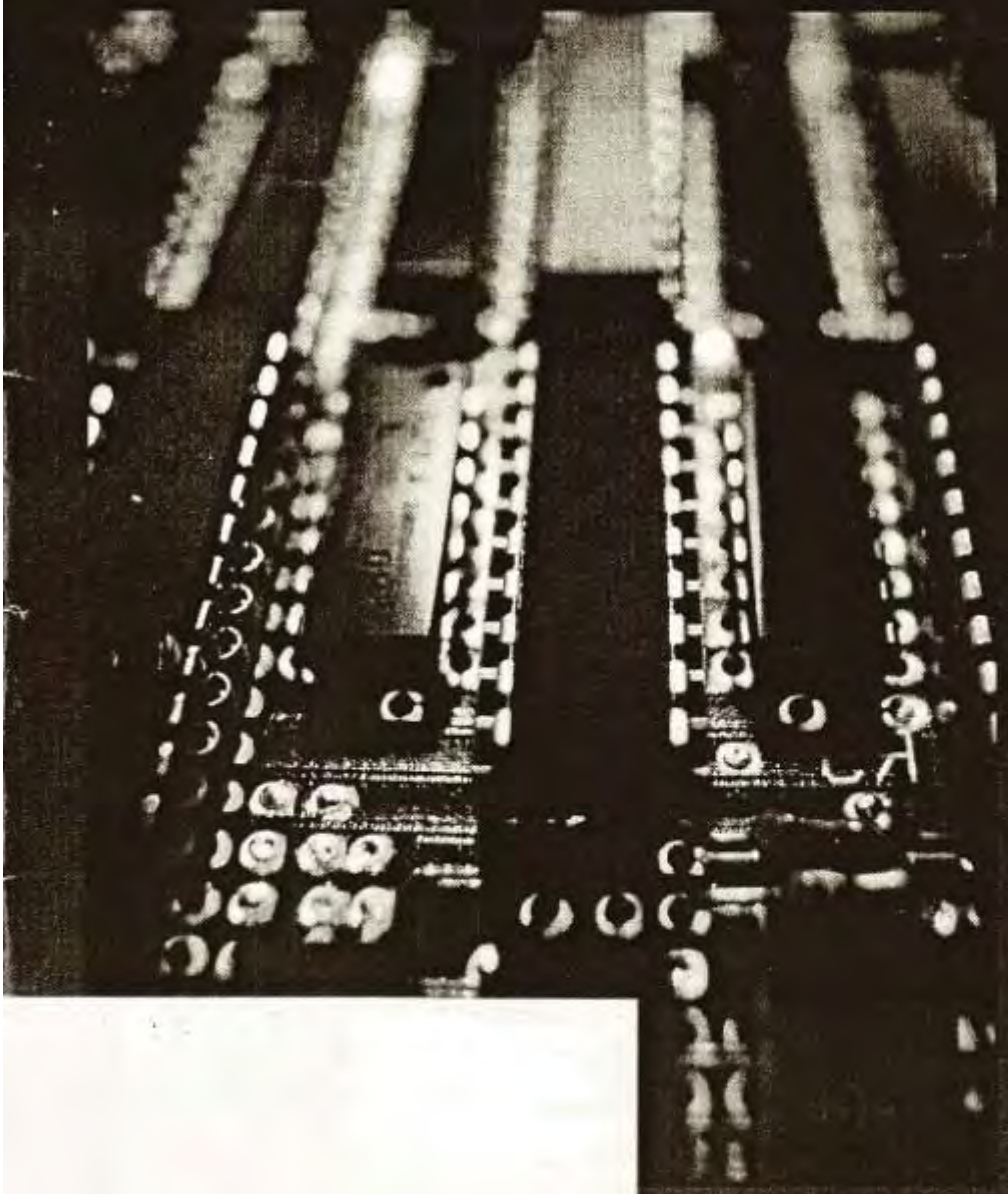
Information Technology AND Libraries

ISSN 0730-9295

Volume 23 Number 2

JUNE 2004



IN THIS ISSUE

Information Technology and Libraries (ISSN 0730-9295) is published quarterly in March, June, September, and December by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Library and Information Technology Association, a division of the American Library Association. SUBSCRIPTIONS: Subscription price, $30, is included in membership dues. Nonmembers may subscribe for $55 per year in the U.S.; $60 in Canada, Mexico, Spain, and other PUAS countries; $65 in other foreign countries. Send subscription requests to *Information Technology and Libraries*, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611; 800-545-2433; fax: (312) 944-2641; subscription@ala.org. Single copies, $20. Periodical-class postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

Information Technology and Libraries publishes material related to all aspects of libraries and information technology, including digital libraries, metadata, authorization and authentication, electronic journals and electronic publishing, telecommunications, distributed systems and networks, computer security and intellectual property rights, technical standards, geographic information systems, desktop applications, online catalogs and bibliographic systems, optical information systems, software engineering, universal access to technology, futuristic forecasting, library consortia, vendor relations, and technology and the arts. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 92 of this issue or on the LITA Web site at www.lita.org/ital. Manuscripts of feature articles, communications, and tutorials should be addressed to Dan Marmion, 221 Hesburgh Library, The University of Notre Dame, Notre Dame, IN 46556-5629. Books for review should be addressed to Tom Zillner, WILS, 464 Memorial Library, 728 State St., Madison, WI 53706. Software for review should be addressed to Andy Boze, Hesburgh Library, The University of Notre Dame, Notre Dame, IN 46556-5629.

ADVERTISING: The Goldman Group, 14497 N. Dale Mabry Hwy., Ste. 205N, Tampa, FL 33618; telephone (813) 264-2772; fax (813) 264-2343.

As a matter of policy, *ITAL*, as the scholarly organ of LITA, does not review LITA publications.

PRODUCTION: ALA Production Services (Troy D. Linker, Angela Renshaw; Ellie Barta-Moran, Kristen McKulski, and Karen Sheets); American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by LITA or the ALA.

Abstracted in *Computer & Information Systems Abstracts, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal: Nauchnaya i Tekhnicheskaya Informatsiya, Otdelnyi Vypusk*, and *Science Abstracts Publications*. Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, NewSearch*, and *Social Sciences Citation Index*. Microfilm copies available to subscribers from Proquest, Ann Arbor, Michigan.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.∞

Information Technology AND Libraries

Volume 23, Number 2 June 2004 ISSN 0730-9295

Abstracts and contents pages of recent issues of *ITAL* can be found on the LITA Web site at www.lita.org/ital.

cover design Kevin Heubusch ● **interior design** Dianne M. Rooney

The Effect of Search Engine Keyword Choice and Demographic Features on Internet Searching Success

Melius Weideman and Corrie Strümpfer

The objective of this project was to determine the effect of keyword choice and demographic features on Internet searching success through empirical research. An experiment was done with 1,109 learners, spread across forty-six institutes of higher education on three continents. A variety of relationships were then inspected: that between the number of keywords used, age, race, and gender, and the searching-success rate. The results proved that the first three of these have measurable effects on searching success, while gender has only a marginal effect.

The human race has been involved in storing data and information for centuries. The effective retrieval of relevant information in a short time has always been a commonly experienced problem in this area. The purpose of this paper is to report on a literature study and an empirical experiment on retrieval of relevant information from the Internet.

Work on storage and retrieval of information started approximately at the beginning of the third millennium B.C. The Sumerians are credited as being the first people to store and classify written materials into library collections, with the purpose of allowing various social groups to function better.[1] Everyday activities and literature were recorded on clay tablets that were stored in special areas, with only a label bearing the opening words of the document as sole method of indexing.[2] The physical creation of these clay labels could be viewed as one of the first implementations of technology towards establishing indexing. In the absence of any advanced technological tools to make information retrieval possible, these libraries were little more than marked collections of documents.

Indexing and simple classification of manuscripts were done during the Middle Ages. As a result of the coding schemes and alphabetical keys in use, the indexers involved in these tasks were surrounded by an aura of mysticism.[3] Work on cataloging started in the Middle Ages, using written card catalogs and guard books.

The current high-powered computer era, where document matching is made through inverted indexes, string, and positional searches, has provided the much-needed technology to empower the storage and retrieval of information. This increasingly powerful technology has removed economic constraint on the searching mechanism: any characteristic of the document can now be matched to a search query. In fact, there is now no technical constraint to prevent an index from including every single term of a given textual document in the index. An early example of this can be seen in bible concordances such as *Strong's Exhaustive Concordance*, first published in 1890.

Melius Weideman (meliusw@yahoo.com) is Head of Research Planning and Capacity Building; **Corrie Strümpfer** (corrie@ctech.ac.za) is a Statistician and Senior Lecturer in Research, Faculty of Business Informatics, Cape Technikon, Cape Town, South Africa.

Other Research

A large amount of work has been done on information retrieval in general and on information retrieval from the Internet specifically. Some of these works are compared and reviewed here.

Early Work

The most ideal representation of a document is simply to include it as the index, but the initial absence and later the limitations of technological tools (such as storage space) made this ideal impossible to achieve.

During the late 1950s and 1960s, noted authors in the area of document-content presentation did landmark work. The controversial Uniterm system sparked interest in the United Kingdom and the United States, leading to the Cranfield tests discussed by Cleverdon and Keen, Robertson, and Tonta.[4] In this system, documents were indexed via a single term (hence the name) that was extracted from the document title or abstract. After some structured tests, Uniterm results were compared to those using more traditional indexing methods. The test apparently broke down owing to the disagreement over relevance judgment, and the results were inconclusive. One group of testers claimed that the Uniterm system worked well, while the other claimed the exact opposite.[5]

The actual Cranfield series of tests was done at the College of Aeronautics, Cranfield, UK. It involved another comparison of performance, this time between the Uniterm system and a modified version of the Universal Decimal Classification (UDC) system. A subset of two hundred documents on aeronautics was extracted and used as the master collection of documents. A further extract of forty documents from this set was made, and these documents were used to generate a set of forty artificial requests. The assumption was that if query number one was submitted, document number one should be returned as being the most relevant document from the collection of two hundred documents. Although certain inherent limitations of the study were evident, it did prove the effectiveness of the Uniterm system above the UDC classification.[6]

This test was followed by the so-called Cranfield I test, which involved a comparison between four different indexing systems. In the Cranfield I test, a collection of

eighteen thousand master documents was used, and it was indexed using each one of the four indexing systems. Using the same source-document principle as used in the previous test, twelve hundred questions were drawn up based on these documents. Searches were done, and the success or failure for each one noted. The performance of the four indexing systems was approximately equivalent, with a success level of between 73.8 percent and 82 percent. Again it was proved that the Uniterm system performed equally well compared to more conventional systems. Many authors leveled extensive criticism against the Cranfield I tests, mostly resulting from the concept of using source documents from which to derive the queries.[7]

Another series of tests, aptly called Cranfield II, was undertaken next. This series was based on different indexing devices and languages, rather than being a test of indexing systems. A total of 211 search requests were elicited from the original authors of each document (out of a sample of fourteen hundred documents). Thus, the measure of effectiveness of this test compared to Cranfield I was explicitly relevance-based. It proved that single-term index languages delivered the best performance, and that precision and recall are inversely related.[8]

Two other studies described by Tonta worth mentioning are Blair and Maron's full-text retrieval system study, and the Dewey Decimal Classification (DDC) online project of Markey and Demeyer. Blair and Maron used precision and recall as performance measures in their study on a forty-thousand-document database, based on fifty-one queries from two lawyers. They proved that recall failures occurred much more frequently than expected, and that the mean precision and mean recall ratios were 79 percent and 20 percent respectively. Markey and Demeyer viewed the DDC system as an online searcher's tool, and concluded that there was no relationship between the precision of online searches and the search satisfaction of the participants.[9]

The experiments done during this research project will revisit the issue of using a single term as search key, which will provide interesting comparisons with the Uniterm system previously covered.

Internet Searching

Fidel and others did a study to analyze the Web-searching behavior of high school students while searching for information for homework assignments. The participants' discussions as they were busy searching were audiotaped and used as later evidence of findings. It was found that searchers were satisfied with their results in general, but complained about slow responses.[10]

A study was done by Hirsh to explore the relevance criteria and search strategies applied by elementary-school children while searching for information. It was found that the searchers displayed little concern for the authenticity of the materials found, and spent a large amount of time on finding pictures. Most searchers also could not formulate effective search queries; they would misspell search terms and did not use Boolean operators.[11] This lack of concern about the validity of resources was echoed by the findings of Grimes and Boening. They found that learners used unevaluated Web resources, and that instructor expectations and actual learner usage of Web resources differed widely.[12]

Marcella and Baxter did a study on information needs and information-seeking behavior in the United Kingdom. This study involved a large number of electronic sources, including the Internet. Only a small percentage of the respondents preferred the use of a computer above traditional methods (books, libraries, friends) to satisfy their information needs, while a large percentage indicated libraries as their preferred source.[13]

Large and Beheshti interviewed fifty schoolchildren to find out how they searched for information on the Web. Sophistication in the information-retrieval strategies of these children was evident, admittedly after having been shown at least two basic search methods.[14]

Ross did an empirical investigation of the subject content of queries submitted to the Excite search engine. Since three or even four dimensions were necessary to map relationships between topic contents, it became clear that the intertopic similarity required a complex set of dimensions to describe it.[15]

There appears to be a pattern of many searchers experiencing failure in their endeavors, while some did have success. Either way, no reports of completely trouble-free searching were found.

Searching Methods

As with any process, information seeking requires a structured methodology to ensure success. Even early authors elaborated on searching strategies to be used in information retrieval long before the advent of the Internet.[16] Since there are so many variables involved, this methodology cannot simply be described as a series of steps that will guarantee success. However, Marchionini did define a sequence of steps followed in information retrieval, with a number of complex relations between them: (1) problem definition; (2) source selection; (3) query formulation; (4) query execution; (5) result examination; (6) information extraction; and (7) reflection.[17]

An early study done by Fidel on users searching for electronically stored information begins by stating a belief that end users will one day do their own searching, provided search processes have been simplified.[18] As a result of the openness of the Internet, the computing community has certainly reached that point.

The same author did an in-depth study in 1991 on the selection of search keys by searchers, described in three articles. In the first, a case study was done with forty-seven professional online searchers, proving that a formal model of searching behavior can be created even with a relatively small number of searchers.[19] In the second article, results proved that searchers based their key selection on thesauri and indexing, and that multidatabase searches prescribed the existence of high-quality thesauri and the switching from one language to another.[20]

In the third article, an important distinction between the way searchers altered their methodology while searching came to light: so-called operational as opposed to conceptual moves (where a move is a modification of a search strategy).[21] Operationalist searchers attempt to apply the most optimal search strategies to achieve precise retrieval. During their interaction, they maximize the system capabilities but retain the specific meaning of the original search request. Conceptualist searchers, however, fit a specific request into a faceted structure, and are mainly concerned with recall. They may also change the specific meaning of the original search request in an attempt to find the best results.[22] Another author designed and tested an interface, increasing the success ratio of searchers. This was achieved by allowing the user to concentrate on concepts while leaving the technical issues, like operators, to a program.[23]

According to a study by Kuhlthau, a searcher passes through six stages during information searching, with certain emotions and thoughts characterizing each phase: initiation, selection, exploration, formulation, collection, and presentation.[24]

Oliver and Oliver proved that high levels of learning and knowledge acquisition take place during searching activities where the context and purpose of the activity are known. The perception that learning automatically takes place during the execution of an information-seeking task was proven to be false.[25]

An empirical study was done by Yuan to investigate the effect of the experience level of end-user searchers on their searching behavior. The searching behavior of a group of law students was monitored over a one-year period. It was found that search experience did have an effect on the following aspects of their behavior: increase of command and feature repertoires; language-usage pattern; increase of search speeds; and change of learning approaches. However, the amount of experience had no positive effect on the number of errors being made or on recovery from those errors.[26]

Large, Tedd, and Hartley identify two basic approaches to the ideal situation of finding the exact and correct result to an online search: search-narrowing tactics when a search produces too many possible answers, and search-broadening tactics when a search produces no or few, unrelated answers.[27]

Operators

The syntax of search engines differs widely, and a wide variety of operators exists to allow focusing of a search query. Lancaster referred to natural-language searching, where no operators are used, in an early reference.[28] Some of the more commonly used operators include phrases (using quotes), inclusion operator (+), and exclusion operator (-). Other, more advanced and less-often used operators include Boolean operators (AND, OR, NOT), proximity operators (NEAR), stemming (*), and field limiters (DOMAIN, TITLE). These operators are discussed by a variety of authors.[29]

Frants and other authors claim that most retrieval systems feature Boolean searching. These authors have shown that criticism of Boolean systems is mostly directed at the methodology employed rather than the Boolean principle itself. However, they regard Boolean and other search principles as being of equal importance.[30] A full description of the usage of Boolean operators, both in general databases and Internet search engines, is given in a book on electronic information seeking.[31]

Hirsh found that during information seeking on electronic resources, children generally either do not use any Boolean operators or experience problems when they do.[32] During a 1996 task-oriented assessment using two medical searching systems, a traditional Boolean system was compared to a natural-language system. No significant differences were found in the success rate of the two, as minimal training was given in both cases.[33]

In summary, a recent study claimed to have found that "most people use few search terms, few modified queries, view few Web pages, and rarely use advanced search features."[34]

Query Formulation

In general, the best way to construct a search query during an Internet search is not obvious to the average searcher. A common approach for these searchers is to simply type one or more consecutive words into the search box. Where some search engines would add an implied OR operator between words (Infoseek), others will insert an AND (AltaVista and Google in their basic searching modes), and some might even treat it as a phrase (Looksmart). These three widely differing approaches would produce very different results that would confuse rather than enlighten an average searcher. As an example, an Internet search using five common words with AND operators between them will produce x answers from a given database. The same search using the same five terms and the same database, but using OR operators between the words, will produce y answers, where y is much greater than x.

It became clear from the literature that many users found the formulation of a search query difficult. Lancaster proved as far back as 1968 that the low quality of query formulations was the main reason for the failure of a search.[35]

During a survey of 316 Excite users, Spink, Bateman, and Jansen stated that few users employed logical operators and even fewer used the syntax correctly. They also had problems with search phrases and with the construction of good search terms and complex search queries.[36]

Kassler, however, claimed that the Internet could be mined for information that could provide a competitive advantage, if attention was given to search strategy and proper use of the available features.[37] A number of attempts were made to rectify this situation by some authors who offered an alternative to searchers constructing their own queries. Frants and Shapiro built an algorithm that automatically constructed a Boolean query. This algorithm showed promise for successful implementation, since it accepted user feedback that was used to correct query formulations.[38]

Gauch and Smith tested an expert system for online search assistance that automatically reformulated queries in an attempt to improve search results. This system decreased the number of queries required, increased precision, and improved the relevance ranking.[39]

During this project a comparison will be made between the length of keywords used and the success achieved during Internet searching.

Searching Success Rate

Lancaster discussed some factors that influence success in online searching many years before the advent of Internet search engines.[40] Lately, however, many authors agree that learners should know how to retrieve information from electronic sources.[41]

A large number of authors claim that it is difficult to find relevant information on the Internet:

- "Currently, search is simply bad."[42]
- "Some respondents seemed confused about what they were to report when asked to list query terms for their search."[43]
- "The user's ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low."[44]
- "As the World Wide Web grows exponentially, discovery and retrieval of useful educational material grows more problematic."[45]
- "Only 33 percent of the Internet users agree or strongly agree with the statement 'It is easy to perform subject searches on the Internet.'"[46]
- "I find it difficult to search information on the Internet. . ."[47]
- "Information seeking is a complex and difficult process for these students, who seek to reduce the task to finding an obvious answer or finding a good Web site. . ."[48]
- "Both novice and experienced searchers were overconfident in their performance."[49]

A large body of evidence points to the fact that most Internet searchers find it difficult to locate and retrieve relevant data on the Internet.

The Status Quo

Saracevic, pondering on fifty years of development in this field, best summarized the status quo of both information science and information retrieval.[50] Three of its general characteristics were identified:

- Information science spans various disciplines.
- Information science is inherently connected to information technology.
- Information science plays an active part in the evolution of the information society.

Ingwersen identified a number of other disciplines that had an influence on information science: mathematics and communication (with an overlap in information theory), epistemology, sociology, and linguistics (with an overlap in socio-linguistics), psychology, artificial intelligence, and computer science.[51]

Saracevic lists three fundamental powerful ideas on which information science is based: information retrieval, relevance, and interaction.[52] He thus clearly classifies information retrieval as being part of information science.

Method

An instrument to measure various aspects about a searcher's experience was designed, tested, refined, and then implemented. The form used required the participant to record a variety of elements about the search being undertaken. A total of 1,109 learners from three continents, spread over forty-six sessions and twenty institutes of higher education, participated in this study. Each group of participants was given thirty minutes to search for one academic topic of their own choice, using any program and methodology they wish. No guidance was given on search engine choice or operators.

The responses of a number of participants had to be omitted due to a variety of external factors. Reasons for these responses being classified as invalid included:

- respondents used directory drilling, not keyword searching;
- information on forms indicated impossible answer; or

- forms were incomplete, missing some vital information.

A total of 540 respondents remained, and the results on their forms were used for this study. Some standard personal data was requested on the form (age, gender, race). Another one of the fields requested that the keyword(s) used be written down; still another asked for the success of the search to be indicated with a binary answer. These values were then inspected to determine the presence or absence of relationships between them.

Results and Analysis

Of the 540 remaining respondents, 372 were male and 152 were female (sixteen participants did not specify gender).

The respondents' age distribution ranged from seventeen to fifty-eight years. These ages were classified into three age groups, the first containing respondents younger than twenty years (155); the second between twenty and twenty-five years (295); and the last one older than twenty-five years (69). There were twenty-one cases where the age of the participant was unknown.

The race distribution was: ninety-seven Asian, seventy-seven mixed race, fifty-four Black, and 295 White. The races of the other seventeen cases were not specified.

A total of 109 respondents used only one keyword in their searching tasks, and the remaining 431 used more than one keyword.

For each one of these four elements (gender, age, race, and number of keywords), an analysis of the results was done. They were then combined and statistically analyzed to determine their influence on searching to conclude this research project.

Gender—Results and Analysis

An indication of searching success categorized by gender was extracted from the results. The distribution of these results is shown in table 1.

Table 1. Distribution of Results Obtained by Respondents by Gender

	Male		Female		Total	
	No.	%	No.	%	No.	%
No Result	219	58.87	97	63.82	316	70.99
Acceptable Result	153	41.13	55	36.18	208	29.01
Total	372	100.00	152	100.00	524	100.00

For both male and female participants, the figure for those who obtained no result was higher than for those who obtained an acceptable result. However, without statistical analysis it was not quite clear whether there was any correlation between gender and the success rate.

A chi-square test of independence showed that there was no significant relationship between the gender of the participant and finding a result, using one or two keywords.

Age—Results and Analysis

The relationship between age and searching success was then evaluated. Table 2 lists the figures, which indicate the achievement of an acceptable result as a function of age.

A superficial inspection of the results seemed to indicate that the higher the age of the participant, the greater the possibility of not finding an acceptable result. Moving through the three groups from young to old, the success rate decreased from 46 percent through 39 percent to 28 percent. Further statistical analysis was required to confirm this perception.

The chi-square test for independence showed that a significant relationship existed between results obtained and the age group of the respondent (probability = 0.0343). It was thus confirmed that younger respondents were more likely to obtain acceptable results than older respondents.

Race—Results and Analysis

Next, the authors correlated searching success and the respondent's race. This relationship is highlighted in table 3.

The results of table 3 were sorted from low to high (left to right) according to the percentage figure in the Acceptable Result row. It thus appeared that Asians provided the lowest success figure (28 percent), followed by those of mixed race (35 percent), and Blacks (41 percent). Whites provided the highest success figure (45 percent). Again, it was considered necessary to confirm this result with further statistical testing.

Using a chi-square test for independence, it was determined that a significant relationship existed between the race of the respondent and the results obtained in the Internet search (probability = 0.0189). White respondents were more likely to obtain results than either Black respondents, Asian respondents, or respondents of mixed race.

Number of Keywords—Results and Analysis

The fourth and final comparison, the number of keywords used by respondents, is summarized in table 4.

A cursory analysis seemed to indicate a large difference in success rate between users who specified only one keyword (low success rate) and those using two or more keywords. More specifically, obtaining an acceptable result was 5.5 times more likely when two or more keywords were used.

A chi-square test for independence showed that there was a significant relationship between the number of keywords used in an Internet search and obtaining results on the search (probability = 0.0228).

Table 2. Distribution of Results Obtained by Respondents by Age

	Age 0–19		Age 20–25		Age 25+		Total	
	No.	%	No.	%	No.	%	No.	%
No Result	84	54.19	180	61.02	50	72.46	325	60.19
Acceptable Result	71	45.81	115	38.98	19	27.54	215	39.81
Total	155	100.00	295	100.00	69	100.00	519	100.00

Table 3. Distribution of Results Obtained by Respondents according to Race

	Asian		Mixed race		Black		White		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
No Result	70	72.16	50	64.94	32	59.26	162	54.92	314	60.04
Acceptable Result	27	27.84	27	35.06	22	40.74	133	45.08	209	39.96
Total	97	100.00	77	100.00	54	100.00	295	100.00	523	100.00

Combined Analysis: Gender, Age, Race, and Number of Keywords

The authors then combined the four elements in an attempt to reach a conclusion. Table 5 represents the results of a binary logistic regression that was fitted to the data to determine which variables used together contributed significantly to the obtaining of results in an Internet search. A stepwise variable-selection procedure showed that the number of keywords, age, and race group of the respondents best determined whether a result would be obtained in an Internet search.

Table 4. Distribution of Results Obtained by Respondents by Choice of Keywords

	One Keyword		Two or More Keywords		Total	
	No.	%	No.	%	No.	%
No Result	76	14.07	249	46.11	325	60.18
Acceptable Result	33	6.11	182	33.77	215	39.82
Total	109	20.18	431	79.88	540	100.00

Table 5. Combined Statistical Results

Effect	DF	Wald Chi-Square	Pr > Chi-Square
Age	1	4.3400	0.0372
Race	3	9.6083	0.0222
Keywords	1	4.0727	0.0436

Conclusion and Recommendations

It became clear that demographic features of Internet searchers do have an effect on the outcome of a search operation. It appears as if race, age, and number of keywords have a statistically measurable influence on the outcome, while gender does not. The number of keywords has the strongest effect (more keywords produce greater success rate), followed by age (younger produces a greater success rate) and race (white users produce a greater success rate than others).

Many studies have been done on gender issues, but none could be found that concentrate on the effect gender has on differences in Internet-search success. This provides one possible area for future research.

The noticeable difference in performance between different races comes as no surprise. Since most of the participants in this study were from South Africa, the effects of the history of racial segregation can probably be blamed for this clear difference. Although the political landscape is changing, the participants of this study still carry the results of the apartheid approach in their background and lack of proper education.

The difference in age-group figures seems to indicate that younger users are more adept at using technology to

provide answers. The exposure younger users have had at an earlier age to computing—and, specifically, the Internet—is the most likely factor contributing to this scenario.

Prior reading has highlighted the fact that researchers working on the Uniterm project found that indexing using a single term produces better results than with multiple terms.[53] The result for the number of keywords appears to be in contrast with early research work done on this system. However, the Uniterm system has to do with the way information is indexed, while this project concentrated on the way information is retrieved.

The reason for this difference lies in the richness of the information found on the Internet, which succeeded the Uniterm system by many years. A quick experiment with single-term searching on Google produced a staggering 41,400,000 answers to the term "weather," and 96,400,000 to "research." Even a slightly more focused search using the phrase "information technology" still yielded 3,580,000 answers. These figures indicate that single-term searching on the Internet does not appear to be a fruitful approach. Future research in this area could focus on the success of using unique, difficult-to-spell single terms as a further refinement to single-term searching. As an example, during a Google search, "derailleur" and "anorexia" each produce only 98,000 and 550,000 answers respectively.

Finally, it was interesting to note that age, race, and keyword length as factors produced not entirely surprising results in terms of searching success. However, the fact that gender does not determine the search-success rate to a large degree is surprising and could be the topic of further research.

References

1. R. Cover, *Electronic Text Corpus of Sumerian Literature.* Accessed Apr. 10, 2001, www.oasis-open.org/cover/etcsl.html.
2. Greek Libraries, *Libraries.* Accessed May 11, 2001, www.libraries.gr/en_messop.htm.
3. P. Ingwersen, *Information Retrieval Interaction* (London: Taylor Graham, 1992).
4. C. W. Cleverdon, J. Mills, and E. M. Keen, *Factors Determining the Performance of Indexing Systems,* vol 1. *Design.* Part 1 (Cranfield: Aslib Cranfield Research Project; 1966); S. E. Robertson, *Conflicting Philosophies.* Accessed Apr. 10, 2001, www.soi.city.ac.uk/research/cisr/ser/ucla/node3.html; Y. A. Tonta, *Failure Analysis in Document Retrieval Systems: A Critical Review of Studies.* Accessed Apr. 18, 2001, http://yunus.hacettepe.edu.tr/~tonta/yayinlar/phd/bolum-3.htm.
5. D. Ellis, *Progress and Problems in Information Retrieval* (London: Library Assoc. Pub., 1996).
6. Ibid.
7. Ibid.
8. Ibid.
9. Tonta, *Failure Analysis in Document Retrieval Systems: A Critical Review of Studies.*
10. R. Fidel et al., "A Visit to the Information Mall: Web-Searching Behavior of High-School Students," *Journal of the American Society for Information Science* 50, no. 1 (1999): 24–37.
11. S. G. Hirsh, "Children's Relevance Criteria and Information Seeking on Electronic Resources," *Journal of the American Society for Information Science* 50, no. 14 (1999): 1265–83.
12. D. J. Grimes and C. H. Boening, "Worries with the Web: A Look at Student Use of Web Resources," *College and Research Libraries* 62, no. 1 (2001): 11–23.
13. R. Marcella and G. Baxter, "The Information Needs and the Information-Seeking Behavior of a National Sample of the Population in the United Kingdom, with Special Reference to Needs Related to Citizenship," *Journal of Documentation* 55, no. 2 (1999): 159–83.
14. A. Large, and J. Beheshti, "The Web As a Classroom Resource: Reactions from the Users," *Journal of the American Society for Information Science* 51, no. 12 (2000): 1069–80.
15. N. C. M. Ross and D. Wolfram, "End-User Searching on the Internet: An Analysis of Term Pair Topics Submitted to the Excite Search Engine," *Journal of the American Society for Information Science* 51, no. 10 (2000): 949–58.
16. F. W. Lancaster, *Information Retrieval Systems: Characteristics, Testing, and Evaluation* (New York: Wiley, 1978).
17. G. Marchionini, *Information Seeking in Electronic Environments* (New York: Cambridge Univ. Pr., 1995).
18. R. Fidel, "Towards Expert Systems for the Selection of Search Keys," *Journal of the American Society for Information Science* 37, no. 1 (1986): 37–44.

19. R. Fidel, "Searchers' Selection of Search Keys: I. The Selection Routine," *Journal of the American Society for Information Science* 42, no. 7 (1991): 490–500.

20. R. Fidel, "Searchers' Selection of Search Keys: II. Controlled Vocabulary or Free-Text Searching," *Journal of the American Society for Information Science* 42, no. 7 (1991): 501–14.

21. R. Fidel, "Searchers' Selection of Search Keys: III. Searching Styles," *Journal of the American Society for Information Science* 42, no. 7 (1991): 515–27.

22. R. Fidel, *Constructing a Pattern Model of Man-Machine Interaction in Online Literature Searching*. Accessed Feb. 22, 2001, www.clis.umd.edu/research/dissertations/fidel.html.

23. M. Weideman, "Internet Searching as a Study Aid for Information Technology and Information Systems Learners at a Tertiary Level" (Ph.D. diss., University of Cape Town, 2001).

24. C. C. Kuhlthau, "Inside the Search Process: Information Seeking from the User's Perspective," *Journal of the American Society for Information Science* 42, no. 5 (1991): 361–71.

25. R. Oliver and H. Oliver, "Using Context to Promote Learning from Information-Seeking Tasks," *Journal of the American Society for Information Science* 48, no. 6 (1997): 519–26.

26. W. Yuan, "End-User Searching Behavior in Information Retrieval: A Longitudinal Study," *Journal of the American Society for Information Science* 48, no. 3 (1997): 218–34.

27. A. Large, L. A. Tedd, and R. J. Hartley, *Information Seeking in the Online Age: Principles and Practice* (London: Bowker-Saur, 1999).

28. Lancaster, *Information Retrieval Systems*.

29. R. Boulton, *An Introduction to Information Retrieval*. Accessed Apr. 10, 2001, www.omsee.com/developer/docs/intro_ir.html; D. P. Habib and R. L. Balliot, *How to Search the World Wide Web: A Tutorial for Beginners and Non-Experts*. Accessed Mar. 6, 1999, www.ultranet.com/~egrlib/totur.htm; R. Hock, "Web Search Engines: Features and Commands," *Online* (May/June, 1999): 24–28; G. R. Notess, "Internet Search Techniques and Strategies," *Online* (July/Aug., 1997): 63–66; G. R. Notess, "On-the-Fly Search-Engine Analysis," *Online* (Sept./Oct., 1999): 63–66; D. Sullivan, "Crawling Under the Hood," *Online* (May/June, 1999): 30–38; and M. Weideman, "The Internet Search-Engine FAQ-Tips Page." Accessed Nov. 4, 2003, www.mwe.co.za/seafaqtips.htm.

30. V. I. Frants et al., "Boolean Search: Current State and Perspectives," *Journal of the American Society for Information Science* 50, no. 1 (1999): 86–95.

31. Large, Tedd, and Hartley, *Information Seeking in the Online Age*.

32. Hirsh, "Children's Relevance Criteria and Information Seeking on Electronic Resources."

33. W. H. Hersh, J. Pentecost, and D. Hickam, "A Task-Oriented Approach to Information Retrieval Evaluation," *Journal of the American Society for Information Science* 47, no. 1 (1996): 50–56.

34. A. Spink et al., "Searching the Web: The Public and Their Queries," *Journal of the American Society for Information Science and Technology* 52, no. 3 (2001): 226–34.

35. Lancaster, *Information Retrieval Systems: Characteristics, Testing, and Evaluation*."

36. A. Spink, J. Bateman, and B. J. Jansen, "Searching the Web: A Survey of Excite Users," *Internet Research: Electronic Networking Applications and Policy* 9, no. 2 (1999): 117.

37. H. S. Kassler, "Mining the Internet for Competitive Intelligence," *Online* (Sept./Oct., 1997): 34–45.

38. Frants et al., "Boolean Search: Current State and Perspectives."

39. S. Gauch and J. B. Smith, "An Expert System for Automatic Query Reformation," *Journal of the American Society for Information Science* 44, no. 3 (1993): 124–36.

40. Lancaster, *Information Retrieval Systems*.

41. J. C. Cronje and P. A. Clarke, "Teaching 'Teaching on the Internet' on the Internet," *South African Journal of Higher Education* 13, no. 1 (1999): 213–26; K. De Jager and Y. Sayed, "Aspects of Information Literacy at Five Institutions of Higher Education in the Western Cape," *South African Journal of Higher Education* 12, no. 2 (1998): 197–203; R. J. Edling, "Information Technology in the Classroom: Experiences and Recommendations," *Campus-Wide Information Systems* 17, no. 1 (2000): 10–15; S. L.Gan, "An Overview of Information Technology and Education in Malaysia," *Journal of Global Information Management* 6, no. 1 (1998): 27–32; P. J. J. G. Marais and A. F. Marais, "Science Education for the New Millennium: An Exciting Challenge," *South African Journal of Higher Education* 13, no. 2 (1999): 82–89; P. Wong, "Leapfrogging across the Millennium: Information Technology in Singapore Schools," *Journal of Global Information Management* 6, no. 1 (1998): 5–13.

42. C. Sherman, "The Future of Web Search," *Online* (May/June, 1999): 54.

43. Spink, Bateman, and Jansen, "Searching the Web: A Survey of Excite Users," 122.

44. Ibid., 125.

45. S. A. Sutton, "Conceptual Design and Development of a Metadata Framework for Educational Resources on the Internet," *Journal of the American Society for Information Science* 50, no. 13 (1999): 1191.

46. H. J. Voorbij, "Searching Scientific Information on the Internet: A Dutch Academic User Survey," *Journal of the American Society for Information Science* 50, no. 7 (1999): 604.

47. Ibid., 605.

48. R. M. Wallace, J. Kupperman, and J. Krajcik, "Science on the Web: Students Online in a Sixth-Grade Classroom," *Journal of the Learning Sciences* 9, no. 1 (2000): 75.

49. D. Wolfram and A. Dimitroff, "Preliminary Findings on Searcher Performance and Perceptions of Performance in a Hypertext Bibliographic Retrieval System," *Journal of the American Society for Information Science* 48, no. 12 (1997): 1145.

50. T. Saracevic, "Information Science," *Journal of the American Society for Information Science* 50, no. 12 (1999): 1051–63.

51. P. Ingwersen, *Information Retrieval Interaction*.

52. T. Saracevic, "Information Science," 1052.

53. S. E. Robertson, *Conflicting Philosophies*; Y. A. Tonta, *Failure Analysis in Document Retrieval Systems*; D. Ellis, *Progress and Problems in Information Retrieval*; Cleverdon, Mills, and Keen, *Factors Determining the Performance of Indexing Systems*.