Encouraging Behavior: A Foray into Persuasive Computing

Elena Agapie
Harvard University
33 Oxford St, Cambridge, MA 02138
eagapie@seas.harvard.edu

Gene Golovchinsky
FX Palo Alto Laboratory, Inc.
3174 Porter Drive, Palo Alto, CA 94304
gene@fxpal.com

Pernilla Qvarfordt
FX Palo Alto Laboratory, Inc.
3174 Porter Drive, Palo Alto, CA 94304
pernilla@fxpal.com

ABSTRACT
Whereas longer queries have been shown to produce better results for information seeking tasks, people tend to type short queries. We created an interface designed to encourage people to construct longer queries, and evaluated it via an exploratory Mechanical Turk experiment. Results suggest that our interface manipulation may be effective for eliciting longer queries, but the effect is compromised when instructions to create longer queries are given.

Author Keywords
Interactive information seeking; query construction; persuasive computing

ACM Classification Keywords
H.5.m. Information interfaces and presentation: Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Exploratory search is a complex cognitive activity that relies on an iterative process to understand and refine the information need and to find pertinent documents. This iterative process consists of multiple steps, some of which involve running multiple keyword queries. The relevant literature shows that longer keyword queries are more effective at retrieving useful documents [2, 3], but it is also equally clear that people tend to create short queries [1].

Our intuition was that this propensity to create short queries could be mitigated through an appropriate interaction design. We created a visualization that reflects the length of the query being constructed in a pleasant, affective, manner to encourage people to create longer queries. The visualization creates a halo around the query input area the color of which depend on the length of the query being constructed.

We drew inspiration from persuasive computing [4, 6], a discipline that explores how to encourage desired outcomes in human-computer interaction. Our design varies visual characteristics of the text input area to create a more pleasing effect for longer queries, thereby implicitly nudging people to type more.

We tested the effectiveness of this visualization by running a Mechanical Turk study that asked people to find information on topics we constructed. Topics were designed to require exploratory search because answers could not be found with a single query. Instead, people were expected to use one set of queries to find one key piece of information, and another to combine it with another. Our between-subjects design controlled the presence or absence of the visual effect, and the presence or absence of an instructional message suggesting that longer queries were more effective.

The experimental results suggest that in the absence of any instructions, people exposed to the visualization typed longer queries than those who did not see it, but the effect was reduced when the instructional message about longer queries being more effective was shown. These are interesting and encouraging results: the interface clearly works as we had expected, but seems to be disrupted by the explicit instructions we provided. We are continuing this line of research to explore the phenomena further.

USER INTERFACE
Our interface design creates a halo around the query text box that varies in color and size with the length of the query being constructed. We chose the halo as a way of giving feedback to users based on several considerations:

1. It is a familiar interface element used to achieve pleasant artistic effects on many web sites. The current version of Twitter, for example, adds such an effect to its input area when it gains focus.
2. It does not compete for attention with the query construction task, as a text-based cue might.
3. It is visually unobtrusive and can be integrated with a variety of interfaces.

The initial state of the text box is shown in Figure 1: a soft pink halo with a radius of about 20 pixels surrounds the text box. As the user starts to type a query, the halo gradually shrinks in size (Figure 2) and becomes progressively less pink. After the query reaches a certain minimum length, the interface settles on a cooler, bluish tone (Figure 4).

Proceedings of HCIR 2012, Cambridge MA, Oct 4-5, 2012

Figure 1. Empty query box.

Figure 2. As the person starts to type, the halo changes.

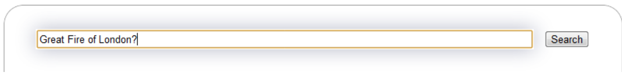

Figure 3. A longer query with a bluer halo.

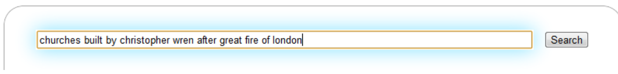

Figure 4. A long query with a bluish halo.

In our web-based implementation, we used the box-shadow CSS property to set the color and size of the halo. We interpolated the color between the two extremes the normalized query word length, which was computing by mapping from $[0, min(length(query), 7)]$ to [0,1].

Non-determinism

While it trivial to compute the query length in words and to set the halo to the associated color, we thought that the sharp transitions at word boundaries would make it easier for people to discover the relationship between the color and what they were typing, and then to ignore the mechanistic effect.

Instead, we chose to mask it somewhat by animating the change through a variable duration (0-1000 milliseconds). The goal was to generate a correlated visual effect that is non-trivial to understand based on casual observation.

EXPERIMENT

While a thorough evaluation of this interface widget would require long-term use by many people, we wanted to test the feasibility of approach quickly to see if it has merit before proceeding to a longer and more thorough evaluation. We built a small search interface on top of the Bing search API and deployed it through Mechanical Turk.

Search tasks

Modern search engines have made fact-finding quite efficient by crowd-sourcing relevance: users' click-through rates are used to adjust the ranking of documents to generate results that are more likely to be useful. This works well for common information needs, but becomes less reliable in the long tail, when there is insufficient information on the relative utility of relevant documents. Thus, when looking for more obscure, uncommon information (as typical of exploratory search), users have to rely on longer queries.

We wanted to simulate this kind of information seeking in a controlled experiment. To this end, we used a-google-a-day (http://agoogleaday.com) search puzzles to create our search tasks. These puzzles are designed to encourage people to learn how to solve complex information needs, and thus were good proxies for our experiment. We selected three older puzzles (from April 2012) to reduce the chances of our participants recognizing them. The following three search tasks were used in the experiment:

1. How many churches were built by the architect of the world's first triple-domed cathedral (and his office) after the Great Fire of London?
2. What tree does a mammal with fingerprints (primates excluded) rely on for food?
3. What material fuses with lime and soda to create an item on your dinner table that's considered to be an amorphous solid?

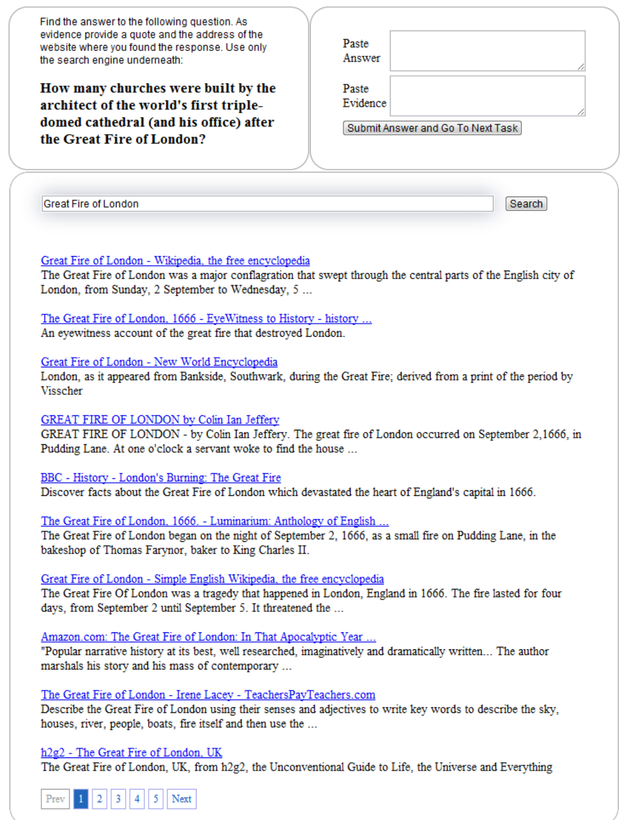

Figure 5. Experimental interface

Experimental software

Our search system included a web-based interface for eliciting queries, for presenting search results, and for collecting the answers to the search tasks. In addition to reporting the answer to the search question, we wanted people to record where or how they found the information. The goal was to discourage people from spamming our experiment and from having them fill in the answer as a guess or based on prior knowledge.

Queries were executed using the Bing search API[1] and results were filtered to remove any reference to the terms

---

[1] https://datamarket.azure.com/dataset/8818F55E-2FE5-4CE3-A617-0B8BA8419F65

'google' and 'google a day.' We also discouraged people from running the question as a query because we wanted them to actually find the answer rather than discovering someone else's google-a-day solution.

Prior to performing data analysis, we removed from consideration queries that repeated the problem statement verbatim (or nearly so), as we were interested in learning about people's willingness to construct queries by typing into the text box.

Before starting the experiment, participants were shown the following instructions:

> We are testing the performance of a new search engine we developed. To test our search engine we will ask you to use it to answer search puzzles. Do not use other search engines or resources other than the ones provided by the search engine results to find the answer to the task.
>
> You can only use the search engine we make available to you in the content of the hit.

Experimental design

Our experiment had a two-factor between-subjects design. The factors were the presence or absence of a halo (*Halo* condition) on the search box and the presence or absence of a statement (*Instruction* condition) following the experimental instructions telling participants that "our system performs better with longer queries." Each subject performed three search tasks in random order in a randomly-assigned experimental condition. Subjects were paid $1.55 through Mechanical Turk upon completion of all tasks. We restricted participants to be based in the United States and required them to have a 98% or better HIT completion rate. Because we were interested in queries people created, we removed 91 queries that were copies of task questions or duplicates submitted within a second of each other; analysis of the remaining 451 queries follows.

Hypothesis

Our experiment was designed to test the hypothesis that the halo visualization around the query box would result in people typing longer queries. A second hypothesis was that telling people that longer queries resulted in better performance would also result in longer queries.

Results

One hundred participants started our experiment, but 39 did not complete it. Table 1 lists the breakdown of participants by experimental condition, the average number of queries by condition, and the average query length per condition. It's worth noting that the average query length for this experiment was considerably longer (5.1 words/query) than those typically reported for web searches [1], which range from two to four words. On average, the participants used 3.2 (SD=1.32) queries to solve the Church task, 2.4 (SD=1.62) to solve the Tree task 2.4 (SD=1.58) queries to solve the Material task.

Table 1. Performance by experimental condition.

Condition	N	No. of Queries		Query length	
		Mean	SD	Mean	SD
Total	61	2.7	1.55	5.1	2.75
Halo	30	2.7	1.78	5.4	2.90
No halo	31	2.6	1.30	4.8	2.58
Instruction	32	2.7	1.61	4.9	2.59
No Instruction	29	2.6	1.48	5.4	2.60
Halo, No Instr.	14	2.6	1.79	6.6	3.28
Halo, Instr.	16	2.8	1.78	4.5	2.23
No Halo, No Instr.	15	2.6	1.16	4.2	2.01
No Halo, Instr.	16	2.7	1.43	5.3	2.89

To test the hypothesis that a halo would result in people typing longer queries, we performed an ANOVA on the query length in words by our experimental factors. For the factor *Halo*, we found a statistically significant main effect ($F(1,447) = 5.1$, $p < 0.05$). For the factor *Instruction*, we saw a borderline-significant main effect ($F(1, 447) = 3.4$, $p = 0.064$). In addition, we observed a statistically-significant interaction ($F(1,447) = 41.7$, $p < 0.001$) between the two factors. These results support our hypothesis that a halo results in people typing longer queries, but do not provide sufficient evidence to support the instruction hypothesis.

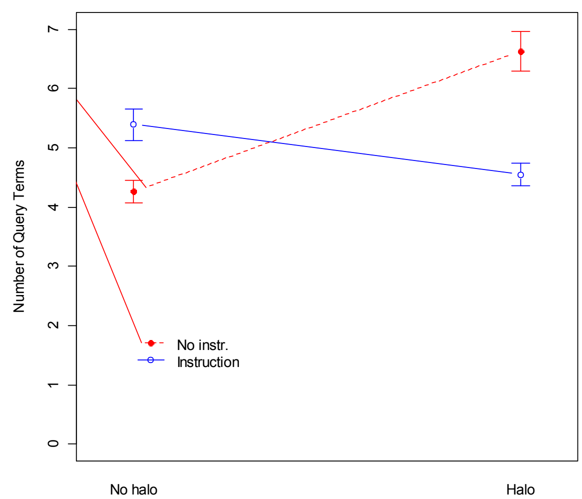

Figure 6. Mean number of queries for the factors Halo and Instruction. Error bars represent ± 1 standard error.

However, we also found a significant interaction effect between the factors *Halo* and *Instruction*. The interaction is illustrated in Figure 6. To further analyze the interaction effect, we performed a Tukey HSD *post hoc* test. The analysis showed that the *Halo with no instruction* condition outperformed all others. It had on average 2.1 words more than *Halo with instruction* ($p<0.001$), on average 2.3 more words than the *No halo with no instructions* condition ($p<0.001$), and finally, 1.2 words more than the *No halo with instructions* ($p<0.01$). These results further strengthen the conclusion from the main effect that a halo results in longer queries. We also found that the *No halo, Instruction* condition on average had 1.1 more words than the *No Halo, No instruction* condition ($p<0.01$). This result indicates that when no other factor is involved, people will enter longer queries with textual instructions.

No other comparisons in the Tukey HSD *post hoc* test show any significant results, although *No Halo, Instruction* produced borderline 0.8 more words than *Halo, Instruction* ($p < 0.055$).

DISCUSSION

What does all this mean? We can clearly reject the first null hypothesis, as the results show that the presence of the halo increases the query length by 0.56 words on average. The interesting, and surprisingly strong, effects come in the interaction between the halo and instruction conditions, where the combination of the instruction and the halo seems to reduce performance over just the halo or just the instructions. The exact reasons for this effect are unclear, and will need a follow-up experiment to investigate.

Our second hypothesis that a prompt would also encourage people to type longer queries was not borne out.

It is also interesting to note the body of work on shortening long queries to improve precision of search results (e.g., [9]). While shortening to improve query clarity [5] and coherence [8] may be useful, in many cases longer queries may be desirable either to improve recall or to refine results in topics with many documents.

NEXT STEPS

We have integrated this visualization into Querium [7], our search interface for CiteSeer, and will be conducting some longitudinal evaluations of these tools. The version integrated into Querium modifies not only the color of the halo, but also its size: as the query gets longer, the halo shrinks somewhat. Some further experiments will be required to fine-tune the color, size, and timing parameters.

We will also be running follow-on studies to explore the nature of the interaction effect we observed. We are also planning to run a similar study that offers real-time textual prompts below the text box to encourage people to type longer queries. It will be interesting to compare the results of the different manipulations to gain a better understanding of people's reactions to such interface features.

It is worth noting that our choice of colors was culturally-specific, as we chose a reddish hue to represent an undesirable state. More research will be required to determine if colors are generally useful for this application, and which colors are best suited to convey desirability (and lack thereof) in different cultures. Finally, equivalent techniques suitable for color-blind users need to be investigated.

CONCLUSION

A number of challenges related to interactive information retrieval have been identified in the literature. Prominent among them is the notion that people are unwilling to create longer keyword queries, despite the fact that such queries can be effective at retrieving useful documents.

This research was motivated by the desire to shape searchers' behavior toward more constructive outcomes using techniques from persuasive computing. As an initial exploration of this space, we created a novel interaction technique to encourage people to create longer keyword queries, and evaluated it with a Mechanical Turk experiment. The encouraging results of our evaluation suggest that this is a promising area for further exploration.

ACKNOWLEDGMENTS

We thank Tony Dunnigan for his suggestions about the design of the halo, and Thea Turner for help with some of the mechanics of data analysis.

REFERENCES

1. Bailey, P., White, R.W., Liu, H., and Kumaran, G. (2010) Mining historic query trails to label long and rare search engine queries. ACM Trans. Web, 4:15:1–15:27, 2010
2. Belkin, N.J., Cool, C., Kelly, D., Kim, G., Kim, J.-Y., Lee, H.-J., Muresan, G., Tang, M.-C., Yuan, X.-J. (2002) Query Length in Interactive Information Retrieval. In Proc. SIGIR 2003 (Toronto, Ont, Canada). ACM Press.
3. Belkin, N.J., Cool, C., Jeng, J., Keller, A., Kelly, D. Kim, J., Lee, H.-J., Tang, M.-C., Yuan, X.-J. (2002) Rutgers' TREC 2001 Interactive Track Experience. In Proc. TREC 2001, pp.465-472. Washington, D.C.: GPO.
4. Communications of the ACM, March 1999.
5. Cronen-Townsend, S., Zhou, Y., and Croft, W. B. (2002) Predicting query performance. In Proc. SIGIR '02, pp. 299–306, New York, NY, USA, 2002. ACM Press.
6. Fogg, B.J. (1998) Persuasive Computers: Perspectives and Research Directions. In Proc. CHI 1998 (Los Angeles, CA). ACM Press.
7. Golovchinsky, G., Diriye, A., and Dunnigan, A. (2012) The future is in the past: Designing for exploratory search. In Proc. IIiX 2012 (Nijmegen, The Netherlands, August 2012). ACM Press.
8. He, J., Larson, M., and de Rijke, M. (2008) Using coherence-based measures to predict query difficulty. In Proc. ECIR 2008, pp. 689–694. Springer.
9. Kumaran, G. and Carvalho, V. R. (2009) Reducing Long Queries Using Query Quality Predictors. In Proc. SIGIR 2009, (Boston, MA), ACM Press.