

IR Research and Innovation in Commercial Online Systems: an Exploratory Survey

Sherry Koshman (E-mail: slkst11@icarus.lis.pitt.edu)
Edie Rasmussen (E-mail: emr@icarus.lis.pitt.edu)

School of Library and Information Science
University of Pittsburgh
Pittsburgh, PA 15260

Abstract

"Conventional" information retrieval systems (IRS), originating in the research of the 1950's and 1960's, are based on keyword matching and the application of Boolean operators to produce a set of retrieved documents from a database. In the ensuing years, research in information retrieval has identified a number of innovations (for example, automatic weighting of terms, ranked output, and relevance feedback) which have the potential to significantly enhance the performance of IRS, though commercial vendors have been slow to incorporate these changes into their systems. This was the situation in 1988 which led Radecki, in a special issue of *Information Processing & Management*, to examine the potential for improvements in conventional Boolean retrieval systems, and explore the reasons why these improvements had not been implemented in operational systems. Over the last five years, this position has begun to change as commercial vendors such as Dialog, Dow Jones, West Publishing and Mead have implemented new, non-Boolean features in their systems, including natural language input, weighted keyword terms, and document ranking. This paper identifies some of the significant findings of IR research and compares them to the implementation of non-Boolean features in such systems. The preliminary survey of new features in commercial systems suggests the need for new methods of evaluation, including the development of evaluation measures appropriate to large-scale, interactive systems.

I. INTRODUCTION

Research in information retrieval (IR) is characterized by a strong focus on the practical implementation of ideas in a real world environment. Underlying most experimental research is the motivation and the optimism that the features tested may eventually be used in the development of large-scale IR systems such as commercial vendors and library online catalogues.

The incorporation of new IR techniques into operational systems has been limited over the years, but this scenario has recently changed. It now appears that the link between theory and practice has strengthened. More innovative IR techniques are being put into practice in software and are manifest in online commercial vendor systems.

This observation forms the basis of the investigation described in this paper. Some of the significant findings of IR research are identified and compared to the implementation of non-Boolean features in operational commercial systems such as Dialog and Westlaw. The preliminary results from a survey of systems are presented, and the paper concludes with a discussion of implications regarding the implementation of new non-Boolean based features in large-scale operational systems and further work regarding an informal vendor survey to examine the motivation behind these implementations.

II. SIGNIFICANT FINDINGS OF IR RESEARCH

"Conventional" information retrieval systems (IRS), originating in the research of the 1950's and 1960's, are based on keyword matching and the application of Boolean operators to produce a set of retrieved documents from a database.

Boolean-based systems require the user to connect search terms with the logical operators AND, OR, NOT using the search syntax of the system. Problems with Boolean systems are frequently identified in the IR literature and include: 1) users find the syntax of entering Boolean queries problematic; 2) document output is not ranked; 3) Boolean queries may yield an overly large number of documents; 4) the query will not identify documents that only partially match document representations (Salton, 1989; Belkin & Croft, 1987).

In ensuing years IR researchers have developed alternatives to the Boolean model. For example, the vector model uses vectors of terms to represent queries and documents. Queries and documents are matched by calculating similarities between their vectors. The vector model accommodates term weights in query and document representations which are usually proportional to term frequency within the document (tf) and inversely proportional to the frequency of terms in the database (inverse document frequency or idf).

Term weights are a significant advance in IR research since they enhance retrieval techniques by facilitating ranked output and relevance feedback. To achieve a ranked document output, query-document similarities are calculated using term weights and the documents are ranked in decreasing importance or similarity to the query (Salton, 1989). Two standard similarity measures are the Dice and Cosine coefficients. Basic features of these measures are that only nonzero term weights in both document and query contribute to the similarity measure, and that they are normalized for document length.

Relevance feedback is based on the assumption that if the user identifies documents that are relevant to an information need, then the query may be modified to make it retrieve more documents similar to the relevant documents (Salton, 1989). Relevance feedback may be accomplished by adjusting term weights or adding new terms from the documents identified by the user (Belkin & Croft, 1987). Document-query similarities are recalculated for the revised query. With a query that more accurately reflects the user's needs, relevant documents should be more highly ranked. The query may be fine tuned by several iterations of the relevance feedback process.

The probabilistic model offers an alternative IR technique. Maron (1988) states that in a typical searching situation there are some documents

that are more likely to be relevant to the searcher than others. Further, the use of document 'clues' or attributes indicate or predict different levels of relevance (Maron, 1988). In a probabilistic model the query-document matching is calculated according to the probability of a document's relevance or nonrelevance to the query and the documents are ranked in order of their probability calculations. This provides a ranked output which may be useful to the user. Maron (1988) describes the appeal of the probabilistic model for commercial vendors, but practical implementation is difficult and it is used most successfully as a framework in developing innovative IR techniques. For example, Croft, Turtle and Lewis (1991) base their inference network research upon the probabilistic model.

IR researchers have also introduced methods to enrich the basic Boolean model such as incorporating term weights in the inverted index (the index structure used in most operational systems) to produce a range of values for a query-document similarity calculation (Salton, 1989). Salton (1988) introduced the extended Boolean model based on a p-norm model which essentially allows the Boolean query to retrieve a set of documents, and then uses a p-value for further matching operations. This research is significant in its application of term weights in the Boolean model. Term weights are more complicated in Boolean systems due to the query's logical operations. However, methods for establishing weights among the

different logical sets of operations in Boolean queries were introduced in order to achieve document output ranking (Salton, 1989).

As shown, IR researchers have introduced a number of innovations which would enhance operational Boolean-based systems, though commercial vendors have been slow to incorporate these changes into their systems. This was the situation in 1988 which led Radecki, in a special issue of *Information Processing & Management*, to examine the potential for improvements in conventional Boolean retrieval systems, and explore the reasons why these improvements had not been implemented in operational systems.

III. OPTIONS AND BARRIERS TO IR IMPLEMENTATIONS

Radecki (1988) called attention to the firm entrenchment of the Boolean model in operational retrieval systems. Boolean systems remain prevalent in operational environments since there is a large financial commitment to the systems in use and any alterations would require a significant investment in system redesign and user retraining. Further, alternative methods had not been tested for feasibility in large-scale systems and some of the techniques put forth by IR researchers did not suggest enhancements great enough to make them worthwhile to implement. Part of the problem may lie in the accessibility of IR research literature in providing information on alternative techniques. Also, users

seemed generally satisfied with Boolean-based search results (Belkin & Croft, 1987).

Smit and Kochen (1988), after a survey of vendors, cited two fundamental problems regarding the barrier to implementing new IR techniques: (1) gaps in the vendors' knowledge regarding practical information retrieval innovations, and (2) financial considerations especially if the vendor was not planning to implement any new techniques. More recently, Ledwith (1992) identified three factors which are prohibitive in applying IR research findings in ranked retrieval models to searching scientific files in large-scale operational systems such as STN. They include: (1) the test collection size and composition may be too small to render research findings generalizable to STN; (2) the broad nature of the search queries used for research purposes do not correspond to the types of queries found on STN; and (3) the overall cost/benefit ratio for improvements in ranked retrieval systems may not prove to be more beneficial to users than techniques found in traditional systems.

Recently, the Association of Information and Dissemination Centers (ASIDIC) conference featured vendors who highlighted innovations in commercial systems. The picture for incorporating innovative retrieval methods has changed as systems such as Dialog, DowQuest, Mead and Westlaw have implemented new, non-Boolean features in their systems,

including natural language input, document ranking and relevance feedback.

If certain vendors exhibit system changes, what are the nature of these changes and how do they draw upon the research conducted in the IR field? To address these questions, a preliminary survey of systems was conducted and an informal vendor survey is being developed to evaluate the innovations and motivations for change. The preliminary system survey includes the following questions:

System Features

1. a) Does the system allow **natural language** input?
 b) What is the input syntax: only search terms or a question format?
2. Are phrases identified?
3. Does the system allow **truncation**?
4. What type of **stemming** procedure is used?
5. How are **query terms weighted**?
6. How are **document terms weighted**?
7. a) How is the **document/query similarity** calculated?
 b) What type of formula is used for the document/query similarity measure?
 c) Is the formula normalized for document length?
8. Does the system employ **relevance feedback**?

9. Is **query expansion** employed?

10. Is there a **document ranking** mechanism?

a) Is there some form of **scaling**?

b) How many documents does it present to the user?

c) Can the user navigate (move up and down) the list of items?

11. Is there intra-document ranking (or by paragraphs)?

Some of these features have been implemented in the commercial systems surveyed and the preliminary findings in regard to innovative retrieval techniques found in these systems are discussed below and summarized in Table 1.

IV. COMMERCIAL SYSTEM FEATURES

The IR research indicates that major commercial database vendors are beginning to incorporate term weighting and ranked output in their systems, starting with DowQuest and followed by WESTLAW's WIN (WESTLAW Is Natural) system. Recently Dialog introduced its Target feature, and Mead Data the FreeStyle capability on LEXIS/NEXIS. (It is worth noting that the WAIS software for searching internet databases is also based on relevance ranking and has the same genealogy, through Thinking Machines, as DowQuest). There are also several stand-alone

software packages using these features, including Personal Librarian and Fulcrum. This paper focuses on DowQuest, Dialog's Target, and Westlaw's WIN system.

A) DowQuest from Dow Jones News/Retrieval

DowQuest was a leader in the field for innovative system change. It was introduced in 1989 by Dow Jones for use on their News/Retrieval system of full text news and business databases (Weyer, 1989). The DowQuest system is supplementary to standard Boolean-based retrieval for the Dow Jones databases, and this search mode is selected by typing "DowQuest" at the prompt. The information is primarily full text business information. The query is entered as list of keywords, as in this sample search for information on the recent IRA bombings at Heathrow Airport: "bombs ira heathrow". The system responds with a "Starter List" of 4 of 16 references, and offers options for relevance feedback, browsing or printing. Entering the number of 1 to 3 articles which are relevant begins a relevance feedback loop, which brings the items chosen to the top of the list and re-ranks the documents. Paragraphs may also be used for relevance feedback. In this instance, in order to find more specific information on the effect of the bombings on British-Irish relations, and John Major's response, item 4 is selected as relevant (See Figure 1). The top 100 words in the document are then used as a query and the data base

is re-ranked. The items which rise to the top of the list have an emphasis on this aspect of the topic.

Figure 1: DowQuest Relevance Feedback Search

DOWQUEST
Copyright (C) 1994 Dow Jones & Company, Inc.

* * * * *

DowQuest searches articles from more than 500 sources of current business information, including The Wall Street Journal, Barron's, Business Week, Fortune and Forbes. All articles are copyrighted by their publisher.

* * * *

TO VIEW INSTRUCTIONS or copyright notices, press (Return).
For additional information, type HELP at any point while using DowQuest.

TO BEGIN A SEARCH: Enter several words that are relevant to the subject you are interested in and press (Return).

ENTER QUERY:
bombs ira heathrow

DOWQUEST STARTER LIST HEADLINE PAGE 1 OF 4

- 1 International: IRA Is Linked To Mortar Fire At . . .
WALL STREET JOURNAL , 03/10/94 (555 words)
- 2 U.K. Police Search For More Mortars At Heathrow Airport
DOW JONES NEWS SERVICE , 03/14/94 (440 words)
- 3 What's News -- World-Wide
WALL STREET JOURNAL , 03/14/94 (834 words)
- 4 Britain Scorns New Talks With IRA --- Peace Feeler . . .
WASHINGTON POST: A SECTION, 03/15/94 (1,010 words)

* * * * *

TO IMPROVE A SEARCH: Type SEARCH and the numbers of up to three articles that are examples of your subject.

TO VIEW AN ARTICLE: Enter article number and press (Return).

TO PRINT ARTICLES: Enter BEST and article number to see "best" section.
Type PRINT and the article numbers.

PRESS (RETURN) FOR MORE HEADLINES. TYPE HELP FOR MORE INFORMATION.
search 4

DOWQUEST SECOND SEARCH HEADLINE PAGE 1 OF 4

- 1 Britain Scorns New Talks With IRA --- Peace Feeler . . .
WASHINGTON POST: A SECTION, 03/15/94 (1,010 words)
- 2 Major Rebukes Call For U.K. To Withdraw From N. Ireland
DOW JONES NEWS SERVICE , 01/03/94 (313 words)
- 3 Troubles Aplenty: In Northern Ireland, The . . .
WALL STREET JOURNAL , 12/16/93 (2,401 words)
- 4 Major Rejects Early Talks With Sinn Fein
WASHINGTON POST: A SECTION, 12/23/93 (402 words)

B) Target from Dialog

Dialog has recently implemented a search capability called 'Target'. This is a document ranking feature that uses the frequency of search terms in the record and other criteria to provide ranked output of 50 titles to the user. The ranking is indicated by a calculated percentage of 'relevance' to the query. While features of the ranking mechanism are proprietary, Dialog indicates that the following elements contribute to document-query similarity:

- quorum: the number of query terms in the record
- proximity: closeness of query terms to each other in the record
- frequency: weight based on $tf \times idf$

There is no relevance feedback, and no normalization for document length. (Dialog recommends that Target should not be used in a mix of databases which includes bibliographic and full text since the results will favour full text because of the document length). Searches are limited to current (1993-1994) records to improve processing speed.

Target is a separate subsystem within Dialog, invoked by typing 'target' at the prompt. Search terms are entered on the command line separated by a space. Although this is a 'natural language' search interface, Dialog allows the incorporation of Boolean and proximity features. An asterisk next to a term indicates that it must be present---effectively a Boolean AND. Synonyms may be grouped together in parentheses, giving the effect of a Boolean OR. Phrases can

be delineated with quotes in the search string which mimics the effect of the proximity operator. There is no stemming, and the system does not recognize non-query terms as stopwords (e.g. "I am interested in...").

In this sample search on the Billings lawsuit against Novell, the terms have been entered as a simple string. The system retrieves the top 50 records (See Figure 2). The display can be customized to include the 'statistical relevance' (as a percentage) and the term frequencies for each item. There is no relevance feedback, though modification of the search repeats the ranking procedure. The statistical relevance appears to be scaled, so that if the best document contains only one or two terms in the query it may still have a ranking close to 100%. A high percent relevance does not guarantee a relevant item.

Figure 2: DIALOG's TARGET Feature

?target

Input search terms separated by spaces (e.g., DOG CAT FOOD). You can enhance your TARGET search with the following options:

- PHRASES are enclosed in single quotes
(e.g., 'DOG FOOD')
- SYNONYMS are enclosed in parentheses
(e.g., (DOG CANINE))
- SPELLING variations are indicated with a ?
(e.g., DOG? to search DOG, DOGS)
- Terms that MUST be present are flagged with an asterisk
(e.g., DOG *FOOD)

Q = QUIT H = HELP

?novell patent lawsuit billings

Your TARGET search request will retrieve up to 50 of the statistically most relevant records.

Searching 1993-1994 records only
(ALL records for files: 30 139 565 583)
Processed 10 of 18 files ...

...Processing Complete

Your search retrieved 50 records.

Press ENTER to browse results C = Customize display Q = QUIT H = HELP

?c

BROWSE output includes: TI,JN,PD
Term frequency/relevance: off

Continuous display for COMPLETE TEXT: off
Custom display options:

- 1 Change BROWSE output to Title Only (usually free)
- 2 Customize BROWSE output with your own choice of display codes
- 3 Reset BROWSE output to the default (i.e., title, journal, date)
- 4 Change COMPLETE TEXT output to continuous display
- 5 Show term frequencies and statistical relevance (%) for each item

Press ENTER for NO CHANGE, or enter option number(s)
(e.g., 1,5) to customize the display. Q = QUIT H = HELP

?5

DIALOG-TARGET RESULTS (arranged by percent RELEVANCE)

----- Item: 1 -----
Processed 10 of 18 files ...
TITLE: Roger *Billings* case about *patent* claim may rattle industry.
(Focus on Computers)
JOURNAL: The Kansas City Business Journal
PUB DATE: Jan 29, 1993
- Statistical Relevance: 96%
- Term Frequency: NOVELL - 16 ; PATENT - 27 ; LAWSUIT - 2 ;
BILLINGS - 35

----- Item: 2 -----
TITLE: *Billings* must be stopped, and you can help. (From the Ether)
(Column)
JOURNAL: InfoWorld
PUB DATE: April 26, 1993
- Statistical Relevance: 95%
- Term Frequency: NOVELL - 13 ; PATENT - 17 ; LAWSUIT - 4 ;
BILLINGS - 27

----- Item: 3 -----
TITLE: B of A caves in to *Billings* NetWare extortion. (Bank of America
pays Roger *Billings* for his bogus network file service *patent*)
(From the Ether) (Column)
JOURNAL: InfoWorld
PUB DATE: Oct 25, 1993
- Statistical Relevance: 95%
- Term Frequency: NOVELL - 8 ; PATENT - 14 ; LAWSUIT - 4 ;
BILLINGS - 33

Press ENTER to continue browsing or enter item number(s) to see full record
M = Modify search T = New TARGET C = Customize display Q = QUIT H = HELP
?
.
.

C) WIN on WESTLAW from West Publishing

WESTLAW is an online system for legal materials from West Publishing in Minnesota (Tenopir, 1993). The WIN system of natural language searching was added in October 1992. It uses term weights for words and recognizes about 10,000 phrases from a dictionary of legal terms. Documents and sections selected from the initial ranked output can be used for relevance feedback. West claims

improvements of at least 15% in precision at standard recall points (Pritchard-Schoch, 1993). The system is able to deal with "conversational" natural language input (Turtle, 1994).

Table 1. A Preliminary Comparison of System Features

Features	DIALOG's Target (1993)	DOWQUEST (1989)	WESTLAW WIN (1992)
Natural Language	Yes	Yes	Yes
- Search Term Format	Keyword Entry	Keyword Entry	Keyword Entry
- Conversational Style	No	No	Yes
Phrases Identified	Manual	Limited	Yes
Truncation	Manual	Yes	No
Stemming Method	No	Yes	Yes
Term Weights	Yes	Yes	Yes
Doc/Query Similarity Formula Used	Based on quorum, proximity and frequency.	Yes	Yes
Formula Normalized	No	Yes	Yes
Relevance Feedback	No	Yes	Yes
Doc Ranking Feature	Yes	Yes	Yes
- Scaling	% of Relevance	No	No
- No. Docs Displayed	50	16	20

V. IMPLICATIONS

The incorporation of techniques for natural language input and ranked output raises some interesting questions: why have vendors suddenly began to implement these features, what is the potential impact on searchers, and what are the implications for evaluation?

A) Vendors

Preliminary findings indicate that the reasons for the current interest in introducing these features on the part of vendors are related to changes in IR research testing and more cost effective methods for practical implementation. The barriers to implementation cited in earlier studies have been largely overcome. The problem with research collection size, for instance, has been lifted with the current emphasis on large scale IR experimentation. Noteworthy is the series of TREC (Text REtrieval Conferences) sponsored by the National Institute of Standards and Technology, which involve the use of test collections of gigabyte size and a set of queries and relevance assessments, providing a standard mechanism for IR researchers around the world to compare their results (Harman, 1993). Research on large databases answers the criticism that experimental techniques will not scale up to sizes comparable to databases available commercially. There has also been an attempt to make the queries available on TREC more reflective of "real" queries, answering another criticism of earlier research.

TREC has also been effective in generating interest in the vendor community as well as in the research community, so that problems of lack of knowledge about research findings and inaccessibility of the research literature can be overcome. Many of the vendors are now involved in TREC projects of their own.

With regard to the financial barriers to change in IR systems, presumably cost-effective mechanisms have been found to implement natural language systems in concert with Boolean systems. Current implementations seem to be operating within the existing file structures. DowQuest, in implementing its system, cited "product differentiation" as one rationale for change, and presumably the current trend for other vendors to follow suit is an attempt to keep up, or to limit the competitive advantage of systems with natural language capability.

It is worth noting that vendors are taking different approaches to designing natural language systems: DowQuest based its system on the Connection Machine Document Retrieval System (CMDRS), developed by researchers at Thinking Machines (Stanfill & Kahle, 1986), and implemented it on their hardware using two Connection Machines. WESTLAW has hired as its technical specialist a former academic researcher, implemented his system (an inference network) and continues to carry out a series of tests to evaluate system parameters. We will explore the mechanisms by which IR research is incorporated by vendors in the course of this study.

B) Searchers

These new manifestations of natural language searching have a significant impact on searchers. In the past, searchers have had several vendor systems available to them, all of which offered basically the same capabilities. All had the same Boolean operators, proximity functionality which differed only slightly, and

similar features such as multi-database searching, and duplicate detection. Searchers could choose their system on the basis of characteristics which were essentially independent of system effectiveness. Moreover, searchers could easily interpret their results in the light of their queries and understand exactly what the system had done.

The number of parameters used in term stemming, term weighting, document-query similarity calculations, and relevance feedback is large, and details are usually proprietary. This means that the retrieval system has become a black box which is difficult for the searcher to interpret. It is not known how searchers will respond to this lack of control. Dialog cites the need to retain searcher control as a rationale for its structured input options.

Searchers now have an option of using Boolean or natural language searching. However most research is based on average precision-recall values and not on individual queries and classes of queries, so that not enough is known about when one approach would be superior to the other.

C) Evaluation

The features described effectively differentiate systems in terms of their performance. Since these natural language systems may give highly variable results, system performance becomes a major issue. How can these new system features be evaluated? Obviously TREC has a role to play here, but it does not really address an operational environment. A realistic evaluation of interactive

systems is exceptionally difficult. New evaluation techniques will have to be found which will allow these systems to be tested and compared (e.g. Su, 1991).

There are also problems inherent in comparing Boolean and natural language systems since they do not produce the same kind of output (e.g. ranked chronologically vs. by presumed relevance, differences in output set size). Some of these problems have been addressed by WESTLAW, where they are comparing WIN to Boolean query processing using test collections ranging in size from 250 MB to 6 GB (Turtle, 1994).

In the final analysis, natural language searching will probably complement rather than replace Boolean searching. While natural language gives on average better results, Boolean searching is superior for some queries, and preferred by some users (Turtle, 1994). However while Boolean systems are relatively static, the potential for improvements in natural language systems exists as researchers explore new techniques and refine existing ones in large scale operational environments.

With further work impending, an informal survey of vendors for details about the implementation of new features will be conducted. The following questions regarding system feature implementation will be asked:

- 1) How and why were these features introduced into the system?
- 2) Are these features based upon any particular piece of research?
- 3) Are there any ongoing research programs to test these new features?

4) Does the vendor have any internal or external documentation to explain these features?

The survey will illuminate the motivation and barriers behind commercial vendor acceptance or rejection of innovative techniques found in IR research. This survey will not only assist in understanding better which links between IR research and operational environments are strongest, but also it will indicate a research agenda for the IR community.

VI. REFERENCES

- Belkin, N. J. and Croft, W.B. (1987). Retrieval techniques. In M.E. Williams (Ed.), *Annual Review of Information Science and Technology*. 22 (pp. 109-145). Amsterdam: Elsevier Science Publishers.
- Croft, W. B. and Turtle, H. (1992). Text retrieval and inference. *Text-Based Intelligent Systems*, 127-155.
- Harman, D. (1993). Overview of the first TREC conference. In B. Korfhage, E. Rasmussen, and P. Willett (Eds.), *Proceedings for the 16th International Conference on Research and Development in Information Retrieval*. (pp. 36-47). New York, NY: ACM.
- Ledwith, R. (1992). On the difficulties of applying the results of information retrieval research to aid in the searching of large scientific databases. *Information Processing & Management*, 28(4), 451-455.
- Maron, M. E. (1988). Probabilistic design principles for conventional and full-text retrieval systems. *Information Processing & Management*, 24(3), 249-255.
- Pritchard-Schoch, T. (1993). Natural language comes of age. *Online*, 17(3), 33-43.
- Radecki, T. (1988). Trends in research on information retrieval - the potential for improvements in conventional Boolean retrieval systems. *Information Processing & Management*, 24(3), 219-227.
- Salton, G. (1989). *Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer*. Reading, MA: Addison-Wesley.
- Smit, P. H. and Kochen, M. (1988). Information impediments to innovation of on-line database vendors. *Information Processing & Management*, 24(3), 229-241.
- Stanfill, C. and Kahle, B. (1986). Parallel free-text search on the Connection Machine system. *Communications of the ACM*, 29(12), 1229-1239.
- Su, L. T. (1991). Evaluation of interactive information retrieval: implication for operational systems and practice. In M.E. Williams (Ed.), *Proceedings of the 12th National Online Meeting*. Medford, NJ: Learned Information.

- Tenopir, C. (1993). Natural language searching with WIN. *Library Journal*, 118(18), 54-56.
- Turtle, H. (In Press). Natural language vs. Boolean query evaluation: a comparison of retrieval performance. In *Proceedings of the 17th International Conference on Research and Development in Information Retrieval*.
- Weyer, S. A. (1989). Questing for the "DAO": DowQuest and intelligent text retrieval. *Online*, 13(5), 39-48.