Analysis of User Behavior for Web Search Success using Eye Tracker data

David F. Nettleton*‡ and Cristina González-Caro†*
*Universitat Pompeu Fabra
Barcelona, Spain
david.nettleton@upf.edu
†Universidad Autónoma de Bucaramanga
Bucaramanga, Colombia
cgonzalc@unab.edu.co
‡IIIA-CSIC
Bellaterra, Spain

***Abstract*—Web log data has been the basis for analyzing user query session behavior for a number of years, but it has several important shortcomings. The main one being that we dont really know what the user is doing – is s/he looking at the screen or doing something else? We have conducted an Eye-Tracking study to analyze user behavior when searching the web and looking for specific information on results and content pages. The goal is to obtain more precise information about the search strategy of the user. Which characteristics make the difference between successful and unsuccessful searches? This research presents results focusing on the number of formulated queries by session, documents clicked, the fixation durations on the documents, and the distribution of the attention in the different areas of the screen, among other aspects.**

***Keywords*-eye tracking; search success; user behavior; web search; visualization**

I. Introduction

Eye-tracking is a state of the art technology for registering ocular activity on a computer screen, which recently has become much more manageable, economical and user friendly. Web logs, on the other hand, have been used widely in the web search analysis community to identify profiles of search categories, user types, document categories, and so on. However, anonymous web log data is nothing more than an approximation of what the user is really doing. The behavior of users who are successful or unsuccessful in their searches may also vary in terms of eye movement and fixation points on the screen.Our hypotheses is that these two groups of users will behave differently, and we can quantify these factors from the web log and eye tracking data together, and analyze the differences statistically.

In the present work, we apply Eye-tracking technology to the study of the web search behavior of successful and unsuccessful users. The goal of interpreting user search behavior is to provide information to web applications, such as search engines, so that they will be able to adapt to different kind of users, user needs and search contexts.

II. Related Work

In recent years, web search behavior has been analyzed in different ways by the web mining community. Most of the studies recognize the importance of how to gain insight into user intention and the factors that influence the users decision processes [1]. Numerous studies have been based on query logs collected by search engines to analyze user's past behavior. Some of these studies work with user profiles, like Sugiyama et al [2] who tried constructing user profiles from past browsing behavior of the user. Agichtein et al [3] worked on modeling the behavior of web search users to predict web search results preferences using query-text features, browsing features and click-through features. Attenberg et al [4] used query logs to study user behavior on sponsored search results. Lee et al [5] and Baeza-Yates et al [6] developed approaches for the automatic detection of user goals in web search.

More recent studies have introduced the Eye-tracking technique in the analysis of user search behavior. Work applying Eye Tracking to web search [7][8], tends to focus on specific aspects of the page content. For example, in [7], Cutrell et. al. conduct a study to evaluate the effect of snippet length on how people use Web search. With respect to work related to eye movement analysis on web pages, two key recent findings are the Golden Triangle, presented by Hotchkiss in [9], and the F-Shaped pattern presented by Nielsen in [10]. Hotchkiss related the Golden Triangle tendency to how a search engine works and presents results. In general, the search engine places the best listings at or near the top. Therefore, users have the tendency, at the beginning of a search session, to restrict ocular scan activity to the area of the page most likely to have the best listings. This area has been termed the Golden Triangle. On the other hand, Nielsen [10] showed that users read web content in an F-shaped pattern. The F-shape reading pattern refers to a viewing order in which users commence by reading across the top line and then looks down the page a little and reads across again and then continues down the left

978-0-7695-4839-5/12 $26.00 © 2012 IEEE
DOI 10.1109/LA-WEB.2012.18

side. The implications of the F-shaped tendency confirm that users do not read all of the content on a Web page. As a consequence of these findings, Nielsen recommended that the most important information should be contained in the first two paragraphs.

Some other studies have analyzed the differences between the search strategies of the users [11] and successful performance in web searching [12]. The experiments and statistics of these papers are derived from web log data, especially the number of results clicked, elapsed time durations on results and content pages, and characteristics of the query and query terms. They do not use eye-tracking technology. Earlier research has mainly focused on the construction of general models of web user search behavior principally based on mining query log data. Individual differences such as successful or unsuccessful searches and the differences between those strategies have not been studied deeply. In this work we apply Eye-tracking technology to the study of the web search behavior of successful and unsuccessful users. In contrast with previous studies, in the present work we allow users to formulate queries on their own, instead of providing a ready-made set of queries for each task. Learning to interpret user search behavior would allow systems to adapt to different kind of users, needs and search settings.

III. Experimental Design

In traditional Web Search Engines users submit a set of keywords and are then given a list of results with descriptions of documents that possibly contain the answer to their information needs. Some of the results shown by the Search Engine will be irrelevant and just few of them will contain links to pages with the correct answers to the users query. Our experimental setting to evaluate eye movement data was designed to simulate a natural web search session. Users were asked to search as they normally would, and the study focused on unconstrained web search by allowing users to navigate multiple pages and to formulate and reformulate all the queries that they want, while searching for information to resolve a specific question. The aim of the research is to discover as many differences as possible among people that are successful in their web searches compared to those that are not.

A. Participants

Fifty seven (57) people participated in the study, 24 men and 33 women. The average age of the participants was 28 years in a range of 18 to 61 years. The participants had a diverse range of professions, background and education levels (undergraduate student was the minimum education level).

B. Apparatus

The device used for measuring eye movements was the Tobii 1750 Eye Tracker. The eye tracker is integrated into a 17 TFT monitor. The tracker illuminates the user with two near infrared projections to generate reflection patterns on the corneas of the user. A video camera then gathers these reflection patterns as well as the stance of the user. Digital image processing is then carried out for extracting the pupils from the video signal. The system tracks pupil location and pupil width at the rate of 50 Hz. The resolution of the tracker is 1280x1024. The system allows free head motion in a cube of 30x15x20 cm at 60 cm from tracker.

Table I
General statistics by query session

	Successful	Unsuccessful	All
Avg. number of queries	1.86	3.26	2.5
Avg. number of pages selected	4.36	8.32	6.34

Table II
Eye-tracking statistics of successful and unsuccessful query sessions

	Successful	Unsuccessful	All
Fix. frequency on SERP	25.41	49.40	37.41
Fix. frequency on CRP	29.47	56.96	43.22
Fix. duration on SERP (Sg)	16.11	28.6	22.36
Fix. duration on CRP (Sg)	17.82	32.39	25.1

C. Query Questions

We designed three questions for the web search experiment. The idea was to create general questions in a non specialized topic. All the questions were defined as informational questions, that is, the intent of the task is to find an answer to a question that has a unique answer [1]. We use informational questions in order to reduce the ambiguity of the results. Each question is distinct and there is no relation or connection between them. The designed questions are:

- Q1. Name of a mechanical machine (not electrical) for calculating, of German origin, which fitted in the palm of a hand. The correct answer to the question is: Curta.
- Q2. Name of the wife of the author of "The Jungle Book" . The correct answer to the question is: Carrie Balestier.
- Q3. Name of a Catalan NGO which works in India and whose founder was recently hospitalized. The correct answer to the question is: Vicente Ferrer Foundation.

Throughout this paper the terms Q1, Q2 and Q3 will be used to make reference to the above three questions, respectively.

D. Procedure and Design

At the beginning of the experiment the users were informed about the purpose and the procedure of the study. Users were asked to perform the search tasks using the web search engine as they would normally do. Two task searches corresponding to two of the designed questions

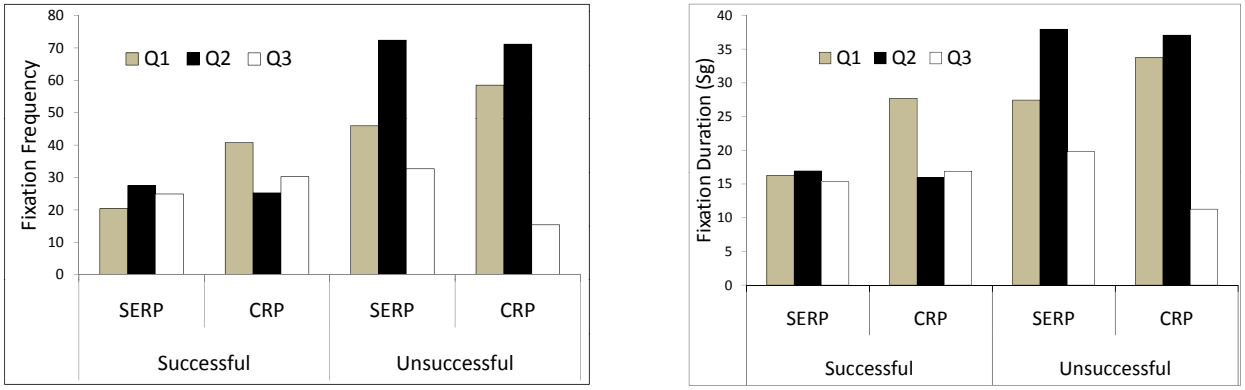

Figure 1. (left) Fixation frequency and (right) Fixation duration for successful and unsuccessful searches by type of question and type of page.

of this study were conducted by each participant. When starting the search task, a description of the question was shown to the users. After reading the question description, the participants clicked a start button and were taken to a search engine page to start the search task. The users had to find the correct answer to the question by searching the web through a commercial Search Engine. Participants were free to click links and scroll page-up/page-down as needed, this navigation freedom adding to the study a high degree of reality to the search tasks. Users were told they should continue the task until completed or until 5 minutes had elapsed.

IV. Empirical Results and Analysis

In this section we analyze the characteristics of the Web searches performed by the participants of this study. In all the analysis, we distinguish the page type, where "SERP" refers to the *search engine result page* and "CRP" refers to the *clicked result page*. The results are divided into two principal subsections, Section IV-A, a basic statistical analysis by question, search success and page type (SERP or CRP), and Section IV-B, an evaluation of the statistics of user gaze coordinates.

A. Basic statistics and analysis by question, search success, and page type (SERP o CRP)

Almost all the participants (56) completed the two task search that they had assigned; only one of them completed half of the task and performed just one of the two query sessions. In total, 57 users performed 67 successful searches and 46 unsuccessful searches. The results of the general statistics (see Table I) show clear trends for successful and unsuccessful searches. On average, users who are successful formulate fewer queries per session and visit a smaller number of documents than unsuccessful users. We conducted ANOVA tests with independent variable type of user (successful, unsuccessful) and with dependent variables number of queries formulated and average number of documents selected. Significant effects were found for both cases: number of documents (F=153.309, $p < 0.001$) and average number of queries (F=229.053, $p < 0.001$). These initial trends suggest that successful users have a more focused search behavior. We can contrast this assumption by evaluating the variables derived from the eye tracking methods, fixation frequency and fixation duration (see Table II). Fixation frequency refers to the number of times that the user fixes his attention in some specific point of the screen. Fixation duration is related to the duration of a fixation [13]. The results show that the time and number of fixations that unsuccessful users invest in SERP and CRP is almost the double of the time and fixations that successful users invest in the same pages. If the web search engine is able to detect this behavior, it has the possibility to bring to the user a more effective service: more clear and intuitive interfaces or highlights on the key information SERP that allows these users to make a better selection of documents.

1) Analysis of web search behavior in terms of task difficulty: Another aspect to consider in the analysis of web search behavior of the users is the difficulty of the task. The percentage of successful searches to each question is a good indicator of the difficulty level of the task search: Q1 (yes=20% / no=80%), Q2 (yes=77% / no=33%)and Q3 (yes=88% / no=12%). By observing these success percentages we can clearly identify question Q1 as the most difficult task search. The general statistics obtained from the query log data, average number of queries formulated (Q1=2.7, Q2=2.5, Q3=2) and average number of document pages selected (Q1=7, Q2=6.3, Q3=4.3), suggest that the search activity of the users is proportional to the difficulty level of the search task. However, the results for fixation frequency and fixation duration of successful and unsuccessful searches suggest that the time that a user spends in SERP and CRP is related with the strategy of search of the user and not with the difficulty of the task. In general, the fixation frequency and fixation duration of successful users for SERP is homogenous for the three questions (see Figure 1). Successful users find the answers for the questions approximately in the

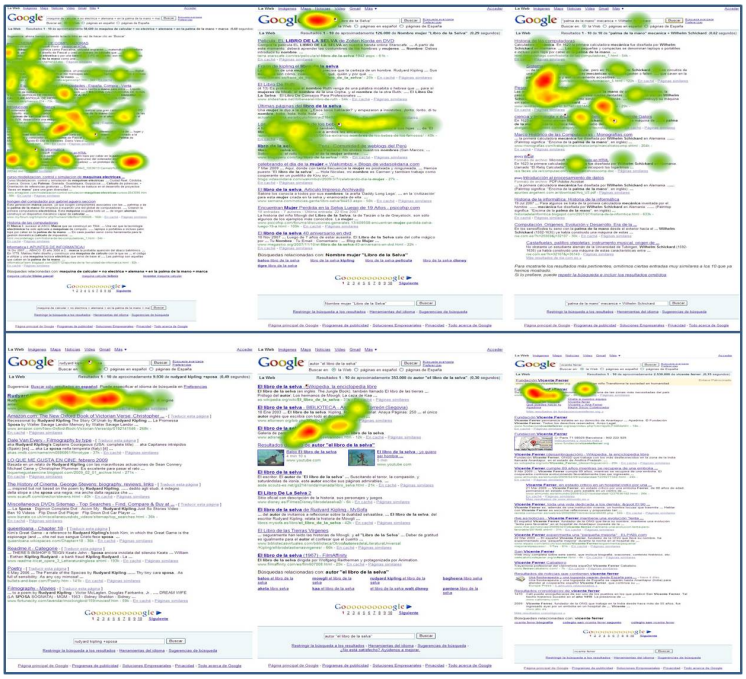

Figure 2. Screen images on top show SERPs of unsuccessful searches. Screen Images on bottom show SERPs of successful searches.

same time and with the same number of fixations. It is not the same situation for unsuccessful users, this kind of user does not have a homogeneous search behavior. The fixation time and fixation duration is different for each question and in any case their behavior is less effective in time and number of fixations when compared with successful users. The gap in the attention that satisfactory and unsatisfactory searches paid to SERP and CRP suggest that these two kinds of search behaviors need different types of service or help from the web search systems. Figure 2 shows some examples of the unsuccessful and successful users, respectively, when they search for information in the SERP.

B. Evaluation of statistics of user gaze coordinates

In the following subsection we evaluate the statistics of user gaze coordinates (x, y) on the screen, in terms of standard deviation and by visualization using scatter plots. In Tables III and IV we see the compiled standard deviation statistics for the gaze x, y coordinates generated from the gaze plots, for SERP (Search Engine Results Page) and CRP (Clicked Result Page), respectively. Statistics were compiled for each question (Q1, Q2 and Q3) separately so as to group similar pages. In Table III we observe, for the SERPs, the standard deviation, and therefore the scatter of where the users look for gaze durations of greater than one second (1000ms). We observe that in all cases the successful searchers have smaller standard deviations than the unsuccessful searchers. This confirms that for the SERPs, a lower gaze scatter is more effective. On the other hand, in Table IV we observe that, for the CRPs, the x coordinate scatters

Table III
AVERAGE STANDARD DEVIATIONS FOR SERPS (GAZE DURATION > 1000MS).

	Q1		Q2		Q3	
	$S*$	$NS*$	S	NS	S	NS
x	88.26	92.47	59.70	82.16	66.71	73.83
y	119.22	139.70	104.32	142.64	121.76	154.67

are fairly similar for successful and unsuccessful searchers, except for question Q1. However we observe a much more significant trend for the y coordinate scatters which show a much greater scatter for the successful searchers. This can be explained by the fact that key information in a CRP may be in different extremes of that page, and therefore the searcher has to make a greater displacement of his/her gaze to focus on the key information. On the other hand, the unsuccessful searcher will not have located that information or restricted his/her inspection of the page to a smaller region. In the case of the SERP (Table III), the search engine tries to locate the key information in the upper left part of the screen. Therefore the successful searchers who find the key information in principal will have smaller gaze scatters. In agreement with the findings of [8], we also observed that the complexity of web page design influences the degree of scan-path variation among different subjects on the same web page. The standard deviations on SERPs, where the web design is more homogenous, are lower than standard deviations on CRPs, where the web page design tends to have different levels of complexity.

Table IV
AVERAGE STANDARD DEVIATIONS FOR CRPS (GAZE DURATION > 1000MS).

	Q1		Q2		Q3	
	S^*	NS^*	S	NS	S	NS
x	120.33	138.18	116.92	118.72	119.03	116.62
y	846.68	501.81	836.79	463.50	169.89	148.83

1) X, Y Distributions of Gaze: In this Section we compare the scatter plots of successful and unsuccessful users, for the three queries (Q1, Q2 and Q3) and for the two types of page (SERP and CRP). It is important to note that we filtered the gaze coordinates so as only to include users whose corresponding gaze duration was greater than 1000 (1 second). The scatter plots a to d of Figure 3 show where the users looked on the page, distinguishing successful searchers on the left, unsuccessful searchers on the right, for search questions Q1 and Q3. We divide by search objective because the same query gives similar SERPs and CRPs, so are more comparable than results/content pages of different search objectives. If we compare the left hand plots (successful users) with the right hand plots (unsuccessful users) we see some clear differences. The successful users for query Q1 have gaze points which are almost completely restricted to an area of x=50 to 350 and y=0 to 600. On the other hand, unsuccessful users have a greater dispersion on both axes. We recall that Q1 was the most difficult question. The successful users for query Q3 also show a similar tendency, with a greater concentration in the area x=50 to 350 and y=0 to 600. We recall that Q3 was the easiest question. In the case of query Q2 (not shown for brevity) the patterns were not so distinguishable between successful and unsuccessful users. We recall that question Q2 had an intermediate level of difficulty.

The scatter plots a and b of Figure 4 are divided in the same manner as those of Figure 3: successful users on the left, unsuccessful users on the right, for question Q1. If we compare the left hand plots (successful users) with the right hand plots (unsuccessful users) we again see some clear differences. The successful users for query Q1 have gaze points which are divided into two bands relative to the y axis (0 to 2000 and 3800 to 5000), and also seem to be displaced on the x-axis (in general greater than x=150). The unsuccessful users, on the other hand, have a sort of normal distribution over the x-axis. The scatter plots of questions Q2 and Q3 (not shown for brevity) also showed tendencies of greater point concentration in given regions, which differentiate successful and unsuccessful searchers.

2) Identification of tendencies in heat maps and gaze plots: We evaluated the Tobii heat maps and gaze plots of results pages and content pages of the users, using a scale from 0 to 5 to indicate the presence of the golden triangle and/or a structured (non random) gaze pattern on the screen. The scale was: 0:None, 1:Very little; 2:Little; 3:Some incidence; 4:Quite evident; 5:Very evident. In Table V we see the compiled results for all the SERPs and CRPs. We manually inspected each screen individually, using the evaluation scale of 0 to 5 for the presence of the golden triangle and structured trends. The values in Table V are the average values for all pages and users. We observe that for all cases, the values for the successful users are higher than those of the unsuccessful users, ranging from a 20% difference for CRP/structured-pattern to a 124% difference for the SERP golden triangle.

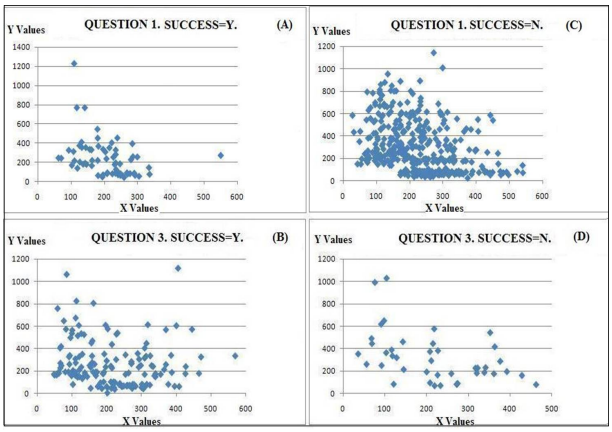

Figure 3. (a to d). SERPs (results pages). Scatter plots of X,Y Gaze coordinates, by question (Q1, Q3) and search success (Yes, NO).

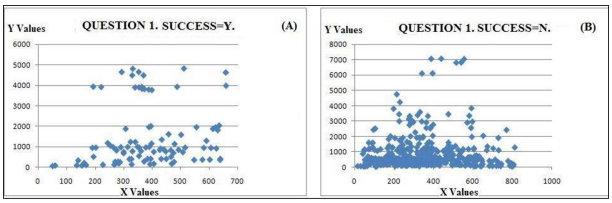

Figure 4. (a, b). CRPs (content pages). Scatter plots of X,Y Gaze coordinates, for question Q1 and search success (Yes, NO).

We can conclude from this data that successful users tend to follow more predefined patterns (such as looking in the golden triangle and being systematic in their eye movement). This is most evident for the SERPs, and for the golden triangle. With reference to Figure 5 (left) we see an example SERP (results page) for a successful search, and in Figure 5 (right) we see an example SERP for an unsuccessful search. In Figure 5 (left), we can see quite clearly a more systematic search concentrated in the upper part of the screen which agrees with the heat map results (see Figure 2), that successful searchers focus in the Golden Triangle for SERP pages. On the other hand, in Figure 5 (right, unsuccessful search) a different path is displayed which seems more chaotic, covering a greater area of the page and involving more time and effort (shown by the greater number of gaze points). This again agrees with the heat maps results for SERP (Figure 2).

As a final comment to this section, we mention that as the

Figure 5. SERPs (Results Pages). Gaze plots for question Q1 and search success (Yes on the left; NO on the right).

Table V
INCIDENCE OF PATTERNS FOR DIFFERENT PAGE TYPES, AND SUCCESSFUL/UNSUCCESSFUL USERS .

	SERP		CRP	
Outcome of Search	Golden Triangle	Structure Trend	Golden Triangle	Structure Trend
Success=YES	2.98	1.93	2.51	2.02
Success=NO	1.33	1.45	1.56	1.69
% difference between Yes and No indexes	124	33	61	20

experiments have dealt with real web pages, noise may be introduced into the gaze statistics by the presence of embedded images, scroll options, and other features/functionality of the web pages.

V. CONCLUSIONS

In this study we have analyzed different types of factors which profile the users and query sessions in terms of successful and unsuccessful searches, taking into account the page type which can be the search engine results page (SERP) or a clicked results page (CRP). The results of the analysis of user search behavior have identified several trends which distinguish between successful and unsuccessful searches. On average, users who are successful formulate fewer queries per session and visit a smaller number of documents than unsuccessful users. The results for fixation frequency and fixation duration of successful and unsuccessful searches suggest that the time that a user spends in SERP and CRP is more related with the strategy of search of the user than with the difficulty of the search task. Based on the scanning behavior of the SERPs and CRPs we have shown that the search strategies of the successful users are more focused and systematic that the search strategies of unsuccessful users.

Now the open question is how Search Engines should react with unsuccessful users? How can Search Engines help this type of users to achieve successful searches? There are some suggested actions that can be derived from this study, for example, Search Engines could detect if a user is spending more time than the established average for a successful search and try to help the user in his task: suggestion of better queries, different presentation of the results and so on. Another important conclusion that we can derive from the eye tracking analysis of the unsuccessful users is the high cognitive effort required to select the results from the list presented by the Search Engine. Through the fixation frequency and fixation duration we can establish that it is not easy for this kind of user to select good results that lead them to the desired information. One suggestion is to enrich the snippets of the search results: if the users spend too much time evaluating the results it could be because they dont find the results sufficiently informative, so Search Engines need to consider this factor and try to offer more descriptive results.

As future work we propose conducting a new experiment, with two versions of the results pages, one with enriched search result snippets, and the other with the default result snippets. Also, by programming an api to the results page, we can detect when the user is spending too much time evaluating the results or showing unstructured search patterns, and offer help at that moment.

ACKNOWLEDGMENT

We would like to thank all the participants of the study. We are especially grateful to AIPO, Tobii Inc. and Alt64,

thanks to which we have temporarily had the eye tracker and the software license.

This research was partially supported by the Spanish MEC (HIPERGRAPH TIN2009-14560-C03-01).

REFERENCES

[1] N. J. Belkin, "Some(what) grand challenges for information retrieval," *SIGIR Forum*, vol. 42, no. 1, pp. 47–54, 2008.

[2] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users," in *Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, 2004, pp. 675–684.

[3] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning user interaction models for predicting web search result preferences," in *SIGIR '06*, 2006, pp. 3–10.

[4] J. Attenberg, S. Pandey, and T. Suel, "Modeling and predicting user behavior in sponsored search," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2009, pp. 1067–1076.

[5] U. Lee, Z. Liu, and J. Cho, "Automatic identification of user goals in web search," in *WWW '05*, 2005, pp. 391–400.

[6] R. Baeza-Yates, L. Calderón-Benavides, and C. González-Caro, "The intention behind web queries," in *SPIRE '06*, ser. LNCS, vol. 4209, 2006, pp. 98–109.

[7] E. Cutrell and Z. Guan, "What are you looking for?: an eye-tracking study of information usage in web search." in *CHI'07*, 2007, pp. 407–416.

[8] B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, and J. K. Newman, "The determinants of web page viewing behavior: an eye-tracking study," in *Proceedings of the 2004 symposium on Eye tracking research & applications*, ser. ETRA '04. New York, NY, USA: ACM, 2004, pp. 147–154.

[9] G. Hotchkiss, S. Alston, and G. Edwards, "Eye tracking study: An in depth look at interactions with google using eye tracking methodology," Enquiro, EyeTools, and Did-It, Tech. Rep., 2005.

[10] J. Nielsen, "F-shaped pattern for reading web content," http://www.useit.com/alertbox/reading_pattern.html, 2006.

[11] A. Aula, R. M. Khan, and Z. Guan, "How does search behavior change as search becomes more difficult?" in *CHI '10*, 2010, pp. 35–44.

[12] A. Aula and K. Nordhausen, "Modeling successful performance in web searching," *Journal of the American Society for Information Science and Technology*, vol. 57, no. 12, pp. 1678–1693, 2006.

[13] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky, "Eye tracking in web search tasks: design implications," in *ETRA '02*, 2002, pp. 51–58.