

Towards Cognitive Optimisation of a Search Engine Interface

Kenneth Treharne

School of Informatics and Engineering
Flinders University of South Australia
PO Box 2100 Adelaide, 5001
treh0003@flinders.edu.au

Darius Pfitzner

School of Informatics and Engineering
Flinders University of South Australia
PO Box 2100 Adelaide, 5001
pfit0022@flinders.edu.au

David M W Powers

School of Informatics and Engineering
Flinders University of South Australia
PO Box 2100 Adelaide, 5001
David.Powers@flinders.edu.au

Abstract

Search engine interfaces come in a range of variations from the familiar text-based approach to the more experimental graphical systems. It is rare however that psychological or human factors research is undertaken to properly evaluate or optimize the systems, and to the extent this has been done the results have tended to contradict some of the assumptions that have driven search engine design. Our research is focussed on a model in which at least 100 hits are selected from a corpus of documents based on a set of query words and displayed graphically.

Matrix manipulation techniques in the SVD/LSA family are used to identify significant dimensions and display documents according to a subset of these dimensions. The research questions we are investigating in this context relate to the computational methods (how to rescale the data), the linguistic information (how to characterize a document), and the visual attributes (which linguistic dimensions to display using which attributes).

1 Introduction

Any search engine must make two kinds of fundamental decisions: how to use keywords/query words and what documents to show and how. Similarly every search engine user must decide what query words to use and then be able to interact with and vet the displayed documents. Again every web page author or designer makes decisions about what keywords, headings and link descriptions to use

to describe documents or sections of documents. This paper presents one set of experiments targeted at the choice of keywords as descriptive query words (nWords¹) and a second set of experiments targeted at explaining the effectiveness of the graphical options available to display and interact with the search results.

2 Term relevance and Cognitive Load

There is considerable literature on visualisation techniques and a number of experimental and deployed search engines that offer a visualisation interface e.g. kartoo.com and clusty.com. However, there is little research to establish the effectiveness of such techniques or to evaluate or optimise the interface. This is surprising given the many theories and studies that target memory and processing limitations and information channel capacity, including many that build on and extend the empirical results summarized in George Miller's famous Magical Number Seven paper (1956). It seems likely that for any search task visualization to realize optimal use of channel capacity, it should permit users to draw on their powerful and innate ability of pattern recognition.

Another important aspect that has never been properly evaluated relates to the question of "which words do people use to describe a document?" Techniques like TFIDF are used in an attempt to automatically weight words according to how important they are in characterizing a document, but their cognitive relevance remains unexplored.

¹ nWords is available at
<http://dweb.infoeng.flinders.edu.au>

Our work will enhance the document search process by improving the quality of the data the user filters while reducing machine processing overheads and the time required to complete a search task. It will also provide fundamental insight into the way humans summarise and compress information. The primary objective of this part of our research is to quantify the number of words a broad spectrum of people use to describe different blocks of text and hence the number of dimensions needed later to present this information visually. Our secondary objective is to enhance our understanding of the approaches taking by users during completion of search tasks. A third objective is to understand the choices a user makes in selecting keywords or phrases to describe or search for a document.

2.1 nWords Results

Results for the nWords experiments explaining use of terms as descriptors or queries using a web based scenario are presented in Table 1 (with standard deviations). It is not only instructive to compare the results across task conditions (with/without access to the text, with/without restricting terms to occur in text), but the difference between the sub conditions where subjects were asked for keywords that described the document (D) versus search query terms (Q) they would use.

Descriptor & Query Term Usage	Task1	Task2	Task3
Number of D's Used	3.79 \pm 3.34	4.02 \pm 2.39	4.98 \pm 3.35
Total Distinct Stems Used in Ds	8.22 \pm 7.40	7.27 \pm 4.71	11.80 \pm 10.63
Average Distinct Stems per D	2.40 \pm 1.83	2.14 \pm 1.64	3.02 \pm 3.46
Distinct D Stems in Top 10 TFIDF	1.79 \pm 1.62	1.85 \pm 1.52	2.62 \pm 1.91
Total Distinct Q Stems Used in Qs	3.28 \pm 1.78	3.81 \pm 1.99	3.85 \pm 1.89
Distinct Q Stems in Top 10 TFIDF	0.98 \pm 0.95	1.19 \pm 1.02	1.38 \pm 0.99
Q Stem / D Stem Intersections	1.83 \pm 1.58	2.52 \pm 1.69	2.70 \pm 1.59

Table 1: Statistics for nWords survey tasks (\pm standard deviation). Descriptors (D), query terms (Q). Task 1 Access to text, Task 2 No access, Task 3 Term must be in text.

From the results of Table 1 we can see that TFIDF is poor at ranking keywords and query words. For the full data pool, only 2.11 ± 1.74 of the top ten TFIDF terms are used in description which best describes a text, whilst only 1.19 ± 0.99 are used in the query building task.

3 Graphical Representations

Little research on perceptual discrimination of dynamic encodings can be found, but a few empirical studies have investigated the human

visual system's ability to detect movement. Motion is a visual feature that is processed pre-attentively. Motion cues can be detected easily and quickly. Overall dynamic encodings outperform static encodings for attracting attention.

3.1 Preliminary Results

Our Miller inspired experiments using web based applets varied properties of icons either statically or dynamically as show in figure 1. In general our preliminary results indicate that dynamic icons yield slower performances in relation to static vs. dynamic variation of any attribute other than size (see error bars in figure 1). Intuition tells us that some aspect of the performance slow down is due to the fact that a dynamically encoded feature requires longer identification time since the encoding is time based. However, we also report that a high degree of volatility or variance was observed in all dynamic conditions. Significant differences were observed between static encodings and their dynamic equivalents in all cases except feature size (and hue - we do not report a flashing hue condition currently).

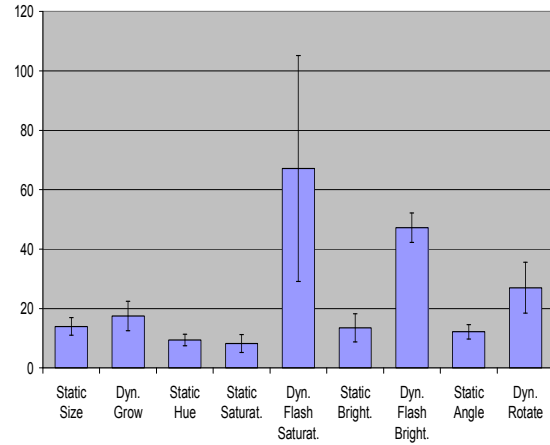


Figure 1 Aggregated subject response times (seconds) across 9 static and dynamic conditions.

An analysis over repeated sessions did not reveal any learning trend. However, we do note that significant outliers were regularly encountered at the beginning of all experiments indicating that first timers took a while to learn how to use the experiment software. Possible outliers in trials midway through the duration of some experiments may indicate that the subject was distracted during the search task. Experiment subjects. This is possible given that subjects administered the test in their own time.