An Analysis of Searchers' Perceptions of Sponsored and Non-sponsored Links Using Nested Design

Ying Zhang

The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering College of Engineering The Pennsylvania State University University Park, PA 16802 yzz114@psu.edu

Bernard J. Jansen

College of Information Sciences and Technology The Pennsylvania State University 311B Information Sciences and Technology Building University Park PA 16802 jjansen@ist.psu.edu

This paper examines the effects of sponsored links' reputation. We pursued this study through a mature design of experiment method - the nested design - in order to investigate this fast growing segment of online advertising. In this study, we firstly analyze the adaptability of the nested design to find out whether the reputation of sponsored links will bias searchers. We then carefully contrived a user study to investigate this phenomenon. We used 56 participants engaged in six ecommerce Web searching tasks. The tasks were extracted from the transaction log of a Web search engine so that they could represent actual ecommerce searching information needs. We controlled the study by switching non-sponsored and sponsored links on half of the tasks for each participant. By doing this, we could use the nested design statistical method to detect the impact of the sponsored links' reputation on the final evaluations of these results. During the study, we collected 2,435 interactions with links from search engine results page and 961 utterance evaluations of these links. The results indicate that the reputation or the location of the sponsored links will show strong significance on the final evaluation of the search engine performance. We give some perspectives and discussion which could benefit search engine providers to improve their services.

Introduction

Web-based commercial information systems are widely accepted and recognized as challenging to develop. Among the various types of approaches to Web-based information system, Web search engines have been regarded as an important area due to performance and capabilities. Generally, Web search engines often present at least two categories of search results on the results page. One set is composed of non-sponsored links that the search engine determines using its proprietary matching algorithm. The other set is composed of sponsored links that appear because a company, organization, or individual purchased the keywords that the searcher used in the search query (Jansen, 2006).

Sponsored search brings the major search engine companies, such as Google, MSN Live, Yahoo! and AOL lots of revenue, reportedly in the billions of dollars (see Bloomberg.com, http://www.bloomberg.com/). However, the key to whether sponsored search is a viable business model comes from the rate of perceived relevance by the users. To maximize the number of searchers coming to the search engine, the search engine companies have to take specific actions to attract visitors to your site from search sites. This is called search engine marketing. However, users have questioned the relevance of the sponsored links because of their advertising characteristics. Therefore, it is a trade off for the search engine companies in terms of the revenue and customers.

In responding to this dilemma for the search engine companies, this paper reports the results of a carefully designed research study that investigates the interaction between the searchers and the sponsored links. Few studies focus on the reputation of the sponsored links, although there have been some in this general area (Hotchkiss, 2004; Marable, 2003). These studies rarely give the statistical results, which will be more reliable and adaptable to universal conditions.

After comparing related studies of the sponsored links area in the next section, we will introduce a statistical method named nested design and discuss its adaptability to our study. Later, we present our user study designed to match the 'nested design' method. Finally, we will give the experimental results and some discussion and implications for the future research.

Related Research

Web research is now a major interdisciplinary area of study, including the modeling of user behavior and Web search engine performance. Some prior research has established a potential disconnect between the perception of sponsored listings by business and users. Businesses spent $8.5 billion on paid search in 2004, and this amount is expected to grow to $16 billion by 2009 (Jarboe, 2005). However, if these sponsored links are not perceived as relevant to the customer the search engines have the risks of loosing their customers for the poor performance of the sponsored listings (Hotchkiss, 2004).

Prior research on sponsored listings focus on the area of searching methods, personal preference, demographic factors and results characteristics (Jansen & Renick, 2006). Web searching can be a complex, iterative process that evolves as the search progresses. Cognitive search costs reflect consumers' efforts to formulate search tasks, select queries, filter incoming information, and integrate new information with existing information to form decision evaluations (Hauser, Urban and Weinberg, 1993). However, empirical studies have shown that the 'typical' Web searcher has little understanding of how search engines retrieve, rank or prioritize links on the results page (Marable, 2003), so the searching paradigms may not be valid. Using results from a user study, Marable (2003) reports that searchers trust search engines to present only unbiased results on the first page, not realizing that 41% of selections were sponsored links. When informed of the nature of the sponsored listings, participants reported negative emotional reactions. Search engines that were less transparent about sponsored search results lost credibility of the customers.

This is in line with results from an 11-month investigation sponsored by the Federal Trade commission recommending that search engine companies clearly mark paid listings on their sites (Hansen, 2002). When users suspect that search engines are intentionally disguising the presence of paid listings, ecommerce searchers may be less likely to consider them. Hotchkiss (2004) used an enhanced focus group format to observe the search behaviors of 24 participants and interviewed them for their reactions to what they saw online. The researcher found that searchers take keywords from the sponsored listing descriptions to use in future iterations of their search process. As the search process becomes more focused, the likelihood that users will consider the paid listings increases.

Other studies have focused on searcher biases toward or against sponsored links and the effect of those biases for businesses and organizations in attaching potential customers. Langford (2000) conducted an investigation of various online advertising media reporting that online-only promotions are of little value in attaching new customers. Greenspan (2004) found that users prefer organic listings relative to sponsored links. This study also raised ethical issues regarding how search engines present sponsored listing. Brooks (2004) found that the likelihood of a searcher selecting a sponsored link is a curvilinear function of its placement on the page. The higher the link's placement in the results listing, the more likely a searcher is to select it.

Investigating search engine loyalty and interaction with Web search engines, iProspect Inc. (2004) surveyed 1,649 Web users. Of the respondents, 60% of Google users reported organic results to be more relevant than sponsored. This was even higher for predominantly Google users (70%). Frequent users of the Web found organic results to be relevant than sponsored results (65% to 56%). More women (43%) than men (34%) found sponsored results to be generally relevant.

From the above prior work, we can tell that sponsored links have a negative evaluation from the most Web searchers, but they have become a major resource for search companies. In order to balance its advantages and disadvantages, this research was designed to detect whether the previous existing attitude toward the sponsored links causes the low evaluation and accumulative negative attitude. If that is, the companies could find some solutions such as adjusting its viewing positions or enhancing the presentation to eliminate the customers' bias.

Research Question

Our research question is whether this existing negative attitude toward the sponsored links by Web searchers results in low relative evaluations.

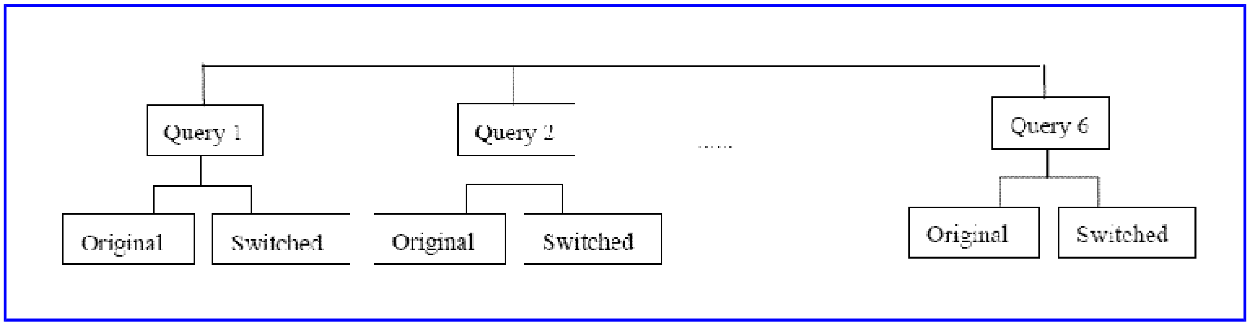
Research Method

Our goal in this research is to exam whether the customers' attitude or perception regarding sponsored links would impact the searchers' evaluation of those links. This is critical and timely research given the increased focus on Web personalization and the desire to provide more supportive Web systems, along with online advertising and targeted advertisements. The research reported in this paper extends results reported in (Jansen & Renick, 2006).

Noting our experiment's purpose, we designed a user study to detect the significance of the sight impact for the sponsored links. Generally, traditional ANOVA method is used to detect the significance of specific factors. However, in order to clear the obscurity caused by other nuisance factors, we have to choose as many potential factors as possible. For example, the searcher's educational background, searching skills, ages, time for doing the user study, etc. In a sense, ANOVA could not completely detect all the potential factors and make an exact conclusion for what we are interested in, which was the method used in (Jansen & Renick, 2006). Therefore, there will be a problem that we could not get a crossed experiment since all the users are different. Fortunately, potential factors could be presented as different searchers. We put all the other potential factors together and make them represented by different users. For this reason, we turn to the nested design, a well designed statistical method.

A 'nested' or 'hierarchical' design is an arrangement when the levels of one factor are similar but not identical for different levels of another factor (see figure 1) (Douglas, 2005). We could use Design-Expert software by properly designating effects in specific ways to do the analysis of a nested design. As mentioned before, in our user study, we had to consider as many factors as possible so that our results will not be hidden by potential undetected factors or interactions. Additionally, since the different queries have already been noted as a significant factor to the final results (Jansen & Renick, 2006; MaKris, 2006), in order to eliminate the effect of it on our result, we consider the query as a general factor. In this case, we use the common factor—query content as our first factor. In order to cover all the aspects of the commercial terms, we extracted a set of ecommerce queries from an approximately one million query Excite transaction log using a modified snowball technique, explained in the Research Design section. We regard this factor of query as a fixed factor. What is more, we could not use a single person to do the research because individual's evaluation will obviously be biased. Under different levels of the queries, we could not use the same group of the searchers since randomly selected users could represent the typical and reasonable results. Therefore, in order to eliminate the potential factors and interactions in the study, we consider another two level factor, which points to different group of searchers assigned different first search engine results page (SERPs).

Figure 1: the nested design for the user study



Research Design

In order to investigate our research questions, we first extracted a set of ecommerce queries from an approximately one million query Excite transaction log (Spink & Jansen, (2004) using a modified snowball technique). In this case, we select six queires representing three categories of ecommerce query types: general, specific and location specific. We then submitted these six queries to Google, a major U.S. search engine, using a software application that submitted the queries and also also retreved the SERP for each query exactly as it would be presented to a human user. The total time from submission to completing of result retrieval took just more than 30 seconds. In order to block the nuisance factors, we then removed all identifying logos, text, uniform resource locations (URL), and HTML code from the Google result pages, replacing them with a fictitious search engine identifier. And in order to simplify the procedure for our study, we disabled all hyperlinks to other result pages and the form submit button. Besides, we removed the redirects in the organic and sponsored results, so the URLs pointed directly to targeted Web site. We then had six SERPs (one for each of the six queries) with ten organic links and five sponsored links. We refer these SERPs as the orginal pages.

We then use each original page to create a second page, referred to as the swiched page. For the switched page, we switched the five sponsored links and the top five organic links. We did this to find the impact of the sponsored links. We manipulated only the top five organic links because most users do not scroll down past the top results on the top results on the page (Hotchkiss et al., 2004; Jansen & Spink, 2003). This process provided us with six Web search engine result pages with what looked like ten organic links and what looked like five sponsored results.

We conducted the study simultaneously at two locations, each a major U.S. university. Both locations followed the same procedure and used the same instruments. We begin with demographic characteristics of the participants. We recruited 56 participants between both campuses. The age range was restricted to 18-29 to focus the study on the demographic most valued by marketers. We explained to each participants the purpose of the study as an investigation into searching methods and obtained informed consent. Before the user study, we used the design expert to generate the design matrix to guide the choose of the level of each factor and assigned the random order to every participant.

After establishing the study environment, we started the research. For each participant, a moderator read the participant a short introduction. For each experimental task, we explained the task to the participant and reminded the participant to think aloud. We used an unrelated practice task to explain the use of the verbal protocol method. We then read the participant one of the six ecommerce searching scenarios, informed him/her that the query had already been entered into the search engine, opened the appropriate Web page, and asked the participant to continue the search. The participants would then continue the search as if they had entered the query. The session for that query would end when the participant took some action that would remove them from the presented results page. We instructed the participants to describe the screen content they were viewing, evaluate its relevance to the task, and explain why they moved from one item to the next.

We presented each participant with all six queries, one at a time. Each participant completed one query before moving to the next. The moderator would read the applicable scenario before moving to the next query. For each participant, three of the result pages were original and three were switched, so that we could get the balanced design. According to the design matrix, we counterbalanced the order of original and switched result pages within each participant's sessions and between each participant. During the searching sessions the moderators did not assist the participants; however, the moderator would answer procedural questions. While the participant was searching, the moderator annotated utterances and user actions using an application that the researchers designed for quantitative and qualitative data capture for Web searching studies, such as this one. In order to get some complementary materials, the participants were asked to completed a demographic questionnaire and answered questions about his/her opinions regarding paid listings in general after their major tasks. Approximately one hour was required to complete the sequence for each participant.

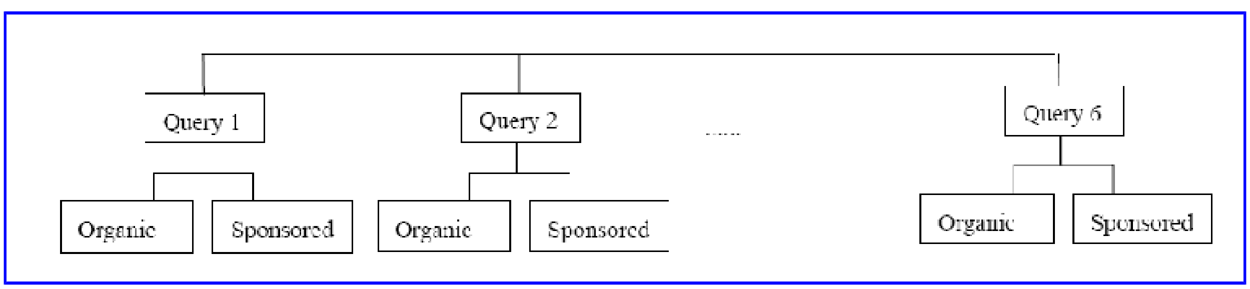
Data Analysis

Note that our experiment is designed to find the impact of the sponsored links relevant to searcher's information needs. As stated in the Research Methods section, we turned to the nested design to do the data analysis. We choose the a reasonable arithmetic to calculate the final evaluation as the responses:

$$Overall\ Evaluation = \frac{sum(V\_flag \times V\_link\ evaluation)}{sum(V\_flag)}$$

In order to detect the impact, we should firstly find the users' general reponses to the sponsored links, and we refer this designed experiment as Experimnet 1 (Exp1). To do that, we sorted the records in terms of the users who were required to evaluate the unswitched SERPs. In this case, factor B referred to whether the evaluation is for the organic or not. And as mentioned before, factor A referred to the type of queries. The complete nested design is shown as Figure 2. After doing this, we could switch to the design shown in Figure 1, which we denote it as Experiment 2 (Exp2) here. For the level of switched SERPs, users thought they were organic links but actually they were not. On the contrary, for the previous design shown in Figure 2, users were not blinded because they give the evaluation for the genuine sponsored links. Thus the major differece between these two designs is whether the users were blinded while other elements are the same.

Figure 2: the nested design for judging the bias to the sponsored links



In our experiment, the factor of selected people to do the evaluation is a random factor that is nested within the queries. The query is a fixed factor because we use a modified snowball technique to generate the specific queries. The expected mean squares (EMS) could be calculated as below:

$$\sigma^2 + n\sigma_B{}^2 + \frac{bn\sum A^2}{a-1} = \sigma^2 + 28\sigma_B{}^2 + 11.2\sum A^2 \quad (2)$$

The factor of the queries, which we referred it as factor A, is fixed and the factor of the types of the SERPs, which we refferred it as factor B, is random. According to the rules for expected mean squares (EMS), we generated Table 1 to calculate the EMS. And then we could get the appropriate test for significance of factor A:

$$F = \frac{\sigma^2 + 28\sigma_B{}^2 + 11.2\sum A^2}{\sigma^2 + 28\sigma_B{}^2} \quad (3)$$

Table 1. the Calculation of EMS

Factor	d.o.f	Expected MS
query	5	$\sigma^2 + n\sigma_B{}^2 + \frac{bn\sum A^2}{a-1} = \sigma^2 + 28\sigma_B{}^2 + 11.2\sum A^2$
type	6	$\sigma^2 + n\sigma_B{}^2 = \sigma^2 + 28\sigma_B{}^2$
Error	324	σ^2
Total	335	

Since the factor B is nested within the factor B, there can be no true interaction AB. Therefore, the sum of squares is $SS_{B(A)} = SS_B + SS_{AB}$. Then to perform the analysis, we should regard factor B and interaction of AB as the error in order to test the significance of the effect of the queries.

After the analysis for the Exp1 in the design expert (Table 2), we could tell that the factor A could be regarded as significant to the final evaluation. Recommended by the Box-Cox Plot (Figure 3), we do the square root transformation to the responses (Figure 4). And the ANOVA table (Table 3) shows that users have bias according to the sponsored links.

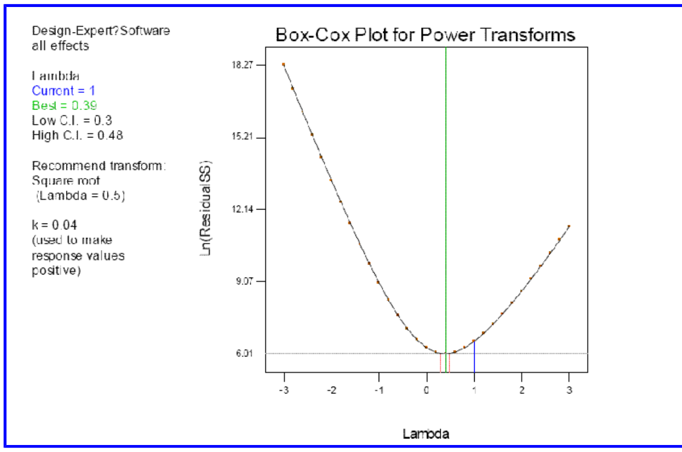


Figure 3: Box-Cox plot before the transformation in Exp1

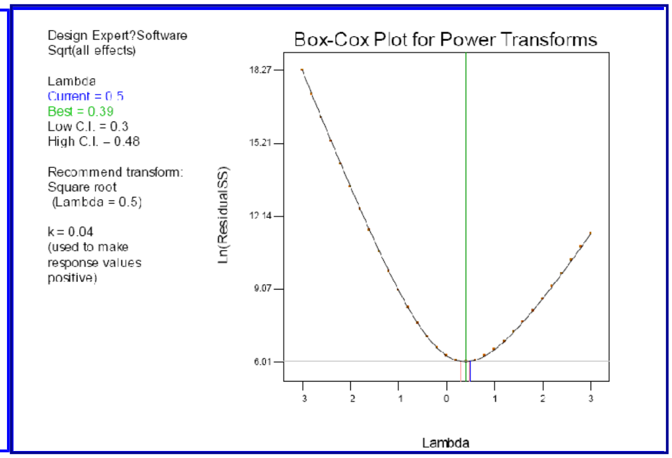


Figure 4: Box-Cox plot after the transformation in Exp1

Table 2. ANOVA table for factor A in Exp 1

Analysis of variance table [Classical sum of squares - Type II]						
Source	**Sum of Squares**	**df**	**Mean Square**	**F Value**	**p-value Prob > F**	
Model	53.56	5	10.71	28.61	0.0004	significant
A-query	53.56	5	10.71	28.61	0.0004	
Residual	2.25	6	0.37			
Cor Total	55.81	11				

Table 3. ANOVA table for factor B in Exp 1

Analysis of variance table [Classical sum of squares - Type II]						
Source	**Sum of Squares**	**df**	**Mean Square**	**F Value**	**p-value Prob > F**	
Block	9.15	5	1.83			
Model	41.25	6	6.87	12.02	< 0.0001	significant
B-type	39.75	1	39.75	69.51	< 0.0001	
AB	1.5	5	0.3	0.52	0.7585	
Pure Error	185.28	324	0.57			
Cor Total	235.68	335				

Then we went to Exp 2 and follow the same steps as Exp 1. Again, recommended by the Box-Cox Plot (Figure 5), we do the 0.86 power transformation to the responses (Figure 6). The ANOVA table (Table 4) shows that users have no bias according to the sponsored links. From the above analysis, we could tell that searchers are blinded by our design because they give the same evaluation to the different pages.

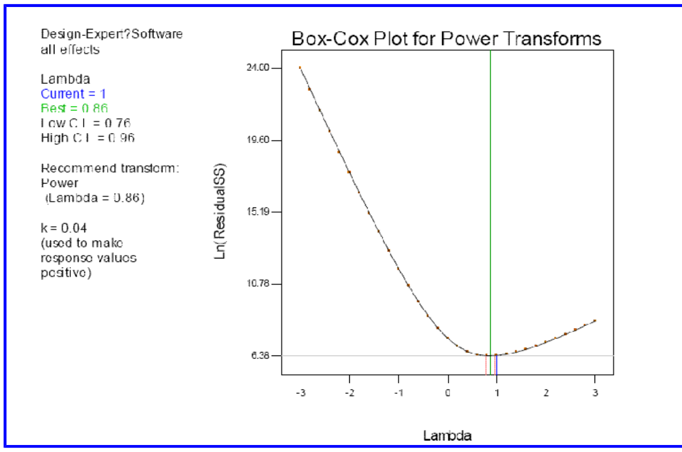


Figure 5: Box-Cox plot before the transformation in Exp 2

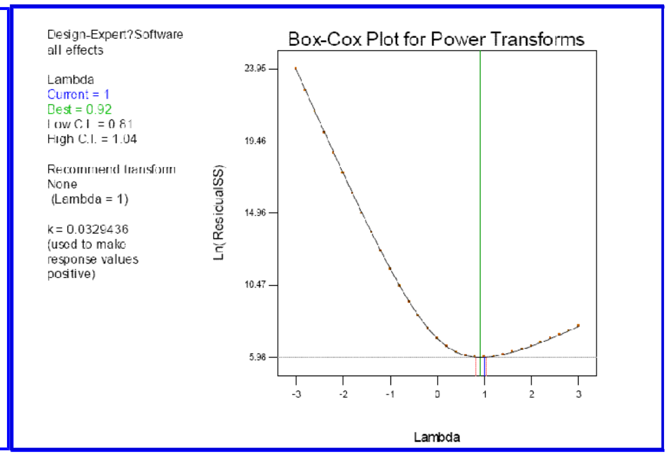


Figure 6 Box-Cox plot after the transformation in Exp 2

Table 4. ANOVA table for factor B in Exp 2

Analysis of variance table [Classical sum of squares - Type II]						
Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Block	84.18	5	16.84			
Model	3.31	6	0.55	0.3	0.9376	not significant
B-type	0.13	1	0.13	0.07	0.7913	
AB	3.18	5	0.64	0.34	0.8862	
Pure Error	599.05	324	1.85			
Cor Total	686.54	335				

Discussion

We conducted a laboratory study investigating searcher biases toward sponsored links. Based on the data analysis results, people will give the negative evaluation for the sponsored links even though the results are exact the same. The major contribution of this paper is that we found the searchers' negative attitude will disturb their genuine judgments. Namely, searchers will be blinded by the reputation of the sponsored links just because of their previous negative attitudes toward commercial implications within the sponsored links. They will give the lower evaluation based on their attitude toward the sponsored links, rather than the links' contents. So, search engine companies could leverage this finding to achieve the marketing lead qualification and more revenue. For instance, perhaps changing the position or the viewings of the sponsored links will be an effective way to increase the possibility of user clicking on the links. Changing their position and viewings may somewhat eliminates searchers' previous biases and inspires their curiosity to enter the links. Another method may be a marketing campaign to education searchers on the relevance of sponsored links.

Additionally, this paper introduced a professional statistical method — nested design — to do the analysis and also explained the adaptability. To some extent, it extended the traditional method to do the data analysis. In our designed experiments, we put all the other potential factors except queries together and make them represented by different users. In a sense, this nested designed method simplifies the procedure and gives more reasonable conclusion than the traditional ANOVA method.

In the future research, we would like to evaluate a broader range of ecommerce queries to identify specific query characteristics that might predict the viewing of sponsored links. This would also facilitate the identification of searcher, system, or content factors that contribute to the present searcher bias against sponsored links. This will aid the expansion of the sponsored search market to wider range of Web searchers, helping to ensure the growth of this market. We could also turn to the information retrieval method and study the inner searching mechanism in order to explore some creative suggestions to improve the search marketing.

References

Brooks, N. (2004, July). The Atlas Rank Report I: How Search Engine Rank Impacts Traffic. Retrieved 1 August, 2004, http://www.atlasdmt.com/media/pdfs/insights/RankReport.pdf

MaKris, Christos, Yannis Panagis, Evangelos Sakkopoulos, Athanasios Tsakalidis. (2006). Category ranking for personalized search. *Data & Knowledge Engineering,* 60, 109-125

Montgomery, Douglas C. (2005). Design and Analysis of Experiments. *John Wiley & Sons, Inc* Wiley, NY

Greenspan, R. (2004). Searching for balance. Retrieved 4 May, 2004, www.clickz.com/stats/big_picture/applications/article.php/3348071

Hansen, E. (2002). FTC wants paid search to shape up. Retrieved 14 January, 2005, http://news.com.com/2100-1023-940598.html

Hauser, J. R., Urban, G. N., & Weinberg, B. D. (1993). How Consumers Allocate their Time When Searching for Information. *Journal of Marketing Research,* 30, 452-466.

Hotchkiss, G. (2004). Inside the Mind of the Searcher. Retrieved 15 March, 2005, http://www.enquiro.com/research.asp

iProspect Inc. (2004, April-May). Search Engine User Attitudes. Retrieved 28 March, 2005, http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf

Jansen, B. J., & Renick, M. (2006). An Examination of Searcher's Perceptions of Organic and Sponsored Links during Ecommerce Web Searching. *Journal of the American Society for Information Science and Technology.* 57(14), 1949-1961

Jarboe, G. (2005, 17 January). Online Ads & Search: Looking Back, Looking Forward. Retrieved 18 January, 2005, http://searchenginewatch.com/searchday/article.php/3459961

Langford, B. E. (2000). The Web Marketer Experiment: a Rude Awakening. *Journal of Interactive Marketing,* 14(1), 40-48.

Marable, L. (2003). False Oracles: Consumer Reaction to Learning the Truth About How Search Engines Work, Results of an Ethnographic Study *Consumer WebWatch.* (Research Report)

Spink, A., & Jansen, B. J. (2004). Web Search: Public Searching of the Web. *Kluwer* New York