Competing for Users' Attention: On the Interplay between Organic and Sponsored Search Results

Cristian Danescu-Niculescu-Mizil[‡1], Andrei Z. Broder[†],
Evgeniy Gabrilovich[†], Vanja Josifovski[†], Bo Pang[†]

‡ Cornell University, Ithaca, NY 14853, USA
† Yahoo! Research, 4301 Great America Parkway, Santa Clara, CA 95054, USA
cristian@cs.cornell.edu | {broder | gabr | vanjaj | bopang}@yahoo-inc.com

ABSTRACT

Queries on major Web search engines produce complex result pages, primarily composed of two types of information: *organic results*, that is, short descriptions and links to relevant Web pages, and *sponsored search results*, the small textual advertisements often displayed above or to the right of the organic results. Strategies for optimizing each type of result in isolation and the consequent user reaction have been extensively studied; however, the interplay between these two complementary sources of information has been ignored, a situation we aim to change. Our findings indicate that their perceived *relative* usefulness (as evidenced by user clicks) depends on the nature of the query. Specifically, we found that for navigational queries there is a clear competition between ads and organic results, while for non-navigational queries this competition turns into synergy.

We also investigate the relationship between the perceived usefulness of the ads and their textual similarity to the organic results, and propose a model that formalizes this relationship. To this end, we introduce the notion of *responsive ads*, which directly address the user's information need, and *incidental ads*, which are only tangentially related to that need. Our findings support the hypothesis that in the case of navigational queries, which are usually fully satisfied by the top organic result, incidental ads are perceived as more valuable than responsive ads, which are likely to be duplicative. On the other hand, in the case of non-navigational queries, incidental ads are perceived as less beneficial, possibly because they diverge too far from the actual user need.

We hope that our findings and further research in this area will allow search engines to tune ad selection for an increased synergy between organic and sponsored results, leading to both higher user satisfaction and better monetization.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms: Measurement, Experimentation, Theory

Keywords: Usefulness of ads, sponsored search, ad selection, textual similarity, responsive ads, incidental ads, navigational queries.

[1]The research described herein was conducted while the first author was a summer intern at Yahoo! Research.

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

1. INTRODUCTION

"...in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."
– Herbert Simon, "Designing Organizations for an Information-Rich World", 1971.

The development of Web search has led to a paradigm shift in information access, as huge amounts of information became accessible to anyone with a basic Internet connection and minimal search skills; indeed, users worldwide send hundreds of millions of queries to Web search engines every day. At the same time, the invention of *search advertising* (or *sponsored search*) turned Web search into an easily monetizable activity, by allowing advertisers to target interested audiences. As a result, the *search engine result pages* (SERPs) today are primarily composed of two types of information: *organic results*, that is, short descriptions ("snippets") and links to relevant Web pages, and *sponsored search results*, the small textual advertisements displayed alongside the organic results. The two types are clearly separated on the SERP, with the ads being marked as "sponsored links" or "sponsored results". Most ads are displayed on the right hand side on the page (*East ads*), but some ads might be shown above the organic results (*North ads*).

There are some fundamental differences between the organic results and the ads. First, for every query, organic results aim to satisfy the most likely user intent(s) as inferred from the query, while ads tend to focus on potential commercial intent(s). Second, while organic results are ranked solely based on relevance, ad ranking jointly optimizes relevance and revenue. Lastly, the "snippets" associated with organic results are usually computer-generated by a summarization algorithm, while most ads are manually crafted, as is the practice in the classic advertising industry. (In fact, following the same practice, the body of sponsored search ads is called a *creative*, supposedly produced by an ad agency.)

Nevertheless, both ads and organic results provide information to the user. While several papers considered the interaction of the user with the ads [2, 14, 13, 7, 5, 8] and the Web results [16, 17, 1] in isolation, to the best of our knowledge, no study considered the interplay between these two sources of information. If we take the click rate as evidence of user perceived usefulness, we can conceivably observe three situations:

1. The clickthrough rates on these two types of results are negatively correlated, possibly indicating that ads and organic results compete for user attention and/or satisfy different intents.

2. The rates are uncorrelated, possibly indicating that users treat ads and organic results as independent sources of information.

3. The rates are positively correlated, possibly indicating mutually reinforcing effects or quality similarity.

Clearly the situation depends on the nature of the user's query and the user's intent, as well as on the characteristics of both types of results. Indeed, we found out that for navigational queries (in the sense of [3]) there is a clear competition between ads and organic results, while for non-navigational queries, the competition turns into synergy, as the two sources together appear to satisfy the user's information need better than either source alone.

To further elucidate these observations, we investigate the relationship between the perceived usefulness of the ads and their textual similarity to the organic results, and propose a model that formalizes this relationship. To this end, we introduce the notion of *responsive ads*, which directly address the user's information need, and *incidental ads*, which are only tangentially related to that need. Our findings support the hypothesis that in the case of navigational queries, which are usually fully satisfied by the top organic result, incidental ads are perceived as more valuable than responsive ones, which are likely to be duplicative. On the other hand, in the case of informational queries, incidental ads are perceived as less beneficial, possibly because they diverge too far from the actual user need.

The main contributions of this paper are as following. We study the interplay between the two main parts of the search engine results page, namely, the organic results and the ads. We juxtapose the clickthrough rates on these two sources of information to identify when their synergy turns into an explicit competition for users' attention. We also introduce the notions of responsive and incidental ads, which help us understand when a high degree of topical similarity is desired between the organic results and the ads, and in which cases such similarity turns out to be detrimental. We believe our findings will lead to richer Web search experience, where the increased synergy between the organic and the sponsored results yields higher overall users' satisfaction.

2. BACKGROUND: TEXTUAL ADVERTISING ON THE WEB

A large part of the $21 billion Web advertising market consists of *textual ads*, the ubiquitous short text messages usually marked as "sponsored links". There are two main channels for distributing such ads.

a. *Sponsored search* places ads on the result pages of a Web search engine, where ads are selected to be relevant to the search query (see [10] for a brief history of the subject). All major Web search engines (Google, Microsoft, Yahoo!) support sponsored ads and act simultaneously as a Web search engine and an ad search engine.

b. *Content match* (or *contextual advertising*) places ads on third-party Web pages.

In this paper we focus on sponsored search, which is an interplay of the following three entities.

1. The **advertiser** provides the supply of ads. As in traditional advertising, the goal of the advertisers can be broadly defined as promotion of products or services.

2. The **search engine** provides "real estate" for placing ads (i.e., allocates space on search results pages), and selects ads that are relevant to the user's query.

3. **Users** visit the Web pages and interact with the ads.

The prevalent pricing model for textual ads is that the advertisers pay a certain amount for every click on the advertisement (pay-per-click or PPC).

The amount paid by the advertiser for each sponsored search click is usually determined by an auction process [9]. The advertisers place *bids* on various search phrases, with the intent that queries matching these *bid phrases* will trigger the corresponding ads; when a query triggers multiple ads, their relative position is determined by their bids, as explained below.

Thus, each ad is associated with one or more bid phrases. In addition to bids and bid phrases, each ad must specify a *title* usually displayed in bold font, a *creative*, which is the few lines of text forming the body of the ad, and a *target URL*, which is the result of clicking on the ad.

In the model currently used by all the major search engines, bid phrases serve a dual purpose: they explicitly specify queries that should trigger the ad and simultaneously put a price tag on a click event. Obviously, these price tags could be different for different queries. For example, a flooring contractor advertising his services on the Internet might be willing to pay a small amount of money when his ads are clicked from general queries such as "home remodeling", but higher amounts if the ads are clicked from more focused queries such as "hardwood floors" or "laminate flooring".

Most often, ads are shown for queries that are quasi-identical to the bid phrases for the ad, thus resulting in an *exact match* between the query and the bid phrase. However, it might be difficult (or even impossible) for an advertiser to list all the relevant queries ahead of time. Hence, search engines, provide *broad* (or *advanced*) match, whereby advertisers allow the search engines to decide what queries are suitable for triggering their ads based on the bid phrases already provided.

Given a query q, the revenue from a click can be estimated as

$$R = \sum_{i=1..k} P(click|q, a_i) \cdot price(a_i, i),$$

where k is the number of ads displayed on the page with search results for q and $price(a_i, i)$ is the click price of the ad a_i at position i. The price depends on the set of ads competing for display on the result page and their respective bids. Several models have been proposed to determine this price, most of them based on generalizations and variants of second price (GSP) auctions. For more details, see [9] and references therein.

Users' attitude towards search advertising has evolved since the inception of Web search some 15 years ago. While today it is an accepted fact that search engines are supported by advertising on the SERP, the first attempt by AltaVista to combine organic and sponsored search results was met with a huge public outcry [19]. Even now, although ads are ubiquitous, all major search engines limit the number of ads and their position to mitigate any negative impact on users' experience.

3. DATA

We formed a *virtual* search result page SERP_v (v indicates "virtual") for each query q as described below. We collected all organic and sponsored search results displayed within one month time frame by a certain set of Yahoo! Search servers for queries issued at least 100 times. In order to mitigate presentation effects, we consider only the North ads, that is, the ads displayed above organic results in a manner relatively similar to the organic ads themselves, within a block marked "Sponsored Results". Furthermore, we discarded all queries for which no North ad was displayed.

All these restrictions aim to reduce the probable causes of observed users' reactions to differences in displayed content and/or users' inherent preference for organic vs. ad results. (A side-effect of this selection is that queries included in our dataset might be of more commercial interest than a random sample since only such queries are likely to trigger North ads.) While North ads may have an advantage over organic results given that they are displayed earlier in the page, note that the average CTR (*click through rate*, defined as the ratio between the number of times a result was clicked and the number of times it was shown) for organic results is still higher than the average CTR for ads. We discuss the *non-commercial bias*, that is, user's preference for organic results, in more detail in Section 4.1.

Among the results collected as described, we kept only those that had been shown for at least N_v times. For our investigation, we need to estimate the CTR for each result, be it organic or sponsored. For reliability, we set N_v to 50 for most analyses conducted in the paper; the only exception is the prediction experiments described in Section 5.4, where we set $N_v = 10$ to keep more datapoints. In what follows, we use SERP_v to denote the virtual SERP that consists of all organic and North ad results that had been displayed for a given query for at least N_v times.

Next, we picked a pair of sponsored and organic results from each SERP_v that were deemed suitable for *meaningful comparisons*. The first possibility we considered was to pick the topmost results in their respective blocks, since they are the most comparable ones in terms of page position. However, it turned out that the notion of "topmost" is not well-defined in a SERP_v: given the vagaries of the ad marketplace, the order of the ads is highly variable; in fact, even the order of the organic results is not always consistent. Thus, we constructed the pair by employing the users' implicit feedback as an indication of which results were the most comparable: for each query q_i, we took the most clicked ad (A_i^*) and the most clicked organic search result (O_i^*) from its SERP_v. We denote the CTR of A_i^* and O_i^* as $ctrA_i$ and $ctrO_i$, respectively. (We omit the subscript i when referring to the results for a random query q.)

The procedure described above yielded a collection of 63,789 (q_i, O_i^*, A_i^*) tuples. All our subsequent analyses were conducted on this dataset, unless specified otherwise. One potential concern was that our observations were specific to our choice of (O_i^*, A_i^*) pairs. To ease such concerns, we also conducted our analyses over all the ads present in SERP_v — a total of 332,607 (q_i, O_i^*, A_i^k) tuples[1] where A_i^k ranged over all North ads displayed for q_i — and we observed the same qualitative results. Thus, while we do not refer back to this extended dataset in the rest of the paper, it is important to keep in mind that all the following discussions are also valid for this alternate setting.

Note that in order to focus on meaningful observations, all plots in this paper include only the points that represent at least 2% of the considered tuples. In addition, in order to avoid disclosing commercially sensitive data, we transformed raw CTR values via a linear transformation into *relative CTR* values.

4. RELATIONSHIP BETWEEN AD AND ORGANIC CTR

Consider a user who is taken to a search result page after issuing a query. Both organic and sponsored search results are presented on the page, vying for the user's attention. Is the user going to click on the organic results or on the ads? Clearly the answer varies from query to query, depending on the nature of the user's information need, as well as the quality of both types of results. Nonetheless, it is interesting to investigate whether there is any interesting signal at the aggregated level.

We start by examining the basic relationship between organic CTR and ad CTR. In particular, we study the relationship between $ctrO$ and $ctrA$, i.e., the highest organic and sponsored search CTR for each query.

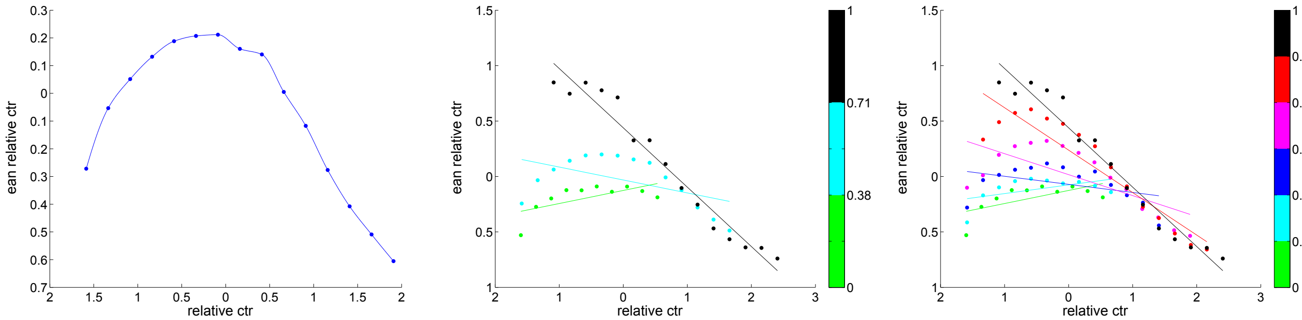
4.1 Non-commercial bias

Position bias has been shown to play an important part in a user's perception of different search results in a SERP [15]. Similarly there can be a bias based on whether the result is displayed in the organic region and or the sponsored region. Indeed, previous work has suggested that users have a bias against sponsored search results [12]. This was confirmed by our findings.

In particular, we examined the set of queries for which the target URLs of A^* and O^* shared the same domain. There are a significant number of such queries (in our case 13.5% of all the queries in our dataset), since some advertisers have an interest in holding prominent positions in both ads and organic results. In such cases, we expected A^* and O^* to be of almost identical quality, and most likely only one of them would be clicked on by a given user. This set of queries was as close as we could get to an equal-quality setting to analyze users' base preferences. We refer to this set as Q_{eq} hereafter.

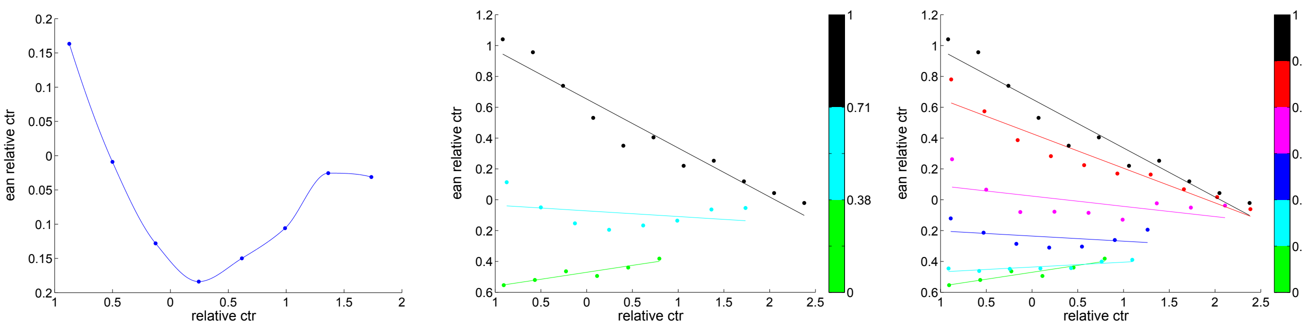
Given that all the ads in our dataset were North ads, if position bias was the dominating factor, we should observe that most users clicked on A^*. As a result, the average $ctrA$ should be much higher than the average $ctrO$ in Q_{eq}. On the other hand, users may prefer the organic results in spite

[1] Recall that SERP_v contained all the North ads that were shown for this query *during one month* — while there were (on average) roughly 5 different North ads in each virtual SERP_v, this did not mean that there were 5 North ads per *physical* SERP.



(a) The overall trend (b) Breakdown for queries classified as navigational (black) and non-navigational (green) with high confidence (c) Breakdown by navigational score

Figure 1: Average ad clickthrough for different values of organic clickthrough. Best linear fits are shown for (b) and (c) in order to emphasize the respective trends.



(a) The overall trend (b) Breakdown for queries classified as navigational (black) and non-navigational (green) with high confidence (c) Breakdown by navigational score

Figure 2: Average organic clickthrough for different values of ad clickthrough. Best linear fits are shown for (b) and (c) in order to emphasize the respective trends.

of the position bias. There can be two main reasons:

1. Some users might not trust ads as much as they trust organic results, knowing that ads were paid to be shown. In the extreme case, some users might be decisively "ad-blind" and simply ignore the entire block of ads altogether.

2. Some users might be more likely to click on the organic result because it reinforces the message sent by the top ad.

Our observations supported a slight preference for organic results: for 52% of the queries in Q_{eq} we observed $ctrO > ctrA$, and the average $ctrA$ was only 95% of the average $ctrO$ in Q_{eq}. That is, in this equal quality setting, users did have a bias towards clicking on organic results in spite of their lower position. We refer to this preference as the *non-commercial bias*.

4.2 Correlations? - The overall trend

In this section, we take a closer look at the relationship between $ctrA$ and $ctrO$. Suppose we have two queries q_i and q_j, and we know $ctrO_i > ctrO_j$, does this tell us anything about $ctrA_i$ vs. $ctrA_j$? In other words, is there a clear pattern of the relationship between $ctrA$ and $ctrO$?

One hypothesis is that users' perception of ads is independent of their perception of organic results, and we should expect no relation whatsoever. That is, the average $ctrA$ for queries with specific $ctrO$ should be relatively constant with respect to the value of $ctrO$ – it would simply be the average $ctrA$ for all queries.

However, we believe that a user has an integrated view of the sponsored and organic search results on a SERP. In particular, we hypothesize that there are two possible underlying forces that can lead to correlations:

- Antagonism: It is natural to assume that a user has limited time and energy to spend on a particular query, and thus she has to chose a selected few results she would click on and explore — in particular, she has to chose between ads and organics.

- Mutual growth: The mechanisms underlying ads retrieval and organic Web retrieval have their differences in most commercial search engines. Still, as queries are better served by the search engine, and the organic and ad results collectively better address the user's information need, both $ctrO$ and $ctrA$ can grow together at an aggregated level even as they compete for user's attention on each individual SERP.

Would one of these forces be the dominating factor, or would they cancel each other out and lead to no correlation? In order to answer this question, we examined the average trend in the data we collected.

Let $Q(x) = \{q_i \mid ctrO_i = x\}$ be the set of queries whose best organic CTR is x, we computed the average $ctrA$ for $Q(x)$ as

$$avgA(x) = \frac{\sum_{q_i \in Q(x)} ctrA_i}{|Q(x)|}$$

Figure 1(a) shows the overall trend of how $avgA(x)$ changes with respect to $ctrO$. We observe a clear non-linear correlation between them, thus ruling out the independence hypothesis. However, neither of the two aforementioned forces explains the trend on its own. For relatively small $ctrO$ values, we observe a mutual growth region, where $avgA(x)$ and $ctrO$ are positively correlated (i.e., positive slope in the cubic[2] fit). This is followed by an antagonistic region, where $avgA(x)$ and $ctrO$ are negatively correlated (i.e., negative slope in the fit). It is important to bear in mind that as the points in Figure 1(a) represent mean values, a positive slope should not be interpreted as $ctrA$ monotonically increasing with $ctrO$ at the individual query level.

Similarly we examined the average $ctrO$, or $avgO(x)$, for different $ctrA$ values (Figure 2(a)). Note that this analysis is complementary to the one summarized in Figure 1(a): since the correlation we observe there is non-linear, it would be possible for $avgO(x)$ to remain constant with respect to different $ctrA$ values — for example, a flat line with zero slope in Figure 2(a) could still be consistent with Figure 1(a). However, we can again observe both a a clear antagonistic region (for lower $ctrA$) and a mutual growth region (for higher $ctrA$).

4.3 Navigational and non-navigational queries

In order to test whether the observed trend is indeed a result of the interplay between the two forces described in the previous section, we would like to *decouple* their effects. Ideally, we wish to separate the queries into two sets: one dominated by the antagonistic force, and the other dominated by mutual growth.

Our insight is to exploit the fact that the intensity of the the antagonistic force is dependent on the amount of time and effort a user is willing to spend on a particular query: the more time one invests, the more results one is willing to explore (in the extreme case, one would click on all sponsored and organic results) and the less competition there is; the less time one dedicates to the the query, the less results one is going to explore and thus, a fiercer competition. It is also natural to assume that the amount of effort invested depends on the type of the information need. In particular, if a user is looking for a particular site, as in the case of a "Pandora Radio" query, he might be willing to invest far less exploration time than a user who is looking for "the meaning of life"[3].

Following this insight we separate the queries into *navigational* and *non-navigational* queries, in accordance with the taxonomy of web search introduced in [3][4]. The purpose of navigational queries is to reach a particular site that the user has in mind, and for such queries there is usually only one "right" result (`www.pandora.com` in the case of our example). We expect users to dedicate less exploration time to navigational queries, and thus for the antagonistic effect to dominate.

We obtained a *navigational score* for each query in our dataset automatically from an internal classifier trained on manually labeled examples. The value of the navigational score was between 0 and 1, representing the confidence of the query being navigational. The classifier was reported to yield an F1 score of 0.842 on queries randomly sampled from the most frequent one million queries in Yahoo! query log. To further confirm the accuracy of the navigational scores, we sampled a small set of 45 queries from our dataset, and asked a human annotator to label them into navigational and non-navigational queries (with high- and low- confidence). The following table shows the statistics of the navigational scores of the queries that received the respective manual labels:

manual label	mean	std. dev.
navig, high-confidence	0.64	0.16
navig, low-confidence	0.64	0.22
non-navig, low-confidence	0.54	0.17
non-navig, high-confidence	0.36	0.09

As we can see, the navigational scores are consistent with the human labels. In fact, out of the 36 queries that received high-confidence labels, with the exception of two queries, all of the queries labeled as navigational by the annotator received higher navigational scores than those labeled as non-navigational. Thus, we considered the navigational scores as a reasonable estimate of the likelihood of a query being navigational. We rank all queries by their navigational scores, and refer to the top decile as *navigational** queries, and the bottom decile as *non-navigational** queries. These represent queries classified as navigational and non-navigational with high confidence.

4.4 Dissecting the trend

Figure 1(b) is a breakdown of Figure 1(a) with respect to navigationality. The relation between $avgA(x)$ and $ctrO$ for navigational* queries is shown in black; in contrast, the same relation for non-navigational* queries is shown in green (the relation for the rest of the queries is shown in light blue).

We successfully decoupled the effects of the two forces: the antagonistic relationship dominates for navigational* queries,

[2]Piecewise cubic Hermite interpolation.

[3]Unless the user is just searching for the Monty Python movie.

[4]This work introduced three main types of queries: navigational, informational, and transactional. For our purposes we grouped the last two types together into the non-navigational class.

as the best line fit for these queries yields a steep and negative slope. In comparison, the positive slope for the non-navigational* queries indicates the dominance of mutual growth. Furthermore, as can be seen in Figure 1(c), there is a smooth transition between these two extreme cases: as the likelihood of the considered queries being navigational lowers, the slope gradually increases (decreasing in magnitude), eventually turning positive for those least likely to be navigational. A similar decoupling of the antagonistic and mutual growth trends can be observed in Figure 2(b) and Figure 2(c), further strengthening our observations.

4.5 Digression: the relation between the navigational score and query difficulty

Intuitively, navigational scores can also play surrogate as a measure of the difficulty of queries: highly navigational queries are typically the easier ones for most commercial search engines.

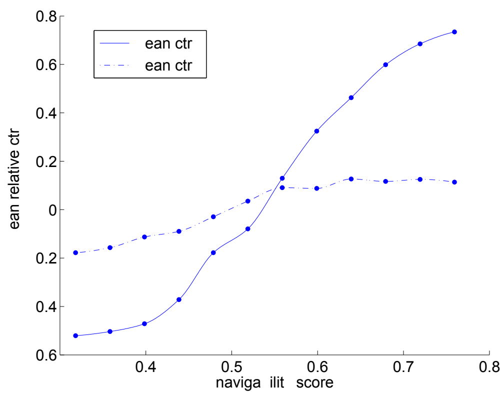


Figure 3: Average organic / ads clickthrough vs. navigational scores.

Indeed, as we can see from Figure 3, the average *ctrO* clearly increases with the navigational scores. In other words, for navigational queries, O^* better satisfies the user's information need. We would like to note that higher *ctrO* does not necessarily *prove* the query is easier. Given the definition of navigational queries, it could be that users' information need is more diverse for non-navigational queries and the CTR was spread out more evenly among different search results. Interestingly, as shown in Figure 3, the average *ctrA* increases with a much flatter slope with respect to navigational scores. The same "concentration of CTR" is not observed for the ads to the same degree.

Nonetheless, both *ctrO* and *ctrA* tend to be lower for queries with low navigational scores. Recall that this is the group of queries where mutual growth dominates. In other words, the mutual growth relationship dominates for those queries that were not best satisfied (collectively) by the best ad and organic results.

4.6 Ads vs. position-two organics

Is this mix of mutual growth and antagonistic relationships something unique between ads and organics? Or are they more generally applicable to any two items on a SERP?

We examined the relationship between the position-one (O_1) and position-two (O_2) organic results. As we can by comparing the plots in Figure 4 , the relationship between

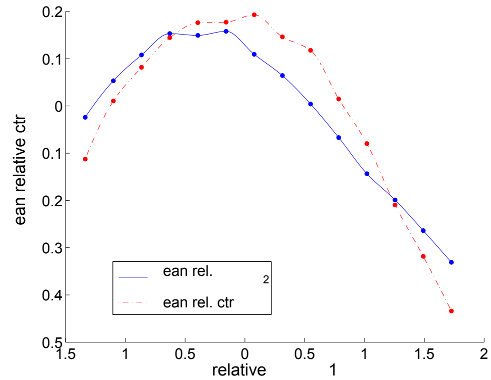


Figure 4: Average CTR of position-two organic results for different values of position-one organic CTR. Compare with average *ctrA* for different position-one organic CTR (dashed).

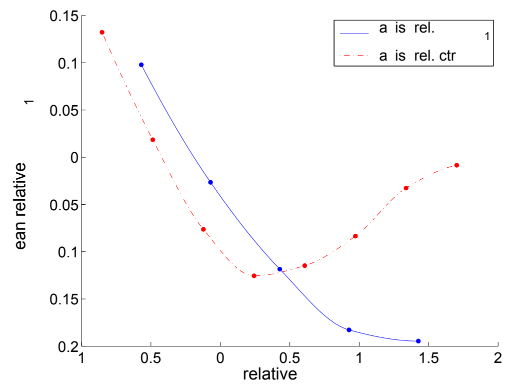


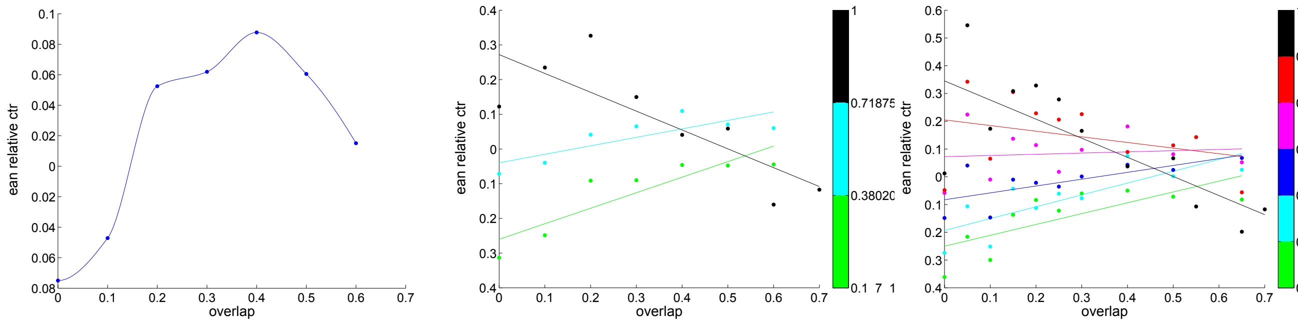
Figure 5: Average CTR of position-one organic results for different values of position-two organic CTR. Compare with average CTR of position-one organic for different *ctrA* (dashed).

O_2 CTR and the O_1 CTR is to some extent similar to the relationship between the ctrA and the O_1 CTR. However, in Figure 5 the mutual growth region is absent. Indeed, the antagonistic relationship is in general more prominent between the two organic results. The explanations behind this difference is beyond the scope of this study.

Note also that had position bias dominated user perception, we would expect the relationship between A^* and O_1 to be similar to the relationship between O_1 and O_2. Instead, A^* behaves more like O_2 in spite of the fact that it is shown above O_1.

5. THE IMPORTANCE OF BEING DIFFERENT

In the previous section, we examined the interplay between the CTR of sponsored and organic search results. The next interesting question to ask is: in the context of a given set of organic search results, which ad is perceived as more useful? If users take an integrated view of the entire search result page, what type of ads would they prefer: ads that are more similar to the organic results or those that provide diversity?



(a) Overall trend

(b) Breakdown for queries classified as navigational (black) and non-navigational (green) with high confidence

(c) Breakdown by navigational score

Figure 6: Average ad-organic overlap vs. ads clickthrough. Best linear fit is shown for (b) and (c) in order to emphasize the respective trends.

Both hypotheses are plausible. Ads that are similar to the organic results have a higher chance of being relevant to the query (assuming that the organic results are a reasonable representation of pages relevant to the query). Indeed, this idea has been exploited in the computational advertising literature. For instance, organic results have been used to classify queries with respect to a taxonomy designed for ad placement, which was subsequently used to improve search advertising [4]. In addition, as users tend to “trust” the non-commercial organic results, being similar to organic results may be subjectively perceived as evidence for high-quality.

On the other hand, ads too similar to the organic results can be perceived as redundant information. As we discussed in Section 4.1, users prefer the organic results when A^* and O^* lead to the same site. Thus, being different from the organic results can have its advantage as it offers something not available from the organic results, which can help to overcome user’s non-commercial bias. Indeed, there have been studies showing that user experience can benefit from diversity in the top organic search results [11, 18, 6]. Since ads are an integrated part of a SERP, it is reasonable to hypothesize that users can also benefit from diversity among sponsored and organic search results.

5.1 Data analysis

For each SERP$_v$, we computed the similarity between A^* and O^*. User studies have shown that the titles of sponsored and organic results play an important role when users are asked to judge the relevance of the results [12]. Based on this finding, we focused on a similarity measure computed from the titles of A^* and O^*. (See Section 5.4 for other similarity measures that we considered.) More specifically, let B_A and B_O be the set of terms (i.e., Bag-of-words) that appeared in the titles of A^* and O^*, respectively. The similarity measure, which we refer to as *overlap*, is defined as the Jaccard similarity coefficient:

$$\text{overlap} = \frac{|B_A \cap B_O|}{|B_A \cup B_O|}$$

For instance, overlap (“Free Radio”, “Pandora Radio - Listen to Free Internet Radio, Find New Music”) $= \frac{2}{9}$.

Figure 6(a) presents the overall trend of how average $ctrA$ changes when A^* has different overlap with the corresponding O^*. What we observed is a non-monotonic trend. Neither of the hypotheses we discussed could fully explain the observed data on its own, since both would lead to a monotonic dependency.

To understand this non-monotonic trend, we exploit the intuition gained from our analysis in Section 4: the way ads and organic results are perceived in the context of each other can change dramatically depending on the user’s information need. Indeed, after dividing the data according to the navigational score (Figure 6(b)) we observed the following dichotomy. For navigational* queries (i.e., queries that are very likely to be navigational), $ctrA$ decreases as the *overlap* increases (shown in black); that is, the preference for diversity dominates. For non-navigational* queries we observe the opposite trend (shown in green): the preference for similarity is salient. Furthermore, as can be seen in Figure 6(c), there is a rather smooth transition between these two opposite types of preferences when the likelihood of the considered queries being navigational varies.

5.2 Model

In this section, we describe a simple model that formalizes our observations, in particular, the non-monotonic relationship between $ctrA$ and *overlap*. Let $\delta_\mathbf{c}$ denote the event that an ad is clicked, and $\mathbf{Ov}$ be its *overlap* to O^*, we model the probability of an ad being clicked given $\mathbf{Ov}$: $P(\delta_\mathbf{c}|\mathbf{Ov} = x)$. Note that what is plotted in Figure 6(a) is an empirical estimate of this probability[5].

Let $\mathbf{QT}$ be the random variable that indicates the type of the query that the ad was shown for:

$$\mathbf{QT} = \begin{cases} n & \text{if the query is navigational} \\ \overline{n} & \text{otherwise} \end{cases}$$

[5]Recall that in the plots we display the *relative* CTR, and thus the negative values. The original range of CTR values was $[0, 1]$.

We have:

$$P(\delta_{\mathbf{c}}|\mathbf{Ov}=x) = \sum_{\alpha\in\{n,\overline{n}\}} P(\delta_{\mathbf{c}}, \mathbf{QT}=\alpha|\mathbf{Ov}=x)$$
$$= \sum_{\alpha\in\{n,\overline{n}\}} P(\delta_{\mathbf{c}}|\mathbf{QT}=\alpha, \mathbf{Ov}=x)P(\mathbf{QT}=\alpha|\mathbf{Ov}=x) \tag{1}$$

$P(\delta_{\mathbf{c}}|\mathbf{QT}=n,\mathbf{Ov}=x)$ is the probability that an ad is clicked given that the query is *navigational* and *overlap* is x. Recall that the black curve in Figure 6(b) is an empirical estimate of this probability. We observe that it is largely monotonically decreasing with respect to x. For simplicity we approximate it with a linear function

$$P(\delta_{\mathbf{c}}|\mathbf{QT}=n,\mathbf{Ov}=x) = b_1 - c_1 x$$

with $c_1 \in (0,1], b_1 \in [c_1, 1]$.

Similarly, based on the green curve in Figure 6(b), we approximate

$$P(\delta_{\mathbf{c}}|\mathbf{QT}=\overline{n},\mathbf{Ov}=x) = b_2 + c_2 x$$

where $c_2 \in (0,1], b_2 \in [0, 1-c_2]$.

Next, we consider $P(\mathbf{QT}=\alpha|\mathbf{Ov}=x)$ by examining the trend in our data. Figure 7 shows the fraction of navigational* queries for each overlap value, relative to the set of all navigational* and non-navigational* queries. We can observe that this dependency is of a linearly increasing type. Based on this observation we approximate

$$P(\mathbf{QT}=n|\mathbf{Ov}=x) = b_3 + c_3 x$$

where $c_3 \in (0,1], b_3 \in [0, 1-c_3]$.

Note also that

$$P(\mathbf{QT}=\overline{n}|\mathbf{Ov}=x) = 1 - P(\mathbf{QT}=n|\mathbf{Ov}=x).$$

Thus, equation (1) becomes:

$$\begin{aligned} &P(\delta_{\mathbf{c}}|\mathbf{Ov}=x) \\ &= (b_1 - c_1 x)(b_3 + c_3 x) + (b_2 + c_2 x)(1 - b_3 - c_3 x) \\ &= -Ax^2 + Bx + C \end{aligned} \tag{2}$$

where

$$\begin{aligned} A &= c_3(c_2 + c_1) \\ B &= c_3(b_1 - b_2) - b_3(c_1 + c_2) + c_2 \end{aligned} \tag{3}$$

This model formalizes the observations summarized in Figure 6 and Figure 7. In particular, since $A > 0$, as a function of x, $P(\delta_{\mathbf{c}}|\mathbf{Ov}=x)$ takes its maximum at $\frac{B}{2A}$. For a wide range of choices of c_i and b_i, we have $0 \leq \frac{B}{2A} \leq 1$, and thus this model explains the bell-shaped curve observed in Figure 6(a).

5.3 Further investigation

Previously, we discussed how the preference for similarity changes into a preference for diversity, as the users interact with the search engine differently. In particular, for a navigational intent, ads that are similar to the organic results tend to receive lower CTR, whereas the opposite trend is observed for a non-navigational intent. This is reflected in the differences between the models for $P(\delta_{\mathbf{c}}|\mathbf{QT}=n,\mathbf{Ov}=x)$ and $P(\delta_{\mathbf{c}}|\mathbf{QT}=\overline{n},\mathbf{Ov}=x)$ we proposed in Section 5.2.

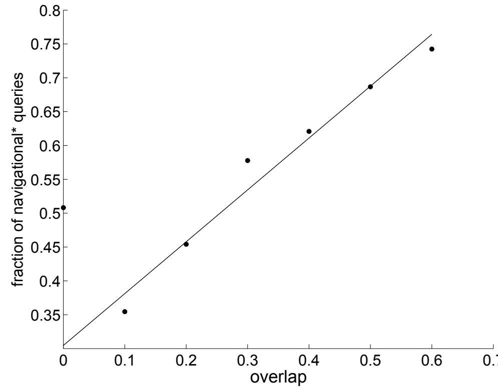


Figure 7: Fraction of navigational* queries for each overlap bin. The fraction is relative to the set of navigational* and non-navigational* queries.

In this section, we propose a possible explanation for this dichotomy.

Consider a user who submitted a navigational query like "pandora radio". The desired result is almost always a particular site (`www.pandora.com` in our case) that would receive the highest CTR among organic results (i.e., O^*). One might expect that the most useful ad would be an ad for "Pandora Internet Radio" — after all, this is the result that the user was expecting when she submitted the respective query. However, our findings indicate the opposite: for navigational queries, the ads that are very similar to O^* tend to receive lower CTR than those that are more different.

To understand the reason behind this seemingly counter-intuitive preference, we introduce the notion of *responsive* and *incidental* ads. Responsive ads directly address the user's information need: a "Pandora Internet Radio" ad would be a responsive ad for a navigational query like "pandora radio", or for a non-navigational query like "free online radio". Incidental ads do not aim to directly address the user's information need. Rather, they are loosely related in that the advertisers expect users who are interested in the query to be also interested in their product. For instance, an ad for "Discounted Bose Computer Speakers" would be an incidental ad for both "pandora radio" and "free online radio". Indeed, incidental ads would not be considered as reasonable results if presented as organic search results.

Based on the intuition that incidental ads would, in general, not be considered as suitable organic results, we assume that ads that were not similar to the organic results are more likely to be incidental[6]. Thus, under this interpretation, our results show that in the case of navigational queries incidental ads are more effective, whereas in the case of a non-navigational query, incidental ads tend to receive lower CTR. Recall the non-commercial bias we discussed in Section 4.3. When a user has a navigational intent, a responsive ad would not offer any incentive to overcome this bias, whereas an incidental ad would provide useful information complementary to what was offered by the targeted

[6]One can envision using a more elaborate classifier that distinguishes between these two types of ads to produce more accurate labels. Given the scope of this study, we consider the similarity-measure as an approximation and leave the classifier as interesting future work.

organic result. More importantly, in the case of navigational queries, the *cost of diverting* from the original information need is very small: the user could always retrieve the targeted organic result with minimal efforts. In contrast, when the intent of the user is non-navigational, this cost would be higher since the goal of the user is not to retrieve a particular known site, but to find out what best satisfy her information need (for example, the best free online radio). As a result, the user would be less willing to interrupt the search for a potentially related need (e.g., a pair of discounted computer speakers) provided by an incidental ad.

5.4 Predicting relative CTR

In the previous section we examined the interplay between the organic and the sponsored search results at an aggregated level. In this section, we investigate the effects of this interplay at the query level: given a SERP_v (as defined in Section 3), we study whether we can better predict which sponsored search result gets higher clickthrough rate by using the context of the organic search results.

More specifically, we define a prediction task as follows. Let $\overline{pos}(a')$ be the average position in which an ad a' has been shown (recall that an ad is not necessarily always shown in the same position). We say that two ads, a' and a'', are shown at a *comparable* position if $|\overline{pos}(a') - \overline{pos}(a'')| < 0.5$. For a given query, we identify the most clicked ad A^*. We then select an ad a shown for the same query at a comparable position that has different CTR, domain and title than A^*. This pair of ads is then arranged in random order as (a_1, a_2). For each query we consider as many pairs of ads as we could find subject to the above criteria. The prediction task is to see whether we can build a classifier to correctly predict if $CTR(a_1) > CTR(a_2)$.

Our goal is to understand whether by using novel features inspired by our observations above, namely, features that reflect the interaction between the sponsored and organic results, we can achieve better CTR prediction than a system employing only traditional features based on the relation between the query and the content of the ad. The set of the 20 features that we introduce is presented in Table 1. The first 18 features in the table reflect the similarity of the ads in the pair to the most clicked organic result O^*; the last 2 features describe the SERP_v from which the respective pair was extracted. We used the Weka[7] implementation of the C4.5 decision tree as the classifier. In what follows, we refer to the classifier using only the features described in Table 1 as ORG_SIM. We compare this classifier against a classifier which ignores the context of the organic results (QSIM), but employs word and phrase features to predict the relevancy of each ad to the query, as described in [4].

Table 2 summarizes the performances achieved by these systems on the prediction task. We can see that ORG_SIM, which uses novel features characterizing the ad-organic interaction, outperforms the QSIM system with an improvement in accuracy of almost 2%. Furthermore, using both types of features in the classifier (ORG_SIM + QSIM) outperformed each system individually. These improvements are all statistically significant ($p < 0.001$ according to a paired t-test). Our results indicate that not only are our proposed features useful for this task, but they also supply complementary information to the conventional feature set.

[7]Available at `http://www.cs.waikato.ac.nz/ml/weka/`

Source	Type	Description
Title	ad	Jaccard coef. of the titles of a_1 and O^*
	ad	Jaccard coef. of the titles of a_2 and O^*
	pair	difference of the previous two values.
	ad	cos similarity of the titles of a_1 and O^*
	ad	cos similarity of the titles of a_2 and O^*
	pair	difference of the previous two values
Snippet	ad	Jaccard coef. of the snip. of a_1 and O^*
	ad	Jaccard coef. of the snip. of a_2 and O^*
	pair	difference of the previous two values
	ad	cos similarity of the snip. of a_1 and O^*
	ad	cos similarity of the snip. of a_2 and O^*
	pair	difference of the previous two values
Domain	ad	edit dist. of the domains of a_1 and O^*
	ad	edit dist. of the domains of a_2 and O^*
	pair	difference of the previous two values
URL	ad	edit distance of the URL of a_1 and O^*
	ad	edit distance of the URL of a_2 and O^*
	pair	difference of the previous two values
SERP_v	serp	navigational score
	serp	the CTR of O^*

Table 1: ORG_SIM features for a pair (a_1, a_2). The features of type *ad* characterize each of the ads individually, the features of type *pair* characterize a given pair of ads, features of type *serp* is common to the all ads displayed on the SERP_v.

Features	Accuracy
QSIM	59.5562
ORG_SIM	61.4570
ORG_SIM + QSIM	62.8222

Table 2: Performance on pair-wise prediction using different feature sets: average accuracy over ten-folds cross-validation.

6. RELATED WORK

Web search and sponsored search has been an area of active research in the past few years. A few studies have been reported that focus on CTR prediction and analysis of click behavior in sponsored search. Regelson and Fain [16] analyze sponsored search data by examining the ad CTR based on its position, bid phrase, and bid phrase volume decile. They also show that using clustering of similar bid phrases by features extracted from their text, can improve the CTR prediction compared to other simpler bucketing of bid phrases as for example volume deciles. Richardson et al. [17] show that features from the ad title, creative, and ad group can further improve the CTR prediction.

In addition to such predictive techniques, Agarwal et al. [1] report a reactive method for click aggregation on multiple levels of a page and ad taxonomy. While this method has been examined in the context of content match, it can easily be adapted to sponsored search.

Web search click data has also been the focus of examination of the search and information retrieval communities. Two main areas of investigation here have been using clicks for direct improvements in ranking [2, 14, 13]; and modeling, analysis and prediction of user clicks in web search [7, 5, 8]. To the best of our knowledge, there have been no previous studies that analyzed the interplay between the organic and the sponsored search results in terms of the CTR.

In [12] Jensen and Resnick report a user study and a survey examining user behavior and perceptions when interacting with the SERP. The first experiment shows that, on average, the users examine the organic results before the ads and that the organic results is perceived as being more relevant and less biased. They also point out that when judging the relevance, the users rely greatly on the titles of the ads and the organic results. While the findings of this study are largely in line with ours, in this work we focus on the actual behavior of large number of users in the real world environment, and quantify the different behavior as a function of the CTR and similarity of the organic results and the ads.

7. CONCLUSIONS

Result pages of major Web search engines routinely include information from multiple sources, where its two most prominent constituents are organic search results and sponsored ads. To the best of our knowledge, no prior study has examined the interaction between these two complementary sources of information, and in this paper we sought to address this issue. Specifically, we used real-life click data from a major Web search engine to understand how their competition for user attention can turn into mutual growth of click through rates.

Recent research on query expansion for advertising advocated using Web search results for query augmentation [4], essentially assuming that topical similarity between the organic results and the ads is desirable. One of the goals of the present study was to verify this hypothesis, and we (somewhat unexpectedly) found that the desired degree of organic-to-ad similarity depends on the type of query at hand. To explain our observations, we introduced the notions of responsive and incidental ads, which correspond to the narrow and the broad interpretation of the user's information need underlying the query.

We believe our findings will allow us to better understand the ways in which different types of queries can be best answered. Our findings also suggest that *jointly* optimizing the selection of organic and ad results might offer users a more productive Web search experience, as well as enable search engine companies to better monetize this experience.

Acknowledgments. We thank Lillian Lee and Alex Niculescu-Mizil for insightful conversations, and the anonymous reviewers for very helpful comments. We also thank Peter Ciccolo, Sergiy Matusevych and Don Metzler for their assistance with the data collection. This paper is based upon work supported in part by the National Science Foundation grant IIS-0910664.

8. REFERENCES

[1] D. Agarwal, A. Z. Broder, D. Chakrabarti, D. Diklic, V. Josifovski, and M. Sayyadian. Estimating rates of rare events at multiple resolutions. In *KDD*, 2007.

[2] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 19–26, New York, NY, USA, 2006. ACM.

[3] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36, 2002.

[4] A. Broder, P. Ciccolo, M. Fontoura, E. Gabrilovich, V. Josifovski, and L. Riedel. Search advertising using Web relevance feedback. In *CIKM'08*, 2008.

[5] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 1–10, New York, NY, USA, 2009. ACM.

[6] H. Chen and D. R. Karger. Less is more: probabilistic models for retrieving fewer relevant documents. In *SIGIR '06*, pages 429–436, New York, NY, USA, 2006. ACM.

[7] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 87–94, New York, NY, USA, 2008. ACM.

[8] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *SIGIR '08*, pages 331–338, New York, NY, USA, 2008. ACM.

[9] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, 2007.

[10] D. Fain and J. Pedersen. Sponsored search: A brief history. In *Second Workshop on Sponsored Search Auctions*, 2006.

[11] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *18th International World Wide Web Conference (WWW2009)*, April 2009.

[12] B. Jansen and M. Resnick. Examining searcher perceptions of and interactions with sponsored results. In *Workshop on Sponsored Search Auctions*, 2005.

[13] T. Joachims. Optimizing search engines using clickthrough data. In *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, New York, NY, USA, 2002. ACM.

[14] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *SIGIR '05*, pages 154–161, New York, NY, USA, 2005. ACM.

[15] L. Lorigo, M. Haridasan, H. Brynjarsdóttir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, and B. Pan. Eye tracking and online search: Lessons learned and challenges ahead. *J. Am. Soc. Inf. Sci. Technol.*, 59(7):1041–1052, 2008.

[16] M. Regelson and D. Fain. Predicting click-through rate using keyword clusters. In *Second Workshop on Sponsored Search Auctions*, 2006.

[17] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: Estimating the click-through rate for new ads. In *WWW'07*. ACM Press, 2007.

[18] A. Swaminathan, C. Mathew, and D. Kirovski. Alta vista invites advertisers to pay for top ranking. In *Technical Report MSR-TR-2008–015, Microsoft Research*, 2008.

[19] N. Y. Times. Alta vista invites advertisers to pay for top ranking. *http://www.nytimes.com/library/tech/99/04/biztech/articles/15online.html*, 1999.