General and Efficient Strategies for Information Retrieval

Suresh K. Bhavnani
School of Information
University of Michigan
Ann Arbor, MI 48109-1092
Tel: +1-734-615-8281
bhavnani@umich.edu

Karen Drabenstott
School of Information
University of Michigan
Ann Arbor, MI 48109-1092
Tel: +1-734-763-3581
karen.drabenstott@umich.edu

Sashi M. Athota
Dept. of Electrical Engineering & Computer Science
University of Michigan
Ann Arbor, MI 48109-1092
Tel: +1-734-213-7601
sashi@umich.edu

ABSTRACT
To address the inefficient use of information retrieval (IR) systems such as search engines and library catalogues, we present a unified framework of strategies for information retrieval. This framework (1) contains a small set of general and efficient information retrieval strategies that are useful across many IR systems, and (2) can be used to identify key missing functionality in IR systems, and to design training approaches that lead to the efficient retrieval of information.

Keywords
Strategies, information retrieval, search, efficient.

INTRODUCTION
Despite huge advances in making information accessible to vast numbers of users, the efficient retrieval of relevant information remains a challenge. User studies of online library catalogues, abstracting and indexing systems, and web search portals, repeatedly show that despite knowledge of basic search techniques, many users do not acquire strategies to find relevant information efficiently [4]. For example, in an extensive study of Excite user queries [5], only 5% queries involved searches using the feature MORE LIKE THIS. This strategy, known as the *Citation Pearl Growing* strategy, is frequently used by expert librarians to quickly retrieve relevant information [3].

To enable users to become more efficient in information retrieval (IR), researchers have identified and taught several other effective search strategies for use in online library catalogues [1], and web search portals [3]. However, the identification of these strategies has been ad hoc, and therefore has not led to a unified understanding of general and efficient strategies. The lack of a unified framework of IR strategies directly affects how we train users, and how we design systems to support effective strategies. This paper presents a unified framework for general and efficient strategies, and discusses its implications to design and training.

A FRAMEWORK FOR GENERAL AND EFFICIENT IR STRATEGIES
Figure 1 shows 13 general and efficient strategies useful for tasks across different IR systems. The strategies fall under two main classes. One set of strategies exploits *regularities* in information spaces. Regularities include the tendency of authors to cite other authors that do related work and could therefore be of high relevance. Such regularities are *emergent* because they are based on the behavior of information authors. For example, Strategy-1 exploits such emergent regularities by reminding users to search for relevant information by accessing works cited by a relevant author. This strategy can be used by following footnotes in books retrieved through a library cataloguing system, through reading the reference list of papers acquired through full text retrieval in abstracting and indexing systems, or by using the FIND MORE OF THE SAME feature in a web search portal. Another form of regularity in an information space is *imposed* to create collections or consistency in the storing of information sources. For example, Strategy-5 exploits the imposed regularities of any indexed collection such as INSPEC containing mainly papers in engineering. The strategy calls for the use of simple keywords in successive queries to narrow down a retrieved set (also known as the *Successive Fractions Strategy* [3]).

The second set of strategies attempts to *enhance* the process of retrieval. For example, it is unreasonable to expect that a search process can automatically organize hits based on the categories that are meaningful to users. Therefore, users must find ways to organize the results of their retrieval at the time of reviewing the results of a search. Strategy-9 therefore advises users to organize results of a search in named hierarchies to enable easy access at a future time. Strategies 10-13 deal with other ways to enhance the retrieval process such as maintaining context during large searches, performing depth-first and breadth-first search, and avoiding searches during high traffic times. As shown in Figure 1, all 13 strategies are generally useful across the three main types of IR systems.

General Strategies	Online Library Catalogues	Abstracting & Indexing Systems	Web Search Portals
Strategies that exploit regularities in information spaces			
Emergent regularities			
1. Identify other authors through references	Retrieve paper and identify other authors	Retrieve paper and identify other authors	Use FIND MORE OF THE SAME to identify other sources
2. Identify works that cite current work	Access citation index to identify authors that cite a book	Access citation index to identify authors that cite a journal paper	Use link indexing to identify sites that point to current site
3. Identify other works of same author	Search by author name of known book	Search by author name of known paper	Search by organization name of known site
Imposed regularities			
4. Identify relevant collections	Select CMU library catalog when looking for books on robotics	Select INSPEC when looking for papers in Engineering	Use ditto.com when searching for graphic images
5. Search a collection using indexing scheme	Search within subject headings	Search within index	Search subject trees
6. Search a collection using keywords on successive sets of recalled hits	Search database using single keyword, then search within that result	Search database using single keyword, then search within that result	Search database using single keyword, then use "search within results" to narrow down search
7. Search multiple collections simultaneously	Use online computer library center (OCLC) to locate book in any university	Use DIALOG One Search to search multiple databases	Search using meta search engines
8. Guess name based on fields or naming conventions	Locate book by typing: guessed name "for dummies"	Locate journal by typing guessed field, "Journal"	Locate web site by typing: "www.", guessed name, ".com"
Strategies that enhance the process of retrieval			
9. Organize hits in hierarchies	Print out hits and organize according to relevant categories	Print out hits and organize according to relevant categories	Organize hits in hierarchical folder structure using SmartMarks
10. Search using context and focus views	Display previous sets in addition to current recall set	Display previous sets in addition to current recall set	Display tree of web sites found and visited
11. Search using depth and breadth	Keep record of sets at each level of search and return to previous sets	Keep record of sets at each level of search and return to previous sets	Display unvisited sets in different view and return to them to continue search
12. Setup dependencies to enable update notifications	Ask librarian to inform you of new books in particular area	Use profiles in DIALOG to set up notification for new papers in selected field	Use SmartMarks to set up notification for changes in selected URLs
13. Check traffic congestion before starting search	Avoid searching library systems during exam times	Avoid using DIALOG during overlap of east and west coast office hours	Avoid using Yahoo during overlap of east and west coast office hours

Figure 1. General and efficient IR search strategies and examples of their use across different information retrieval systems.

APPLICATIONS OF THE STRATEGY FRAMEWORK

The strategy framework enabled us to systematically analyze whether 6 web search portals (Yahoo, Lycos, Infoseek, HotBot, Excite, and AltaVista) supported the 13 strategies. We found that none of the web search portals supported strategies 9-13. Furthermore, even the better-known strategies (1-9) were not consistently supported. For example, neither Yahoo nor Excite provide a way to identify web sites that link to a page of interest (Strategy-2). These general strategies can therefore help designers systematically identify and make salient missing functionality to include on their web sites.

Because the framework provides the rationale behind each strategy, it can be used to develop training. Based on previous success in teaching general strategies in combination with specific commands [2], we hope to teach users not only how to efficiently retrieve information from an IR system, but also to transfer those skills across other IR systems. Towards that goal, we are currently developing a taxonomy of search tasks, which would help us to explicate how to sequence specific strategies in the framework for specific classes of tasks.

CONCLUSION

This paper presented a framework of general and efficient IR strategies. The framework can be used to systematically design functionality, and to develop a comprehensive approach to training with the goal of providing knowledge and developing systems that allow for the use of general IR search strategies. This should make users more efficient in their quest for relevant information regardless of the type of IR system they use.

REFERENCES

1. Bates, M.J. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13, 5, (1989), 407-424.
2. Bhavnani, S.K., Reif, F., John, B. Beyond command knowledge: Identifying and teaching strategic knowledge for using complex computer applications. *Proceedings of CHI'2001* (in press).
3. Drabenstott, K. Web search strategies. In *Saving the user's time through subject access innovation*, (ed.) Wheeler, W., The Board of Trustees of the University of Illinois, 2000, 114-161.
4. Fidel, R. Searchers' selection of search keys: I. The Selection Routine. *Journal of American Society for Information Science* 42,7, (1991), 490-500.
5. Jansen, B. J., Spink, A, Bateman, J., and, Saracevic, T. Real life information retrieval: A study of user queries on the web. *SIGIR Forum* 32, 1, (1998), 5-17.