Sponsored search: an overview of the concept, history, and technology

Bernard J. Jansen* and Tracy Mullen

College of Information Science and Technology,
University Park, Pennsylvania 16802, USA
Fax: +01-814-865-6426 Fax: 814-865-6426
E-mail: jjansen@acm.org
E-mail: tmullen@ist.psu.edu
*Corresponding author

Abstract: The success of sponsored search has radically affected how people interact with the information, websites, and services on the web. Sponsored search provides the necessary revenue streams to web search engines and is critical to the success of many online businesses. However, there has been limited academic examination of sponsored search, with the exception of online auctions. In this paper, we conceptualise the sponsored search process as an aspect of information searching. We provide a brief history of sponsored search and an extensive examination of the technology making sponsored search possible. We critique this technology, highlighting possible implications for the future of the sponsored search process.

Keywords: web search engines; sponsored search; paid search; web advertising; sponsored links; sponsored results; web searching; online auctions; online advertising; click fraud; electronic business.

Reference to this paper should be made as follows: Jansen, B.J. and Mullen, T. (2008) 'Sponsored search: an overview of the concept, history, and technology', *Int. J. Electronic Business*, Vol. 6, No. 2, pp.114–131.

Biographical notes: Bernard J. Jansen is an Assistant Professor at the College of Information Sciences and Technology at The Pennsylvania State University. He has more than 120 publications in the area of information technology and systems, with papers appearing in a wide range of journals and conferences. Several agencies and corporations have supported his research. He has received several awards and honours, including an ACM Research Award and six application development awards, along with other writing, publishing, research, and leadership honours.

Tracy Mullen is an Assistant Professor at the College of Information Sciences and Technology at The Pennsylvania State University. She has previously worked at Lockheed Martin, Bellcore, and NEC Research and received her PhD in Computer Science from the University of Michigan. Her research focus is on decentralised decision making via computational markets and multi-agent systems. Her research papers have been published in *ACM Transactions on Internet Technology, Decision Support Systems, Electronic Commerce Research, IEEE Computer, Mathematics and Computers in Simulation and Operating Systems Review*, among others.

1 Introduction

Web search engines are indispensable tools for interacting on the web. In addition to addressing information requests, modern web search engines are navigational tools that take people to specific websites or are an aid in browsing. People also employ search engines as applications to carry out ecommerce transactions. People continue to employ search engines in new and increasingly diverse ways, and search engines are constantly trying to improve the retrieval aspects of their services. One novel innovation for improving web retrieval has been sponsored search (i.e., where content providers pay search engines for traffic from the search engine to their websites). With sponsored search, major web search engines such as Yahoo!, MSN/Live.com, Google, and Ask have significantly altered online commerce. Battelle (2005) provides an overview of the factors that have led to the development of these sponsored web search platforms.

The impact of sponsored search on the accessibility of information and services on the web has been enormous. Sponsored search has played a critical role in supporting access to the many free services (i.e., spell checking, currency conversion, flight times, desktop searching applications, etc.) provided by search engines that have rapidly become essential to so many web users. Without the workable business model of sponsored search, it is doubtful if the major web search engines could finance anything close to their current infrastructures. These infrastructures provide the capability to crawl billions of webpages, index several billion documents (e.g., textual, images, videos, news papers, blogs, and audio files), accept millions of web queries per day, and present billions of links per week.

Sponsored search has also provided a workable business model for meta-search engines, which are extremely beneficial for searches needing high recall and requiring a thorough coverage of a topic. Sponsored search provides an effective method for overcoming the inherent biases in the technical implementation of particular web search engines (Introna and Nissenbaum, 2000) as well by allowing content providers to move their links to the first Search Engine Results Page (SERP) at relatively low cost. In doing so, sponsored search is an essential tool vital to the success of many businesses. It is fair to say that without sponsored web search, the web search engine market – indeed the web! – would look far different than it does today.

Certainly, sponsored search has drawn some criticism. The Federal Trade Commission (Hansen, 2002) reported that search engines do not adequately label sponsored links. Marable (2003) reported that most searchers did not recognise sponsored links. The Pew Internet and American Life Project (Fallows, 2005) reported that 38% of searchers reported that they were unaware of the distinction between sponsored links and non-sponsored links. In the same study, fewer than 17% of survey respondents reported that they could always tell which links were sponsored and which were non-sponsored. When searchers do recognise sponsored links, they tend to find them less relevant.

Using a lab study, Jansen and Resnick (2006) reported that 65% of the study participants did not typically view sponsored listings, stating that they considered them less relevant than the non-sponsored listings. Jansen and Resnick (2006) reported that participants were unconcerned whether the listings were sponsored or non-sponsored. Their primary concern was relevance. In fact, when searchers did view and evaluate links in response to given queries, the ratings of the sponsored links were identical to the non-sponsored links. Also, Jansen (2006) showed that sponsored and non-sponsored link are equivalent in terms of relevance.

Additionally, sponsored search has suffered from efforts to subvert the system with click fraud (Kitts et al., 2005b).

Given the implications of sponsored search as the predominant business model for web search engines and as a key component of the content provided to searchers, there are few academic papers addressing sponsored search in a comprehensive manner. This paper addresses this need. We present an overview of the processes related to sponsored search from an information searching (Wilson, 2000) perspective. We then present an historical overview of sponsored search in relationship to web advertising. Building on this, we then discuss in detail the technology behind sponsored search with a focus on the online auction process. Finally, we end with the implications and challenges for sponsored search.

2 The concept of sponsored search

The process of sponsored search is continually evolving and gaining complexity as a means of satisfying both web searchers' desire for relevant information and content providers' desire for targeted traffic to their websites. Although the payment process has undergone several incarnations, the conceptual elements of sponsored search have remained essentially the same (Jansen and Molina, 2006). These elements are (see Figure 1):

- *Provider*. A person or organisation interested in generating user traffic to a particular website for some specific purpose. We use the term 'provider' rather than advertiser to highlight that one can view sponsored search as a version of providing relevant content to a searcher, and not solely as an advertising medium.
- *Provider content*. A set of keywords (representing concepts) along with the associated Uniform Resource Locators (URLs), titles, and descriptions, typically referred to as an advertisement or a sponsored link. Although, these terms have a heavy commercial interpretation, their use has become commonplace within the sponsored search domain. Therefore, we use them in this paper when referring to the provider content displayed on the SERP.
- *Provider bids*. Bids for specified keywords that are a monetary valuation of traffic to a particular website by a provider.
- *Search engine*. A search engine that serves the advertisement in response to user queries on SERP, relevant websites, or email.
- *Search engine review process.* A method utilised by a search engine to ensure that the provider's content is relevant to the targeted keyword on contextual material.
- *Search engine keyword and content index.* A mechanism that matches provider's keywords to user queries or to contextual material.
- *Search engine user interface*. An application for displaying provider content as links in rank order to a searcher. Typically, the interface displays the sponsored links with non-sponsored links on a SERP, within email messages, or along side content on a web page.

- *Search engine tracking.* A means of matching keywords to queries, gathering provider's content, bids, metering clicks, and charging providers based on searcher clicks on their displayed links.
- *Searcher*. An agent (i.e., human or automated surrogate) that actually clicks on a sponsored link that is deemed relevant.

Figure 1 The participants, goals, and process of sponsored search on the web

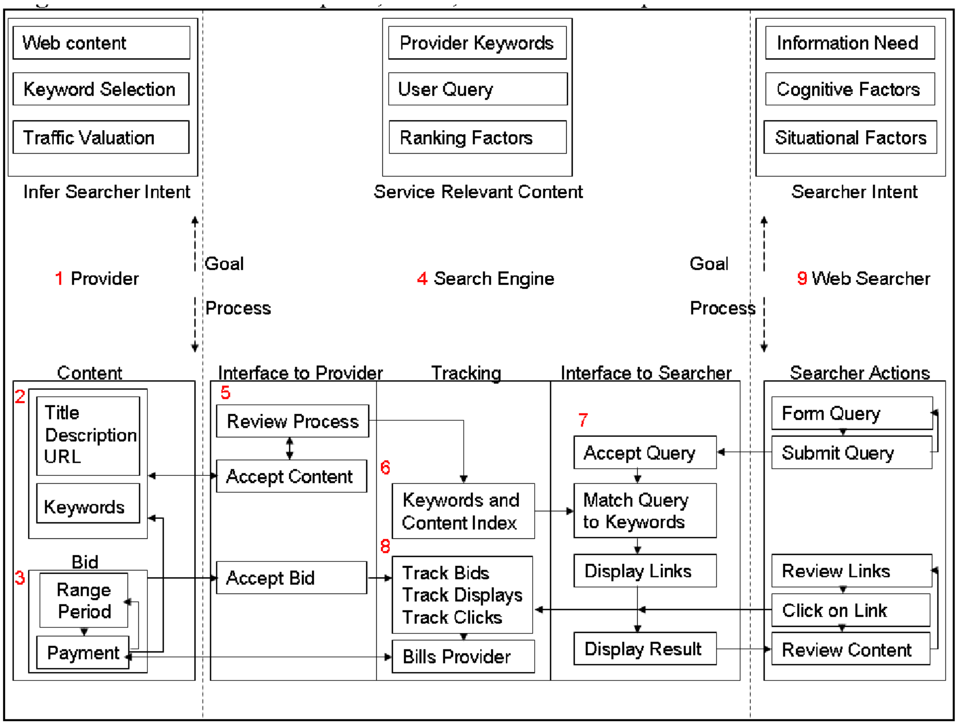


Figure 1 presents the sponsored search process as an aspect of information searching rather than strictly as an advertising venue. Along the upper half of Figure 1, the three major participants (providers, search engines, and searchers) have mutually supporting goals. The web searcher has some information needs bounded by cognitive, affective, and situational factors. Content providers select terms and search phrases (i.e., keywords) that they believe:

- are likely to be submitted by searchers
- will be applicable to their web content
- will be relevant to the underlying intent of the searcher.

These content providers also tailor the presentation of the search results to conform to the targeted searchers, possibly presenting several variations linked to particular sets of queries within a given campaign. This content is also known as a sponsored listing to differentiate it from the organic, or non-sponsored, listings on the SERP.

The search engine can also serve these sponsored links on a vast network of websites that are deemed relevant to the provider's content or in search engine hosted email services. This is known as contextual advertising, sometimes referred to as content targeting. The idea of this approach is that visitors to these websites will also be interested in the content provider's website. Other than occasionally mentioning it when it directly affects web searching, contextual advertising is outside the scope of this paper.

Search engines provide the mechanism for this sponsored search process to occur, shown in the lower half of Figure 1. The pay-per-click model is the most common payment method, although others such as pay-for-impression, pay-per-action, and pay-per-call also exist. The content providers normally pay the search engines to present their tailored web result (title, description, and URL) whenever a searcher submits one of these terms and clicks on the link. They can also pay the search engine when a user clicks on their link when serviced on another website. The provider can tailor this matching algorithm from exact targeted matches to very loose matches to account for various spellings and misspellings as well as term usage. The search engine determines when a match occurs between the searcher's query and a keyword bid upon by the provider.

The provider pays the search engine because of a bid placed on the keyword. A bid usually includes a maximum price per keyword and can include period of activation, language, geographical and other constraints. Multiple content providers may want to pay a search engine for the same keyword or phrase. When this occurs, ranking (i.e., which result goes on top) is handled by an electronic auction that determines the order of the advertisements.

Various search engines factor in other elements into their ranking schemes, such as the Click-Through-Rate (CTR) of advertisements, although this innovation was developed by Google. This approach helps address the concern that search engines will present less relevant content to the searcher solely for profit. It also ends up benefiting the search engine and the content provider. Since the search engine is typically paid on a per-click basis, total revenue is a function of Cost-Per-Click (CPC) and CTR for advertisements placed in the SERPs. Therefore, including CTR in the ranking algorithm increases overall revenue, which means more profits for the search engine and more clicks for content providers.

When multiple providers bid on a particular term or phrase, the result is often higher minimum and maximum bids. This is not always true, though. Generic single keywords often have many bidders, but relatively low maximum bids. The search engine sets the minimum bid on any keyword. For competitive markets, the bids can get much higher. For links serviced on the contextual network, the search engine splits the revenue per click with the website owner. The price that providers pay per click for contextual links is typically lower than those on the SERP.

The point of sponsored search is to provide a mechanism for providers to get searchers to visit their websites. When a searcher submits a query, reviews the SERP, and clicks on a sponsored link, the searcher's browser displays the provider's web page pointed to by the link. The search engine tracks this click, along with all the other clicks within a given period. At the end of this period, the search engine bills the provider, providing various statistics concerning the outcome of the provider's campaign. From an analysis of these statistics, the provider can alter bids, maximum budget per period, and selection of keywords in real time. By engaging in and 'buying' search phrases, these content providers become active participants in the information seeking process of searchers in a very dynamic way.

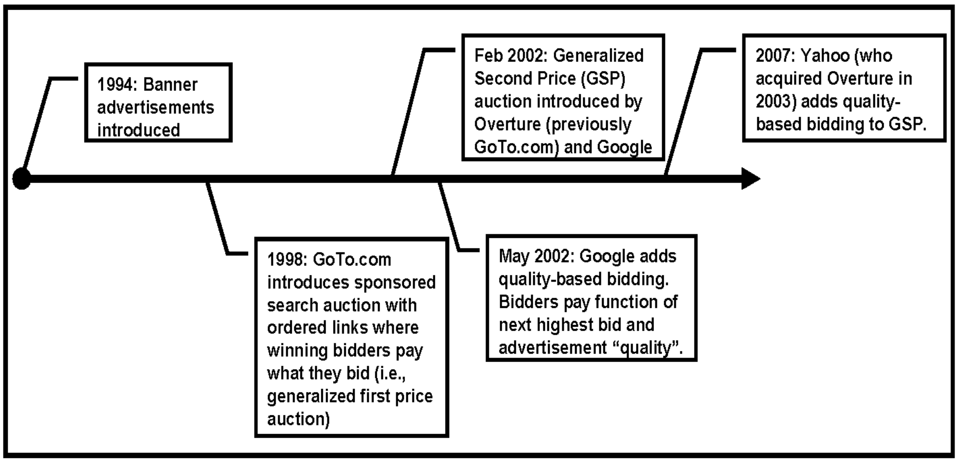
This accounting is one reason that sponsored search is so popular for businesses and organisations. In most models of offline advertising, there is little accountable with the cost being reach (i.e., how large is the potential audience for a particular advertisement). Sponsored search is also a viable revenue model for search engines, and it appears to provide relevant content to web searchers.

Sponsored search has undergone a variety of incarnations since its inception.

3 History of sponsored search auctions

Web advertising before 1998 consisted of banner advertisements generally priced by the number of impressions delivered (i.e., Cost-Per-Thousand (CPM) pricing). GoTo.com (renamed Overture in 2001, and acquired by Yahoo! in 2003) created the first sponsored search auction, and Google's first sponsored search auction followed in 2002 (Fain and Pedersen, 2006). Between them, Yahoo! and Google have advanced the sponsored search auction format since, as shown in Figure 2. Additionally, there are several other players in the sponsored search market, including Ask, LookSmart, and Microsoft.

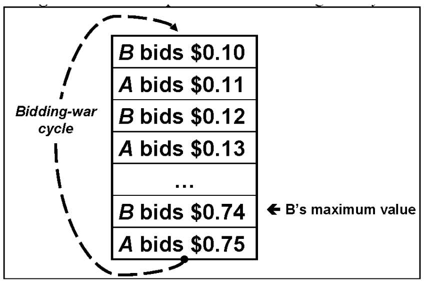
Figure 2 Timeline for sponsored search auctions



GoTo.com's original sponsored search auction was a Generalised First-Price (GFP) auction; the advertiser who bids the highest wins the top slot and pays what it bid. The second highest bidder wins the second slot and pays its bid amount, and so forth. While conceptually straightforward, this auction format led to bidding-war cycles that consumed advertiser time and diminished search engine profits (Edelman and Ostrovsky, 2007).

Example (first-price auction): Suppose advertiser *A* will pay up to \$ 1 for the keyword 'coffee', while advertiser *B* values the same keyword at \$ 0.74. If *B* starts by bidding the lowest possible price, say \$ 0.10, then *A* would bid \$ 0.11 to win the first advertisement slot. Advertiser *B* would respond by bidding \$ 0.12, and so forth. Once *A* bids \$ 0.75, then *B* will not bid \$ 0.76 since it only values the keyword at \$ 0.74. To acquire the second advertisement slot, *B* simply has to bid $ 0.10. Now, *A* only needs to bid $ 0.11 to win the first slot, and so the cycle starts over, see Figure 3.

Figure 3 First price auction bidding-war cycle



Developers of Google's AdWords platform changed the pricing scheme from a first price auction to a more stable second price auction. In a single-item second price auction, the highest bidder wins but only pays the second-highest bid price plus some small delta.

Example (second-price auction): Consider our previous example where advertiser *A* is willing to pay up to $ 1 for the keyword 'coffee', and advertiser *B* will pay up to $ 0.74. In a sealed-bid environment, recall that only the auctioneer knows the value of all bids. If *A* bids $ 1, and *B* bids $ 0.74, then in a second-price auction, *A* pays $ 0.74 + $ 0.1 (the minimum delta) while *B* pays the minimum bid amount of $ 0.10. Now suppose A and *B* bid strategically rather than honestly. If *A* bids $ 0.78 (i.e., less than what the advertisement is worth to A) and *B* bids $ 0.80 (i.e., more than what the advertisement is worth to *B)*, then *B* wins with a price of $ 0.79. This is 5 cents more than the slot is worth to *B*. On the other hand, if *A* bids $ 0.70 and *B* bids $ 0.65 then *A* gets the slot for $ 0.66. However, *B,* with a bid of $ 0.74 could have won the slot for $ 0.71, and effectively loses 3 cents of value. However, when the bidders can see, or infer, their competitors' bids, clearly there are other strategic possibilities. For example, if *A* truthfully bids $ 1, then *B* can safely bid $ 0.99 instead of $ 0.74, forcing *A* to pay $ 1 for the advertisement slot instead of $ 0.75.

Developers of Google's AdWords platform changed the pricing scheme from a first price auction to a more stable second price auction. In a single-item second price auction, the highest bidder wins but only pays the second-highest bid price plus some small delta.

Google auctions extend this single-item second price auction to a multi-item second price auction, known as a Generalised Second Price (GSP) auction, to bring more stability to the auction bidding, increase profits, and help reduce (but not eliminate) strategic bidding.

Google also changed the standard allocation scheme. Instead of ranking advertisements by bid price alone, they compute a quality score derived from the bid amount, the advertisement's CTR, keyword relevancy along with landing-page and site quality. The CTR measures the rate at which web searchers click through to the advertiser's website when shown an advertisement. Google's other quality-based criteria serve to penalise sites that use deceptive practices or have exceptionally poor websites. These criteria are assessed through both human and automated methods. In this way, Google can protect the user's search experience and also increase their profits, since users are more likely to click on relevant advertisements, and thereby generate Google revenue.

These two auction mechanism changes have helped make Google's auction more stable and more profitable than the original first-price auction. Yahoo! Search Marketing (known then as Overture) quickly updated its pricing scheme to second-price after Google in 2002 and implemented quality-based bidding in the form of its Quality Index (2007). Thus, sponsored search vendors began with basic market institutions, and in the process of fixing deficiencies and aligning the auction mechanism more closely to domain requirements, ended up designing novel auction mechanisms. These sponsored search auctions have effectively become large-scale laboratories for experimental and computation economics (Roth, 2002), driving new research in game theory.

4 The technology of sponsored search

In this section, we present some foundational material concerning properties of sponsored search auction. We then reflect on the economic problem domain landscape, the existing sponsored search market institutions, and finally consider participant issues including strategic behaviour, bidding expressiveness, and the use of software agents.

4.1 Auction properties

McAfee and McMillan (1987) define an auction as "a market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from the market participants". Designing market institutions falls under the rubric of economic mechanism design. Mechanism design constructs the 'rules of interaction' for economic transactions that will yield some desired outcome (Varian, 1995). The desired outcome may be to maximise seller profit, minimise buyer cost (as in a procurement auction), allocate resources efficiently, discourage coalitions, and so forth. The resultant interaction rules should ensure that each player has the incentive to behave as the designer intends. Clearly, to successfully design a market institution, one must consider both institution design, and participant behaviour.

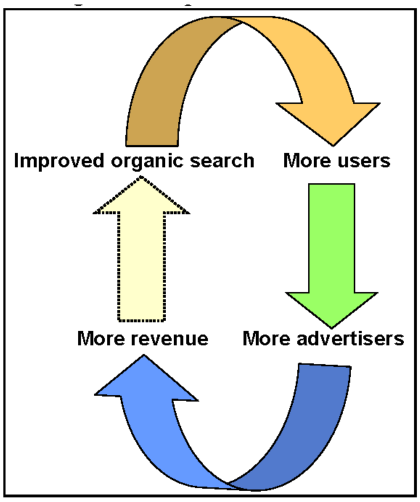
4.1.1 Competitive landscape

An individual sponsored search auction is actually part of a much larger competitive landscape. Within a given search engine, advertisers (conforming to the jargon of the field, we refer to a content provider bidding on a keyword as an advertiser) can buy placement (a.k.a. position) across multiple keywords, times, and geographic locations, which means that bidders are competing across multiple auctions. At an even more macro level, advertisers can choose to take their business to any one of several search engine companies. As shown in Figure 4, search engines make money by attracting advertisers, but to attract advertisers, they must attract users, and to attract users, they generally also have good non-sponsored (a.k.a., organic, non-paid, algorithmic) search results. As Michael Schwarz points out, the main competitor to Yahoo! sponsored search is Google's organic search results (and vice versa) – so the different sponsored search companies are not competing directly (Balcan et al., 2006).

By keeping sponsored and non-sponsored search listings clearly marked, search engines can avoid diminishing their credibility with users. Diminished credibility could reduce their market share and other user-based revenue such as licensing opportunities

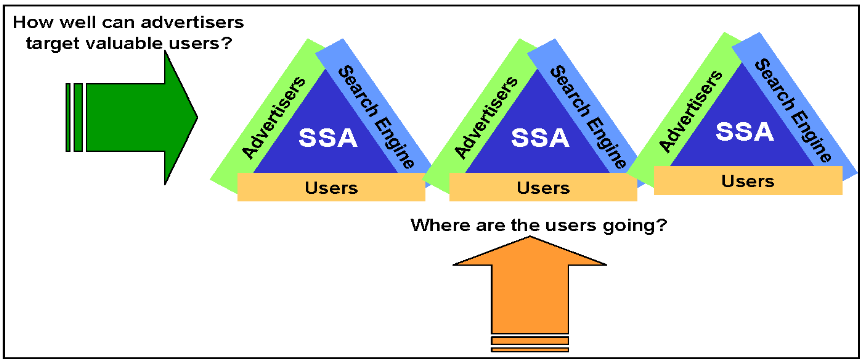
(Bhargava and Feng, 2002). As Bhargava and Feng observe, in a broader context, search engines are information gatekeepers (Baye and Morgan, 2001), who connect content providers with users, and must, therefore, balance their revenue streams.

Figure 4 High-level sponsored search revenue chain



While one key consideration for sponsored search companies is attracting users, another is allowing advertisers to target the appropriate users and to manage advertiser risk within the Sponsored Search Auction (SSA), see Figure 5. In this area, the sponsored search companies are competing directly with each other based on the ability to express useful preferences across keywords and searches, and the kinds of payment mechanisms selected.

Figure 5 Key considerations for successful sponsored search



4.1.2 Market institutions

In sponsored search auctions, the advertisement placement (i.e., the rank at which the sponsored line appears) are the goods being auctioned. Since higher slots generally have higher click through rates (assuming advertisement quality is the same), advertisers bid to get their advertisement placed higher and pay based on user clicks. Once a user types in a search query, an auction is formed using bids that match the search term. Advertisers specify if they want their keywords to match the user's search query exactly (e.g., do not match if the user types in extra keyword terms), more broadly, or they can exclude negative matches (e.g., match 'truffles' but not if combined with the keyword 'chocolate'). Advertisers can also specify criteria such as geographic region or time period in which they would like their advertisements to display. The advertisement with the highest ranking is placed at the top of the sponsored search listings, the next highest under it and so on. If the search term has no bids, or no bids above the minimum bid, then no advertisements are assigned to the page, and the inventory is wasted. This makes advertising placements on SERPs a very perishable commodity (think 'airline seats').

Two market design questions are how to allocate advertisements to slots? (i.e., how to rank them) and how to set slot prices? (i.e., what price to charge). In a generalised first price auction, advertisements are ranked strictly by bid price, so that the higher the bid, the higher the placement. This is also true of a second price auction. For each click on an advertisement, bidders pay their bid price. However, strategic players can cause price cycles that both diminish search engine revenue and result in higher auction monitoring costs for the advertisers. Google's AdWords altered the allocation and pricing aspects of the GoTo auction model so that the ranking allocation depends on the bid amount and the advertisement relevance or quality. So, allocation was not longer a function of just price, it just is a multivariable function of price and 'quality score'.

The quality measure includes expected CTR, along with other quality indicators as described above. Given two advertisements with the same bid, the one with a higher click through rate will be placed higher. Thus, advertisements are effectively ordered by both the search engine's expected revenue and the users' perceived aggregate interest. Before 2007, Yahoo's Search Marketing GSP model used the bid amount as the only factor in ranking. When all bidders have identical CTRs and other quality measures, then these two auction types are identical. Depending on the correlation between advertisers' willingness to pay and their advertisement quality, both Google's and Yahoo's particular ranking mechanisms can outperform each other under specific conditions, although Google's appears to require fewer assumptions (Lahaie, 2006; Feng et al., 2007).

In second-price auctions, a bidder's payment is based only on the allocation and the bids of others, not on the bidder's own bid. Auction theory shows that under certain conditions (e.g., sealed-bid, one-shot), a second price auction is *incentive compatible*, meaning that the optimal strategy for bidders is to bid their true valuation (Vickrey, 1961). This eliminates time-consuming strategic game play and ensures that the item is sold to the bidder who values it the most. The VCG auction is a multi-item extension to the single-item second price auction that retains the incentive compatibility property by charging each bidder the opportunity cost imposed upon the other bidders. Developed using game theory, the Vickrey-Clark-Groves (VCG) auction is an incentive compatible multi-unit version of the second price auction. For each click on an advertising position, a bidder pays the externality (or displacement costs) imposed on the other bidders by its presence.

Example (VCG auction): Suppose Advertiser *A* bids \$ 10 per click, while advertiser *B* bids \$ 5, and advertiser *C* bids \$ 1. There are two slots available, and the first slot has gets an average of 100 clicks per hour, the second gets 50. The auction allocates the first slot to *A*, and the second slot to *B*. Under the GSP auction, *A* pays \$ 500, and *B* pays \$ 50 (ignoring the \$ 0.01 minimum bid increment). To determine the VCG payment, we must consider what would have happened if *A* had not been present, and if *B* had not been present. Consider first that if *A* was not present, then *B*'s advertisements would get the first slot and receive 50 more clicks, which *B* values at \$ 5 per click. So *A*'s presence means that *B* loses \$ 250 in advertising value. Similarly for advertiser *C*, who would have been in the second slot with 50 clicks that it values at \$ 1 per click. So *A*'s presence costs *C* \$ 50. Therefore, *A*'s total VCG payment based on the externality imposed on advertisers *B* and *C* would be \$ 250 + \$ 50 = \$ 300. Consider, now, if *B* was not present, then *A*'s position would be unaffected, but *C* would be pushed out of slot 2, losing 50 clicks per hour that it values at \$ 1 per click, for a total of \$ 50. So the externality imposed by *B*'s presence, and thus *B*'s VCG payment is \$ 50.

The basic GSP auction (without quality-based bidding) was empirically designed as a simple extension of the one-unit second-price auction. In a GSP auction, for each click on an advertising position, the bidder pays the next highest bidder's price plus some specified increment. So if a user clicks on the advertisement in the *k*th position, the advertiser is charged the $(k + 1)$st bid. The VCG and GSP auctions would be identical if the search engines only offered one advertisement position. However, when the auction becomes a multi-unit one, with multiple advertisement positions, the two auctions no longer have the same behaviour and properties.

Not surprisingly, the VCG auction has the stronger theoretical pedigree, including truth-telling as an equilibrium dominant strategy. This reduces incentives for strategising and thus makes bidding more straightforward for the advertiser. However, it has practical drawbacks such as low seller revenue, vulnerability to collusion and other fraudulent auction behaviour (Ausubel and Milgrom, 2005). While the GSP auction is known to be non-dominant-strategy-solvable, its price setting mechanism is more transparent and recent research suggests that it may have other properties, such as seller profit maximisation, that are more valuable to search engine companies (Edelman et al., 2007).

Most formal modelling of sponsored search auctions reduces the complexity of the full-blown sponsored search auction environment by focusing on one auction for one keyword. Bidders compete for a vector of slots in an infinitely repeated auction, and equilibrium in the dynamic game occurs if there is a fixed vector of bids such that bidders no longer want to change their bids. Under such simplified conditions, the GSP auction has been shown in theory to have a restricted equilibrium such that advertisers neither want to raise their bid to move up to the next highest slot, nor lower their bid to move to the next lower slot. In other words, the bidder is paying the lowest amount necessary to maintain ranking given the auction rules.

This property was studied independently as *locally envy-free equilibrium* by Edelman et al. (2007), or equivalently, as *symmetric Nash equilibrium* by Varian (2006). This is an ex-poste equilibrium relying on the fact that the sponsored search auctions are essentially a continuously repeated game. The minimum revenue for the GSP equilibrium is at least as good as the revenue from VCG, and its maximum revenue is the same as for Nash equilibrium. Under these assumptions, the GSP auction will always get at least as much revenue as the VCG, and will most likely achieve more. An underlying question is: how do bidders reach this equilibrium? Edelman et al. (2007) show a simple myopic strategy that leads to an envy-free equilibrium with VCG payoffs. In the real world, of course, click through rates can vary by both advertisement and position, so that a bidder with a high overall keyword valuation may prefer to get a slightly lower spot at a low price over getting the top slot for their true valuation, resulting in strategic behaviour (Edelman and Ostrovsky, 2007). Clearly, this is an avenue for future research efforts.

While it makes the problem complexity more manageable, there is another downside to simplifying the modelling of sponsor search as a single auction with a single keyword. As Balcan et al. (2006) has pointed out, if you get more profit from one keyword auction, you may be squeezing it out of the profits for another related keyword auction. Introducing budget constraints can help bring these kinds of tradeoffs back into the model (Abrams, 2006; Borgs et al., 2007; Lahaie, 2006; Szymanski and Lee, 2006). However, introducing budget constraints changes the auction properties. For example, the VCG mechanism is not truthful when there are hard budgets, and indeed it is not possible to design a non-trivial truthful auction that allocates all of its units (Borgs et al., 2005). Budget constraints also mean that the budget must be allocated across both markets and time under conditions of imperfect information (e.g., we do not know the user queries ahead of time). This can be mitigated by algorithms for matching advertisers to keywords that consider an advertiser's daily budget and provide the optimal tradeoff between the current bid and the remaining budget (c.f., Mehta et al., 2007).

Finally, the black box nature of quality-based bidding means that advertisers face a more complex and less transparent market landscape. Search engines control the actual calculation of an advertisement's quality score, so advertisers must simply trust the outcome of search engines rankings and bid management techniques can no longer rely on knowing competitor bid values. Recently, in response to advertiser concerns, Google Adwords released a tool to help make the score less opaque called the Keyword Analysis (2007) page. However, it remains to be seen how much transparency search engine companies will provide.

4.1.3 Payment mechanisms

Advertiser payments can take the form of pay-per-impression, pay-per-click, or pay-per-action. Pay-per-impression, generally priced by Cost Per Mille (CPM), comes directly from traditional print media where the most straightforward measurement available is how many copies of the material consumers bought. Although online pay-per-impression would seem to be inherently more accurate than print, a recent New York Times paper points out

> "far from solving the squishy-numbers problem, the internet seems to have added more confusion. Many advertisers pay web publishers each time their ad gets an impression, meaning that it is viewed by a reader, but each company uses its own methodology to count impressions." (Story, 2007)

However, online advertising's additional measurement capabilities do support additional payment opportunities such as pay-per-click and pay-per-action. Of course, each of these payments types can be converted probabilistically to each other. Search engine bid management companies such as iProspect (www.iprospect.com) allow customers to specify their objectives in terms of maximising revenue, conversions, profits subject to budget constraints with very few customers targeting the objectives of maximising clicks or impressions (Kitts et al., 2005a). Such middlemen provide a valuable service by allowing advertisers to select the payment mechanism that works best with their goals. They must also adapt their techniques and approach to the more opaque quality-based bidding markets.

While multiple payment options may address varied advertiser goals, they also allow different risk-sharing payment models between the search engine and the advertiser. Thus, in the pay-per-click model, the search engine and advertiser share in the overall risk. The search engine assumes the risk that the searcher will not click on a displayed advertisement and the advertiser assumes the risk that a searcher will generate value following a click. In the pay-per-impression model, the search engine bears no risk; it will receive revenue whether the user clicks on the advertisement or not. Finally, in the pay-per-acquisition, the search engine bears all of the risk. The user must not only click on the advertisement, but must also engage in a conversion activity such as product purchase, form completion, white paper download, event registration, or product inquiry. Factors such as whether an advertiser is a large or small company may affect an advertiser's willingness to assume risk, and thus affect the pool of advertisers that a search engine attracts.

The multiple payment mechanisms also provide different levels of protection against fraud. In the case of pay-per-click, click fraud means that the advertiser is taking most of the risk, at least in the short term. Click fraud occurs when an advertisement is clicked with the sole intent of generating a charge to the advertiser. It can be done automatically by computer scripts or directly by humans, and can occur for a variety of reasons, including a competitor's desire to minimise the impact of an ad campaign, simple vandalism, or a desire by a publisher to increase their income. With pay-per-action, the search engine takes on the risk of both fraud and the risk that the advertiser has a product and a marketing approach that will attract customers. Of course, since Google Analytics and other software programs can track conversions, search engines could still rank pay-per-action and pay-per-click advertisements based on estimated revenue. While this could potentially lead to return fraud, where fraudulent buyers click, buy, and then return physical items to deplete an advertiser's budget, this does not seem to be a very cost-effective approach.

4.1.4 Participant behaviour

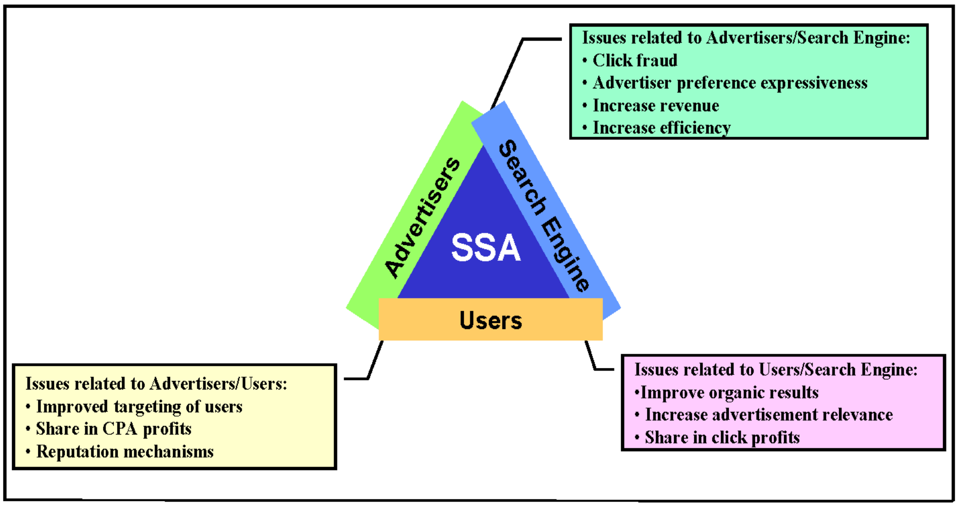
A typical auction has two participant types, namely buyers and sellers. A sponsored search auction involves three participants whose interactions have a significant impact on how well the system serves high-quality, profitable advertisements. Ideally, search engines and advertisers establish risk-sharing incentive mechanisms that allow them to achieve goals such as maximising profit, minimising cost, and user satisfaction. This interaction is illustrated in Figure 6.

Notice that current auction mechanisms do not explicitly consider how the user might play an active role in the system. Two possibilities include incorporating reputation into advertisements as a quality signal or the ability for users to share in the advertising revenue they generate.

4.1.5 Bidder strategies

Advertisers must determine how to allocate advertising funds across campaigns, then across keywords within a campaign, and finally, they must select their individual keyword bidding strategies. Strategies to optimise this landscape can be very complex and difficult for humans to operate, and any change in auction rules may result in a change of bidding strategies on the part of the advertisers. Software bots could assist with implementing these strategies, but search engines generally restrict the bidding information available to the software, and require a review of any automated bidding code. Even restricting bidding strategies to the single keyword strategy space is complex. Keyword markets are fluid and an optimal strategy for one keyword market (e.g., 'coffee') may not be the optimal strategy for another (e.g., 'buy coffee'). Most research has considered single-keyword single-auction bidding strategies. Unless a market is incentive compatible, strategic bidding behaviour can have an effect on market efficiency, stability and profits.

Figure 6 Issues at the intersection of sponsored search players



Although not so relevant now that quality-based bidding dominates, strategic behaviour has been observed empirically in non-quality-based bidding sponsored search auctions (Edelman and Ostrovsky, 2007), including bidding war cycles (Asdemir, 2006) and spiteful bidding behaviour such as gap jamming (Brandt and Weiss, 2002; Ganchev et al., 2006). In gap jamming, bidders raise their bids to a point just below their competitors so that the competitors will pay the maximum amount and more quickly deplete their budget. In return, bidders may protect themselves from this behaviour by increasingly shading their bid under their truthful valuation or investing in bid management companies or software. Proposed solutions to vindictive bidding include using stochastic auctions and contingent-payment auctions (Meek et al., 2005). Stochastic auctions allow all bidders to win with some probability, so that vindictive bidders will periodically be forced to pay a higher amount, thus reducing their incentive. Stochastic auctions also mean that even low bidders can win occasionally, and the search engine can acquire a more accurate click-through rate estimate, in effect trading off optimising immediate revenue for optimising future revenues. Finally, bidders with near equal slot valuations can 'share' the slot (Borgs et al., 2007). In a contingent-payment auction, if a bidder-specific contingency occurs, the winning bidder pays for the item, otherwise it does not. This also makes it more difficult for a spiteful bidder to run down a competitor's budget.

To analyse what are essentially repeated games (i.e., auctions) in game theory, most researchers first fix a bidding strategy for an auction, and then analyse the equilibrium of the strategy. Szymanski and Lee (2006) compare three types of auction mechanisms, namely, GFP, GSP, and VCG, by pairing each auction with possible bidding strategies. Since VCG is incentive-compatible, they pair it with truthful bidding. For GSP, they consider two strategies. The first uses a Locally Envy-Free (GSP-LEF) equilibrium strategy where advertisers only increase (or decrease) their bid if the payoff of swapping with the bidder above (or below) exceeds the marginal cost. They also consider the case where agents use a dynamically adjusting bid strategy, setting current bids based on the bid price that would maximise their utility in the previous round. For the GFP, the agents only use the dynamically adjusting bid strategy. Their results show that the VCG and GSP-LEF generate the most revenue since they avoid the unstable bid patterns of the dynamic adjustment strategy when there is no Return On Investment (ROI) constraint. However, when a minimum ROI threshold is introduced, this results in a decrease of a bidder's maximum bid price, which in turn results in decreased auction revenues. The authors find that the GSP-LEF revenue is most stable in the face of changes in bidders' ROI threshold.

Borgs et al. (2007) fix the advertiser heuristic of equalising ROI across all keywords to maximise utility, then show that existing auction mechanisms show bidding-war price cycling similar to that in Figure 3. They propose a stochastic model to reduce this instability, adding a random amount to the bids before running the auction, and show that GFP provably converges to stable equilibrium prices and that GSP experimentally converges.

Finally, an advertiser's overall ROI optimum may differ from any individual keyword's ROI optimum placement, leading the advertiser to reapportion funds between keyword auctions (Kitts et al., 2005a). One proposed algorithm finds the optimal set of keywords to maximize expected profit subject to a budget constraint assuming that keyword cost is fixed (Rusmevichientong and Williamson, 2006).

4.1.6 Bidding expressivity

Bidding expressivity concerns how to best translate advertiser needs into an appropriate bidding language. For example, a given advertiser may want say that its advertisement can be placed anywhere in slots 1–10, but not below slot 10, or targeted towards a particular user demographic. Driven by customer interactions, bidding management software platforms have been designed to help translate across any expressivity gaps between the sponsored search auctions and the advertiser needs. Of course, the more expressive the bidding language the more complex the auction mechanism or the middleman software must be, so this expressivity comes at both a computational and advertiser 'comprehension cost'. Finding the right tradeoff between the two is key. CombineNet, a company that creates combinatorial auctions for procurement, carefully tailors bidding expressiveness to the stakeholder needs (Sandholm, 2006).

Even-Dar et al. (2007) show how auction mechanisms that employ more expressive context-based bidding generally increase social welfare by allowing advertisers to more narrowly target their audience (by zip code, income, likely task, etc.). Advertisers are also generally willing to pay a premium for more targeted advertisements. However, for some situations the competition is split across contexts such that either the search engine receives less overall revenue or the premium exceeds advertiser's price tolerance. One might also expect that users would value context-based advertisements more than un-contextual advertisements, assuming the targeting methods are sensitive to user privacy. To aid advertisers in finding their best advertising terms, Bartz et al. (2006) developed an algorithm based on logistic regression and collaborative filtering models that predicts terms relevant to a set of seed terms that describe an advertiser's product or service. Singh et al. (2007) find that using mediators can provide bidding management and aggregation services to a specialised portion of advertisers, increasing both search engine profits and market efficiency.

5 Conclusion

Sponsored search has had, and appears set to continue to have, an effect on how people find and use information and services on the web. Software agents can affect an auction system in several ways. Although often viewed solely as an advertising medium, sponsored search is presented in this paper as a special case of information searching. We have reviewed the historical incarnations of sponsored search, leading to the expectation that sponsored search will continue to evolve. The technology powering the sponsored search process is extremely complex, with multiple moving parts. Additionally, there is an aspect of adversarial techniques, as with most web information systems; specifically, click fraud (Jansen, 2007; Kitts et al., 2005b).

Since advertisers must compete in a competitive landscape rather than just a single stand-alone auction, some potential research questions in this area include:

- How can advertisers make sense of this complex environment? Through more expressive bidding languages? Better bid management tools? Software agents that translate between the advertising domain and the sponsored search auction requirements?
- What kinds of auction mechanisms and trading strategies work best in this environment for the search engine and/or for the advertiser? Can restricting the allowable bidding strategies mean that some auction properties, such as incentive compatibility, could be guaranteed for a wider class of mechanisms.
- Another research area addresses what effect paid search has as an interpersonal and mass communication tool.
- Can searchers trust sponsored links? What are searcher perception issues?
- What are the long-term effects of sponsored search on the overall economic market?
- Can sponsored and non-sponsored links be effectively integrated into one listing (Jansen and Spink, 2007)?
- What will be the effect of social networking driven advertisements, such as Facebook advertisements?

In whatever form sponsored search evolves, it is apparent that its effect on the structure of the web, online commerce, and information availability on the web has been, and will continue to be, immense.

References

Abrams, Z. (2006) 'Revenue maximization when bidders have budgets', *Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Miami, Florida.

Asdemir, K. (2006) 'Bidding patterns in search engine auctions', *Second Workshop on Sponsored Search Auctions*, Ann Arbor, MI.

Ausubel, L.M. and Milgrom, P. (2005) 'The lovely but lonely vickrey auction', in Crampton, P., Steinberg, R. and Shoham, Y. (Eds.): *Combinatorial Auctions*, MIT Press, Cambridge, MA, pp.17–40.

Balcan, M-F., Chen, J., Devanur, N. and Kumar, A. (2006) *Models for Sponsored Search: What are the Right Questions?* (transcript of panel discussion), The Second Workshop on Sponsored Search Auctions.

Bartz, K., Murthi, V. and Sebastian, S. (2006) Logistic regression and collaborative filtering for sponsored search term recommendation, *Second Workshop on Sponsored Search Auctions*, Ann Arbor, MI.

Battelle, J. (2005) *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, New York, Penguin Group.

Baye, M.R. and Morgan, J. (2001) 'Information gatekeepers on the internet and the competitiveness of homogeneous product markets', *American Economic Review*, Vol. 91, pp.454–474.

Bhargava, H.K. and Feng, J. (2002) 'Paid placement strategies for internet search engines', *11th International Conference on World Wide Web*, ACM Press, Honolulu, Hawaii, USA.

Borgs, C., Chayes, J., Etesami, O., Immorlica, N., Jain, K. and Mahdian, M. (2007) 'Dynamics of bid optimization in online advertisement auctions', *16th International World Wide Web Conference (WWW2007)*, Banff, Alberta, Canada.

Borgs, C., Chayes, J., Immorlica, N., Mahdian, M. and Saberi, A. (2005) 'Multi-unit auctions with budget-constrained bidders', *6th ACM Conference on Electronic Commerce*, Vancouver, BC, Canada.

Brandt, F. and Weiss, G. (2002) 'Antisocial agents and Vickrey auctions', in Meyer, J-J.C. and Tambe, M. (Eds.): *Intelligent Agents VIII*, Springer-Verlag, Berlin/Heidelberg, Vol. 2333, pp.335–347.

Edelman, B. and Ostrovsky, M. (2007) 'Strategic bidder behavior in sponsored search auctions', *Decision Support Systems*, Vol. 43, pp.192–198.

Edelman, B., Ostrovsky, M. and Schwarz, M. (2007) 'Internet advertising and the generalized second price auction: selling billions of dollars worth of keywords', *American Economic Review*, Vol. 97, pp.242–259.

Even-Dar, E., Kearns, M. and Wortman, J. (2007) 'Sponsored search with contexts', *Third Workshop on Sponsored Search Auctions, 16th International World Wide Web Conference (WWW2007)*, Banff, Canada.

Fain, D.C. and Pedersen, J.O. (2006) 'Sponsored search: a brief history', *Bulletin of the American Society for Information Science and Technology*, Vol. 32**,** pp.12, 13.

Fallows, D. (2005) *Search Engine Users*, Pew Internet and American Life Project.

Feng, J., Bhargava, H.K. and Pennock, D.M. (2007) 'Implementing sponsored search in web search engines: computational evaluation of alternative mechanisms', *Informs Journal on Computing*, Vol. 19, pp.137–148.

Ganchev, K., Kulesza, A., Tan, J., Gabbard, R., Liu, Q. and Kearns, M. (2006) 'Empirical price modeling for sponsored search', *Third Workshop on Sponsored Search Auctions*, Banff, Canada.

Hansen, E. (2002) *FTC Wants Paid Search to Shape Up*, CNET news.com.

Introna, L. and Nissenbaum, H. (2000) 'Defining the web: the politics of search engines', *IEEE Computer*, Vol. 33, pp.54–62.

Jansen, B. (2006) 'Paid search', *IEEE Computer*, Vol. 39, pp.88–90.

Jansen, B.J. and Molina, P. (2006) 'The effectiveness of web search engines for retrieving relevant ecommerce links', *Information Processing and Management*, Vol. 42, pp.1075–1098.

Jansen, B.J. (2007) 'Click fraud', *IEEE Computer*, Vol. 40, pp.85, 86.

Jansen, B.J. and Resnick, M. (2006) '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.

Jansen, B.J. and Spink, A. (2007) 'Sponsored search: is money a motivator for providing relevant results?', *IEEE Computer*, Vol. 40, No. 8, pp.50–55.

Keyword Analysis (2007) https://adwords.google.com/support/bin/answer.py?answer=76846, available 1 November.

Kitts, B., Laxminarayan, P., Leblanc, B.J. and Meech, R. (2005a) 'A formal analysis of search auctions including predictions on click fraud and bidding tactics', *1st Workshop on Sponsored Search Auctions*, Vancouver, BC, Canada.

Kitts, B., Leblanc, B., Meech, R. and Laxminarayan, P. (2005b) 'Click fraud', *Bulletin of the American Society for Information Science and Technology*, Vol. 32, pp.14–16.

Lahaie, S. (2006) 'An analysis of alternative slot auction designs for sponsored search', *The Seventh ACM Conference on Electronic Commerce*, Ann Arbor, MI.

Marable, L. (2003) *False Oracles: Consumer Reaction to Learning the Truth About How Search Engines Work, Results of An Ethnographic Study*, Consumer WebWatch.

Mcafee, R.P. and Mcmillan, J. (1987) 'Auctions and bidding', *Journal of Economic Literature*, Vol. 25, pp.699–738.

Meek, C., Chickering, D.M. and Wilson, D.B. (2005) 'Stochastic and contingent-payment auctions', *1st Workshop on Sponsored Search Auctions*, Vancouver, BC, Canada.

Mehta, A., Saberi, A., Vazirani, U. and Vazirani, V. (2007) 'AdWords and generalized online matching', *J. ACM*, Vol. 54, No. 5, October, p.22.

Quality Index (2007) http://help.yahoo.com/l/us/yahoo/ysm/sps/start/overview_qualityindex.html, available 30 March.

Roth, A.E. (2002) 'The economist as engineer: game theory, experimentation, and computation as tools for design economics', *Econometrica*, Vol. 70, pp.1341–1378.

Rusmevichientong, P. and Williamson, D.P. (2006) 'An adaptive algorithm for selecting profitable keywords for search-based advertising services', *7th ACM Conference on Electronic Commerce*, Ann Arbor, ACM Press, Michigan, USA.

Sandholm, T. (2006) 'Expressive commerce and its application to sourcing', *Eighteenth Conference on Innovative Applications of Artificial Intelligence*, Boston, Massachusetts.

Singh, S.K., Roychowdhury, V.P., Gunadhi, H. and Rezaei, B.A. (2007) 'Capacity constraints and the inevitability of mediators in adword auctions', *Internet and Network Economics, Third International Workshop (WINE 2007)*, San Diego, CA, pp.318–332.

Story, L. (2007) 'How many site hits? Depends who's counting', *New York Times*, Late Edition (East Coast) Ed., New York Times, New York City.

Szymanski, B.K. and Lee, J-S. (2006) 'Impact of ROI on bidding and revenue in sponsored search advertisement auctions', *Second Workshop on Sponsored Search Auctions*, Ann Arbor, MI.

Varian, H.R. (1995) 'Mechanism design for computerized agents', *USENIX Workshop on Electronic Commerce*, New York.

Varian, H.R. (2006) *Position Auctions*, UC Berkeley.

Vickrey, W. (1961) 'Counterspeculation, auctions, and competitive sealed tenders', *Journal of Finance*, Vol. 16, pp.8–27.

Wilson, T.D. (2000) 'Human information behavior', *Informing Science*, Vol. 3, pp.49–55.