Interactive TREC-4 at Georgia Tech

Aravindan Veerasamy
veerasam@cc.gatech.edu
College of Computing
801, Atlantic Drive
Georgia Institute of Technology
Atlanta, Georgia 30332-0280
Phone: 404-894-8791
Fax: 404-894-9442

Abstract

At Georgia Tech, we investigated the effectiveness of a visualization scheme for Information Retrieval systems. Displayed like a bar-graph, the visualization tool shows the distribution of query words in the set of documents retrieved in response to a query. We found that end-users use the visualization for two purposes:

- to gain specific information about individual documents – such as the distribution of different query words in that document.
- to gain aggregate information about the query result in general – such as getting a sense of the direction of the query results.

In general they used the visualization tool as much as the title and full text in the process of deciding if a document addresses the given search topic. In structured post-session interviews with searchers, we also obtained information about what the searcher liked, what was frustrating to them, and what they wanted in the system.

1 Introduction

At the TREC-4 interactive experiments at Georgia Tech, we were interested in investigating the effectiveness of a visualization scheme for IR systems that we have developed. The visualization scheme, as given in Figure 2, is intended to provide more information to the user about the query results in addition to just the title and full text. In ranked output systems, the naive end-user has little knowledge about why the system retrieved and ranked the documents in a given way in response to a free-form text query. This problem does not arise in boolean systems since there is no element of surprise in why a particular document was retrieved. The above-mentioned lack of knowledge in ranked output systems can be quite disturbing when a user is not able to get the set of documents he/she needs and does not know enough about the system to modify the query to get the documents he/she needs. It is with this in mind that we have developed a visualization scheme that shows the distribution of query words in the retrieved documents. This visual display of distribution information provides a good overview of the retrieved set of documents with respect to the free-form user query.

For TREC-4 we were interested in investigating how end-users used the visualization scheme. We were also interested in finding what aspects of the system were frustrating, what aspects they liked and what they wanted in the system. We have yet to do a thorough statistical analysis of the trace data to quantitatively determine the ways in which users with visualization tool acted different from the users without the visualization tool. What we report here is our observations of user interactions, information from structured interviews, and questionnaires.

In the next section we give a brief description of our system. Then we describe our experimental design followed by our observations as it relates to the visualization tool. Then we discuss user's frustrations, likes and wants.

2 System Description

For our study, we used the INQUERY retrieval engine from University of Massachusetts, Amherst [CCH92]. We built a simple graphical user interface on top of INQUERY using Tcl/Tk [Ous94]. There are two versions of our system – one with the visualization, and one without. In our base system, as shown in Figure 1, there are three windows: the top left window is for entering and editing the query. The titles of retrieved documents are displayed immediately below that window. Thirty titles can be displayed in one screen. One can scroll down to a maximum of 150 document titles. Mouse-clicking a title brings up the full text of that document in the window at the bottom right. By clicking the "Next Query Word" button in the full text window, one can position the full text display such that the next occurrence of query word in the document is at the top of the window.

One can save documents and mark documents for relevance feedback by clicking the "Save?" and "Rel?" buttons immediately to the left of the title in the title display window. The only operator that is allowed is the adjacency operator: A hyphen between two words specifies that the two words must appear right next to each other in the same order in a document in order for the word-combination to contribute to the retrieval of that document. There is no negation operator. Automatic stemming and stopping are performed.

The visualization tool is displayed in another window as shown in Figure 2. It consists of a series of vertical column of bars. There is one column of bars for each document. The leftmost vertical column of bars corresponds to the document ranked 1 and the rightmost vertical column corresponds to the document ranked 150 with all the intermediate ranks lying in between. In each vertical column there are multiple bars – one each for each query word. The height of the bar at the intersection of query word row and a document column corresponds to the weight of that query word in that document. Thus if there are a handful of query words that convey the crux of the query and is very important for a document to contain these query words, one can quickly see from the visualization which retrieved documents have those important words. One can also see how many of the retrieved documents have those words in combination to get a feel for the overall goodness of query results. The effects of modifying the query, like adding a query word, would clearly be shown in the visualization. One can quickly take stock of how useful the query modification turned out. Moving the mouse cursor over the vertical columns would highlight the column directly beneath the mouse cursor and simultaneously highlight the title corresponding to that document in the title display window.

Apart from the query words typed in by the user, the visualization also shows the distribution information for words added by the system due to relevance feedback. In summary, all the words internally used by the system in computing the query results are shown in the visualization. The words in the visualization are also stopped and stemmed.

3 Experimental Setup

The searchers for our study were undergraduate student volunteers from a course on library searching at Georgia Tech. All the searchers had prior computer experience – a majority of them more than 4 years. All the students were majoring in an engineering discipline. They had differing levels of experience with the Georgia Tech Electronic Library catalog – a boolean online public access catalog.

All the users were asked to fill out a background questionnaire. They were given a tutorial on how to use the system. They were then asked to do a practice search on topic 224 for 15 minutes. Following that they were asked to find as many documents as they can that address the given information problem without too much rubbish (as specified by the interactive track guidelines). This was followed by another intermediate tutorial and then a search for a second topic. Immediately after each of the two real searches, they filled out a search evaluation questionnaire. Finally, there was a structured interview.

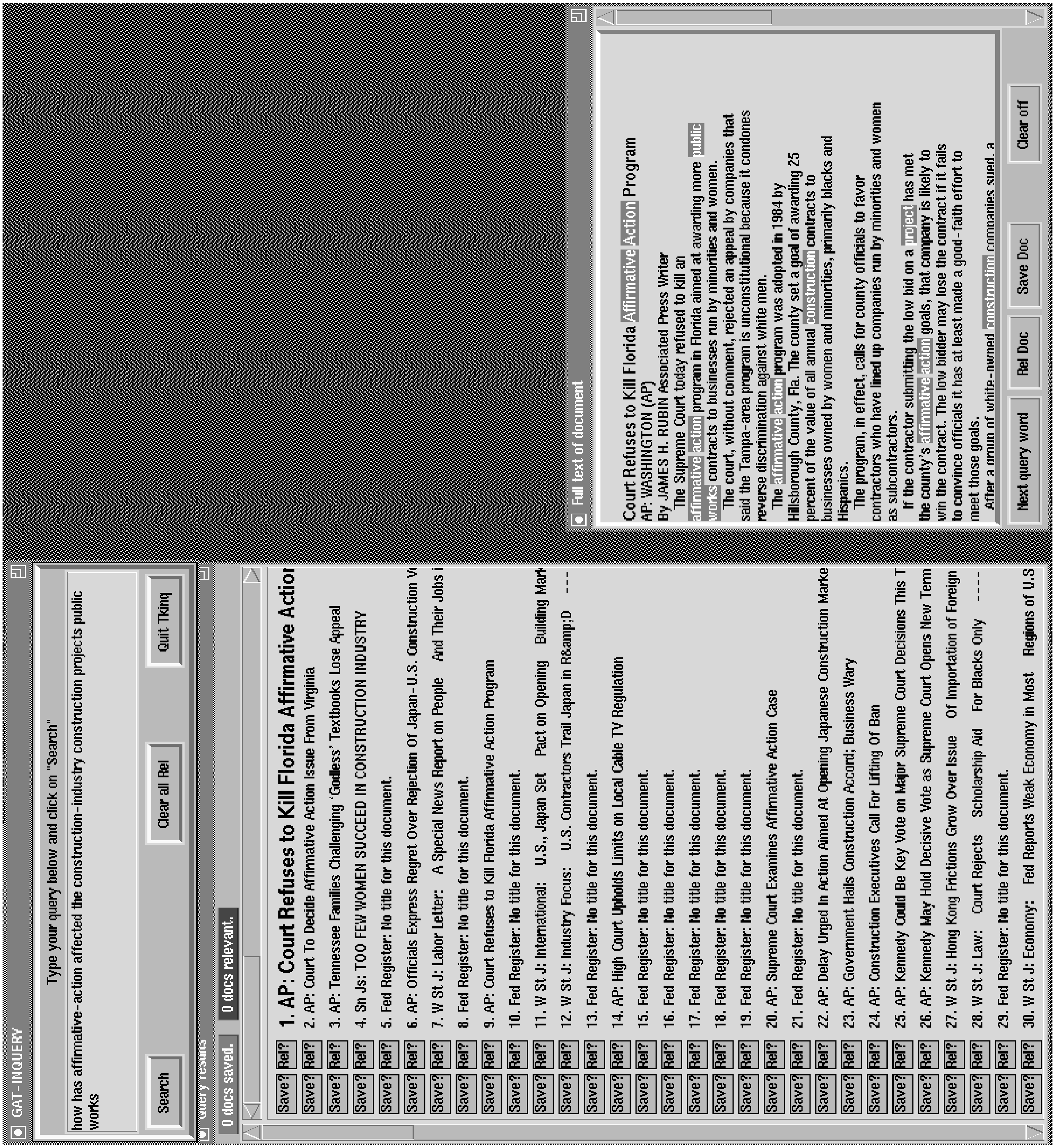

Figure 1: Sample querying session. The window in the top-left corner is the query entry window. Immediately below that is another window where the titles of retrieved documents are displayed. To the bottom right is another window where the full text of documents are displayed.

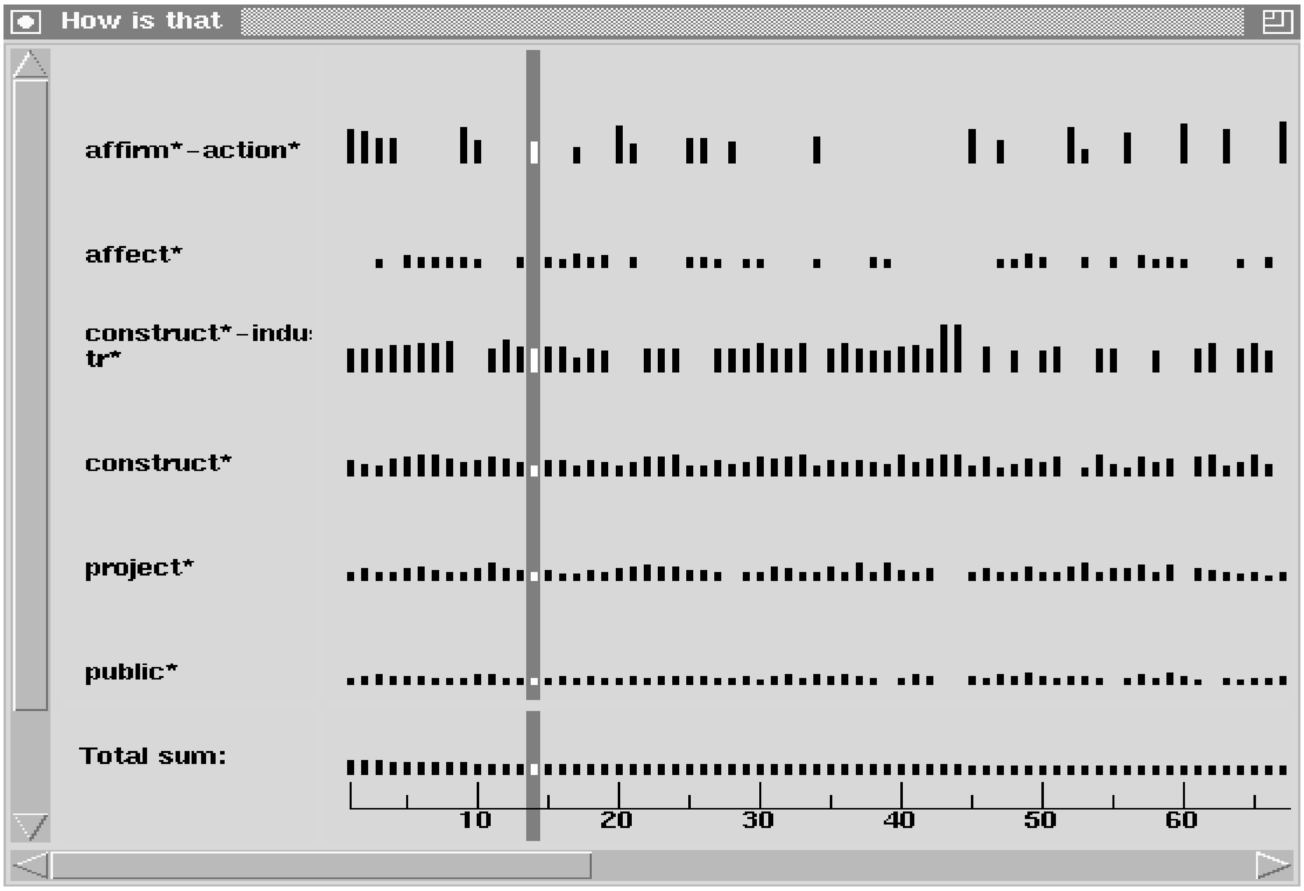

Figure 2: Visualization of results. The highlighted vertical column corresponds to document ranked 14. The title of document ranked 14 document will also be highlighted in the title display window. Clicking the highlighted vertical column brings up the full text of that document.

The searchers were divided into three groups. In each group there were 12 searchers. In the first group (hereafter named "w:w", since both first and search topics are searched WITH visualization), the searchers used the visualization tool for all the searches and the tutorial. In the second group (hereafter named "wo:w", since the first search topic is searched WITHOUT and second topic WITH visualization), the initial tutorial, the practice search and the first search was done without the visualization tool. The intermediate tutorial introduced the visualization tool and the search for the second topic was done with the visualization tool. In the third group (hereafter named "wo:wo", since both the search topics are searched WITHOUT the visualization), all the tutorials and searches were done without the visualization tool. The intermediate tutorial for the w:w and wo:wo groups was a dummy tutorial to compensate for the intermediate tutorial of the wo:w group.

Since each searcher searched for two topics and there were 12 searchers in each group, all the 24 topics were covered by each of the three groups. The 24 topics were randomly divided into 12 pairs and each pair was searched by 3 searchers, one each from the w:w, wo:w and wo:wo groups. The idea was to compare the performance among the three groups to find out the effects of the visualization scheme. Only 24 of the 25 topics for the interactive track were given to end-users in the study. The remaining one topic (topic 223) was searched by the author using the visualization tool.

The searchers were asked to think aloud as they used the system. For the most part, there was an observer in the same room using a different computer and simultaneously observing the searcher. Based on such observations while the user session was in progress, we felt that huge searcher differences in interpreting the query combined with huge differences in the nature of the search topics will greatly confound the effects of the visualization tool. As a result, we decided to run a second study. In the second study, we picked topic 242 for the practice search and the practice search was extended to 30 minutes. The intermediate tutorial was removed. We picked topics 203 and 236 for all the searchers. There were two groups of searchers for the second study – the first group had the visualization tool and second group did not have the visualization tool. There were 18 searchers in each group. By keeping the two search topics constant for all these searchers, we expected to eliminate the effects of search topic difference. It turns out that the searcher variability in interpreting the search topic is so huge among searchers that it is not fair to compare different searchers using different systems unless the search topic is extremely clear and specific.

4 End-users view of the visualization tool

A vast majority of the users mentioned visualization as one of the aspects of the system that they liked. They mentioned using the visualization tools in the following ways.

- Some searchers mentioned using it to see the importance of query words in the retrieved documents – as given by the height of the bar. They mentioned that they were more likely to look at the full text of a document if it has a higher concentration of the important query words.
- Most of the searchers mentioned using it most frequently to see the co-occurrence of important query words in the retrieved documents. They mentioned it being easier to use the visualization tool to look for the co-occurrence information than going through the full text of documents in search of occurrences of the important query words.
- Many searchers felt that the visualization in conjunction with the document title gives a fairly good idea of what the document is about. If the title looks promising and the visualization shows that the document has the right combination of query words, one is tempted to look at the full text of the document.
- They mentioned using it to get a quick overview of the number of retrieved documents a query words appears in. They mentioned using it as a checkpoint to see if a query has turned out the way they had expected it to. If not, they were tempted to readjust the query to get a better result. This happens often when some of the crucial query words are not well represented in the retrieved documents. In that case, one is tempted to add synonyms or words related to those crucial query concepts.
- Some of the searchers mentioned that the visual nature of the distribution information was much easier to identify things than reading text information. This suggests that the mental effort of reading textual information as being much higher than interpreting a simpler visual pattern, and given a choice, the users are more likely to choose the latter.
- **Disadvantages:** A few searchers mentioned that relying heavily on the visualization can also hurt as follows: They mentioned that using the bar-graph to pick out a document containing certain query words may not be indicative of the content of the document – just as the title may not be a good indicator of content. An exemplar case is searcher 35 on the topic of "status of nuclear proliferation treaties". Since almost all of the retrieved documents had something

to do with "nuclear proliferation", the searcher mentioned using the visualization tool to pick those documents containing the query word "status" – only to see that the usage of "status" in the document was not in the context of nuclear proliferation treaties. Then the searcher started paying little emphasis on the presence of "status" in documents. Although relying on that information was initially detrimental, one tends to learn when and how to rely on the visualization. We believe that the presence in retrieved documents of adjectives, adverbs and verbs from the query may not be good content indicators especially when they have a high collection frequency. And relying on the visualization to select documents that have these adjectives, adverbs and verbs from the query may not help.

In summary, the visualization tool seems to help in the following ways:

- to gain more information about specific documents in addition to the title before looking at the full text. Higher concentration of important query words in a document suggests a closer look at the document.
- to gain aggregate information about the query result. The absence of important query words in a vast majority of the retrieved documents suggests query reformulation by adding synonyms and other related concepts.

5 Likes, Frustrations and Wants of users

Apart from the visualization, we were also interested in finding if there are any specific facilities that the users wanted, what features they liked, and what aspects were frustrating. While interpreting the following, we wish to reiterate that all the searchers had some amount of experience with the Georgia Tech Electronic Library catalog which is a character-based-command-driven interface to a boolean system. Some of the features they liked may arise out of the fact that they have had little experience with ranked output systems and the only other major information retrieval system they know is a character based interface to a boolean system.

5.1 Likes

- A vast majority of the searchers with the visualization mentioned that the visualization tool and relevance feedback as the two major aspects of the system they liked. Searchers without the visualization mentioned relevance feedback as the most important feature they liked.

- A number of searchers found the fact that all the information (like the user query, titles of documents and the document full text) is displayed simultaneously in one screen to be very useful. In the Georgia Tech library system, one has to switch between screens to get different types of information. There seems to be a significant mental overload in the context switch between screens. Having simultaneous access to all information seems to bring about a rich interplay between the different sources of information.

- many searchers mentioned that the mouse-based graphical nature of the interface is a significant improvement over a command line based interface.

- many searchers also mentioned that the free-form textual queries without having to worry about any syntax leads to a free flow of thought. "I like the fact that I can type in whatever comes to my mind ... knowing that it will ignore all the junk words like a, an, the, etc...".

- The "Next Query Word" feature was also liked by many searchers. They liked it because they did not have to scroll through a long document looking for occurrences of query words. (All the occurrences of query words in a document are highlighted by the system).

5.2 Frustrations

- A number of searchers mentioned that it was frustrating when the system takes a long time to get the full text of a large document. Similarly, they were also frustrated when it takes a long time to evaluate a query with a large number of relevance feedback documents. The longest delay for evaluating a query was about 2 minutes (when there are about 30 relevant documents). Most of the query evaluations took less than 20 seconds. They said that they understand that the system has to process a lot information (when there a number of relevant documents), but it was frustrating nevertheless.

- Some searchers said that it was frustrating to spend some time reading through the full text of a document and when they are halfway, realizing that they had already seen the same/similar document.

- While some searchers seemed to like having access to 150 retrieved documents, some others mentioned that 150 documents is too much especially when most of the 150 are not relevant. They seem to have the opinion that if some documents are definitely not relevant to the query, then they should not be shown. Thus, this problem is not alleviated even if one reduces the number of documents displayed. They seem to be quite sensitive about precision. They are not as sensitive about recall – since they are usually satisfied if they get a few documents concerning the topic. Based on our observations, we believe that when the non-relevant documents consistently come from a particular subject area, and when the user is not in a position to remove those documents, they tend to get more frustrated. Using subject classification schemes (where available) to negate disinteresting subject areas would help in this regard.

- In our system, when the title for a document is not available, the message "No title for this document" is displayed instead of the title in the title display window. Many of the federal register documents do not have a title and this is quite annoying to some searchers since they do not have any idea about the document content. This makes it difficult to decide whether to request the full text or not. In cases where the full text is requested, the document happens to be large and hence takes a lot of time to retrieve, thereby adding to the frustration.

- Some federal register documents do not have anything worthwhile – they consist of a listing of subject areas or table of contents. Some searchers wondered why these documents were in the database in the first place.

- Some searchers mentioned a general dislike towards federal register documents partly because they felt that many of them did not have any important piece of information, partly because in general they have no title, partly because it took too long to retrieve them.

- Some searchers were frustrated when a document that they know as non-relevant keeps coming up in the query result. The fact that they were not able to delete the document from the display seemed to add to the frustration.

5.3 Wants

Many of the frustrations mentioned above seemed to directly translate into wants for removing the causes of frustration. In addition to those wants, we observed the following:

- Many searchers expressed a desire to remove certain query words that were added by the system from relevance feedback documents – especially when they are proper names and when they are not necessarily what they are looking for.

- A number of searchers wanted a keyboard equivalent of mouse actions. This is not to say that they did not want mouse actions. It seems to be a significant effort for these searchers to move the right hand out of the keyboard, reach over to the mouse, look at the screen to position the mouse cursor, click the mouse button and move back to the keyboard.
- When asked if they felt a need to have access to an online thesaurus, some searchers expressed a desire for it and some did not. Some of those who did not want a thesaurus mentioned that relevance feedback seemed to alleviate the need for a thesaurus.
- Many searchers wanted to be able to specify that the system should definitely avoid retrieving certain documents in subsequent query iterations. They wanted to have a negative relevance feedback where the system avoids all documents like a particular nonrelevant document.

6 Acknowledgments

The tremendous help from Prof. Nick Belkin regarding the experimental setup and questionnaire design is highly appreciated. Many thanks to Prof. Scott Hudson and Prof. Shamkant Navathe who were instrumental at every stage of the interface development. Special thanks to Prof. Bruce Croft for letting us use the INQUERY retrieval system. Support from the ARPA contract No. F33615-93-1-1338 is appreciated.

References

[CCH92] J.P. Callan, W.B. Croft, and S.M. Harding. The inquery retrieval system. In *Third International Conference on Database and Expert Systems Applications*, September 1992.

[Ous94] John K. Ousterhout. *Tcl and the Tk Toolkit*. Addison-Wesley, 1994.