Information Search: The Intersection of Visual and Semantic Space

Franklin P. Tamborello II and Michael D. Byrne
Rice University
Department of Psychology
6100 S. Main Street, MS-25, Houston, TX 77005 USA
+1 713 348 4856
{tambo, byrne}@rice.edu

Abstract
In the context of an information search task, does the visual salience of items interact with information scent? That is, do things like bold headlines or highlighted phrases interact with local semantic cues about the usefulness of distal sources of information? Most research on visual search and highlighting has used stimuli with no semantic content, while studies on information search have assumed equal visual salience of items in the search space. In real information environments like the Web, however, these things do not occur in isolation. Thus, we used a laboratory study to examine how these factors interact. The almost perfectly additive results imply that good information scent cannot overcome poor visual cues, or vice versa, and that both factors are equally important.

Categories & Subject Descriptors: Experimentation; Human Factors; Performance.

General Terms: Design; Experimentation; Human Factors; Performance.

Keywords: World Wide Web and Hypermedia; Information Architecture; Visual Design; Content Strategy / Content Creation.

INTRODUCTION
When searching through a website for information, users must typically navigate through a large body of text, as well as headlines printed in larger type, links to other pages that are typically printed in other colors, and likely also some graphics. Users typically have a goal formed about what information they are looking for. They then use semantic clues (information scent) in the document at hand to inform their decision about where to look for the sought-after information next. These semantic clues about where to search for information all occur within an area affected by the visual manipulations described above. But studies of visual search typically use stimuli without semantic content. Likewise, studies of semantic search tend to assume equal visual salience among stimuli. As the example above illustrates, these two tasks rarely occur in isolation in real life.

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.
ACM 1-59593-002-7/05/0004.

In such a situation, certain objects should pop out from the body of text in preattentive processing of basic visual features because they differ on one or another basic visual feature [13, 14]. But they also tend to differ in their implications for information scent (e.g., links to other pages). Intuition would say that a large, red headline that is a link and has a high degree of information scent should be clicked on. But what happens when links are highly visually salient, but low in information scent? Do they fool users into wasting time following a fruitless path? Do users notice cues with high information scent even if they are not visually salient?

Pirolli and Card [9] have put forth a theoretical framework, Information Foraging Theory (IFT), for understanding search through an information space. Information foragers use the semantic clues at hand to make educated guesses as to whether or not following a given path would be profitable in their search for information. The information in the cues used to guide this decision is termed information scent. Cues which are highly semantically related to the target have high scent, while those unrelated have low scent.

One method for assessing information scent is Latent Semantic Analysis (LSA). LSA is a general theory of and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large text corpus [7,8]. It operates on the principle that the aggregate of all the contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other [8]. LSA's ability to induce word meaning similarities both from co-occurrences and contexts in which a particular word does not actually appear makes it a powerful inductive knowledge tool.

Blackmon and colleagues [1, 2] have used LSA to assess the usability of websites. Their paradigm had participants read goal statements instructing them to find a particular web page, then navigate a web site in order to find the specified page. They found that participants had a higher percentage of correct first clicks when headings and links on the web site were better related to the goal statements as measured by LSA.

Furthermore, Blackmon et al. [2] developed a technique for predicting where users of a website would click during an information search task. Cognitive Walkthrough for the Web

(CWW) compares LSA scores for links and headings on a page to contextually rich descriptions of user goals. There were some links, "unfamiliar links", which CWW identified as likely being links that a user would have insufficient knowledge about to assess that link's relatedness to the user's search goal. There were also "goal-specific competing links". CWW indicated that these were not necessarily related to the target link, but were approximately as related to the goal statement as the target link. Blackmon et al. found that participants had a much lower rate of successful first clicks when a trial had unfamiliar links or goal-specific competing links than when it had neither. But CWW only examined semantic factors affecting website navigation. Visual features present in the web pages were restricted to links organized into regions within the web page topped by a heading. There was no manipulation of visual features.

On the topic of visual features, most investigators agree that certain fundamental features of vision (such as color, brightness, and movement) are processed in parallel relatively early in the visual pathway of humans [13,14]. Given a visual search task using many items, those items that can be distinguished by one of those basic features tend to "pop out" from the field of other stimuli. For example, when searching through a field of green T's, the time to find a red T remains the same no matter how many distractor green T's are present. If, however, the target item(s) can only be distinguished by a conjunction of features (such as a certain color and brightness combination), then visual search will be slow and effortful.

Highlighting by color is one way to apply the pop-out effect. Fisher and Tan [4] performed two experiments assessing the effects of highlighting types and validity (percentage of trials using highlighting in which the highlighting is or is not predictive of the target's location) on search times. Subjects searched for a target digit in a horizontal array of five digits. Fisher and Tan determined that highlighting by color can significantly speed search when the target, and not a distractor, is highlighted.

Donner et al. [3] conducted research on highlighting using more complicated displays. They discovered a benefit to valid highlighting on a display that was poorly formatted, but no time cost to invalid highlighting. However, for a display reformatted to better match task requirements, neither valid nor invalid highlighting was significantly different from no highlighting, preventing reliable attribution of the overall difference in response time to either cost or benefit of highlighting.

Perhaps closest in spirit to our study is the research of Pirolli, Card, and Van Der Wege [10,11]. They conducted studies to discern how information scent might affect visual search, though visual salience was not manipulated. They found that search performance was highly modulated by information scent. In low scent conditions, users dispersed their visual search over more nodes of the interface than users of the Microsoft Windows Explorer file browser. But in conditions of high information scent, users of both interfaces dispersed their visual search over narrow regions. Also, eye movements in high scent conditions were about 25% longer than in low scent conditions. But under conditions of poor information scent, density of items in the HTB adversely affects the visual search and navigation process. Though Pirolli and colleagues found evidence of information scent guiding visual search, all of their stimuli were equally visually salient. It would be informative to know how the visual search task changes when some items appear visually distinct from other items, as in highlighting.

Given these results, we wanted to create an information environment in which both visual and semantic factors were operating. In our experiment, participants read a brief description of a particular word. Their goal was then to search a matrix of words to find the one word that best matched the description they just read. Half the time none of the words in the word matrix were highlighted, but on the other half of the trials, some of the words were highlighted. Half of those times the "target" word was among those highlighted, half of the time they were not. We also manipulated the the degree of semantic relatedness between the target and the description.

METHOD

Participants

Forty-nine undergraduates at Rice University were recruited to fulfill experiment participation requirements, 28 of whom were female. Subjects' age ranged from 18 to 25 years, with a mean of 19.7.

Design

The experiment incorporated a within-subjects two by three factorial design. The factors were information scent (high and low) and highlighting condition: control (no highlighting present), salient (the target was among the highlighted items), and non-salient (the target was not among the highlighted items).

Stimuli and Materials

Words used in the experiment were chosen from seven- and eight-letter words obtained from the Frequency of English Words corpus [5] of the Oxford Psycholinguistic Database software application [12]. Target descriptions were paragraphs of 100 to 150 words. They were constructed such that the first few sentences defined the target word. The remaining sentences served as examples in which the target word could be used, perhaps in place of synonyms.

Information scent between target descriptions and target words was assessed using the application provided by the Science and Applications of Latent Semantic Analysis Group (SALSA) [6]. The college freshman text corpus was used as it most closely matched the knowledge level of the participants who were tested. The document-to-term, one-to-many comparison method was used to assess the relatedness of target descriptions to their matched set of targets and distractors. High-scent targets had an average cosine in relation to their target descriptions of 0.31, with 0.01 for low-scent targets. The average cosine of a set of a target and distractors for their respective target description was 0.02. Furthermore, all low scent targets had at least one distractor in its trial that had a larger cosine. A relatively high cosine is about 0.60, while the average cosine between unrelated words is ≈ .07 ± .04 [7].

The target and distractors for each trial were displayed in a six-word by six-word matrix, with 146 pixels (approximately 4.2 degrees visual angle at a seating distance of 25 inches from the computer monitor) intervening between the center of each word width-wise, and 110 pixels (about 3.2°) between the center of each word vertically. All text was displayed in 12 pt. Times font and was therefore each word was approximately 0.4° tall by 1.3° wide. All non-highlighted words were printed in black and highlighted words were printed in bright red.

Half of the experiment's trials had no highlighting (standard condition), while the other half were split evenly between validly highlighted trials (the target was among the highlighted items) and invalidly highlighted trials (the target was not among the highlighted items). Selection of non-highlighted items was random on each trial.

The experiment was programmed in Macintosh Common Lisp and run on Apple eMac computers. The experiment display area encompassed the computers' entire display (16 inches viewable) at a resolution of 1024 x 768 pixels.

Procedure

Ten practice trials and 72 test trials were administered. At the beginning of each trial, the computer displayed the target description until the participant clicked an on-screen button using the mouse. The computer then cleared the target description from the monitor and displayed the target and distractors matrix, as well as a mouse cursor. The location of the target word varied randomly from trial to trial with the condition that each position displayed a target word twice throughout the experiment. Trials were divided evenly among the six scent by highlighting conditions, as well as among the six test blocks.

Participants were instructed to click on the word that most closely matched the target description. Response times and accuracy were recorded, but no feedback was given for response time. Participants were, however, informed about incorrect responses with a beep and a pause. The pause penalty encouraged them to read the target descriptions more carefully on future trials. Once the participant had clicked the mouse button on a selected item, the screen turned blank until the participant pressed a button to indicate he was ready to read the next target description.

RESULTS

Generally, subjects were faster when scent was high versus low, and fastest for valid highlighting, then no highlighting, and slowest for invalid highlighting. The effect was additive such that bad scent added approximately 4 seconds to search time regardless of highlighting type. Bad highlighting also added about 4 seconds relative to good highlighting. See Figure 1 for the main results.

Response time and error data were analyzed using repeated measures ANOVA. Predictably, subjects found targets more quickly when information scent was high rather than low $F(1, 48) = 104.689$, $p < 0.001$. Valid highlighting lead to faster search times than did no highlighting, which in turn was faster than invalid highlighting $F(2, 48) = 46.854$, $p < 0.001$.

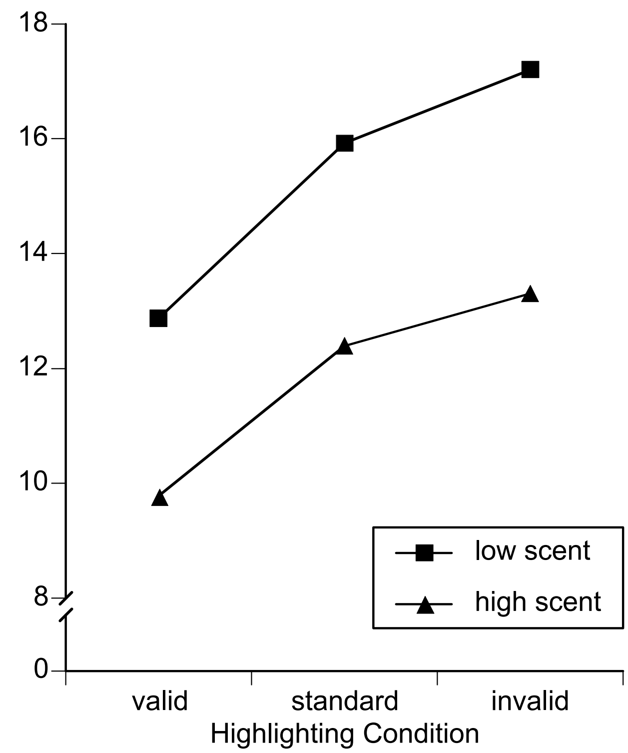

Figure 1. Response time as a function of scent and highlighting.

We also analyzed response time as a function of location of the target item, with row coded from top to bottom and column from left to right. Response time reliably increased as both row and column increased, linear contrast on row $F(1, 48) = 52.87$, $p < .001$ and linear contrast on column $F(1, 48) = 13.00$, $p = .001$. This is not surprising given that Westerners are accustomed to reading left-to-right, top-to-bottom.

The pattern of results for error rate was analogous to response time: subjects tended to perform better on both dimensions simultaneously. Subjects made more errors when information scent was low $F(1, 48) = 84.454$, $p < 0.001$. They also committed fewer errors when highlighting was valid than when there was no highlighting, which caused fewer errors than invalid highlighting. There was also an effect of target row such that subjects made relatively more errors when the target was located in the middle rows of the matrix, $F(1,49) = 9.458$, $p = 0.003$.

DISCUSSION

The substantial effects observed for highlighting and information scent reaffirm their influence in information search tasks. The complete lack of interaction between scent and highlighting on response times suggest that the effects are strictly additive. With a penalty of about four seconds each for bad highlighting and bad scent, the two combined could make it very difficult for a user to navigate an information environment. Indeed, as Figure 1 illustrates, the effects are so large that searches in the bad scent, bad highlighting condition took nearly twice as long to complete as searches in the good scent, good highlighting condition.

These results imply that good visual design and good information architecture (in terms of menu labels and structure) are independent and roughly equally important. Poor menu label choices cannot be fully compensated for by adding highlighting and vice versa. While this result may not appear particularly surprising we believe that this is the first time it has been empirically documented.

Similarly, the results further imply that it is possible that good design on both factors can save users substantial time. Using CWW or some IFT-informed analysis of link scent can indeed save users substantial time, and so can highlighting of likely targets. Note that our results show there can be some cost to highlighting the wrong items, but that the gains associated with correctly highlighting potential targets can be larger than the costs. Thus, if an analysis of users' probable goals indicates that a few links are more likely to be targets than others, highlighting them should, on average, benefit users.

The position effects should be taken to mean that prominent search cues, both highlighting and information scent, should be placed where users are likely to see them first: toward the top-left of a document (for Western users). This is already implemented to some extent as headlines tend to be placed above text bodies. But it might prove useful to provide a brief synopsis, list of links, or keywords at the top of the document that capitalize on and concentrate information scent and highlighting.

Future research in this domain might manipulate information scent on a finer scale than the "high" and "low" levels used in this study and might look at the effects of highlighting in search spaces of different sizes. With 36 possible choices, the displays used here are moderately large relative to current Web pages. Highlighting may be somewhat less effective in smaller domains and somewhat more effective in larger ones, depending on the proportion of the items that are highlighted.

ACKNOWLEDGMENTS

The authors would like to thank Alana Abernathy for her assistance in data collection and analysis.

REFERENCES

1. Blackmon, M.H., Kitajima, M., Polson, P.G. (2003). Repairing usability problems identified by the cognitive walkthrough for the web. *Proceedings of CHI 2003, 5*(1), 497 – 504.
2. Blackmon, M.H., Polson, P.G., Kitajima, M., Lewis, C. (2002). Cognitive walkthrough for the web. *Proceedings of CHI 2002, 4*(1), 463 – 470.
3. Donner, K.A., McKay, T., O'Brien, K.M., Rudisill, M. (1991). Display format and highlighting validity effects on search performance using complex visual displays. *Proceedings of the Human Factors Society 35th Annual Meeting,* 374 – 378.
4. Fisher, D.L., and Tan, K.C. (1989). Visual displays: The highlighting paradox. *Human Factors*, *31*(1), 17 – 30.
5. Kucera, H. & Francis, W.N. (1967). Computational analysis of present-day American English. Providence: Brown University Press.
6. *LSA @ CU Boulder.* (1998, October) Retrieved November 25, 2003, from http://lsa.colorado.edu
7. Landauer, T.K., and Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211 – 240.
8. Landauer, T., Foltz, P., and Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25,* 259-284.
9. Pirolli, P. and Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643 – 675.
10. Pirolli, P., Card, S., and Van Der Wege, M. (2001). Visual information foraging in a focus + context visualization. *Proceedings of CHI 2001, 3*(1), 506 – 513.
11. Pirolli, P., Card, S., and Van Der Wege, M. (2003). The effects of information scent on visual search in the hyperbolic tree browser. *ACM Transactions on Computer-Human Interaction, 10*(1), 20 – 53.
12. Quinlan, P. (1992). Oxford Psycholinguistic Database (Version 1.4) [Computer software]. York, U.K.: University of York.
13. Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*. 97 – 136.
14. Wolfe, J.M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2). 202 – 238.