1982) Models of dialogue for information retrieval. In: I. Friberg, ed. *Proceedings* of the 4th International Research Forum in Information Science. Skrifter från Högskolan i Borås, Borås, Sweden: 15-36.

Belkin, N.J., Cool, C., Stein, A. & Thiel, U. (1993) Scripts for information seeking strategies. *This volume.*

Belkin, N.J., Marchetti, P.G. & Cool, C. (1993) BRAQUE: Design of an interface to support user interaction in information retrieval. *Information Processing and Management,* 29,4: in press.

Belkin, N.J. et al.. (1990) Taking account of user tasks, goals and behavior for the design of online public access catalogs. In: *Proceedings* of the 53rd ASIS Annual Meeting. Medford, NJ, Learned Information: 69-79.

Belkin, N.J. & Vickery, A. (1985) *Interaction in information systems.* London: The British Library.

Croft, W.B. & Thompson, R.H. (1987) I^3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science,* 38,6: 389-404.

Ellis, D. (1989) A behavioral approach to information retrieval system design. *Journal of Documentation,* 45,3: 171-212.

Hancock-Beaulieu, M. (1990) Evaluating the impact of an online library catalogue in subject

searching behaviour at the catalogue and at the shelves. *Journal of Documentation,* 46,4: 318-338.

Ingwersen, P. (1992) *Information retrieval interaction.* London, Taylor Graham.

Oddy, R.N. (1977) Information retrieval through man-machine dialogue. *Journal of Documentation,* 33,1: 1-14.

Salton, G. & Buckley, C. (1990) Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science,* 41,4: 288-297.

Tißen, A. (1993) How to use cases in information seeking processes. *This volume.*

Thiel, U. (1990) *Konversational graphische Interaktion mit Informationssytemen: Ein sprechakttheoretischer Ansatz.* Ph.D. Dissertation, University of Konstanz, Germany.

Walker, D.E. (1971) *Interactive bibliographic search: The user/computer interface.* Montvale, N.J., AFIPS Press.

Wersig, G. (1979) The problematic situation as a basic concept of information science in the framework of the social sciences - a reply to N. Belkin. In: *New trends in informatics and its terminology.* Moscow, VINITI: 48-57.