
The Effect of Brand Awareness on the Evaluation of Search Engine Results

Bernard J. Jansen

College of Information Sciences
and Technology
The Pennsylvania State University
University Park, PA 16802
jjansen@acm.org

Mimi Zhang

College of Information Sciences
and Technology
The Pennsylvania State University
University Park, PA 16802
mzhang@ist.psu.edu

Ying Zhang

Department of Industrial and
Mechanical Engineering
College of Engineering
The Pennsylvania State University
University Park, PA 16802
yzz114@psu.edu

Abstract

In this paper we investigate the effect of search engine brand (i.e., identifying name that distinguishes a product from its competitors) on evaluation of system performance. Our research is motivated by the large amount of search traffic directed to less than a handful of Web search engines, even though many are of equal technical quality with similar interfaces. We conducted a laboratory experiment with 32 participants measuring the effect of four search engine brands while controlling for the quality of search engine results. Based on average relevance ratings, there was a 25% difference between the most highly rated search engine and the lowest, even though search engine results were identical in both content and presentation. We discuss implications for search engine marketing and the design of empirical studies measuring search engine quality.

Keywords

User intent, Web queries, Web searching, search engines

ACM Classification Keywords

H.3.3 [1] Information Search and Retrieval – *Search process*: Measurement, Experimentation, Human Factors

Copyright is held by the author/owner(s).
CHI 2007, April 28–May 3, 2007, San Jose, California, USA.
ACM 978-1-59593-642-4/07/0004.

Introduction

There has been a rapid growth in the search engine market since its inception. Search engines continue to attract large number of Web searchers and consistently rank as some of the heavily visited sites in the market in terms of the number of visitors. There are approximately 4,000 search engines on the Web; however, only a handful dominate in terms of usage.

From a technological point, this clustering is interesting because studies report that the performance of most major search engines is practically the same [c.f., 1]. Performance is typically defined as returning relevance results. Performance is measured by precision, which is the ratio of relevant documents to the total number 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., 7].

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 motivate our research.

Review of Literature

Searching engine interfaces contain branding elements, such as symbols, logos, and names. A brand is the intangible sum of an organization's attributes, which can include an organization's name, history, reputation, and advertisement. A brand is also identifying symbol,

sign, name, or mark that distinguishes an organization or a product from its competitors. Therefore, good branding can results in loyal customers.

However, the effect of branding on technology design has not been well acknowledged, a CHI 2001 panel on branding being an exception [5]. Park, Harada, and Igarashi [6] report that the users' perceptions of a product's brand affect their perceptions of mental demand. While there may be some recognition that branding is important in the marketing of product, there has been little research in to the brand effect on the evaluation of system performance.

In this research, we measure the effect of brand perception on user perception of the performance of Web search engines.

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 designed a study that altered the brand of search engines for a set of queries while controlling for the quality and display of the results. We report the specifics of our design in the following section.

Research Design

Data Preparation

To investigate our research questions, 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

Procedures During Study

For each participant, a moderator read the participant a short introduction, explained to each participant that they 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.

representing four searching domains: medical, entertainment, travel, and ecommerce. We developed searching scenarios around each of the four queries. The four queries used are: *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, AI²RS, and got screen captures of the AI²RS results pages for each of the queries.

Using the cleaned Google results and the images from the AI²RS, 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.

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.

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.

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 results shown in Tables 1 and 2.

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.

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 [3, 4].

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 [2]. There have been other studies of search engine performance, but we wanted to control for variation in the quality of results.

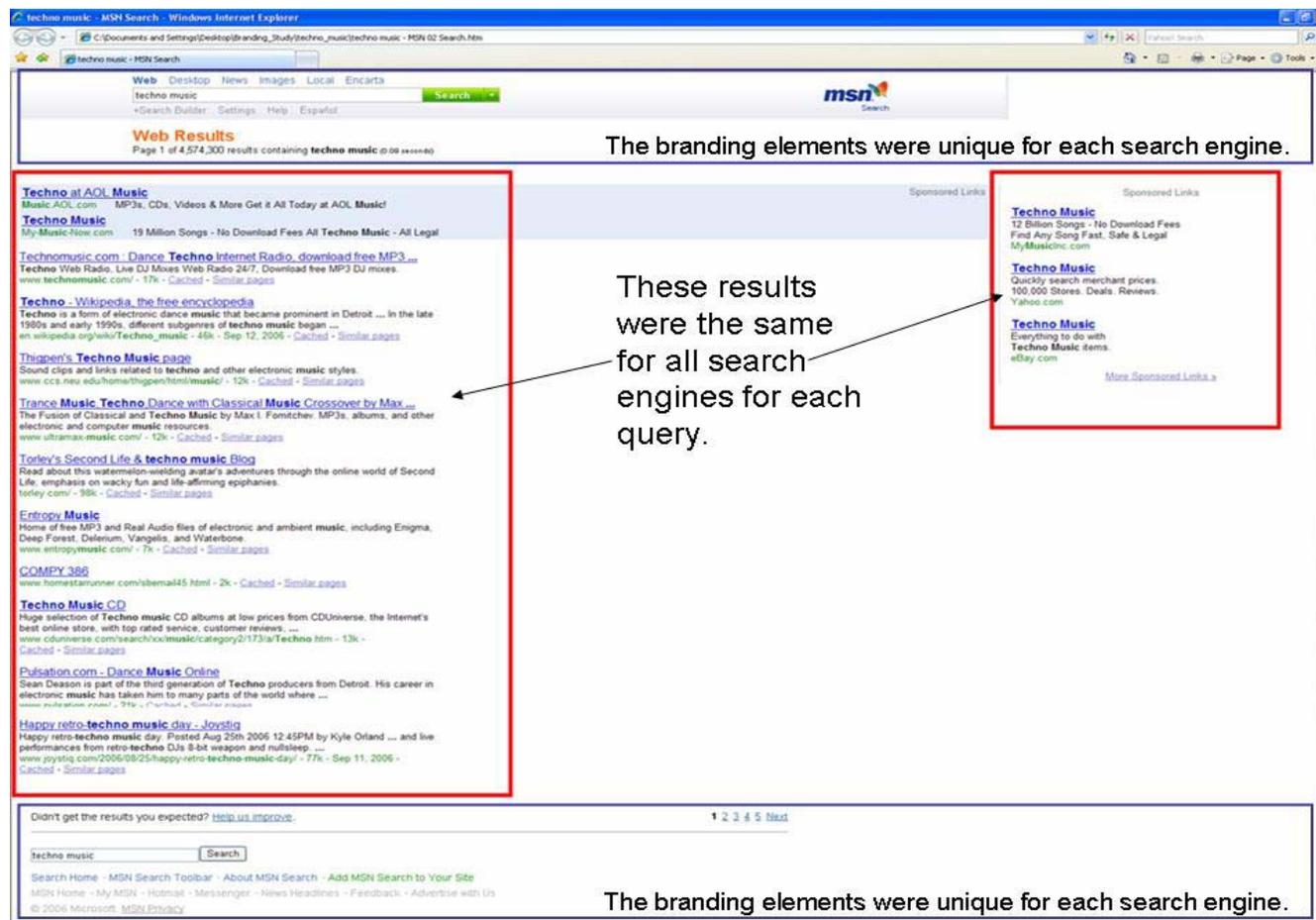


figure 1. Example of Experimental Search Engine Results Page.

Discussion

In this experiment studying the effect of branding on evaluation of system performance, regardless of which search engine a participant used for a particular domain, the results for each domain were designed to be the same.

However, there were dramatic differences in how participants rated the performance of each search engine using relevance of retrieval results. Performance evaluations varied by more than 25% between the top-most rated search engine and the bottom. Again, this difference was noted even though all the results were identical.

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.

Search Engines	Queries				Average
	camping mexico	laser removal	manufactured home	techo 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.

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% of the results were judged relevant to the query.

Search Engines	Queries				Difference from Average
	camping mexico	laser removal	manufactured home	techo 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.

In Table 2, we present the difference in average precision ratings for each search engine. **AI²RS, the unknown brand fared the worst – with an average precision rating of 10% under the average. Yahoo! had the highest rating at 15% 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 AI²RS search engine, who's average precision was 10% 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.

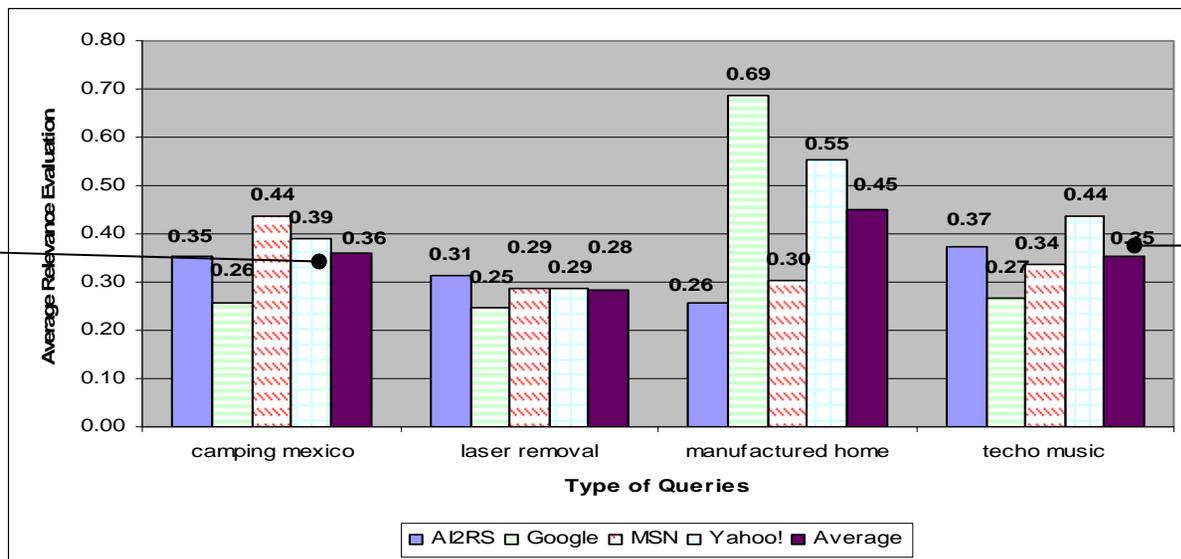
Conclusion

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.

Of all the search engines, **Yahoo!** had the overall best average precision of **0.42**. This was **15.3% better** than the average of all four search engines.

Implications

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. 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



The overall average for all search engines over all queries was approximately 0.36.

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

Acknowledgements

We thank the study participants for their time. The Air Force Office of Scientific Research (AFOSR) funded portions of this research.

References

[1] Eastman, C. M. and B. J. Jansen, "Coverage, Ranking, and Relevance: A Study of the Impact of Query Operators on Search Engine Results," *ACM Transactions on Information Systems*, vol. 21, pp. 383 - 411, 2003.
 [2] Hothkiss, G., "Eye Tracking Report: Google, MSN, and Yahoo! Compared," Enquiro, Kelowna, BC, Canada 2006.
 [3] Jansen, B. J. and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," *Information Processing & Management*, vol. 42, pp. 248-263, 2005.

[4] Jansen, B. J., A. Spink, and T. Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing & Management*, vol. 36, pp. 207-227, 2000.
 [5] Marcus, A., "Branding 101," *interactions* vol. 11, pp. 14 - 21, 2004.
 [6] Park, S., A. Harada, and H. Igarashi, "Influences of personal preference on product usability," in *Proceedings of Conference on Human Factors in Computing Systems: CHI '06 extended abstracts on Human factors in computing systems*, pp. 87 - 92.
 [7] Wildemuth, B. M. and A. R. Carter, "The Perceived Affordances of Web Search Engines: A Comparative Analysis: SILS Technical Report 2002-02," Accessed on 1 January 2005 on the World Wide Web at <http://ils.unc.edu/ils/research/reports/TR-2002-02.pdf>.