Web Searching: A Process-Oriented Experimental Study of Three Interactive Search Paradigms

Simon Dennis
Human Factors Research Centre, University of Queensland 4072, Australia.
E-mail: s.dennis@humanfactors.uq.edu.au

Peter Bruza and Robert McArthur
Distributed Systems Technology Centre, University of Queensland 4072, Australia.
E-mail: {bruza,mcarthur}@dstc.edu.au

This article compares search effectiveness when using query-based Internet search (via the Google search engine), directory-based search (via Yahoo), and phrase-based query reformulation-assisted search (via the Hyperindex browser) by means of a controlled, user-based experimental study. The focus was to evaluate aspects of the search process. Cognitive load was measured using a secondary digit-monitoring task to quantify the effort of the user in various search states; independent relevance judgements were employed to gauge the quality of the documents accessed during the search process and time was monitored as a function of search state. Results indicated directory-based search does not offer increased relevance over the query-based search (with or without query formulation assistance), and also takes longer. Query reformulation does significantly improve the relevance of the documents through which the user must trawl, particularly when the formulation of query terms is more difficult. However, the improvement in document relevance comes at the cost of increased search time, although this difference is quite small when the search is self-terminated. In addition, the advantage of the query reformulation seems to occur as a consequence of providing more discriminating terms rather than by increasing the length of queries.

Introduction

There is a plethora of technology for searching the Web. However, when one considers the actual paradigms offered, then these can be broadly classified into four categories:

1. Unassisted keyword search: the user enters one or more search terms and the associated search engine returns a list of document summaries (usually ranked). Popular examples of this paradigm include Google (www.google.com) and AltaVista (www.altavista.com).
2. Assisted keyword search: It has been widely reported that the average Web query is somewhere around two terms (Jansen & Pooch, 2000). Such short queries are probably poor descriptions of the user's actual interest. As a consequence, some search engines produce suggestions based on the user's initial query. These suggestions most often represent expansions of the initial query, thereby attempting to help the user to identify a more precise formulation of his or her interest. For example, entering "surfing" into the AltaVista search engine produces a list of suggestions such as "surfing pictures" and "kite surfing." The number of suggestions are normally limited, but recently search mechanisms have appeared that embody query (re)formulation assistance as a core of their search mechanism, for example, Vivisimo (www.vivisimo.com) and GuideBeam (www.guidebeam.com).
3. Directory-based search: in this paradigm, the information space is divided into a hierarchy of usually manually constructed categories through which the user can browse starting from broader categories and navigating down through the hierarchy (directory) to more specific ones. This paradigm differs from the previous two because there is no query formulation required. Yahoo (www.yahoo.com) is the most prominent example of this paradigm.
4. Query-by-example: after a search mechanism has retrieved a list of document summaries, the user can select an abstract (s)he finds interesting, which is then used as the basis of a new query to the associated search engine to produce a new list of document abstracts. The query-by-example paradigm is often flagged by words like "more like this document."

Most Web-based search technology combines more than one of the above paradigms.

Are users satisfied when searching the Web? Studies of Web query logs have not been able to answer this question. For example, the Silverstein et al. study observed that most internet query sessions consist of a single query; one study

DOI: 10.1002/asi.10015

stated that 63.7% of sessions consist of a query but, " . . . how many of these simple sessions are so short because the information need was fulfilled so easily, because the user gave up in despair . . . or because the user was unaware of the usefulness of modifying the query" (Silverstein, Henzinger, Marais, & Moricz, 1999). Similar sentiments are echoed by Spink, Wolfram, Jansen, & Saracevic, (2001), "Were the users so satisfied by the results that they did not need to view more pages? Were a few answers good enough? Is the precision of Web search engines that high? Are the users after precision? What proportion was relevant in relation to the (first) X sites (retrieved)? Or did they just give up? Using only transaction log analysis, we cannot determine the answers to these questions." Most systematic studies of Web search conducted thus far have involved the analysis of query logs [see Jansen & Pooch (2001) for a review].

One way that user satisfaction can be addressed is through the use of surveys. One survey[1] of 566 Web searchers found that it takes about 12 minutes before "Web-rage." In addition, 71% of users stated that they become frustrated whether they are successful or not. Eight-six percent of users held the opinion that "a more efficient way to search the Web for accurate information should be in place." Contrasting these results is a survey of 33,000 Web users in the first quarter of 2000 (NPD Search and Portal Site Study http://searchenginewatch.com/reports/npd.html). In response to the question, "How often do you find what you are looking for?" almost 60% replied "most of the time," whereas only 2.6% replied never. Just over 21% stated that they found what they were looking for "every time." Some care must be taken when viewing the results of such surveys. Surveys based on self-assessment can be unreliable thereby hampering the production of reliable and verifiable conclusions.

Gordon and Pathak (1999) have shown that some Web search engines are more effective than others, but there have not been many attempts to analyze the effectiveness of Web searching in a rigorous fashion. This article addresses this need. To this end we compare the search effectiveness of three of the four paradigms mentioned above:

1. Unassisted keyword search as offered by the Google search engine (http://www.google.com)
2. Directory search as offered by Yahoo (http://www.yahoo.com/)
3. Assisted keyword search as offered in the form phrase-based query reformulation by the Hyperindex Browser[2] (http://www.guidebeam.com)

One motivation of this study was to compare the effectiveness of straight keyword search against browsing a directory. Our assumption in this regard was that directory-based search would lead to higher relevance of documents as the underlying collection has been vetted by editors. Also, by drilling down the directory to a leaf node containing a relevant document, we expected that other documents associated with the leaf node would also have a high likelihood of relevance. Another motivation was to investigate whether query reformulation assistance does lead to more effective queries compared to unassisted keyword search. Our underlying assumption here was query reformulation assistance would lead to more expressive queries (i.e., longer), which would translate into better document summary rankings.

In addition to studying effectiveness of Web searching, our aim was to trial an evaluation framework for interactive search. This framework focuses on evaluating aspects of the search process, not only the end result. As a consequence, evaluation criteria within this framework not only measure relevance, but time spent in a particular search state, time to bookmark the first relevant document and the cognitive load imposed by various states during search. The work presented here goes beyond our earlier work by employing the evaluation framework for the self-terminated search and for a broader range of queries.

The Search Mechanisms

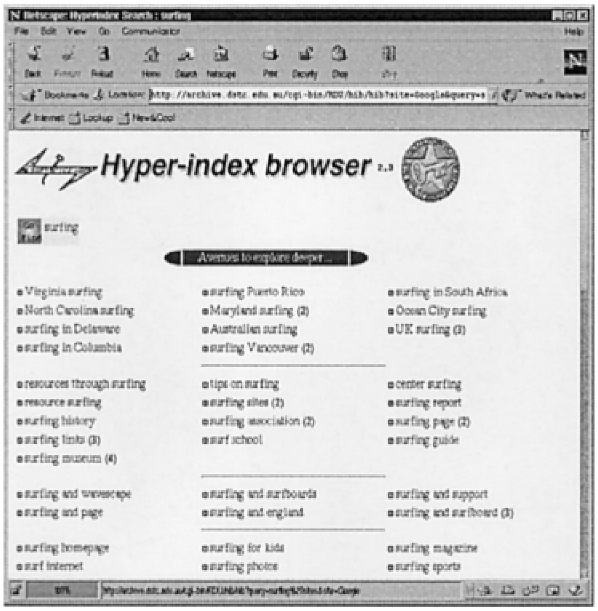
Google was chosen as the mechanism to support standard Internet query search because it seems to be one of the more effective newer generation search engines. In addition, its interface is less cluttered than the more established search engines. Yahoo was chosen, as it is the most established, and probably most comprehensive Internet directory.

Short queries on the WWW are a well-known phenomenom. For this reason a number of query formulation aids have appeared in conjunction with Web-based search engines. The Hyperindex Browser (HiB) is a Web-based tool used in conjunction with a search engine producing reformulations of a query in the form of linguistically well-formed phrases (Bruza & Dennis, 1997). The user enters an initial query that is passed onto the associated search engine. The resultant document summaries are not presented directly to the user but are analyzed using a shallow natural language parsing technique. Phrases are derived from the documents in the result set and displayed to the user as candidate refinement possibilities for the initial query. For example, a query "surfing" would produce refinements like "Australian surfing," "tips on surfing," "surfing and surfboards" (see Fig. 1). The user can then select any of these refinements to become the new query, and the process repeats. In this way, the user can navigate to a desired phrase in a process termed Query-By-Navigation (QBN; Bruza & Van der Weide, 1992).

At any point a desired phrase can be used to retrieve a ranking of document summaries. In this experiment, the HiB was configured to use the Google search engine as its associated retrieval mechanism.

[1] http://www.zdnet.com/zdnn/stories/news/0.4586,2667216,00.html.

[2] The Hyperindex Browser has since evolved into GuideBeam (www.guidebeam.com).


Netscape: Hyperindex Search : surfing
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://archive.dstc.edu.au/cgi-bin/RDU/hib/hib?site=Google&query=s
What's Related
Internet Lookup New&Cool
Hyper-index browser 2.3
Avenues to explore deeper...

surfing

- Virginia surfing
- North Carolina surfing
- surfing in Delaware
- surfing in Columbia
- surfing Puerto Rico
- Maryland surfing (2)
- Australian surfing
- surfing Vancouver (2)
- surfing in South Africa
- Ocean City surfing
- UK surfing (3)

- resources through surfing
- resource surfing
- surfing history
- surfing links (3)
- surfing museum (4)
- tips on surfing
- surfing sites (2)
- surfing association (2)
- surf school
- center surfing
- surfing report
- surfing page (2)
- surfing guide

- surfing and wavescape
- surfing and page
- surfing and surfboards
- surfing and england
- surfing and support
- surfing and surfboard (3)

- surfing homepage
- surf internet
- surfing for kids
- surfing photos
- surfing magazine
- surfing sports

FIG. 1. Query refinements produced by HiB.

The Paraphrase Search Assistant has a similar philosophical basis to the HiB in that it employs the interactive use of linguistic phrases to support query reformulation (Anick & Tipimeni, 1999). A major difference between that study and ours is the evaluation methodology. Anick's evaluation focused on log file analysis to compute how often phrases were used and within which search tactic (e.g., phrases that "narrow," or act as "suggestions"). In another related approach, WordNet has been used to expand internet queries with some improvements in precision (Moldovan & Mihalcea, 2000).

Modeling the Search Process

At any given moment during a typical interactive Internet search users can be viewing a number of different types of information and have a number of different options in terms of what they do next. For instance, they may be viewing a home page, a list of document summaries, or a document. Similarly, they may choose to move to a new document, back to a summaries page, or back to the original query box. The first stage in understanding the search process is to enumerate what states a user may be in and what transitions they may choose given their current state. For Google, Yahoo, and HiB, these are depicted in Figures 2, 3, and 4, respectively.

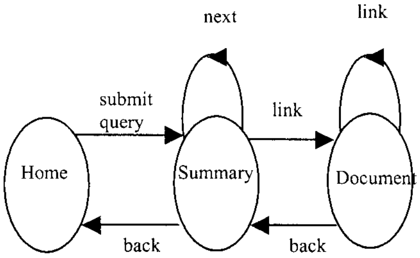


FIG. 3. Search state transition diagram for Yahoo.

Home State

This state is where the user begins the search process. In Yahoo, the home page displays the top-level directory. If the query is to the Google search engine, the home page contains the query box, and when the query is entered it leads to the state in which the user peruses document summaries (see Fig. 2). If the initial query was to the HiB, this leads to a state in which the user peruses query refinements (see Fig. 4).

Refinement State

The refinement state is specific to the HiB. In this state the user sees a focus representing the current query description and candidate refinements of this focus (see Fig. 1), each of which can be selected to make the current focus more specific. The current focus can also be selected. This results in the focus being used as a query to return document summaries. This transition is termed a "beam down," as it transports the user from the space of potential queries (the hyperindex) to the level of the document summaries (see Fig. 4).

Directory State

This state is specific to Yahoo. In this state, users see a set of possible categories under which the information they

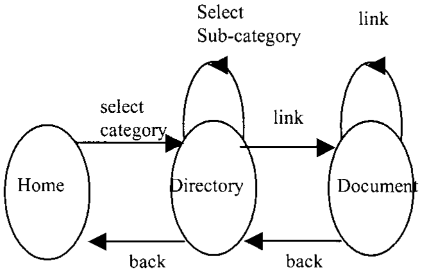


FIG. 2. Search state transition diagram for Google.

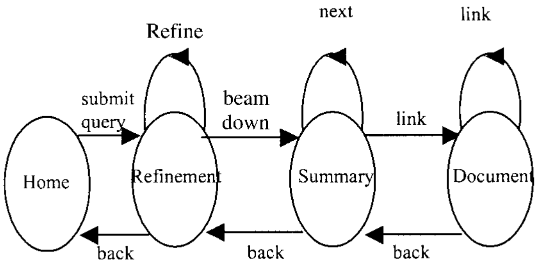


FIG. 4. Search state transition diagram for HiB.

require may appear, and they must chose the most likely of these (see Fig. 3). Subjects using Yahoo were not permitted to use Yahoo's query facility.

Summary State

In this state, the user peruses document summaries, which consists of a title, a short abstract, and links to the actual documents. The link can be activated to transfer the user to the document state. The user can also traverse to the next page of summaries.

Document State

The state in which the user is perusing a Web document. If the user finds the document relevant, it can be bookmarked. Alternatively, the user can follow a link to another document.

In all states, the back button can be activated to transfer the user to the previous state or to a different page within the same state. In addition, the home state is accessible from any state if the user wishes to start afresh. The document state is a *termination* state when the user has bookmarked a document that satisfies the information need. Any state can be a termination state due to a time out (not due to the page timing out on the network, but because subjects were given 5 minutes per query).

Performance Criteria for Interactive Search

There are a number of dependent variables that can be used as measures of search effectiveness, for example, the traditional information retrieval measures such as recall and precision. However, these measures have a couple of important limitations that should be noted. First, the document collection is the WWW where recall is impossible to measure. Moreover, most searches on the WWW are not concerned with finding all of the relevant material. For this reason, we have chosen to measure the quality of the search results by having independent raters judge, on a seven-point scale, the quality of the documents that were perused by subjects during their search. The mean of the ratings for all of the documents that the subject peruses during a search is termed "relevance rating."

Second, when discussing interactive search it is important to record how long it takes a searcher to find the material for which they are looking. In the experiment subjects bookmarked what they considered to be relevant pages as they conducted the search, and we used the time to the first bookmark as an indicator of how long the engine was taking to present relevant pages.

Third, and perhaps most importantly, it is crucial to measure the demands placed upon the user while interacting with the search mechanism—measures that capture the users' experience during the search process. The first measure that we chose to collect was the amount of time that a user spends in each state of the search process—composing queries, assessing refinements, reading through document summaries, and evaluating the documents themselves (see previous section).

In addition, however, we can look to the human factors literature for methods that more directly reflect the amount of cognitive load experienced by the user. We chose to employ a dual task methodology. As such measures are not common within the information retrieval literature we briefly discuss the measurement of cognitive load in the following section.

Cognitive Load and Dual Task Methodology

The dual task technique has a long history in the human factors and cognitive psychology literature (see Wickens, 1992, for a summary). The idea is to have subjects do two tasks simultaneously. For instance, they might be asked to monitor a sequence of digits for repeats (primary task) while at the same time detecting tones played through headphones (secondary task; Posner & Bois 1971; Posner & Klien 1973). If people have a constant size pool of cognitive resources upon which they can draw, the amount of effort they employ on the primary task will be inversely proportional to their performance on the secondary task. That is, as a subject is required to focus more on the primary task they will have fewer cognitive resources "left over" to employ on the secondary task and performance will be compromised.

When employing dual-task methodology, it is important to keep in mind its limitations. First, the interpretation of high cognitive load can be problematic. In general, high load leads to increased fatigue, poor learning, and an increased probability that users will fail to attend to relevant detail (i.e., a loss of situational awareness). However, if load is high because the user is engaged in task-relevant activities, such as deciding between relevant refinements, then high load may result in superior performance as measured by relevance or time to first bookmark. As a consequence, it is important to view cognitive load in the light of relevance and search time data.

Second, current theories of human resources contend that there are multiple pools and that tasks that draw on different pools will not necessarily impact upon each other. In this case, the secondary task performance may not be sensitive to differences in primary task demands, so it is important that one chooses a secondary task that is sufficiently similar to the primary task (in this case Internet search).

Third, as people become practiced at a task, so that it becomes automatized, the resources needed to complete this task may decrease. For this reason, it is important not to supply subjects with too many trials when employing dual-task methodology.

Finally, the secondary task should not precipitate a change in the strategy that subjects use on the primary task. For instance, if it were the case that the tone detection task caused subjects to process the letter sequences in an entirely different fashion, then we are no longer measuring the task of interest.

A large variety of secondary tasks have been employed, from finger tapping (Michon, 1966), to random number generation (Baddeley, 1966). After several pilot experiments, we chose a digit-monitoring task in which subjects listened to a stream of digits (from one to five) and responded when a digit was repeated. This task seemed to impose a considerable but not overwhelming load on the subjects, and provided two simple measures of secondary task performance: reaction time (time to respond when a dual digit is heard) and the miss rate (how many times a dual digit is not responded to). Furthermore, in a previous study (Dennis, McArthur, & Bruza, 1998), this task was successfully employed to demonstrate that subjects were less loaded when perusing query refinements than when perusing the document summaries generated by the Excite search engine. So there was a reasonable expectation that it may provide a sensitive measure for the purposes of this investigation.

Experiment One[3]

Subjects

Fifty-four subjects were recruited from the undergraduate psychology pool at the University of Queensland, and received credit for their participation. A pretest questionnaire was administered to gather demographic, computer usage, and computer attitude information: 21 of the subject were male, 31 female, and three did not respond. The mean age was 20.02 years, with a range of 17–37 years. Forty-seven of the subjects owned their own computer. Subjects were asked to rate on a five-point scale how long they had used computers, how often they use computers, and how often they use the Internet. In addition, they were asked 13 questions on a seven-point scale. The first 10 of these were combined to provide a measure of attitude towards computers. The last three were combined to provide a measure of computer anxiety. On all demographic and computer-related measures one-way analysis of variance was conducted to ensure that there were no significant differences between the subjects assigned to each of the search engines. The experiment was conducted between August and November 1999.

Design and Materials

A single-factor design was used. Search engine was manipulated between subjects and could be Yahoo, Google, or the Hyper Index Browser (HiB).

Eighteen queries were carefully generated to be a broad brushstroke of interesting Internet queries (see Appendix A). These were divided into three sets of six queries, and each subject saw one of the sets.

Procedure

Before completing the experiment subjects were given two questionnaires to complete. The first collected demographic, computer attitude, and computer anxiety information.

In the second, subjects answered a series of domain knowledge questions about the queries for which they were going to be searching. For instance, if they were going to be searching for pages of women's wave surfing (q1.3) then they were asked for the names of women surfers that they knew. Their responses were rated as either high- or low-domain knowledge by the experimenter. Our intention in asking these questions was to attempt to factor out variations between subjects in domain knowledge and to see if these variations might favor one engine over another. In particular, we hypothesized that the engines that provide some structure to the information domain such as Yahoo and the HiB might help subjects with little domain knowledge. Unfortunately, analysis of the results indicated no differences, most probably as a consequence of the domain questions being an insufficiently precise measure, and so we will not report any of these results.

During the experiment, subjects conducted searches using the PRISM browser developed at the Key Center for Human Factors and Applied Cognitive Psychology. The browser allows an experimental procedure to be administered, and records the URLs that subjects visit, the amount of time they spend in each, whether they bookmark a page, and the degree of cognitive load they are under as they conduct the search (as measured using a dual digit monitoring task). The software also classifies pages into different types using pattern matching on the URL of each page. In this experiment, pages were classified into Yahoo, Google, or HiB home pages, Yahoo directory pages, Google Summary pages, HiB refinement pages, HiB summary pages, and document pages (c.f. state transition diagrams).

At the start of the search phase, subjects were given written instructions on how to use the engine to which they had been assigned, including the restrictions that we placed on each of the engines. In the case of Yahoo, they were told to use only the directory mechanism and not the query box.

They were also given instructions on completing the dual task. During each search a random series of digits between one and five were played into their headphones. So that they would not become accustomed to the rate of the digits and hence switch attention to the dual task in a rhythmic fashion rather than maintaining attention on the dual task, the digits were timed to have a mean interdigit interval of 5 seconds, with a uniform random variation around this mean of 1.5 seconds. Subjects were required to hit the escape key when a digit was repeated. To further ensure that subjects would have to maintain attention on the dual task and that data was collected uniformly across all phases of a search, a double digit was forced every five iterations if one did not occur by chance. In pilot testing, these values seemed to provide a balance between the collection of enough data to monitor

[3] This experiment was previously reported in Bruza, Dennis, and McArthur (2000).

cognitive load while ensuring that the subject continued to see the Internet search as their primary task.

After reading through the instructions, subjects conducted seven searches. The first was a practice trial to allow them to become familiar with the dual task procedure, and none of the data from this trial is reported. Each question appeared on the screen. When they had read and understood the question they clicked on a continue button that took them to the home page of the engine to which they had been assigned. They then had 5 minutes to find as many pages relating to they query as they could. Any that they felt were relevant they were asked to bookmark. The question remained in a text box at the top of the browser throughout the experiment. Subjects completed the entire experiment within 1 hour.

Results

Time to first bookmark

There were significant differences between the amounts of time it took for subjects to register their first bookmark as a function of engine, $F(2, 51) = 23.59$), $p < 0.001$. Yahoo took the longest, with a mean of 137 seconds. Next was the HiB with a mean to first bookmark of 117 seconds and the fastest engine was Google with a mean time of 75 seconds.

Relevance

Relevance judgements are made in the context of the quality of the documents that the subject is seeing and subjects will adjust their criteria appropriately (raising it if they are seeing many relevant documents and lowering it if they are seeing many irrelevant documents). Because search engine is a between subjects factor in this experiment, no direct comparisons of user-generated relevance judgements (of which bookmarks are one type) can be made. For this reason, independent raters were employed, and we have not analyzed the number of pages bookmarked across engines.

Independent relevance judgements were compiled for 504 documents perused for Yahoo subjects, 794 for Google, and 648 for HiB. (In reality, the figures for documents accessed were higher, but due to the dynamic nature of the WWW it was not possible to assess relevance for some URLs). Figure 5 shows the mean relevance ratings from the independent raters of the documents that were retrieved by subjects for each of the queries and as a grand mean. These results include all documents that the subjects saw, not only those they bookmarked. Note that these analyses were conducted with the unique documents as the random factor. As a consequence, because different queries generate different numbers of documents they are weighted differentially in the analysis of the grand mean.

One-way analysis of variance was applied to the grand mean and to each of the queries, and the * or ** postceding each title (in Fig. 5) indicate the results that were significant at the .05 and .01 levels, respectively. Table 1 shows the results of post hoc analyses of the differences between engines for those engines with a significant one-way ANOVA.

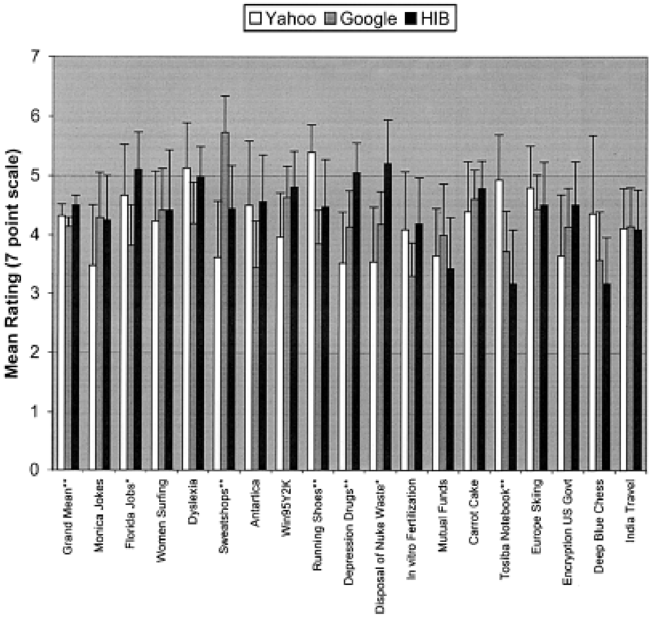


FIG. 5. Relevance of the pages that subjects visited as a function of Query for Experiment 1. Error bars represent the 95% confidence intervals. * = $p < 0.05$, ** = $p < 0.01$.

Time in state

Figure 6 shows the time spent in each state as a function of engine, and Figure 7 shows the these figures as percentages of total time taken. First, note that the amount of time spent in the home state differed by only 2 seconds across engines. Second, there was only a mean difference of 6 seconds between the time spent in Google summaries versus the time spent in the HiB summaries. The extra time spent in the HiB refinement state was being taken primarily from the time spent in documents of Google. Finally, note that the subjects spent the least time in documents when using the HiB, followed by Yahoo and then Google.

Dual task

Dual-task performance can be measured either in terms of the number of times the subjects fail to respond when a digit was repeated (the miss rate) or in terms of the amount of time it takes for them to respond when a digit is repeated. Previous work has found the miss rate to be a useful measure (Dennis, McArthur, & Bruza, 1998). In this experiment, however, miss rate was not significant, $F(2, 51) = 0.30$. However, reaction time did show a significant difference between the Google summaries state (mean = 1546 ms), and the HiB refinement state (mean = 1815 ms), $F(1, 34) = 4.21$, $p = 0.05$, but not between the Google summary state and the HiB summary state (mean = 1752), $F(1, 34) = 2.14$, or the Google summary state

TABLE 1. Means and significance levels of relevance differences for the grand means and significant queries.

	Yahoo	Google	HiB	HiB versus Google	Yahoo versus Google	HiB versus Yahoo
Grand Mean	4.3	4.1	4.5	.002	.146	.174
Florida jobs	4.7	3.8	5.1	.010	.114	.411
Sweatshops	3.6	5.7	4.4	.026	.001	.150
Running shoes	5.4	3.9	4.5	.184	.001	.044
Depression drugs	3.5	4.1	5.0	.027	.236	.004
Disposal nuclear waste	3.5	4.2	5.2	.031	.236	.005
Toshiba notebooks	4.9	3.7	3.2	.351	.020	.005

and the Yahoo directory state (mean = 1766), $F(1, 18) = 2.86$.

Discussion

The time to first bookmark results are not surprising. The HiB involves one or more refinement states before the user beams down to the document summaries. A subject must traverse through the Yahoo directory from a root classification, to a leaf classification, before document summaries are seen. Proceeding through the refinement or directory states takes time—time that the Google subjects will not experience, because a query directly results in a ranking of document summaries. This also explains why Google subjects perused more documents than either HiB or Yahoo subjects.

In terms of relevance, Table 2 shows that Google was superior to both HiB and Yahoo on the Sweatshops query. "Sweatshop" is an unambiguous term with fairly high discriminating power. Such terms feed well into statistical ranking techniques, and this may explain Google's superiority. Contrast this with the "depression drugs" and "radioactive waste" queries. These queries involve more general terms that open up the problem of vocabulary mismatch. For example, in the query dealing with the disposal of radioactive waste, the HiB readily produces a refinement "radioactive waste management," which is a useful phrase for retrieving relevant documents. None of the subject's queries formulated for the Google search engine employed the term "management." Further support comes from the number of unique terms submitted to the retrieval engine through either Google or the HiB. For Google there were 166, while for the HiB there were 198. Important to note here is that the HiB query term vocabulary is in part derived from the titles of retrieved documents. In this sense, our study contributes to the evidence that phrases generated from titles or abstracts assist in the discovery of useful query terms (Spink, 1994).

More generally, this experiment shows that query reformulation support via the interactive use of phrases (HiB) does lead to higher relevance of documents perused than unassisted search (Google). The improved relevance does not stem from HiB refinements (average 2.95 terms) being longer than Google queries (average 2.97 terms). An interesting point to note is that the HiB subjects spent the least time perusing the documents, but those they did peruse tended to be those with higher relevance. One possibility in this regard is that the users tended to prefer perusing query refinements (i.e., gaining an overview of the underlying document space) rather than trawling through the documents themselves.

Yahoo performed best with the shopping-related queries "Toshiba Notebooks" and "Running Shoes." This is due to the fact that the Yahoo directory editors have optimized such queries. If these queries were omitted from the analysis, then HiB would be superior to Yahoo as well as Google with respect to average relevance. Interestingly, this study did not reveal that directory-based search improved relevance over standard query-based Internet search.

These experiments demonstrate that cognitive load complements the use of relevance judgements and time to first

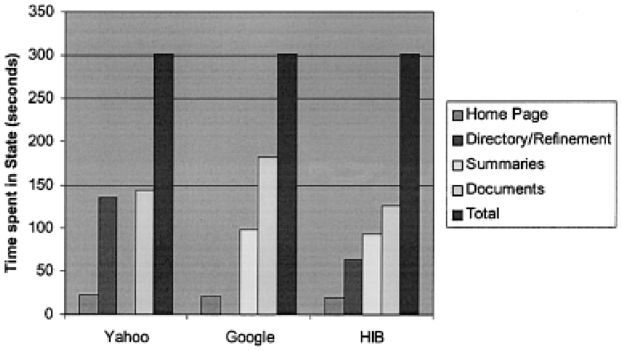


FIG. 6. Time spent in state as a function of engine.

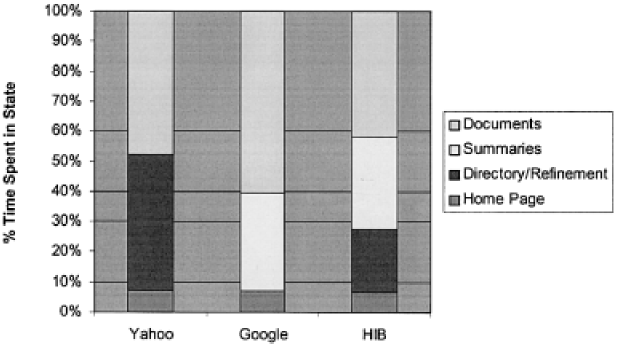


FIG. 7. Percentage of time spent in state as a function of engine.

TABLE 2. Means and significance levels of relevance differences for the grand means and significant queries.

	Yahoo	Google	HiB	HiB versus Google	Yahoo versus Google	HiB versus Yahoo
Grand mean	3.23	4.52	4.71	.243	0.001	0.001
Monica Lewinsky	1.94	4.00	2.58	.018	.259	.002
Florida jobs	2.88	4.80	5.36	.402	.001	.001
Dyslexia	2.82	6.25	5.80	.398	.001	.001
Clothing Sweatshops	1.70	5.22	4.40	.421	.001	.001
Antivirus Software	1.44	4.31	5.00	.421	.001	.001
Measuring Creativity (H)	3.00	4.44	5.00	.319	.003	.028
Encryption Technology	2.67	6.44	4.20	.010	.198	.001
Deep Blue	1.54	5.38	4.83	.559	.001	.001
Travel India	3.00	5.15	4.75	.612	.124	.003
Lockerbie Plane Crash (H)	2.91	4.06	6.00	.055	.017	.208

(H) indicates the hard queries.

bookmark for evaluating interactive search processes. Relevance by itself, or even with the inclusion of a simplistic measure of effort such as the time to first bookmark, would have missed how much effort a user must employ when using these systems. Even though the HiB produced higher average relevance than Yahoo and Google, it came at a "cost" via the higher cognitive load when perusing HiB refinements. This finding was converse to our intuition: as the hyperindex allows the user to view the underlying document space at higher level of abstraction (by navigating through query refinements derived form it), we felt that the cognitive load while perusing query refinements could be less than when perusing document summaries, where there tends to be more detail. One possible explanation for the higher cognitive load was that most subjects were not familiar with the HiB browser and its associated query by navigation metaphor. The unfamiliarity may have induced increased cognitive load. Another factor leading to increased load may have been related to the query refinements not being ordered in any obvious fashion, for example, alphabetically, or with respect to probability of relevance.

It has been reported that users tend not to reformulate (Jansen, Spink, Bateman, & Saracevic 1998; Silverstein, et al., 1999). HiB subjects spent about 20% of their time in the refinement state. We conclude that users will reformulate when there is satisfactory support for this, but note that they may only take advantage of the support in a limited way by refining, on average, once after the initial query.

One can ponder to what extent the experimental methodology chosen influenced the results. In particular, users were given a fixed time (5 minutes) per query (the motivation for this was to provide a uniform basis for analysis). Typical Internet search does not follow this pattern. For this reason a second experiment was conducted based on self-terminated search.

Experiment 2

The second experiment featured self-terminated search across the same search states investigated in experiment one. In addition,

1. Query refinements were ordered according to probability of relevance (see McArthur & Bruza, 2000, for details of the ranking function). Our hypothesis is that ordering the refinements according to probability of relevance will reduce cognitive load.
2. There was anecdotal evidence that the HiB performs better with "hard" queries, that is, queries whereby it is difficult to *a priori* formulate query terms in response to an information need. For the second experiment, 12 queries from the TREC-8 *ad hoc* and small Web track topics were identified as being difficult. These queries were to be used in addition to those used in the first experiment. Our hypothesis is that such queries would benefit most from query formulation assistance.

Subjects

Fifty-seven subjects were recruited from the undergraduate psychology pool at the University of Queensland, and received credit for their participation. The experiment was undertaken in early 2000.

Design and Materials

The design for the second experiment was the same as for the first, with engine as a between subjects factor. Each set of queries from experiment one (normal queries) was augmented with four queries taken from the TREC-8 collection (hard queries, see Appendix B). In total, there were 30 queries—18 normal queries and 12 hard queries.

Procedure

The procedure for experiment two was identical to that in experiment one except that:

1. When subjects bookmarked a relevant page the browser moved directly on to the next question, that is, the search was self-terminated.
2. After the completion of the six normal queries, subjects completed an additional four hard queries. Note we chose to add the hard queries after the normal queries to

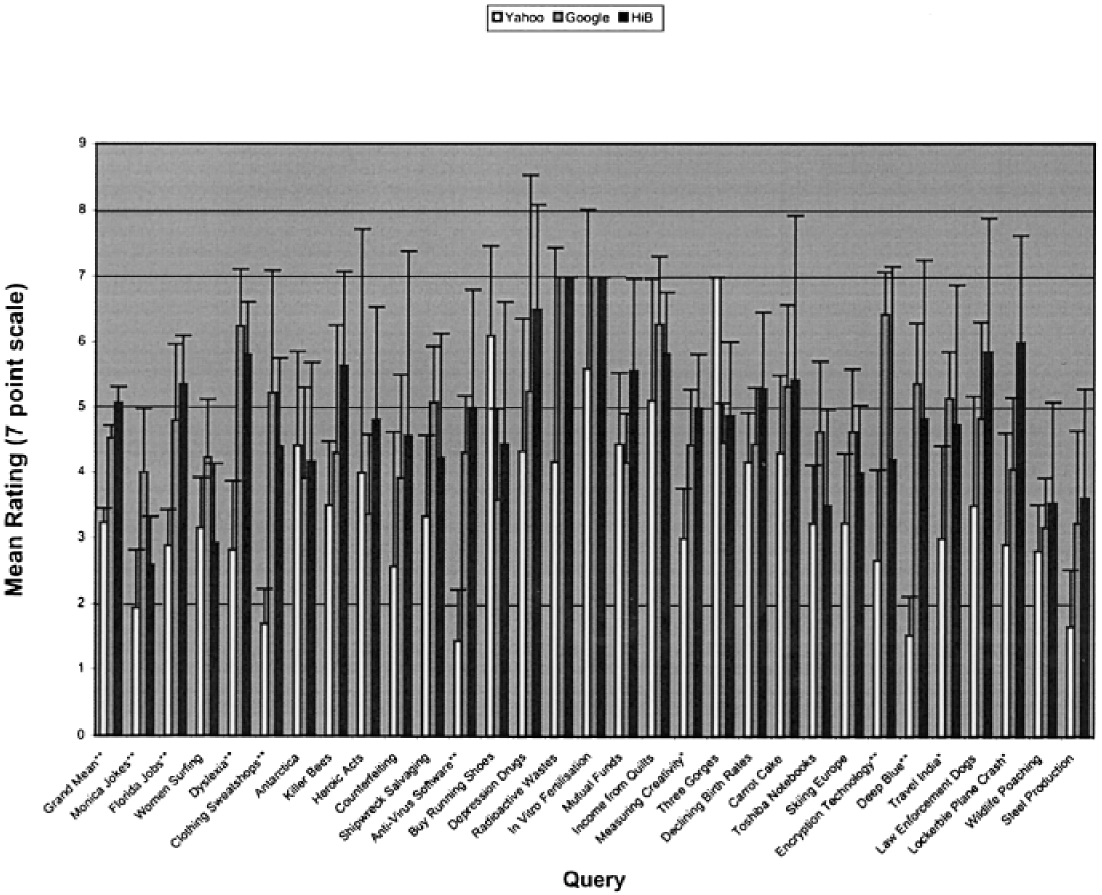


FIG. 8. Relevance of the pages that subjects visited as a function of Query for Experiment 2. Error bars represent the 95% confidence intervals. * = $p < 0.05$, ** = $p < 0.01$.

ensure that any comparisons of performance on the original queries between Experiments one and two were not a consequence of changes in strategy induced by the hard queries. These queries were also randomized for order.

3. Because the questionnaire data had not contributed in the first experiment, we chose not to collect this data in the second experiment.

Results

Time to first (and only) bookmark

Again, there were significant differences between the amounts of time it took for subjects to register their bookmarks as a function of engine, $F(2, 54) = 8.8$, $p < .001$. Yahoo took the longest with a mean of 154 seconds. Next was the HiB, with a mean to first bookmark of 126 seconds and the fastest engine was Google with a mean time of 108 seconds. Yahoo took significantly longer than the HiB, $F(1, 37) = 4.91$, $p = 0.033$ and than Google $F(1, 37) = 26.09$, $p < 0.001$, but there was no significant difference between the HiB and Google, $F(1, 37) = 2.61$. So the same pattern of results was observed. Note that these times are somewhat longer than those recorded in the first experiment, where subjects were able to bookmark multiple pages and were therefore likely to be less conservative than in the second experiment.

Relevance

Relevance judgments were compiled for 362 documents perused for Yahoo subjects, 426 for Google, and 273 for HiB. Figure 8 shows the mean relevance ratings from the independent raters of the documents that were retrieved by subjects for each of the queries and as a grand mean.

With the inclusion of the hard queries in this experiment the performance of Yahoo has dropped while Google has improved. There was a main effect of engine $F(2, 1055) = 51.24$, $p < 0.001$, which occurred because both the HiB and Google outperformed Yahoo. There was no significant difference between the HiB and Google, overall.

When the type of query (hard vs. normal) was added to the analysis (see Fig. 9) there was a significant interaction with engine $F(2, 1055) = 5.53$, $p = 0.004$. This interaction occurred because Google produced higher relevance documents on the normal queries $F(1, 425) = 48.93$, $p < 0.001$, while Yahoo, $F(1, 361) = 40.93$, $p < 0.001$ and the HiB, $F(1, 272) = 4.77$, $p = 0.03$, produced higher relevance documents on the hard queries. For the normal queries, the relevance of the documents provided by

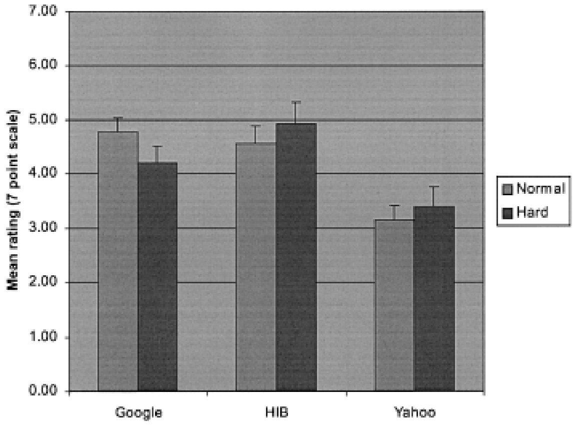


FIG. 9. Mean relevance rating as a function of engine and query type. Bars represent the 95% confidence intervals.

the HiB was higher than those provided by Yahoo $F(1, 399) = 41.70$, $p < 0.001$, as were the documents provided by Google $F(1, 256) = 76.28$, $p < 0.001$. There was no significance difference between the HiB and Google engines $F(1, 390) = 1.19$. For the hard queries the pattern was somewhat different. The HiB outperformed both Google $F(1, 307) = 8.43$, $p = 0.004$, and Yahoo $F(1, 234) = 31.52$, $p < 0.001$. In addition, there was a significant advantage for Google over Yahoo $F(1, 310) = 10.70$, $p = 0.001$. These results are in accordance with our hypothesis that the additional structure provided by query reformulation will be of greatest aid when the information need is difficult to specify.

Time in state

Figure 10 shows the time spent in each state as a function of engine and Figure 11 shows the same data as percentages of the total time. Overall subjects spent about half as much time in the search task as they were required to in the first experiment. However, the decrease in total time did not affect all states equally. The main impact of self-terminating search appears in the document perusal phase where the mean time decreased from 151 seconds to just 45 seconds. A similar percentage decrease is seen in the summary state of the HiB. In the first experiment subjects spent about 92 seconds, whereas in the second experiment this was reduced to about 35 seconds. By contrast, the amount of time subjects are spending in the refinement and directory states is relatively intact. For the HiB, the time in the refinement state was 63 seconds in the first experiment, and about 63 seconds in the second experiment. For Yahoo, the time in the directory phase was 136 seconds for the first experiment, while in the second experiment it was 119 seconds. These results suggest that when subjects were attempting to go faster they chose to invest their time in the perusal of refinements and directories rather than summaries and documents.

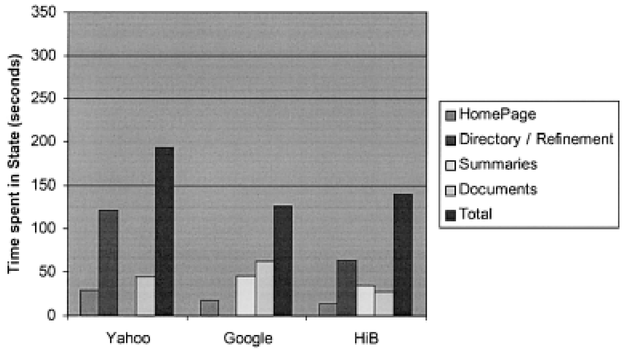


FIG. 10. Time spent in state per query as a function of engine.

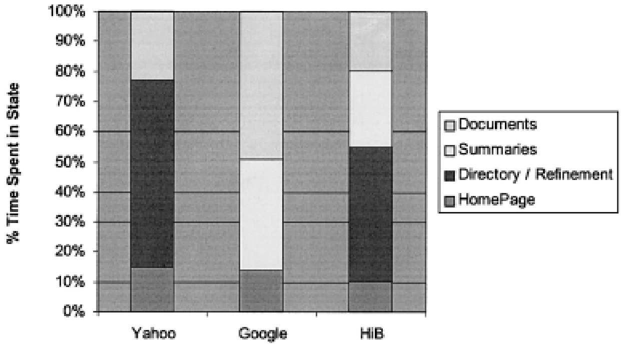


FIG. 11. Percentage of time spent in each search state as a function of engine.

Dual task

In experiment two, neither the miss rate, $F(2, 51) = 0.829$, $p = 0.442$, nor the reaction time data, $F(2, 51) = 0.286$, $p = 0.752$, produced significant results, suggesting that the digit monitoring task was not sufficiently sensitive to measure any differences in load that occurred across engines or states.

Discussion

Search Effectiveness

In the experiments reported in this article we employed three measures of search effectiveness: the amount of time required for subjects to find a document they considered relevant, the relevance of the documents subjects were required to peruse, and the degree of cognitive load experienced by the subject as indicated by the miss rate and reaction time on a secondary digit monitoring task.

In both experiment one and experiment two, subjects took longest to bookmark a page in Yahoo, followed by the HiB and then Google (although the difference between Google and HiB was not significant in the second experiment). As outlined in the first discussion, these results are not surprising. When using the HiB, subjects must traverse at least one refinement state before being presented with summaries, and this will take time. In addition, the HiB

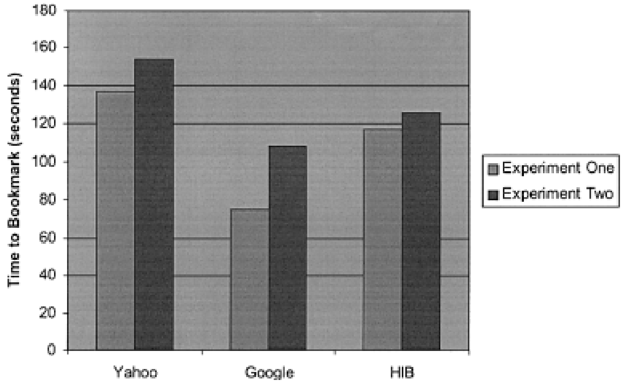


FIG. 12. Time to bookmark as a function of engine for Experiments one and two.

must itself make a call to the Google engine to retrieve the results from which the hyperindex is built, so any network delays will affect the HiB more severely. The directory structure of Yahoo will often require several steps as subjects work their way through the hierarchy, and again, this will take time.

It is interesting to note, however, that the times to bookmark were longer in the second experiment where search was self-terminated, and that the difference between Google and the HiB was not as great under these conditions (see Fig. 12). It is probable that subjects were more conservative in the second experiment because they knew that they would receive only one opportunity to bookmark a page. This would have encouraged the subjects, and Google users in particular, to spend more time perusing summaries and documents before deciding upon the most relevant item; hence, the overall increase for all subjects and the dramatic increase in time spent by the Google subjects. The real searching situation probably lies somewhere between experiments one and two. While real search is self-terminated, it is not the case that a searcher may choose one and only one document. What these results demonstrate, however, is that the differences between the times to use Google compared against the HiB may not be as significant as one might assume.

In the first experiment, subjects were given a fixed time period of 5 minutes to find as many relevant articles as they could. Furthermore, they were given queries that were designed to reflect an average Internet user's information need. Under these conditions, Yahoo and the HiB provided more relevant documents than Google. In the second experiment where search was self-terminated and some more difficult queries were included it was Google and the HiB that out performed Yahoo. The first point to note about these results is that, regardless of the testing conditions, the query reformulation process employed by the HiB is successfully narrowing the search space to a more relevant set of documents. However, it would appear that the usefulness of keyword search as compared against directory-based search depends on the amount of time pressure on the search process. When using Google, the extra time that subjects were required to spend in the first experiment would allow them to progress further down the relevance ranked list of summaries, which would in turn, compromise Google's performance. By contrast, Yahoo takes longer to use and the extra time allowed subjects to find the relevant set.

The experiments reported in this article did not show any overall differences in cognitive load between the search mechanisms. Whether these results are caused by a lack of power of the digit monitoring task or reflect the fact that load does not vary dramatically across engines is difficult to discern.

Unassisted vs. Assisted Keyword Search

In self-terminated search, the average query length for Google queries was 2.86 terms; "hard" queries were slightly longer: 3.15 terms versus 2.68 terms for "normal" queries. For the HiB, the average beam down query length[4] was also 2.90 terms, with no difference between "hard" and "normal" queries. The overall averages were equivalent to those observed in terminated search (Google: 2.97 terms; HiB: 2.95 terms).

Query reformulation (i.e., assisted keyword search) does seem to achieve increased relevance over unassisted keyword search. Our hypothesis that query reformulation would be more effective for "hard" queries was borne out by the results in self-terminated search, however, the increased relevance was not a product of query refinements being longer than unassisted queries. This is the same observation made for terminated search, whereby query formulation assistance yielded increased relevance over unassisted keyword search. The increased relevance was a product of the *particular* terms being produced by the query refinement mechanism. For example, one HiB subject began with the query "artificial fertilization," then selected the refinement "in vitro fertilization" and then used the refinement "fertilization IVF" as a beam-down query. This example demonstrates how query refinement can help users identify useful query vocabulary without having to come up with it themselves. There were examples where query refinement helped counter spelling errors. For example, one subject started with the query "mesuring creativity," which lead to the refinement "measurement of creativity," which they then selected yielding the refinement "model for the measurement of creativity."

In addition, a query refinement of three terms seems to be specific enough, on average, for subjects to use for retrieving document summaries. This has implications for the design of query refinement mechanisms for Web searching. Based on our findings, producing refinements of more than three terms would seem unnecessary in many cases, as users would tend not to use them. Effort should be directed at producing query refinements of three terms providing a

[4] These are refinements that HiB subjects chose to use for retrieving document summaries.

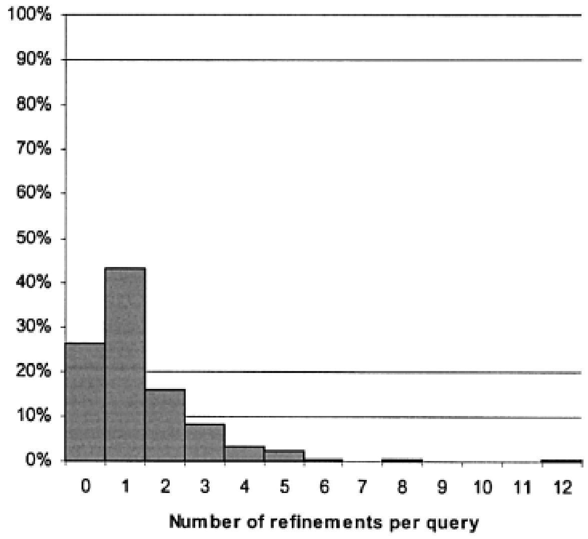


FIG. 13. Histogram of number of refinements per query.

good cross-section of possibilities. These possibilities essentially provide a map of a part of the underlying document space, which can help the user orient themselves.

During self-terminated search, HiB users refined on average 1.6 times after an initial query and beaming down to view document summaries. This is higher than observed for terminated search where there was, on average, 1.0 refinement. Refining a query can be envisaged as starting with an initial query and then navigating through a space of candidate queries. During a query session, the user may refine a number of times, view document summaries and then come back and refine again. They may also initiate a refinement sequence in a number of ways, and go back and use previously seen query refinements, as the following query log fragment illustrates

```
(start): antidepressants
(start): Depression
(refinement): support depression
(back): Depression
(refinement): treatment of depression
(start): Prozac
(refinement): prozac in depression
(back): Prozac
(refinement): prozac substitutes
(refinement): substitutes for pharmaceuticals
(back): prozac substitutes
(refinement): prozac substitutes depression
  books
(back): prozac substitutes
(back): Prozac
(refinement): antidepressant prozac
(refinement): antidepressant drugs
(beam down): antidepressant and drugs depres-
  sion
```

Figure 13 shows that there is quite a variation in the number of refinements per query. Interestingly, there was no discernible difference between “hard” queries and “normal” queries. One could have expected that there may be significantly more refinements for “hard” queries.

In summary, query reformulation does seem to produce more effective searching, but this advantage would seem to come from the provision of more discriminating search terms rather than through the provision of longer queries, and seems to be dominated by the initial set of refinements rather than relying on an extensive refinement path.

The Process of Interactive Search

As outlined in the introduction, one of our primary purposes has been to map the nature of the cognitive processes that people employ as they engage in the search task as it is structured by each of the engines. We have two measures—time in state and cognitive load—that we can use to provide insight.

In the first experiment, subjects spent most of their time perusing documents. The focus on documents was particularly pronounced in the Google engine where over 60% of the time was spent in this state. However, when search was self-terminated, subjects were prepared to sacrifice the time they spent looking at documents to maintain the amount of time they were spending on the home page and in the refinement and directory states. One possible explanation is that subjects start the search process by attempting to assess the scope of the information domain. Once this assessment is complete they may then feel confident to evaluate documents within this domain. Because mapping the information domain will take the same amount of time, regardless of the search conditions (fixed time or self-terminated), the states that reflect this process (refinement and directory) will be unaffected, and any additional time available will be spent on the evaluation of additional documents in a quest for more relevant pages to bookmark.

We have employed cognitive load in three experiments. Dennis et al. (1998) found a significantly lower cognitive load in the refinement state in comparison with the document summary state. The difference was detected via miss rate. The experiments used the Excite search engine, which does not use query-biased summaries. As a consequence, it is possible that the user is more heavily loaded in the document summary state, as significant processing is sometimes required to determine how the query terms are related to the document summary being perused. The first experiment in this article shows the opposite result. This time the cognitive load is significantly higher in the refinement state in comparison with the document summary state. The difference manifested via reduced reaction times—no difference in miss rate was detected. Google employs query-biased document summaries, which might offer part of an explanation for this apparent reversal of results. In the second experiment, which involved both self-terminated search, and ranking of query refinements, no significant results were detected in cognitive load. In short, there is insufficient evidence to conclude if there are significant variations of cognitive load across the states during the search process.

Evaluating Cognitive Load as a Measure for Search Engine Assessment

Given the disparity of these results, what can be concluded about cognitive load and its deployment in an experimental setting? Cognitive load has the primary advantage that it is a measure amenable to statistical tests of significance. In this sense, it follows in the tradition of recall and precision, which are also "hard" measures. Cognitive load also has the advantage that it can be measured in various search states and can provide insight into what the user is experiencing during the search process. Compared to measuring recall and precision, the measurement of cognitive load involves extra experimental infrastructure, for example, head phones and specialist software;[5] however, once the infrastructure is in place, it is straightforward to run the experiments and analyze the results.

Cognitive load should be used carefully from a methodological point of view. We feel that the digit monitoring task was sufficiently taxing, without being overly taxing, but does not generate a datum per impression. Consequently, we needed several trials to produce enough data for statistical analysis. This was not so much of a concern within the 5-minute duration used in the terminated search experiment, but became more of an issue in experiment two in which searches could be very short. In hindsight, we could have used a dual task of the form "is the digit divisible by three," which generates a datum for measuring cognitive load on every impression. In short, further experimentation is required to determine the worth of cognitive load as an evaluation measure.

Conclusions

In conclusion, directory-based search using Yahoo does not seem to offer increased relevance over keyword-based search (with or without query formulation assistance), and also takes longer. Query reformulation using the HiB can significantly improve the relevance of the documents through which the user must trawl versus unassisted keyword search using Google, and seems particularly effective for information needs for which the formulation of suitable query terms is difficult. However, the improvement in document relevance comes at the cost of marginally increased search time and possibly increased cognitive load when perusing query refinements.

Our experiments reveal that users prefer to use queries of about three terms in length for retrieving document summaries, even in the presence of query refinements that are longer. This has implications for the design of query refinement mechanisms for assisting Web search. Producing query refinements of longer than three terms would seem unnecessary. It is the quality of the terms in the query refinement that promotes relevance, not their length.

Cognitive load can be measured in different states of the search process, and thus has the potential for producing a more fine-grained analysis. Our experiments did not reveal many differences across search states, and no differences across search mechanisms. Further experimentation is required to ultimately determine the worth of cognitive load as an evaluation measure.

Acknowledgments

The work reported in this article has been funded in part by the Australian Research Council, Key Center for Human Factors and Applied Cognitive Psychology at the University of Queensland, and part by the Cooperative Research Centers Program through the Department of the Prime Minister and Cabinet of Australia. We gratefully acknowledge the research assistance of Rohan Clarke, Bernd Irmer, Israel Keys, Naomi Norman, and Sonia Twigg.

References

Anick, P.G., & Tipirneni, S. (1999). The paraphrase search assistant: Terminological feedback for iterative information seeking. In Proceedings of the 22nd annual international ACM SIGIR Conference (SIGIR'99) (pp. 153–159).

Baddeley, A. (1966). The capacity for generating information by randomization. Quarterly Journal of Psychology, 18, 119–130.

Bruza, P.D., & Dennis, S. (1997). Query re-formulation on the Internet: Empirical data and the hyperindex search engine. In Proceedings of the RIAO97 Conference—computer-assisted information searching on internet, Centre de Hautes Etudes Internationales d'Informatique Documentaires (pp. 488–499).

Bruza, P.D., & van der Weide, T.H. (1992). Stratified information disclosure. The Computer Journal, 35(3), 208–220.

Bruza, P.D., Dennis, S., & McArthur, R. (2000). Interactive Internet search: Keyword, directory and query reformulation mechanisms compared. Proceedings of the 23rd annual international ACM SIGIR Conference (pp. 280–287).

Dennis, S., McArthur, R., & Bruza, P.D. (1998). Searching the World Wide Web Made Easy? The cognitive load imposed by query refinement mechanisms. In: Proceedings of the third Australian document computing symposium (ADCS'98), Department of Computer Science, University of Sydney, TR-518 (pp. 65–71).

Jansen, B.J., & Pooch, U. (2001). A review of Web searching studies and a framework for future research. JASIST, 52(3), 235–246.

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

Michon, J.A. (1966). Tapping regularity as a measure of perceptual load. Ergonomics, 9, 401–412.

Moldovan, D.I., & Mihalcea, R. (2000). Using WordNet and Lexical operators to improve Internet searches. IEEE Internet Computing, 4(1), 34–43.

Posner, M.I., & Boies, S.J. (1971). Components of attention. Psychological Review, 78, 391–408.

Posner, M.I., & Klien, R.M. (1973). On the functions of consciousness. In S. Kornblum (Ed.), Attention and performance IV. New York: Academic Press.

[5] We used the PRISM browser, which is freely available for research use: contact s.dennis@humanfactors.uq.edu.au.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a very large Web search engine Query Log. SIGIR Forum, 33(3), 1999.

Spink, A. (1994). Term relevance feedback and query expansion: Relation to design. In proceedings of the 17th annual international ACM SIGIR conference (SIGIR'94) (pp. 81–90).

Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). Searching the Web: The public and their queries. JASIST, 52(3), 226–234.

Wickens, C.D. (1992). Engineering psychology and human performance. New York: Harper Collins.

Appendix A: Queries Used in Experiment One

1.1 Find pages listing jokes referring to Monica Lewinsky.
1.2 You are planning to move to Florida. Find pages listing jobs in the Florida area.
1.3 Find pages containing women's wave surfing competition results over the last 2 years.
1.4 Find pages about dyslexia.
1.5 Find pages that discuss clothing sweatshops.
1.6 Find pages that describe current or planned explorations or scientific investigations of Antarctica.
2.1 You own a personal computer that runs Windows '95. Find pages describing software that will test if it is Y2K compliant.
2.2 Find pages from which you can buy a pair of running shoes (online or at an address provided by the page).
2.3 Find pages that inform you which drugs are used to treat depression.
2.4 Find pages that discuss the disposal of long-lived radioactive wastes.
2.5 Find pages that discuss in vitro fertilization.
2.6 Are there any reliable or consistent predictors of mutual fund performance?
3.1 Find recipes for different varieties of carrot cake.
3.2 Find prices of Toshiba notebook computers.
3.3 You want to go skiing in Europe. Find pages describing a package holiday.
3.4 Find pages that discuss the concerns of the United States government regarding the export of encryption technology.
3.5 What makes Deep Blue capable of beating a human chess player?
3.6 Find pages that provide information regarding traveling in India.

Appendix B: Additional Queries for Experiment Two

1.7 Identify instances of attacks on humans by Africanized (killer) bees. Relevant documents must cite a specific instance of a human attacked by killer bees. Documents that note migration patterns or report attacks on other animals are not relevant unless they also cite an attack on a human.
1.8 Find accounts of selfless, heroic acts by individuals or small groups for the benefit of others or a cause. Relevant documents will contain a description of specific acts. General statements concerning heroic acts are not relevant.
1.9 What counterfeiting of money is being done in modern times? Relevant documents must cite actual instances of counterfeiting. Anticounterfeiting measures by themselves are not relevant.
1.10 Find information on shipwreck salvaging: the recovery or attempted recovery of treasure from sunken ships. A relevant document will provide information on the actual location and recovery of treasure; on the technology that makes possible the discovery, location, and investigation of wreckages that contain or are suspected of containing treasure; or on the disposition of the recovered treasure.
2.7 In what ways have quilts been used to generate income? Documents mentioning quilting books, quilting classes, quilted objects, and museum exhibits of quilts are all relevant. Documents that discuss AIDS quilts are irrelevant, unless there is specific mention that the quilts are being used for fundraising.
2.8 Do any countries other than the United States and China have declining birth rates? To be relevant, a document will name a country other than the United States and China in which the birth rate fell from the previous year. The decline need not have occurred in more than 1 preceding year.
2.9 Find ways of measuring creativity. Relevant items include definitions of creativity, descriptions of characteristics associated with creativity, and factors linked to creativity.
2.10 What is the status of the Three Gorges project? A relevant document will provide the projected date of completion of the project, its estimated cost, or the estimated electrical output of the finished project. Discussions of the social, political, or ecological impact of the project are not relevant.
3.7 What is the impact of poaching on the world's various wildlife preserves? A relevant document must discuss poaching in wildlife preserves, not in the wild itself. Also deemed relevant is evidence of preventive measures being taken by local authorities.
3.8 What are new methods of producing steel? Relevant documents will discuss the processes adapted by entrepreneurs who have organized so-called "minimills," and are producing steel by methods that differ from the old furnace method of production. Documents that identify the new companies, the problems they have encountered, and/or their successes or failures in the national and international markets are also relevant.
3.9 What legal actions have resulted from the destruction of Pan Am flight 103 over Lockerbie, Scotland, on December 21, 1988? Documents describing any charges, claims, or fines presented to or imposed by any court or tribunal are relevant, but documents that discuss charges made in diplomatic jousting are not relevant.
3.10 Find information on the use of dogs worldwide for law enforcement purposes. Relevant items include specific information on the use of dogs during an operation. Training of dogs and their handlers are also relevant.