Title
Using LSA Semantic Fields to Predict Eye Movement on Web Pages

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Stone, Benjamin
Dennis, Simon

Publication Date
2007

Peer reviewed

Using LSA Semantic Fields to Predict Eye Movement on Web Pages

Benjamin Stone (bpstone@psychology.adelaide.edu.au)
School of Psychology, University of Adelaide
Adelaide, SA 5005 AUSTRALIA

Simon Dennis (simon.dennis@adelaide.edu.au)
School of Psychology, University of Adelaide
Adelaide, SA 5007 Australia

Abstract

This paper outlines a new method for estimating the visual saliency different areas displayed on a web page. Latent Semantic Analysis is used to calculate Semantic Fields values for any (x, y) coordinate point on a web page based on the structure of that web page. These Semantic Field values were then used to predict eye-tracking data that was collected from 49 participants' goal-orient search tasks on a total of 1842 web pages. Semantic Field values were found to predict the participants' eye-tracking data.

Keywords: LSA; Semantic Fields; LSA-SF; web pages; eye-tracking; visual saliency.

Introduction

Combining approaches

A review of both the Display-based and Semantics-based research into web user's visual search of web page hyperlinks has indicated that the user's search processes are influenced by: text semantics, element position, aesthetic qualities of elements, and environmental learning (Brumby & Howes, 2003, 2004; Chi et al. 2003; Faraday, 2000, 2001; Cox & Young, 2004; Kaur & Hornof, 2005; Ling & van Schaik, 2002, 2004; McCarthy, Sasse & Rigelsberger, 2003; Pearson & van Schaik, 2003; Pirolli & Fu, 2003; Rigutti & Gerbino, 2004; Blackmon, Kitajima & Polson, 2005; Grier, 2005; Pirolli, 2005). As is described more fully below, Semantics-based researchers have, to varying degrees, started to incorporate characteristics of the web-page display into their models. Moreover, several researchers have highlighted the importance of this combined approach to modelling users navigation through web sites (Blackmon et al, 2002; Pirolli and Fu, 2003; Chi et al. 2003; and, Kaur & Hornof, 2005).

The Cognitive Walkthrough for the Web (CWW) is a theory-based tool designed to assess the usability of websites (Blackmon, Kitajima & Polson, 2005). To this end, CWW simulates web user's navigation through a website using the CoLiDeS (Comprehension-based Linked model of Deliberate Search) model. Furthermore, CoLiDeS is based on Kintsch's Construction-Integration theory of comprehension. CWW approaches the problem of modelling web user's link following behaviour in a somewhat similar fashion to the Bloodhound Project (Chi et al., 2003). Like the Bloodhound Project's use of page position to inform calculation of probable link choice (Chi et al., 2001), the CWW process takes some aspects of the web page's display structure into consideration by grouping screen areas into regions. Also, like the Bloodhound Project, the semantic content of the each web page is evaluated statistically against the web user's target goals. However, instead of using the WUFIS (Web User Flow by Information Scent) algorithm, the CWW uses Latent Semantic Analysis (LSA) to compare semantic content. Furthermore, similar to the Bloodhound Project's close relative SNIF-ACT, which is a model based on the ACT-R cognitive architecture (Pirolli & Fu, 2003; Pirolli, 2005), the CWW does not limit the content of its statistical semantic analysis to the documents in the website. To this end, both CWW and SNIF-ACT also incorporate a corpus of documents that is considered to represent a user's knowledge base. Once a web page has been segmented into regions or sections, the model generates a description of each section, and these descriptions are then compared, using LSA, with the users goals and knowledge base. The section with the highest similarity to these user components is then selected for further analysis. Link texts in the selected section are again evaluated against the web user's goal and knowledge base using LSA. After this evaluation, the model then follows the hyperlink with the highest utility.

Latent Semantic Analysis (LSA)

LSA is a statistical method of textual evaluation that allows the researcher to derive meaning from a set of documents (Landuaer & Dumais, 1997; Landauer, MacNamara, Dennis & Kintsch 2007). Linear algebraic methods, such as Singular Value Decomposition, enable the researcher to determine the semantic similarity between words and sets of words contained within a corpus of documents. In a way, the corpus of documents acts as a knowledge base. For example, the Touchstone Applied Science Associates (TASA) document corpus represents literature that students may have been exposed between grade 3 and the first year of college. Moreover, the research described in this paper uses the TASA corpus as a best approximation to the knowledge-base of the first year university students who have participated in this study.

LSA - Semantic Fields (LSA-SF)

In this paper, an alternative method of modelling human behaviour in web page environments is reported. The LSA-

SF model is based on the assumption that each element displayed on a web page (such as hyperlinks, content text, and images) will influence how the page elements surrounding it are perceived and what value a web page user will assign to them. In this way, a group of hyperlink elements that are close in spatial proximity may be recognised by the web page user as a navigation menu. For instance, it is not that an individual hyperlink is recognised as a menu, but its proximity and similarity in appearance to the set of hyperlinks around it, which forms their overall Gestalt or structure into a navigation menu. LSA-SF models this relation between web page elements by using a decay function to distribute the utility[1] of a given element over a web page. In this way, items that are closely positioned on a web page, such as hyperlinks in a navigation menu, accumulate more utility or saliency than items placed further apart.

Rather than predicting hyperlink-clicking behaviour, although this is an area that will be examined in future research, the method presented here attempts to predict the precursor to hyperlink clicking, eye movement during goal-directed search on web pages. To this end, LSA-SF is used to generate heat-maps of the semantic content on each web page. Put another way, the heat maps are used to predict the saliency or 'draw' associated with any given (eye) point on a web page. Interestingly, LSA-SF method generally performs better at predicting eye movement than would be expected by chance alone.

Method

Participants

In this study 49 participants, 27 males and 22 females, searched for information on three web sites. The participants were recruited from either a first-year pool volunteering in exchange for partial course credit, or other members of Adelaide University who where paid $30 for their time. Participant ages range from 16 to 57 (*M*=22y3m *SD*=1y). Also, there was a positive skew in the samples age distribution, with 93 percent of participants were younger than 31 years old.

All participants had both previous computer and Internet experience. Self-reported years of Internet experience ranged from 4 to 17 years (*M*=8.65, *SD*=0.37), with self-reported frequency of Internet use ranging between 2 to 75 hours per week (*M*=14.24, *SD*=1.86). Based on these self reports, this group appeared on average to be very experienced users of the Internet.

[1] Similar to the CWW, LSA-SF also incorporates a knowledge base, the TASA database, with LSA cosines generated from comparisons between goal text, each web page element's text, and the TASA database.

Apparatus

Behavioural recording equipment and software The IETracker program was developed to record both participants' behaviour during website search tasks and web page display characteristics. These observations are accomplished by programmatically controlling, and integrating, Microsoft Internet Explorer (Version 6) and the ViewPoint EyeTracker PC-60 QuickClamp System. User and site specific data collected during this exploratory experiment included: eye-tracking; hyperlink clicking; and web page composition (location, semantics, images, colour, style, and size of web page elements). All data was then stored in a Postgresql database for later analysis.

Websites Three websites were chosen from the Internet:
www.greencorps.com.au (Green Corps Australia)[2]
www.missionaustralia.com.au (Mission Australia)
www.whitelion.asn.au (White Lion Australia)

Static versions of these sites[3] were fetched in December 2005 and stored locally on the experimental computer. This was done to standardize the pages viewed for each participant. These websites are similar in the type of service they provide, in that they all offer services to disadvantaged members of the community. They were chosen because they were sufficiently complex that searching for information on these sites would be a non- trivial task for participants. Moreover, these website were each designed using table-based HTML which is well suited to the IETracker software.

In the case that a participant might click on a hyperlink linking to a PDF document, these hyperlinks where updated to display a web page instructing the participant that they had found their search target, or that they had not and should click the back button and keep searching.

Distributed over the three websites there were 114 unique pages viewed by participants. Collectively, the participants viewed 1910 web pages, however eye-tracking calibration difficulties lead to the exclusion of 68 of these pages, leaving a total of 1842 web pages in this analysis. On average, each search task required the participants to visit four pages on a given web site. Furthermore, the time spent on each page varied from 3 seconds to a 1 minute 44 seconds, with the median time spent searching an individual page recorded at 11 seconds.

An Hitachi CM823F 21 inch monitor was used to display the websites to participants with screen resolution set to 1280 by 1024 pixels.

[2] The Green Corps Australia URL is no longer used, the Australian Government has change both the website and its URL which can be viewed here: http://www.greencorps.gov.au. The static version of this website used in this research can be viewed here: http://www.psychology.adelaide.edu.au/mall_lab/lsa-sf_sites/

[3] The static versions of these websites can be found here: http://www.psychology.adelaide.edu.au/mall_lab/lsa-sf_sites/

Procedure

Participants were required to search for three target pieces of information on each of the three web sites. For example, one of the tasks for the Mission Australia web site was worded:

"Who is currently the Chief Operating Officer of Mission Australia?"

Each task was read aloud to the participants twice before they commenced their search. Moreover, they were asked to signal the experimenter (with a hand gesture) at any time they wanted the search task repeated aloud. A three by three Latin square design was used to control for order effects in the display of each website. Also, a nested three by three Latin square design was used for the same purpose to guide the presentation order of each of the target tasks.

After an initial calibration procedure, using the Viewpoint eye tracking software, participants where given their search task in the manner described above and commenced their search. Given the physical structure of the Viewpoint Eye-Tracker (which uses a mounted camera with forehead and chin rests) and some participant's predisposition to fidget, it was necessary to perform additional eye-tracking calibration during the search tasks. This additional calibration required the participants to focus on targets in nine separate regions of the monitor. Moreover, these targets were automatically displayed on the screen by the IETracker program after each hyperlink clicked during the participant's search task. To elaborate on this point further, if a participant clicked through four pages in their search for the target information, then four extra calibration procedures were performed during this search task; each calibration performed after leaving the page that was clicked on and before the next page was displayed to the participant.

The algorithm used to calculate the offsets (O_t) used to reposition the eye-tracking data is:

$$O_t = \frac{\sum_i o_i e^{-\lambda d_i}}{\sum_i e^{-\lambda d_i}} \qquad (1)$$

Where O_i is the median offset (x, y) of the eye-tracking data when viewing calibration point *i* from the actual center of calibration point *i*. And, d_i is the distance between the eye tracking point *t* and median offset (x, y) of the eye-tracking data when viewing calibration point *i*. Therefore, as λ gets closer to 1, the O_t shifts the eye tracking point towards the calibration point *i*. as λ gets smaller than 1, the O_t shifts the eye-tracking point towards the average of the median offsets (x, y) for all calibration points.

Finally, participants were instructed that when they believed that they had found the target information that they should then click on the 'HOME' icon (which is an image of a house on Internet Explorer's menu bar). This was followed by one more round of automated calibration, and some degree of relief for the participants.

Calculating the LSA Semantic Fields LSA-SF values were calculated for each unique web page that the participants visited in this experiment. The rationale underlying the concept of the LSA-SF, is that the information displayed on one region of a web page (or any document for that matter), has an influence which goes beyond it's immediate location. Moreover, this influence will decay as the point of focus is moved away from its source.

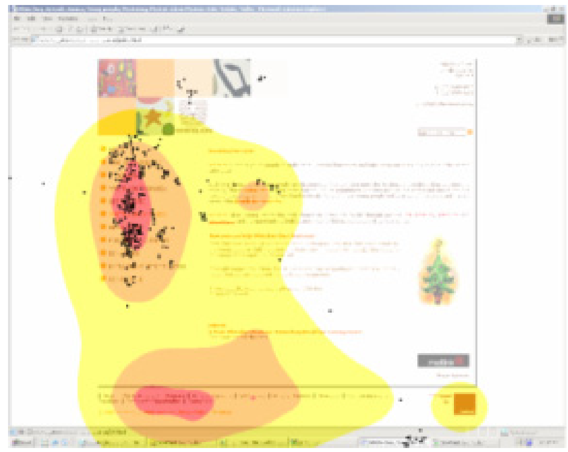

Figure 1: Example of a LSA-SF Map with a participant's eye data super-imposed using black dots

To this end, LSA cosines are calculated between a given task's search text and the semantic content of text-based web page elements. The concept of a LSA-SF is probably best illustrated using a LSA-SF Map (see Figure 1.) that can be likened to a heat map based on semantic content of a web page. Using these LSA cosines and calculating the distance between the centre of an element and any given point on the web page (x, y), LSA-SF values can then be generated by weighting (λ) the values returned by the decay function in the following algorithm:

$$sf(x,y) = \sum_i L_i e^{-\lambda d_{i(x,y)}} \qquad (2)$$

Where sf(x, y) denotes the LSA-SF value calculated for a given point (x, y), L_i is the LSA cosine of the search task text and an element's text, and d_i (x, y) is the distance between the center of a web page element and the given point (x, y).

Results

There are two analyses of the accuracy of the LSA-SF model that are reported here. They differ in the basic information on which the LSA-SF values were generated. In the first case the LSA-SF values were generated from the hyperlink text that was on the web page. While in the second case, the LSA-SF values (in addition to using the hyperlink text) also included the text from other elements on the page and the text descriptions of images that web designers include in some image's alternative text field.

The accuracy of both LSA-SF methods were compared against random points that were equivalent in frequency to the number of eye points[4] recorded by a participant on any given page. To this end, for any given web page viewed by a participant, 1000 trials were run in which randomly assigned points, equal to the number of observed eye points collect for that page view, were distributed over the web page. For each of the 1000 random trials, the average LSA-SF value was calculated. Finally, a 'LSA-SF Overall Mean Value' (OMV) was derived from the aforementioned 1000 LSA-SF averages associated with each page. Overall, 1842 OMVs where calculated, one for each page viewed by participants.

If the average LSA-SF values calculated for the participant's actual eye data was not significantly greater than that which can be generated by chance (the OMV), then the LSA-SF method would be have failed to predict the eye movements of participants at a rate that was equivalent to chance.

Table 1: Eye-based LSA-SF values compared to LSA-SF Overall Mean Values (OMVs) of the 1000 random trials associated with each page view.

		Worst Case		Best Case	
Method	Data	*M*	*SD*	*M*	*SD*
Links	Eye	62.83	36.10	69.90	38.58
	OMV	29.07	8.15	29.07	8.15
All Text	Eye	91.82	37.56	100.04	39.05
	OMV	50.78	14.25	50.78	14.25

Worst case to the best case scenarios

As was outlined in the method section of this paper, some participant's disposition to movement made it necessary to record eye-tracking calibration data after each web page that was viewed. Based on these calibration data, the weightings used in the recalibration process of eye-tracking points resulted in several possible scenarios for the eye-tracking data. We present the 'best' and 'worst' case scenarios of this recalibration process (see Table 1). It should also be emphasized, that while the categorization 'best' and 'worst' is based on the average LSA-SF values for these recalibrated eye data, the actual recalibration process is prior to, and in no way influenced by, the production of LSA-SF values for a page.

Hyperlink-based LSA-SF

Firstly, it needs to be noted that because not all web pages had hyperlinks on them[5], it was not possible to generate "Hyperlink-based" LSA-SF values for all of the web pages analyzed. Of the 1812 web pages analyzed, regardless of the weights used for the calibration eye data, Paired-Samples t-tests showed that the average LSA-SF values generated by the human eye data were significantly greater than the LSA-SF OMVs generated using the random points data sets (Worst Case: $t(1811) = 43.92$, $p<.001$; Best Case: $t(1811) = 49.65$, $p<.001$). Figure 2 illustrates that using the Link-based LSA-SF method, for most web pages viewed (92.11%) the average eye-based LSA-SF value was greater than the LSA-SF OMV that was randomly generated for that web page view.

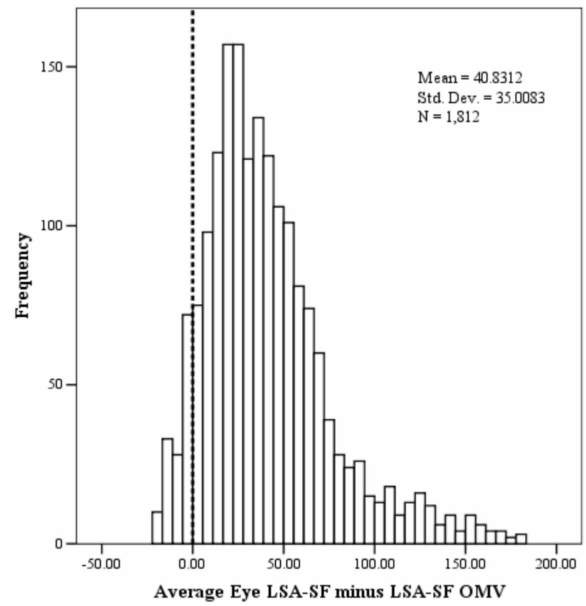

Figure 2. Eye LSA-SF average minus the LSA-SF OMV for each page viewed by participants using the Link-based LSA-SF method (Best Case).

All text LSA- SF

Adjusting the method used to generate the LSA-SF values to include text from all page elements including hyperlink text and alternate text from image fields improved the performance of the LSA-SF method. Again, in both the worst and best case scenarios of calibration weighting employed, Paired-sample t-tests indicated that the average LSA-SF values recorded for the human eye data were significantly greater than the LSA-SF OMVs (Worst Case: $t(1841) = 50.01$, $p<.001$; Best Case: $t(1841) = 58.23$, $p<.001$). As can be seen in Figure 3, when using the All Text-based LSA-SF method, for most web pages viewed by participants (94.73%), the average eye-based LSA-SF value was greater than the LSA-SF OMV that was randomly generated for that web page view.

[4] 'Eye points' refer to all eye-tracking data (fixations, saccades, etc.) collected during a participant's viewing of a web page.

[5] As was mentioned in the method section, commonly followed links that connected to PDF files were replaced with links to web pages giving the participant instruction such as "You have found your search target. Click the HOME button". These pages had no hyperlinks on them.

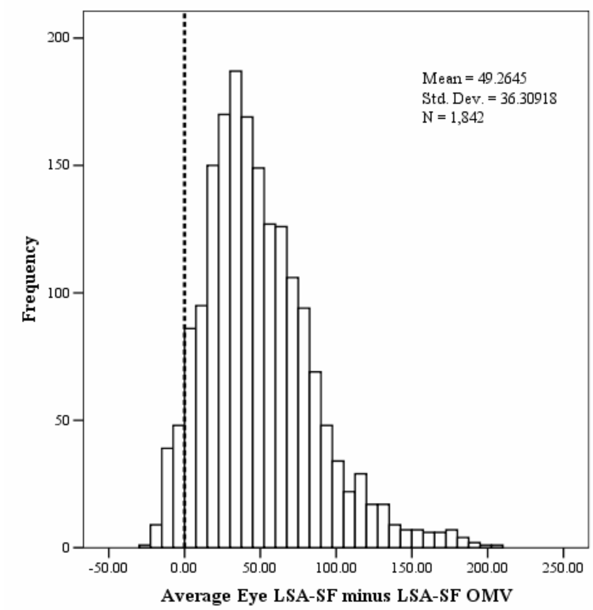

Figure 3. Eye LSA-SF average minus the LSA-SF OMV for each page viewed by participants using the All Text-based LSA-SF method (Best Case).

Discussion

The results of this study indicate that the LSA-SF method predicts eye-tracking data collected in this research. To further exemplify the LSA-SF method, Figure 1 displays the actual eye data, collected from one participant in this research. In this image, eye data (in black) has been super-imposed onto the LSA-SF Map associated with the web page that the participant was viewing. It should be noted that within the menus on this page there are several layers of heat. This layering of heat in the menus has resulted from the text contained in individual menu hyperlink elements different LSA cosines when they were compared to the search task text. Furthermore, it can be seen clearly in this image, that most of the eye-tracking data fell into the areas of greater accumulated utility or heat.

Other sources of heat

While the heat maps that were generated in this study were based on the text and its position on the page, there are other ways these maps could have been generated. For example, a similar analysis of 'information density' was conducted by Granka and colleagues that examined hues and colour contrast of the display in relation to screen position (Granka et al., 2004; Granka, Hembrooke & Gay, 2006). In future analysis, methods that examine the display's colour contrasts may be used in conjunction with the type of textual analysis that was performed in this study.

The addition of rules

In the future, the LSA-SF method could also be augmented using rules like those described by Rigutti and Gerbino (2004). These authors present the WebStep model that predicted differences in user behaviour when following a navigation bar option as opposed to links embedded in the content text. They suggest users will judge that embedded links lead to specific information, whereas navigation links lead to broad categories. In their model, perceived distance to the target information moderates use of both the navigation bar and embedded links. When the distance was deemed to be small, their model predicts embedded links, as opposed to navigation bar links, are more likely to be chosen by the user. However, as the perceived distance to the target information increases, the probability that a navigation bar link will be chosen by user increases, and the probability that an embedded link will be chosen by the user decreases. Furthermore, utility of navigation bar links decrease as the position of the navigation bar is placed lower in the visual display. In the context of the LSA-SF model, LSA-SF values could reflect perceived distance to the target information. However, these values would be moderated by the height of the source element on the page.

Benefits of the LSA-SF to eye-tracking research

Some eye-tracking studies of human behaviour in web page environments have examined user's scan-paths either over the whole web page (Josephson & Holmes, 2002) or in sub-sections (such as menus) of the web page (Cox & Young, 2004). While, in other work, researchers have examined the frequencies of participants' eye gazes in different regions (center, top, bottom, left, right) of web pages (McCarthy, Sasse & Rigelsberger, 2003). Common to these approaches is the practice of defining static rectangular regions or areas of interest (ROI) with which to interpret their eye-tracking data. This serves at least three purposes. Firstly, it gives the researcher a unit to measure. Secondly, eye-tracking systems come with software that allows the researcher to define these ROI and count the eye points that fall within them. Thirdly, there is error associated with the recording of eye points, so the use of broad ROI of interest provide a buffer for this error.

The LSA-SF method offers eye-tracking researchers who are studying visual media a more fluid way of automatically defining ROIs. Moreover, the unit of measure becomes more precise, moving eye-tracking studies from a nominal to a continuous scale of measurement. Finally, the LSA-SF algorithm buffers the error associated with eye-tracking, because an eye-point which is 'out' by several degrees will still gauge the heat generated from the element beside it.

Summary

In this paper we have presented a new method called LSA-SF that detects structure contained within web pages. This is done by combining knowledge of both the semantic content displayed by page elements and this content's relative position in relation to other textual web page elements. LSA was used to assign utility to a web page's textual elements, based on comparisons between search goal text and the TASA corpus of documents. Moreover, in this analysis the TASA corpus acted as a knowledge base for the

LSA algorithm. LSA cosines derived in this manner were then distributed over the page using a decay function, and the accumulated value of these decaying cosines was calculated at any given point on the web page. This facilitated the production of semantic heat-maps that display estimates of the visual saliency for all points on the web page.

The LSA-SF values produced in the manner were then used to predict human eye-tracking data in a large set of web pages. To this end, it was found that the LSA-SF values did predict the participants' eye-tracking data. Furthermore, the tasks performed by participants seemed to have good ecological validity, a situation that is facilitated by the virtual environment created by Internet browsers.

Acknowledgments

Many thanks go to Assistant Professor Michael Lee from Irvine University for purchasing the ViewPoint EyeTracker system. Also, thank you to Dr Dan Navarro, Professor Marilyn Hughes Blackmon, and the anonymous reviewers for both taking the time to read through this paper and their excellent suggestions.

References

Blackmon, M. H., Kitajima, M., & Polson, P. G. (2005). Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 31-41). New York: ACM Press.

Brumby, D. P. & Howes, A. (2003). Interdependence and past experience in menu choice assessment. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (p. 1320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brumby, D. P. & Howes, A. (2004). Good enough but I'll just check: Web-page search as attentional refocusing. *Proceedings of the Sixth International Conference on Cognitive Modelling* (pp. 46-51). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information need and actions on the web. *Proceedings of ACM Conference on Human Factors and Computing Systems* (pp. 490-497). New York, NY: Association for Computing Machinery

Chi, E. H., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J. & Cousins, S. (2003). The bloodhound project: automating discovery of web usability issues using the InfoScentπ simulator. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 505-512). New York, NY: Association for Computing Machinery

Cox, A. & Young, R. M. (2004). A rational model of the effect of information scent on the exploration of menus. *Proceedings of the Sixth International Conference on Cognitive Modelling* (pp. 82-87). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaur, I. & Hornof, A. J. (2005). A comparison of LSA, wordNet and PMI-IR for predicting user click behavior. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 51-60) New York, NY: Association for Computing Machinery.

Granka, L., Hembrooke, H., Gay, G., & Feusner, M. (2004). Correlates of Visual Salience and Disconnect. *Unpublished research report*, Cornell University Human-Computer Interaction Lab. Retrieved April 29, 2007, from http://www.hci.cornell.edu/projects/eye_tracking.htm

Granka, L., Hembrooke, H., & Gay, G. (2006). Location location location: Viewing patterns on WWW pages. *Proceedings of the 2006 symposium on Eye tracking research & applications* (p. 43). San Diego, CA: Association for Computing Machinery

Josephson, S. & Holmes, M. E. (2002) Attention to repeated images on the World-Wide Web: Another look at scanpath theory. *Behavior Research Methods, Instruments, & Computers,* 34, 539-548

Landauer, T. K., & 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.

Landauer, T. K., McNamara, D. M., Dennis, S., & Kintsch, W. (2007). *Handbook of Latent Semantic analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ling, J. & Van Schaik, P. (2002). The effect of text and background color on visual searches of Web pages. *Displays*, 23, 223-230.

Ling, J. & Van Schaik, P. (2004). The effect of link format and screen location on visual search of web pages. *Ergonomics*, 47, 907-921.

McCarthy, J., Sasse, M.A., & Rigelsberger, J. (2003). Could I have the menu please? An eye tracking study of design conventions. *Proceedings of HCI 2003* (pp. 401-414) London, UK: Springer-Verlag.

Pearson, R. & Van Schaik, P. (2003). The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task. *International Journal of Human-Computer Studies,* 59, 327-353.

Pirolli, P. (2005). Rational analyses of information foraging on the Web. *Cognitive Science, 29,* 343-373.

Pirolli, P. L. & Fu, W-T. (2003). SNIF-ACT: A model of information foraging on the world wide web. *Ninth International Conference on User Modeling* (pp. 45-540) New York, NY: Springer-Verlag.

Rigutti, S. & Gerbino, W. (2004). Navigating within a web site: the WebStep Model. Proceedings of the Sixth International Conferenceon Cognitive Modeling (378-379) Mahwah, NJ: Lawrence Erlbaum Associates.