

Sponsored Search: Is Money a Motivator for Providing Relevant Results?

Bernard J. Jansen, The Pennsylvania State University

Amanda Spink, Queensland University of Technology

Analysis of data from a major metasearch engine reveals that sponsored-link click-through rates appear lower than previously reported. Combining sponsored and nonsponsored links in a single listing, while providing some benefits to users, does not appear to increase clicks on sponsored listings.

Search engines have become indispensable to interacting on the Web. In addition to processing information requests, they are navigational tools that can direct users to specific Web sites or aid in browsing. Search engines can also facilitate e-commerce transactions as well as provide access to noncommercial services such as maps, online auctions, and driving directions. People use search engines as dictionaries, spell checkers, and thesauruses; as discussion groups (Google Groups) and social networking forums (Yahoo! Answers); and even as entertainment (Googlewhacking, vanity searching).

In this competitive market, rivals continually strive to improve their information-retrieval capabilities and increase their financial returns. One innovation is *sponsored search*, an “economics meets search” model in which content providers pay search engines for user traffic going from the search engine to their Web sites.¹ Sponsored search has proven to be a successful business model for Web search engines, advertisers, and online vendors, as well as an effective way to deliver content to searchers.² The “Impact of Sponsored Search” sidebar describes some of the model’s notable benefits.

Most Web search engines display separate listings of both sponsored and nonsponsored links, also known as algorithmic or organic links, on the search-engine results

page (SERP). Search engines use proprietary algorithms that consider dozens of factors generally unknown to the user, such as PageRank score and anchor text, to retrieve nonsponsored links. However, the economic motivation behind sponsored links on the SERP is clear.

Research indicates that people have a negative bias against sponsored links.³ Assuming that searchers want relevant results in response to their queries, and that sponsored links are as relevant as nonsponsored results, are Web search engines doing users a disservice by highlighting sponsored results so prominently? Would combining sponsored and nonsponsored results in a single listing, as some metasearch engines do, benefit searchers as well as result in an increase in clicks on sponsored links?

To explore these questions, we analyzed the transaction log of more than 7 million records from a major metasearch engine that combines sponsored and nonsponsored links in a single listing on its SERP. We examined users’ click-through patterns and considered the implications of the results for sponsored search platforms and sponsored link presentation.

SPONSORED SEARCH

Bill Gross of Idealab (www.idealab.com) created the sponsored-search paradigm in 1998 with the founding of GoTo.com, which later became Overture and is now

Yahoo! Search Marketing. Google developed its own sponsored-search technology, settling a patent infringement lawsuit with GoTo.com in 2004. Other Web search engines have also developed their own platforms. Google AdWords and Yahoo! Search Marketing are currently the largest entities, accounting for the majority of sponsored Web search traffic, but there are several other players in the sponsored-search market space.^{4,5}

How it works

Sponsored search uniquely combines input from content providers, Web search engines, and users. Content providers select search phrases they believe will best link user queries to their Web sites; most sponsored search platforms give content providers the capability to tailor presentation of the sponsored link to conform to targeted queries. The Web search engine matches a searcher's query to the keywords the content providers select, displaying the corresponding sponsored link. In most cases, content providers pay the search engines whenever a user submits one of these terms and then clicks on the sponsored link; sometimes the user must go one step further and carry out some specified action on the Web site.

Multiple content providers might want to employ a search engine for the same term or phrase. In these cases, an electronic auction ranks the sponsored links, with the highest bidder generally getting the topmost rank, the second-highest bidder getting the next rank, and so on. The more providers that want to display their links in response to a term or phrase, the higher the minimum and maximum bids.

The major Web search engines employ other ranking elements besides bidding price, such as which sponsored link gets more clicks. In practice, the link with the most clicks is often the most relevant. If the link is relevant, the user is a potential good customer for the content provider. In addition, a sponsored link with the most clicks will generally produce the most profit for the search engine. Thus, both Web search engines and content providers have a monetary incentive to strive for relevant content.

Conceptually, key-phrase selection can be viewed as a dynamic form of Web site metatagging with the focus on the user. However, because they can change search phrases, their bid price, the degree of term matching, temporal restrictions, geographical limits, or even the amount they will spend in a given period, content providers become active participants in the sponsored search process.

User perceptions

A 2002 investigation by the US Federal Trade Commission recommended that search-engine companies clearly mark sponsored listings on their sites.⁶ The study reported that phrases such as "Recommended

Impact of Sponsored Search

Sponsored search has played a critical role in supporting access to the nonsponsored links that have become essential to Web users.

Without the revenue that sponsored search generates, the major search engines could not finance anywhere near their current infrastructures. These infrastructures provide the capability to crawl billions of Web pages, index several billion documents (including text, images, videos, news articles, blogs, and audio files), accept millions of Web queries per day, and present billions of links per week—not to mention the spell-checking applications, "free" e-mail services, online word and spreadsheet applications, and news feeds.

Sponsored search also provides a workable business model for metasearch engines, which are extremely helpful for searches requiring high recall and thorough topical coverage.

In addition, sponsored search effectively overcomes the inherent biases in particular Web search engines¹ by letting content providers move their links to the first search-engine results page at relatively low cost. In doing so, it serves as an essential tool vital to the success of many businesses.

It is fair to say that without sponsored search, the search-engine market—indeed the Web itself—would look far different than it does today.

Reference

1. L. Introna and H. Nissenbaum, "Defining the Web: The Politics of Search Engines," *Computer*, Jan. 2000, pp. 54-62.

sites," "Featured Listings," "Premier Listings," "Search Partners," or "Start Here" inadequately inform searchers of the nature of the links; even more ambiguous terms were "Products and Services," "News," "Resources," "Featured Listings," or "Spotlight." Although it was not clear why these terms were inadequate or on what the report's conclusion was based, the implication was that users might be less likely to consider search engines if they suspect them of intentionally disguising the presence of sponsored listings.

However, empirical studies indicate that the typical user has limited understanding of how search engines retrieve, rank, or prioritize links on the SERP or even care about these factors. Interestingly, this includes sponsored as well as nonsponsored links. Leslie Marable reported that searchers in one 2003 study did not realize that 41 percent of links on the SERP were sponsored search listings.⁷ According to a 2005 survey by the Pew Internet & American Life Project, 38 percent of respondents were unaware of the distinction between sponsored and nonsponsored links, and less than 17 percent

could always distinguish them.⁸ It is a reasonable assumption that a higher percentage of respondents would have been able to “point out the ads” in a SERP.

A 2003 laboratory study indicated that 65 percent of participants did not typically view sponsored listings, regarding them as less relevant than nonsponsored listings.³ However, when participants viewed and evaluated sponsored links in response to given queries, the ratings of these links were identical to the nonsponsored ones. A large-scale study conducted earlier this year confirmed that, despite user preconceptions, sponsored and nonsponsored results are actually equivalent in terms of relevance.⁹

RESEARCH STUDY

These prior studies point to a major potential problem with sponsored search. Online retailers are primarily interested in directing qualified customers to their Web sites, but if users, however misguided, have a bias against sponsored links, they may thus be less likely to select them. To address this issue, we wanted to find out how combining sponsored and nonsponsored links on the SERP would impact user click-through behavior.

Dogpile

We obtained a transaction log file from Dogpile (www.dogpile.com), a metasearch engine that combines both types of links from multiple search engines into a single listing. Dogpile is owned by Infospace Online (www.infospace.com), which also provides local search and online directory services. According to Nielsen//NetRatings, Dogpile was the eighth-most-popular search engine in 2006, with just over 30 million searches performed per month (www.clickz.com/showPage.html?page=3624821). That same year it earned a J.D. Power and Associates award for having the highest customer-satisfaction rating (www.submitexpress.com/news/shownews.php?article=157).

Dogpile does not crawl the Web like typical search engines. When a searcher submits a query, Dogpile concurrently submits it to multiple other Web search engines, collects the results from each, removes duplicate results, and uses a proprietary algorithm to aggregate the remaining results into a combined ranked listing. In this way, Dogpile integrates the results of the four leading Web search services—Ask.com, Google, MSN, and Yahoo!—along with approximately 18 other search engines.

As Figure 1 shows, Dogpile’s interface has separate tabs for searching the Web, images, audio, video, news, yellow pages, and white pages. It also offers query-reformulation assistance with suggestions in an “Are you looking for?” section. Note that the SERP integrates sponsored and nonsponsored links in one listing using numerous factors designed to provide the most relevant results. The specific mixture of sponsored and unspon-

sored results generally depends on the nature of the search (www.dogpile.com/info.dogpl/search/help/faq.htm). Dogpile is upfront with users concerning this integration—it labels sponsored links as such and indicates which search engine they are from.

As Figure 1 shows, the query “IEEE Sponsored Search” returns five sponsored and six nonsponsored links on the SERP above the fold—the visible portion of the screen that does not require scrolling by the user. At first thought, it may seem strange that any sponsored links appear at all for this query, but the phrase “sponsored search” has a heavy commercial bias, which probably is what generated the sponsored links. In contrast, the query “IEEE” returned only three sponsored and eight nonsponsored links in the top 11, with all three sponsored links being for courses concerning IEEE exams.

Methodology

The original transaction log contained 7,142,874 records, representing a portion of searches executed on 15 May 2006. Each record contained several fields, including

- user identification—a user code that the Web server automatically assigns to identify a particular computer;
- cookie—a small file that the Dogpile server automatically assigns to identify unique users on a particular computer;
- time of day—measured in hours, minutes, and seconds as recorded by the Dogpile server;
- query terms—the terms exactly as entered by the user;
- vertical—the content collection that the user selects to search, with Web being the default;
- sponsored—whether the user click was on a sponsored link;
- organic—whether the user click was on a nonsponsored link; and
- rank—the position in the results listing of the clicked link.

We imported the original flat ASCII transaction log file into a relational database and generated a unique identifier for each record. We removed records in which users visited the Dogpile homepage but did not execute a search as well as records with corrupted data. We used four fields—time of day, user identification, cookie, and query—to locate a particular user’s initial query and then re-created that user’s chronological series of actions during the searching session.

Our database terminology was similar to that used in other Web transaction log studies:¹⁰

- term—a series of characters separated by white space or other separator,

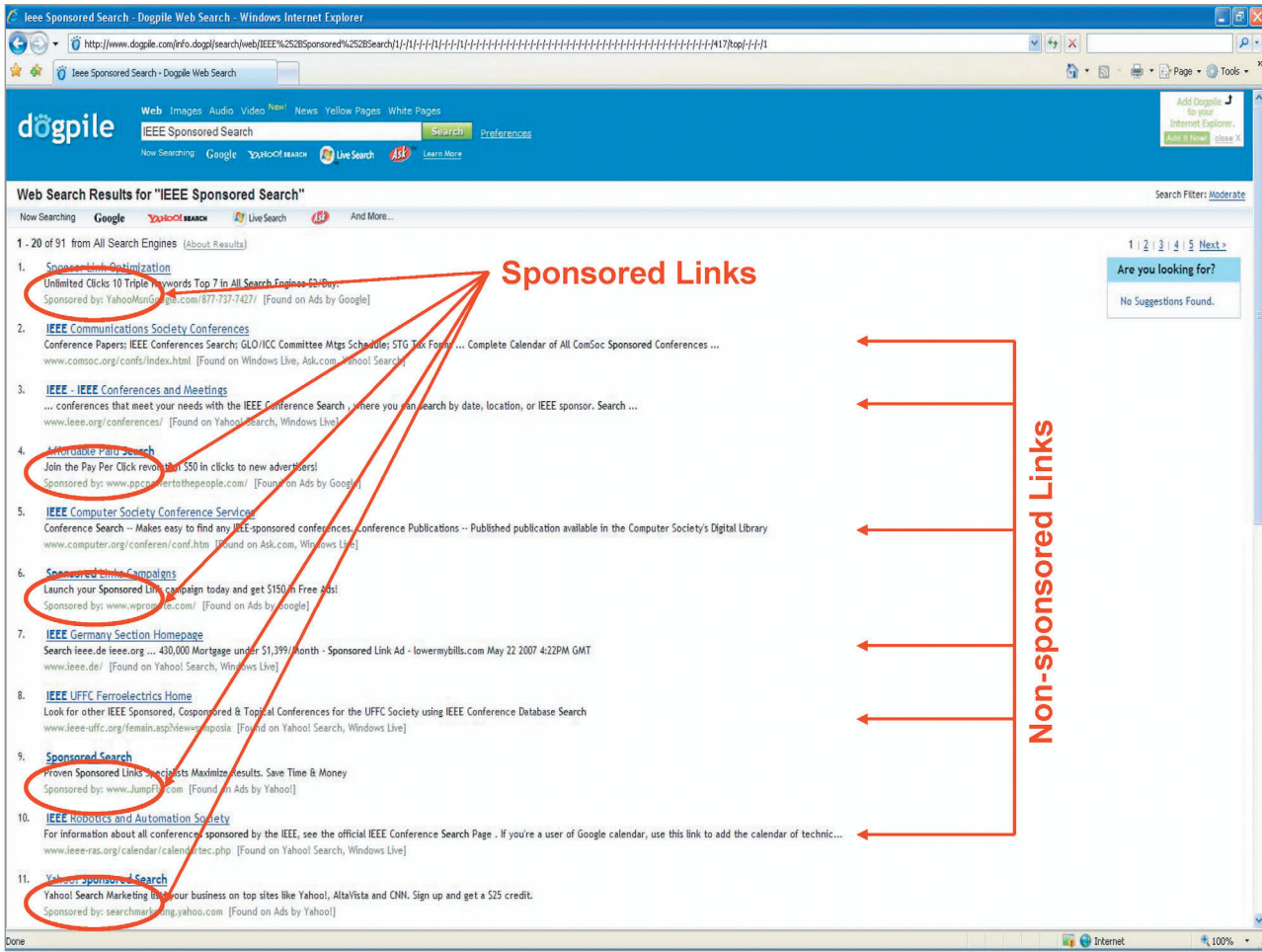


Figure 1. Dogpile interface. The search-engine-results page combines sponsored and nonsponsored links in one listing.

- query—a string of terms submitted by a searcher in a given instance,
- query length—the number of terms in the query (including traditional stop words),
- session—a series of queries submitted by a user during one interaction with the search engine, and
- session length—the number of queries submitted by a user during a defined period of interaction with the search engine.

The transaction log contained queries from both human users and agents. To filter out agent submissions, we excluded all sessions with 100 or more queries, an approach consistent with previous Web search studies.¹⁰ This cutoff is substantially greater than the mean search session¹¹ for human Web searchers, which helped ensure that we were not excluding any human sessions. It probably introduced some agent or common user terminal sessions, but we were satisfied that we had included most of the queries submitted by human searchers.

Transaction log applications usually record SERP views with an identical user identification and query but with a new time stamp for each visit. This permits the

calculation of results page views, but it also introduces duplicate records that skew the query calculations. To correct for these duplicate queries, we collapsed the transaction log upon user identification, cookie, and query. We then calculated the number of identical queries by user (number of SERPS viewed) and stored this information in a separate field within the log.

The resulting database contained 1,874,397 queries from 666,599 users (identified by unique IP address and cookie) containing 5,455,449 total terms with 4,201,071 total interactions. These interactions included submitting a query, viewing a SERP, and clicking on a URL. This data made it possible to automatically calculate which clicks from the SERP listing were on sponsored and nonsponsored links.

Study results

We first examined overall Web search behavior from the Dogpile log data. The aggregate statistics presented in Table 1 are consistent with those observed in prior studies.¹⁰ The general Web search characteristics are short queries of three terms or less, short sessions of one or two queries, and a power-law distribution of terms,

Table 1. Dogpile transaction log aggregate statistics.

Category	Number	Percent
Users	666,599	
Queries	1,874,397	
Total interactions (queries, page views, and click-throughs)	4,201,071	
Terms		
Unique	360,174	6.6
Total	5,455,449	
Mean terms per query	2.83	
Terms per query		
1 term	352,285	52.8
2 terms	114,391	17.2
3+ terms	199,923	30.0
	666,599	100.0
Users modifying queries	314,314	47.15
Repeat queries (submitted more than once by two or more searchers)	152,771	11.6
Unique queries (submitted only once in the entire data set)	1,159,764	88.4
Session size		
1 query	352,285	52.8
2 queries	114,391	17.2
3+ queries	199,923	30.0
	666,599	100.0
Boolean queries	42,138	2.2
Other query syntax	95,232	5.1
Terms not repeated in data set (208,804 terms, which is 58 percent of the unique terms)	208,804	3.8
Use of 100 most frequently occurring terms (100 terms, which is 0.0003 of the unique terms)	1,011,729	18.5
Use of other 126,208 terms (151,370 terms, which is 42 percent of the unique terms)	5,246,645	96.2
Unique term pairs (occurrences of term pairs within queries from the entire data set)	2,753,468	

Table 2. Proportion of clicks on sponsored and nonsponsored links.

Interaction type	Occurrences	Percent (including no clicks)	Percent (excluding no clicks)
Clicks on sponsored links	430,068	10.2	15.8
Clicks on nonsponsored links	2,290,804	54.5	84.2
No clicks	1,480,199	35.2	
Total	4,201,071	100.0	
Total (discounting no clicks)	2,720,872		100.0

with a small number of terms used quite often and a long tail of terms used infrequently. The use of Boolean and other advanced query operators is minimal.

Using these aggregate statistics, we calculated user click-through patterns as shown in Table 2. Column one indicates the type of interaction: clicks on sponsored links, clicks on nonsponsored links, and no clicks—times when a user submitted a query but did not click on any result. Column two lists the raw number of incidents of each type of interaction, column three shows the percentages of each type of interaction, and column four shows the percentage of clicks on sponsored and nonsponsored links if no clicks are excluded from the total.

Including all types of interactions, searchers clicked on a sponsored link approximately 10 percent of the time and on a nonsponsored link about 54 percent of the time. Approximately 16 percent of interactions in which the user actually clicked on a link were executed on sponsored links, with about 84 percent of the clicks on nonsponsored links. Users did not click on a result about 35 percent of the time. This figure may seem high, but according to Internet marketing research company comScore, Dogpile had one of the highest click-through rates of any major search engine in 2004 (www.comscore.com/press/release.asp?press=325). Typically, search engines experience non-click-through rates of approximately 45 percent.

IMPLICATIONS AND FUTURE RESEARCH

Popular press reports indicate that 25 to 30 percent of all clicks are on sponsored links (see, for example, www.internetnews.com/xSP/article.php/3502611). Previous user studies similarly reveal that about 30 percent of Web searchers will click on sponsored links over a series of searches, with approximately 35 percent of all clicks being on sponsored links.³

Although certain key phrases might achieve these and possibly higher sponsored-link click-through rates, our study indicates that for most queries the rates are substantially lower. Given that sponsored links are not prominently labeled on Dogpile, we expected that integrated listings would result in higher than average sponsored-link click-through rates. The fact that the opposite occurred suggests that the reported click-through rates of sponsored results are inflated. However, even if they are correct or even in the ballpark, integrating sponsored and nonsponsored links does not seem to raise overall sponsored-link click-through rates and in fact might decrease them.

Nevertheless, such integration could yield two substantial benefits. First, given the reported negative bias of users for sponsored links³ despite their demonstrated relevance for user queries,⁹ separating sponsored from nonsponsored links might deny users relevant results. Second, separate links can result in duplicate links on the SERP that push other relevant links below the fold and thereby reduce the chance of users clicking on them. Integrating sponsored and nonsponsored links could help overcome ranking bias and improve screen real estate management.

Certain limitations in our study might restrict the ability to generalize our conclusions. One issue is that Dogpile users might not be representative of the overall Web population. However, we found in a previous study¹⁰ that characteristics of queries across search engines are fairly consistent. In addition, overall user statistics from the Dogpile transaction log were similar to those reported elsewhere.¹¹ Therefore, we believe that the findings from this research study extend to other search engines. Nonetheless, we would like to obtain similar data from other major search engines such as Google, MSN, and Yahoo!.

Also, we do not know the exact percentage of sponsored versus nonsponsored links actually displayed in response to this set of queries at the time users submitted them. The click-through rate on sponsored links might represent the actual percentage of displayed sponsored links. It would also be interesting to see what the sponsored-link click-through rate is for e-commerce-related queries only.

The commercial aspects of Web search are here to stay. Google AdWords and Yahoo! Search Marketing provide sponsored links on SERPs, and Google AdSense and Yahoo! Content Match provide links on Web sites. With billions of dollars at stake, adversarial information-retrieval techniques are emerging, with click fraud affecting sponsored search and link farms and other techniques impacting nonsponsored search. As such, sponsored search is a rich area for research and development.

To continue to improve, Web search engines must obtain greater knowledge of user behavior. This includes understanding searchers' underlying intentions and how they relate to the sponsored search paradigm. If search engines can more accurately determine what users are looking for based on queries and other interactions, designers can leverage this knowledge to implement algorithms and interfaces that better help searchers achieve their goals. This might lead to technology that more effectively integrates sponsored and nonsponsored links, using money as one factor in providing relevant results. ■

Acknowledgments

We thank Infospace.com for providing the data for this analysis, without which we could not have conducted this research. We also thank the anonymous reviewers for a useful critique, and Naren Ramakrishnan for helpful suggestions, including the "economics meets search" phrase. We encourage other search-engine and information companies to consider processes to work with the academic research community to improve the overall Web search experience and organizational business models.

References

1. B.J. Jansen, "Paid Search," *Computer*, July 2006, pp. 88-90.
2. S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, Apr. 1998, pp. 107-117.
3. B.J. Jansen and M. Resnick, "An Examination of Searchers' Perceptions of Nonsponsored and Sponsored Links During E-commerce Web Searching," *J. Am. Soc. for Information Science and Technology*, Dec. 2006, pp. 1949-1961.
4. J. Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, Portfolio, 2005.
5. D.C. Fain and J.O. Pedersen, "Sponsored Search: A Brief History," *Bull. Am. Soc. for Information Science and Technology*, Dec. 2005/Jan. 2006, pp. 12-13.
6. E. Hansen, "FTC Wants Paid Search to Shape Up," 28 June 2002, CNET News.com; <http://news.com.com/2100-1023-940598.html>.
7. L. Marable, "False Oracles: Consumer Reaction to Learning the Truth about How Search Engines Work," research report, 30 June 2003, Consumer Reports WebWatch; www.consumerwebwatch.org/pdfs/false-oracles.pdf.
8. D. Fallows, "Search Engine Users," 23 Jan. 2005, Pew Internet & American Life Project; www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf.
9. B.J. Jansen, "The Comparative Effectiveness of Sponsored and Nonsponsored Links for Web E-Commerce Queries," article 3, *ACM Trans. Web*, May 2007.
10. A. Spink and B.J. Jansen, *Web Search: Public Searching of the Web*, Kluwer Academic, 2004.
11. B.J. Jansen, 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, no. 2, 2000, pp. 207-227.

Bernard J. Jansen is an assistant professor in the College of Information Sciences and Technology at the Pennsylvania State University. Jansen received a PhD in computer science from Texas A&M University. Contact him at jjansen@acm.org.

Amanda Spink is a professor on the Faculty of Information Technology at Queensland University of Technology, where she also co-leads the Information Research Group. Spink received a PhD in information science from Rutgers University. Contact her at ah.spink@qut.edu.au.