

# USERS' INTERACTIONS WITH THE *EXCITE* WEB SEARCH ENGINE: A QUERY REFORMULATION AND RELEVANCE FEEDBACK ANALYSIS

Amanda Spink, Carol Chang & Agnes Goz  
School of Library and Information Sciences  
University of North Texas  
[spink@lis.admin.unt.edu](mailto:spink@lis.admin.unt.edu)

Major Bernard J. Jansen  
Department of Electrical Engineering and Computer Science  
United States Military Academy  
[jjansen@acm.org](mailto:jjansen@acm.org)

Please Cite: Spink, A., Bateman, J., and Jansen, B. J. 1998. [Searching heterogeneous collections on the web: Behavior of Excite users.](#) *Information Research: An Electronic Journal*, 5(2).

## ABSTRACT

We conducted a transaction log analysis of 51,473 queries and 18,113 user sessions on *Excite* - a major Web search engine. The purpose of the study was to examine the use of query reformulation and relevance feedback by *Excite* users. There has been little research examining how reformulation and relevance feedback is used during Web searching. A total of 985 user search sessions (5% of all user sessions) were examined. A sample subset of 181 user sessions (2.9% of all 6045 sessions including more than one query or 1369 queries), were qualitatively analyzed to examine patterns of user query reformulation. All 804 sessions (4% of all sessions) using relevance feedback were analyzed. Results show limited use of query reformulation and relevance feedback by Web search engine users. Given the high level of research activity in relevance feedback techniques, there were a surprising small percentage of sessions including relevance feedback. Implications for system design are discussed.

[See Other Publications](#)

## INTRODUCTION

This paper investigates two important information retrieval (IR) techniques in the Web context – query reformulation by users and their use of relevance feedback options. Research shows that IR system users reformulate their queries during their search interactions to differing degrees (Efthimiadis, 1996). Little research has examined query reformulation by Web users. Relevance feedback is a classic information retrieval (IR) technique that reformulates a query based on documents identified by the user as relevant. Relevance feedback has been a major and active IR research area and reported to be successful in many IR systems (Harman, 1992; Spink & Losee, 1996). However, little research has investigated the use of relevance feedback options by Web users. The study of user queries to Web search engines is an important area of research to improve Web-based information access and retrieval.

The next section of the paper briefly discusses the related Web studies.

### Related Web Studies

A growing body of research is investigating many aspects of users' interactions with the Web. User oriented Web research generally includes experimental and comparative studies, user surveys, and user traffic studies. Experimental and comparative studies show little overlap in the results retrieved by different search engines based on the same queries (Ding & Marchionini, 1996, Lawrence & Giles, 1998). Many differences in search engine features and performance (Chu & Rosenthal, 1996) and users' Web searching behavior (Tomaiuolo & Packer, 1996) have been identified. A growing number of studies are comparing novice and expert Web searchers and some studies found regular patterns in Web users surfing behavior (Huberman, Pirolli, Pitkrow & Lukose, 1998). Many surveys of Web users have been conducted, either library based (Tillotson, Cherry & Clinton, 1995) or distributed via newsgroups. Pitkrow and Kehoe (1996) found major shifts in the characteristics of Web users over four surveys, including a growing diversity of Web users based on age, gender, and access through both the office and home computers. Hoffman and Novak (1998) found that white and African Americans use the Web differently. Whites were twice as likely to use the Web and have a computer in their home than African Americans. The Spink, Bateman and Jansen (1999) survey of *Excite* users shows many users perform successive searches of the Web on the same topic. However, few studies have examined the users' interactions with the Web or their use of query reformulation and relevance feedback.

The next section of the paper lists the specific research questions addressed in the study.

## RESEARCH QUESTIONS

The aim of our study of *Excite* user queries was to identify the:

1. Frequency and patterns of user query reformulation.
2. Use of relevance feedback.
3. Difference between relevance feedback users and the overall population of *Excite* users in term of searching characteristics?

The next section of the paper outlines the background on the *Excite* data corpus.

## RESEARCH DESIGN

### Excite Data Corpus

Founded in 1994, *Excite, Inc.* is a major Internet media public company that offers free Web searching and a variety of other services. The company and its services are described at its Web site, thus not repeated here. Only the search capabilities relevant to our study are summarized. *Excite* searches are based on the exact terms that a user enters in the query, however, capitalization is disregarded, with the exception of logical commands AND, OR, and AND NOT. Stemming is not available. An online thesaurus and concept linking method called Intelligent Concept Extraction (ICE) is used, to find related terms in addition to terms entered. Search results are provided in a ranked relevance order. A number of advanced search features are available. Those that pertain to our results are described here:

- As to search logic, Boolean operators AND, OR, AND NOT, and parentheses can be used, but these operators must appear in ALL CAPS and with a space on each side. When using Boolean operators ICE (concept-based search mechanism) is turned off.
- A set of terms enclosed in quotation marks (no space between quotation marks and terms) returns answers with the terms as a phrase in exact order.
- A + (plus) sign before a term (no space) requires that the term must be in an answer.
- A – (minus) sign before a term (no space) requires that the term must NOT be in an answer. We denote plus and minus signs, and quotation marks as modifiers.
- A page of search results contains ten answers at a time ranked as to relevance. For each site provided is the title, URL (Web site address), and a summary of its contents. Results can also be displayed by site and titles only. A user can click on the title to go to the Web site. A user can also click for the next page of ten answers.
- In addition, there is a clickable option *More Like This*, which is a relevance feedback mechanism to find similar sites. When *More Like This* is clicked, *Excite* enters and counts this as a query with zero terms. When using the *Excite* search engine, if the user finds a documents that is relevant, the user need only "click" on a hyperlink that implements the relevant feedback option. It does not appear to be any more difficult than normal Web navigation. In fact, one could say that the implementation of relevance feedback is one of the simplest IR techniques available on *Excite*.

The *Excite* transaction log record of 51,473 queries contained three fields of data on each query. **Time of Day:** measured in hours, minutes, and seconds from midnight (U.S. time) on 9 March 1997. With these three fields, we were able to locate a user's initial query and recreate the chronological series of actions by each user in a session:

1. **User Identification:** an anonymous user code assigned by the *Excite* server.
2. **Query Terms:** exactly as entered by the given user.

Focusing on our three levels of analysis, sessions, queries, and terms, we defined our variables in the following way.

1. **Session:** A session is the entire series of queries by a user over time. A session could be as short as one query or contain many queries.
2. **Query:** A query consists of one or more search terms, and possibly includes logical operators and modifiers.
3. **Term:** A term is any unbroken string of characters (i.e. a series of characters with no space between any of the characters). The characters in terms included everything – letters, numbers, and symbols. Terms were words, abbreviations, numbers, symbols, URLs, and any combination thereof.

The basic statistics related to queries and search terms are given in Table 1.

Table 1. Numbers of users (sessions), queries, and terms.

Number of Users (Sessions)	18, 113
Number of Queries	51,473
Number of Terms	113,793
Mean Queries Per User	2.84
Mean Terms Per Query	2.21
Mean Number of Pages of Results Viewed	2.35

### Data Analysis

We conducted a transaction log analysis of 51,473 queries and 18,113 user sessions on *Excite* - a major Web search engine. The purpose of the study was to examine the use of query reformulation and relevance feedback by *Excite* users. Specific analysis was conducted on a total of 985 user sessions (5% of all sessions). We quantitatively and qualitatively analyzed the 18,113 user sessions and conducted specific analysis of 985 user sessions, including:

- A sample subset of 181 sessions (1% of all user sessions and 1369 queries) that included query reformulation were qualitatively analyzed to examine user patterns of query reformulation.
- All 804 user sessions (4% of all user sessions) including relevance feedback were analyzed. Of the 51,473 queries only about 6% (2,543) could have been from *Excite's* relevance feedback option. This is a surprisingly small percentage of the queries relative to usage in IR systems. From our analysis, we identified queries resulting from the relevance feedback option and isolated the sessions (i.e., sequence of queries by a user over time) that contained relevance feedback queries.

Working with these user sessions, we classified each query in the session by query type. We identified patterns in these sessions composed of transitions from query type to and to another query type. By classifying these session patterns we hope to gain insight in the current use of query reformulation and relevance feedback on the Web.

The next section of the paper presents the results of our analysis.

## RESULTS

This paper is part of a large project analyzing the *Excite* data set and extends findings reported in Jansen, Spink, and Saracevic (1998a, b, c, in press). The first section of the paper reports findings from the analysis of user query formulation.

### Query Reformulation Analysis

We examined users' query modification and patterns of query reformulation.

### Query Modification

Table 2 shows the distribution of the number of queries per user session.

Table 2. Number of queries per user session.

Queries Per User Session	Number of User Sessions	Percent of Users Sessions
1	12,068	67
2	3,501	19
3	1,321	7
4	583	3
5	287	1.6
6	144	0.80
7	79	0.44
8	32	0.18
9	36	0.20
10	17	0.09

11	7	0.04
12	8	0.04
13	15	0.08
14	2	0.01
15	2	0.01
17	1	0.01
25	1	0.01

Some users entered only one query in their session and others entered successive queries. The average session was 2.84 queries in length. This means that 33% (6045) of users went on to either modify their query, view subsequent results, or both.

We also examined how users modified their queries. These results are display in Table 3.

Table 3. Changes in number of terms in successive queries

Increase in Terms	Number	Percent
0	3909	34.76
1	2140	19.03
2	1068	9.50
3	367	3.26
4	155	1.38
5	70	0.62
6	22	0.20
7	6	0.05
8	10	0.09
9	1	0.01
10	4	0.04
Decrease in Terms	Number	Percent
-1	1837	16.33

-2	937	8.33
-3	388	3.45
-4	181	1.61
-5	76	0.68
-6	46	0.41
-7	14	0.12
-8	8	0.07
-9	2	0.02
-10	6	0.05

In Table 3 we concentrate on the 11,249 queries that were modified by either an increase or a decrease in the number of terms from one user's query to that user's next query (i.e., successive queries by the same user at time T and T+1). Zero change means that the user modified one or more terms in a query, but did not change the number of terms in the successive query. Increase or decrease of one means that one term was added to or subtracted from the preceding query. Percent is based on the number of queries in relation to all modified (11,249) queries.

Query modification was not a typical occurrence. This finding is contrary to experiences in searching regular IR systems, where modification of queries is much more common. Having said this, however, 33% of the users did go beyond their first query. Approximately 14% of users entered (3) three or more queries. These percentages of 33% and 14% are not insignificant proportions of system users. It suggests that a substantial percentage of Web users do not fit the stereotypical naïve Web user. These sub-populations of users should receive further study. They could represent sub-populations of Web users with more experience or higher motivation who perform query modification on the Web. We can see that users typically do not add or delete much in respect to the number of terms in their successive queries. Query modifications were done in small increments, if at all.

The most common modification is to change a term. This number is reflected in the queries with zero (0) increase or decrease in terms. *About one in every three queries that is modified still had the same number of terms as the preceding one.* In the remaining 7,338 successive queries where terms were either added or subtracted, about equal numbers had terms added as subtracted (52% to 48%) - thus users go both ways in increasing and decreasing number of terms in queries. *About one in five queries that is modified has one more term than the preceding one, and about one in six has one less term.*

The next phase of the study investigated user patterns of query modification. We analyzed subset of 181 user sessions - 33% of all sessions that included 2 (two) or more queries in the data set.

### Patterns of Query Modification

Table 4 shows the basic data for the qualitative query modification analysis.

Table 4. Basic data.

Number of User Sessions	Number of Queries	Number of Terms	Mean Terms & Range
181	1369	3015	2.19 0-10

We analyzed 181 user sessions to examine how successive queries differed from other queries by the same user during the same session. Each query was first classified as either:

- *Unique Query (U)*: Unique query by a user.
- *Modified Query (M)*: Subsequent query in succession (second, third ...) by the same user with terms added to, removed from, or both added to and removed from the unique query.
- *Next Page(N)*: When a user views the second and further pages (i.e., a page is a group of 10 results) of results from the same query, *Excite* provides another query, but a query that is identical to the preceding one.
- *Relevance Feedback(R)*: when a user enters a command for relevance feedback (*More Like This*), the *Excite* transaction log show that as a query with zero terms.

Table 5 shows the number of occurrences of each type of query.

Table 5. Number of occurrences of each query type.

Query Type	Number of Queries	Percentage of Queries
Unique Query (U)	340	25%
Modified Query (M)	274	20%
Next Page (P)	642	46%
Relevance Feedback (R)	123	9%
Total	1379	100%

- The mean number of queries per user was 2.19 with a range from 2-10.



- Overall, users performed limited query modification. 1 in 5 queries were modified queries.
- 1 in 2 queries were requests for the next page of results.
- Less than 1 in 10 queries were relevance feedback.

We next examined how the succession of queries differed amongst users. For example, if a user enters a unique query (U) followed by three modified queries (M) and finally a next page (N) – this is represented as the shift pattern UMN. The user shifted from a unique query to modified queries to looking at the second 10 retrieved Web sites.

Table 6 shows the number of occurrences of each session shift pattern in the data set.

Table 6. Patterns of user sessions.

Pattern	Number of User Sessions	Percentage of User Sessions
UP	52	28%
UM	18	10%
UPM	8	4.2%
UPU	8	4.2%
UMP	8	4.2%
UPMP	7	3.6%
UPR	5	2.6%
UR	4	2%
UPMPM	2	1%
UPMPMP	2	1%
UMPMP	2	1%
UMUM	2	1%
RU	2	1%
Other (Unique)	59	31%
Total	181	100%

The data analysis shows that:

- The most common user session was a unique query followed by a request to view the next page of results with no query modification.
- 1 in 2 users viewed the next page of results before modifying their query.
- 1 in 2 user sessions included query modification
- 1 in 4 users included a unique query followed by a next page and conducted no query modification.
- Not a lot of subject change – 73% of user sessions included one (1) topic and 27% included two (2) topics.
- 1 in 3 user sessions were unique in their query patterns, e.g., only 1 user entered the pattern NMPRMP – a unique query followed by a modified query, then a next page, a relevance feedback, followed by a modified query and a next page.
- Relevance feedback was not used extensively during user sessions.

The following sections contain results from the analysis of all 804 relevance feedback sessions.

### Use and Frequency of Relevance Feedback

From the 51,473 query transaction log, a maximum of 2,543 (6%) of queries could have been from relevance feedback. There are more complicated IR techniques, such as Boolean operators and term weighting, that are used more frequently (Jansen, Spink, & Saracevic, 1998, in press). We found it surprising that this highly touted and widely research IR feature, implemented in straightforward fashion, was so seldom utilized.

A note should be made on queries with zero terms. As mentioned, when a user enters a command for relevance feedback (*More Like This*), the *Excite* transaction log counts that as a query, but a query with zero terms. Thus, the last row represents the largest possible number of queries that used relevance feedback, or a combination of those and queries where user made some mistake that triggered this result. Assuming they were all relevance feedback, only 5% of queries used that feature – a small use of relevance feedback capability. In comparison, a study involving IR searches conducted by professional searchers as they interact with users found that some 11% of search terms came from relevance feedback (Spink & Saracevic, 1997), albeit this study looked at human initiated relevance feedback. Thus, in these two studies, relevance feedback on the Web is used half as much as in traditional IR searches. This in itself warrants further study, particularly given the low use of this potentially highly useful and certainly highly vaunted feature.

Given the way that the transaction log recorded user actions, the relevance feedback option was recorded as a null (i.e., empty and with no terms) query. However, if a user entered an empty query, the empty query would also be recorded the same (i.e., empty and with no terms). For this study, we counted these as mistakes. Using a purely quantitative analysis, we had previously reported that there were 2,543 queries representing the maximum possible amount of relevance feedback queries. For this study,

we had to separate the relevance feedback queries from null queries. Therefore, we reviewed the transaction log data and removed all queries that were not relevance feedback. The vast majority of null queries that were the first query in a session. Obviously, these queries could not be a result of relevance feedback. If a determination could not be made, the query was considered as a result of relevance feedback. The results are summarized in Table 7.

Table 7: Percentage of relevance feedback queries.

Classification	Number of Queries	Percentage of Queries
Relevance Feedback	1597	63%
Null Queries	946	37%
Total	2543	100%

From Table 7, one sees that fully 37% of the possible relevance feedback queries were judged not to be relevance feedback queries but instead a blank first query. This result in itself is very interesting and noteworthy. It implies that something with the complexity of the interface, system, or network is causing users to enter null queries just under 40% of the time. From observational evidence, some novice users "click" on the search button prior to entering terms in the search box thinking that the button takes them to a screen for searching. Additionally, Peters (1993) states that users many times enter null queries.

### Patterns of Relevance Feedback

Regardless of the reason for the mistakes, the maximum possible relevance feedback queries were 1,597. These queries resulted from all 804 user sessions including an occurrence of relevance feedback, with an average of 1.99 relevance feedback queries per user session. Working with these user sessions, we classified each query in the session as belonging to one of the types listed previously in Table 5.

We first looked at the number of occurrences of each type of query.

### Query Analysis

As stated, we classified queries within each session. The number of occurrences of each query for all 804 user sessions is displayed in Table 8.

Table 8: Occurrences of occurrences of query types.

Query Type	Number of Queries	Percentage of Queries
Relevance Feedback (R)	872	40.6%

Next Page (P)	693	32.2%
Modified Query (M)	467	21.8%
Unique Query (U)	116	5.4%
Total	2148	100%

There were 2148 queries in the 804 user sessions including relevance feedback. As to be expected, *Relevance Feedback* occurred by far the most (872 occurrences) followed by *Next Page* (693 occurrences). This indicates that there was a number of viewing of subsequent results by users. There were also a large number of modified queries, indicating the addition, removal, or change of query terms.

### Query Transitions

We then examined the occurrence of each query type within each user session as opposed to the overall totals.

- The shortest session was two queries since every session had to at least consist of *Query*

### *Relevance Feedback.*

- The maximum session length was seventeen shifts in query type.
- If a query type occurred in succession, we counted it as only occurring once. For example, a session of *Query -> Relevance Feedback -> Relevance Feedback*, the *Relevance Feedback* query would be counted as only occurring one time within that session. We did this to simplify the pattern and isolate the state transitions from one state to another. However, using the above example, if the *Relevance Feedback* occurred lasted in the session, following another query type, it would be counted again. For example, *Query -> Relevance Feedback -> New Query-> Relevance Feedback*. In this example, *Relevance Feedback* would be counted twice.
- Given that there were no one query session in this sample (i.e., the shortest session was *Query -> Relevance Feedback*, a two query session), there were 239 two query sessions, the largest group. However, there were 251 three query sessions, 120 four query session, 82 five query sessions, followed by a fair number of six and seven query sessions. After that, there is a sharp drop-off in session length. As the length of the session increases, the occurrences of relevance feedback decreased.
- For the sessions of two and three queries, the relevance feedback state is the dominant Query type. As the length of the sessions increased, the occurrences of relevance feedback as a percentage of all query types decreased.

Beginning with session of five queries or more, relevance is no longer the query type with the most occurrences. This needs further investigation as to the cause of this decreased use of relevance feedback.

### Relevance Feedback Sessions Analysis

Given the low occurrences of relevance feedback queries, we attempted to determine if the session containing relevance feedback was successful or not. Without access to the users, this was difficult. However, we can make some generalizations. If the user utilized relevance feedback and then quit searching, we gave relevance feedback the benefit of the doubt and counted it as a success (i.e., assumed the user found something of relevance). This pattern seems consistent with what one would expect of a successful search session. Possibly, many times these sessions were not successful, so this count represents the maximum number of successful sessions. If the user utilized relevance feedback and returned to the exact previous query, we assumed that nothing of value was found (i.e., an unsuccessful session). This pattern seems consistent with what one would expect of an unsuccessful search session (i.e., assumed the user nothing of relevance).

There were some sessions where the user used relevance feedback and returned to a similar but not exact query. Since the relevance feedback query could have provide some terms suggestions, we classified these session as partially successful. The results of this analysis are summarized in Table 9.

Table 9: Classification of relevance feedback sessions.

Classification	Number of Occurrences	Percentage
Successful	509	63%
Unsuccessful	156	20%
Partially Successful	139	17%
Total		100%

As one can see, giving relevance feedback the benefit of the doubt, fully 63% of the relevance session could be construed as being successful. If the partially successful are included, then approximately 80% of the relevance feedback sessions provide some measure of success. This is a fairly high percentage; although as mentioned, we are presenting the maximum possible number of successful sessions. The question then becomes, why is relevance feedback not used more on the Web search engine? In order to gain greater insight to this behavior, we compared the population if RF users to the larger population in our data set to see if there was some difference that set this sub-set apart from the larger population.

### Comparison to Larger Population of *Excite* Users

We first examined the query construction of relevance feedback users to the query construction of the general population. The actual numbers from the larger population are unimportant. The important item of comparison is the percentages. The actual numbers are available in (Jansen, Spink, & Saracevic, 1998, in press). The comparison is displayed in Table 10.

Table 10: Terms per query.

Terms Per Query	Number of Queries in RF Population	Percent of RF Queries	Percent in General Population
0	872	21%	6%
1	972	23%	31%
2	1045	25%	31%
3	635	15%	18%
4	310	7%	7%
5	195	5%	4%
6	70	2%	1%
7	36	1%	0.94%
8	23	1%	0.44%
9	3	0%	0.24%
> 10	22	1%	0.36%
Total		100%	100%

There appears to be little different between the relevance feedback users and the population in general in number of query terms, other than zero term queries (e.g., the relevance feedback queries) of course. The average number of terms per query was 1.98 for the relevance feedback population and 2.2 for the larger population. Assuming that lengthier queries are a sign of a more sophisticated user, it appears that the relevance feedback population does not differ significantly from the larger population of *Excite*, and possibly, Web users.

Next, we examined the number of queries per user. This data is displayed in Table 11.

Table 11: Queries Per Session.

Query Per User	Number of Users	Percentage of RF Users	Percentage of General Population
1	3	0.36%	67%
2	375	45%	19%
3	223	27%	7%
4	97	12%	3%
5	64	8%	2%

6	34	4%	0.80%
7	11	1%	0.44%
8	4	0.48%	0.18%
9	8	0.97%	0.20%
10	6	0.72%	0.09%
11	1	0.12%	0.04%
> 12	1	0.12%	0.04%
	827	100%	100%

Concerning the number of queries per user, the relevance feedback population had significantly longer sessions than the population at large. The median number of queries per user for the relevance feedback population was 2 and for the larger population it was 1. There were also a significant number of relevance feedback users with sessions of 3, 4, 5, and even 6 queries. In the larger population, there is a steep drop-off at 2 queries per user. This may indicate that relevance feedback users were more persistent in satisfying their information need and therefore more willing to invest the time and effort to use not only relevance feedback but also longer sessions in general.

## DISCUSSION

We conducted a transaction log analysis of 51,473 queries and 18,113 users of *Excite*, a major Web search engine. Of the over 51,473 queries only about 6% were from *Excite's* relevance feedback option. This is a small percentage of the queries. In order to gain insight into the possible causes of this phenomenon, we analyzed the user sessions that contained the approximately 2,543 relevance feedback queries. Given the way that the transaction log recorded user actions, the relevance feedback option was recorded as a empty query. Fully 37% of the possible relevance feedback queries were judged not to be relevance feedback queries but instead a null query.

We isolated states within each user session, identifying 4 possible query types, which are: *unique query*, *relevance feedback*, *modified query*, and *next page*. Of these query types, not counting the *unique query* state, *relevance feedback* was the most common, occurring 872 times. There was an average of 1.99 relevance feedback queries per user session.

Most users entered only a single query. A third of users went beyond the single query, with a smaller group using either query modification or relevance feedback, or viewing more than the first page of results. We then examined the occurrence of each query type in user sessions; the shortest user session was two queries. We see that the distribution of query type shifts as the length of the user session increase. For the user sessions of two and three queries, the relevance feedback query is dominant. As the length of the queries increase, the occurrences of relevance feedback as a percentage of all query types decreases. Given the low occurrences of relevance feedback queries, we attempted to determine if the user sessions containing relevance feedback were successful or not. Given relevance feedback the benefit of the doubt, fully 63% of the relevance feedback

sessions could be construed as being successful. If the partially successful user sessions are included, then almost 80% of the relevance feedback sessions provide some measure of success.

We then compared the relevance population to the larger *Excite* population. We first examined the query construction of relevance feedback users to the query construction of the general population. There appears to be little difference between the relevance feedback users and the population in general. Both populations had average query lengths of about two terms.

Next, we examined the number of queries per user. The relevance feedback population had significantly longer queries than the population at large. The median number of queries per user for the relevance feedback population was 2 and for the general population it was 1.

## CONCLUSION

The data and analysis suggest that relevance feedback is successful for Web users, although only a small percentage of Web users take advantage of this feature. On the other hand, although it is successful 63% of the time, this implies a 37% failure rate or at least a not totally successful rate of 37%. This may be one reason relevance feedback is so seldom utilized. Its success rate on the Web is just too low. It points to the need for an extremely high success rate before Web users consider it beneficial. As for characteristics of the relevance feedback population, they do not differ in terms of query construction, but they exhibit more doggedness in attempting to locate relevance information. This is manifested in longer sessions. This could be for several reasons, perhaps the subjects they are searching for are more intellectually demanding. Unfortunately, a cursory analysis of the query subject matter and terms does not support this conclusion. It probably points to some detail of sophistication in searching techniques. So, relevance feedback options appear to attract a more sophisticated Web user.

If these results can be generalized to other Web search engines other than *Excite*, it points to the need to tailor the interface if the goal is to increase the use of relevance feedback. Also, the precision of the relevance feedback option must be increased.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Graham Spencer, Doug Cutting, Amy Smith and Catherine Yip of *Excite, Inc.* in providing the data and information for this research. Without the generous sharing of data by Excite Inc. this research would not be possible. We also acknowledge the generous support of our institutions for this research.

## REFERENCES



Chu, H., & Rosenthal, M. (1996). Search engines for the World Wide Web: A comparative study and evaluation methodology. *Proceedings of the 59<sup>th</sup> Annual Meeting of the American Society for Information Science, Baltimore, MD* (pp. 127-135).

Ding, W., & Marchionini, G. (1996). A comparative study of Web search service performance. *Proceedings of the 59<sup>th</sup> Annual Meeting of the American Society for Information Science, Baltimore, MD* (pp. 136-142).

Efthimiades, E. (1996). Query expansion. *Annual Review of Information Science and Technology, 31*, 121-188.

Harman, D. K. (1992). In: Belkin, N. J.; Ingwersen, P., Pejtersen, A.M., eds. *SIGIR 92: Proceedings of the Association for Computing Machinery Special Interest Group on Information Retrieval (ACM/SIGIR) 15<sup>th</sup> Annual International Conference of Research and Development in Information Retrieval; 1992 June 21-24* (pp. 1-10).

Huberman, B. A., Pirolli, P., Pitkow, J. E., & Lukose, R. M. (1998). Strong regularities in World Wide Web surfing. *Science, 280*(5360), 95-97.

Jansen, B. J., Spink, A., & Saracevic, T. (in press). Real life, real users, and real users: A study and analysis of user queries on the Web. *Information Processing and Management*.

Jansen, B. J., Spink, A., & Saracevic, T. (1998a). Failure analysis in query construction: Data and analysis from a large sample of Web queries. *Proceedings of the Third ACM Conference on Digital Libraries, June 1998, Pittsburgh, P.A.* (pp. 289-290).

Jansen, B. J., Spink, A., & Saracevic, T. (1998b). Searchers, the subjects they search, and sufficiency: A study of a large sample of Excite searches. *Proceedings of WebNet 98 Conference, November 1998, Orlando, FL*.

Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998c). Real life information retrieval: A study of user queries on the Web. *SIGIR Forum, 32*(1), 5-17

Lawrence, & Giles, C. L. (1998). Searching the World Wide Web. *Science, 280*(5360), 98-100.

Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech, 42*:11:2, 41-66.

Pitkow, J. E., & Kehoe, C. M. (1996). Emerging trends in the WWW user. *Communications of the ACM, 39*(6), 106-108.

Spink, A., Bateman, J., & Jansen, B. J. (1999). Searching the Web: Survey of Excite users. *Internet Research: Electronic Networking Applications and Policy*.

Spink, A., & Losee, R. M. (1996). Feedback in information retrieval. In: M. Williams (Ed.), *Annual Review of Information Science and Technology*, 31, 33-78.

Spink, A., & Saracevic, T. (1997). Interaction in information retrieval: Selection and effectiveness of search terms. *Journal of the American Society for Information Science*, 48(8), 728-740.

Tillotson, J., Cherry, J., & Clinton, M. (1995). Internet use through the University of Toronto: Demographics, destinations and users' reactions. *Information Technology and Libraries*, September, 190-198.

Tomaiuolo, N. G., & Packer, J. G. (1996). An analysis of Internet search engines: Assessment of over 200 search queries. *Computers in Libraries*, 16(6), 58-62.