Interactive Exploration as a Formal Text Retrieval Method: How Well can Interactivity Compensate for Unsophisticated Retrieval Algorithms

Nipon Charoenkitkarn, Mark Chignell, and Gene Golovchinsky
Department of Industrial Engineering
University of Toronto

Introduction

Our goal in participating in TREC-3 was to see if it is possible to achieve adequate performance using relatively simple retrieval techniques, coupled with the type of interactivity at the user interface that is typical of a browsing system.

As the results reported in this paper show, our system was able to perform large scale text retrieval on routing queries, albeit not at a level to challenge well established text retrieval systems. However, since ST-PatTREC can handle both browsing and querying styles of interaction in a single environment, and since ST-PatTREC did reasonably well in comparison with other systems in category B of the competition, the present findings demonstrate that it should be possible to develop general-purpose information exploration systems that allow for both browsing and querying styles of interaction. The results that we obtained provide a baseline for what can be done when a browsing style of query formulation is combined with simple text retrieval methods. In addition to discussing the methods used and results in TREC-3, future plans are outlined for using visualizing and other styles of browsing in information exploration.

Information Exploration (Browsing AND Querying)

Current methods of accessing information tend to fall into two separate categories, depending on whether a browsing or querying style of interaction is used. In browsing, the goal is to navigate through a collection of documents or information nodes, looking for interesting or relevant information.

In contrast to the point and click selection that typically occurs in browsing, querying requires the user or system to specify a search goal that can be matched in some way against the database, in order to identify a set of relevant documents. One of the characteristics of querying situations is that it is much easier to define criteria of success. For instance, if one can measure the relevance of documents in some way, then standard measures such as precision and recall can be assessed.

The approach that we used in TREC-3 grew out of our earlier research on information exploration. Waterworth and Chignell (1991) identified three dimensions of information exploration. The first dimension of information exploration (structural responsibility) concerns who is responsible for searching and who must consequently be concerned with structure.

The second dimension of information seeking behaviour (target orientation) arose from the distinction between browsing and querying. According to Waterworth and Chignell, “Browsing is distinguished from querying by the absence of a definite target in the mind of the user.” Thus the distinction between browsing and querying is not determined by the actions of the user, or by the configuration of the system, but by the cognitive state of the user. One hypothesis stated in this early conceptualization of the information exploration task is that there is a continuum of user behaviours varying between querying and browsing that is characterised by the level of specificity of the user's information seeking goals. In our recent research we have been studying the degree to which users can readily mix browsing and searching styles of information exploration when given appropriate tools. The ST-PatTREC system used in TREC-3 allowed users to intermix browsing and querying styles of interaction.

A third dimension of information seeking behaviour can be identified based on the method of interaction used in the interface to the information system. Descriptive interfaces have generally been associated with querying behaviour, and referential (point and click) interfaces have generally been associated with hypertext browsing, but there is no intrinsic correlation between interaction method, target orientation, and structural responsibility. Thus the three-dimensional description of information seeking activity can be represented as in Figure 1.

The Interactive Querying Approach

One of the points made by Waterworth and Chignell (1991) is that the space of possible information exploration systems has not been fully explored. Most existing systems tend to fall into two general categories, text retrieval (i.e., systems that focus on command-based querying) and hypertext (systems that emphasize referential, or point and click, interaction within a navigational browsing framework).

Information retrieval and hypertext systems as currently envisioned are only two extreme cases within a family of possible information exploration systems. Our research is looking at the properties of intermediate members of that family that combine the functionalities of browsing and querying. One way of making querying more like browsing is to increase the interactivity of querying. If sufficiently rapid and incremental feedback is provided, there is very little cost to trying out a large number of different queries. The result is a more exploratory style of interaction that starts to approximate the interactivity of hypertext. The analogy with hypertext can then be reinforced by allowing queries to be marked up directly on text. Hits then appear as menus of hypertext links from the marked up concept specified by the graphical query to corresponding concepts in other documents.

User centeredness of hypertext systems can also be enhanced. Hypertext can be envisioned as a special kind of retrieval process where the goal is to provide the user with a link to the article or node that is most relevant given the context and status of the current transaction. A fundamental problem in manually authored (static) hypertext is that the links that are likely to be relevant in a particular context must be predicted in advance. The user's range of action is often limited to choosing among the available links at each step in the browsing process.

One way to increase user centeredness of hypertext is to give users more control over selection, as well as creation of links. The result would then be a form of dynamic hypertext in which the user expresses a concept of interest and the system identifies the nodes that address or relate to that concept. In this approach, link authoring is done at run-time, on the basis of selected concepts. Clearly, this type of user centeredness places considerable responsibility on the user to formulate and express concepts that are used as the basis for creating links. However, with appropriate support tools and user interfaces, such reader-based (dynamic) linking has considerable potential for enriching and customizing the information exploration task.

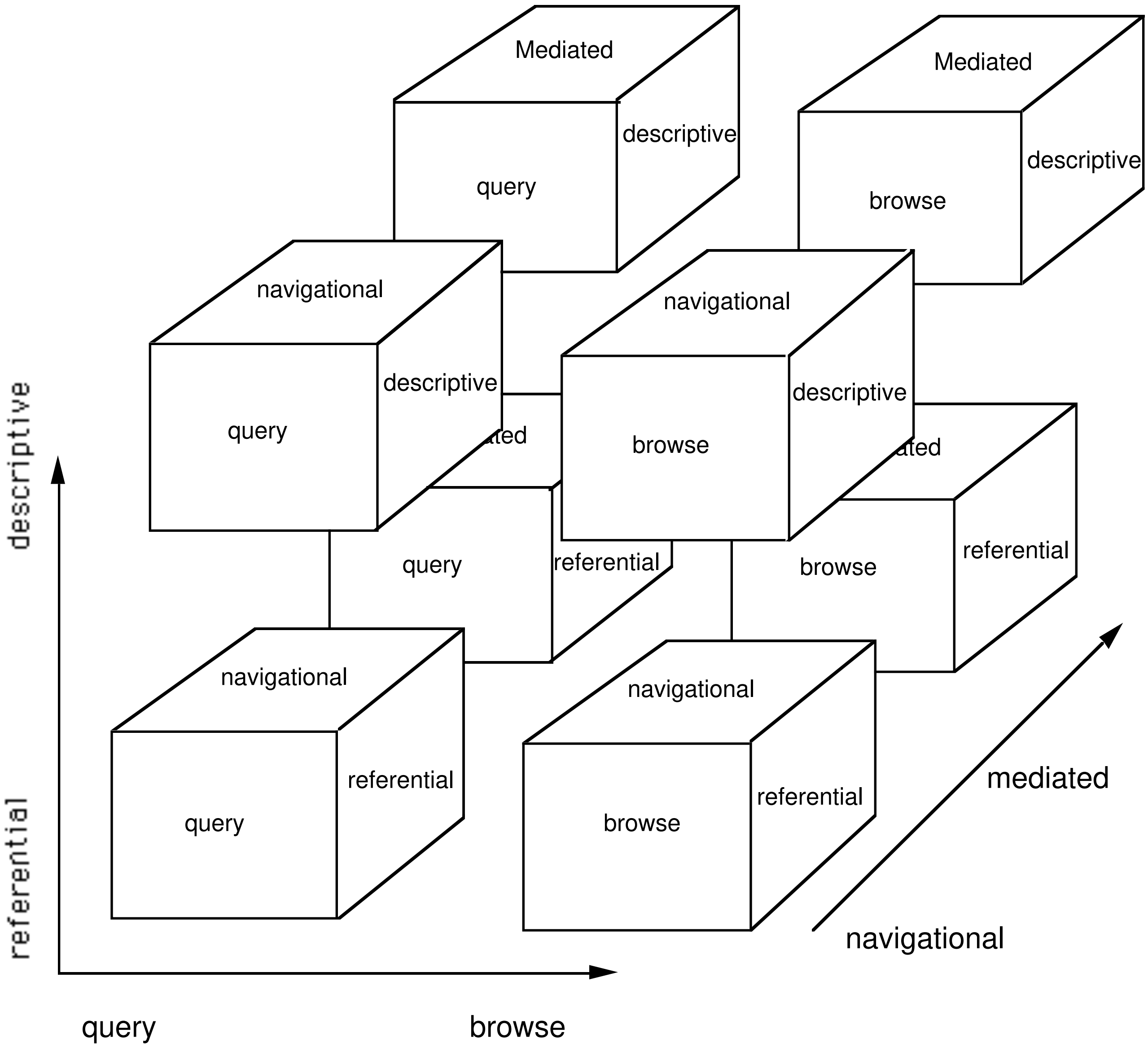

Figure 1. A Model of Information Exploration

Different types of feedback may be relevant at different stages of search or browsing. In the early stages of exploring a topic, users may not fully understand what they are interested in, or what terminology is used in the database. Thus the concept of a search lifecycle may be appropriate, with statistical information about co-occurrence of terms being more useful input from the user early in the search, and relevance feedback (based on accumulated evidence about topics of interest to the user) being more useful for topic reformulation later on in the search.

Some forms of dynamic feedback may approximate forms of authored information that have proved to be useful in books and other information retrieval structures and systems. For instance, feedback concerning term co-occurrence information provides some of the functionality of a hierarchical index. The first term acts to specify an entry in the index. The coordinate terms for that first term serve as sub-topics listed under the entry. One can then explore further levels of the index hierarchy by looking for the coordinate terms of the first and second terms joined together. For instance one might have "transport" as the first term. A coordinate term might then be "roads." This would be equivalent to a hierarchical index where *road transport* was a sub-entry of *transport*. A coordinate term of "roads" and "transport" jointly might then be "trucking," corresponding to a sub-sub-entry under *transport*. Not many books include more than two levels of indexing because of the effort required to produce highly nested indices. The use of coordinate terms or more sophisticated techniques for assessing relationship and cohesion can simulate some of the functionality of a hierarchical index without authoring (manual indexing) effort. This process capitalizes on hierarchical indexing implicit in the text.

Support Tools for the User: ST-PatTREC

ST-PatTREC[1] was developed to implement a text browsing system capable of handling moderately large (hundreds of megabytes) databases and providing enhanced feedback to the user. ST-PatTREC uses a graphical notation developed by the second author (Golovchinsky, 1993) for specifying queries against databases capable of supporting them. It maintains a history of user requests and provides a workspace for storing components of queries.

The number of nodes returned can be controlled in ST-PatTREC by reducing the scope of the proximity operator (represented by the AND operator) until the desired number of nodes is displayed. For example, assume a user set a target of 10-15 nodes. A query with proximity of 500 characters may retrieve 100 nodes, but when the proximity criterion is reduced to 150 characters, a more manageable set of about 10 nodes might be retrieved. Furthermore, this set would likely be more relevant, since the terms would tend to be more closely related in the text. Giving users dynamic control over the volume of data should encourage exploration by reducing some of the uncertainty about the effects of commands.

IR Algorithms

A detailed description of the ST-PatTREC user interface and functionality is provided in the appendix to this paper. In this section we will briefly review the IR algorithms that ST-PatTREC uses, focusing on query execution and ranking.

ST-PatTREC uses exact matching of non-negated Boolean queries expressed in disjunctive normal form. Queries in ST-PatTREC can be entered by directly marking up text (drawing lines between selected words to indicate the AND operator). Non-negated disjunctive normal form Boolean command strings, e.g., (cancer AND drug) OR (cancer AND therapy) can be created by building AND clusters of terms, where each AND cluster contains a collection of words linked by lines (as explained by Golovchinsky and Chignell, 1993). The AND operator was actually executed as a proximity operator in ST-PatTREC. In the default condition, the AND condition was true if the words covered by the AND appeared within 500 characters of each other. In addition though, the user was provided with a proximity slider so that the scope of the proximity could be varied directly by the user.

---

[1]ST represents the Smalltalk interface, Pat is the full-text search engine from OpenText Foundation, and TREC is the Text REtrieval Conference sponsored by NIST for which this software was developed.

A very simple form of document ranking was used for each final query. The first challenge was to obtain a set of 1000 documents, given that many of our queries would extract considerably less than a 1000 documents (sometimes less than 100 documents). Once the set of 1000 documents (or as close as we could get to 1000 documents) was obtained, the documents were ranked according to the number of times that the query terms appeared in their text, normalized by the length of each document. This ranking was based on exact matching of terms (i.e., no truncation was used).

The selection of the set of 1000 documents was done manually using the following algorithm that could have been implemented in software had more time been available. This algorithm was generally adhered to, but other methods were also tried if necessary to produce the right number of documents.

> **Step 1**. Use the exact final query with the exact proximity that was specified. This provided the initial set of documents.
>
> **Step 2**. Use the exact final query with the proximity expanded so that it equalled the length of the document. That is, the AND became an actual Boolean AND rather than a proximity operator.
>
> **Step 3.** Use truncation AND/OR loosening of the query. For instance, loosening might be done by removing a term from an AND cluster.

The procedure for expanding the hit set to the necessary 1000 documents can be illustrated with an example. The original routing query that we identified for one topic was:

> (Japan AND market AND access) OR (Unfair AND Talk) with proximity = 200

The first expansion of this query involved removing the proximity restriction so that the scope of the AND operator was within entire documents. The query was then loosened by removing the word "market" from the first AND cluster so that it became (Japan AND Access). At this stage we still did not have the required number of documents. The final expansion in this case involved conflating two terms across the AND clusters to obtain (Japan AND unfair), which yielded the desired number of documents.

TREC-3 Evaluation Strategy for ST-PatTREC

Queries were constructed based on the work of two searchers. The searchers carried out query formulation using the ST-PatTREC system. They worked independently and then reviewed their results to select a routing query based on the precision and recall measures obtained on the training data. In our study, the Wall Street Journal data was the only data used in training.

Given that we used relatively simple text retrieval algorithms[2] , how could we expect to obtain reasonable results in comparison with much more sophisticated retrieval algorithms? Our hope was that our searchers would be able to come up with better search terms based on extensive browsing of the training data. To the extent that the TREC routing topics already included rich statements of the search concept, along with key terms, the task of compensating for rudimentary IR algorithms through better selection of terms based on extensive browsing was made more difficult.

---

[2] The version of ST-PatTREC used in TREC-3 did not even have a truncation operator, which has since been added to the system.

We tried to increase the likelihood of including meaningful terms in the query by cross-validating queries and terms across two separate halves of the training database. Each half of the database was approximately 250 MB in size. We carried out browsing and query formulation on one half of the database and then ran the resulting query on the other half of the database to see how much "shrinkage" of the precision and recall scores occurred. We then carried out query formulation on the other half of the database. The routing query was chosen that seemed to maximize precision and recall across both halves of the database. An example of the variability that can occur in the split half comparisons is shown in Table 1. Numbers in parentheses show the corresponding values of precision and recall that occurred in the second half of the database.

In general, Table 1 shows a fair amount of consistency between the two halves of the database in terms of recall and precision associated with the exact final query. The large differences that occurred (e.g., the increase in recall from .67 to .87 for the first query) are generally due to changing the proximity, or seem to be associated with a tradeoff between recall and precision (e.g., the query in the second row of the table has a higher recall in the second half of the database, but a lower precision). Based on generally strong correlations on recall and precision between the two halves of the training database, we expected that better recall and precision on the training database would be predictive of better recall and precision on the test database.

Query	Recall	Precision	Proximity
(drug AND cancer)	.67 (.87)	.16 (.13)	128 (275)
(drug AND cancer OR cancer AND research)	.69 (.75)	.15 (.12)	85 (85)
(drug AND cancer OR cancer AND research OR cancer AND development)	.72 (.78)	.14 (.10)	85 (110)
(drug AND cancer OR cancer AND therapy)	.69 (.78)	.14 (.14)	128 (110)

Table 1. Performance of Query for Topic 122 across Two Halves of the Training Database.

Results

ST-PatTREC did reasonably well in comparison with the other two systems in category B of the competition. Table 2 summarizes the results that we obtained across the 50 topics. In addition to the topic number (top), the table shows the elapsed time for each query (time), the number of iterations (iter), the precision of the query on the training data (tpre), the recall of the query on the training data (trec), the corresponding recall (erec) and precision (epre) scores for the test data, the average precision for the test data (avgpre), the number of relevant documents (numr) that existed in the test database for that topic (as assessed by the TREC organizers), the number of terms in the final query (numt), and the number of OR operators in the final query (numor). The ranking (rank) of our system on a three point scale (best, median, worst) based on feedback provided to us about how our system compared with the other systems in our section of the competition, is also shown, as are the ratings that the searchers made concerning their familiarity with the general search topics (exp). All precision and recall scores shown in Table 2 except average precision scores (avgpre) are based on exact final queries with exact proximities. The ranks are based on those of relevance retrieved at top 100 articles. If our rank was tied with another system, we used the average rank (e.g. tied for first = 1.5).

top	time	iter	tpre	trec	epre	erec	numr	numt	numor	rank	avgpre	exp
101	33	14	.13	.67	.15	.22	9	8	2	1.50	.2973	8
102	22	10	.22	1.00	.17	.20	5	7	2	1.00	.2246	8
103	17	8	.00	.00	.02	.05	20	7	2	3.00	.1025	5
104	15	6	.00	NA	.00	NA	0	6	2	2.00	.0000	5
105	36	10	.14	1.00	.00	NA	0	8	2	2.00	.0000	5
106	60	10	.07	.45	.24	.42	12	5	1	2.00	.2063	5
107	40	6	.42	.25	.43	.12	24	4	1	2.00	.1629	5
108	12	18	.26	.23	.28	.07	67	5	1	3.00	.1377	5
109	39	7	.49	.92	.97	.82	132	5	4	1.00	.8952	8
110	29	10	.38	.71	.46	.52	102	4	1	3.00	.4595	7
111	36	15	.39	.49	.43	.30	231	11	5	2.00	.4488	7
112	18	4	.27	.36	.33	.22	72	6	2	2.00	.3161	5
113	51	7	.10	.71	.09	.57	21	8	3	1.50	.1386	8
114	42	12	.07	.29	.10	.03	40	6	1	3.00	.1563	7
115	23	7	.09	.73	.18	.64	25	7	4	2.00	.1548	5
116	33	11	.05	.44	.00	NA	0	5	2	2.00	.0000	5
117	20	9	.13	.39	.10	.20	15	4	1	2.00	.1762	6
118	57	10	.12	.34	.23	.16	183	6	2	3.00	.1652	7
119	29	10	.16	.41	.23	.19	139	6	2	2.50	.2159	7
120	20	12	.00	.00	.25	.01	89	4	1	1.50	.1016	5
121	45	6	.00	.00	.08	.03	118	7	2	2.00	.1157	5
122	20	10	.15	.80	.17	.80	30	4	1	1.00	.2204	5
123	18	7	.16	.49	.40	.54	125	6	2	2.00	.3976	5
124	22	6	.14	.78	.05	.67	21	8	3	2.00	.1037	5
125	17	5	.21	.64	.46	.61	101	7	3	1.00	.5791	5
126	59	13	.37	.45	.30	.27	67	6	2	2.00	.2621	5
127	20	7	.04	.56	.08	.62	24	6	2	2.00	.1857	7
128	39	12	.23	.60	.25	.22	27	8	3	1.00	.2913	5
129	14	7	.06	.86	.25	.01	36	6	2	2.00	.1891	7
130	28	8	.70	.36	.38	.17	59	3	1	2.00	.2295	7
131	54	13	.12	.42	.00	.00	1	3	0	2.00	.0038	4
132	21	6	.28	.72	.18	.71	28	8	3	2.00	.3737	4
133	21	1	.47	1.00	.20	.91	11	1	0	1.50	.5123	8
134	20	1	.89	.94	.50	.83	6	1	0	1.50	.7553	5
135	27	7	.63	.84	.53	.66	76	8	3	2.00	.4793	5
136	60	3	.59	.80	.71	.45	11	8	3	2.00	.5451	6
137	22	12	.39	.52	.22	.23	43	5	2	3.00	.1972	6
138	49	10	.06	.75	.08	1.00	16	4	1	1.50	.1559	7
139	16	6	.NA	.00	.67	.20	10	5	1	2.00	.2198	7
140	18	5	.33	.50	.00	.00	7	2	0	1.00	.0618	7
141	60	18	.00	.00	NA	.00	1	3	0	1.00	.0106	6
142	37	7	.42	.44	.86	.29	123	7	2	1.00	.4965	7
143	40	9	.12	.47	.14	.16	19	6	2	2.00	.2143	6
144	12	8	.00	.00	.00	NA	0	4	1	2.00	.0000	7
145	37	9	.46	.51	.41	.52	21	4	2	1.00	.7523	6
146	16	5	.00	NA	.11	.67	3	4	1	1.50	.3269	7
147	55	10	.21	.43	.40	.21	98	11	5	3.00	.2087	6
148	20	7	.56	.64	.54	.94	78	4	3	2.00	.7951	4
149	55	7	.46	.28	.33	.05	58	4	1	1.00	.1511	5
150	48	10	.38	.64	.35	.55	155	4	2	1.00	.4903	6

Table 2. Evaluative and Descriptive Data for the ST-PatTREC Searches

Table 3 shows the results obtained by our system (marked with asterisks) compared with other systems in our section of category B, as provided to us by the TREC organizers. While it is difficult to make meaningful comparisons across different databases and topics, our average precision aggregated across all the topics was .27, within the range obtained by the Okapi system in TREC-2 (Robertson et al. 1993). The Okapi average precisions varied between a high of .36 and a low of .142 (aggregating across all the different topics with different weightings and algorithms). Our average precision results also appeared to be competitive with results obtained in TREC-2 using combination of evidence (Belkin et al. 1993).

We also looked at how well recall and precision on the training set predicted recall and precision on the test data. We found evidence of predictive relationships on both recall and precision scores (calculated from documents retrieved by exact final queries with exact proximities). That is, if a query achieves a high recall score or high precision score on the training data it seems to do well on the test data as well. Table 4 shows the matrix of intercorrelations for the data shown in Table 2. The correlation between the precision measures on the training data versus the corresponding measures on the test data, and the correlation between recall measures for training and test data were both .70

We then looked to see which other factors would predict when our system would perform well. We looked at the relationship between the ranking of our system (rounded up to 1, 2, or 3) across the 50 topics, and a variety of predictive measures, including the training precision and recall, the elapsed time for the query formulation process, the number of iterations used, the number of terms in the final routing query, etc. Separate analyses of variance were carried out for each of these predictors as the dependent variable, with the independent variable being the ranking of the topic. None of these ANOVAs were significant at the .05 level. However, the number of terms in the final query ($F[2, 47]=2.42$) and the number of iterations ($F[2, 47]=2.70$) were both borderline significant ($p <.10$). Figures 2 and 3 show the corresponding error bar charts. It can be seen in Figure 3 that the possible significant difference for iterations is due to the fact that more iterations tend to be associated with relatively poor performance of the routing query. This seems reasonable to us, since a large number of iterations indicated that the searchers were having trouble in identifying a suitable query. Similarly a relatively low (less than 5) terms in the query tended to be associated with better performance. This suggests that our system worked best when a fairly simple query could be expressed with a few diagnostic terms.

Relatively few of the correlations shown in Table 4 are significant. Not surprisingly, perhaps, there was a high ($r=.83$) correlation between the number of terms in the final routing query and the number of OR operators used. However, the correlation between number of iterations and time taken in performing the query was low (about .3). The number of iterations was inversely related to recall on both the training and test data, confirming the impression that a large number of iterations is generally a sign of problems in query formulation.

There was also a significant ($r=.45$) correlation between the number of relevant documents in the test database and the number of ORs in the final routing query. This seems reasonable, since a broad topic might be expected to cover a number of disjunctive concepts.

We also looked at what sort of topics our method did better or worse in. The two searchers each independently rated their familiarity with the general topics used in the searches (e.g., science and technology, finance, international politics, etc.) on a 5-point scale. We added the two sets of ratings to obtain a 10-point rating that represented the joint familiarity of both searchers with the general topics. We then correlated the expertise ratings of

Topic	Rel.	Relevant Retr. @100 Best	Median	Worst	Relevant Retr @1000 Best	Median	Worst	Average Precision Best	Median	Worst
101	9	7*	7	4	9*	9	4	0.3120	0.2973*	0.0980
102	5	5*	4	4	5*	5	4	0.5520	0.5187	0.2246*
103	20	8	7	3*	18*	18	11	0.1194	0.1172	0.1025*
104	0	0*	0	0	0*	0	0	0.0000*	0.0000	0.0000
105	0	0*	0	0	0*	0	0	0.0000*	0.0000	0.0000
106	12	9	8*	1	11*	10	2	0.2453	0.2063*	0.0066
107	24	15	14*	10	21	17*	17	0.3117	0.1629*	0.1615
108	67	29	27	18*	61	59*	43	0.3138	0.2091	0.1377*
109	132	94*	89	13	129	127*	41	0.8952*	0.8804	0.0668
110	102	53	52	50*	102*	102	69	0.5480	0.4595*	0.4175
111	231	78	59*	40	222*	221	123	0.4488*	0.4140	0.4130
112	72	37	36*	6	66*	65	21	0.3639	0.3161*	0.0368
113	21	14*	14	8	21*	21	9	0.2068	0.1444	0.1386*
114	40	24	24	14*	40*	38	27	0.4086	0.3713	0.1563*
115	25	18	16*	8	24*	23	9	0.2662	0.1548*	0.0998
116	0	0*	0	0	0*	0	0	0.0000*	0.0000	0.0000
117	15	14	10*	8	15	14*	8	0.5962	0.2636	0.1762*
118	183	67	59	32*	154	115*	106	0.5081	0.4247	0.1652*
119	139	36	26*	26	124*	108	44	0.2827	0.2159*	0.1366
120	89	13*	13	9	84*	72	13	0.1016*	0.0926	0.0276
121	118	52	14*	2	114	107*	39	0.4721	0.1157*	0.0190
122	30	20*	13	12	26*	18	12	0.2690	0.2316	0.2204*
123	125	58	49*	24	121*	105	49	0.4521	0.3976*	0.1002
124	21	12	11*	4	20*	14	8	0.1532	0.1037*	0.0387
125	101	61*	46	42	101	100*	44	0.5791*	0.4037	0.3042
126	67	43	32*	25	55	49*	31	0.4598	0.3337	0.2621*
127	24	22	15*	12	24	24*	12	0.3712	0.2350	0.1857*
128	27	22*	10	5	25*	18	10	0.2913*	0.1724	0.0247
129	36	20	18*	13	33	31*	16	0.2750	0.1891*	0.1387
130	59	30	23*	19	59	52*	23	0.3379	0.2360	0.2295*
131	1	0*	0	0	1*	1	1	0.0072	0.0057	0.0038*
132	28	23	22*	15	28	27*	15	0.5139	0.4022	0.3737*
133	11	10*	10	5	10*	10	5	0.7622	0.5123*	0.3690
134	6	6*	6	1	6*	6	1	0.7553*	0.4451	0.0833
135	76	53	50*	30	76	63*	37	0.6848	0.4793*	0.2947
136	11	10	9*	4	11	10*	6	0.5451*	0.4997	0.0651
137	43	30	20	18*	41	38*	22	0.6102	0.2496	0.1972*
138	16	12*	12	11	16*	16	12	0.2786	0.2146	0.1559*
139	10	8	5*	4	8	5*	4	0.2198*	0.1640	0.1224
140	7	5*	4	4	7*	5	4	0.3813	0.3805	0.0618*
141	1	1*	0	0	1*	1	0	0.0106*	0.0064	0.0000
142	123	69*	42	31	99*	89	62	0.4965*	0.3113	0.1784
143	19	12	10*	5	18*	17	8	0.2573	0.2143*	0.0471
144	0	0*	0	0	0*	0	0	0.0000*	0.0000	0.0000
145	21	20*	12	2	21*	15	6	0.7523*	0.2259	0.0116
146	3	3*	3	1	3*	3	1	0.4019	0.3333	0.3269*
147	98	38	37	32*	77	66	59*	0.2889	0.2661	0.2087*
148	78	71	70*	44	77*	77	45	0.7951*	0.7870	0.5345
149	58	17*	16	10	40*	39	12	0.1511*	0.1141	0.0563
150	155	56*	53	19	140*	111	56	0.4903*	0.3862	0.1160

Table 3. TREC-3's Routing Results for Category B.

Correlation Coefficients (Coefficient/(Cases) / 2-tailed Significance)
"." is printed if a coefficient cannot be computed

	TIME	ITER	TPRE	TREC	EPRE	EREC
TIME	1.0000	.3023	.0355	-.0760	.1266	-.1770
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .	P= .033	P= .809	P= .607	P= .386	P= .239
ITER	.3023	1.0000	-.3251	-.3225	-.2044	-.4384
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .033	P= .	P= .023	P= .025	P= .159	P= .002
TPRE	.0355	-.3251	1.0000	.3858	.6963	.3029
	(49)	(49)	(49)	(47)	(48)	(45)
	P= .809	P= .023	P= .	P= .007	P= .000	P= .043
TREC	-.0760	-.3225	.3858	1.0000	.1333	.6995
	(48)	(48)	(47)	(48)	(47)	(45)
	P= .607	P= .025	P= .007	P= .	P= .372	P= .000
EPRE	.1266	-.2044	.6963	.1333	1.0000	.2250
	(49)	(49)	(48)	(47)	(49)	(45)
	P= .386	P= .159	P= .000	P= .372	P= .	P= .137
EREC	-.1770	-.4384	.3029	.6995	.2250	1.0000
	(46)	(46)	(45)	(45)	(45)	(46)
	P= .239	P= .002	P= .043	P= .000	P= .137	P= .
NUMR	.1564	.1612	.1933	-.1021	.4520	-.0353
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .278	P= .263	P= .183	P= .490	P= .001	P= .816
NUMT	.1567	.1641	-.2065	.1008	.0540	-.0994
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .277	P= .255	P= .155	P= .495	P= .713	P= .511
NUMOR	.0934	.0035	.0498	.2609	.2957	.2087
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .519	P= .981	P= .734	P= .073	P= .039	P= .164
RANK	-.0524	.1856	-.1504	-.2690	-.1805	-.2539
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .718	P= .197	P= .302	P= .065	P= .215	P= .089
AVGPRE	-.0512	-.3509	.6984	.4551	.7620	.6165
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .724	P= .012	P= .000	P= .001	P= .000	P= .000
EXP	.0152	-.0135	-.0153	.1607	.1059	.0057
	(50)	(50)	(49)	(48)	(49)	(46)
	P= .917	P= .926	P= .917	P= .275	P= .469	P= .970

Table 4. Matrix of Intercorrelations of Data in Table 2.

	NUMR	NUMT	NUMOR	RANK	AVGPRE	EXP
TIME	.1564 (50) P= .278	.1567 (50) P= .277	.0934 (50) P= .519	−.0524 (50) P= .718	−.0512 (50) P= .724	.0152 (50) P= .917
ITER	.1612 (50) P= .263	.1641 (50) P= .255	.0035 (50) P= .981	.1856 (50) P= .197	−.3509 (50) P= .012	−.0135 (50) P= .926
TPRE	.1933 (49) P= .183	−.2065 (49) P= .155	.0498 (49) P= .734	−.1504 (49) P= .302	.6984 (49) P= .000	−.0153 (49) P= .917
TREC	−.1021 (48) P= .490	.1008 (48) P= .495	.2609 (48) P= .073	−.2690 (48) P= .065	.4551 (48) P= .001	.1607 (48) P= .275
EPRE	.4520 (49) P= .001	.0540 (49) P= .713	.2957 (49) P= .039	−.1805 (49) P= .215	.7620 (49) P= .000	.1059 (49) P= .469
EREC	−.0353 (46) P= .816	−.0994 (46) P= .511	.2087 (46) P= .164	−.2539 (46) P= .089	.6165 (46) P= .000	.0057 (46) P= .970
NUMR	1.0000 (50) P= .	.3071 (50) P= .030	.4471 (50) P= .001	.1467 (50) P= .309	.3504 (50) P= .013	.0905 (50) P= .532
NUMT	.3071 (50) P= .030	1.0000 (50) P= .	.8307 (50) P= .000	.2449 (50) P= .087	−.0603 (50) P= .677	−.0233 (50) P= .873
NUMOR	.4471 (50) P= .001	.8307 (50) P= .000	1.0000 (50) P= .	.1163 (50) P= .421	.2885 (50) P= .042	−.0336 (50) P= .817
RANK	.1467 (50) P= .309	.2449 (50) P= .087	.1163 (50) P= .421	1.0000 (50) P= .	−.2791 (50) P= .050	−.1068 (50) P= .461
AVGPRE	.3504 (50) P= .013	−.0603 (50) P= .677	.2885 (50) P= .042	−.2791 (50) P= .050	1.0000 (50) P= .	.0893 (50) P= .537
EXP	.0905 (50) P= .532	−.0233 (50) P= .873	−.0336 (50) P= .817	−.1068 (50) P= .461	.0893 (50) P= .537	1.0000 (50) P= .

Table 4. (Cont.)

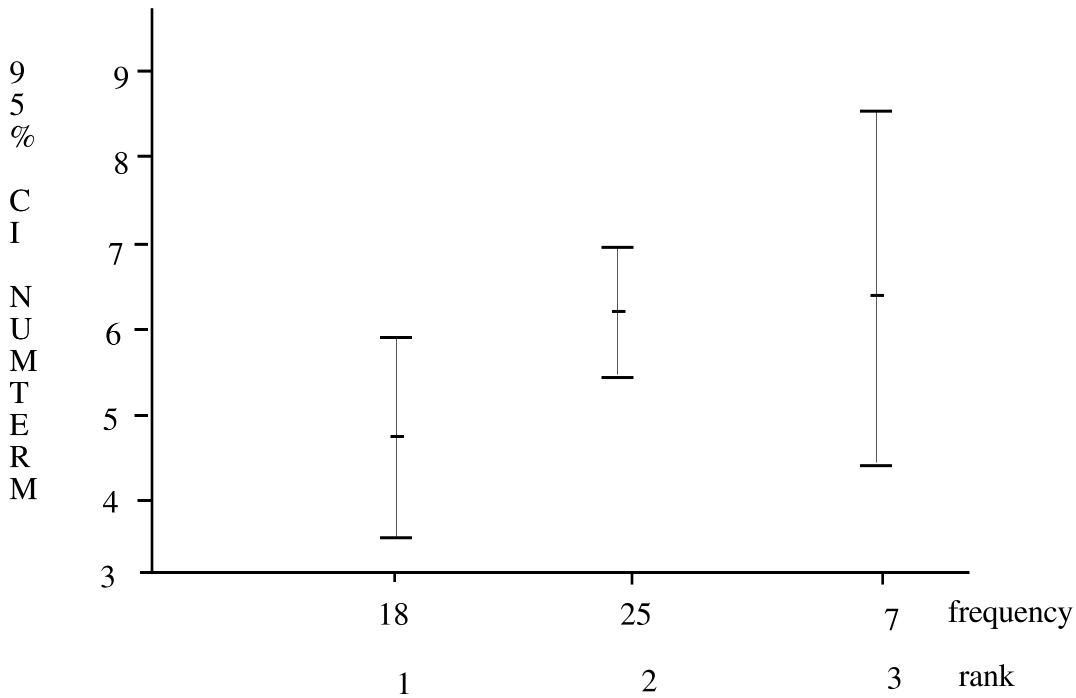

Figure 2. Effect of Number of Terms in Final Query on Performance Rankings.

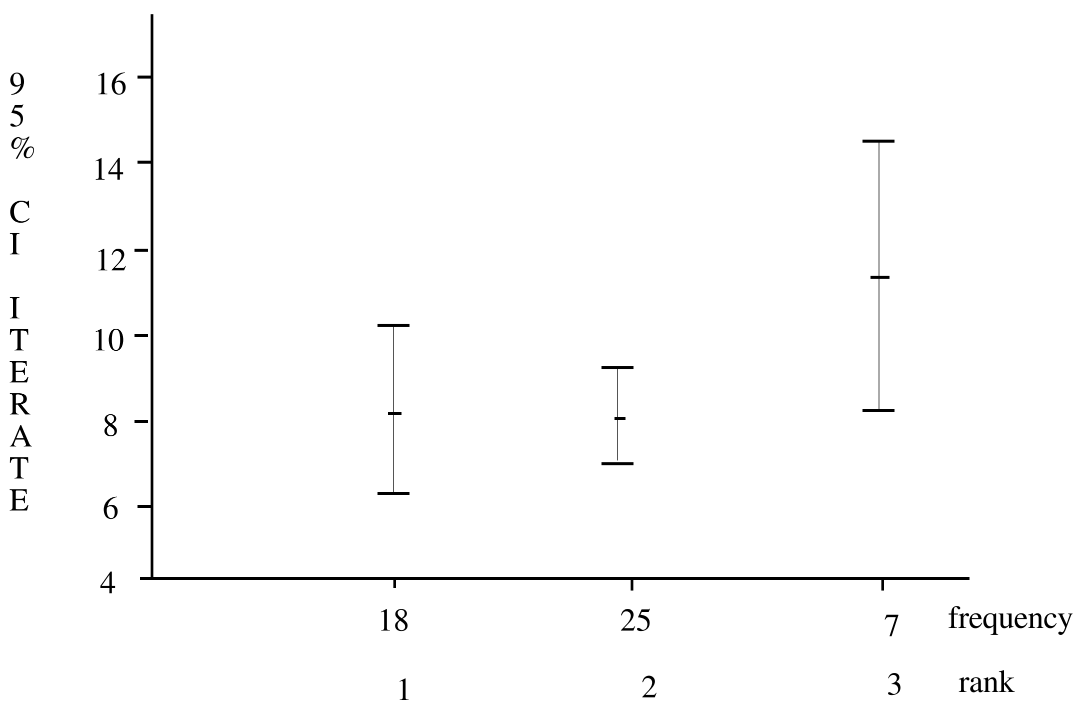

Figure 3. Effect of Number of Iterations on Performance Rankings.

familiarity with each topic with the corresponding measures of precision and recall on the training and test data, respectively. None of these correlations were significantly greater than zero ($r<.10$ in every case). Thus we have no evidence that domain expertise was a factor in the searches.

The final analysis that we carried out looked at what predicted performance of the routing queries in terms of average precision. A stepwise multiple regression was carried out using query formulation time, number of iterations, precision of the query for the training data, recall of the query for the training data, the number of relevant documents for the topic actually in the test database, and the number of terms used in the final query as predictors. Training precision was the strongest predictor of average precision of the routing query ($r=.54$). The other two predictors that entered the final regression equation were the number of relevant documents and training recall (multiple R for the three predictor equation was .79). None of the other predictors significant improved the three term prediction equation. The beta weights in the three term equation were as follows:

training precision	.54
number of relevant documents	.29
training recall	.29

The precision and recall of the exact final query on our training data were strongly related to precision on the test data. The other predictor of performance on the test data appeared to be number of iterations. A number of different attempts to form a search query is generally indicative of problems in query formulation that tend to show up in the eventual performance of the routing query. In the following section we will look at some of the lessons learned concerning the usability of the system.

Lessons Learned

The two searchers spent many hours working on the 50 topics to form the final routing queries. Much of this time was wasted due to deficiencies in the ST-PatTREC system. One problem was that the graphical queries for complex disjunctions were sometimes parsed incorrectly, requiring the users to try out different input strategies in order to get the query they wanted. This problem has since been fixed by allowing users to type in Boolean queries directly. We also found that the demands of the TREC evaluation on our system were greater than expected. As a research prototype, the system seemed to work quite well, but in the face of many hours of focused searches, the system would often crash, adding to the time and effort required. Thus there has been some effort to improve the reliability of the system and make it more suitable for large scale searches in the future.

One of the features that we missed the most was truncation during the query formulation. We used truncation during the query expansion phase, but in the interactive querying that preceded the generation of the final routing query truncation was not available. Truncation was one of the features that did not seem to be a "natural" part of the browsing interface, but turned out to be important for querying styles of search. Although our original philosophy had been to construct queries by directly marking up the text in a point and click style of interaction, we soon found that direct input of queries as text was more convenient in constructing complex queries. It was also easier to combine older queries into new queries using text entry in a workspace than to try and modify the markup once a certain number of terms were involved in the query.

The online dictionary didn't seem to add any value during the query generation process and was not used. Similarly, the online thesaurus (WordNet) was not used. We found that the material in the online dictionary and thesaurus was too general, so that they did not add value for specific queries (at least queries on topics relevant to the Wall Street Journal database). In contrast, the coordinate terms turned out to be very useful. Knowing the coordinate terms made it much easier to identify terms that could be ANDed to the current term.

The searchers found that interactive feedback of recall and precision was both useful and motivating. We found that the ability to select terms directly from document text was useful, because the searchers could follow the terminology actually used in the database. For the TREC search tasks, the best working style seemed to be selection of the terms in the text and graphical markup for simple queries followed by textual commands to combine the simple queries into complex query. In later versions of the system we have since added this capability and found it to be very useful. The TREC experience made it clear to us that trying to build complex queries was just too difficult with the graphical querying methods that we had originally developed. The searchers also found that the documents judged to be relevant by experts were very useful and they consulted extensively in looking for appropriate search terms.

Future Extensions

Our experience in using ST-PatTREC in the TREC-3 experiment showed that the ability to form complex queries is fairly important in achieving good performance under the TREC-3 rules. The detailed search topics provided the relatively large number of documents that were relevant for queries (typically over 100), and the requirement to create a routing query tended to de-emphasize the role of browsing in the interaction in favor of query construction. We are currently looking at methods of reinstating the role of browsing, either within the TREC-3 rules, or else in terms of a modified approach that changes the task and evaluation structure so as to encourage more browsing.

One of the problems with large databases and topics that have large numbers of relevant items is that interactive browsing of the actual documents is too costly, in both time and effort. However, browsing of a summarized version of the document space might still be feasible. Two forms of summarization are the use of brief abstracts, and the use of visualizations of the concept space in which documents reside. Automatic abstracting is difficult, and articles such as those available in the TREC database tend not to have written abstracts. For some stories or articles, lead paragraphs might serve as abstracts or overall summaries, while for others, the simple heuristic of picking the first one or two paragraphs may not be adequate. Recently there have been proposals to use the cohesiveness of text (e.g., Halliday and Hasan, 1976) to develop paths or chains through the text. However, although promising, this research is still in the early stages.

In the near term concept visualization may be the most useful approach for text summarization. In one technique (Damashek, 1994), inter-document similarity is assessed based on frequencies of n-grams in documents. The advantage of using N-grams is that they are less affected by incorrect spellings (Cavnar, 1993). Using one of a number of scaling techniques (e.g., singular value decomposition, or multidimensional scaling) the documents can then be located in a low dimensional space and clustered.

We carried out an informal experiment at the TREC-3 conference to see how well the Acquaintance system (Damashek, 1994) would work as visualization tool for handling TREC-3 topics. We began by using ST-PatTREC to form a query. We used the final routing query that we had developed for topic 122. This query was (cancer AND drug) OR (cancer AND research). This query yielded about 15% precision and 74% recall. We then took the 268 documents that were retrieved by this query for half of the training (Wall Street Journal) database and passed them to the Acquaintance system where they were visualized using the N-gram similarity scaling and clustering method. The document clusters were then examined for relevance to the topic and non-relevant items were pruned from the visualization. This pruning was done fairly quickly, taking less than five minutes. The resulting set contained 19 documents, of which 9 or roughly 47% were relevant. The visualization method was successful in improving the precision, something that had been a problem for the ST-PatTREC system in the TREC-3 experiment, but at the expense of recall (since only 9 of the 54 relevant documents or approximately 17% were relevant). While this recall is low, it should be possible to explore the visualization further and find additional clusters of relevant documents. Thus one future direction is to use text browsing to develop a relatively large set of candidate documents with high recall and low precision, and then pass this to a visualization process to prune irrelevant document clusters and thereby improve precision. We could then characterize the steps in this overall process as:

1. Text browsing
2. Query formulation
3. Visual Browsing of Document Clusters

We will also be carrying out research that investigates the role of browsing and querying in information exploration more closely. We are currently developing a revised information exploration system called BrowsIR that incorporates some of the lessons learned from our participation in TREC-3 as well as from subsequent usability studies using ST-PatTREC. We will then carry out a study to see how people use a flexible information exploration system when faced with different types of tasks. In particular, we will look at how instructions that emphasize browsing or querying styles of interaction tend to affect retrieval performance using a flexible information exploration system. The experiment will be carried out using TREC databases and topics.

One further avenue that we will explore is the coordination of text retrieval on local and distributed databases. We are currently connecting our text browsing tool to the world wide web. We will then carry out TREC searches on both the TREC and world wide web data. Our goal will be to study the effectiveness of wide area network (WAN) text searching as a supplement to the information contained in the TREC databases.

Conclusions

Overall we were satisfied with the performance of our system in TREC-3. Given the changes that we have since made to the system based on our experience, we expect that its effectiveness should be improved now that truncation and textual entry methods to combine simple queries into complex queries have been added.

One important result is the consistency between the evaluation scores for the training data and the test data. Our system seemed to work best with short queries consisting of a few highly diagnostic terms. The extreme case of this was two topics where the query consisted of only one term ("Hubble" for one topic and "genome" for the other). It is not clear how much this effect was due to the fact that construction of complex queries was too difficult in the version of our system that we were using, and how much it is a general effect. One clear diagnostic result is that a large number of querying iterations was a sign of trouble in our system. The searchers wrote down the query and the recall and precision scores when they felt that they had an interesting result. We used the number of entries that they wrote in the search log as indicators of the occurrence of iterations in the query formulation. While the implications of this result are unclear, one interpretation is that there is a law of diminishing returns in interactive querying, and that searches which do not yield useful results relatively quickly are unlikely to improve with significant extra effort. Certainly our experience as searchers was that some searches were difficult, while others seemed easy. Putting a lot of extra time and effort into the difficult searches didn't seem to improve the results much in most cases.

Acknowledgments

This research was supported by the Information Technology Research Center of Excellence of Ontario, the National Science and Engineering Research Council of Canada, and by the Canadian International Development Agency. We would also like to thank Professor Frank Tompa of the University of Waterloo, and Open Text Corporation, for making the Pat text retrieval engine available to us for our research.

References

Belkin, N.J., Kantor, P., Cool, C., and Quatrain, R. (1993). Combining Evidence for Information Retrieval. In D. Harman (Ed.), *Proceedings of the Second Text REtrieval Conference (TREC-2)*, NIST Special Publication 500-215, National Institute of Standards and Technology, Gaithersburg, Maryland.

Cavnar, W.B. (1994). N-Gram-Based Text Filtering for TREC-2. In D. Harman (Ed.), *Proceedings of the Second Text REtrieval Conference (TREC-2)*. NIST Special Publication 500-215, National Institute of Standards and Technology, Gaithersburg, Maryland.

Damashek, M. (1994). Gauging similarity via N-Grams: Language-Independent Sorting, Categorization, and Retrieval of Text. Unpublished manuscript, U.S. Department of Defense, Ft. George Meade, Maryland.

Fawcett, H. (1989). *PAT 3.3 User's Guide*. , Waterloo, Ontario, Canada: UW Centre for the New Oxford English Dictionary.

Golovchinsky, G. (1993).*Queries-R-Links: Query-based browsing in a full-text retrieval system.",* MASc thesis, University of Toronto

Golovchinsky, G. and M.H. Chignell. (1993). Queries-R-Links: Graphical markup for text navigation. in *Proceedings of INTERCHI '93*. . Amsterdam, The Netherlands:

Halliday, M.A.K. and Hasan, R. (1976) Cohesion in English. London: Longman.

Shneiderman, B. (1987). *Designing the User Interface*. Reading, Massachusetts: Addison Wesley.

Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M.M., and Gatford, M. (1993). Okapi at TREC-2. *Proceedings of the Second Text REtrieval Conference (TREC-2)*, NIST Special Publication 500-215, National Institute of Standards and Technology, Gaithersburg, Maryland.

Waterworth, J.A. and Chignell, M.H. (1991). A Model of Information Exploration. *Hypermedia*, ***3****(1)*, 35-58.

Appendix. Interactive System Description

I. System Description

Figure 4 shows a "typical" screen from the system. The screen contains a number of regions or windows which allow the user to build queries and view the results of queries within a single environment. One of the features of this system is that it does not have different modes (e.g., forming queries vs. viewing documents).

The window on the lower left is the text window where documents are displayed. Queries can be constructed directly by clicking on words in the text to add them to the query. Dragging lines between words then connects them in the query with an AND operator.

Words can also be added to a query by typing them in the margin. For instance, in Figure 4 the word "degenerative" has been typed in the margin. The current query is displayed in the small gray field above the text window. In this case the current query is (research AND cancer OR degenerative). This display uses parentheses to disambiguate the scope of the AND and OR operators. The query displayed as text corresponds to the graphical query marked on the text of the document. Each of the terms in the query is shown within a box. The terms cancer and research are linked by a line, which indicates the AND operator. The fact that "degenerative" is not linked with a line to the other terms shows that it is ORed with them.

Above the query display area is a quite button, and open file button (this button wasn't used during the TREC-3 experiment), and a new query button, which gave the user a quick way to initialize the query (the alternative was to explicitly delete the terms in the current query using point and click selections). To the right of these buttons was a pull-down menu showing a list of the TREC search topic statements. The pull-down menu is currently showing topic 104 as being selected. The topic statement was normally used as the starting point for each search. The topic statement could be marked up like any other document.

The search controls at the top of the screen contained a proximity slider and a relevance judgment slider. The proximity slider was used to vary the proximity. In addition to moving the slider, the user could also type the new proximity value directly into the field above the slider (thus making it possible to have proximities of greater than 1000). In practice, the default proximity of 500 characters was used most of the time, with proximity mainly being adjusted to fine tune the query.

Output from the online dictionary is shown in the lower right part of the display in a window labeled "definition". In this case the dictionary reports that there is no definition for "degenerative". The lack of useful feedback from the online dictionary in this case was fairly representative of its effectiveness throughout the searches.

The coordinate terms were shown in one of the two empty windows in the lower right of the screen. Above them is a window that provided the feedback on recall and precision of the current query. This information was updated every time a hit was selected. The window above this allowed users to access the online thesaurus. The information from the thesaurus was then displayed in the definition window (the same place that displayed the online dictionary information).

The window to the left of the WordNet window displayed the list of documents judged to be relevant by NIST as per their procedure. Each line (choice) in the window represents a document. If the user clicks on one of the choices in the window, the corresponding document is shown in the text window.

The two windows above the expert judged relevance window are the query history window and the query construction workspace. The query history showed a listing of the previous queries that had been executed. The workspace provide a place where users could type words or terms that could later be pasted in to the margin. Alternatively the users could type words directly into the margin.

The large window above the query history showed the hit list. Each line in this list represented one document in the set of hits. When a line was clicked on in the hit list, the corresponding document was displayed in the text window, with the current query marked up on it. Thus users could edit the query by modifying the graphical query that overlaid the document in the text window.

The style of interface was a mixture of graphical user interface and command interface. Queries could be constructed using point and click interactions, but terms could also be typed in the margin or entered in the workspace. Since TREC-3 we have revised the system so that it retains the graphical user interface while adding more command-based functions and making it easier for users to construct complex queries.

II. Experimental Conditions

The TREC-3 experiment was carried out over an intense one week period. After some pilot testing with two experienced librarians, the actual searches were carried out by the two researchers. The two searchers worked on all fifty of the topics. One searcher was 29, the other 38 years of age. One of the searchers has had considerable searching experience, while the other searcher had only limited experience. However, the goal of our system was to minimize the need for experience and to make searching more productive for novices. To a large extent, there did not appear to be a significant difference between the search results obtained for the two searchers in our experiment. In terms of educational level, both searchers had more than eight years of post-secondary education. One searcher had an undergraduate degree in computer engineering, the other in psychology. Neither of the searchers had any experience or expertise with respect to any of the 50 TREC-3 topics. One searcher was an associate professor at the University of Toronto, the other was a Ph.D. student at the same university.

The searchers knew that they were participating in the TREC-3 evaluation and were familiar with the structures and the rules of TREC-3. Thus their goal was to construct a query that would maximize the recall and precision on the training data. To do this they could use any or all of the features and functions of ST-PatTREC.

The training process consisted of trying the system out and seeing what worked. Due to lack of time, and the fact that the searchers were also trying to iron out the bugs in the system, etc., there was little if any time for training. Our impression was that the earlier topics ended up being training for the later topics, and that our skill as searchers using ST-PatTREC increased significantly during the experiment. Thus we would say that no formal training was used.

III. Search Process

1. Clock time (i.e. real, elapsed time) per search, from the time the searcher was given the topic, until the final query was saved, in seconds.
Mean: 32.44 mins
Median: 28.50 mins
SD: 15.45 mins
Range: 12-60 mins

2. Number of documents "viewed" in during the search.
a. Definitions of viewing:
For the purposes of our system, viewing was defined as seeing a document in the text window. Users had to click on the hit list to bring a document into the text window and thus there was a clear indication of their intention to view documents.

b. Number of items viewed per search(repeat for each viewing category).
Mean: 18.18
Median: 16.5
SD: 7.912
Range: 3-42

3. Number of iterations per search.
a. Definition of iteration:
The searchers wrote up a search log that contained significant queries, along with recall and precision measures for those queries. An iteration was defined operationally as a new entry in the search log.

b. Number of iterations per search.
Mean: 8.9
Median: 8.5
SD: 3.388
Range: 3-18

4. Number of terms used in queries.
a. Number of terms in the first query of a search, per search.
Mean: 1
Median: 1
SD: 0
Range: 1

b. Number of terms in the final query of a search, per search.
Mean: 5.571
Median: 6
SD: 2.17
Range: 1-11

5. Use of system features.
Not available

6. Number of user errors made per search.
Not applicable

7. An Example of the system in Use: Topic 122

The following interaction shows how the interaction occurred for topic 122. This topic was concerned with new or alternative cancer drugs therapies. The title of the topic was "RDT&E of New Cancer Fighting drugs. The concepts provided in the topic statement included cancer, leukemia, drug, and chemotherapy. Figure 5 shows portions of the transaction log that our system produced for the interaction. The interaction started with the selection of the term "cancer" at 6:09 pm. on July 24. The transaction took almost exactly 20 minutes. This was a relative short interaction for the TREC-3 topics with our system, since the average topic took just over half an hour.

The user started by typing the words cancer, followed by drug, in the margin. The words were then linked together to create an AND. The term "development" was added to the query and then removed. At this stage an article was selected from the hit list. The user then added the term "research" to the query, first ORing it with cancer AND drug and then ANDing it to create the query research AND cancer AND drug. A document was then selected from the resulting hit list. According to the log, the user then had about 10 seconds to review the text of the document before the query was modified by removing the word "research". The user would also have noted the feedback for this query, which was a precision of .16 and a recall of .35. In contrast, the query "drug AND cancer" had provided a recall of .81 and a precision of .14, thus clearly being superior to the version of the query which added "research" to the AND combination. Although this log does not show it, the user had continuous updates on the precision and recall scores for each query executed. Thus, the evolution of the search strategy in constructing the test query was guided to a large extent by this feedback.

The user then tried adding "research" in a different version of the query, leading after an set of interactions over a period of almost three minutes to the query:

drug AND research AND cancer OR cancer AND development

A considerable amount of time was then spent trying to find a term that would improve the precision without adversely affecting recall, but without much success. The final enhancement to the query came when the term therapy was used (therapy was one of the coordinate terms for "cancer"). The final query eventually chosen for this topic was:

cancer AND drug OR cancer AND therapy

One interesting feature of this interaction is that of the four concept terms provided as part of the TREC-3 topic definition, only two (cancer, drug) were actually used in this session. Not using "leukemia" or "chemotherapy" was probably an oversight on the part of the user. When we amended the final query to include these terms , i.e., "cancer AND therapy OR cancer AND drug OR chemotherapy OR leukemia" the resulting precision and recall on the training database were .13 and .89, respectively. Thus "human errors" in forgetting to use, or consider, useful terms could have lowered the performance of queries using our system in some cases.

After this interaction, the final query submitted for the TREC-3 competition was "drug AND cancer" OR "cancer AND therapy". For the first 100 documents, 20 out of the 30 relevant documents were retrieved from this query, reflecting a recall of .67 and a precision of .20 for this subset of documents. This results were in line with the experiment gained in forming the query on the training data where using exact matching, the recall was .81 and the precision was .13. Expanding to 997 documents then yielded a further 6 relevant documents for a recall of .87. However the precision was then only .026. However, given that there were only 30 relevant documents, and 1000 documents were to be retrieved, the maximum precision that could have been obtained at that point was only about 3 percent in any case.

```
18:09:06 query cancer
18:09:06 addterm cancer
18:09:17 query drug OR cancer
18:09:17 addterm drug
18:09:22 query drug AND cancer
18:09:22 addop drug AND cancer
18:09:46 query development OR drug AND cancer
18:09:46 addterm development
18:09:54 query drug AND cancer
18:09:54 removeterm development
18:10:02 selecthit hitlist 3124737 3122285
18:10:02 selectdoc 3122285
18:11:45 selecthit hitlist 28740917 28736070
18:11:46 selectdoc 28736070
18:12:43 query research OR drug AND cancer
18:12:43 addterm research
18:12:51 query research AND drug AND cancer
18:12:51 addop research AND drug
18:12:52 selecthit hitlist 20201575 20196494
18:12:53 selectdoc 20196494
18:13:03 query drug AND cancer
18:13:03 removeterm research
18:13:12 query drug AND cancer OR research
18:13:12 addterm research
18:13:18 query drug AND cancer OR research OR cancer
...(this section of the log deleted to save space)...
18:24:51 query therapy AND cancer OR cancer
18:24:51 addterm cancer
18:24:55 query therapy AND cancer OR cancer OR drug
18:24:55 addterm drug
18:25:00 query therapy AND cancer OR cancer AND drug
18:25:00 addop cancer AND drug
18:25:01 selecthit hitlist 10646564 10646443
18:25:02 selectdoc 10646443
18:25:35 selecthit hitlist 7206040 7205634
18:25:36 selectdoc 7205634
18:26:24 newquery
18:26:30 query therapy cancer/cancer drug
18:26:30 query therapy AND cancer OR cancer AND drug
18:26:37 selecthit hitlist 3124615 3122285
18:26:37 selectdoc 3122285
18:28:05 query development Cancer/research cancer/Cancer drug/Cancer test
18:28:06 query development AND Cancer OR research AND cancer OR Cancer AND
          drug OR Cancer AND test
18:28:51 selecthitoffset2 31463575 31463482
18:28:51 selectdoc 31463482
18:29:06 selecthitoffset2 117419014 117418921
18:29:07 selectdoc 117418921
logout
```

Figure 5. The Transaction Log for Topic 122.