# Information-Seeking Time: Only a Subset of Home Page Elements Matters

**Sara Rigutti*[1], Walter Gerbino[1], Carlo Fantoni[1,2]**
[1]University of Trieste, Department of Life Sciences, Psychology Unit *Gaetano Kanizsa*, Trieste, Italy
[2]Istituto Italiano di Tecnologia, Center for Neuroscience and Cognitive Systems, Rovereto, Italy
*srigutti@unit.it

**Abstract**

During goal oriented web navigation does the competition for web selection depend on all navigation options or only those options which are more likely to be functional for information seeking? Here we provide evidence in favour of the latter alternative. Within a representative set of real web sites of variable breadth, the time required to reach a goal located at the depth of two clicks from the home page is accounted for by *C*, an objective measure of the complexity of the start page, based on the number of links weighted by the number and type of embedding web elements. Our results demonstrate how focusing on links while ignoring other web elements optimizes the deployment of attentional resources necessary to navigation.

**Keywords:** web navigation; information-seeking time; complexity; layout; information scent; selective attention.

## Introduction

According to Pirolli & Fu (2003) Web Information-Seeking Behavior (WISB) is a matter of *information scent* (Pirolli & Card, 1999), a construct grounded on semantic rather than procedural knowledge, including an estimate of the relevance of all links within a web page. The Comprehension-based Linked model of Deliberate Search (CoLiDeS; Kitajima, Blackmon, & Polson, 2000) takes a step further: it assigns information scent for a user goal not only to links but also to sub-regions within a web page, according to the user's prior knowledge of layout conventions. Users select one patch of information within the web page and ignore all the rest, through a mechanism of *selective attention* that optimizes the deployment of cognitive resources. Real web sites actually include not only labels and text, but also different types of web objects (navigation bars, canvas, footers, embedded links) placed in different page regions (top, bottom, central, left, right) to display, group, and emphasize the relevance of labels. Experimental evidence points to the existence of several effects of web page layout on WISB; e.g., selection time increases as the target label deviates farther from the top-left region of the web page (Rigutti, Gerbino, & Fantoni, 2007; van Schaik & Ling, 2001; Tamborello & Byrne, 2005).

Despite different perspectives on WISB, scientists agree in considering web page complexity as a valuable predictor of information-seeking time (IST). Gwizdka and Spence (2006) defined the complexity of a web page in terms of the overall number of links within the page (i.e., navigation choices). Several researches corroborated the predictive power of such a definition by finding that an increase in the number of links per page decreases link selection accuracy and increases IST (Blackmon, Polson, Kitajima, & Lewis, 2002; van Schaik & Ling, 2012). A knowledge based approach to WISB, as formalized by CoLiDes, sets the basis for an alternative hypothesis: IST is accounted for by the complexity of the home page, defined by those only options and links which are more likely to be functional for information seeking, rather than by the entire set of web features which are detectable within the page (i.e., words, images, links, segregated graphical regions).

### A knowledge based measure of web complexity

Here we provide and test a new measure of web page complexity (*C*), by formalizing three basic constraints at the basis of a knowledge based approach to WISB: [1] only those navigation links (NLs) within relevant web elements are encoded for web page complexity; [2] NLs are weighed according to the weight of the web element they belong to, and [3] according to the number of relevant web elements within the web page. In other words, the complexity *C* of any NL depends on the type and number of web elements embedding it. According to evidence on selection time within hierarchical menu trees (Lee & MacGregor, 1985) we assumed that IST increases as *C* increases.

Our experiment was designed to test the following hypotheses:

(H1) the time required to search for an information item located at a given depth (starting from the home page) does not depend on the simple number of NLs or web elements (NE) they belong to, but rather on their weighted combination as formalized by *C*: any independent and combined effect of NL and NE on IST should be entirely accounted for by our synthetic *C* measure;

(H2) our *C* index explains a larger amount of variance of the IST distribution than alternative metrics of web page complexity that are not intended to formalize selective attention processing of the page (e.g., overall number of links, number of segregated graphical regions, number of words, number of images).

## Method

We tested our expectations by measuring the *IST* for a representative set of web sites extracted from the population of small and medium enterprises' sites of our regional district (Regione Autonoma Friuli Venezia Giulia). The visual complexity of the home page of selected sites displayed a large variability. Real web sites were used to

provide the experiment with high ecological validity. The entire set of web sites was downloaded in a local directory; then, the XAMPP platform was customized to store log files and extract the navigation paths as well as the timing associated to each web-page click with millisecond precision. Each web site was displayed and navigated using Google Chrome Browser.

The experimental navigation session consisted in the sequential presentation of trials in a randomized order. In each trial the participant was asked to perform an information-seeking task on one of selected web sites. To keep the structural complexity of the task constant, we selected all information goals two clicks away from the home page (a depth constraint related to information architecture) and all end-target items (i.e., links to the page displaying the information goal) in the central region of the intermediate page (a visuo-spatial constraint).

In the home page, we avoided possible biasing effects of link visibility by selecting target items included in the page portion displayed at the onset; i.e., without scrolling. Furthermore, all target items were superordinate words, as needed to balance their semantic access.

Twenty students of the University of Trieste participated in the experiments.

## Results and discussion

To test our main hypothesis that IST is affected only by a subset of web elements (i.e., those relevant for information seeking), we calculated the empirically grounded knowledge based measure of complexity *C* for each web site. Then, we analyzed valid individual ISTs using a step-wise procedure that contrasted linear mixed-effect (*lme*) models of increasing complexity (Bates & Sarkar, 2007), depending on the number of fixed effects, modelled by our candidate continuous predictors (*C, NL, NE* and/or standard complexity indices) and their combination.

### Test of H1: Knowledge based *C* vs. NE and NL

To investigate how structural elements of a web page relevant for information seeking can determine the speed of web information search, we analyzed the independent and conjoint effects of the number of navigation web elements (NE) and links embedded within them (NL). In a first *lme* model we disregarded *C* and found that individual IST was positively affected by NE and NL. Then, we repeated the same analysis including *C* as a third independent covariate, to control for its effects. The main effects of both NE and NL became non significant when *C* was included: the likelihood of IST was thus completely explained by *C*. In other terms, the present investigation provided no evidence that navigation choices and web elements contribute by themselves to the perceptual response beyond what *C* can explain.

Furthermore, a model of IST with *C* as the only covariate finely describes the IST metric obtained in our experiment: the best linear fit describing the relationship between average predicted and average observed search times was indeed a line with unitary slope and null intercept.

### Test of H2: Knowledge based *C* vs. standard complexity indices

To support our second hypothesis we tested an *lme* model inspired by previous work (Michailidou, Harper, & Bechhofer, 2008), which provided us with a compound measure of home page complexity based on a weighted linear combination of several metrics of web page complexity alternative to *C*, such as the overall number of links, bytes, segregated graphical regions, words, and images. The simplest *lme* model accounting for the largest amount of variance amongst all candidate models resulted to be one including the independent and conjoint effects of the number of segregated graphical regions, and of the overall number of links. However, consistently with *H2*, by adding *C* as a covariate, both the main effects of these two metrics and their interaction became equal to zero. Again *C* resulted to be the only significant predictor of IST.

In summary, we obtained two findings: (a) consistently with *H1*, IST was accounted for by only the subset of navigation elements whose effectiveness is formalized by *C* (a weighted combination of web options and objects available in the start page that, according to the user's knowledge of layout conventions of web sites, are more likely to be relevant for the achievement of information goals); (b) consistently with *H2*, all other elements and metrics of artefact complexity do not affect search time beyond what relevant elements can explain.

## Conclusion

Many web site designs, implicitly, and the semantic approach to WISB, explicitly, share the expectation that site search is faster when the home page of a web site includes a smaller number of web options. Our study did not provide clear empirical support for such an expectation. Similar search times were indeed found for sites whose start page included a different number of selection choices but equal number of web objects, while different search times were found for sites whose start page included an equal number of selection choices but a different number of web objects.

In our experiment IST depended on only a subset of web elements and links, as predicted by *C*, a measure of start page complexity based on knowledge of layout conventions. Our *C* measure is consistent with a mechanism of selective attention not necessarily including an estimate of the relevance of all links within a web page. It thus provides a convenient and general way to model the label relevance, being free from user's dependent parameters and requiring a minimal amount of knowledge about WISB. Moreover, our *C* measure is consistent with an efficient search strategy that avoids the large computational effort necessary for a selection based on an assess-all strategy, as used by several information navigation models.

## Acknowledgments

Part of this research has been supported by a grant of the European Social Fund 2007/13 – S.H.A.R.M. project – awarded to SR (p.i. WG).

## References

Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. *R package version* 0.99875-6.

Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the web. *Proceedings of CHI 2002* (pp. 463–470). New York, NY: ACM Press.

Gwizdka, J., & Spence, I. (2006). What can searching behaviour tell us about the difficulty of information tasks? A study of web navigation. *Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology*. Silver Spring, MD: American Society for Information Science and Technology.

Ivory, M. Y., Sinha, R. R., & Hearst, M. A. (2000). Preliminary findings on quantitative measures for distinguishing highly rated information-centric web pages. *Proceedings of the Sixth Conference on Human Factors & the Web*.

Kitajima, M., Blackmon, M. H., & Polson P. G. (2000). A Comprehension-based Model of Web Navigation and Its Application to Web Usability Analysis. In S. McDonald, Y. Waern & G. Cockton (Eds.), *People and Computers XIV - Usability or Else! (Proceedings of HCI 2000).* London: Springer-Verlag.

Lee, E., & MacGregor, J. (1985). Minimizing user search time in menu retrieval systems. *Human Factors, 27*, 157–162.

Michailidou, E., Harper, S., & Bechhofer, S. (2008). Visual complexity and aesthetic perception of web pages. *Proceedings of the 26th annual ACM international conference on Design of communication* (pp. 215-224). New York, NY: Association for Computing Machinery.

Paap, K., & Roske-Hofstrand, R. (1986). The optimal number of menu options per panel. *Human Factors, 28*, 377–385.

Pirolli, P., & Card, S.K. (1999). Information foraging. *Psychological Review*, *106*, 643–675 .

Pirolli, P., & Fu, W. T. (2003). SNIFT-ACT: A Model of Information Foraging on the World Wide Web. *User Modeling 2003* (pp. 45-54). Springer Berlin Heidelberg.

Rigutti, S., Fantoni, C., & Gerbino, W. (2007). Searching for a target word in a web page: the three components of information seeking behaviour, VSS, Sarasota, FL, USA, maggio 2007, *Journal of Vision*, 7: 720a, http://journalofvision.org/7/9/720/, doi:10.1167/7.9.720.

Tamborello, F. P., II, & Byrne, M. D. (2005). Information search: The intersection of visual and semantic space. *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1821–1824). New York, NY: Association for Computing Machinery.

van Schaik, P., & Ling, J. (2001). The effects of frame layout and differential background contrast on visual search performance in Web pages. *Interacting with Computers, 13*, 513–525.

van Schaik, P., & Ling, J. (2012). An experimental analysis of experiential and cognitive variables in web navigation. *Human-Computer Interaction, 27*, 199–234.