What the Query Told the Link: The Integration of Hypertext and Information Retrieval

Gene Golovchinsky
FX Palo Alto Laboratory
3400 Hillview Avenue, Bldg. 4
Palo Alto, CA 94304
+1 415 813-7361
gene@pal.xerox.com

ABSTRACT
Traditionally hypertexts have been limited in size by the manual effort required to create hypertext links. In addition, large hyper–linked collections may overwhelm users with the range of possible links from any node, only a fraction of which may be appropriate for a given user at any time. This work explores automatic methods of link construction based on feedback from users collected during browsing. A full-text search engine mediates the linking process. Query terms that distinguish well among documents in the database become candidate anchors; links are mediated by passage–based relevance feedback queries. The newspaper metaphor is used to organize the retrieval results.

VOIR, a software prototype that implements these algorithms has been used to browse a 74,500 node (250MB) database of newspaper articles. An experiment has been conducted to test the relative effectiveness of dynamic links and user–specified queries. Experimental results suggest that link–mediated queries are more effective than user–specified queries in retrieving relevant information. The paper concludes with a discussion of possible extensions to the linking algorithms.

KEYWORDS: Dynamic hypertext, information retrieval, information exploration, browsing, relevance feedback, newspaper metaphor.

1 INTRODUCTION

Remde et al. [31] enumerate several deficiencies of traditional, linear text that motivated the creation of hypertext interfaces. Among other reasons, they suggest that it is "… too hard to find information in ordinary text," that it is "… too hard to acquire information in a sequence other than that determined by the author," and that it is "… extremely difficult to integrate and update large bodies of frequently changing information from many different sources."

These are criticisms of the user interface used to present linear text, yet the same criticisms have been levied against hypertext interfaces. Problems of disorientation [7], constraints on the reader imposed by static links [20], a limitations of the size of hypertext collections (e.g., [4]) are all well documented. In particular, limitations due to the difficulty of creating large numbers of manually–authored links have produced a number of efforts of automatic construction of hypertext structure.

1.1 Document structure and hypertext

Explicit document structure — often expressed as SGML tags — has been used to segment text into hypertext nodes and to create the hierarchical skeleton that organized the database (e.g., [30], [15], [4], [33]). A variety of systems have used automatic link suggestions to aid the author in creating links (e.g., [15], [2], [4], [27], [32]). These approaches were quite successful, but the dependence on human authors to supervise the link creation process limited the amount of information that could be processed. Thus HEFTI [4] was used to convert a 400 page book, less than 400KB of text. Similarly, Robertson et al. [32] report results for 60, 90 and 100 page documents.

Two classes of problems may be expected when such techniques are applied to multi–gigabyte collections of text containing hundreds of thousands of nodes. The amount of time required to supervise link creation, particularly in growing collections, is prohibitive large. Robertson et al. [32] for example, report 5 to 45 hour times to convert 60 to 100 page documents on a PC. The HEFTI process handled about 400KB in about a day. The time required to create links via such processes will render these techniques impractical for multi–gigabyte collections.

The second problem affects the reader more than the author. Very large collections of texts can be expected to contain vast numbers of potentially useful links. Unfortunately, most links would not be useful most of the time, and their sheer number can overwhelm even the most careful user interface design. Thus alternatives to semi–automatic link generation must be found.

1.2 Information retrieval and hypertext

The link creation problems described above suggest that more traditional information retrieval techniques must be used to accommodate large collections of nodes. Information retrieval methods have been applied to hypertext databases in several ways: link information has been used to inform retrieval algorithms (e.g., [8], [14], [19], [13], [36]), term co–location information has been used to suggest links to human authors (e.g., [2], [4], [32]), and queries have been used to retrieve hypertext nodes (e.g., [9], [12], [6], [5]). Relevance feedback has also been used to guide retrieval and to infer links among documents (e.g., [3]). Most systems that use queries as navigational aids use them to identify relevant neighborhoods in the hypertext, and then rely on manually–created links to support further navigation.

IR techniques have been used to segment long articles into shorter, more focused nodes (e.g., [34], [22]). Similarity among passages has been used to create links between specific nodes. This work, however, has focused on text segmentation techniques rather than on the hypertext interface. Although it is clear that such approaches are promising, little evidence has been published to date regarding their integration into interactive hypertext systems and about the effectiveness of such techniques in support of interactive browsing.

SuperBook, one of the more successful query–mediated browsing systems, used keyword queries instead of static hypertext links as a navigation mechanism [31]. Information was presented to the user in several windows, including table of contents (TOC), query, and text viewed. Users could use the TOC hierarchy to arrive at the desired section, or they could type in queries (or select keywords in the text). Search results were used to annotate the TOC to indicate relevant passages. Thus the system achieved hypertext–like browsing by combining TOC–based navigation with full–text search.

SuperBook's reliance on the table of contents to organize the browsing session limits it to providing access to highly–structured documents. Although an extension to SuperBook that works across documents has been demonstrated [28], it still relies on hierarchical structure of each document to support local navigation. SuperBook has been shown to be an effective interface for IR tasks when browsing structured collections [12], but alternatives to the book metaphor must be found to support browsing through loosely–structured hypertext collections.

One such alternative — the newspaper metaphor — is discussed in the following section, and VOIR, a prototype that implements it, is described. Some experimental results from an evaluation of VOIR are presented, and the paper concludes with a discussion of possible extensions and applications of this query–mediated hypertext interfaces.

2 VOIR

This section describes VOIR (Visualization of Information Retrieval), a prototype newspaper–based dynamic hypertext interface. The section first introduces the newspaper metaphor and discusses its implementation in VOIR. A description of VOIR's linking interface follows, and the discussion concludes with an overview of VOIR's visualization features.

2.1 The newspaper metaphor

Newspapers such as the Wall Street Journal are designed to present a variety of different, loosely–related articles in a manner that supports browsing and selective reading. The front page of each newspaper section provides an overview of the contents. It presents summaries of articles, with references to other pages where additional details are discussed. The layout of each broadsheet provides cues to the relative importance of articles: important articles are usually placed near the top of the page, and more column space is allocated to them. These layout features serve to alert the reader to potentially useful information, and to structure interaction with a text that does not possess an overall narrative.[1]

These features of a newspaper make it an appropriate vehicle for displaying hypertext information [18]. Users can capitalize on their familiarity with newspapers to browse hypertext collections. In addition to providing similarity–based structure, the newspaper metaphor can support the notion of landmark nodes [29] and hypertext links. The front page of a newspaper serves as a landmark around which semantically–related articles are organized. Articles split among several pages are connected with links. Overviews of contents are quite common. This

[1] Each article, of course, has an internal structure. The newspaper merely serves to bind these largely–independent narratives together.

suggests that newspaper–style interfaces should also be appropriate for some non–news hypertexts (e.g., on–line help).

It is possible to synthesize newspaper–like layouts to display documents related to a particular topic. Kamba et al. [25] demonstrated Krakatoa Chronicle, an electronic newspaper prototype that retrieved newspaper articles from an on–line source, displayed them in a multi–column layout, and recorded users' browsing patterns within the retrieved articles. Browsing patterns (e.g., the time spent reading a particular article) were used in a relevance feedback loop to amend a representation of the user's interests. This prototype showed the potential of the newspaper metaphor to organize the display of retrieved information. Their focus was on display, rather than on interaction, however. Each page took about one minute to compose, and users had to flip to a different screen to specify a new search topic or to modify the existing one explicitly. Thus the Krakatoa Chronicle appears to be designed to support filtering (query–routing) tasks: given a specification of a user's information need, the software can synthesize a daily (or hourly) newspaper.

VOIR was designed to encourage interactivity and to support iterative exploration tasks. Thus emphasis was placed on fast response time and ease of interaction. Users could specify their search intent in a variety of ways, including selecting visible text passages with the mouse, typing queries, or selecting hypertext links. Each interaction caused the system to display a new collection of articles. VOIR's screen (Figure 1) was divided into eight independently–scrollable text columns, each containing a retrieved article. This choice was somewhat arbitrary and depended in part on limitations of screen resolution and size. The exact number of columns used in a particular interface may, in principle, be determined by users based on their preferences and on the characteristics of their displays and of their tasks.

A full–text search engine (Inquery[2]) was used to index and retrieve documents. The rank order of retrieved articles was used to arrange them on each page, the highest–ranked article occupying the top-left column, the second–ranked occupying the next column to the right, etc. The size allocated to each column also reflected the article's importance.[3]

2.2 Dynamic links

VOIR identifies anchors dynamically based on users' topic specifications. Two approaches — the statistical and the heuristic — are used by the system to identify anchor candidates. Inverse document frequency (*idf*) scores are calculated for each query term, and terms with *idf* scores above a certain threshold are used as anchors. Intra–document term frequency is not used: the goal is not to find terms that are characteristic of any specific document, but to find terms that discriminate well among documents in a collection. The heuristic approach treats capitalized words (as they occur in the text) as anchor candidates because these terms (e.g., proper names, names of places, companies, etc.) may have specific meaning to users independently of their *idf* scores.

Query terms that match either the statistical or the heuristic criterion described above become anchors when they occur in retrieved articles. When an anchor is selected, its context (e.g., the sentence that contains it) is used to expand the previous query. That is, content–bearing terms that occur in the sentence containing the selected anchor are added to the terms of the previously–executed query. Term weights depend on the age of the term (on the number of links that have been followed since the term has been last introduced into the query), and on whether the term is an anchor or not. Terms that have not reappeared after three link selections are dropped from the query. Schemes that give more weight to anchors and to recently–added terms produce better recall and precision scores than uniform weighting schemes [17].

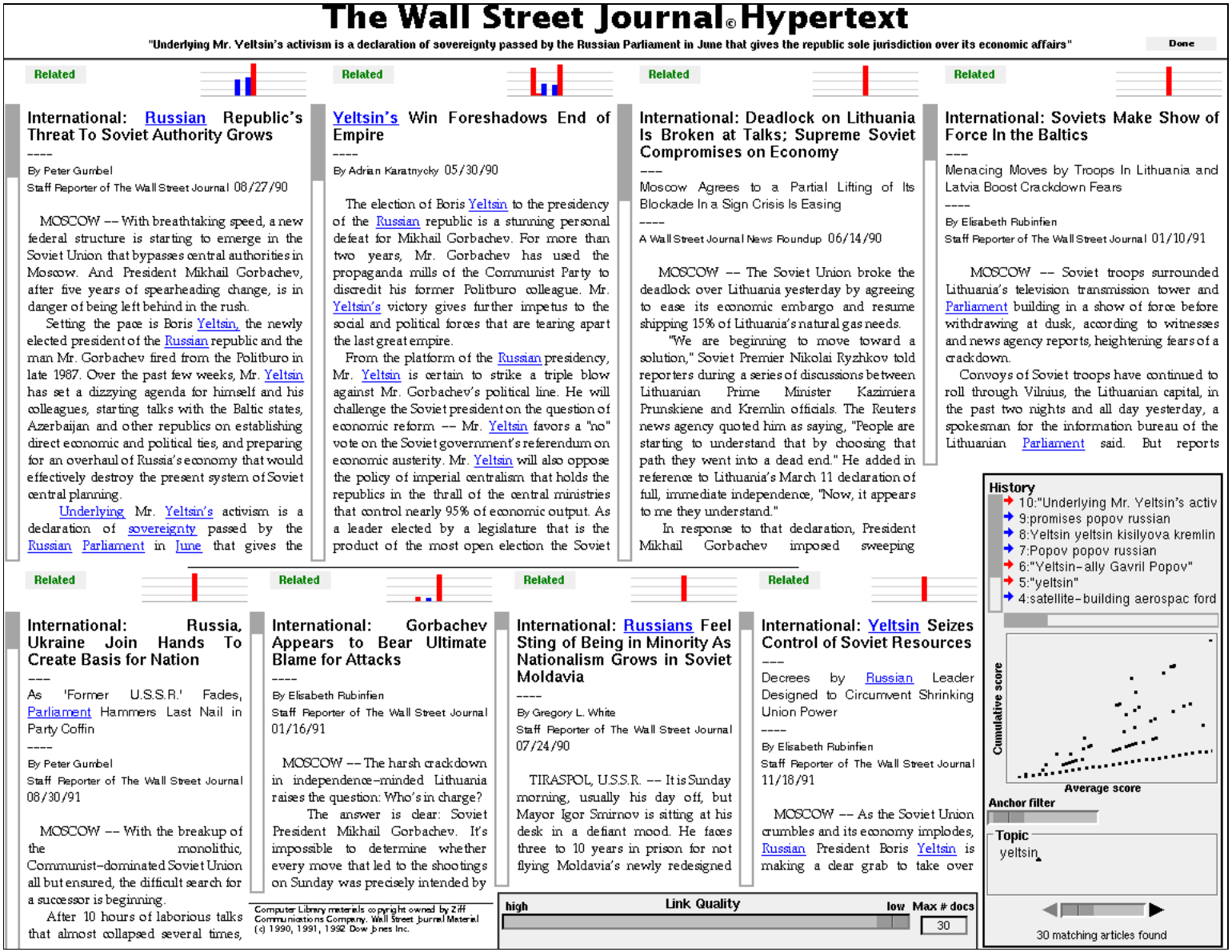
These anchor selections implement a sort of passage–based relevance feedback mechanism. Relevance feedback typically uses the entire document as a source of highly–discriminating terms [35], whereas this approach restricts candidates to terms occurring near the selected anchor. This allows the user to have more control over the evolution of the query. This approach is related to passage–based relevance feedback described by Allan [1]. The difference is that in VOIR relevant passages are identified by the user (via anchor selections) whereas Allan used best–matching passages as proxies for entire documents when performing relevance feedback.

---

[2]

[3]It is possible to make column size dependent on the scores used to measure article relatedness to a give query [25].

Figure 1. VOIR showing a browsing session about Yeltsin.



The sequence of interactions is depicted in Figure 2, where oval–shaped bubbles represent users' actions, and rectangles represent system responses. The user specifies a topic by typing or selecting an set of terms. The system retrieves and displays a collection of articles that matches the query. It also marks some of the query terms as anchors in the newly–displayed articles. The user now selects an anchor, causing the system to add terms from the context of the anchor to the query. The corresponding set of articles is then displayed, and new anchors are shown to the user. At any time during this browsing process, a new topic may be specified via drag–selection or typing. Terms comprising the new topic will replace the previous set of terms as anchor candidates.

VOIR also supports context–independent links. These links use document–based relevance feedback to identify other documents similar to the one for which a context–independent link has been selected.

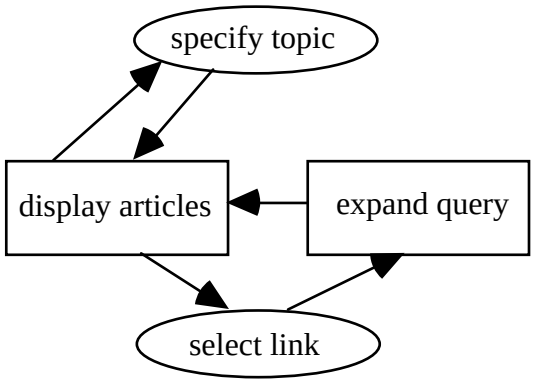


Figure 2. VOIR browsing interactions

2.3 Visualization

Graphical representations of browsing history play an important role in supporting information exploration tasks in VOIR. Disorientation in hypertext environments may occur when users follow cycles in the hypertext's graph [7]. Users may realize that they've seen certain information before, or they may treat previously viewed nodes as new information. VOIR provides a visualization of each node's display history within the current browsing session. The retrieval ranks of each node with respect to the sequence of queries and link selections are presented in a bar chart display next to each article (Figure 3).

Figure 3 depicts several typical situations which users may encounter during a browsing session. The node in Figure 3a has been displayed for the first time, and it is quite relevant to the search topic. Figure 3b shows a node being shown for the second time (and not being as relevant) after a number of links had been followed. Figure 3c characterizes a node that is quite central to the topic of interest, as indicated by the moderate number of tall bars. Link types are color–coded in the display history histograms: context–independent links are displayed in green, context–specific ones in blue, and new contexts (passage or typed queries) in red.

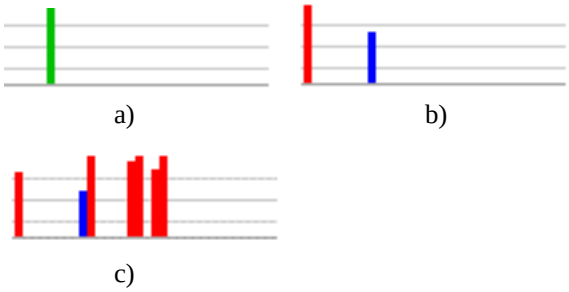


Figure 3. Bar chart displays of the retrieval histories of different articles.

Figure 1 shows several examples of these bar chart displays. The user is exploring a collection of Wall Street Journal articles, looking for information about Yeltsin and the independence of the Russian Republic. The screen displays articles related to a query about the declaration of sovereignty by the Russian Parliament. The histograms reveal at a glance that articles one, two and six have been viewed before, while the rest have been retrieved for the first time. The second article has been ranked quite high once before, as revealed by the first tall bar, whereas articles one and six have only been marginally relevant during the browsing session so far.

These graphical displays are designed to provide users with quick overviews of the relationships between nodes displayed as a result of a link traversal or query and previously–seen nodes. Users may follow several context–specific links, identify one or more landmark nodes (nodes with dense histograms), and then use context–independent links to display nodes similar to the one being viewed.

3 EXPERIMENT

A behavioral experiment was conducted to evaluate the performance of VOIR in an information–exploration task. Part of the TIPSTER collection [21] consisting of about 74,500 Wall Street Journal articles (about 250MB of text) was used as the experimental database. The experiment was designed in part to test the relative efficiency of dynamic (query–mediated) links versus natural language queries. Two instrumented versions of VOIR were created. They differed slightly from the original prototype: one version used queries only, and both had simplified controls. Up to 32 articles were retrieved for each query or link selection.

The experiment was a three–way between–subjects design. The experimental manipulation consisted of three levels of the interface factor (query only, links with explanation, and links without explanation). Subjects in the query condition were able to type queries and to make passage selections (i.e., to select text visible on the screen by dragging over it). Search terms were highlighted with a yellow background in the retrieved documents. Subjects in the two other conditions were presented with link anchors instead of highlighted terms. In one link condition, subjects were told that link selections were translated into query expansion, whereas in the other link condition, subjects were merely told that the interface included links.

3.1 Subjects

24 subjects participated in the experiment, and were paid $20 upon completion. Subjects were either graduate students at the University of Toronto, or recent graduates. All had prior experience with computers, and most had used a Web browser. Subjects did not have any formal training in information retrieval strategies.

3.2 Method

The experimental task consisted of retrieving as many relevant articles as possible within fifteen minutes given a topic description that characterized relevant articles. Subjects were given time to familiarize themselves with each set of search criteria prior to starting the search session. Each subject performed six fifteen minute searches. The topics and associated relevance judgments had been provided as part of our participation in the TREC workshop [21]. These relevance judgments were used to calculate precision and recall scores for each topic.

Three sets of recall and precision measures were used to characterize performance. *Retrieved* recall and precision were used to measure the effectiveness of queries made by the users; these correspond to traditional IR measures. In addition, *viewed* recall and precision were used to measure the numbers of relevant articles displayed on the screen, since not all retrieved articles were always viewed. Finally, *judged* recall and precision were used to characterize subjects' ability to identify articles relevant to the search topic. *Judged recall* was obtained by dividing the number of relevant articles selected by users by the total number of relevant articles, and *judged precision* was the ratio of the number of relevant articles selected to the total number of articles selected. *Retrieved* measures were designed to measure system performance, while *viewed* and *judged* measures assessed users' behavior. See [16] for a more thorough discussion of these measures.

3.3 Results

This experiment was designed to assess the effects of the different interface conditions on user performance and behavior, and to evaluate the performance of the various types of queries available to the users. No significant differences in recall or in precision were found for the interface factor (query, naïve, and informed), but significant differences in behavioral strategies were detected among users. Cluster analyses of browsing behavior indicated that subjects who adopted a skimming — rather than reading — strategy obtained better *judged* recall without sacrificing precision. These results are discussed further in [16].

The second purpose of this experiment was to evaluate performance of the various query types, and to examine their effects on users' behavior. Analysis of variance was performed with query type (context link, passage selection or typed query) as the main effect, and average recall and precision measures as the dependent measures. Averages were obtained by dividing the recall and precision scores for each topic by the number of interactions of each type (typed, passage selection or link) performed during that session. This normalization was required to de–couple user strategy (many vs. few selections) from the results of each selection. Data for this analysis were pooled from all three interface conditions.

Several important effects were found for the query type factor. Compared with passage selection, context links resulted in higher *retrieved* recall ($F[2,255]=4.63$, $p<0.011$), higher *retrieved* precision ($F[2,253]=4.55$, $p<0.012$), higher *viewed* recall ($F[2,255]=5.26$, $p<0.006$), and higher *viewed* precision ($F[2,253]=4.03$, $p<0.019$). Typed queries resulted in higher *viewed* recall than passage selection ($F[2,255]=4.63$, $p<0.011$) and higher *viewed* precision ($F[2,255]=5.26$, $p<0.006$) than passage selections. The differences in means were greater for links than for typed queries, and differences between links and typed queries were not significant. These results are summarized in Table 1. A total of 213 context link selections were made, compared with 274 passage selections and 362 typed queries. (These numbers include the query interface condition, in which context links were not available.)

Differences	Recall	Precision
Link > Passage	*retrieved, viewed*	*retrieved, viewed*
Typed > Passage	*viewed*	*viewed*
Link > Typed	Higher means, no sig. diff.	

Table 1. Comparison of performance by query type.

3.4 Discussion

The differences in recall and precision for the query types tested in this experiment suggest that query–mediated links can be effective information exploration tools when used in conjunction with typed queries or passage selections. Query expansion algorithms triggered by anchor selection were shown to be more effective than passage selection queries, and slightly (but not significantly so) better than typed queries. Further improvements in query–mediated link performance may be expected by incorporating more sophisticated query expansion techniques such as those described in [1].

The experimental task was designed as a compromise between experimental control and ecological validity. Although it is possible to construct more direct comparisons of the different query types, such experimental procedures tend to make the experimental task less representative of some “real” tasks. It is worth noting that experimental subjects enjoyed using the interface, and several had expressed the desire to use it for their own research. They found the interfaces intuitive, and reported that they liked having multiple mechanisms for expressing their search intent.

This research also suggests that it is possible to construct hypertext interfaces to very large text collections, and to preserve the interactivity and directness of hypertext interfaces while providing users with the power and flexibility of sophisticated information retrieval algorithms. Much additional research is necessary to determine which aspects of interfaces facilitate exploration, and what implications such interfaces have on the design of search engines.

4 EXTENSIONS

The information presentation techniques described above may be applied to a variety of information retrieval systems. In particular, there appears to be some synergy between newspaper metaphor–based interfaces and cluster–based retrieval techniques such as Scatter/Gather [10]. The newspaper metaphor may be applied readily to displaying document clusters identified with Scatter/Gather techniques. Hearst and Pedersen used Scatter/Gather to partition document sets retrieved by weighted sum queries [23]. They found improvements in performance between interfaces using Scatter/Gather and interfaces presenting the same documents in ranked lists. In the newspaper metaphor, each cluster may be associated with a separate newspaper section. Overviews could be created from excerpts from articles closest to cluster centroids. The clustering algorithm would be used to group articles together based on inter–document similarity; this grouping should convey a greater sense of semantic relatedness than is possible with only ranked document sets.

An additional benefit from the integration with such retrieval techniques is the availability of terms characteristic of retrieved clusters, terms that were not necessarily specified in users' queries. These characteristic terms may be used as anchors instead of (or in addition to) the salient query terms. Thus searching an abstracts database for "information retrieval" could identify a cluster of articles related to hypertext. The term "hypertext" could then be made into an anchor when displaying documents in that cluster. Other clusters would have other characteristic terms. These techniques are expected to approximate semantic links better than the existing term–frequency approach. Similar results may be obtained with LSI [11] or with association thesauri [24].

5 CONCLUSIONS

This paper describes VOIR, a query–mediated hypertext interface designed to support information exploration tasks in very large text databases. It describes a technique for mediating links with passage–based relevance feedback queries. Experimental results indicate that link–based queries produced better performance than passage selections, and that subjects found dynamic hypertext interfaces intuitive to use. This research suggests that integrating hypertext interfaces with full–text search engines can produce effective solutions for a class of information exploration tasks.

This work also has implications for models of information exploration (e.g., [37]) that posit a distinction between selecting anchors and forming queries. Interfaces such as the one described here suggest that the distinction between hypertext and information retrieval can become progressively blurred. As more sophisticated (e.g., agent–based) techniques are integrated into information exploration interfaces (e.g., [26]), some distinctions between semantic and statistical links should also disappear.

6 ACKNOWLEDGMENTS

This research was conducted as part of the author’s Ph.D. research in the Department of Mechanical and Industrial Engineering at the University of Toronto. The author wishes to thank Mark Chignell his helpful comments. This research was funded by a grant from the Information Technology Research Centre of Excellence of Ontario (ITRC).

REFERENCES

1. Allan, J. (1995) Relevance Feedback with Too Much Data, In *Proceedings of SIGIR '95* (Seattle, Wash.) pp. 337–343.
2. Bernstein, M. (1990) Link apprentice. In *Proceedings of ECHT '90*, INRIA, France, Cambridge Series on Electronic Publishing, pp. 212-223.
3. Boy, G. (1991) Indexing Hypertext Documents in Context. In *Proceedings of Hypertext '91*. San Antonio, Texas. ACM Press. pp. 51-61.
4. Chignell, M.H., Nordhausen, B. Valdez, F. and Waterworth, J.A. (1991) The HEFTI Model of Text to Hypertext Conversion. *Hypermedia*, 3 (3), pp. 187-205.
5. Christophides, V. and Rizk, A. Querying Structured Documents with Hypertext Links using OODBMS. In *Proceedings of ECHT '94*, Edinburgh, UK. ACM Press. pp. 186-197.
6. Clitherow, P., Reicken, D., and Muller, M. (1989) VISAR: A System for Inference and Navigation in Hypertext. In *Proceedings of Hypertext '89*. Pittsburgh, PA. ACM Press. pp. 293-304.
7. Conklin, J. (1987) Hypertext: an introduction and survey, *IEEE Computer*, 20 (9), pp. 17-41.
8. Croft, W.B. and Turtle, H. (1989). A Retrieval Model for Incorporating Hypertext Links. In *Proceedings of Hypertext '89*, Pittsburgh, Penn., ACM Press, pp. 213-224.
9. Crouch, D.B., Crouch, C.J., and Andreas, G. (1989) The User of Cluster Hierarchies in Hypertext Information Retrieval. In *Proceedings of Hypertext '89*, Pittsburgh, Penn., ACM Press, pp. 225-237.
10. Cutting, D.R., Karger, D.R., Pedersen, J.O. and Tukey, J.W. (1992) Scatter/Gather: A cluster–based approach to browsing large document collections. In *Proceedings of SIGIR '92*. pp. 318-329.
11. Dumais, S.T., Furnas, G.W., Landauer, T.K., Deerwester, S. and Harshman, R. (1988) Using Latent Semantic Analysis to Improve Access to Textual Information. In *Proceedings of CHI '88*, ACM Press, pp. 281-285.
12. Egan, D.E., Remde, J.R., Gomez, J.M., Landauer, T.K., Eberhardt, J. and Lochbaum, C.C. (1989) Formative Design-Evaluation of SuperBook. *ACM Transactions on Information Systems*, 7 (1) pp. 30-57.
13. Frei, H.P. and Stieger, D. (1992) Making Use of Hypertext Links when Retrieving Information. In *Proceedings of ECHT '92*, Milan, Italy, ACM Press, pp. 102-111.
14. Frisse, M.E. and Cousins, S.B. (1989) Information Retrieval From Hypertext: Update on the Dynamic Medial Handbook Project. In *Proceedings of Hypertext '89*, ACM Press. pp. 199-212.
15. Furuta, R., Plaisant, C., and Shneiderman, B. (1989) A spectrum of automatic hypertext construction, *Hypermedia*, 1(2), pp. 179-195.
16. Golovchinsky, G. (1997a) Queries? Links? Is There a Difference? In Proceedings of CHI '97 (Atlanta, GA).
17. Golovchinsky, G. (1997b) From Information Retrieval to Hypertext and Back Again: The Role of Interaction in the Information Exploration Interface. Unpublished Ph.D. Thesis, University of Toronto.
18. Golovchinsky, G. and Chignell, M.H. (in press) The Newspaper as an Information Exploration Metaphor, *Information Processing and Management.*
19. Guinan, C. and Smeaton, A.F. (1992) Information Retrieval from Hypertext Using Dynamically Planned Guided Tours, In *Proceedings of ECHT '92*, Milan, Italy, ACM Press, pp. 122-130.
20. Halasz, F.G., Moran, T.P., and Trigg, R.H. (1987) NoteCards in a Nutshell. In *Proceedings of ACM CHI + GI '87*, Toronto, Ontario. ACM Press. pp. 45-52.
21. Harman, D. (1995) Overview of the Third Text REtrieval Conference (TREC–3). In *Proceedings of the Third Text REtrieval Conference (TREC–3)*, D.K. Harman, ed. National Institute of Standards and Technology Special Publication 500–225, Gaithersburg, Maryland. pp. 1-19.
22. Hearst, M.A. (1994) Multi-Paragraph Segmentation of Expository Text. In Proceedings of the 32nd Meeting of the Association for Computational Linguistics, Los Cruces, NM.
23. Hearst, M.A. and Pedersen, J.O. (1996) Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. In *Proceedings of ACM SIGIR '96*, August, 1996, Zurich, Switzerland.
24. Jing Y. and Croft W.B. (1994) *An Association Thesaurus for Information Retrieval*, UMass Technical Report *94-17*, Center for Intelligent Information Retrieval, UMASS – Amherst. http://ciir.cs.umass.edu/info/psfiles/irpubs/jingcroftassocthes.ps
25. Kamba, T., Bharat, K. and Albers, M.C. *The Krakatoa Chronicle — An Interactive, Personalized Newspaper on the Web*. Georgia Tech Technical Report GIT-GVU-95-25.

26. Laurel, B., Oren, T. and Don, A. (1990) Issues in Multimedia Interface Design: Media Integration and Interface Agents. In *Proceedings of CHI '90*, Seattle, Wash., ACM Press, pp. 133-139.
27. Lelu, A. and Francois, C. (1992) Hypertext paradigm in the field of information retrieval: a neural approach. In *Proceedings of ECHT '92*, Milan, Italy, pp. 112-121.
28. Lochbaum, C. (1993) SuperCat in the SuperBook Document Browser System. Demonstration at *Hypertext '93*, Seattle, Wash., ACM Press.
29. Nielsen, J. (1990) *Hypertext and Hypermedia*. Academic Press.
30. Raymond, D. and Tompa, W.F. (1987) Hypertext and the New Oxford English Dictionary, In *Proceedings of Hypertext '87*, Chapel Hill, NC, ACM Press, pp. 143-153.
31. Remde, J.R., Gomez, L.M., and Landauer, T.K. (1987) SuperBook: An automatic tool for information exploration – hypertext? In *Proceedings of Hypertext '87*, San Antonio, TX. ACM Press. pp. 175–188.
32. Robertson, J. Merkus, E., and Ginige, A. (1994) The Hypermedia Authoring Research Toolkit (HART). In *Proceedings of ECHT '94*, Edinburgh, UK. ACM Press. pp. 177-185
33. Salminen, A., Tague–Sutcliffe, J., and McClellan, C. (1995) From Text to Hypertext by Indexing. *ACM Transactions on Information Systems* 1 (13), pp. 69-99.
34. Salton, G. and Allan, J. (1993) Selective text utilization and text traversal. In Proceedings of *Hypertext'93*, pages 131-144.
35. Salton, G. and McGill, M. (1983). *Introduction to Modern Information Retrieval.* McGraw-Hill.
36. Savoy, J. (1993) Searching Information in Hypertext Systems Using Multiple Sources of Evidence, *International Journal of Man–Machine Studies* 6 (38), pp. 1017-1030.
37. Waterworth, J.A. and Chignell, M.H. (1991) A Model for Information Exploration. *Hypermedia* 3 (1) pp. 35-58.