Searching the World Wide Web Made Easy? The Cognitive Load Imposed by Query Refinement Mechanisms

Simon Dennis
School of Psychology; University of Queensland
Brisbane, 4072, Australia
s.dennis@psy.uq.edu.au

Robert McArthur
Distributed Systems Technology Centre, University of Queensland
Gehrmann Laboratories, Brisbane, 4072, Australia
mcarthur@dsrc.edu.au

Peter Bruza
School of Information Systems, Queensland University of Technology
GPO Box 2434, Brisbane, 4001, Australia
bruza@icsis.qut.edu.au

Abstract

This article addresses the effectiveness of search reformulation using query refinement mechanisms on the Internet. Cognitive load was measured using a secondary digit-monitoring task. The load was found to be lower when using the refinement mechanisms than when perusing document summaries - suggesting that the development of refinement mechanisms can make Internet searching easier. Two refinement mechanisms, one based on statistical term co-occurrence and the other on a shallow natural language parsing technique were tested. No difference in load was found, possibly because of the limited time that subjects spent in the refinement process.

Keywords: information retrieval evaluation, internet searching.

Introduction

Short queries on the WWW result in large, imprecise result sets. If longer queries could be elicited from the user, then the precision in the retrieval could be improved. For this reason a number of query formulation aids have appeared in conjunction with web-based search engines. For example, the AltaVista and Excite search engines produce lists of keywords that the user can use to make the initial query more specific. Similarly, the Hyper Index Browser\(^1\) (HIB) provides the user with a list of phrases that include the initial query terms and that can be used in subsequent calls to the retrieval engine.

The challenge for query refinement mechanisms is to provide an interface to the search space that makes searching easy and productive for the user through the careful selection of information to present to the user. In this paper, we intend to address two key issues.

Firstly, what are the merits of remaining in the refinement processes rather than perusing document summaries? The commitment to query refinement mechanisms hinges on the assumption that it is easier and more productive for users to process the refinements than to peruse a list of document summaries. Perhaps, however, there is no substantive difference between refinement and summary processing and the query refinement mechanisms are unlikely to lead to more effective search.

Secondly, how should a refinement mechanism choose candidate refinements for the users’ perusal? AltaVista and Excite produce candidate keywords for query refinement using term co-occurrence statistics. The refinements selected are words that tend to occur with the target terms. By contrast, the Hyper Index Browser (HIB) selects phrases that contain the initial query. For example, if the query is “Internet”, refinements like “Internet security”, or “guide for Internet” are presented. The second question we will address is whether the statistical or linguistic approach to refinement generation leads to more effective search.

In the discussion above, we have mentioned terms like “easy”, “productive” and “effective” in conjunction with search. There are a number of dependent variables we might use as measures of these factors. We could use the traditional

---

\(^1\) http://www.dsrc.edu.au/cgi-bin/RDU/hib/hib
information retrieval measures such as recall and precision. However, these measures have a couple of important limitations that should be noted. Firstly, our domain of interest is the World Wide Web where recall is difficult to measure accurately because it relies on identifying all of the relevant documents in a collection. Secondly, and perhaps more importantly, recall and precision tell us little about the demands placed upon the user while refining his or her query. We would prefer a measure that captures more closely the users search experience.

One possibility is to consider the amount of time that a user spends on a search task. Long searches might be considered less effective. 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.

Cognitive Load and Dual Task Methodology

There are a number of reasons why the direct measurement of search time may not accurately reflect the amount of effort employed by the user. Firstly, users are capable of employing greater cognitive resources in response to task demands (Norman & Bobrow [9], Eysenck & Keane [6], Wickens [10]). Provided the user can command ample resources to complete the search task, an inferior engine might require the user to apply effort which is not reflected in the amount of time they spend on the task. Secondly, two search mechanisms might provide different amounts of data to the user (Wickens [10]). The engines we will consider in this article (the HIB and Excite) typically generate different numbers and types of refinements. Extra time might be spent with the HIB because it generates a larger number of refinements, but each of those refinements may be easier to process requiring less effort from the user. Again, total time may not be sensitive to such differences.

For these reasons, we chose to employ the secondary task, or dual task technique. The dual task technique has a long history in the human factors literature (see Wickens [10] for a summary). The idea is to have subjects do two tasks simultaneously, to use any excess resources that might be available. The amount of effort they employ on the primary task (i.e. the Internet search) is inversely proportional to their performance on the secondary task.

A large number of secondary tasks have been employed, from finger tapping (Michon [8]), to random number generation (Baddeley [1]). After piloting, 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 and the miss rate.

When employing dual task methodology, it is important to keep in mind its limitations. Firstly, 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. Secondly, the secondary task should not precipitate a change in the strategy subjects use on the primary task. For instance, if it were the case that the digit-monitoring task caused subjects to process only a couple of refinements, where they would have processed many more had they not been required to perform the secondary task, then we are no longer measuring the task of interest. One of the primary purposes of this research is to determine the extent to which the dual task methodology and this particular secondary task are suitable for measuring search engine performance.

For the purposes of this study we assume that there is a relationship between cognitive load and retrieval effectiveness. Consider when a user has a high cognitive load, then it is likely that (s)he may miss potentially relevant information, thus affecting recall. The detailed nature of the relationship between cognitive load and retrieval effectiveness is a large research question. The study in this paper makes a small step in fleshing out this area.

Refinement versus Summary Load

Our first main hypothesis is that cognitive load is lower when the user is viewing candidate refinements (whether statistical or linguistic) for a query than when they are viewing document summaries. Document summaries typically include a list of candidate document titles with a small automatically generated precise of each page. For most queries, this list is long, often in excess of 1000 documents, and searching the summaries can be tedious. Furthermore, evaluating each of the summaries is difficult because the process of precis extraction is imperfect and is not customized to the user's information need. In contrast, the refinement mechanisms provide a relatively small number of candidate queries, and these candidate queries are typically only a few terms in length making them potentially easier to process.

Statistical versus Linguistic Refinements

The second of our main hypotheses is that linguistic refinements will decrease cognitive load and improve effectiveness of search as compared to statistically generated refinements. In this section, we will outline
how both statistical and linguistic refinements are generated.

Generating Statistical Refinements
If terms co-occur frequently, then there must be a relationship between them. This hypothesis is the basis of automatic thesaurus construction, but can also be used to produce candidate terms for query refinement. An often-employed technique is based on the vector space model. Given a query vector, term vectors are ranked according to similarity with the query vector. Those term vectors above a given threshold are deemed sufficiently similar to the query vector to warrant displaying as candidate terms for query refinement.

The Excite search engine uses the vector space model (Cutting [5]). For this reason it is reasonable to assume that Excite employs the above strategy (or variant thereof) to produce candidate terms for query refinement and we used Excite to generate the statistical refinements to be used in the experiment.

Generating Linguistic Refinements
The Hyperindex Browser (HIB) is a query formulation tool which attempts to produce reformulations of a query as linguistically well formed expressions (Bruza [2]; Bruza & van der Weide [4]; Bruza & Dennis [3]). The user enters an initial query, which is passed onto a retrieval mechanism, and a summary set is generated. The documents in the result are not presented directly to the user but are analyzed using a shallow natural language parsing technique (see also Greffner et al. [7] for a related technique). Phrases, called index expressions, 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 "surfing in australia", "internet surfing", "wave ski surfing". The user can then select any of these refinements to become the new query.

The Research Browser

The Interface
For the purposes of measuring cognitive load and timing users when refining queries or perusing documents or document summaries, an experimental web browser interface was constructed.

There are six buttons prominent above the HTML window - Back, HIB, Excite, Bookmark, Start/Finish and Stop (see Figure 1). The Back button returns the subject to the previous Web page. The HIB button takes subjects directly to the HIB starting page (see Figure 1), while the Excite button takes subjects to the Excite starting page. Each of these pages allows the subject to enter a set of search terms.

Clicking Search presents a page of refinements (see Figure 2). Subjects can choose to click on the focus (the terms that got them to the current page), click on a refinement, click on an enlargement, or type in new search terms.

![Hyper-index browser](image)

The Hyperindex browser is a tool to aid exploration, especially when you can't exactly prescribe what you are looking for!

Just enter a single broad term or phrase (no quotes)

database to browse: [WWW 3]
Enter topic of interest: [search]

Figure 1: HIB Start Page

Clicking on the focus link moves the user into the document summaries - the focus is used as a query and the resultant document summaries are presented to the user. In this experiment, the Excite search engine was employed to retrieve document summaries. Activating a refinement or enlargement link modifies or replaces the current query under focus with the refinement or enlargement that was selected. Refinements are modifications of the query which make it more specific. Enlargements broaden the query.

The Bookmark button saves the URL of the current page. The Start/Finish button starts and stops the timing of a query session.

Audio files are played to the subject at random intervals (3.5 seconds ± 1.5 seconds). The subject hears a voice saying a random number between one and five inclusive. The system forces a repeat to be played - that is, one digit played that is the same as the immediately preceding digit - every five digits played. Consequently, there was a maximum of 17.5 seconds between the playing of repeated digits.

When a subject clicks on any of the first three buttons, or on a link within the HTML page, all buttons except the Stop button are disabled until the entire page is loaded from the network, parsed, and is
displayed in the HTML window. Consequently, subjects cannot begin processing the page until it has been completely loaded, allowing us to distinguish when users are searching and when they are waiting for network transfers.

Figure 2: HIB Refinement Page

The Search Engine State Diagram
When using browser interface, the user is residing in one of four possible states, or is in transition between these states (see figure 3).

Figure 3: The search engine state diagram.
The possible states and their transitions are:

Start State In this state, the user enters an initial query. This always leads to a refinement state when using the Excite browser, but can lead to both a refinement (see Figure 2), or summary state in the case of the HIB browser. The HIB automatically transports the users to the document summaries when it cannot generate any refinements of the query description under focus.

Refinement State In this state the user sees a focus representing the current query description and candidate refinements of this focus, each of which can be selected to make the current focus more specific. The HIB also has enlargements, which allow the focus to be broadened. (This facility was not employed during the experiments). The current focus can also be selected. This results in the focus being used as a query to return document summaries. This event moves the user to the summary state. Alternatively, if the user clicks on a refinement generated by the HIB that cannot be further refined the user is transported automatically to the document summaries retrieved by using the activated refinement as query.

Summary State In this state, the user peruses document summaries retrieved from the Excite search engine, which consist of a title, short abstract, and links to the actual documents. The link can be activated to transfer the user to the document state.

Document State The state in which the user is perusing a web document. If found 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. In addition, the start 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. Figure 3 depicts the search engine state diagram. It provides the conceptual framework for interpreting and discussing the experimental results.

Experiment Subjects and Design
Ten subjects, two female and eight male, participated in the experiment for payment. In general, the
subjects had substantial computing experience. Seven use a computer daily and all had been using computers for more than a year. Similarly, most of the subjects had substantial experience on the Internet. Six used the Internet daily and only two indicated that they did not use the Internet on a regular basis. None of the subjects had used the HIB previously, but five had used the Excite search engine. When asked about their experience with the engines after the experiment, four preferred the Excite engine, three the HIB and three were uncertain. It should be noted, however, that both engines used the Excite summary mechanism and some subjects commented that had found it difficult to recall which engine they were using for some questions.

A 2x2 factorial design was employed. The factors were the refinement strategy used (statistical/Excite or linguistic/HIB) and the search state (either refinement or summary). Both factors were within subjects.

Procedure and Materials

Each subject began with a tutorial on the use of the research Web browser. This tutorial explained each of the buttons on the interface (see the description above) and provided the subjects with practice at performing the dual tasks.

Then each subject attempted to answer eight questions (see appendix A). Half of the subjects used the HIB for the even numbered questions and Excite for the odd numbered questions, while the other half used the HIB for the odd numbered question and Excite for the even numbered ones. The questions were arranged in this way to control for possible order effects. To answer a question the subject bookmarked a page that they felt contained the relevant information.

While using the engine to answer a question, the subjects were required to monitor a series of digits, which they heard through headphones. The digits ranged from one to five and the subjects were instructed to click the right mouse button when they heard a digit repeated. Reaction times were measured from the start of the playing of a repeated digit to the right button click. Subjects were informed that the digit-monitoring task was important and they should endeavor to click the right mouse button as quickly and accurately as possible.

Finally, the subjects were asked to fill in the subject information questionnaire.

Results

Figure 4 shows the time subjects spent in the refinement and summary states as a function of the refinement strategy. Subjects took longer in the summary states than they did in the refinement states.

Figure 4: Reaction Time as a Function of Search State and Refinement Strategy.

Sampling of the subjects reaction times to the repeated digits was done randomly and consequently some cells were missing in the analysis. For this reason, it was decided to analyze each reaction time as a separate event (rather than doing it on a per subject basis) and to make comparisons between the HIB and Excite refinement reaction times and the refinement and summary state reaction times.

No difference was found between the HIB (M=1867 milliseconds) and Excite (M=1800 milliseconds) refinement reaction times F(1, 39)=0.029, p=0.865. Nor was there a difference between the reaction times as a function of refinement (M=1823 milliseconds) versus summary state (M=1885 milliseconds) F(1,186)=0.059, p=0.808. There was a significant difference between the transition reaction times and the states reaction times (Transition M=1283, State M=1847, F(1, 9)=21.986, p=0.001), indicating that reaction time was faster when processing requirements were low. Log reaction times were also examined because reaction time distributions tend to show substantial skew, which can undermine the assumption of the analysis of variance technique. However, the log reaction time analyses required no change to the interpretation, so the raw scores are reported.

In addition to reaction time, one can examine the number of times subjects fail to respond to a double
digit as an indication of cognitive load. The miss rate
was calculated as the number of misses divided by
the total number of double digits played (the number
of hits plus the number of misses). There were few
misses in the refinement state, so it was not possible
to compare the linguistic and statistic refinement
mechanisms. It was, however, possible to look at the
miss rate as a function of state (see figure 5).

![Graph of Mean Miss Rate as a function of state]

**Figure 5:** Mean Miss Rate as a function of State.
The bars indicate the standard error.

Using the Wilcoxon matched-pairs signed ranks
test there was a significant difference between the
miss rate during the refinement state and the
summary state ($T=3, p < 0.05$).

We were also interested in whether there would be
a difference in search outcomes as a function of the
refinement strategy. The mean number of
bookmarked items was 0.92 for Excite and 1.12 for
the HIB. This difference was not significant,
however, $F(1, 69)=0.784, p=0.379$. We also looked at
the total time spent on queries as a function of
refinement strategy. Again there was no difference
$F(1,69)=0.330, p=0.567$.

In an effort to determine to what extent subjects
were using the refinement engines we examined the
breakdown of the chains as a function of chain length.
On the majority of occasions (62%) subjects did not
refine their queries.

**Discussion and Conclusion**

In the introduction, we outlined two main
hypotheses. The first of these, that refinement
mechanisms reduce cognitive load, was supported.
This result provides a sound basis for further work in
the area. Refinement mechanisms can make searching
the Internet easier.

However, we were unable to find support for the
second hypothesis, that linguistic refinements induce
less load than statistical refinements. Our ability to
discern the load associated with statistical and
linguistic refinements was hampered by subjects
reluctance to use the refinement mechanisms.
Regardless of whether they were using the HIB or the
Excite refinements subjects spent on average less
than ten seconds considering refinements. After
entering an initial query, which produced a page of
candidate refinements, they often proceeded directly
to the document summaries. The users could not
ignore the refinements as these were displayed on a
separate page to the summaries, and in the case of
Excite (and on most occasions for the HIB) users
were required to view the refinements before moving
to the summaries.

One possible explanation is that the users were not
convinced about the benefits of refining. They may
have been unaware of the connection between a more
expressive query and precision in the retrieval result.
Also, it may be that the subjects have been influenced
by previous exposure to search engines that only
provide summaries. An interesting avenue for further
exploration is to test whether novice users or users
with additional training employ the refinement
mechanism.

This paper makes two main contributions. Firstly,
the current results suggest refinement mechanisms do
reduce the cognitive load experienced by the user.
Secondly, we have demonstrated the use of dual
task methodology in the assessment of refinement
mechanisms. The digit-monitoring task has been
shown to be appropriate as a secondary task and both
reaction time and miss rate are sensitive to load
variations while conducting internet search tasks.

**Acknowledgements**

The work reported in this paper has been funded in
part by the Cooperative Research Centers Program
through the Department of the Prime Minister and
Cabinet of Australia.

**References**

information by randomization. Quarterly Journal of
Psychology, 18, 119-130.

for Searching in Hypermedia. In A. Rizk and N.
Streitz and J. Andre, editors, Proceedings of the
European Conference on Hypertext - ECHT 90,
pages 102-122, Cambridge University Press.

formulation on the Internet: Empirical Data and the
Hyperindex Search Engine. In Proceedings of the
Appendix A: Questions
1) You are planning to move to Florida. Use the engine to find two pages listing jobs in the Florida area.
2) You need to replace a piece of pipe below the kitchen sink. Find two companies that could supply what you need. Bookmark their homepages.
3) Find three pages providing recipes for different varieties of carrot cake.
4) Find two pages listing indicators that a person has dyslexia?
5) Find a page that describes what happened to the singer Melanie who was popular in the 1960s.
6) Find one page describing each of the two major streams of Zen Buddhism in Japan (one page for each stream is sufficient).
7) Find a page that clarifies the term "wiking" in the context of world war two Germany.
8) Find a page listing the current rankings of male tennis players on the ATP tour.