Understanding the Relationship of Information Need Specificity to Search Query Length

Nina Phan
University of Melbourne
Melbourne, Australia
ninska123@hotmail.com

Peter Bailey
CSIRO ICT Centre
GPO Box 664, Canberra ACT
AUSTRALIA
peter.bailey@csiro.au

Ross Wilkinson
CSIRO ICT Centre
GPO Box 664, Canberra ACT
AUSTRALIA
ross.wilkinson@csiro.au

ABSTRACT

When searching, people's information needs flow through to expressing an information retrieval request posed to a search engine. We hypothesise that the degree of specificity of an IR request might correspond to the length of a search query. Our results show a strong correlation between decreasing query length and increasing broadness or generality of the IR request. We found an average cross-over point of specificity from broad to narrow of 3 words in the query. These results have implications for search engines in responding to queries of differing lengths.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation

General Terms

Experimentation, Human Factors

Keywords

Information need, search query length, IR request specificity

1. INTRODUCTION

People use search engines by expressing their information need as a (textual) search query – the information retrieval request. The search engine then returns a ranked list of documents. The person then reviews the list and browses to individual documents, or issues a new search.

From a search engine's perspective, a query may be the entire expression of a person's nuanced information need. But usually it is just a few words in length. To improve the quality of the ranking, it would be valuable to know if the specificity of the IR request and, by implication, the person's original information need is narrow or broad. We investigate only searchers' subjective assessment of specificity. An example broad information need might be "information about cows", and a corresponding IR request "cows".

Ingwersen and Järvelin characterise six dimensions to IR requests [2]; one of which is "generic vs. specific". In their critique of current IR research practices, they identify a need to gather further evidence on these dimensions and their effect on retrieval performance. They also characterise information needs in multiple dimensions, including the quality of current knowledge and knowledge state specificity. The specificity of an IR request is likely to be related to these two dimensions of the searcher's original information need, but it may be a complex relationship.

Ramírez and de Vries use the terms "broad" and "narrow" for these concepts while investigating request type as a contextual feature [4]. They performed manual classification of a number of information need descriptions, with a view to segmenting the request population for evaluating performance and user satisfaction.

Several researchers have examined query logs and length relationships. For example, Lau and Horvitz explore query refinement behaviour on large search logs, observing that queries become longer as query refinement occurs [3].

SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.
ACM 978-1-59593-597-7/07/0007.

2. EXPERIMENTS

We carried out three experiments to investigate our hypothesis that the specificity of an IR request is related to its search engine query length.

2.1 Experiments 1 and 2

For the first two experiments, participants were asked to read a search query, imagine a corresponding information need (for which they might have typed the query into a search engine), and then make a judgement on a 4-point scale ((very) narrow to (very) broad) about the information need's specificity. These instructions may fail to adequately separate specificity from other dimensions of the information need and subsequent IR request.

The first experiment used archived search query logs of a national government search engine (govsearch.australiagov.au). This dataset was chosen as least likely of the query logs we had available to be noticeably biased towards particular populations of users.

The queries were placed in buckets by length (1 ... 9, & 10+). A random selection of 100 queries was made, 10 from each bucket, and vetted to remove queries with mis-spellings and acronyms that participants might fail to understand. Random replacements were made until 10 queries per bucket were present. On registration, a unique subset of 50 queries was randomly and uniformly chosen for each participant (i.e. 5 per bucket).

Using a basic web form with four radio buttons to judge their specificity, participants were presented with their 50 queries (one at a time), displayed from their selection in random order to avoid any biases due to length.

1500 queries were judged by 30 participants. We combined

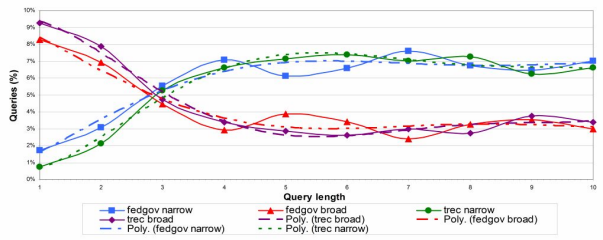

Figure 1: Expts 1 & 2 – Specificity vs query length

the counts of people voting for very broad and broad ratings, and for very narrow and narrow ratings. We used Pearson's Correlation to assess the association between specificity and query length. These were well above the minimum correlation coefficient of 0.463. Thus we may state that there is a statistically highly significant relationship (99% confidence level) between narrow/broad specificity and query length ($r_{narrow}(28) = 0.766, r_{broad}(28) = -0.766, \rho < 0.01$). The Kendall tau coefficients were narrow $\tau = 0.828$, 2-sided $p = 0.0015$; broad $\tau = -0.828$, 2-sided $p = 0.0015$.

Graphing the normalized results (Figure 1) with polynomial lines (order 5) of best fit we detect an intersection between broadness and narrowness at 3 words. To summarise, as query length increases, the corresponding information need is more likely to be perceived to be narrow.

The second experiment used the same method as the first. A random sample of 100 topics from a series of collections used in TREC Ad Hoc and Web tracks was used as the dataset; we also had fewer participants (16). The purpose was two-fold: to validate that the results obtained could be replicated, and to use a dataset that other groups without access to query logs could use too.

The minimum Pearson's correlation coefficient is 0.623 ($\rho < 0.01$). Again, the results calculated exceed this minimum value to indicate a correlation ($r_{narrow}(14) = 0.737$, $r_{broad}(14) = -0.737, \rho < 0.01$). Computing the Kendall tau coefficient confirms this result, though less strongly than in Experiment 1, and not at a 5% confidence level. (narrow $\tau = 0.449$, $p = 0.088$; broad $\tau = -0.449$, $p = 0.088$.) More noise was observed at large query lengths, which might smooth out with more queries or more participants. The normalized results are shown in Figure 1, and are similar.

2.2 Experiment 3

To address possible biases or flaws in the method of the previous experiments (for example, participants were judging queries from other people and had to imagine an information need), another experiment was undertaken. This experiment required participants to imagine their own narrow and broad information needs, and express these as search queries. A basic web form displaying text boxes (two for narrow, two for broad) was given to participants to fill in.

We had 30 participants, and placed the queries into buckets similar to the other experiments (1 . . . 5 & 6+ words). The minimum Pearson's correlation coefficient is 0.463 for $\rho < 0.01$ and 0.361 for $\rho < 0.05$. The Pearson's result for narrowness is $r_{narrow}(28) = 0.288$, and the Kendall tau correlation is $\tau = 0.298$, 2-sided $p = 0.546$. We conclude short queries are not correlated with narrowness. However, the Pearson result for broadness is above this minimum – $r_{broad}(28) = -0.846, \rho < 0.01$ and the Kendall tau correlation is $\tau = -0.867$, $p = 0.024$. Peaks of 2 and 4 words exist for broad and narrow.

From this, we see that queries are more likely to be broader as the length of a query decreases. The experiment also confirmed that users tend to use short queries in general, as shown in the workload distribution for the national government's large (16000+) query log (overlaid in Figure 2). In other words, uniform distribution of query lengths does not occur in real-life situations. We found 3 words to be the cross-over point again.

However, users may provide narrow queries at all query lengths. This result accords with general intuitions about web search – that people often have very narrow information needs (e.g. the home/best/known page finding task [1]) which can be expressed and satisfied with short queries.

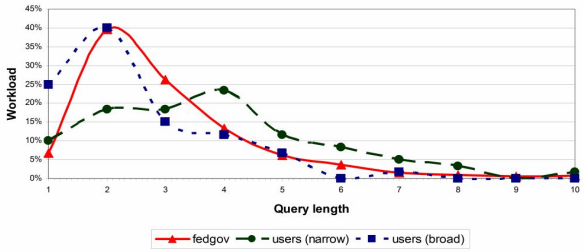

Figure 2: Expt 3 – Workload (percentage of queries)

3. CONCLUSIONS & FUTURE WORK

We hypothesised that there is a correlation between query length and the degree of specificity of a query. Our experiments partially validate this theory (for broadness), and indicate a cross-over of 3 words for queries being 2-3 times more frequently about narrow than broad information needs.

Deeper analysis of query characteristics to identify features which could be extracted to indicate narrowness in short queries is indicated. For example, detection of best page queries through anchor text matching. This would address the issue of short but specific queries. We plan to conduct experiments to investigate searcher preference for different result ranking when information needs are broad.

4. ACKNOWLEDGMENTS

We thank the many participants who carried out judgements, our colleagues for advice, and the reviewers.

5. REFERENCES

[1] N. Craswell, D. Hawking, and S. Robertson. Effective site finding using link anchor information. In *Proceedings of ACM SIGIR 2001*, pages 250–257, 2001.

[2] P. Ingwersen and K. Järvelin. *The Turn*. Springer, 2005.

[3] T. Lau and E. Horvitz. Patterns of search: analyzing and modeling web query refinement. In *UM '99: Proceedings of the seventh international conference on User modeling*, pages 119–128. Springer-Verlag, 1999.

[4] G. Ramírez and A. P. de Vries. Relevant contextual features in XML retrieval. In *Proceedings of the 1st international conference on Information Interaction in Context*, pages 95–110. ACM Press, 2006.