

# The Effect on Click-through of Combining Sponsored and Non-sponsored Search Engine Results in a Single Listing

Bernard J. Jansen

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

Amanda Spink

Faculty of Information Technology  
Queensland University of Technology  
Gardens Point Campus, GPO Box 2434  
Brisbane QLD 4001 Australia  
ah.spink@qut.edu.au

## ABSTRACT

Most major Web search engines typically present sponsored and non-sponsored results in separate listing on the search engine results page. In this research, we investigate the effect of integrating both sponsored and non-sponsored results into a single listing. The premise underlying this research is that searchers are primarily interested in relevant results to their queries. Given the reported negative bias that searchers have concerning sponsored results, separate listings may be a disservice to Web searchers by not directly them to relevant results. Some meta-search engines do combine sponsored and non-sponsored results into a single listing. Using a Web search engine log of more than 7 million interactions from hundreds of thousand of users from a major Web meta-search engine, we analyze the click through patterns of both sponsored and non-sponsored listings from various perspectives. We also classify queries as *informational*, *navigational*, and *transactional* based on the expected type of content destination desired and analyze click through patterns of each. Our findings show that about 80% of Web queries are *informational* in nature, approximately 10% each being *transactional*, and *navigational*. Combining sponsored and non-sponsored links does not appear to increase clicks on sponsored listings. In fact, it may decrease such clicks. We discuss how one could use these research results to enhance future sponsored search platforms and search engine results pages.

## Categories and Subject Descriptors

H.3.3 [1] Information Search and Retrieval – *Search process*

## General Terms

Measurement, Experimentation, Human Factors

## Keywords

User intent, Web queries, Web searching, search engines

## 1. INTRODUCTION

Commercial Web search engines usually present at least two categories of search listings on the search engine results page (SERP). One category is the non-sponsored (a.k.a., organic or algorithmic) links that the search engine determines using its proprietary matching algorithm. The other category is the sponsored links that appear because a company, organization, or

individual bid on the keyword or series of keywords that the searcher entered in the search query.

Search engines are essential tools for locating information on the Web. In addition to addressing information needs, modern Web search engines are navigation tools that take users to specific Websites or an aid in browsing. People also employ search engine as applications to conduct ecommerce transactions. Search engines provide access to content collections of images, songs, and videos rather than directly presenting a particular information object. They provide entrance to non-ecommerce transactional services such as maps, online auctions, driving directions, or even other search engines. Search engines execute social networking functions, such as with Yahoo! Answers. Web search engines also function as spell checkers, thesauruses, and dictionaries. They provide entertainment with games, such as Google Whacking or vanity searching. People are continually employing search engines in new, novel, and increasing diverse ways.

From a user point of view, search engines need to provide relevant results in response to some user input for each of these user goals. Search engines use dozens of factors in determining how to score relevance and to rank the retrieved results. Typically, the user has no idea what factors lead to a particular result being retrieval and ranked relative to another document – with one exception.

Most search engines provide separate listings of sponsored and non-sponsored results. For this one (and only one) category, searchers are informed in broad terms how the result was retrieved. Why are not other categories (e.g., anchor text, page rank) of results highlighted as well? Certainly, one understands the underlying motivation, namely to emphasize the economic incentive that caused the sponsored links to appear. Research has shown that searchers have a bias against sponsored links.

However, assuming that searchers want relevant results in response to their queries and that the sponsored links are as relevant as non-sponsored results, are search engines doing searchers a disservice by highlighting these sponsored results so prominently? What would be the effect of combining the sponsored and non-sponsored results in a single listing? These questions motivate our research.

In this paper, we present results from a research study analyzing a Web search engine transaction log of over 7 million records representing click through data of both sponsored and non-sponsored results. One the SERP of this search engine, the sponsored and non-sponsored results are combined in a single listing.

In this research, we address overall search characteristics of these Web searchers. We categorize the searching expressions based on intent relative to type of Web content implied. We examine click through patterns of sponsored and non-sponsored listing from various perspectives. We discuss the implications for sponsored search platforms and sponsored link presentation. In the next section, we present background information and related research.

## 2. RELATED STUDIES

### 2.1 Sponsored Search Studies

Major Web search engines such as Yahoo! and Google have significantly altered online commerce. The specific characteristics of the Web for ecommerce and Web-based retailing are fundamentally transforming the way in which consumers and vendors interact. Pachauri [38] discusses a review of streams of research in the ecommerce and highlights future research questions for this on-going transformation. Research focuses on both the sponsored and non-sponsored listings.

Battelle [3] provides an overview of Google Adwords and Yahoo! Search Marketing and the factors that have led to the development of these sponsored search platforms. See Jansen [18, 19] for an overview of and Fain and Pedersen [9] for a history of sponsored search. Feng [11] discusses paid placement strategies for Web search engines. Feng, Bhargava, and Pennock [12] present the search engine mechanisms for implementing sponsored search. Lui and Chen [34] analyze sponsored search as weighted unit-price-contract auctions. Feng [11] discusses the role of gatekeepers in sponsored search. The research that we report in this paper focuses on the searcher interaction area of sponsored search, specifically the use of combining sponsored and non-sponsored links in Web search engines. This a different approach than much prior research, which has focused on more clearly labeled two categories of search engine results.

An investigation supported by the Federal Trade Commission recommended that search-engine companies clearly mark paid listings on their sites [14]. The study reports that phrases such as “*Recommended Sites*,” “*Featured Listings*,” “*Premier Listings*,” “*Search Partners*,” or “*Start Here*” inadequately inform searchers of the nature of the links, although the report was not clear on why these were inadequate. Even more ambiguous terms were *Products and Services*, *News*, *Resources*, *Featured Listings*, or *Spotlight*. When users suspect that search engines are intentionally disguising the presence of sponsored listings, ecommerce searchers may be less likely to consider them.

Interestingly though, empirical studies have shown that the “typical” Web searcher has limited understanding of how search engines retrieve, rank or prioritize links on the results page [35]. This includes sponsored and non-sponsored links. Using data obtained during a user study, Marable [35] reports that searchers did not realize that 41% of links on the SERP were sponsored search listings. When informed of the nature of the sponsored listings, participants reported negative emotional reactions. Web search engines that were less transparent about sponsored search results lost credibility with this sample of users.

Examining user perceptions, Hotchkiss [15] used an enhanced focus group format to observe the search behaviors of 24 participants and interviewed them for their reactions to what they saw on the SERP. In general, the participants rated the sponsored

listings as lower quality. The researcher reports that as the search process becomes more focused, the likelihood that users will consider the sponsored listings increases. Hotchkiss [15] also reported that there were 4 distinct types of searchers, and these search patterns affected the portion of SERP seen and the likelihood of conversion (i.e., the searcher buys something). The researcher states that novice users have particular trouble identifying sponsored links and that half of the participants were suspicious that payments influence even the non-sponsored links. The study results also indicated that many searchers visually ignored or did not see the sponsored listings, partly due to their screen location on the right side of the page.

In a follow-on study, Hotchkiss, Garrison, and Jensen [16] conducted survey research with 425 respondents who overwhelmingly choose links offering sources of trusted, unbiased information. More than 77% of participants also favored non-sponsored links more than the sponsored links. Even in an ecommerce-like scenario, survey respondents still choose non-sponsored over sponsored links.

Similarly, Greenspan [13] also found that users prefer non-sponsored listings more than sponsored links. The study also raised ethical issues regarding how search engines present sponsored listings, with Greenspan [13] reporting that users are more likely to select sponsored listings with search engines that do not clearly identify them as such, suggesting that they might not have selected them had they known these links were sponsored.

Examining the effect of rank of sponsored links, Brooks [5] establishes that the likelihood of a searcher selecting a sponsored listing is a curvilinear function of its placement on the page (i.e., based on rank). The higher the link’s placement in the results listing, the more likely a searcher is to select it. The study reports similar results with non-sponsored listings. Generally, the difference between the first position and the tenth position is a 20% - 30% drop in click through (i.e., customer that actually visits a Web site by clicking on a link from a SERP) for the listing. In a related study, Brooks [6] reported that the conversion rate (i.e. customers that actually buy something) drops nearly 90% between the first and tenth position. There appears to be an intrinsic trust value associated with the rank of a listing as presented by the search engine.

Dobrow [8] reported study participants are significantly more likely to recall the name of the company from a search listing compared to a banner ad, tile ad, and three search listings on the same page. Therefore, even if study participants do not select the link, there is some marketing benefit of the sponsored listing. Investigating search engine loyalty and interaction with Web search engines, iProspect Inc. [17] surveyed 1,649 Web users. Of the respondents, 60% of Google users reported non-sponsored results to be more relevant than sponsored. This was even higher for predominantly Google users (70%). Frequent users of the Web (four or more years of Internet use) found non-sponsored listings to be more relevant than sponsored listings (65% to 56%). More women (43%) than men (34%) found sponsored listings to be generally relevant.

The Pew Internet and American Life Project [10] reported that 38 percent of searchers reported that they were aware of the distinction between sponsored links and non-sponsored links.

Less than 17 percent of survey respondents report that they can always tell which links are sponsored and which are non-sponsored.

Jansen, Brown, and Renick [22] report that 65% of the study participants did not typically view sponsored listings, viewing them as less relevant than the non-sponsored listings. However, the researchers also report that participants were unconcerned whether the listings were sponsored or non-sponsored. Their primary concern was, again, relevance.

In this research, we focus on melding sponsored and non-sponsored links on the SERP, rather than highlighting their differences. This research focus is important because prior research has established a potential disconnect between the perception of sponsored listings by business. Web searchers appear to be suspicious of sponsored links and may see these links as less relevant than non-sponsored links. Thus, they are less likely to select them. Yet, businesses see sponsored search as the future of Web marketing. Commercial and other organizations (along with some individuals) spent \$8.5 billion on sponsored search in 2004, and this amount is expected to grow to \$16 billion by 2009 [31]. Studies show that search engines are effective at returning relevant listings for Web ecommerce searching [24]. Jansen [20] shows that sponsored and non-sponsored results are equivalent in terms of relevance. However, sponsored links are primarily transactional in that businesses are trying to gauge the intent of the searcher. These businesses are only interested in getting qualified customers to their Websites that are interested in transactions, either now or sometime in the future.

## 2.2 Automatic Query Classification

Understanding the intent of Web searchers is a growing research area. Broder [4] surveyed 3,190 users of and analyzed 400 queries from AltaVista. Broder proposes the three broad Web queries classifications of navigational, informational, and transactional. Using survey results, Broder reports that approximately 73 percent of queries were informational, about 26 percent were navigational, and an estimated 36 percent were transactional, with some queries placed in multiple categories. From the log analysis, Broder reports that 48 percent of the queries were informational, 20 percent navigational and 30 percent transactional. The remaining 2 percent were not reported.

Rose and Levinson [39] manually analyzed approximately 1,500 queries from AltaVista, classifying queries as informational, navigational, and resource with hierarchical sub-categories. Rose and Levinson report that approximately 62 percent of the queries were informational, 13 percent navigational, and 24 percent resource.

While Broder [4], Rose and Levinson [39] relied on manual classification, Lee, Liu, and Cho [33] attempted automated classification as informational and navigational. The researchers used 50 queries, throwing out 20 queries (40%) that they deemed “unclassified”. Counting all 50, their success rate was 54%.

Kang and Kim [32] endeavor to classify queries as either topic (i.e., informational) or homepage finding (i.e., navigational) using selected TREC topics (50 information and 150 navigational) and portions of the WT10g test collection. The researchers report a classification rate of 91% overall, tailoring the algorithm for the particular set of queries and utilizing documents within the test

collection, which is an impractical approach when dealing with millions of queries and Web pages.

Dai and fellow researchers [7] investigated classifying Web queries in terms of commercial intent. Baeza-Yates, Benavides, and Gonzalez-Caro [2] used supervised and unsupervised learning to classify 6,042 Web queries as either informational, not informational, or ambiguous, achieving precision classification of more than 50 percent. In a related study, Nettleton, Calderon, and Baeza-Yates [37] used 65,282 queries and click stream data. The researchers then label these clusters as information, navigational, or transactional. The researchers conducted no verification of the classification.

## 2.3 Synthesis of Prior Work

From a review of existing literature, we note a small but increasing interest in sponsored search literature from a user perspective. Most of this research has focused on how users perceive sponsored results or investigations into how relevant are sponsored results. This prior work notes that searchers have a bias against sponsored results and that sponsored results are at least as relevant to user queries as non-sponsored results.

In terms of query classification, efforts at classification of Web queries have usually involved small quantities of queries manually classified with little effort in automatically classifying of queries for user intent.

However, no prior work that we could locate investigated the effects of separate listings for sponsored and non-sponsored links on user behavior. None of the efforts concerning identifying user intent focused on the aspects of click through or sponsored results. Understanding the underlying user goals and intents concerning sponsored listings is critical for the further advancement of Web systems.

This synthesis of prior work defines and motivates our research questions, which we present in the following section. Given the implications of sponsored search as the predominant business model for Web search engines, the results of this research could have substantial impact on the future development of and use of sponsored links.

In the following section, we present our research questions. We follow with a description of our data set and data analysis methods. We then present our results, along with discussion of these results. We conclude with directions for future research and implications for the design of Web searching systems.

## 3. RESEARCH OBJECTIVES

The following are our research objectives:

1. *Investigate the click through patterns of searchers when the sponsored and non-sponsored links are combined in a single listing on the SERP.*

For this research question, we analyze a transaction log file from Dogpile, a meta-search that combines results (both sponsored and non-sponsored) from multiple search engines into a single listing.

2. *Compare differences in click through patterns of informational, navigational, and transactional Web queries.*



## 4.1 Characteristics of Web Queries

For research question one, we logged the searches executed on Dogpile.com on 15 May 2006. The original general transaction log contained 7,142,874 records, representing a portion of the searches executed on Dogpile.com that date<sup>1</sup>.

Each record contained several fields, including:

- *User Identification*: a user code automatically assigned by the Web server to identify a particular computer
- *Cookie*: an anonymous cookie automatically assigned by the Dogpile.com server to identify unique users on a particular computer.
- *Time of Day*: measured in hours, minutes, and seconds as recorded by the Dogpile.com server.
- *Query Terms*: terms exactly as entered by the given user.
- *Vertical*: the content collection that the user selects to search (e.g., *Web*, *Images*, *Audio*, or *Video*) with *Web* being the default.
- *Sponsored*: whether or not the user click was on a sponsored link or not.
- *Organic*: whether or not the user click was on a non-sponsored link or not.
- *Rank*: the position in the results listing of the clicked link.

We imported the original flat ASCII transaction log file of 7,142,874 records into a relational database. We generated a unique identifier for each record. From the original transaction log, we removed records with NULL queries (i.e., no search was executed) and records with corrupted data. We used four fields (*Time of Day*, *User Identification*, *Cookie*, and *Query*) to locate the initial query and then recreated the chronological series of actions by a user.

We were only interested in queries submitted by humans and the transaction log contained queries from both human users and agents. Therefore, we removed all the agent submissions that we could identify using an upper cut-off similar to that used in prior work [36, 40]. We used an interaction cut-off by separating all sessions with 100 or fewer queries into an individual transaction log to be consistent with the approach taken in previous Web searching studies [26, 29, 41]. This cut-off is substantially greater than the mean search session [30] for human Web searchers. This increased the probability that we were not excluding any human searches. This cut-off probably introduced some agent or common user terminal sessions; however, we were satisfied that we had included most of the queries submitted primarily by human searchers.

Transaction log applications of Web search engines usually record result pages viewing as separate records with an identical

user identification and query, but with a new time stamp (i.e., the time of the second visit). This permits the calculation of results page viewings. It also introduces duplicate records that skew the queries' calculations. To correct for these duplicate queries, we collapsed the transaction log upon user identification, cookie, and query. We calculated the number of identical queries by user, storing in a separate field within the transaction log. This collapsed transaction log provided us the records by user for analyzing user queries without skewing by the result list viewing.

After processing the transaction log, the 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.

## 4.2 Automatic Classification of Web Queries

To address research question two, we derived characteristics of *information*, *navigational*, and *transactional* queries, implemented our characteristics in an algorithm (i.e., program), and executed this program on a Web transaction log.

The complete approach is presented in [21], with an abbreviated version presented here. To determine characteristics of Web queries, we selected random samples of queries from seven transactions logs of three Web search engines and manually classified them in one of three categories (*information*, *navigational*, and *transactional*). We then identified characteristics for each category that would serve to define the queries in that category. This was an iterative process with multiple rounds of "query selection – classification – characteristics refinement". These high level classifications are the same as presented by [4] and are similar to those reported by [39]. We define the intent within each category as:

- *Informational Searching*: The intent of Web information searching is to locate content concerning a particular topic in order to address an information need of the searcher. The content can be in a variety of forms, including data, text, documents, and multimedia.
- *Navigational Searching*: The intent of Web *navigational* searching is to locate a particular Web site. The Web site can be that of a person or organization. It can be a particular Web page, site or a hub Website. The searcher may have a particular Web site in mind, or the searcher may just "think" a particular Web site exists.
- *Transactional Searching*: The intent of Web *transactional* searching is to locate a Web site with the goal to execute a service in order to obtain some other end product. Examples include purchase of a product, execute a Web service, or download multimedia.

By utilizing seven transactions logs from three Web search engines, we believe that we obtained results that are generalizable across multiple search engines and user demographic populations. The defining characteristics derived are:

### *Navigational Searching*

- queries containing company/business/organization/people names
- queries containing domains suffixes

---

<sup>1</sup> We expect to make this search engine transaction log available to the research community once the current non-disclosure agreement expires and upon successful negotiation with Infospace. Other search engine logs are available at: [http://ist.psu.edu/faculty\\_pages/jjansen/](http://ist.psu.edu/faculty_pages/jjansen/).

- queries length (i.e., number of terms in query) less than 3
- queries with “web” as the source
- searcher viewing the first search engine results page

#### Transactional Searching

- queries containing terms related to movies, songs, lyrics, recipes, images, humor, and porn
- queries relating to image, audio, or video collections
- queries with “audio”, “images”, or “video” as the source
- queries with “download” terms (e.g., download, software, etc.)
- queries with “entertainment” terms (pictures, games, etc.)
- queries with “interact” terms (e.g., buy, chat, etc.)
- queries with “obtaining” terms (e.g., lyrics, recipes, etc.)
- queries with movies, songs, lyrics, images, and multimedia or compression file extensions (jpeg, zip, etc.)

#### Informational Searching

- queries containing informational terms (e.g., list, playlist, etc.)
- queries length (i.e., number of terms in a query) greater than 2
- queries that do not meet criteria for navigational or transactional
- queries that were beyond the first query submitted
- queries where the searcher viewed multiple results pages
- queries with natural language terms
- uses question words (i.e., “ways to,” “how to,” “what is,” etc.)

Some *navigational* queries were quite easy to identify, especially those queries containing portions of uniform resource locators (URLs) or even complete URLs. Although it may seem counter intuitive to some, it has been noted in prior work that many Web searchers type in portions of URLs into search boxes as a shortcut to typing the complete URL in the address box of a browser [29].

We also classified company and organizational names as navigation queries, assuming that the user intended to go to the Website of that company or organization. Naturally, there may be other reasons for a user entering a URL or proper name, which we address in the Discussion section. We also noted that most navigation queries were short in length and occurred at the beginning of the user session.

Identification of *transactional* queries was primarily via term and content analysis, with identification of key terms related to transactional domains, such as entertainment and ecommerce.

With the relatively clear characteristics of *navigational* and *transactional* queries, *informational* became the catchall by default. However, we did note characteristics that indicated *informational* searching. The most pronounced was the use of natural language phrases. *Informational* queries were also more likely to be longer and sessions of information searching was longer in terms of number of queries submitted.

For each of these classifications, we developed databases of key terms relating to characteristics of each classification. We employed these databases of key terms in our program to classify automatically the Web queries. For conditional characteristics such as query length and session length, we used program

variables. We then used the program we create to classify each query according to the characteristics developed in research question one. The algorithm for the classification is:

#### Algorithm: Web Query Classification based on User Intent

##### Assumptions:

1. Transaction log is sorted by IP address, cookie, and time (ascending order by time).
2. Search engine result page requested are removed.
3. Null queries are removed.
4. Queries are primarily English terms.

##### Input:

Record  $R_i$  with IP address ( $IP_i$ ), cookies ( $K_i$ ), query  $Q_i$ , source  $S_i$ , and query length  $QL_i$ ;

Record  $R_{i+1}$  with IP address ( $IP_{i+1}$ ), cookies ( $K_{i+1}$ ), query  $Q_{i+1}$ , source  $S_{i+1}$ , and query length  $QL_{i+1}$ .

$I$ : conditions of information query characteristics

$N$ : conditions of information query characteristics

$T$ : conditions of information query characteristics

Variable:  $B$ : Boolean // (if query matches conditions, ‘yes’ else ‘no’)

Output: Classification of User Intent,  $C$

begin

While not end of file

Move to  $R_{i+1}$

Compare ( $IP_i, K_i, Q_i, Fi, \text{ and } QL_i$ ) to  $N$

If  $B$  then  $C = N$

Elseif Compare ( $IP_i, K_i, Q_i, Fi, \text{ and } QL_i$ ) to  $T$

If  $B$  then  $C = T$

Elseif Compare ( $IP_i, K_i, Q_i, Fi, \text{ and } QL_i$ ) to  $I$

If  $B$  then  $C = I$

( $R_{i+1}$  now becomes  $R_i$ )

Store values for  $R_{i+1}$  as  $IP_i, K_i, Q_i, S_i, \text{ and } QL_i$

end loop

## 4. RESULTS

### 5.1 Research Question 01

For research question one (*Investigate the click through patterns of searchers when the sponsored and non-sponsored links are combined in a single listing on the SERP.*), we examined searcher behavior as recorded by the Dogpile log. We present an aggregate statistical analysis of the data in Table 1. Comparing these searching statistics to those reported in prior works, the results indicate the user searching characteristics are consistent with those observed on other Web search engines.

Concerning click through patterns, we present results in Table 2.

**Table 2. Proportion of Clicks on Sponsored and Non-sponsored Links**

Interaction Type	Occurrences	%	%
Sponsored	430,068	10.2%	15.8%
Organic	2,290,804	54.5%	84.2%
No Click	1,480,199	35.2%	
Total	4,201,071	100.0%	
Total (discounting No Clicks)	2,720,872		100.0%

**Table 1. Dogpile Transaction Log (15 May 2006) Aggregate Statistics**

<b>Users</b>	666,599	
<b>Queries</b>	1,874,397	
<b>Total Interactions (Queries, Page Views, and Click Throughs)</b>	4,201,071	
<b>Terms</b>		
<i>Unique</i>	360,174	6.6%
<i>Total</i>	5,455,449	
<b>Mean terms per query</b>	2.83	
<b>Terms per query</b>		
<i>1 term</i>	352,285	52.8%
<i>2 terms</i>	114,391	17.2%
<i>3+ terms</i>	199,923	30.0%
	666,599	100.0%
<b>Mean queries per user</b>		
<b>Users modifying queries</b>	314,314	47.15%
<b>Repeat Queries (queries submitted more than once by two or more searchers)</b>	152,771	11.6%
<b>Unique Queries (queries submitted only once in the entire data set)</b>	1,159,764	88.4%
	1,312,535	100.0%
<b>Session size</b>		
<i>1 query</i>	352,285	52.8%
<i>2 queries</i>	114,391	17.2%
<i>3+ queries</i>	199,923	30.0%
	666,599	100.0%
<b>Boolean Queries</b>	42,138	2.2%
<b>Other Query Syntax</b>	95,232	5.1%
<b>Terms not repeated in data set (208,804 terms, which is 58% of the unique terms)</b>	208,804	3.8%
<b>Use of 100 most frequently occurring terms (100 terms which is 0.0003 of the unique terms)</b>	1,011,729	18.5%
<b>Use of other 126,208 Terms (151,370 terms which is 42% of the unique terms)</b>	5,246,645	96.2%
<b>Unique Term Pairs (occurrences of terms pairs within queries from the entire data set)</b>	2,753,468	

We see from Table 2, that approximately 16 percent of the clicks (where the searcher actually clicked on a link) were executed on the sponsored links, with about 84 percent of the clicks being on the non-sponsored links. These percentages were discounting the no clicks (i.e., times where a query was submitted but the searcher did not click on any result). No clicks occurred approximately 35 percent of the time. Including no clicks, searchers clicked on a sponsored link approximately 10 percent of the time and on non-sponsored links about 54 percent of the time.

We could locate no published work on what overall percentage of queries result in a click on a sponsored link. However, popular press reports that 25 percent of clicks are on sponsored links (e.g., <http://www.internetnews.com/xSP/article.php/3502611>). Results from user studies also confirm that about 30 percent of searchers

will click on sponsored listings [c.f., 22, 25]. Using these percentages, it appears (perhaps counter intuitively) integrating sponsored and non-sponsored links in the same listing lowers overall sponsored link click through. However, based on calculations of click through rate versus impressions for many key terms on Yahoo! and Google's sponsored search platforms, the ration of sponsored to non-sponsored links appears lower than 25 or 30 percent that has been reported.

As far as we could determine, the research presented here is the first published accounting of actual click through rates for sponsored links. As such, there is limited benchmarking that one can do. The overall implications are that actual sponsored link click through rates are probably in the neighborhood of 15 percent, rather than the 25 to 30 percent reported elsewhere.

## 5.2 Research Question 02

For research question two (*Compare differences in click through patterns of informational, navigational, and transactional Web queries.*), we implemented our characteristics in a program. We then executed the program on the Dogpile search engine transaction log, with Table 3 presenting the results.

**Table 3. Results from Automatic Classification of Queries**

Classification	Occurrences	%
<i>Informational</i>	3,502,013	83.4%
<i>Navigational</i>	321,611	7.7%
<i>Transactional</i>	377,447	9.0%
	4,201,071	100.0%

Table 3 shows that nearly 84 percent of Web queries were classified as informational in intent, with *transactional* queries representing about 9 percent of Web queries, and *navigational* queries representing approximately 8 percent of the queries in the log. Our analysis found a surprising high percentage of *informational* queries.

Results presented in prior work indicated that *navigational* intent was a significantly portion of Web searching [4, 29]. As an example, Broder [4] reports *navigational* queries of 24 percent based on approximately 3,100 survey responses and 20 percent based on an analysis of 400 Web queries. The low occurrence of *transactional* queries in our results is also somewhat surprising. Broder [4] reports *transactional* queries of 36% based on survey responses and 30% based on the analysis of Web query. Jansen and Spink [27] report that e-commerce-related queries ranged from

12 percent to 24 percent based on analysis of approximately 2,500 queries from multiple transaction logs.

The variation in reported percentage of *navigational* and *transactional* queries may be related to the size of the samples used in prior studies and the power log distribution of Web queries [1]. For example, Jansen, Spink, and Pederson [29] report on the most frequently occurring queries, so *navigational* queries may be more prevalent in the more frequently occurring queries than the entire distribution, especially those in the long tail. A similar effect may be happening with *transactional* queries. Moreover, Broder [4] used a random but small sample of queries. Rose and Levinson [39] classified only the initial query in the user's session, which may have skewed results. Additionally, there may be e-commerce queries that are not *transactional* in nature, but may represent effort to contain information about product. This could account for the differences with e-commerce topics reported in [29]. Finally, [4, 29, 39] were all executed on data from AltaVista, which may be non-representative. In fact, Jansen and Spink [27] report statistic differences with AltaVista transaction log results relative to results from other Web search engines.

We see in Table 4, *navigational* queries had the highest occurrence of click through on sponsored links (more than 19 percent), and *transactional* queries had the lowest sponsored link click through (slight more than 14 percent). It appears that sponsored links can provide relevant results for searchers to navigate to Websites. Also, the sponsored link click through rates for information queries was also quite high, indicating that these links can possibly provide relevant results to Web searchers.

**Table 4. Click Through of Informational, Navigational, and Transactional Queries**

	Count of Sponsored	%	Count of Organic	%	
<i>Informational</i>	357,127	15.70%	1,917,402	84.30%	2,274,529
<i>Navigational</i>	37,333	<b>19.31%</b>	156,030	80.69%	193,363
<i>Transactional</i>	35,608	14.08%	217,372	<b>85.92%</b>	252,980
	430,068	15.81%	2,290,804	84.19%	2,720,872

**Table 5. Click Through of Informational, Navigational, and Transactional Queries by Rank**

Rank	<i>Informational</i>	%	<i>Navigational</i>	%	<i>Transactional</i>	%
1	381,286	10.9%	95,318	<b>29.6%</b>	38,436	10.2%
2	264,836	7.6%	26,124	8.1%	27,597	7.3%
3	191,379	5.5%	16,148	5.0%	21,044	5.6%
4	159,283	4.5%	10,893	3.4%	16,708	4.4%
5	126,553	3.6%	7,428	2.3%	13,766	3.6%
6	111,636	3.2%	5,832	1.8%	12,054	3.2%
7	88,974	2.5%	4,307	1.3%	10,359	2.7%
8	83,196	2.4%	3,658	1.1%	9,410	2.5%
9	66,912	1.9%	2,951	0.9%	7,876	2.1%
10	66,614	1.9%	2,485	0.8%	7,311	1.9%
>10	733,860	<b>21.0%</b>	18,219	5.7%	88,419	<b>23.4%</b>
No Click	1,227,484	35.1%	128,248	39.9%	124,467	33.0%
	3,502,013	100.0%	321,611	100.0%	377,447	100.0%

\* highest percentage rank is bolded

In Table 5, we illustrate the click through of each category by rank of the listing as shown on the Dogpile SERP. The default listing on Dogpile is 20 results, so we can assume that most SERP contained this number of links. From Table 5, we see that nearly 30 percent of clicks for *navigational* queries were on the first result. This noticeable different user behavior, relative to *informational* and *transactional* queries, (where a substantial percentage of the clicks occurred on links 11 and higher) may assist in identifying user intent from implicit feedback measures. It also points to the issue that Web search engines are effective at ranking results for navigational queries but may be less effective at ranking for informational and transactional. This is understandable as the user intent behind these classes of queries is much more complex and dimensional.

## 5. DISCUSSION

In this research, we investigated the effect of integrating sponsored and non-sponsored links within the same SERP listing. In order to investigate this research question, we analyzed a transaction log from a Web meta-search engine that displayed integrated search results. We examined overall user statistics from the transaction log, which were similar to user search statistics reported elsewhere [30, 41]. Therefore, we believe that the findings from this research are generalizable to other search engines.

We also employed a three-category classification – *informational*, *navigational*, and *transactional* – of Web searching that is useful in identifying the intent of the searcher. This classification is based on our own analysis and on prior published work, most notably that of [4, 39]. Additionally, this research uses a much larger data set of queries than prior work [2, 32, 33], does not depend on external content, and can be implemented in real time. This makes it a viable solution for Web search engines to identify user intent of searchers, there by improving system performance.

From our results, we noted that slightly less than 16 percent of clicks are on sponsored links. This appears to substantially lower than reported in popular press comments. Why this is so cannot be answered by transaction log analysis, pointing to the need for future research. However, it may also indicate that popular press comments concerning click through rates of sponsored results are not correct. One would expect that integrated listings of sponsored and non-sponsored listing would results in higher sponsored link click through rates. The fact that they are not may indicate that the common held benchmark is not correct.

In classification of queries, a high portion was *informational*, indicating that searchers continue to view Web search engines primarily as traditional information retrieval systems. *Transactional* queries were approximately 9 percent, with *navigational* queries at slightly less than 8 percent.

*Navigational* queries had the highest sponsored link click through rate of more than 19 percent (*informational* at nearly 16 percent and *transactional* at 14 percent). Although one might expect *transactional* queries to have the highest percentage of sponsored link clicks, major commercial entities buy sponsored links for branding purposes, which many times may be relevant results for *navigational* queries (e.g., walmart, dell, ibm, target).

We examined click through rates at various result ranks by query classifications. Based on this analysis, users submitting

navigational queries clicked on the number one result nearly 30 percent of the time, compared to approximately 10 percent for *informational* and *transactional* queries. This points to the need for further analysis to identify other implicit feedback [23] metrics to assist in identifying user intent. *Informational* and *transactional* queries had higher click through rates at ranks higher than 10 (21 and 23 percent respective) relative to *navigational* queries (approximately 6 percent).

In analyzing our results, we are aware of certain limitations that may restrict the ability to generalize our conclusions. One issue is that the Dogpile user population may not be representative of Web search engine users in general. Therefore, their queries would not be representative of the general Web population. However, Jansen and Spink [27] report that characteristics among queries across search engines is fairly consistent. In addition, we developed our defining characteristics for each of the three categories from seven other transaction logs from three other search engines. Therefore, we would expect results similar to these from other datasets. We would certainly like to apply our classification methods on current data from other major search engines if one could obtain this data (i.e., Google, MSN, Yahoo!).

## 6. CONCLUSION

Sponsored search is a critically important business model that finances the ‘free’ search that millions of users of Web search engines have come to depend upon. Web search engine companies are continually exploring new and novel methods of both providing relevant results and increasing financial returns.

In this research, we explore the effects of integrating sponsored and non-sponsored links into one listing on the SERP. Studies have shown that Web searchers do not understand how search engines rank results. Web search engines do not generally disclose how non-sponsored results are listing and ranked. Certainly, exploring this line of research is an area worth pursuing.

Additionally, in order for Web search engines to continue to improve, they must leverage an increased knowledge of user behavior, especially efforts to understand the underlying intent of searchers and how this intent relates to sponsored links. If Web search engines can determine search goals based on queries and other interactions, designers can leverage this knowledge by implementing algorithms and interfaces to better help searchers achieve their goals.

## 7. ACKNOWLEDGEMENTS

The Air Force Office of Scientific Research (AFOSR) and the National Science Foundation (NSF) funded portions of this research. We would like to thank especially Infospace.com for providing the data for this analysis, without which we could not have conducted this research.

## 8. REFERENCES

- [1] C. Anderson, *The Long Tail: Why the Future of Business is Selling More of Less*. New York: Hyperion, 2006.
- [2] R. Baeza-Yates, L. Calder´on-Benavides, and C. Gonz´alez-Caro, "The Intention Behind Web Queries," in *String Processing and Information Retrieval (SPIRE 2006)*, Glasgow, Scotland, 2006, pp. 98-109.

- [3] J. Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed our Culture*. New York: Penguin Group, 2005.
- [4] A. Broder, "A Taxonomy of Web Search," *SIGIR Forum*, vol. 36, pp. 3-10, 2002.
- [5] N. Brooks, "The Atlas Rank Report I: How Search Engine Rank Impacts Traffic." vol. 2004: Atlas OnePoint, 2004.
- [6] N. Brooks, "The Atlas Rank Report II: How Search Engine Rank Impacts Conversions." vol. 2005: Atlas Institute, 2004.
- [7] H. K. Dai, Z. Nie, L. Wang, L. Zhao, J.-R. Wen, and Y. Li, "Detecting Online Commercial Intention (OCI)," in *World Wide Web Conference (WWW2006)*, Edinburgh, Scotland., 2006, pp. 829-837.
- [8] L. Dobrow, "Internet 2.0 The next level of the ultimate customer-based tool," in *Ito! Magazine*. vol. 2005: Ito!, 2004, pp. 19-23.
- [9] D. C. Fain and J. O. Pedersen, "Sponsored Search: A Brief History," *Bulletin of the American Society for Information Science and Technology*, vol. 32, pp. 12-13, 2006.
- [10] D. Fallows, "Search Engine Users." vol. 2005: Pew Internet & American Life Project, 2005.
- [11] J. Feng, "Paid Placement Strategy for Internet Search Engines," in *the 11th International World Wide Web Conference (WWW 02)*, Honolulu, Hawaii, 2002.
- [12] J. Feng, H. K. Bhargava, and D. M. Pennock, "Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms," *INFORMS Journal of Computing*, Forthcoming.
- [13] R. Greenspan, "Searching for balance." vol. 2004: ClickZ stats, 2004.
- [14] E. Hines, "FTC wants paid search to shape up." vol. 2005: CNET news.com, 2002.
- [15] G. Hotchkiss, "Inside the Mind of the Searcher." vol. 2005: Enquiro Search Solutions Inc, 2004.
- [16] G. Hotchkiss, M. Garrison, and S. Jensen, "Search Engine Usage in North America." vol. 2005: Enquiro, 2004.
- [17] iProspect Inc., "Search Engine User Attitudes." vol. 2005: iProspect, 2004.
- [18] B. J. Jansen, "Paid Search," *IEEE Computer*, vol. 39, pp. 88-90, 2006.
- [19] B. J. Jansen, "Paid Search as an Information Seeking Paradigm," *Bulletin of the American Society for Information Science and Technology*, vol. 32, pp. 7-8, 2006.
- [20] B. J. Jansen, "The Comparative Effectiveness of Sponsored and Non-sponsored Results for Web Ecommerce Queries," *ACM Transactions on the Web*, forthcoming.
- [21] B. J. Jansen, D. L. Booth, and A. Spink, "Modeling User Intent in Web Searching," forthcoming.
- [22] B. J. Jansen, A. Brown, and M. Resnick, "Factors Relating to the Decision to Click-on a Sponsored Link," *Decision Support Systems*, forthcoming.
- [23] B. J. Jansen and M. D. McNeese, "Evaluating the Effectiveness of and Patterns of Interactions with Automated Searching Assistance," *Journal of the American Society for Information Science and Technology*, vol. 56, pp. 1480-1503, 2005.
- [24] B. J. Jansen and P. Molina, "The effectiveness of Web search engines for retrieving relevant ecommerce links," *Information Processing & Management*, vol. 42, pp. 1075-1098, 2006.
- [25] B. J. Jansen and M. Resnick, "An examination of searchers' perceptions of non-sponsored and sponsored links during ecommerce Web searching," *Journal of the American Society for Information Science and Technology*, vol. 57, pp. 1949-1961, 2006.
- [26] B. J. Jansen and A. Spink, "An Analysis of Web Searching By European Alltheweb.com Users," *Information Processing & Management*, vol. 41, pp. 361-381, 2005.
- [27] B. J. Jansen 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.
- [28] B. J. Jansen, A. Spink, C. Blakely, and S. Koshman, "Web searcher interactions with the Dogpile.com meta-search engine," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1875-1887, 2006.
- [29] B. J. Jansen, A. Spink, and J. Pedersen, "Trend Analysis of AltaVista Web Searching," *Journal of the American Society for Information Science and Technology*, vol. 56, pp. 559-570, 2005.
- [30] 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, pp. 207-227, 2000.
- [31] G. Jarboe, "Online Ads & Search: Looking Back, Looking Forward." vol. 2005: Searchengine Watch, 2005.
- [32] I. Kang and G. Kim, "Query type classification for Web document retrieval," in *26th annual international ACM SIGIR conference on Research and development in information retrieval*, Toronto, Canada, 2003, pp. 64 - 71.
- [33] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," in *The World Wide Web Conference*, Chiba, Japan, 2005, pp. 391-401.
- [34] D. Liu and J. Chen, "Designing Online Auctions with Past Performance Information," *Decision Support Systems*, vol. 42, pp. 1307-1320, 2006.
- [35] L. Marable, "False Oracles: Consumer Reaction to Learning the Truth About How Search Engines Work, Results of an Ethnographic Study," *Consumer WebWatch*, 2003, pp. 1-66.
- [36] A. Montgomery and C. Faloutsos, "Identifying web browsing trends and patterns," *IEEE Computer*, vol. 34, pp. 94-95, July 2001.
- [37] D. F. Nettleton, L. Calderon, and R. Baeza-Yates, "Analysis of Web Search Engine Query and Click Data from two Perspectives: Query Session and Document," in *the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006)* Philadelphia, Pennsylvania, 2006.
- [38] M. Pachauri, "Researching Online Consumer Behaviour: Current Positions and Future Perspectives," *Journal of Customer Behaviour*, vol. 1, pp. 269-300, 1 June 2002.
- [39] D. E. Rose and D. Levinson, "Understanding User Goals in Web Search," in *the World Wide Web Conference (WWW 2004)*, New York, NY, USA, 2004, pp. 13-19.
- [40] C. Silverstein, M. Henzinger, H. Marais, and M. Moricz, "Analysis of a Very Large Web Search Engine Query Log," *SIGIR Forum*, vol. 33, pp. 6-12, 1999.
- [41] A. Spink and B. J. Jansen, *Web Search: Public Searching of the Web*. New York: Kluwer, 2004.}