

Using Keywords to Improve the Display of a Search Results List

Offer Drori

The Hebrew University of Jerusalem, Jerusalem, Israel

offerd@cc.huji.ac.il

Abstract

This article presents the results of a comparative study of various user interfaces used to obtain search results from textual databases. This study is one of a series carried out at the Hebrew University of Jerusalem. It focuses on the display of text elements in search results and their impact on user behavior. The findings indicate several distinct benefits in the interface developed for the experiment, in which the search results display consists of the document title, keywords, and several lines from the text that fulfill the search criteria. These user benefits include ease of use, increased confidence that the parameters defined for the search query will yield the correct results and allow successful completion of the task, and satisfaction with the relevance of the information displayed.

Keywords: search results, using keywords, displaying list, information retrieval

Introduction

Problem and Background

Information retrieval in a text retrieval system includes, in most cases, a definition of the search strategy. The search result presents a list of existing documents that fulfill the search strategy criteria (Blair & Maron, 1985). The search results generally include a list of documents, with larger databases yielding a greater number of documents that fulfill the search criteria (Drori, 1999) (Sievert, 1996). The Internet, as it exists today, is a good example of a large, unfocused information database. Several search engines have been developed in order to cull information from the Internet, each of which utilizes a proprietary method to collect material, based on searching for terms that appear in documents listed in the database (Chowdhury, 1999) (Regan, 1998). Most search engines use similar interfaces to display the search results, generally in the form of a list of documents that fulfill the search criteria with an attempt to partially rank the results. The display usually includes the document title, its Internet address, and frequently includes the first few lines of the text. Most search operations on today's Internet search engines yield search results lists of hundreds and even thousands of documents, although the average user tends to examine only the first 10-20 results (Kirsh, 1998). In this state of affairs, the user seeking information may bypass some search results that could be crucial to successful completion of the task. A highly-motivated user, on the

other hand, may devote substantial time to reading extensive sections of numerous documents that appeared in the search results before determining that they are not relevant to completing the task. The purpose of this paper is to address the issue of the display of the list of search results. A brief review of previous studies will be presented, after which the discussion will turn to the results of a compara-

Using Keywords

tive study that examined various interfaces for performing the operations described above.

Attempts at a Solution in Previous Studies

There are currently two accepted methods for displaying a list of search results from textual databases in web search engines. The first is to display the list of document titles that fulfill the search criteria, and the second is to display the document titles accompanied by the first few lines of the text. The use of document titles assumes that the document title is the best means of disclosing the contents of the document. Studies show that this is not the case in reality. In her 1984 study of journals in the field of library science and information, Peritz showed that in 21% of the cases the article title was found to be not informative (Peritz 1984). The concept of expanding the displayed search results by adding the first lines of the document under the title was intended to address this shortcoming. It provided the user with additional information that hinted at the contents of the document, thus allowing an assessment of its relevance.

Various field studies examined other interfaces that could be used to display search results from textual databases. Display of the results in a number of hierarchical levels was carried out using Venn diagrams (Hertzum & Frokjaer, 1996). The findings demonstrated that users processed information and worked more quickly when the display listed the titles than when the interface was in the form of a Venn diagram.

In Project Superbook, on the other hand, the interface was designed as an electronic book (Egan et al., 1989). It displays three vertical windows: one provides the table of contents, the second a search definition, and the third contains the texts of the documents. For most search tasks, using the Superbook was more effective than consulting the same book in print. Another interface, using the Scatter/Gather Browsing method, simultaneously displays ten windows of information. The windows themselves constitute a cluster of collections, each of which includes several documents that have significant terms in common. The titles of the first three documents are displayed under the words. The user can opt to expand the information (scatter) or contract it (gather) by means of an appropriate command (Pirolli et al., 1996) (Pitkow & Pirolli, 1996). This system has advantages for browsing a large unfamiliar document database.

With the TileBars method, the search results are displayed in the form of graphics: Several squares and colors displayed with each document in the results list indicate the degree to which the document is relevant to the search query (Hearst, 1995). Another method, the incremental search, displays a list of documents, particularly a list of words or subjects retrieved from several databases. This method displays the list of documents that fulfill the search criteria in an immediate response as the user presses additional letters.

Concurrent with the methods described above, and on the assumption that human knowledge is, for the most part, textual, this writer has designed and tested a third interface to display a search results list of documents from databases. In addition to the document title, this interface displays several lines, which may appear at any point throughout the text, that fulfill the search criteria (Drori, 1998). This interface relies on the fact that articles in most contemporary databases do not have keywords defined in advance. In other words, most texts are not scientific documents that have been subjected to an accepted control process. Based on this assumption, the information that should be displayed for the user is the information in the document that is most relevant to the task, and not merely information displayed because of its fixed place in the document (i.e., the first several lines of text). For this reason, it may be presumed that the most relevant location in the body of the document is the paragraph or lines that fulfill the search criteria, which are the reason for inclusion of the document in the list of search results. A study of this subject carried out at the Hebrew University of Jerusalem found that displaying lines from the text that fulfill the search criteria has advantages over a display limited to the first lines of the document (Drori, 2000). For this study, the keywords in the document were provided in addition to the title of the document and the relevant lines that fulfilled the search criteria. A description of the study, which examined this hy-

pothesis (the best display interface for search results includes titles, relevant lines, and keywords) and its results, is the principal subject of this article.

Questions in this Study

The following questions of the study are discussed in this paper:

1. What is the most important information* to display in the list of search results?
2. In comparing various methods of displaying information in a list of search results, which method is preferred by the user?

* “Important information” refers to information that contributes to an efficient search strategy, i.e., ease of use, increasing confidence as the search progresses that the search criteria as defined will yield the correct response, time required to successfully complete task, etc.

Methodology

Experiment and Sample Environment

The experiment was carried out testing a number of defined problems in an academic environment using a sample of 51 end users. Each participant was asked to obtain the correct answers for each of three questions that required conducting a search in an English-language database. The questions were: 1. What was the date of birth of the British physicist Newton? 2. What price did the United States pay Russia for the

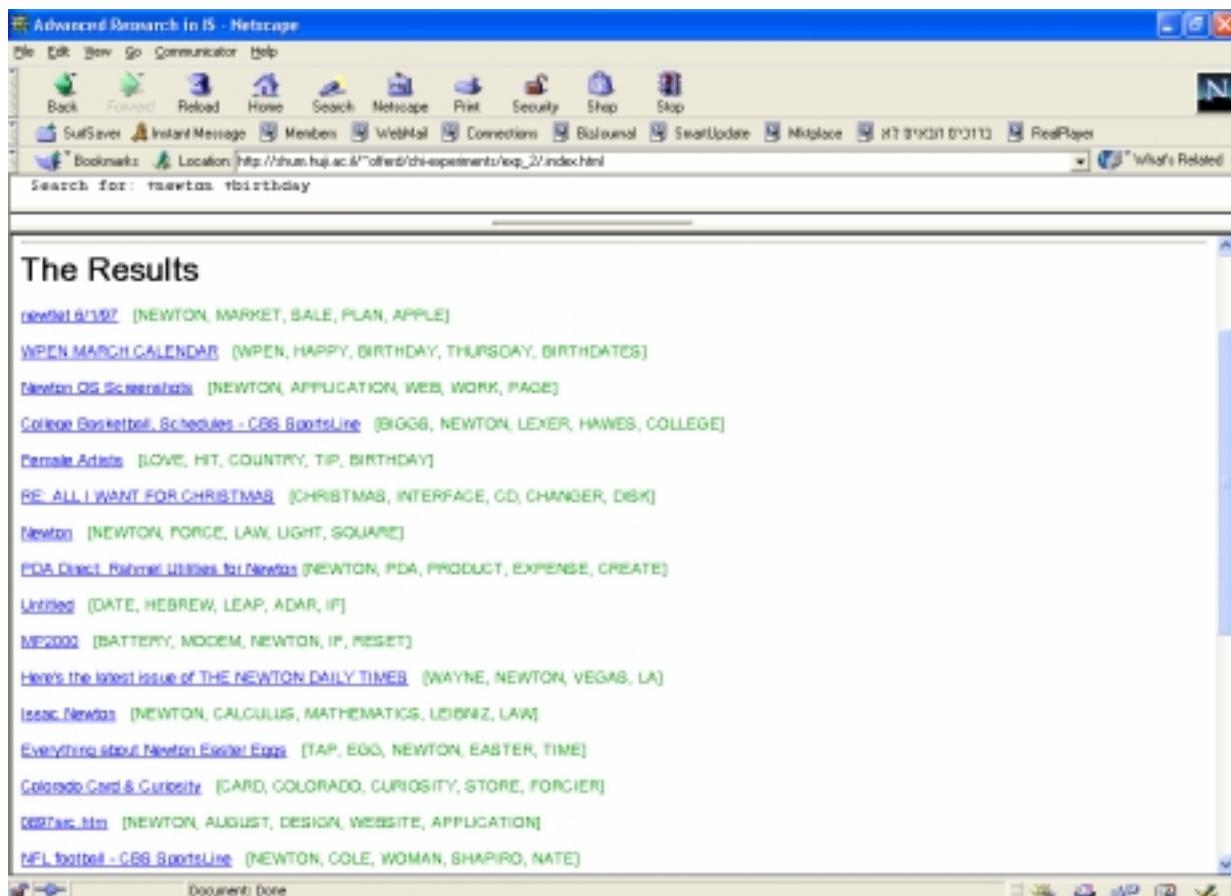


Figure 1: Snapshot "A" interface
("Titles" in blue, "Keywords" in green)

Using Keywords

Alaska Territory? 3. Find the telephone number of a florist in Atlanta who offers delivery services. The layout of the experiment included building three information databases, with each database containing 30 English-language documents for a given subject. All of the documents in the database were collected from the Internet and their contents were relevant to the search task. Use of the information database was intended to avoid distractions in the course of the experiment, and to provide each participant in the experiment access to the same number of identical documents for each search task. Each task yielded the same number of responses (30). The search tasks used three types of user interfaces. All participants were experienced in using all the interfaces, and they each executed all three tasks using all three interfaces. The participants had similar backgrounds and similar levels of experience with computer systems and Internet search engines (one group was composed of MBA students and another was composed of students studying for an MLS in information science). The average age of the participants was 31 years.

Types of Information and Methods of Retrieval

Three types of interfaces were used. The study compared the manner in which the textual database search results were displayed. Two interfaces resembled search engines commonly found on the Internet (although they do not generally add keywords as used in this study).

The first interface tested in the experiment, Method A, displayed a list of document titles and keywords (see figure 1).

The second interface tested in the experiment, Method B, displayed a list of document titles, the first lines

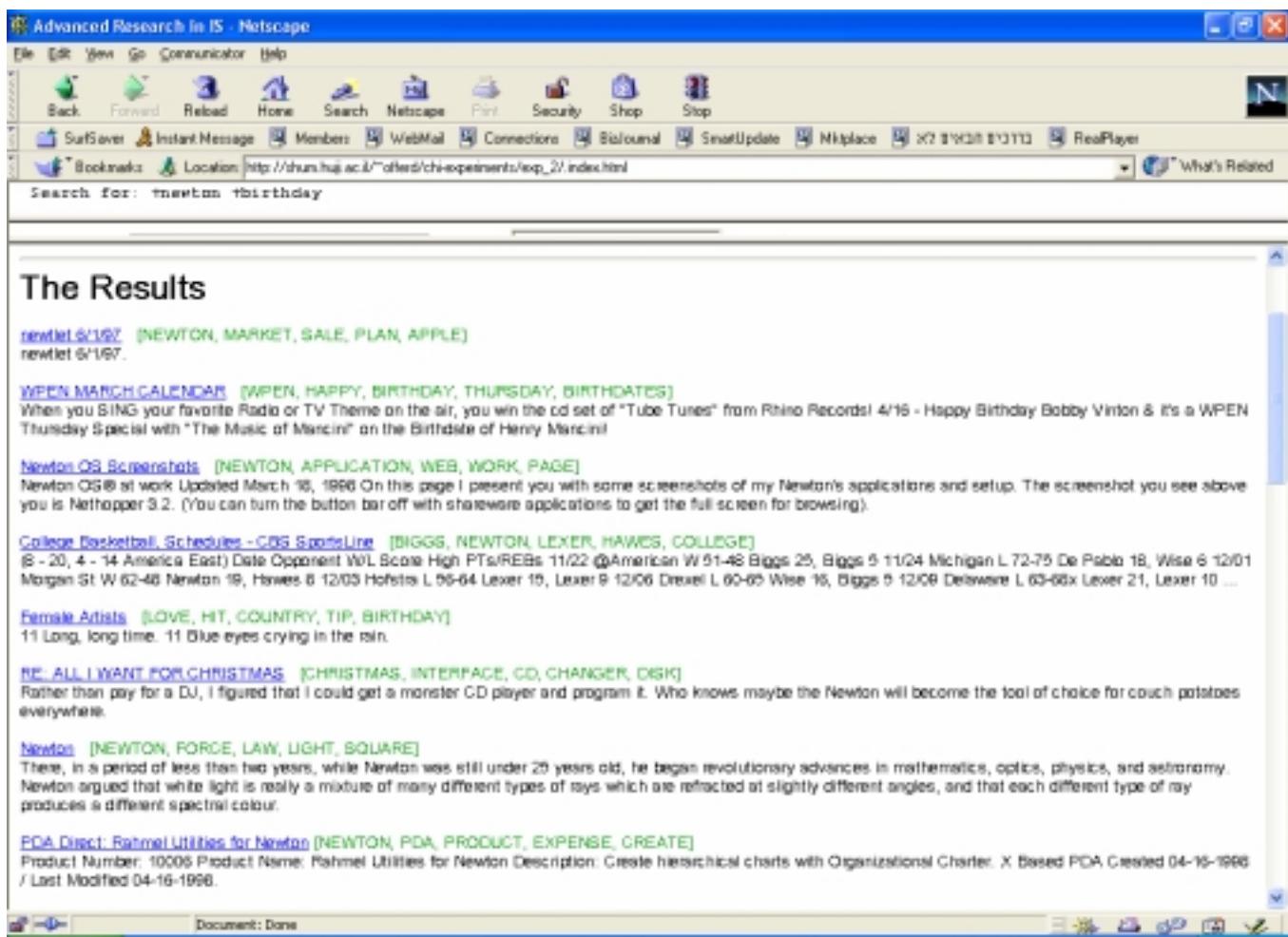


Figure 2: Snapshot "B" interface

(“Titles” in blue, “Keywords” in green, “First lines” in black, “Search terms” in red)

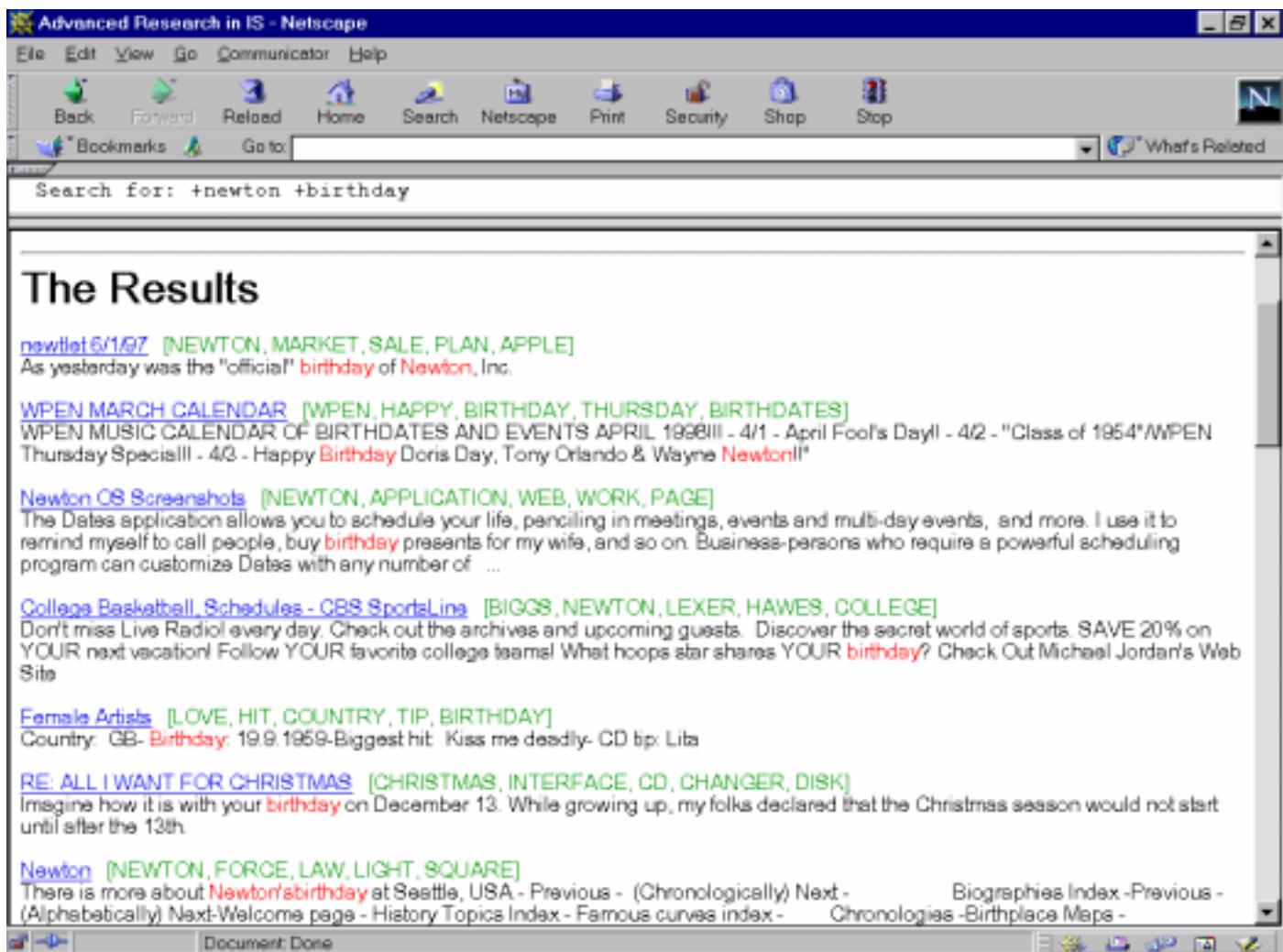


Figure 3: Snapshot "C" interface

(“Titles” in blue, “Keywords” in green, “Lines in context” in black, “Search terms” in red)

of the document, and keywords (see figure 2).

The third interface, Method C, was developed specifically for the experiment. This interface displayed document titles, several lines from the body of the text that fulfilled the search criteria, and keywords. The query terms were highlighted (see figure 3).

The three interface methods compared several subjects of interest to the researchers: the time required for each interface to perform the search tasks, how participants felt about each interface method in terms of ease of use, satisfaction with the results obtained, increasing confidence as the search progressed that the search criteria they defined for the search query would yield the correct response, and the degree to which the respective interface method displayed information that was relevant to the search task.

Information gathered was based on two sources, objective and subjective. The objective information, regarding search time and the accuracy of the result, was measured by computer. The subjective information was obtained from a computer feedback sheet that each participant was required to complete at the end of the experiment, which recorded the participant's sense of satisfaction and other data. Table 1 presents the criteria examined in the experiment, and evaluation ratings that were used by the participants. There are several methods to examine subjective feelings that concern users of an information retrieval system (Doll et al., 1994) (Meadow et al., 1995) (Ellis et al., 1995) (Tague-Sutcliffe, 1996) (Dunlop,

Using Keywords

2000). We could not, however, find a questionnaire that was exactly suited to our needs. A questionnaire was specially developed and tested it in a pre-test experiment.

Subject Tested for each Interface	Table of Values for the Response
Ease of use	1-10 (1-Not at all easy 10-Very easy)
Preference for a specific search engine interface	1-10 (1-No preference 10-High preference)
Increasing user confidence as the search progressed that the search criteria as defined would yield the correct response	1-4 (1-Lack of confidence 10-Very confident)
Relevance of search results	1-4 (1-Not relevant 10-Very relevant)
Degree to which search results were perceived as misleading (i.e., continuing the search as it was defined would not yield a correct response)	1-4 (1-Not misleading 10-Very misleading)
Contribution of keywords to search effectiveness	1-10 (1-Low contribution 10-Very high contribution)
Time required to retrieve document containing correct response	Measured in seconds
Correctness of answer	User's selected response verified by examiner

Table 1: List of Criteria and Their Values

Methods of Analyzing Data

Several tools were used to conduct a statistical analysis of the data: the Pearson test, the Wilcoxon test for pairs, and the McNemar test (SAS/STAT, 1990) for symmetry of the matrix. Most of the results obtained were absolutely high and showed a clear relation among the variables tested. The final results related to the comparison of data among the various methods. Since three interface methods were examined, and since the opinion of the same user was rendered with respect to each of the interface methods, it was possible to perform a comparative analysis between pairs of methods to determine relations between them. An analysis of the data first referred to the correctness of the response, and then to the time required for the user to retrieve it. In addition, answers to questions on the feedback sheets were analyzed.

Course of the Experiment

All of the participants received a sheet listing the three questions that comprised the search task. One document in the database contained the correct answer to each question. Participants used the search engine at the experiment site and retrieved a list of responses that fulfilled the search criteria.

Results

Test results for each test and a discussion of the findings are presented in this section.

Ease of Search

This category tested the degree to which each participant felt at ease performing the search task.

Methods	N	Mean	Std. Error	P (for paired T-Test)
C vs. B	51	1.07	0.23	<0.0001
C vs. A	51	1.33	0.33	<0.0001

Table 2: Ease of Use – Comparison of Method C with Methods A and B

Table 2 demonstrates that 13% of the participants preferred Method C over Method A, and 11% of the participants preferred Method C over Method B.

Preference for Specific Interface Method

This category examined the degree to which each participant preferred that Internet search engines be designed using a specific interface method to display search results.

Methods	N	Mean	Std. Error	P (for paired T-Test)
C vs. B	51	1.50	0.18	<0.0001
C vs. A	51	2.39	0.29	<0.0001

Table 3: Preference for Specific Interface Method: Comparison of Method C with Methods A and B

Table 3 demonstrates that participants chose Method C as preferable both to Method A (24%) and to Method B (15%).

User Confidence in the Defined Parameters of the Search Query

This category examined the degree to which each participant felt increasingly confident, as the search progressed, that the search criteria as defined would yield a correct response to the question. Tables 4, 5, and 6 demonstrate preferences for the compared methods.

Table 4 compares Methods B and C, and Table 5 compares Methods A and C. Table 6 compares Methods A and B. In the McNemar test, the values of the oblique axis designate symmetry. The values in the lower left triangle designate the preference for Method C, while the values in the upper right triangle designate the preference for Method B (Table 4). The data in the tables is presented in percentages. The reported values ranged from 1 to 4, where 1=lack of confidence and 4=very confident.

		Method B				
		1	2	3	4	
		1	0	1.96	0	0
Method		2	0	5.88	1.96	3.92
C		3	0	0	13.73	3.92
		4	1.96	1.96	60.78	3.92

Table 4: Confidence – Symmetry of Matrix in Method C vs. Method B

Using Keywords

Table 4 demonstrates that 65% of the participants felt significantly increasing confidence as the search progressed that the defined search criteria would yield the correct answer when using Method C, compared with only 8% who felt confident that the defined search query would yield the correct answer when using Method B ($P<0.001$).

Method A					
	1	2	3	4	
Method	1	0	0	1.96	0
C	2	0	7.84	1.96	1.96
	3	1.96	3.92	5.88	5.88
	4	1.96	58.82	3.92	3.92

Table 5: Confidence – Symmetry of Matrix in Method C vs. Method A

The comparison of Method C and Method A resulted in similar findings. As Table 5 demonstrates, here, too, 65% of the participants felt very confident as the search progressed that the defined search criteria would yield the correct answer when using Method C, compared with 8% who felt similar confidence when using Method A ($P<0.001$).

Method B					
	1	2	3	4	
Method	1	1.96	0	1.96	0
A	2	0	5.88	3.92	0
	3	0	3.92	11.76	1.96
	4	0	3.92	60.78	3.92

Table 6: Confidence – Symmetry of Matrix in Method B vs. Method A

The comparison of Method A and Method B, illustrated in Table 6, demonstrates that 69% of the participants felt confident when using Method B that the defined search query would yield the correct answer, compared with only 6% of the participants who felt similarly confident when using Method A ($P<0.001$).

Degree of Relevance of Information Accompanying the Titles to the Search Query

This category tested the degree to which the information accompanying the document title was relevant to the search query. The McNemar test was also applied here. Data is presented in percentages.

Method B					
		1	2	3	4
	1	1.96	0	1.96	0
Method	2	0	5.88	3.92	0
C	3	0	3.92	11.76	1.96
	4	0	3.92	60.78	3.92

Table 7: Relevance – Symmetry of Matrix, Method C vs. Method B

Table 7 demonstrates that 65% of the participants felt that the supplementary information displayed in the list of responses arranged using the Method C interface was very relevant to the search query. In practice, this meant that the participants were successful in obtaining the correct answer to the question. In contrast, only 2% of the participants felt that the supplementary information displayed in the list of responses displayed in the Method B interface was relevant to the search query ($P<0.001$). Method A did not display additional information to accompany the document title and is therefore not included in this analysis.

Contribution of Keywords to Search Effectiveness

This category examined the degree to which keywords play a role in the effectiveness of the search.

Methods	N	Mean	Std. Error	P (for paired T-Test)
C vs. B	51	1.09	0.21	0.0001
C vs. A	51	0.80	0.27	0.0053

Table 8: Keywords - Difference between Method C and Methods A and B.

Table 8 demonstrates that Method C has an 11% advantage over Method B and an 8% advantage over Method A.

Time to Obtain Correct Answer for Various Tasks

Table 9 presents the time, in seconds, that was required to obtain correct answers after filtering out incorrect answers. Data relevant to each search task was collected for each of the interface methods. Because the various tasks differed considerably from each other in levels of difficulty, they could not be compared.

Task (name of the task)	N	Method A	Method B	Method C	Average Time for Task (sec.)
1 (F)	21	116	124	84	108
2 (N)	22	50	81	105	79
3 (A)	24	149	146	144	146

Table 9: Time Required to Obtain Answer Using Different Interface Methods, by Task

Analyzing the average time required to obtain the right answer indicates the degree of difficulty of the task itself. The results of our analysis demonstrate that the second task (finding the purchase price of the Alaska Territory) was ranked as the easiest, the first task (finding Newton's date of birth) was ranked as

Using Keywords

intermediate, and the third task (finding the telephone number of an Atlanta florist who offers delivery services) was the most difficult.

Table 10 presents the relative advantage of each method. Method A was the fastest interface for the easiest task. Method C was the fastest interface for both the intermediate and the most difficult tasks. There was only a negligible difference in the search times for Methods A and B for the most difficult task.

Task	Average time for task (sec.)	Degree of difficulty according to time	Percentage of wrong answers	Degree of difficulty according to wrong answers
1	108	Intermediate	8.79	Intermediate
2	79	Easy	6.19	Easy
3	146	Intermediate - Difficult	18.18	Hard

Table 10: Degree of Difficulty of the Task, Measured by Times and Percentage of Incorrect Answers

Discussion

This section discusses the conclusions that can be drawn from the results of this study and from analysis of the different categories. The conclusions are presented in the same order in which the task questions were introduced above.

First Question

"What is the most important information to display in a list of search results?"

The findings demonstrate a clear preference for the document title (displayed in all the interfaces) accompanied by several lines of the document that fulfill the search criteria (Method C). This method of displaying the search results was clearly preferred to the interface that displayed the first few lines of the document (Method B). Moreover, since the accompanying text was identified because keywords used in the search query were included in the document listing in the database, we can conclude that keywords added to the titles of documents listed in a database contribute to the effectiveness of the search. Users felt at ease using Method C, as demonstrated by increasing confidence as the search progressed that the defined search criteria would yield the correct response, and they were satisfied with the relevance of the information. Regarding the time required to find information, the findings demonstrate that Method C was preferred when the search task was at an intermediate level of difficulty, such as task 1 in the experiment (finding Newton's date of birth). The other interface methods were more efficient for simple tasks, such as task 2 (finding the purchase price of the Alaska Territory). For complex tasks, such as task 3 (finding the number of an Atlanta florist who offers delivery services), Method C had no advantage over Method B.

Second Question

When comparing various interfaces for displaying information in a list of search results, which method does the user prefer?

Three display interfaces were tested: Method A – document titles and keywords; Method B – document titles, keywords, and the first few lines of the document; Method C – document titles, keywords, and several lines from the text that fulfilled the search criteria. The study found that:

1. Users felt more at ease in the course of the search when using interface Method C, in comparison with Methods A and Method B.
2. Users indicated a preference for search engines that used interface Method C to display search results over search engines that used Methods A or B.
3. Method C gave users more confidence as the search progressed that the defined search criteria would yield the correct response than they felt with Method A or Method B. A related experiment found that using Method B gave users more confidence than they felt using Method A. The significance of this finding is that if the choice is limited to these two interfaces, then Method B is preferred, although not as significantly as the preference for Method C over Method B.
4. The information accompanying the title as displayed in the Method C interface was found to be significantly more relevant than the information accompanying the title in Method B. We can conclude, therefore, that displaying several lines from the body of the text that fulfill the search criteria is deemed more relevant than merely displaying the first few lines of the document.
5. The experiment found that the addition of keywords to the displayed results increased search effectiveness by 10% for Method C, in comparison with Methods A and B.
6. The findings with respect to the time required to obtain the correct answer were more ambiguous, because measuring search times was a more complex undertaking. Each search task was concluded at a different time and the average times for finding answers were different for the various interface methods, indicating that each task benefits from a different interface method. The level of difficulty of each task was analyzed (Table 10) to account for the difference in the time required to obtain the correct answer. The average times were significantly different for each task: 79 seconds (task 2), 108 seconds (task 1), and 146 seconds (task 3). The time required for each task indicates its level of difficulty, since the database and the number of documents searched were identical for every task.

Another measure to ascertain the level of difficulty of each task was obtained by analyzing the percentage of incorrect answers obtained for each query. The findings demonstrated that 18.18% of the results for task 3 were incorrect, *vis à vis* 8.79% of incorrect results for task 1 and 6.19% incorrect results for task 2. Assuming that the number of incorrect results indicates the level of task difficulty, the findings indicated a range of levels of difficulty of the respective task.

The experiment found that, based on the time required to obtain the correct answer, each interface method was distinctly preferable for the different levels of difficulty of the respective tasks. Method A was preferred for easy tasks (i.e., task 2 – finding the purchase price of the Alaska Territory). Method C was preferable for tasks of intermediate difficulty (i.e., task 1 – finding Newton's birthday). We assume that the fact that the search time using Method C was longer for easy tasks is a consequence of the extent of information that is displayed in this interface method (document title, keywords, and several lines from the body of the text that fulfill the search criteria). The title of the document displayed sufficient information to successfully complete the easy tasks, and therefore Method A, which displayed only the title and keywords, was found most effective for such queries. Method C required users to read several lines from the text in addition to the title, resulting in a longer search time than needed to successfully complete an easy task.

Conclusion

As we saw in this and in the previous study (Drori, 2000), the first lines of the document are not necessarily a faithful reflection of the entire document. We also saw that in most cases the document contains other lines that better reflect its overall contents, and these are the lines that contain keywords which were

Using Keywords

defined in the search query. In addition, this study shed light on the extent to which the addition of keywords to document titles listed in a database contributes to the effectiveness of searches of that database. Method C demonstrated that a display combining keywords and relevant lines from the body of the text yields the best search results. Display of the relevant lines helps the user complete the specified task, and the keywords help the user assess the text in general. We can thus infer that the display of keywords, even if they were not included in the search query or the document, allow an initial assessment of the relevance of the search result. In addition, the display of relevant lines that fulfill the search criteria allow the user to refine the search. This finding makes it clear that writers and database developers would be well-advised to add keywords to texts listed in databases. Existing algorithms can be used to add keywords to texts that were not accompanied by keywords at the outset (Rijsbergen, 1979) (Salton, 1989) (Spark, 1971).

Implementation and Future Studies

Textual information databases are coming into wider use. Questions of cost, and the absence of centralization and control of contents, have led to an increase in the quantity of information that is not keyed or catalogued. This is especially true for the Internet. This study has shown the potential benefits of accurately defining the most important information to display when a search query produces a large number of results. Interface Method C, in which the document title is displayed together with keywords and several lines from the body of the text that fulfill the search criteria, offers users several distinct advantages. The improvements outlined in this study are not difficult to implement on today's Internet search engines and in dedicated systems such as those used by information providers and internal organization systems. The information components to be displayed to the user can be defined at almost every level of textual retrieval systems and conventional systems that provide lists of documents according to defined criteria.

A future study can focus the question of whether information displayed in the results list allows users to evaluate the document environment as well the information in the specific document. Knowing the document environment may allow users to assess whether the document they are viewing is relevant to the subject of their search. Such a study should identify the types of information about the document environment that could enhance the effectiveness of the search and should also examine, as this study did, the impact of the interface on the user's sense of satisfaction and confidence.

Acknowledgements: I would like to thank Avner Ben Hanoch for his help, and the reviewers for their helpful comments.

References

Blair, D., Maron, M. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM*, New York: ACM, 289-299.

Chowdhury, G. (1999). The Internet and information retrieval research: A brief review. *Journal of Documentation*, 55 (2) 209-225.

Doll, W. et al. (1994). A Confirmatory Factor Analysis of the End-User Computing Satisfaction Instrument. *MIS Quarterly*, 453-461.

Drori, O. (1998). The user interface in text retrieval systems. *ACM SIGCHI Bulletin*, New York: ACM, 30 (3) 26-29.

Drori, O. (1999). Integration of text retrieval technology into formatted (conventional) information systems. *Software Engineering Notes (SEN)*, New York: ACM, 24 (1) 78-80.

Drori, O. (2000). The benefits of displaying additional internal document information on textual database search results lists, *Proceedings of ECDL2000 – Fourth European Conference on Research and Advanced Technology for Digital Libraries* (September 2000, Lisbon, Portugal), *Lecture Notes in Computer Science*, No. 1923, Berlin: Springer-Verlag, 69-82.

Dunlop, M. (2000). Reflections on Mira: Interactive Evaluation in Information Retrieval. *Journal of The American Society for Information Science*, 51(14) 1269-1274.

Egan, D.E. et al. (1989). Behavioral evaluation and analysis of a hypertext browser. *CHI '89 Proceedings*, New York: ACM, 205-210.

Ellis, K. et al. (1995). Critical Issues in Systems Theory and Practice. New York: Plenum Press.

Hearst, M.A. (1995). TileBars: Visualization of term distribution information in full text information access. *CHI '95 Proceedings*, New York: ACM, 59-66.

Hertzum, M., Frokjaer, E. (1996). Browsing and querying in online documentation: A study of user interface and the interaction process. *ACM Transaction on Computer-Human Interaction*, New York: ACM, 3 (2) 136-161.

<http://www.dcs.gla.ac.uk/Keith/Preface.htm>

Kirsh, S. (1998). Infoseek's experiences searching the Internet. *SIGIR Forum*, New York: ACM, 32 (2), 3.

Meadow, C. et al. (1995). A Study of User Performance and Attitudes with Information Retrieval Interfaces. *Journal of The American Society for Information Science*, 46(7) 490-505.

Peritz, B. (1984). On the informativeness of titles. *International Classification*, Frankfurt: INDEKS, 11 (2), 87-89.

Pirolli, P. et al. (1996). Scatter/gather browsing communicates the topic structure of a very large text collection. *CHI '96 Proceedings*, New York: ACM, 213-220.

Pitkow, J., Pirolli, P. (1996). Life, death, and lawfulness on the electronic frontier. *CHI '97 Proceedings*, New York: ACM, 213-220,

Regan, B. (1998). Reasons why finding information on the Web is so difficult. *IdeaBytes*, New York: Giga Information Group, Inc.

Rijsbergen, C. (1979). *Information Retrieval*. London, England,

Salton, G. (1989). Automatic Text Processing – The transformation, Analysis, and Retrieval of Information by Computer. Reading, Massachusetts: Addison-Wesley, 232-236, 384-385.

SAS/STAT User's Guide. (1990). Version 6 Fourth Edition, Volume 1, SAS Institute Inc. Cary, NC, USA.

Sievert, A. (1996). Full-text information retrieval: Introduction. *Journal of The American Society for Information Science*, New York, Wiley, 47 (4), 261-262.

Spark, J. (1971). Automatic Keyword Classification for Information Retrieval, London, England.

Tague-Sutcliffe, J. (1996). Some perspective on the Evaluation of Information Retrieval Systems. *Journal of the American Society for Information Science*, 47(1) 1-3.

Biography

Offer Drori has a Ph.D. in computer science and is a lecturer at the School of Business Administration, School of Engineering and Computer Science of The Hebrew University of Jerusalem.

Homepage: <http://shum.huji.ac.il/~offerd/>, E-mail: <mailto:offerd@cc.huji.ac.il>

Head of information database systems, SHAAM Information Systems.

Chairman, Special Interest Group on Text Retrieval Systems (SIGTRS): <http://sigtrs.huji.ac.il>.

Areas of professional engagement and research include: Online information systems, system analysis and design, large information databases, information retrieval, human-computer interaction.

Membership in: ACM (Association for Computing Machinery), Israel Association of System Analysts, and IPA (Information Processing Association of Israel)