A MODEL FOR INFORMATION EXPLORATION

JOHN A. WATERWORTH

Institute of Systems Science, National University of Singapore, Heng Mui Keng Terrace, Kent Ridge, Singapore 0511

MARK H. CHIGNELL

Department of Industrial Engineering, University of Toronto, Toronto, Ontario M5S 1A4 Canada

A three-dimensional model of information exploration is presented. By pointing to three distinct dimensions of exploration, we attempt to clarify the respective roles of the human and the system in browsing and information retrieval, and to characterise alternative interaction styles to maximise retrieval effectiveness. We illustrate the applicability of our model of exploration by describing eight paradigmatic cases of information exploration that represent different vertices of the model. We also briefly describe an initial experiment that attempted to assess the pragmatic impact of various combinations of exploration features. We then discuss methods for integrating hypermedia and information retrieval into general information exploration systems and discuss some of the processes that people use in information exploration. We conclude with a plea for the development of hybrid information systems combining exploration features in the most appropriate way according to the task needs of users.

1. INTRODUCTION

Despite the currency of broad definitions of information retrieval such as that of Salton and McGill[1], 'Information retrieval is concerned with the representation, storage, organisation, and accessing of information items', the concept has become closely related with keyword-based querying of indexed bibliographic databases using Boolean logic. However, information exploration is a broader activity that is carried out in a variety of ways by different people (elsewhere we refer to this as 'information seeking'; *see* Waterworth and Chignell[2]. For instance, browsing through books, library shelves, or hypermedia documents are all forms of exploration activity. Our main goal in this paper is to introduce a new model that covers a broad range of information exploration styles and strategies.

What do we mean to convey by the term 'information exploration'? In a recently published dictionary based on current British usage (COBUILD [3]), the word 'explore' is defined as follows: 'If you explore a place, you travel there because you have not been there before, or because nobody has been there before, in order to find out what it is like'. So a key element of exploration according to common usage of the term is the notion of travelling to a region to find what (information) is located there. If one is actively concerned with selecting the route to a

given location during the process of exploration, then one can be said to be engaged in navigation: 'If you navigate, you work out which direction to go while you are travelling . . .' (COBUILD [3]). From hypermedia enthusiasts, we frequently hear the terms 'navigate' and 'browse' used synonymously. But this obscures an important distinction between choosing routes and the reasons for making those choices. Contrast the definition of 'browse' with that of 'navigate' above: 'If you browse you look at several things . . . in a casual, unhurried way, in the hope that you might find something interesting' (COBUILD [3]). By our account, browsing is opposed to querying, which we would redefine as 'exploration with a specific target in mind'.

Our model of information exploration brings out this important distinction between responsibility for selecting routes and the purpose of the exploratory behaviour. We also resolve a long standing confusion between these two aspects of locating information and the manner in which choices made during exploration are expressed, i.e. the interaction method. We go on to illustrate the use of the model with paradigmatic examples of each type of exploration behaviour, and with a preliminary experiment that examines the effectiveness of different types of information exploration.

After the model description and experimental test we then address the issue of how information exploration systems may be developed. We propose an extended model of hypermedia that includes mediated search, along with a referential style of querying based on index linking that may provide a smooth transition between browsing and querying. We conclude with a discussion of patterns of exploration behaviour, showing how Ellis's[4] model of information seeking may be incorporated within our information exploration approach.

2. A THREE-DIMENSIONAL MODEL OF INFORMATION EXPLORATION

2.1 Structural responsibility

Discussions concerning user interfaces to information technology tend to blur distinctions between user perspectives on the task and the perspective from the system's point of view. For instance, the apparently intuitive concept of navigation reflects this distinction. Navigation is unstructured from the system perspective but structured from the user perspective. This dichotomy is a direct result of which agent (i.e. the user or the system) is responsible for carrying out the search. In the case of navigation, users are responsible for controlling the search process and as a result it is they, rather than the system, that must be aware of the

structure of the information. This role is reversed in the case of traditional information retrieval, where it is the system that is responsible for searching and which must consequently be concerned with structure. The issue of who is concerned with structure represents a primary dimension of exploration that we will refer to as *structural responsibility.*

2.2 Target orientation

We find evidence for a second dimension of exploration is contrasting the activities of browsing and querying. Browsing is distinguished from querying by the absence of a definite target in the mind of the user. We refer to this second dimension of exploration as *target orientation.* Thus the distinction between browsing and querying is not determined by the actions of the user, or by the configuration of the system, but by the cognitive state of the user. Presumably, there is a continuum of user behaviours varying between querying and browsing that is characterised by the level of specificity of the user's informational goals. Given the existence of this continuum, it may be inappropriate to arbitrarily classify user behaviour as either browsing or querying and build systems that reflect this strict dichotomy. This view of browsing suggests the need to merge conventional information retrieval with browsing, rather than the seemingly prevalent view that browsing is best implemented as user search within an unstructured (associative) network.

The view that browsing involves an absence of target specificity in search is orthogonal to the distinction between navigation and information retrieval. Thus our analysis has so far revealed two dimensions of information exploration in hypermedia systems. The definition of browsing focusses on the distinction between targetted and discovery-based information exploration. In contrast, the definition of navigation emphasises the responsibility of the user, as opposed to the system, for dealing with structure.

Since, as we have defined them, these two dimensions are clearly orthogonal, both browsing and querying can occur as part of navigation or information retrieval processes. Thus users may use queries for navigation (although this possibility has generally been ignored in previous approaches to hypermedia and information retrieval), and may be said to browse the contents of a database using information retrieval techniques. This indicates four distinct exploratory activities that may occur (see Table 1) but, as we discuss below, it is possible to develop combinations of these activities that represent intermediate points on the respective dimensions.

TABLE 1. *Two dimensions of information exploration*

	Targetted	*Discovery*
User Handles Structure	NAVIGATIONAL QUERY	NAVIGATIONAL BROWSING
System Handles Structure	MEDIATED QUERY	MEDIATED BROWSING

2.3 Interaction method

A third dimension of exploration arises from the method of interaction used in the interface to the information system. While methods of interaction may be differentiated in various ways, a major distinction can be drawn between descriptive interfaces, where the user describes what is wanted, and referential forms of interaction, where the user selects or refers to what is wanted (generally using some variant of a menu). Descriptive interfaces have generally been associated with querying behaviour in traditional information retrieval style, whereas referential interfaces have generally been associated with browsing during

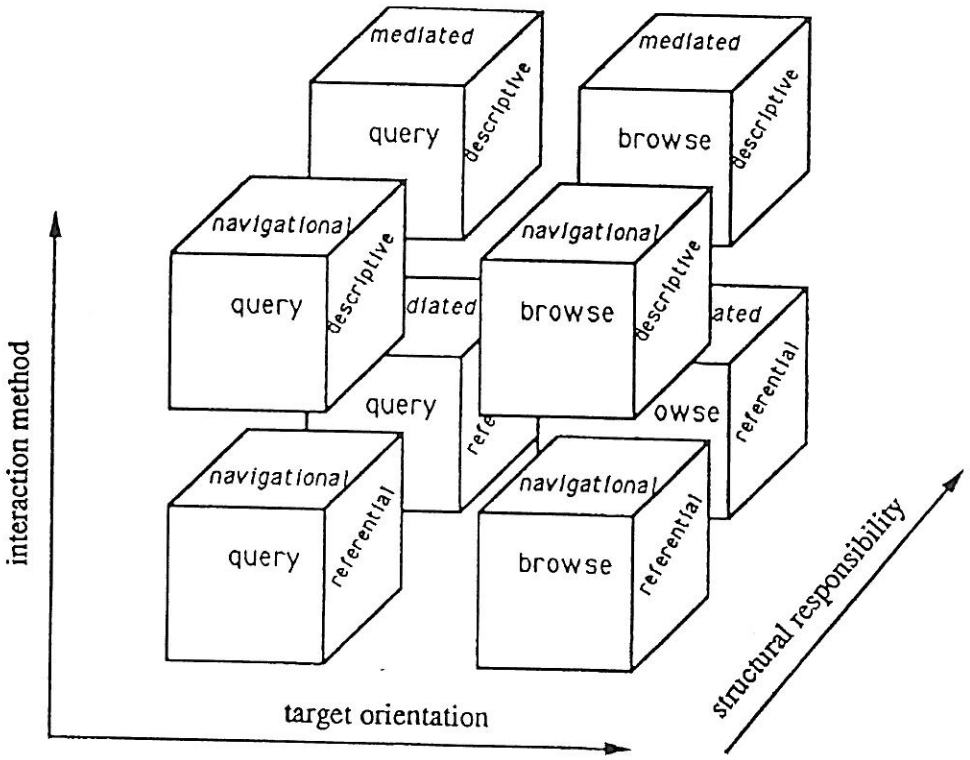

FIG. 1. *3D model of information exploration*

navigation, but there is no intrinsic correlation between the interaction method and target orientation or structural responsibility. Thus we have a third dimension of information exploration, *interaction method,* as shown in Fig. 1.

In our view, this three dimensional model clarifies the role of concepts such as navigation, browsing and querying in terms of fundamental components of information exploration. Other approaches tend to incorporate both structural awareness (on the part of the user) and a discovery target orientation within the definition of navigation as an activity. By focussing on navigation as a process where structural responsibility rests with the user, and contrasting it with mediated search, it is possible to consider research approaches to improving the effectiveness of browsing and navigation such as:

(i) Increase navigation effectiveness by increasing the awareness of structure.
(ii) Increase the effectiveness of browsing by assisting the process of choosing between different elements in the current structure.

The fundamental problem of navigational browsing as defined here is that one has to decide on where to go next within an information structure without having a clear a *priori* target in mind. This places a great deal of reliance on the immediate context of the surrounding structure in navigation decisions. Navigation requires an awareness of what the structure is and a familiar, or easily usable, mechanism for selecting where to go next based on that awareness.

3. PARADIGMATIC EXAMPLES OF INFORMATION EXPLORATION

In this section we describe various examples of information exploration in terms of our model. For present purposes we ignore the role of interaction method which, as we noted earlier, is orthogonal to our other two dimensions. We thus focus here on four styles of exploration:

(i) Navigational browsing,
(ii) Navigational querying,
(iii) Mediated browsing, and
(iv) Mediated querying.

3.1 Navigational browsing

Discussions of users interacting with hypertext systems typically assume a type of navigational browsing (e.g. Yankelovich *et al*,[5]) where one topic tends to lead into another. In this type of exploration the user is moving through the information without a clear target in mind. Often in this type

of situation there is a motivation to find information, but the user has the sense that while they can't formulate their information need as a query, they will recognize relevant information when they see it.

Consider the case of someone who wants to write an essay on Renaissance painting and the cultural influences that affected it. The following is an example of a navigational browsing style of interaction. Note that we are not necessarily concerned with any system actually in existence, but rather with how a form of navigational browsing might be implemented in principle.

> The user is presented with a menu of topics, and chooses — *Artists*.
>
> The user is presented with a choice of different categories of artist, and chooses — *Renaissance painters*.

Note here that we are already carrying out navigation, even though we are using menus. This is because the user is interacting directly with the system structure. The question of whether the user is browsing or not at this stage is a little more problematic. In some senses the user is operating in a querying fashion since the topic of 'Renaissance art' is in fact a target, but within this topic the user will have to browse. Thus in this example the user navigates to the topic neighbourhood in a querying fashion and then browses for information within that neighbourhood. This two-stage type of information exploration (where one finds the neighbourhood first) appears to be a useful strategy for dealing with large amounts of information (Shute *et al.*[6]). Browsing within a well-defined subtopic (neighbourhood) will generally be much more efficient than browsing the complete information system.

> The system shows a menu of Renaissance painters, the user chooses — *Michelangelo*

At this point the Michelangelo "node" is displayed and the user finds out that Michelangelo was a sculptor as well as a painter.

> The system shows links leading to a number of topics on Michelangelo, the user chooses — *Sistine Chapel*

The user now reads the information on this topic and finds out that one of Michelangelo's masterpieces (the ceiling of the Sistine Chapel) was commissioned by the church, thereby explaining the content of that work. By browsing information on other Renaissance painters he will find a similar influence. Depending on what information is available, he may also discover that the sponsors of paintings often appeared along with members of their families as historical or biblical figures in the

paintings. Thus by browsing the information, the user might eventually reach the conclusion that Renaissance art was strongly affected by religious subjects and by the economics of sponsors and their preferences.

This example demonstrates that querying may precede browsing in information exploration, and that relevant information may be obtained even when a target is not clearly defined a *priori*.

3.2 Navigational querying

In the next paradigmatic example of exploration, an a *priori* target is present, but the style of search is still navigational. Assume that we again have an information system on artists, where different pieces of information are connected by menus and reference links. In this example, the user knows that he wants to get information on the religious influence on Renaissance art. In this case, rather than choosing renaissance artists, he might begin by choosing 'art and religion' to get an appropriate background on the topic. He might then choose a topic such as '*Christian art prior to 1600*'.

The point here is that navigation, whether it be through menu selection or link selection, is actually a process of choosing where to go next. In any exploration situation the user will actually be somewhere on the continuum between knowing exactly what they want to find (querying) and having only an extremely vague idea of what they are looking for (browsing). Thus in our navigational query example, the user has a more detailed definition of his own requirements. In the case of navigational browsing he only knew that he was looking for cultural influences on Renaissance art, whereas with navigational querying he could define the topic fairly precisely. In both cases though, the search mechanism involved direct choices by the user (where the choices could be amongst menu items or links).

3.3 Mediated browsing

In the case of mediated browsing our hypothetical user again knows only that he wants to get information on cultural influences on Renaissance art. The difference here though is the agent of search, which has now become the system rather than, as in navigation, the user. So the user makes queries instead of navigational choices. Thus the effective question that the system poses the user at each point in the search is not 'where do you want to go next?' but 'what sort of information do you want me to get?' Thus we might have the following interaction:

User: Search Renaissance

System: Over 200 topics retrieved, be more specific
User: Search Renaissance and Art
System: The following topics were retrieved:
1. Perspective in Renaissance Art
2. Leonardo da Vinci
3. The golden age of Florence
. . .

After reading about interesting looking topics, the user might refine the query until the topic became clearer to him or until he has viewed a sufficient number of information items for his purpose.

3.4 Mediated querying

In mediated querying the user has a much better sense of what the target of the search is. Presumably this will allow better specification of the query, provided that there is some understanding of how to form the query, what descriptors are appropriate, and so on. There is plenty of evidence, however, to suggest that querying approaches such as Boolean retrieval are difficult for many users (e.g. Borgman,[7]) and thus navigation may be needed as a supplement to mediated search, even in cases where the user knows what the information target is.

In the following cases we show examples of successful and unsuccessful mediated querying.

User: search (Renaissance and Art) and (Religion or Christianity)
System: the following nodes are retrieved.
1. art in churches
2. church sponsorship of Art
3. Islamic art during the Renaissance
4. the Sistine chapel

Notice that in this example the user formed a query that directly produced relevant information. In both examples there might be some trial and error in getting the right query. For instance,

User: search Renaissance
System: More than 200 topics retrieved, please be more specific
or
User: search Renaissance and Religion and Christianity
System: No topics retrieved.

These examples illustrate both the good and bad features of mediated

querying. Firstly, if the information has been indexed the right way, a good query might get you straight to the relevant information. On the other hand, the demands of building the right query can hinder searching in some cases. For instance, in the example above, the user appears to be using the AND operator (Religion and Christianity) when really they should probably be using 'or'. This is a fairly common problem in naive searchers.

4. EXPERIMENTAL COMPARISON OF FOUR EXPLORATION STYLES

As an example of a study that attempted to clarify some of the issues we have raised, we briefly describe in this section an initial experiment on exploration styles in relation to task demands. In fact, this serves to illustrate the difficulties of empirical work in this area, though we remain convinced of the need for further behavioural work.

In this study we compared users' performance in solving two types of problem when forced to adopt either a mediated or a navigational style of exploration. The problems were designed to encourage either a browsing or a querying approach, since one set of problems was meant to require more interpretive answers than a second set. In other words, we were testing the following four combinations:

(i) Enforced mediated search with 'query' problems,
(ii) Enforced mediated search with 'browse' problems,
(iii) Enforced navigation with 'query' problems, and
(iv) Enforced navigation with 'browse' problems.

4.1 Materials

The material used was 'Hypertext on Hypertext' published as a Special Section of the July 1988 Communications of the ACM, and featuring papers from the Hypertext '87 conference. We used the Hyperties version produced by Ben Shneiderman and his team at the HCI Laboratory, University of Maryland.

Questions designed to encourage a querying approach were as follows:

Q1. How many times does the word 'tailorability' occur?
Q2. What is the title of the Conklin article published in 1987?
Q3. What are the 4 basic components of hypertext/hypermedia according to the article by Halasz?
Q4. Who is Ben Shneiderman?

Questions designed to encourage a browsing approach were as follows:

B1. Estimate about how many links there are between pages of text in the whole system.
B2. What are the major characteristics of hypertext/hypermedia?
B3. Choose four sentences from the text that best capture the essence of hypertext/hypermedia.
B4. What are the main advantages of using hypertext/hypermedia?

4.2 Subjects and Method

Our subjects were 16 Computer Science undergraduate students working on projects at the Institute of Systems Science, Singapore. Although all were Computer Scientists, none were specialists in the field of hypermedia or information retrieval.

Each of the subjects completed all 4 problems of each type and encountered enforced query ('Please follow the links to answer the following questions') and enforced mediated search ('Please search for the following') problems. So subjects were aware of the structural responsibility aspect of the task, but not of the two problem types to which they were exposed. Order of administration of problem and search conditions were counterbalanced across subjects to control order effects, each subject attempting two problems for each of our four experimental conditions according to a within subjects experimental design.

The main dependent variable was the time taken to complete a problem, or abandon a problem. A secondary measure was subjects' ratings of how difficult they found locating a particular solution under the various conditions.

4.3 Results

The ratings indicated that subjects generally found the 'browse-encouraging' questions more difficult than the 'query-encouraging' questions whichever strategy was used to try and find an answer. This was confirmed by the timing data from the subjects, where query questions were generally solved significantly more quickly than were browse questions ($F[1,15]=4.70, p<0.05$).

The data on exploration strategy were less clear cut, but suggested that mediated search is efficient and popular for solving not only specific (query) problems, but also for more general exploration (browsing) to extract the information needed to arrive at more interpretive types of answer. On average, navigation was rated as more difficult than searching, and also resulted in longer completion times, although these effects only approached significance for completion time scores

($F[1,15]=3.22, p<0.10$). This may reflect the very high variability in the amount of time different subjects spent on each problem.

Table 2 illustrates the completion time results for each problem type and exploration strategy. This illustrates the finding that browse problems were generally more difficult than search problems, and the suggestion that searching is more effective than navigation for both types of problem.

TABLE 2. *Browsing versus searching for two problem types (mean problem solving time)*

	Question Type		
Strategy	Browse	Query	Both
Navigate	1172s	950s	1061s
Search	1063s	496s	779s
Both	118s	723s	

4.4 Discussion

There are problems in carrying out controlled studies that contrast different information exploration strategies. In the first place, there is something intrinsically paradoxical about setting 'browse-encouraging' problems and then taking success as the speed with which an answer could be found. Recall the definition of 'browse' cited earlier: 'It you browse you look at several things . . . in a casual, unhurried way, in the hope that you might find something interesting'.[3] Insofar as browsing is, by definition, unhurried, a successful browsing session would be indicated by long completion times, not short as our study assumed. Nor would success be judged by whether a particular answer was found, but rather by what interesting items were found along the way.

This points to a fundamental problem with assessing which information exploration strategies are best for browsing. Our results suggest that mediated search is more effective than navigation in finding answers to two types of problem (though it should be stressed that only the effect of problem type was significant). But it seems likely that neither of these problem types could be said to elicit true browsing behaviour, since we have defined browsing under our model as exploration with no definite target in mind.

In this study we forced users to either navigate or use mediated search. In this situation, and which particular questions to answer, navigation appears to have no advantages for either problem type. This suggests that the strong emphasis on navigation found in the hypermedia

literature may be inappropriate. We do suspect, however, that navigation may have advantages not identified by our small study. Users needing to gain an overview of a subject area, or wishing to take advantage of serendipitous discovery, might be cases in point. Even so, it is not certain that these aims can best be achieved through navigation. Non/navigable structural overviews, and a degree of random selection of contents for display, might achieve these two aims, but further experimentation is needed to resolve this issue. It seems pretty clear, though, that any information exploration system should provide mediated search in addition to navigation, if users are ever going to want to access specific items of information.

Further studies are required to tap different aspects of information exploration behaviour, especially in relation to loosely formulated informational needs. More data on user satisfaction with different approaches to exploration would be valuable here. Additionally, instead of comparing the situation where users may either search or navigate, we need to develop hybrid systems that combine or merge the two in various ways, and then compare user performance and satisfaction in a wider variety of exploration situations.

5. DEVELOPING INFORMATION EXPLORATION SYSTEMS

In this section we expand on our taxonomic model by considering directions for the development of practical systems that integrate desirable features for information exploration. We have already pointed to the need for systems which combine traditional information retrieval capabilities with the flexible linking characteristics of hypermedia. We develop this theme in three ways. Firstly, we suggest that extending the notion of a hypermedia link will provide the necessary additional structure to support navigational querying behaviour. Secondly, we re-emphasize the view that mediated search should be integrated into the hypermedia paradigm. Index linking is proposed as a way of avoiding the heavy reliance on descriptive interaction that might be expected from this integration. Finally, we focus on patterns of behaviour that would typify various aspects of the information explorer's task.

5.1 Hypermedia and information retrieval

One of the dilemmas for hypermedia enthusiasts is how to promote hypermedia as a general information seeking environment. For historical reasons, perhaps, hypermedia has developed as a counterpoint, rather than a complement, to mainstream information retrieval. Thus, while conventional information retrieval systems have used mediated querying,

hypermedia developers have emphasized navigational browsing (Fig. 2). How can our conceptualization of hypermedia and information retrieval be extended to provide more general information exploration mechanisms?

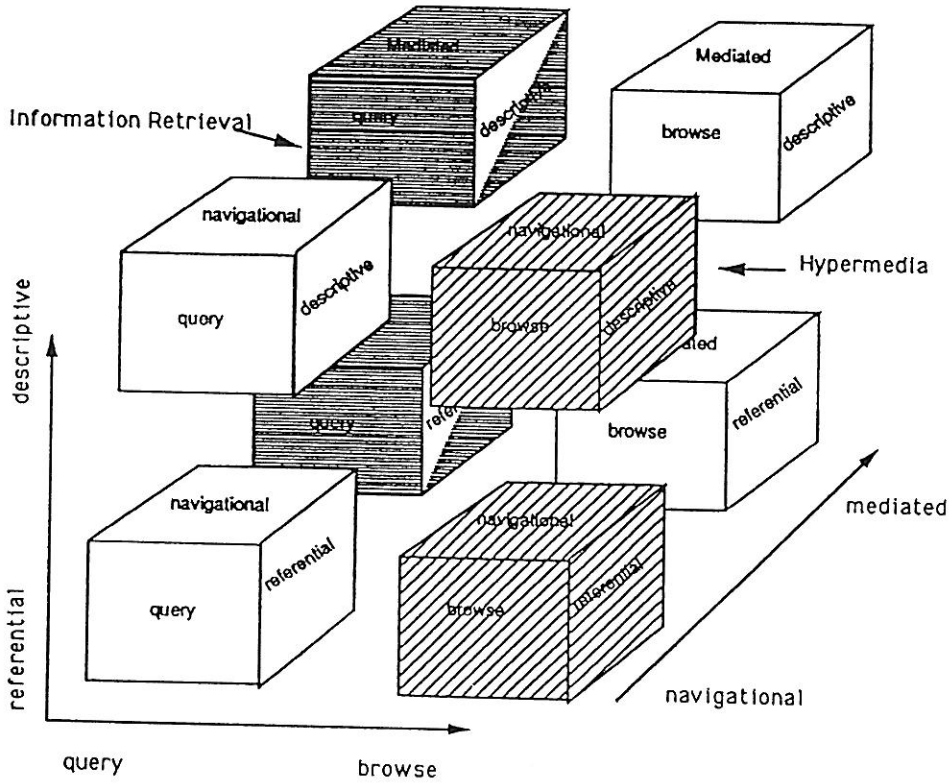

FIG. 2. *Three dimensions of information exploration: highlighting the components emphasized in information retrieval and hypermedia, respectively.*

In our view, the hypermedia model can be extended to provide mediated search and querying capabilities, if we extend our view of what a hypermedia link is and how it functions.

The most frequently used hypermedia link is a reference link of the form 'Node A refers to Node B.' The form of reference is often unspecified, but is assumed to be associative in nature. Unspecified reference links seem to be most appropriate for undirected browsing, which may work for entertainment (e.g. interactive fiction), or even serendipitous learning, but perhaps not for more structured tasks.

Some have looked to a more structured set of reference links as a way of expanding the types of task that hypermedia can handle. Various link

types may be used to promote structure such as paths and hierarchical links. Fig. 3 (after Parsaye *et al*,[8]) shows three embedded representational structures that may be used for different tasks, or different aspects of the same task. The linear path is useful for following a line of reasoning, while the network provides a browsing capability and the hierarchy provides an index structure and allows the user to explore categorical relationships.

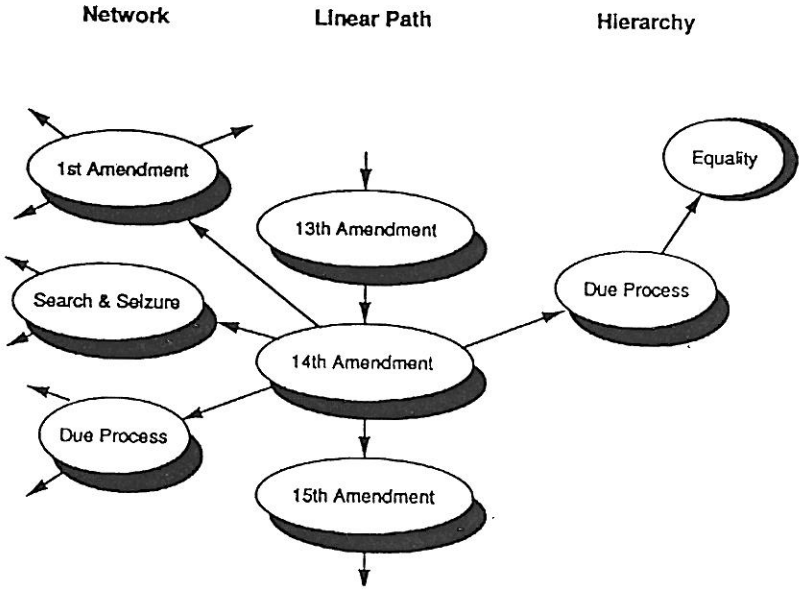

FIG. 3. *Different types of organizational structuring of the fourteenth amendment.*

Additional structure might be achieved by seeing hypermedia as a more general knowledge representation environment (Chignell and Valdez, in press[9]). Here the hypermedia would have the type of linking representative of concepts and their interrelations that has been used in the semantic network approach to knowledge representation (e.g. Norman and Rumelhart,[10]).

The structure provided by these sophisticated linking techniques may permit a style of navigational querying where the user traverses links to find specific information (as in the Renaissance article discussed above). However, even in this style of interaction there appears to be a need for an entry point into the hypermedia. Some developers have got around this problem by simply defining a start-up node or 'gateway' that is the same for all users at all times. Often this gateway node is actually a 'welcome screen' which then leads into a menu.

Command menus can be thought of as a compromise between navigational and mediated search. here the choice element of referential selection (usually reserved for navigation) is preserved, but the effect of choice is to implement a descriptive command, similar to those that are written in mediated search. Nonhierarchical menus that contain lists of alternative topics provide a presentation style that facilitates information exploration by partitioning the available navigation choices from the nodes in which they occur. Thus menus tend to serve as signposts in hypermedia, and as bridges between navigation and mediated search.

The use of menus in a combined model of information exploration is one of the most challenging aspects of integrating hypermedia and information retrieval. Shneiderman[11] has used the term embedded menu to describe the use of highlighted text to denote the available links within a hypermedia node. These embedded links can of course be shown as highlighted icons in a graphical display. However, we would like to distinguish here between what we shall refer to as intrinsic menus and extrinsic menus.

Intrinsic menus point to related information based on the content of the current node. In contrast, extrinsic menus tend to point to topics (usually related to or following on from the current node topic) that exist as other nodes in the hypermedia. Thus extrinsic menus supplement the signalling in intrinsic menus and reference links by providing more general access to topics within hypermedia. Topics within an extrinsic menu may vary in relatedness from being closely related (reference linked) to topics in the current node to being almost totally unrelated. In the latter case, the unrelated menu items may serve as convenient access points to general regions or topics in the hypermedia thereby breaking the 'tyranny of the reference link.' These unrelated menu items are then functionally equivalent to a mediated search on the topic that they represent. The selection of such a menu item might lead directly to the most salient node on that topic, or it might lead to a set of nodes on that topic being retrieved, which can then be browsed in some fashion.

Expanding the reference link to include a more general knowledge representation function, along with using extrinsic menus to improve access within hypermedia, can obviously assist navigation. Extrinsic menus also provide a convenient way of merging mediated search and navigation. However, basic navigation necessarily is a 'link by link' process. This type of process is inefficient if the user has some information target in mind, and more conventional mediated querying is available. This may explain why searching for information in a linear fashion is often preferred to navigating hypermedia by link traversal.[12,13]

5.2 Index linking

Another approach to making hypermedia more usable is to incorporate mediated search into the hypermedia paradigm.[14,15] One strategy for doing this is to index the nodes as if they were documents in a bibliographic information retrieval system. Index-based mediated search may then be used to access relevant nodes in the hypermedia. However, the descriptive style of interaction used in Boolean search systems and relational databases has been shown to be difficult for many users. It is possible then to define a more usable style of mediated search that provides appropriate topic access starting points? Our solution to this problem is to propose an 'index linking' process, as described below.

We assume that the nodes in the hypermedia system are described by a set of index terms. These index terms may be used as the basis for automatic linking, but here we will focus on their use for information exploration in an existing hypermedia document. We can imagine the index as a separate entity that is connected to the hypermedia by 'index links' that emanate from each node to the terms in the index that it is indexed by. The important point to note in the index linking model is that mediated search is launched from inside the hypermedia rather than being launched from outside the information as is generally the case in information retrieval.

Thus, the user may be at a node, and he spots an index link that appears relevant. he may now move to the index via that link. On reading the content of that index node and determining that it is relevant he may then add it to his 'hypermedia query'. At this point, all the nodes that are indexed by that term are assigned a higher weight (made relatively more accessible), as they would be in other information retrieval systems. In index linking though, the index itself is organized as navigable hypermedia. Thus the user may trigger mediated search by browsing through the index and marking index nodes that appear to be relevant. In other words, he has achieved what is normally attempted through descriptive interaction (which can be problematic for the non-specialist), by means of referential interaction.

An index term may be an aggregate that explains the use of the term in different contexts. For instance the term mercury may explain its use in different contexts (the name of a planet, an element of the periodic table, etc.). The user may then select the index term within one of more of these contexts. This type of context selection and restriction serves some of the functionality of the NOT operator in Boolean querying.

Working within the index, the user sets up weightings that will then apply in the hypermedia. Restriction of context is handled by selection

from within aggregated index nodes. At this point the index weighting functions as an additive linear model. However, the user may at any point return to the hypermedia either by returning to one of the nodes previously visited or by moving to one of the highly weighted nodes based on the current selection of nodes within the index. Once within the hypermedia, the user may refine the query by providing relevance feedback of two types. Firstly, through the act of link traversal, nodes visited will have the weightings of their index terms boosted. Second, by operating on the index terms directly, individual terms will be flagged as relevant (this can happen within a hypermedia node or within the index).

Our description of index linking makes no commitment to a particular style of user feedback. There are a number of ways in which user feedback might modify term weights and link availability. For instance, relevance feedback and term weighting methods might be used.[16] Or, more speculatively, Bayesian updating of belief, or connectionist models (Frisse,[15] p.884).

Given an appropriate method of user feedback, index linking is a model for combining navigation and mediated search into an integrated paradigm. With index linking the navigational style of movement between nodes is preserved, but there is also the possibility of mediated search and of an interesting form of constrained navigation where the nature of the current query limits navigational choices while the act of navigation itself modifies the currently defined query.

Ideally, information exploration strategies should be implemented in an environment where smooth transitions between browsing and querying, and between navigation and mediated search are possible. Index linking appears to be a useful strategy for making this possible. An initial approach in this direction has been described by Thompson and Croft.[17] Their I3R system contains a model of the information seeking process involving three main steps:

characterizing the user (e.g. expert or novice)
characterizing the information need
searching for relevant documents

The search for relevant documents then consists of search for documents, followed by an evaluation phase which may in turn be followed by further search. Within the search process itself, users may build queries directly or may paste together query phrases based on words extracted from documents within the system. In addition, users may search by concepts (index terms) as well as by documents. One of the interesting features of the I3R system is that it is organized as a blackboard system, with a number of 'experts' including a control expert, a user model

builder, a domain knowledge expert, and a browsing expert. However, smooth integration between browsing and querying is not yet available in current information retrieval or hypertext systems.

5.3 Patterns of exploration behaviour

The model of information exploration we have presented is based on a logical analysis of the information seeking operations possible in information retrieval systems and hypertext. However, the effectiveness of an information exploration system is ultimately determined by the behavioural patterns adopted by its users. As we have suggested, exploration should be a flexible combination of browsing and querying, navigation and mediated searching; the point of such combinations, and the basis for the design of exploratory systems, is the information exploration behaviours that can be supported. In this section we expand on our taxonomic model of the dimensions of exploration by considering in more detail various exploration behaviours in relation to that model, drawing heavily on the work of Ellis.[4]

Based on extensive interviews with social science researchers, Ellis[4] found evidence for a behavioural model of information seeking. Ellis's study was influenced by the fact that it was conducted with researchers who used conventional (i.e. non-hypertext) information retrieval systems. Thus the findings may underestimate the amount of browsing that would occur in a more balanced information exploration environment that included both browsing and querying capabilities. Even so, the component information seeking processes proposed by Ellis appear to be highly compatible with the general model of information exploration described in this paper.

Ellis's model of information seeking encompasses six component activities:

Starting
Chaining
Browsing
Differentiation
Monitoring
Extracting

5.3.1 Starting

Starting is a process of looking for key references or expert guidance that can provide the right orientation to a search. According to Ellis, starting might involve discussion of a list of possible search terms with an expert,

or identification of a book or article that is particularly relevant and a good starting point for the exploration.

From our perspective, we can contrast starting with a query versus starting with a browse. In a querying style of search starting involves formulating the search target or topic. In a more browsing style of search this involves identifying the starting point for the search (i.e. the initial context for browsing). The starting process might consist of menu-based selection of a start node, followed by browsing of hypertext to get a sense of the topic of interest.

5.3.2 Chaining

Chaining consists of following referential connection between material. Obviously navigation through sequences of hypertext links is somewhat analogous to chaining. However in chaining the user tends to follow a sequence of links of a particular kind. For instance, a form of chaining can be achieved using citation indices, where one can either move backward from one text to a document that is cited within it, or forward from a citation to the texts that cite it. In either case the user traverses a chain of citation links.

The distinction between forward and backward chaining during information exploration is also applicable in navigation through a hypertext that contains asymmetric links (where one moves backward by finding the nodes that have links to the current node). We shall refer to the navigation process where a variety of link types are used as 'linking' to distinguish it from link specific types of 'chaining.'

5.3.3 Differentiation.

In browsing and querying, differences between sources of information may be used as filters on the nature and quality of the material examined, or to restrict the context so that only relevant information is made available. Ellis refers to this process as differentiation. For instance, certain authors, publishers, or journals are usually more highly regarded in a topic area, and thus are given more attention during information exploration. In mediated search, differentiation can be achieved by restricting the search to certain databases, and in browsing by highlighting nodes that have been created by highly regarded authors or that have been extracted from particular sources.

5.3.4 Monitoring

Users may also wish to maintain awareness of developments in a field through the monitoring of particular sources. In a general information

exploration environment that is updated over time, users should be able to review the new information that is relevant to their interests. One way to do this is to maintain a profile of the user's interest. New information that matches this profile is then passed on to the user. In mediated search systems this profile consists of a query and new documents are matched against this query. Since monitoring implies that the user has a target in mind, it is applicable to querying rather than browsing. A navigational approach to monitoring is, however, possible. Nodes that are new, or that have not been previously visited by the user may be highlighted. Navigational monitoring would then consist of navigating through the new information in the network from time to time.

5.3.5 Extracting and evaluating

Extracting, where the user reads through the information source to find material of interest, is the final behavioural component of Ellis's model. Extraction is a major problem with information retrieval systems that only contain abstracts or titles. The user must locate the hard copy of the original source which in some cases may take considerable time. In contrast, extraction is an integral part of hypertext navigation, since the contents of the current node may be viewed in their entirety.

One vital additional behaviour to consider is evaluation. Through evaluating outcomes, the user determines the quality of the information found and the effectiveness of the retrieval strategy (including the index terms if mediated search is being used). For instance, relevance feedback may be used to modify a query based on the user's evaluation of mediated retrieval. Evaluation may also occur during navigation, although it may not be public or observable in terms of selection of documents or explicit modification of a query. After reaching a node a user may decide that it is not leading in a promising direction and backtrack to a prior node, or the user may find that a particular type of link tends to lead to more interesting nodes and start 'chaining' along that type of link.

5.3.6 Target specificity

A major distinction in our model is between querying and browsing, distinguished by the degree of target specificity in the mind of the user. From a behavioural point of view, the starting point of querying is target identification, while browsing begins with a starting context which is typically much less specific (e.g. go to the table of contents of the encyclopaedia and look for an interesting topic).

Fig. 4 provides a visual description of the information exploration

behaviours we have considered above, in relation to the distinction between browsing and querying. This complements the static model of the dimensions of information exploration that we showed in Fig. 1.

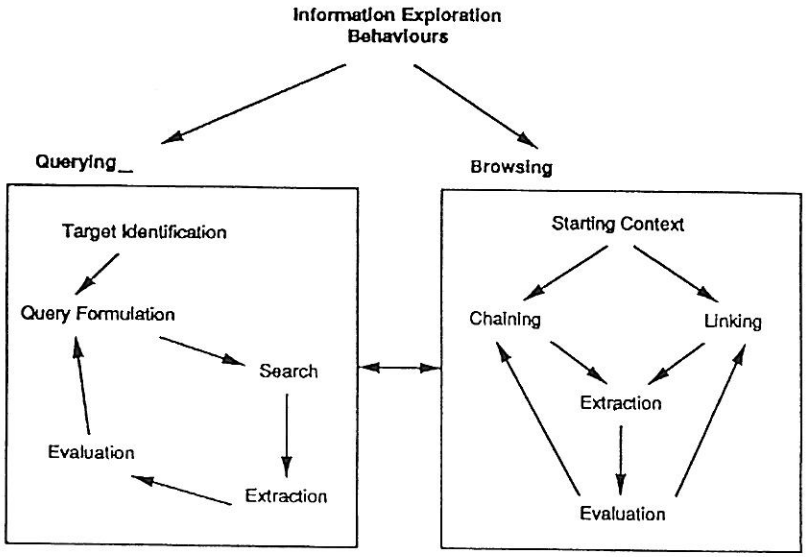

FIG. 4. *A characterization of information exploration behaviours*

In querying, target identification is followed by one or more cycles of query formulation, search, extraction, and evaluation. For browsing, selection of the starting context is followed by either chaining or linking and then by extraction, evaluation and further chaining or linking. The double headed arrow between querying and browsing in the figure indicates that during information exploration users should be able to switch strategies without too much difficulty.

6. CONCLUSIONS

Our three dimensional model provides a focus for empirically testing hypermedia information retrieval tools in relation to task needs. Similarly, the model provides an approach to testing the closely related questions of how information can best be structured for different tasks, and of how different structures can best be conveyed to users and navigated by them. Some directions for future empirical research on these usability issues are described elsewhere.[2]

We believe that it is possible to develop systems that capture all three of our dimensions of information exploration. As an initial step, a concept-based hypermedia system with mediated search could be

developed. A prototype hypermedia system along these lines has already been developed at the Institute of Systems Science.[14,18]

However, the fact that systems can be defined which merge querying with browsing and navigation with mediated search says little about how they should be built in order to enhance usability and information exploration effectiveness. Adding new capabilities for information exploration may not always make users more productive. For instance, after several years of development hypermedia systems have yet to prove themselves clearly superior over linear text in a broad range of tasks. If anything linear text appears to have fared better in the few comparisons that have been carried out.

Much more human factors research is needed to determine how integrated information exploration systems should be constructed. We know very little about how people will explore information if tools that combine the features of hypermedia and of traditional information retrieval are available. This is not surprising since analogous tools have not previously been available. However, even in the absence of such tools types of exploration behaviour can be observed and classified. For instance, researchers carry out browsing in non-hypertext environments by looking through tables of contents of books, abstracting journals such as *Current Contents*, or by simply browsing through library stacks. More work is needed to identify the larger patterns of exploration that achieve users' task goals.

The failure of users to adapt to the Boolean querying requirement of many information retrieval systems stands as an instructive reminder of the need to craft information technologies to match the capabilities and requirements of the human user. The index linking approach introduced above promotes a hybrid form of exploration that includes a simpler style of querying that complements navigational browsing.

We predict that hybrid forms of exploration will increase in popularity as hypermedia and information retrieval systems become more sophisticated and sensitive to users' varied task requirements. We have attempted to represent the universe of interactive possibilities from which these hybrids will be designed. However, development of future information systems should be constrained by logical and behavioural analysis of information exploration. Based on the existing empirical evidence and our model of information exploration, we make the following recommendations.

If users know exactly what they are looking for (i.e. they are querying) then following links is not an effective way of finding the information and good mediated retrieval is needed. In addition,

structuring, such as the use of menu hierarchies, will help narrow down the topics of interest.

If users don't know exactly what they are looking for (i.e. they are browsing) then following linear sequences is more efficient than jumping around. Structuring may also help, but there is a flexibility trade-off where structure may 'channel' the search and restrict the breadth of exploration. If the user has at least a rough idea of the topic of interest mediated retrieval may help narrow down the area of exploration.

We can identify two major trends in information exploration systems, which may be referred to as 'expertext' and 'pilotext.' In practice, actual systems should consist of collections of features that can be mixed and matched according to the application needs.

Expertext is implicit hypermedia containing rules to aid information selection (through the use of conditional links), structure that is hidden from user (with an emphasis on mediated information retrieval), and associated text and media that are used to explain advice. Expertext would typically be aimed at consultative or Intelligent Tutoring System applications. Expertext and its advantages are reviewed by Rada and Barlow.[19]

Pilotext is explicit hypermedia featuring user-initiated navigation (although 'autopilots' may be used too), sophisticated structural models and browse tools (because of an emphasis on navigation). The target applications for pilotext tend to be explorational, such as education reference and publicly accessed information.

Expertext, Pilotext, and related technologies such as hybrid information systems may well revolutionize the ways in which people carry out information exploration. However, programmatic research is needed to determine what functionalities and structures will be most useful for information exploration. In this paper we have sought to provide a guiding framework for this research through our information exploration model, and to provide some initial proposals on how hypermedia and information retrieval may be combined.

ACKNOWLEDGEMENTS

We would like to thank Chung Tze Min and Sue Bowles for carrying out the experiment reported in this paper. Tze Min also undertook the statistical analysis of results. We acknowledge the assistance of Chua Tat Seng and Elaine Lai in discussing the properties of the ISS Hypermedia shell with us. This paper was written while Mark Chignell was a Visiting Scientist at the Institute of Systems Science. Material in Section 2 was presented at 'IT Works 90' Seminar, Singapore, June 1990.

REFERENCES

1. SALTON, G. *and* McGILL, M.J. *Introduction to modern information retrieval.* New York: McGraw-Hill, 1983.
2. WATERWORTH, J.A. *and* CHIGNELL, M.H. A manifesto for *hypermedia usability research. Hypermedia, 1*, 1989, 205-234.
3. *Collins COBUILD English Language Dictionary.* London: Collins, 1987.
4. ELLIS, D. A behavioural approach to information retrieval system design. *Journal of Documentation*, 45, 1989, 171-212.
5. YANKELOVICH, N., HAAN, B. *and* MEYROWITZ, N. Intermedia: The concept and the construction of a seamless information environment. *IEEE Computer*, January, 1989.
6. SHUTE, S.J., SMITH, P.J., KRAWCZAK, D.A., CHIGNELL, M.H. *and* SATER, M. Enhancing cognitive compability in a knowledge-based information retrieval system. *In: 1986 Fall Industrial Engineering Conference Proceedings*, 1986, 111-115.
7. BORGMAN, C.L. Why are online catalogs so hard to use? Lessons learned from information-retrieval studies. *Journal of the American Society for Information Science, 37*, 1986, 387-400.
8. PARSAYE, K., CHIGNELL, M.H., KHOSHAFIAN, S. *and* WONG, H.K.T. *Intelligent databases: object-oriented, deductive hypermedia technologies.* New York: Wiley, 1989.
9. CHIGNELL, M.H. *and* VALDEZ, J.F. Knowledge engineering for hypermedia. *In*: J.A. Waterworth, *ed. Multimedia: technology and applications.* In press.
10. NORMAN, D.A. *and* RUMELHART, D. *Explorations in cognition.* San Francisco: Freeman, 1975.
11. SHNEIDERMAN, B. User interface design for the HyperTies electronic encyclopedia. *In: Proceedings of Hypertext'87, Chapel Hill, North Carolina*, 1987.
12. McKNIGHT, C., DILLON, A. *and* RICHARDSON, J. A comparison of linear and hypertext formats in information retrieval. *In:* R. McAleese and C. Green, eds. *Hypertext: State of the Art.* Oxford: Intellect, 1990.
13. MONK, A.F., WALSH, P. *and* DIX, A.J. A comparison of hypertext, scrolling, and folding. *In*: R. Winder, *ed. People and Computers IV, Proceedings of HCI'88.* Cambridge: Cambridge University Press, 1988.
14. WANG, X. *and* CHUA, T.S. *Support for search and dynamic linking in a hypertext environment.* Internal Report. Singapore: Institute of Systems Science, National University of Singapore, 1990.
15. FRISSE, M.E. Searching for information in a hypermedia medical handbook. *Communications of the ACM, 31*, 1988, 880-886.
16. SALTON, G. *Automatic text processing: the transformation, analysis and retrieval of information by computer.* Reading, Mass.: Addison-Wesley, 1989.
17. THOMPSON, R.H. *and* CROFT, W.B. Support for browsing in an intelligent text retrieval system. *International Journal of Man-Machine Studies, 30*, 1989, 639-668.
18. CHUA, T.S. *and* LAI, E.P.M. Composition editor for hypermedia environment. *In*: J.A. Waterworth, *ed. Multimedia: technology and applications.* In press.
19. RADA, R. *and* BARLOW, J. Expert systems and hypertext. *The Knowledge Engineering Review, 3*, 1988, 285-302.

ADDITIONAL READING

KOH, T.T. *and* CHUA, T.S. On the Design of a frame-based hypermedia system. R. McAleese and C. Green, eds. *Hypertext: State of the Art.* Oxford: Intellect, 1990.

WRIGHT, P. *and* LICKORISH, A. An empirical study of two navigation systems for two hypertexts. R. McAleese and C. Green, eds. *Hypertext: State of the Art.* Oxford: Intellect, 1990.