THE EFFECT OF BRAND ON THE EVALUATION OF IT SYSTEM PERFORMANCE

Bernard J. Jansen
The Pennsylvania State University
jjansen@acm.org

Mimi Zhang
The Pennsylvania State University
mzhang@ist.psu.edu

Carsten D. Schultz
University of Hagen
carsten.schultz@fernuni-hagen.de

ABSTRACT

In this research, we investigate the branding effect of an information technology system on the evaluation of system performance. Using Web search engines, we conducted a laboratory experiment measuring the effect of four search engine brands while controlling for the quality and presentation of search engine results. There was a 25 percent difference between the most highly and the lowest rated search engines based on average relevance ratings by users among the four search engines presenting identical results in both content and presentation. A positive brand is worth approximately fifteen percent in user perception of performance. Users place a high degree of trust in major search engines, but they are more engaged in the searching process when using lesser known search engines. It appears that branding affects overall Web search from various perspectives. We discuss implications for search engine marketing and the design of empirical studies measuring search engine performance.

Keywords

branding, brand, Web searching, search engines

INTRODUCTION

There are numerous search engines on the Web; however, only a handful dominates in terms of usage (http://searchenginewatch.com/showPage.html?page=3626208). From a technological point, this clustering of traffic is interesting because studies report that the performance of most of the major search engines is practically the same (c.f., Eastman et al. 2003). Performance is typically defined as returning useful results and is measured by precision, which is the ratio of relevant documents to the total number of document returned at some point in the results listing. The interfaces of most search engines are also similar, namely a text box, some verticals (i.e., tabs for searching the Web, Images, Audio, etc.). In studies of search engine interface usability, the results among search engines has been similar (c.f., Wildemuth et al. 2002). Given the similarity in terms of technology and interface design, why do only a small number of search engines dominant Web traffic? Do other elements affect the evaluation of a search engine's performance? Seeking the answers to these questions motivates our research.

In a series of user studies concerning Web searching (Jansen 2006; Jansen et al. 2005a), the participants completed pre-surveys concerning their Web searching habits. One question addressed which search engine the participant used and why. One response was surprising – *popularity*. Nearly 14 percent of the participants listed *popularity* being as being a reason for using a search engine, making comments such as '*Google, who doesn't!*' and '*It is the most widely known*'. The popularity of *popularity* being a major reason for search engine usage led us to investigate brand (i.e., identifying name that distinguishes a product from its competitors) as a possible reason to explain the clustering of traffic around a handful of an array of extremely similar Web search engines. This approach is supported by statistics showing that Google and Yahoo! are some of the most talked about brands on the Web (http://www.e-consultancy.com/news-blog/363695/google-yahoo-apple-and-microsoft-most-talked-about-brands-online.html). In this research, we measure the effect of brand on user perception of the performance of Web search engines.

In the following sections, we briefly review the concept of branding and its lack of emphasis in Web search engine design. We then present our research questions and approaches, following this with our research results. We end with the implications of our findings for the effect of brand on information technology (IT) system performance evaluation.

BRANDING LITERATURE

Depending on the perspective, the understanding of a brand differs across the various academic and practical disciplines. A narrow brand perspective centers on tangible brand features, such as name, design, or symbol, while intangible features, such as values, ideas, and personality, are included by a broader brand perspective (e.g., de Chernatony et al. 1998; Haigh et al. 2004; Stern 2006). In the present research, we adopt a broad brand perspective. Attributing a brand with intangible features also recognizes the importance of an individual or customer's brand perception.

Research centering on the individual's perspective of brand has examined various impacts of brand elements. Brand effects have been studied as antecedences of online trust relating to the vendor, to the Website, and the product, as well as a means to communicate the trustworthiness of an e-vendor (for an extensive analysis, see Schultz 2007). These brand concepts are strongly interrelated and represent various stages and aspects of an individual's brand perception and processing. Ha and Perks (2005) examined the relationship of brand experience, brand familiarity, customer satisfaction, and brand trust in the online environment. In another study, Esch and fellow researchers (2006) propose a conceptual model to relate perceptual (brand awareness and brand image) and relationship (brand satisfaction, brand trust, and brand attachment) variables to current as well as future purchasing behavior.

RESEARCH OBJECTIVES

Our research objective is: *How does branding affect overall user evaluation of results retrieved by Web searching systems?*

To address this research question, we conducted a study that altered the brand of search engines for a set of queries while controlling for the quality and display of the results. Our hypothesis is: *There will be no difference in the evaluation of links among search engines.*

We report the specifics of our design in the following section.

RESEARCH DESIGN

Data Preparation

To investigate our research question, we first extracted a set of e-commerce queries from an approximately one and half million queries Web search engine transaction log using a modified snowball technique. From these queries, we selected four queries representing four searching domains: medical, entertainment, travel, and ecommerce. We developed searching scenarios around each of the four queries. The four queries used were: *camping mexico*, *laser removal*, *manufactured home,* and *techo music*.

We then submitted these four queries to a major U.S. search engine (i.e., Google) using a software application that not only submitted the queries but also retrieved the first search engine results page (SERP) for each query exactly as it would be presented to a human user. The total time from submission to completion of result retrieval took approximately 30 seconds. We then removed all identifying logos, text, URLs, and HTML code from the Google result pages. We removed the redirects in the results, so the URLs pointed directly to the targeted Web site. This left us with four cleaned results pages.

We then got screen captures of SERPs from Google, MSN Live Search, and Yahoo!, all major and well-known Web search engines, for each of the four queries. Additionally, we developed an in-house search engine, AI^{2}RS, and got screen captures of the AI^{2}RS results pages for each of the queries.

Using the cleaned Google results and the images from the AI^{2}RS, Google, MSN Live Search, and Yahoo!, we developed four experimental SERPs for each of the four queries. At the end of this process, we had sixteen experimental SERPs, four from each search engine for each of the four queries. However, regardless of the search engine branding elements, the results were identical across all search engines for each query. Figure 1 shows the building of an experimental SERP.

We cropped each SERP image using only the branding elements at the top of the SERP (i.e., logo, search box and button) and bottom (i.e., results page hyperlinks) of each image. We then built a hyperlink page structure to hold the top and bottom images. For the search engine results, we used the cleaned Google results.
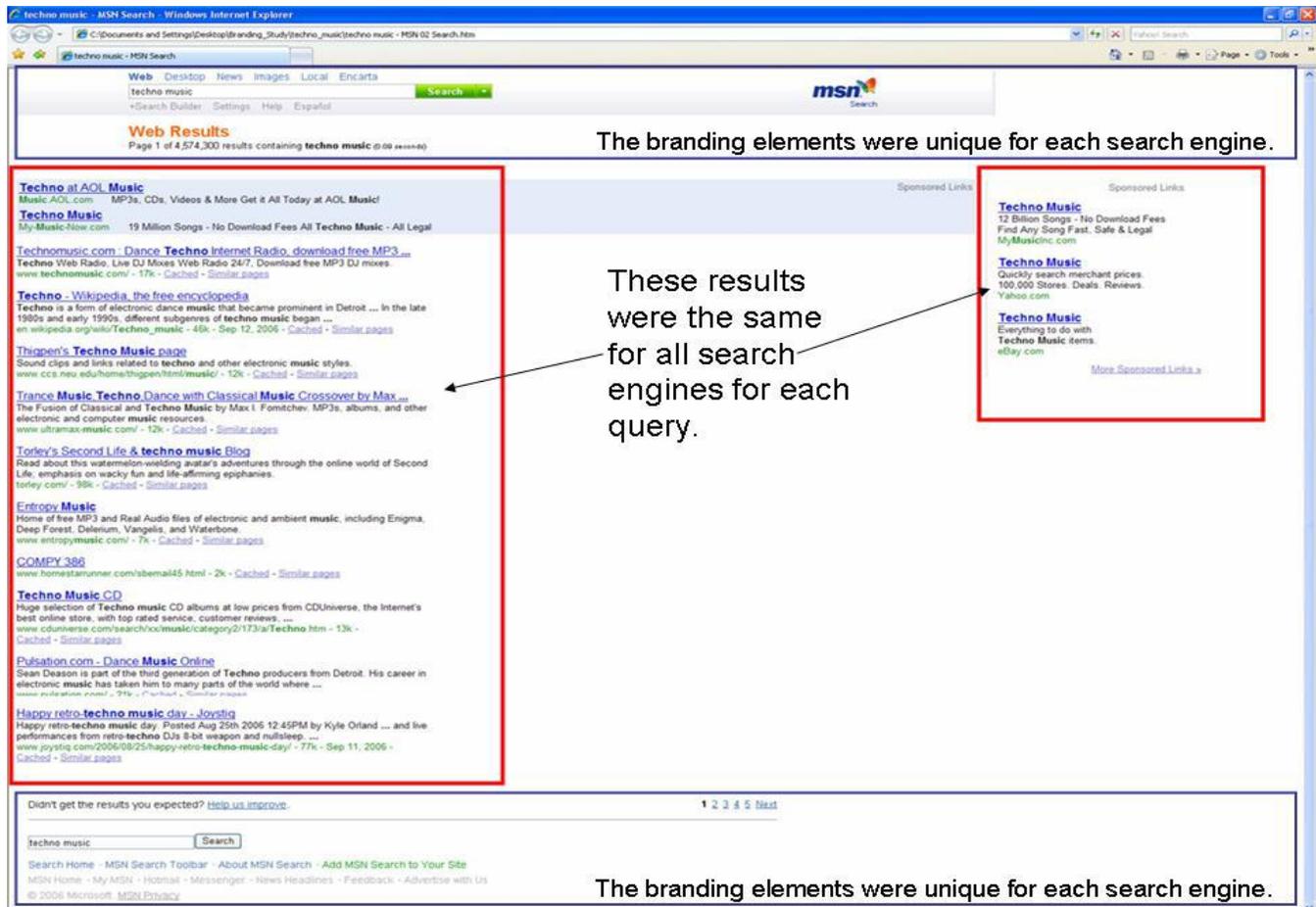
Figure 1. Example of an Experimental Results Search Engine Page

Our goal in this process was to be able to isolate the effect of the branding variable while controlling for the number of results, result presentation, and quality of research. We used only the first SERP for each query because most searchers only view the first results page (Jansen et al. 2005b; Jansen et al. 2000).

We decided to use one style of results formatting because prior work has noted that minute differences in the presentation of search engine results can affect how users interact with those results (Hothkiss 2006). There have been other studies of search engine performance, but we wanted to control for variation in the quality of results.

Study Procedure

We recruited 32 participants from a major US university. The age range was 18 to 25 years. There were 8 females and 24 males. Prior to the search tasks, the participants completed a demographic questionnaire and answered questions about his/her Web searching include the search engine(s) most frequently used. Concerning what search engines participants reportedly used, Google was mentioned by 31 participants, Yahoo! by 10, Dogpile by 2, and AltaVista, Naver, and MSN by one participant each. Participants would list more than one search engine, which is why the total is more than 32.

We presented each participant with all four 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. We counterbalanced the order of search engines and the order of the searching scenarios to eliminate ordering effects. For each participant, a moderator read the participant a short introduction, explained to each participant that he/she would be conducting some searches using Web search engines, and reminded the participant to think aloud. We used an unrelated practice task to explain the think aloud protocol.

We then read the participant one of the four searching scenarios, informed him or her that the query already had been entered into the search engine and results returned, and asked the participant to continue the search. The participant would then continue the search as if he or she 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 without returning (i.e., submit a new query, go to a new

results page, go to a different search engine, etc.). The moderator 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.

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 during Web searching studies such as this one. After the participant had completed all four query sessions, the moderator returned the participant to the first query, and the participant visited all Web pages for each query that the participant had not visited during the session. The participant evaluated the Web document and presented a basis for the evaluation. The moderator collected these Web document evaluations again using the data collection application. Approximately one hour was required to complete the sequence for each participant.

RESULTS

We now return to our research question (*How does branding affect overall user evaluation of results retrieved by Web searching systems?*) with precision results and differences among results shown in Tables 1 and 2.

Search Engines	Queries				Average
	camping mexico	laser removal	manufactured home	techno music	
AI^{2}RS	0.35	0.31	0.26	0.37	0.32
Google	0.26	0.25	0.69	0.27	0.36
MSN	0.44	0.29	0.30	0.34	0.34
Yahoo	0.39	0.29	0.55	0.44	0.42
Average	0.36	0.28	0.45	0.35	0.36

Table 1. Comparison of Average Precision Scores by Query and by Search Engine

Search Engines	Queries				Difference from Average
	camping mexico	laser removal	manufactured home	techno music	
AI^{2}RS	-2.0%	10.9%	-42.9%	5.7%	-10.3%
Google	-28.5%	-12.6%	52.2%	-24.5%	0.7%
MSN	21.9%	0.8%	-32.3%	-5.1%	-5.7%
Yahoo	8.6%	0.8%	23.0%	24.0%	15.3%
Average	0.0%	0.0%	0.0%	0.0%	

Table 2. Comparison of Differences of Average Precision Scores by Query and by Search Engine

We see from Table 1 that the average precision rating for the search engines across all four domains was 0.36, meaning that about 36 percent of the results were judged relevant to the query.

In Table 2, we present the difference in average precision ratings for each search engine. AI^{2}RS, the unknown brand fared the worst – with an average precision rating of 10 percent below the average. Yahoo! had the highest rating at 15 percent above average. Surprising, given the stated preference by the participants, Google's rating were only slightly better than average.

It certainly appears that lack of a brand was a detrimental factor for the AI^{2}RS search engine, with an average precision 10 percent below average. Google, used most often by the study participants, after analyzing the demographic questionnaires, had an average precision just above the norm. However, Google was below average in three of the four domains. Yahoo! performed the best with above average precision ratings across all four domains.

It appears that even though Google is the most commonly used engine for searching. Yahoo! has a positive branding awareness. This may help explain why Yahoo! has endured and prospered in a competitive marketplace where so many other search engines (c.f., Excite, Northern Light, and Infosearch) have come and gone. Of all the search engines, Yahoo! had the overall best average precision of 0.42. This was 15.3 percent better than the average of all four search engines.

Concerning our hypothesis (*There will be no difference in the evaluation of links among search engines*), a binomial test using a mid-point value shows that there were no significant differences in the evaluation of links among the search engines. The average precision for all 128 SERPs was 0.36. The average precision was calculated as the sum of the participants' link

evaluations divided by 640, resulting from the 128 SERPs multiplied by 15 links. We normalized the user ratings so that the precision range would be 0 (i.e., all links not relevant) to 1 (i.e., all links relevant).
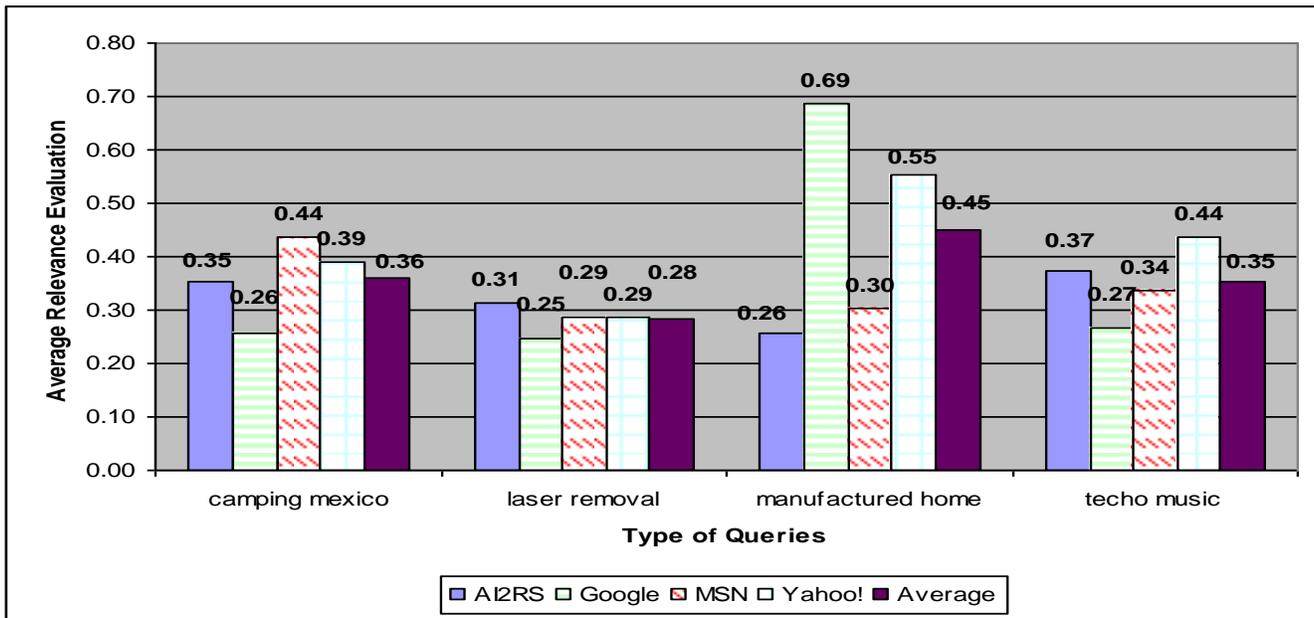

Figure 2. Graphical Comparison of Average Precision Scores by Query and by Search Engine

When we look at just the links that the participants examined during the searching session (rather than all the links), the average precision scores for the SERPs were naturally higher with an average of 0.66 across all search engines. Yahoo!'s average precision was 0.67, Google's was 0.69 MSN's was 0.71, and AI^{2}RS's was 0.64. Therefore, the participants seemed to be choosier concerning links with the non-main stream search engines.

Given Google's standing in the searching market at the time of the study, it may be surprising to some readers that Yahoo! fared so well. However, other research also supports Yahoo!'s positive image in the marketplace (c.f., Freed 2007). However, when the subjects were asked to examine each link outside of the normal searching scenarios, the Google brand really had a positive impact, as their average precision increased. Again, prior research shows that Web searchers place a lot of trust in the ranking of links by Google, and one can assume other major search engines (Pan et al. 2007).

IMPLICATIONS

In this experiment, we studied the effect of branding on how users evaluate search engine performance. Regardless of which search engine a participant used for a particular domain, the results for each query were the same. However, there were dramatic differences in how participants rated the performance of each search engine using relevance of retrieval results. Performance evaluation results varied by more than 25 percent between the top-most rated search engine and the bottom rated. Again, this difference was noted even though all the results were identical in both content and presentation. The implications of these research findings give empirical weight to the notion that affective and cognitive user perceptions affect user interaction with systems. Therefore, product brand is an important usability variable in system design and evaluations of search engines.

Brand appears to be a big positive for the two mainstream search engines (Google and Yahoo!). They obviously have a significant marketplace advantage in the service branding. Google, used most often by the study, had an average precision just above the norm. However, Google was below average in three of the four domains. Therefore, there is certainly a domain effect. Yahoo! performed the best with above average precision ratings across all four domains. It appears that even though Google is the most commonly used engine for searching. Yahoo! has positive brand awareness. This may help explain why Yahoo! has endured and prospered in a competitive marketplace where so many other search engines (e.g., Excite, Northern Light, and Infosearch) have come and gone. The implications of these research findings give empirical weight to the notion that affective and cognitive user perceptions affect user interaction with systems and interactions. Therefore, product brand is an important usability variable in system design and evaluations.

Finally, the implications for Website designers and content providers are clear. It does not matter what search engine sends a site the traffic. Once the user leaves the search engine, the branding aspects of the Website take over. The relevance of the content to the user's query or information need, the user's perceived professionalism of the page, the user trust in the site, along with other factors such as load time all affect the user's positive or negative view of the Website brand. Therefore, once the user is at the Website, the onus is on the content providers to convert the visit into providing value.

CONCLUSION

In this research, we investigated the effect of branding on the evaluation of the system performance of Web search engines. Study findings show that branding as a perception of product has a dramatic effect on user's evaluation of system results. Future research involves in-depth quantitative and qualitative analysis of experimental data, a series of experiment to tease apart the nuanced relationship between perception of system performance and product brand, and how to incorporate branding into the system design.

ACKNOWLEDGMENTS

REFERENCES

1. de Chernatony, L., and Riley, F.D. "Defining A "Brand": Beyond The Literature With Experts' Interpretations," *Journal of Marketing Management* (14:4/5) 1998, pp 417-443.
2. Eastman, C.M., and Jansen, B.J. "Coverage, Ranking, and Relevance: A Study of the Impact of Query Operators on Search Engine Results," *ACM Transactions on Information Systems* (21:4), October 2003, pp 383 - 411.
3. Esch, F.-R., Langner, T., Schmitt, B.H., and Geus, P. "Are brands forever? How brand knowledge and relationships affect current and future purchases," *Journal of Product and Brand Management* (15:2) 2006, pp 98-105.
4. Freed, L. "American Customer Satisfaction Inedx Annual E-BusincessReport," ForeSee Results, pp. 1-11.
5. Ha, H.-Y., and Perks, H. "Effects of consumer perceptions of brand experience on the web: Brand familiarity, satisfaction and brand trust," *Journal of Consumer Behaviour* (4:6) 2005, pp 438-452.
6. Haigh, D., and Knowles, J. "How to define your brand and determine its value," *Marketing Management* (13:3) 2004, pp 22-28.
7. Hothkiss, G. "Eye Tracking Report: Google, MSN, and Yahoo! Compared," Enquiro, Kelowna, BC, Canada
8. Jansen, B.J. "Using temporal patterns of interactions to design effective automated searching assistance systems," *Communications of the ACM* (49:4) 2006, pp 72-74.
9. Jansen, B.J., and McNeese, M.D. "Evaluating the Effectiveness of and Patterns of Interactions with Automated Searching Assistance," *Journal of the American Society for Information Science and Technology* (56:14) 2005a, pp 1480-1503.
10. Jansen, B.J., and Spink, A. "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," *Information Processing & Management* (42:1), Available online 7 January 2005 2005b, pp 248-263.
11. Jansen, B.J., Spink, A., and Saracevic, T. "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing & Management* (36:2) 2000, pp 207-227.
12. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., and Granka, L. "In Google We Trust: Users' Decisions on Rank, Position, and Relevance," (12:3) 2007, p article 3. http://jcmc.indiana.edu/vol12/issue13/pan.html
13. Schultz, C.D. *Consumer Trust in E-Commerce - An Analysis of Means Communicating Trustworthiness From a Buying Transaction Life Cycle Perspective* Verlag Dr. Kovac, Hamburg, 2007.
14. Stern, B.B. "What Does Brand Mean? Historical-Analysis Method and Construct Definition," *Journal of the Academy of Marketing Science* (34:2) 2006, pp 216-223.
15. Wildemuth, B.M., and Carter, A.R. "The Perceived Affordances of Web Search Engines: A Comparative Analysis: SILS Technical Report 2002-02," University of North Carolina, School of Information and Library Science, 2002.