ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

MASTER OF SCIENCE PROGRAM (MSc)
In INFORMATION SYSTEMS

MASTER THESIS

"Data Analysis and Bidding Strategies in Sponsored Search Auctions"

Sotiri Maria

Registration Number: M3090027

ATHENS, FEBRUARY 2011

MASTER OF SCIENCE PROGRAM (MSc)
In INFORMATION SYSTEMS

MASTER THESIS

"Data Analysis and Bidding Strategies in Sponsored Search Auctions"

Sotiri Maria

Registration Number: M3090027

Supervising Professors:
Markakis Evangelos, Stamoulis D. George

ATHENS UNIVERSITY OF ECONOMICS AND BUSSINESS
DEPARTMENT OF INFORMATICS

ATHENS, FEBRUARY 2011

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ανάλυση Δεδομένων και Στρατηγικές Υποβολής Προσφορών σε Δημοπρασίες με Λέξεις Κλειδιά»

Σωτήρη Μαρία
Αριθμός Μητρώου: Μ3090027

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2011

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ανάλυση Δεδομένων και Στρατηγικές Υποβολής Προσφορών σε Δημοπρασίες με Λέξεις Κλειδιά»

Σωτήρη Μαρία

Αριθμός Μητρώου: M3090027

Επιβλέποντες Καθηγητές:
Μαρκάκης Ευάγγελος & Σταμούλης Δ. Γεώργιος

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2011

Preface - Abstract

Sponsored search advertising has evolved into a multi-billion industry and is the primary source of income for the search engine providers, with lion's share for Google and significant market shares for Yahoo! and Microsoft's Bing.

The majority of search engines use auctions in order to sell advertising slots, i.e., "space" in the search result page, to advertisers. The primarily used auction mechanism is the Generalized Second Price auction, which typically generates more revenue for the search engines than most other known mechanisms, (such as the Vickrey-Clark-Groves mechanism). In parallel, GSP encourages strategic bidding; each advertiser does not benefit from reporting truthfully the expected profit from a potential customer as a bid in the auction. As a result, several strategies have been proposed in order to maximize the expected net benefit of the advertiser or the return of the investment, and this is a topic that is worth further investigation.

In this thesis we provide an overview of the sponsored search auction market, its structure and we examine its mechanisms over the years, since the first sponsored search auction till nowadays and present their characteristics, their similarities and differences.

We perform an analysis of Overture's paid placement auctions between June 2002 and June 2003, through a dataset provided as part of the Yahoo! Webscope program for use for academic research purposes. We look up present research concerning the Click Through Rate estimation and try different distributions in calculating its reduction as we move down on the list of sponsored search results. Our findings indicate that the geometric distribution may fit real data very well in some cases, while in other cases smaller or larger deviations may occur.

Finally we map the principal bidding strategies and by means of programming we introduce a new one, and we perform simulations to elaborate its performance.

Πρόλογος

Οι δημοπρασίες με λέξεις κλειδιά έχουν εξελιχθεί σε μια βιομηχανία πολλών δισεκατομμυρίων και αποτελούν την κύρια πηγή εσόδων για τις μηχανές αναζήτησης, με τη μερίδα του λέοντος να την κατέχει η Google, ενώ σημαντικά μερίδια αγοράς να κατέχουν οι Yahoo και Bing (μηχανή αναζήτησης της Microsoft).

Η πλειοψηφία των μηχανών αναζήτησης προσφεύγουν σε δημοπρασίες με σκοπό την πώληση στους διαφημιστές χώρου διαφήμισης, δηλαδή διαφημιστικών περιοχών στην σελίδα των αποτελεσμάτων αναζήτησης. Ο μηχανισμός που χρησιμοποιείται κυρίως για τις δημοπρασία με λέξεις κλειδιά είναι η Γενικευμένη Δημοπρασίας Δεύτερης Τιμής, καθώς, κατά κανόνα, δημιουργεί περισσότερα έσοδα για τις μηχανές αναζήτησης από την πλειοψηφία των μέχρι σήμερα γνωστών μηχανισμών. Παράλληλα, η Γενικευμένη Δημοπρασίας Δεύτερης Τιμής ενθαρρύνει τη στρατηγική υποβολής προσφορών: κάθε διαφημιστής δεν επωφελείται αν υποβάλει ειλικρινή προσφορά σχετικά με το αναμενόμενό του κέρδος. Ως αποτέλεσμα, κατά καιρούς έχουν προταθεί διάφορες στρατηγικές προκειμένου να μεγιστοποιηθεί το αναμενόμενο καθαρό όφελος του διαφημιζόμενου ή η απόσβεση της επένδυσης, κάτι το οποίο αποτελεί θέμα το οποίο χρήζει περαιτέρω έρευνας.

Στην παρούσα διπλωματική εργασία παρέχουμε μια επισκόπηση των δημοπρασιών με λέξεις κλειδιά και της δομής τους, να εξετάσουμε τους μηχανισμούς που χρησιμοποιούνται στο πέρας του χρόνου, από την πρώτη δημοπρασία αυτής της μορφής μέχρι σήμερα, και να παρουσιάσουμε τα χαρακτηριστικά τους, τις ομοιότητες και τις διαφορές τους.

Επιπλέον αναλύουμε δεδομένα από δημοπρασίες με λέξεις κλειδιά της Overture, που πραγματοποιήθηκαν τη χρονική περίοδο μεταξύ Ιουνίου 2002 και Ιουνίου 2003, μέσα από ένα σύνολο στοιχείων που παρέχονται ως μέρος του προγράμματος Yahoo! Webscope για χρήση για ακαδημαϊκούς. Ανατρέχουμε σε δημοσιευμένες έρευνες σχετικά με τα ίδια δεδομένα και ειδικότερα με ποσότητες σχετικές με τον υπολογισμό του ρυθμού κλικ (Click Through Rate estimation) και δοκιμάζουμε διαφορετικές φθίνουσες κατανομές καθώς προχωράμε προς τα κάτω στον κατάλογο των χρηματοδοτούμενων αποτελεσμάτων αναζήτησης. Επιβεβαιώνουμε ότι πραγματικά δεδομένα ικανοποιούν σε μεγάλο βαθμό τη γεωμετρική κατανομή σε πλήθος περιπτώσεων.

Τέλος, εκθέτουμε τις υπάρχουσες στρατηγικές υποβολής προσφορών και εισάγουμε και δοκιμάζουμε μία νέα σε προγραμματιστικό επίπεδο. Στο πλαίσιο αυτό εκτελούμε πειράματα/ προσομοιώσεις με στόχο την αξιολόγηση της απόδοσής τους.

Acknowledgements

At this point I would like to thank my professors and supervisors Evangelos Markakis and George D. Stamoulis for their guidance, patience, support and fundamental remarks throughout this period of time. Last but not least, I would like to thank my parents, my brother and my friends for their continuous support, deep understanding, apposite remarks and valuable encouragement.

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τους καθηγητές και επιβλέποπτές μου Ευάγγελο Μαρκάκη και Γεώργιο Δ. Σταμούλη για την καθοδήγηση, την υπομονή τους, την υποστήριξή τους και τις θεμελιώδεις παρατηρήσεις τους όλο αυτό το χρονικό διάστημα. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου, τον αδελφό μου και τους φίλους και φίλες μου για τη συνεχή υποστήριξή τους, τη βαθιά κατανόησή τους, τις εύστοχες παρατηρήσεις τους και την πολύτιμη ενθάρρυνσή τους.

Contents

1

Introduction to Sponsored Search Advertising

History of Online Advertising

Online advertising is a form of promotion that uses the Internet and World Wide Web for the expressed purpose of delivering marketing messages to attract customers. Examples of online advertising include contextual ads on search engine results pages, banner ads, Rich Media Ads, Social network advertising, online classified advertising, advertising networks and e-mail marketing, including e-mail spam.

There are plenty of benefits regarding online advertising: the immediate publishing of information and content is not limited by place or time. Moreover, online advertising allows for the customization of advertisements, including content and posted websites, offering efficiency of advertiser's investment. For example, AdWords, Yahoo! Search Marketing and Google AdSense enable ads to be shown on relevant web pages or alongside search results of related keywords (Wikipedia "Online Advertising").

This Early Web advertising was sold on the basis of “impressions,” by analogy with the print ads one sees in newspapers or magazines: a company like Yahoo! would negotiate a rate with an advertiser, agreeing on a price for showing its ad a fixed number of times. Thus the only form of online advertisement that was available was in the form of banner ads. However it was soon realized that if the ad shown to a user wasn’t tied in some intrinsic way to his behavior, then one of the main benefits of the Internet as an advertising venue, compared to print or TV is lost. After all, search engine queries are a potent way to get users to express their intent - what it is that they’re interested in at the moment they issue their query - and an ad that is based on the query is catching a user at precisely this receptive moment. (Easley D., Kleinberg J. , 2010)

Web advertising has grown and changed considerably since its beginnings in the early 1990s. Online advertisement offers customers a more interactive way to shop and buy, while at the same time, they give sellers a more layered method of reaching the public. Looking at the history of Web advertising shows us how far the medium has come, and where it is headed.

In 1989 Tim Berners-Lee invented the World Wide Web. However, it was not until 1994, when the first Web advertising appears, after the creation of Web browsers, starting with Netscape, which allowed more complex Internet sites. In 1995, large corporations and brands, such AT&T, Saturn, Time and Proctor & Gamble, begin to invest in both their own websites and online advertising. 1997 can be seen as the year when the Internet gained broad commercial acceptance as a sales medium. Companies flocked to the Web to create e-commerce sites, and content sites (usually online versions of newspapers or magazines), offering advertising space. As advertising began to saturate the Web, marketers tried new tactics to gain attention. By the end of 1997, pop-up and pop-under ads became common. In 1999 online spending reached nearly $1 billion in the second quarter. The "dot-com crash" in 2001 leaded to a sharp decline in online advertising spending. However, Google's AdWords program, introduced in 2002, gave a new boost to the text-based advertising. A decade after the first online ads, in 2004, spending on Web advertising reached nearly $9.6 billion. Nowadays, major corporations move their advertising efforts into new areas by making wide use of social media such as Twitter and Facebook (Beest).

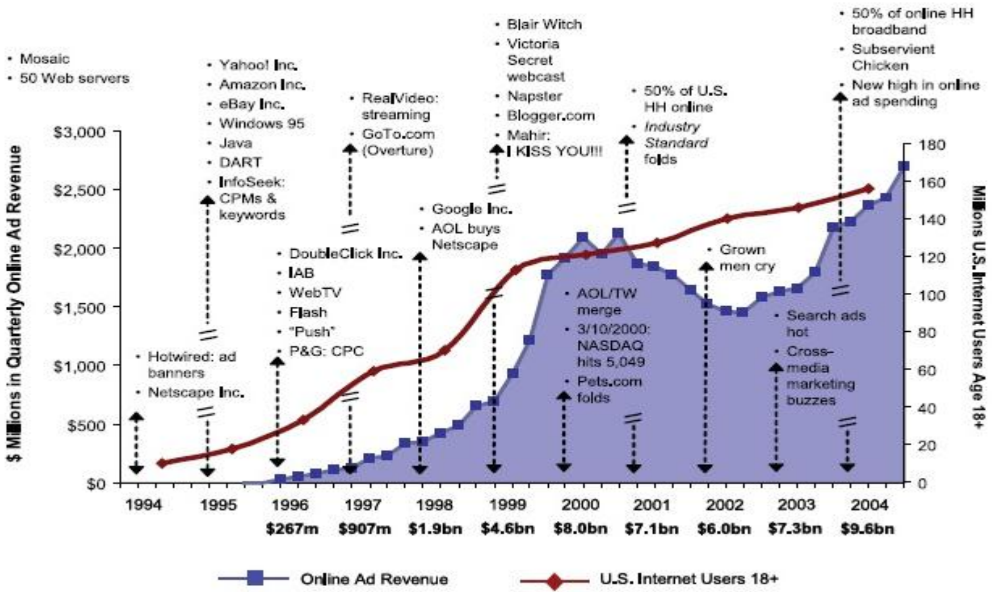

Figure 1: Brief History of Online Advertising until 2004
Source: Rick E. Bruner et al., "The decade in Online Advertising 1994-2004", April 2005, www.doubleclick.com

Sponsored Search Advertising Characteristics

In the last decade, the most popular advertising method is via sponsored search auctions, which has become a particularly profitable market for search engines. The idea behind sponsored search is that for queries with commercial interest, large search engines, such as Google, Yahoo!, Bing (formerly known as Live Search, Windows Live Search, and MSN Search) and others, allow a certain number of ads to be displayed on the top or on the side of the unpaid search (organic or algorithmic) results, although this is not the same for all search engines. The returned set of links, called sponsored links, are typically displayed in sets above the organic or algorithmic results (mainline slots), in sets besides the organic results (sidebar slots). The main advantage of such ads is that an advertiser is displaying his ad to users who have expressed interest for the specific query and therefore are more likely to be interested in his product (Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B., 2010).

The way that the advertiser is charged from the search engine depends on the charging scheme the later has selected. The most popular charging schemes are the following three, although the most popular model that is being used in almost all sponsored search auctions is the *Pay per Click* model.

Cost per Click (CPC): This scheme is also known as *Pay per Click* (PPC). Advertisers pay every time a user clicks on their sponsored link, which means that the user is redirected to their website. This scheme enables the advertisers to refine searches and gain only direct relevant traffic to their websites. Therefore the search engine is motivated to display advertisements which have a high click probability as such ads will generate more revenue for the search engines. Among PPC providers, Google AdWords, Yahoo! Search Marketing, and Microsoft adCenter are the three largest network operators, and all three operate under a bid-based model. Cost per click (CPC) varies depending on the search engine and the level of competition for a particular keyword.

Cost per Impression (CPI): This scheme is also known as *Cost per Mille* (CPM). It is used for measuring the worth and cost of a specific e-marketing campaign. This charging scheme is best suited for contextual ads – ads which appear in websites with high traffic that contains either relevant content or content of general interest.

Cost per Transaction (CPT): This scheme is also known as *Cost per Action* (CPA). Advertisers typically pay only when a user fulfills a transaction (such as a purchase or a form submission) and it is quite common in the affiliate marketing sector of online advertising (Drosos D., Markakis E., Stamoulis G. D., 2010).

Due to the critical influence of search engines on Web users' actions, many firms have realized the importance of gaining a high position on the search results for specific queries. Entire niche industries exist touting services to boost a Web page's ranking on the popular search engines, in part by reverse engineering the search engines' information retrieval algorithms. The expectation of increased traffic from good placement on a search page has led to the creation of a market for sponsored search (or paid placement—we use the terms interchangeably, as dictated by context) where search engines can charge a fee for prominent positioning within a "sponsored" section in the results page. For example, a digital camera retailer may pursue to participate in the auction for the keyword "digital cameras" (Feng J., Bhargava H. K., Pennock D. M., 2007).

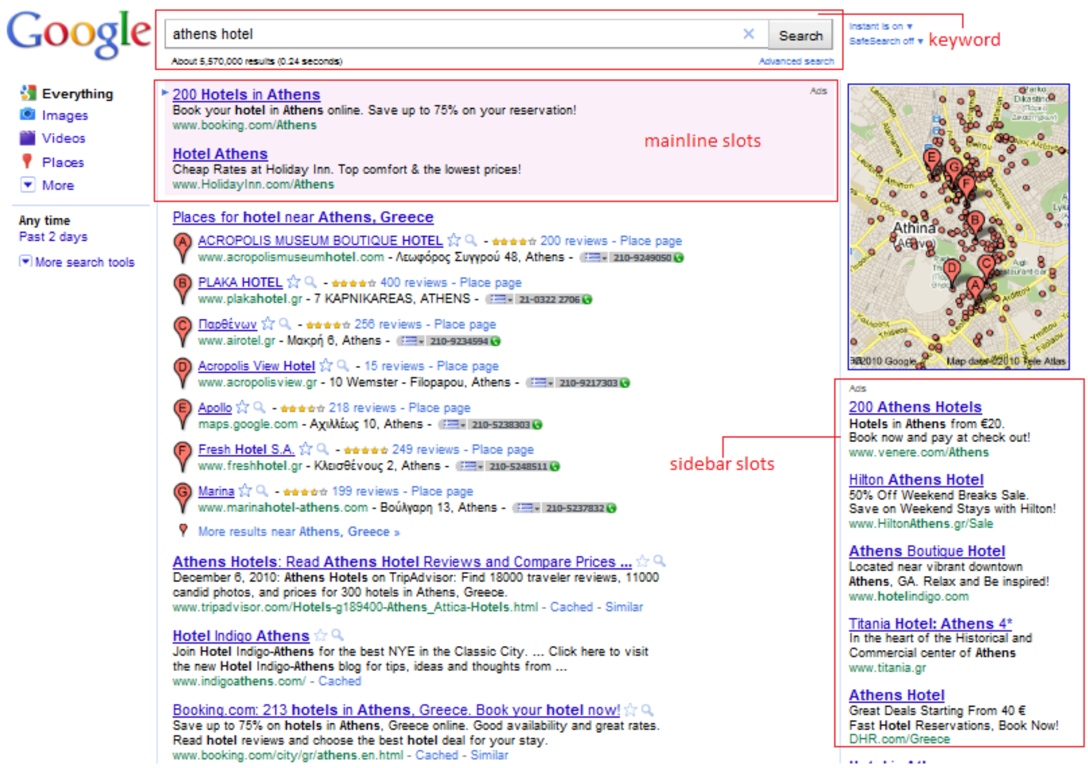

Figure 2: Mainline and sidebar slot results in Google AdWords

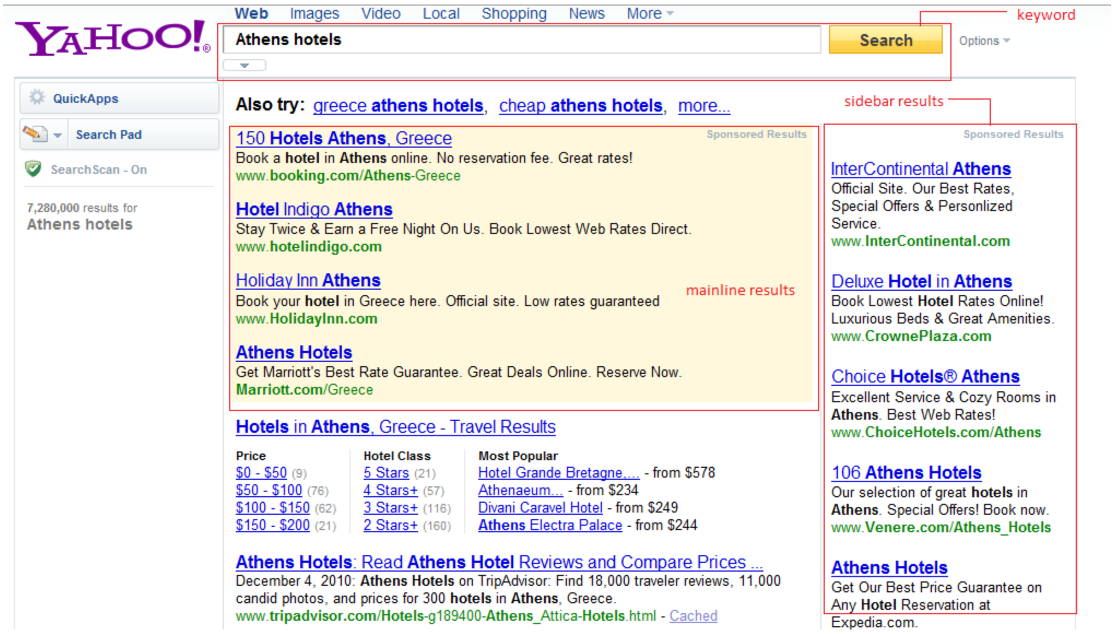

Figure 3: Mainline and sidebar slot results in Yahoo!

Nowadays there are over a dozen search engines, the dominant ones, Google, Yahoo!, Bing and Baidu, which primarily searches the Chinese Web. According to Hitbox (HitBox) Google is the most popular search engine worldwide with share that reaches 84,72%, while Yahoo!, Baidu and Bing follow with 6,42%, 3.76% and 3,14% respectively. On July 29, 2009 (Microsoft and Yahoo seal web deal, 2009), Microsoft and Yahoo! announced that they had made a 10-year deal in which the Yahoo! search engine would be replaced by Bing. Therefore Bing's global share is 9.57% when considering that searches at both Yahoo and Bing are actually powered by the Bing search engine.

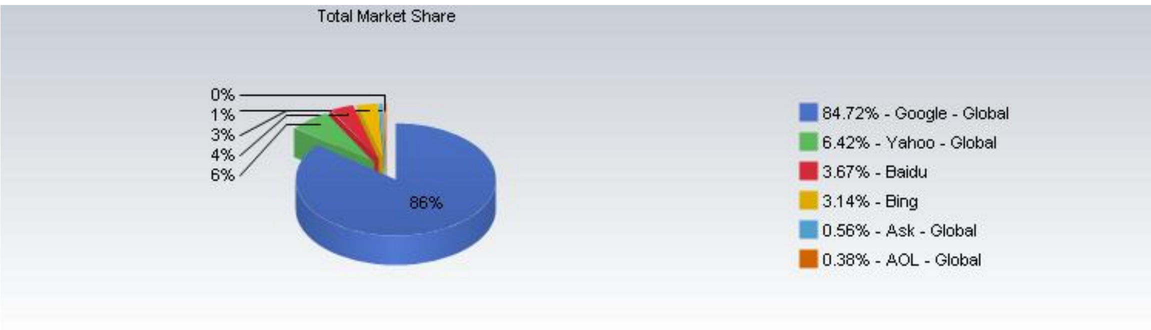

Figure 4: Search engine market shares in November 2010

Sponsored search has evolved into a multi-billion industry and is the primary source of income for the search engine providers. As stated in SEMPO State of Search Engine Marketing Report 2010 (SEMPO, 2010) Google dominates, with 97% of companies paying to advertise on Google AdWords. Fifty-six percent of advertisers and 62% of agencies say that keywords have become more expensive in Google over the last year, while only 32% of advertisers observed an increase in Yahoo, and 29 % in Bing. Finally, even though 2009 was a slow year overall due to the poor economic climate, overall the market is estimated to grow by 14% in 2010, and reach a value of $16.6 billion. Therefore we can easily understand the power of search engines in terms of financial growth and strength and the significance of sponsored search auctions industry to the global economy.

Sponsored Search Market Features

A combination of features makes the market for Internet advertising unique and differentiates it from other traditional centralized markets, such as this of electromagnetic spectrum. First, bids can be changed at any time, leading, each time, to different appearing results. An advertiser's bid for a particular keyword will apply every time that the keyword is entered by a search engine user, until the advertiser changes or withdraws the bid. For example, the advertiser with the second highest bid on a given keyword at some instant will be shown as the second sponsored link to a user searching for that keyword at that instant. The order of the ads may be different next time a user searches for that keyword, because the bids could have changed in the meantime. Specialized software which is responsible for monitoring bid campaigns in an automated way is widely used (Edelman B., Ostrovsky M., Schwarz M., 2007).

Secondly, search engines effectively sell flows of perishable advertising services rather than storable objects: if there are no ads for a particular search term during some period of time, there are no ads displayed.

Third, the utilities of the bidders are affected by special parameters that exist only in the context of sponsored search auctions, most notably the click-through-rate, the quality factor of the advertiser and the convention rate. Those parameters are influenced by the behavior of the whole population of users who conduct a search. Therefore the values of these parameters are likely to vary between industries, geographical regions, the current date or time and the advertisers that participate. Convention rates, for instance, vary vastly between different times of the day.

Finally, unlike other centralized markets, it is usually quite unclear how to measure what is being sold, as there is no “unit” of Internet advertisement that is natural from the points of view of all involved parties. From the advertiser’s perspective, the relevant unit can be defined as the cost of attracting a customer so that the latter makes a purchase (corresponding to pay-per-transaction scheme). From the search engine’s perspective, the relevant unit is the revenues assembled every time a user performs a search for a particular keyword (reflecting the pay-per-view scheme). However, nowadays, the prevailing scheme is the pay-per-click scheme and the advertiser is charged accordingly.

2

An Overview of Sponsored Search Auctions

Introduction to Sponsored Search Auctions

A sponsored search auction, frequently referred as keyword auction or position auction, is an economically efficient mechanism of allocating a limited number of advertisement slots to advertisers in a dynamic way. The characterization "dynamic" suggests that advertisers can change their bid at any time, and a new auction clears every time a user enters the search query. In this way, advertisers can adapt to the changing environment of internet, boosting or withdrawing their products according to demand.

In this chapter we provide mathematical formulations for the underlying mechanisms, and we introduce key concepts related to sponsored search auctions, such as the notion of Click Through Rate and payoff equation. In addition, we will review the available ranking and pricing methods used in sponsored search auction. We will present and analyze the most prevailing search auction mechanisms, their evolution throughout years and alternative mechanisms. Finally we will present an analysis of equilibria in Generalized Second Price (GSP) auctions. All these will offer an overview to sponsored search advertising.

Sponsored Search Auction Terms and Conditions

In every sponsored search auction there is a set of n potential advertisers $N = \{1, \ldots, n\}$ who compete for a set K of k slots, i.e. $K = \{1, \ldots, k\}$, where slot 1 indicates the slot on the top of the list and slot k is the last slot of the list . Typically we have that $n \geq k$ and $k \geq 2$. We will consider only the sidebar slots, ignoring the slots on the top of the organic results, and that the *pay per click* scheme is employed. Each advertiser i is asked to submit a bid b_i that expresses the maximum price he is willing to pay per click. The actual valuation of advertiser i will be denoted by v_i and can differ from b_i.

The payoff to bidder i for being in position j if the price od slot j is p_j, also known as payoff equation, is:

$$CTR_j^i\left(u_i - p_j\right) + \omega_i^j \tag{1}$$

where, $CTR_j^i > 0$, is the *Click Through Rate* that bidder i anticipates if he is in position k. This is the total number of clicks bidder i will receive in the time period for which the positioning resulting from the auction is valid. This can be analyzed into a product of two components: a component which denotes the click rate of an advertiser i.e., the probability that someone will click on an ad by an advertiser i, and a component which represents the slot click rate (Blumrosen L., Hartline J. D., Nong S., 2008). The advertiser click rate is commonly referred as quality factor q_i of the advertiser and indicates the click rate that the advertiser receives independently of the slot and can be used to model the relevance of the advertisement to the query keyword as well as its quality. Finally, $\omega_i^j \geq 0$ is the impression value of being in position j for bidder i. The impression value describes the value that bidder i derives from merely being seen in position$j$, independent of whether a search engine user clicks on bidder i's link. We have in mind that companies derive value from the fact that a sponsored search link reminds customers of the existence of their company, and that it makes users more likely to buy in the future, even if those users do not click on the link and make a purchase at the time of their search. The impression value is thus similar to the value that advertisers derive from other forms of advertising, such as television advertising, that are less targeted than sponsored search advertising. Since our focus is on the payoff derived from clicks, in the remainder of this thesis, we will assume that $\omega_i^j = 0$.

A restrictive assumption implicit in equation (1) is that click rate, value per click, and impression value for bidder i in position k do not depend on the identity on the bidders that win other positions. In practice, this identity might matter. Bidder i might attract a larger click rate in second place if the bidder in the top position is a large, widely known company than if the bidder in the top position is small and not well-known. In auction theory, this is known as an “allocative externality”. It is well-known that such externalities may create multiple equilibria in single unit auctions (Jehiel P., Moldovanu B., 2005).

Equation (1) assumes that bidders know click rates, quality factors, values per click, and impression values. However, in case bidders are uncertain about these variables, all three variables can be replaced by their expected value, equation.

Ranking Mechanisms for Sponsored Search

One of the major questions concerning an auction, is the way search engines design and conduct them (Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B., 2010). The first thing needed is a rule that ranks the bidders and thus determines the allocations of the advertisers' ads on the available slots. In particular this rule computes for each bidder a score, which is a function of his bid and possibly of other parameters (most notably the Click-Through-Rate (CTR)), and ranks bidders in decreasing order. The CTR of an ad is the probability with which a user will click on the ad and as mentioned in the previous sections it can be affected by the ad itself (due to the content of the text being displayed and/or the identity of the advertiser), by the slot that the ad is occupying (higher slots typically receive more clicks) and by several other factors, such as the presence of other competing advertisers or the nature of the advertiser himself (brand loyalty). The ranking rule is complemented by a pricing rule determining the amount that a bidder being allocated a certain slot for his ad will ultimately have to pay upon receiving a click. This allocation depends on the willingness of each advertiser to appear in the top positions of a search page as it is reported through the bid they submit to the search engine. The definition of an auction mechanism is therefore the joint choice of a ranking rule and a pricing rule.

Depending on the ranking rule selected, there are two alternative policies:

- Rank-by-bid: the k higher bidders win the k slots in the order of their bids. If the order is $b_1 > b_2 > \cdots > b_n$ then each winner of a slot $i \leq k$ is charged per click a price $b_{i+1} + \varepsilon$, where the exact value of ε is provided by every search engine for every currency and it usually equals $\varepsilon = 0{,}01$.

- Rank-by-revenue rule, each bid b_i is weighted by a quality score w_i of bidder i, which reflects the probability that a user will click on the ad of bidder i.

The rationale is that ranking only by the bid may lead to displaying ads with very low probability of attracting clicks and therefore lowering the total revenue of the search engine. On the contrary the rank-by-revenue rule takes into account the expected revenue from each bidder. After sorting the advertisers by the product $w_j b_j$, the k highest advertisers get the k slots accordingly, and each of them pays again the amount that is necessary to bid in order to keep his current position.

Historically, Yahoo and Google used different rankings of candidate advertisements. Up until 2007, Yahoo ranked candidates according to the rank by bid rule. Google ranked candidates using the rank by revenue rule. Currently, Yahoo, Google, Bing, and most other web search engines use value per impression as the primary factor to rank the candidate advertisements (Linden G., Meek C., Chickering M., 2006).

Pricing Mechanisms for Sponsored Search

In this section we provide an overview of the second component of adword auctions, which is the pricing rule. The positive correlation which exists between top placement and increased traffic on search engines creates a significant demand among businesses for top placement when a query is conducted, especially as far as popular and commercially-relevant search terms are concerned. However, since Web users face negative utility if the search engine becomes impartial, most search engines limit the number of paid placement requests they accept. Thus, the sponsored slots are a scarce resource that need to be allocated carefully (Feng J., Bhargava H. K., Pennock D. M., 2007).

The history of sponsored search auctions (Edelman B., Ostrovsky M., Schwarz M., 2007) is of interest from the perspective of whether, how, and how quickly markets come to address their structural shortcomings. Notably, the Internet advertising market evolved much faster than other markets. This may be an outcome of on the one hand the competitive pressures on mechanism designers present, and on the other hand the much lower costs of entry and experimentation, as well as the advances in the understanding of market mechanisms, and improved technology.

The early internet advertising, back in 1994, was notably different from the current as advertisements were largely sold on a per-impression basis (Cost-per-Impression or/and Cost-per-Mille schemes). Contracts were negotiated on a case-by-case basis resulting to large minimum contracts for advertising purchases (typically, a few thousand dollars per month), and slow market entry. As for the pay-click charging schemes, there are essentially two pricing rules that have been used, the Generalized First Price and Second Price rules.

The Generalized First Price Auction

Understanding the potential of internet advertising, Overture (later part of Yahoo!) introduced in 1997 a completely new model of selling Internet advertising. In the original Overture auction design, each advertiser submitted a bid denoting the advertiser's willingness to pay on a per-click basis, for a specific keyword. In this way, the advertisers had targeted ads instead of banner ads that would be shown to everyone visiting a Web site, and they could therefore specify which keywords were relevant to their products and how much each of those keywords (or, more precisely, a user clicking on their ad after looking for that keyword) was worth to them. Another innovating element was that ads were no longer sold per 1,000 impressions but per one click at a time, which meant that every time a consumer clicked on a sponsored link, the corresponding advertiser's account was automatically billed the amount of his last bid. This mechanism became widely known as *Generalized First-price Auction* mechanism, as for each particular keyword he has. The sponsored links were arranged in descending order of bids, making highest bids the most prominent. The ease of use, the very low entry costs, and the transparency of the mechanism quickly led to the success of Overture's paid search platform, and it was quickly adopted by major search engines, such as Yahoo! and MSN.

However, the Generalized First-Price auction mechanism was proven far from perfect. Both search engines and advertisers quickly realized that the mechanism was unstable due to a series of reasons. The fact that bids could be changed very frequently was responsible for the existence of bidding cycles. Suppose there are two slots on a page and three advertisers. An ad displayed in the first slot receives 200 clicks per hour, while the second slot gets 100. Advertisers 1, 2, and 3 have valuations of $10, $4, and $2, respectively. Given that advertiser 2 bids $2.01, he gets a slot. Then advertiser 1 will not want to bid more than $2.02, as he does not need to pay more than that to get the top spot. But then advertiser 2 will want to revise his bid to $2.03 to get the top spot, advertiser 1 will in turn raise his bid to $2.04, and so on. However, when the bid reaches \$4, then advertiser 2 will return to \$2.01 in order to maintain his slot but at the same time minimize his cost. Respectively advertiser 1 will revise his bid to \$2.02 and the bidding cycle starts over. Clearly, there is no pure strategy equilibrium in the one-shot version of the game, and so if advertisers best respond to each other, they will want to revise their bids as often as possible. Moreover, some Overture advertisers apparently used an “auto-bid” system, in order to automatically adjust an advertiser’s bid to achieve desired placement and to avoid overbidding. When two or more advertisers activated auto-bidders, their bids tended to form a distinctive “saw-tooth” pattern of gradual rises in price followed by sudden drops, as seen in figure 5. Clearly, this saw-tooth pattern reduces market efficiency: the bidder who values the first spot the most spends only half the time at the top, and even less if there are more than two bidders competing for the top spot. This indicates that the use of this particular bidding pattern could have substantially reduced Overture’s revenue (Edelman B., Ostrovsky M., 2007).

Another weakness observed in generalized first price auctions was that bidding one's true value is a not a dominant strategy; on the contrary bidders shade their bids below their true value which leads once more to revenue losses for the search engines. Last but not least, under the generalized first-price auction, the advertiser who could react to competitors’ moves fastest had a substantial advantage. The mechanism therefore encouraged inefficient investments in gaming the system, causing volatile prices and allocative inefficiencies.

Google was the first to address to these problems when it introduced its own pay-per-click system, AdWords Select, in February 2002. The new mechanism was introduced with the name *Generalized Second-Price Auction* or *GSP*

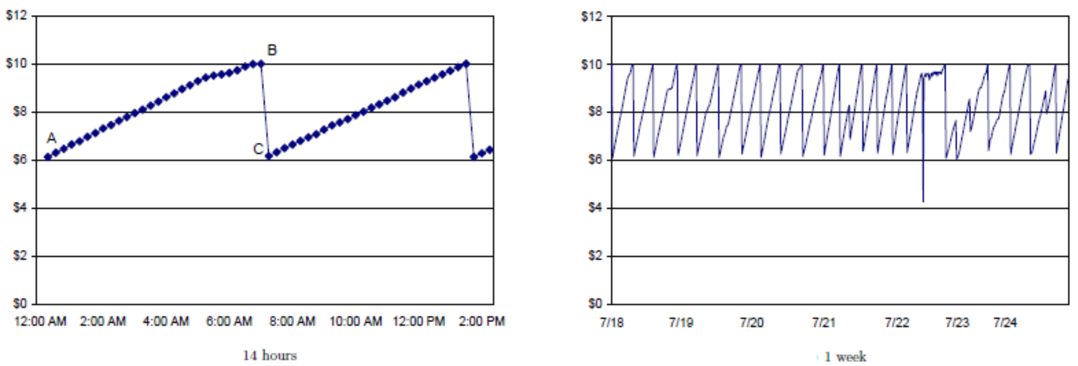

Figure 5: Bidding cycles (“saw-tooth” pattern) in first price auctions.

The Generalized Second Price Auction

The Generalized Second Price auction mechanism is, nowadays, used by the major search engines in order to generate the sponsored link list that will be displayed whenever a user enters a query to the search engine.

In its simplest form, the highest bidder pays per click the second highest bid, the second highest bidder pays per click the third highest bid, etc. The k -th highest bidder pays per click the $k+1$-th highest bid if there is such a bid. Otherwise, if $n=k$, the k-th highest bidder pays nothing. In particular, the pricing rule introduced by Google in 2002 which is until now in use is that each bidder pays the amount that is necessary to bid in order to keep his current position. For the advertiser who obtained slot i, the payment p_i should satisfy the following:

$$w_i b_i \geq w_{i+1} b_{i+1} \leftrightarrow p_i = \frac{w_{i+1} b_{i+1}}{w_i} + \varepsilon \qquad (2)$$

By the time this rule was introduced, w_i was taken to be equal to the estimated CTR, q_i, of the advertiser. The latter approach can be generalized by considering the quality weight to be equal to a power of the CTR. This form of functional dependence allows to remove more easily irrelevant advertisements, such as those with low CTR, that are accompanied by high bids (as irrelevant ads diminish the trust of customers and are therefore undesirable) (Linden G., Meek C., Chickering M., 2006). At the moment however Google's quality score, w_i, is not solely dependent on the CTR but also on other qualities of the advertiser, including also the text of the ad. The exact method of determining w_i is not publicly available. In 2007 the revenue-based ranking rule was also adopted by Yahoo! and Microsoft Live (now Bing). As ε is not an important parameter of the mechanism, from this point forward, we will make the assumption that $\varepsilon = 0$.

Generalized Second Price and Vickrey-Clark-Groves Auction

GSP mechanism looks similar to the Vickrey-Clark-Groves, or simply VCG, mechanism, because both mechanisms set each agent's payments based only on the allocation and bids of other players, and not on that agent's own bid. However GSP is not identical to VCG. Unlike GSP, where a bidder in slot i will be charged according to the bid of slot $i+1$, in VCG each advertiser's payment is equal to the negative externality that he imposes on others, assuming that bids are equal to values, by taking one of the slots away from them (Edelman B., Ostrovsky M., Schwarz M., 2007).

A few remarks are worth noticing as far as VCG auctions in comparison to GSP auctions are concerned. First, unlike GSP, VCG auctions do have an equilibrium in dominant strategies and truth-telling is generally an equilibrium strategy. With only one slot, VCG and GSP would be identical. With several slots, the mechanisms are different. Moreover, if all advertisers were to bid the same amounts under the two mechanisms, then each advertiser's payment would be at least as large under GSP as under VCG (Edelman B., Ostrovsky M., Schwarz M., 2007). Finally VCG auctions eliminate phenomena of *strategic behavior,* so that advertisers, who do not bid their private values but their bids are computed with the use of algorithms that implement a strategy, are diminished. Strategic behavior may lead to a significant shrinkage of the revenue of the search engines.

Surprisingly, nevertheless, both Google and Yahoo! still use the GSP mechanism instead of VCG mechanism due to a series of reasons. The VCG auction is a complex mechanism that can be hard for the advertisers to understand and for the auctioneer to implement. The implementation of the VCG would require from each advertiser to submit a bid for every slot and the auctioneer to solve a more difficult optimization problem (although the number of slots is small, the auction is real time). Furthermore, the VCG mechanism produces substantially less revenues for the search engine than those yielded by the GSP auction and the bidders may be slow to stop shading their bids and report their private values. Finally, the design, implementation and test of a new auction platform impose switching cost which may be unacceptable to both the search engines and the advertisers.

Equilibria of the GSP auctions

In this section we provide a brief overview of equilibrium analysis. Advertisers bidding on Yahoo! and Google can change their bids very frequently. Thus, we can think of these sponsored search auctions as continuous time or infinitely repeated games in which advertisers originally have private information about their types but gradually learn the values of others and can adjust their bids repeatedly (Edelman B., Ostrovsky M., Schwarz M., 2007). As sponsored search auctions are essentially games among advertisers, the ideal situation for the search engine is to ensure that the advertisers have no incentive to misreport their valuations. This would eliminate the possibility of potential manipulations of the used mechanism by the advertisers. However, simple examples demonstrate that neither the rank-by-revenue nor the rank-by-bid GSP mechanisms are truthful (Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B., 2010).

In principle, the sets of equilibria in such repeated games can be very large resulting to quite complex strategies required to support them. Moreover, usually, there is more than one equilibrium set of bids and among these equilibria are some that produce a socially non-optimal assignment of advertisers to slots (Easley D., Kleinberg J. , 2010). Even though in theory, advertisers could implement such strategies, in practice, multiple campaign and management of multiple keywords as well as the exclusion by the search engines of the use of automated robots which would ease the procedure, make this impossible.

We will focus on a subset of Nash equilibria that are called *Symmetric Nash Equilibria* (SNE) (Varian H. R., 2007) or *Locally Envy Free Equilibria* (Edelman B., Ostrovsky M., Schwarz M., 2007). Symmetric Nash Equilibria have specific properties of interest for the search engine and for the advertisers, and therefore they could be the equilibria that the search engine would prefer to attain (Edelman B., Ostrovsky M., Schwarz M., 2007), (Lahaie S., 2006) and (Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B., 2010). Below, we present the main research results related to such equilibria.

We first set up the model in which these equilibria are studied. The equilibrium analysis was performed for the rank-by-bid rule (but can be easily generalized to the rank-by-revenue rule), where the CTR of advertiser i for position s is assumed to be the same for all advertisers, and to depend only on the position slot s, i.e., the bidder-dependent part of CTR, q_i, is assumed to be the same for all bidders. Denoting the CTR for slot s by θ_s, assume that $\theta_1 > \theta_2 > \cdots > \theta_k > 0$. To simplify notations, let us also renumber the bidders so that v_s is the valuation of the bidder assigned slot s. Following the GSP principle, the price paid by advertiser $s \leq k$ (that is, at position s) is $p_s = b_{s+1}$ (since $\varepsilon = 0$ as mentioned before) and his utility is:

$$(v_s - p_s)\theta_s = (v_s - b_{s+1})\theta_s$$

We assume that all advertisers are likely to learn all relevant information about each other's values. In addition, since bids can change anytime, the stable bids (equilibria) must be static best responses to each other, otherwise an advertiser with a not so static bid will have the incentive to change it. At a Nash equilibrium, no advertiser would have an incentive to obtain a different slot. More formally:

A bid vector is a Nash equilibrium if for every slot s and for the advertiser at this slot, it holds:

$$(v_s - p_s)\theta_s \geq (v_s - p_j)\theta_j \qquad \forall\, j > s$$

$$(v_s - p_s)\theta_s \geq (v_s - p_{j-1})\theta_j \qquad \forall\, j < s$$

These equations denote that no bidder who wins a position has an incentive to deviate for a lower or a higher position, no bidder that wins no position has an incentive to deviate and win a position and no bidder that wins a position has an incentive to deviate and win no position.

A symmetric Nash equilibrium is a set of bids that satisfies:

$$(v_s - p_s)\theta_s \geq (v_s - p_j)\theta_j \qquad \forall\, j, s$$

The rationale behind this notion becomes more clear if we look at pairs j, s such that $s = j + 1$. If the bidder at slot s starts raising slightly his bid so as to increase the payment of the bidder above him, then bidder j can underbid him as a retaliation and essentially this means that they will have swapped their bids. The right hand side of the equation expresses the payoff of a bidder s if bidders j and s swap their bids. Symmetric Nash equilibria capture the notion that there should be no incentives for such swapping of bids between any pair of players.

It is straightforward to verify that if a bid vector satisfies the inequalities of the previous equation, then it will be a Nash equilibrium. Hence the class of SNE is a subclass of the set of Nash equilibria. The following key properties are satisfied by SNE and can be helpful for the search engine (Varian H. R., 2007):

- At an SNE, there is monotonicity in the valuations of the winning bidders, i.e., $v_{s-1} \geq v_s \quad \forall\, s$.

- If a set of bids satisfies the SNE inequalities for slots $s-1$ and $s+1$, then it satisfies these inequalities for all s.
- There is an SNE maximizing the search engine revenue among all possible Nash equilibria.

Finally, it is interesting to compare the GSP mechanism with the classical Vickrey-Clarke-Groves (VCG) auction, which is truthful and where each bidder pays for the externality that he is causing to the other bidders. The following recursively defined configuration has been shown to be an SNE:

$$b_j^* = \begin{cases} 2b_2^*, & j=1 \\ \gamma_j b_{j+1}^* + (1-\gamma_j)v_j\,, & 2 \le j \le k \\ v_j, & k < j \le n \end{cases} \tag{3}$$

where, $\gamma_j = \frac{\theta_j}{\theta_{j+1}}$ (for $\gamma_1$we take $\theta_0 = 2\theta_1$). It is interesting that b_1^* can actually be any quantity greater than b_2^* since it does not affect the price of any slot. Except for this degree of freedom, this Nash equilibrium does not involve over-bidding. That is, the various bids do not exceed the corresponding advertisers' valuations.

Bidding vector b^* is known as the VCG equilibrium of GSP. This SNE was shown in to be (Edelman B., Ostrovsky M., Schwarz M., 2007) the worst SNE for the search engine in terms of revenue, and the best for the advertisers. In other words, the engine revenue under GSP is always better than when using the truthful VCG mechanism, which provides an . This is summarized in the following theorem (Aggarwal G., Feldman J., Motwan R., 2006).

The bidding vector b^* defined by (3) is an SNE. In this equilibrium the assignment and the payments are identical to the dominant strategy equilibrium of the VCG mechanism. Furthermore, in any other SNE, the revenue is at least as high as the revenue of b^*.

For a more extensive analysis regarding equilibria in sponsored search auctions we refer the reader to (Borgers T., Cox I., Pesendorfer M., 2007), (Edelman B., Ostrovsky M., Schwarz M., 2007), (Edelman B., Schwarz M., 2010) and (Varian H. R., 2007).

3

Sponsored Search Auction Data Analysis

In the following chapter we examine Overture's paid placement auctions between June 2002 and June 2003 (Overture was acquired by Yahoo! in the second half of 2003, after the end of our sample). The data was provided as part of the Yahoo! Webscope program for use solely for academic research purposes. The dataset was produced from Yahoo!'s records and has been reviewed by an internal board to assure that no personally identifiable information is revealed. Under these circumstances, the dataset was rendered inappropriate for any further analysis, reverse engineering or processing of the data or any correlation with other data sources so that it could be used to determine or infer personally identifiable information (Yahoo! Webscope dataset ydata-ymusic-user-artist-ratings-v1_0).

Description of the dataset – Problems encountered

The Yahoo! Webscope program included a series of datasets of different categories. The dataset which we analyzed was part of the *Advertising & Markets Data* and more specifically "A3. Yahoo! Search Marketing Advertiser Bid-Impression-Click data on competing Keywords, version 1.0"

This dataset contains a small sample of advertiser's bid and revenue information over a period of 4 months (123 days). Bid and revenue information is aggregated with a granularity of a day over *advertiser account id*, *key-phrase* and *rank*. Apart from *average bid*, *impressions* and *clicks* information is also included. Sequence of keywords make a key-phrase. A key-phrase can belong to one or more key-phrase categories. Advertiser account id is represented as a meaningless string. Key-phrase is represented as sequence of meaningless strings, where each string represents a keyword or key-phrase category.

day	anonymized account_id	rank	anonymized keyphrase (expressed as list of anonymized keywords)	avg bid	impressions	clicks
10	2f79e07e-d4b0-4fd0-9b5c-785cee142f09	5	14e71a9195e7b4dc ac26bfe54a8a8f38 3db691494440189b	740	8	1
119	33cfc84f-f0d2-4af6-8a54-7b1e57d0ed59	2	1a4d1506c569d26d 79021a2e2c836c1a 55c43ba572f24c9f	220	6	2
55	6a3f1af4-14f8-4ee2-a700-e86ebf49e9fc	5	13e34d4363aefd0c cd74a8342d25d090 bda6b5235e5592c1 a84f1e914754570a	230	1	1
58	f90a1f59-dfb1-42ef-bf41-759a0e498b1e	2	10b6b278aaa6d3ac 79021a2e2c836c1a 1b3660568d34cf1b	210	9	1
100	33cfc84f-f0d2-4af6-8a54-7b1e57d0ed59	1	100e12ed2a2712d7 9ae15116a91320a6 79021a2e2c836c1a	550	8	1
87	33cfc84f-f0d2-4af6-8a54-7b1e57d0ed59	1	10b6b278aaa6d3ac e52cf59b14d35f1f 5b9834ef1db9cdf7 79021a2e2c836c1a 6d4a2eeae5820969	290	1	1

Figure 6: Yahoo! Webscope A3 dataset snippet

According to the information provided by the Yahoo! Webscope program, there are 6 key-phrase categories. These are the following:

- de84da9dfd5a336d
- 79021a2e2c836c1a
- cd74a8342d25d090
- 3db691494440189b
- aef4ee042bea9c6b
- fc4f04e287746c48

During the analysis of the dataset we came across a series of problems. First, due to the anonymity of the key-phrases, as it was stated by the Webscope program itself, no valuable conclusion could be reached concerning the nature of the keywords. There have been observed, nevertheless, anonymized key-phrases with similar or related string-names, as seen in figure 7, for which we can presume that they are related keywords or keywords of the same category. For example key-phrases coded as "3db691494440189b 3e3f6346413117e2" and "3db691494440189b 3e3f6346413117e2 204804677e80f854" could correspond to "buy iPhone" and "buy iPhone 4" respectively.

day	anonymized account_id	rank	anonymized keyphrase (expressed as list of anonymized keywords)	avg bid	impressions	clicks
1	5c9aa1f2-1c09-418e-b79b-836e50e3eb09	7	3db691494440189b 3554bb31d2f1e395	850	20	1
3	5c9aa1f2-1c09-418e-b79b-836e50e3eb09	8	3db691494440189b 30e0a355cdf8a591	1290	121	1
4	5c9aa1f2-1c09-418e-b79b-836e50e3eb09	5	3db691494440189b 3e3f6346413117e2	8.841.463.414.634.140	141	3
4	5c9aa1f2-1c09-418e-b79b-836e50e3eb09	2	3db691494440189b 3e376b5dcd422f6a	550	14	1
9	5c9aa1f2-1c09-418e-b79b-836e50e3eb09	2	3db691494440189b 3e3f6346413117e2 204804677e80f854	990	28	1
21	5c9aa1f2-1c09-418e-b79b-836e50e3eb09	1	3db691494440189b 3806857fc4520347	550	4	1

Figure 7: Anonymized keyphrases similarities

However there is no certified, documented or secure way to conclude that these terms are indeed related in any way to each other. The only information we can extract is that both key-phrases belong to the category under the string-name

“3db691494440189b”. On the other hand neither category “3db691494440189b” can provide us with any useful information in any way, as it is also anonymized.

Moreover, there has been a particular difficulty when dealing with average bids (*avg bid* column). As seen in figures 7 and 8, there are values which present a rather awkward format. These deviant values are probably badly formatted; what so ever there is no official explanation from Yahoo!’s part. Another problem related to *avg bid* values is that there is no documented detail about the way it is calculated in a specific day. We have verified that the column *avg bid* is not correlated with *clicks*, so we can confirm that the avg_bid is not the price paid. We have no further useful information, though, in order to reach some conclusion regarding bidding behavior or strategies used.

day	anonymized account_id	rank	anonymized keyphrase (expressed as list of anonymized keywords)	avg bid	impressions	clicks
47	e185153f-ffaf-4c6d-8c74-4b638ad24a87	2	cd74a8342d25d090 3381856224f1c922 2ac01c3be4bef1a8	43.782.608.695.652.100	24	2
108	e185153f-ffaf-4c6d-8c74-4b638ad24a87	1	5715fbb855c88e5a 79021a2e2c836c1a	4.167.567.567.567.560	66	10
80	e185153f-ffaf-4c6d-8c74-4b638ad24a87	2	85394825e2a7c6f4 79021a2e2c836c1a	40.588.235.294.117.600	19	1
83	e185153f-ffaf-4c6d-8c74-4b638ad24a87	1	85394825e2a7c6f4 79021a2e2c836c1a	43.090.909.090.909.000	65	2

Figure 8: Deviant avg bid values

Dataset Analysis

According to Edelman and Ostrovsky (Edelman B., Ostrovsky M., 2007), during the period between June 2002 and June 2003, Overture generally operated a first-price auction and changed to GSP soon later, when Overture was acquired by Yahoo!. However, as far as we do not examine the bidding behavior, the mechanism used plays no particular role and does not affect our findings.

Taking into consideration the restrictions and problems of the dataset, mentioned above, we proceeded in the dataset analysis. The fields we were concentrated involved click through rate estimation for slots 1 to 8, as well as click through rate estimation for key-phrases/ key-phrase categories which have been identified as popular among the total data provided.

The dataset is consisted by *77850272* single records, each one of which contained the following fields: day, anonymized account_id (denoting the anonymized bidder), rank (denoting slot awarded), anonymized key-phrase (expressed as list of anonymized keywords), avg bid, impressions and finally clicks. Primary key of the data is a combination of fields date, account_id, rank and key-phrase. Average bid, impressions and clicks information is aggregated over the primary key.

Due to the huge bulk of the data, and the difficulty in parsing and analyzing it, we divided it into 78 subsets of 1000000 records (the last one contained the remaining 850272 records). However the results refer to the total number of records instead of a statistical random selection of them.

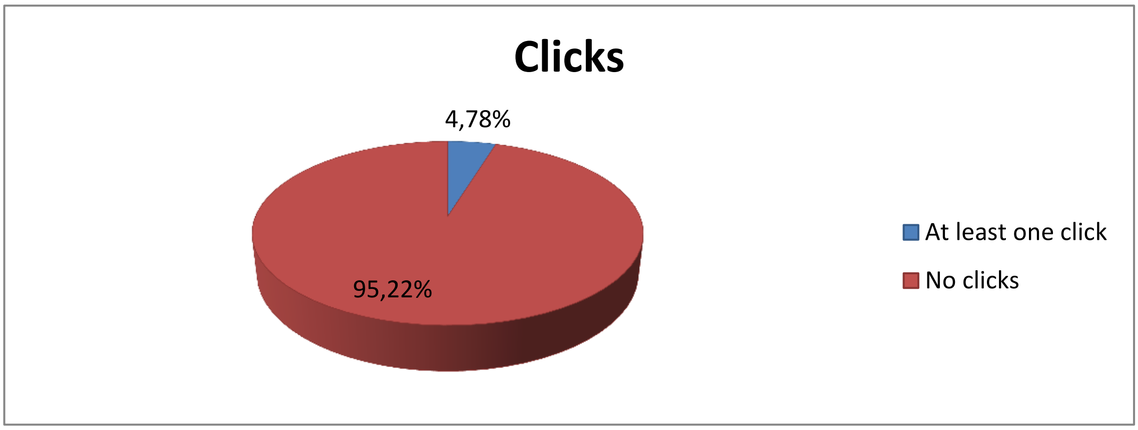

Chart 1: Percentage of advertisements at least once clicked

As seen in Chart 1, the final percentage of the advertisements which were clicked at least once is 4,78%. This is consistent with Richardson's, Dominowska's and Ragno's findings published in their paper (Richardson M., Dominowska E., Ragno R., 2007), where it is stated that CTR is 5%. CTR is defined as:

$$CTR = \frac{x}{y}$$

where x is the number of times an ad is clicked on and y the number of impressions of the page on which that advertisement appears. In this result helps the big volume of data we analyzed.[1]

As with search results, the probability that a user clicks on an advertisement declines rapidly, as much as 90%, with display position. For the first 8 slots the CTR is decreased from 0.28828 for the 1^{st} slot to 0.03152 for the 8^{th} slot. The following chart 2, depicts the reduction of CTR: *0.28828* for slot 1, *0.18934* for slot 2, *0.14539* for

[1] In the paper of Richardson M., Dominowska E. and Ragno R. (2007), it is stated that an ad with a true CTR of 5% must be shown 1000 times before we are even 85% confident that our estimate is within 1% of the true CTR.

slot 3, *0.09504* for slot 4, *0.07566* for slot 5, *0.04706* for slot 6, *0.03662* for slot 7 and *0.03152* for slot 8, which totally gives a coverage of 90.89%.

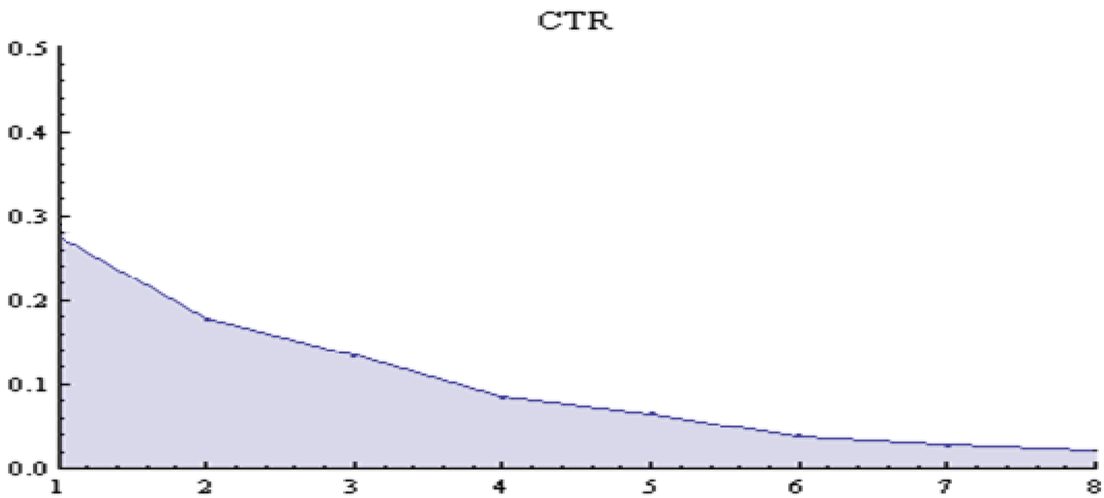

Chart 2: Reduction of CTR throughout slots 1 to 8 for clicked on advertisements

More specific, from the total advertisements clicked on, more than 50% of them appear in the 3 top slots and the rest of them in the slots 4 to 8, as seen in chart 3.

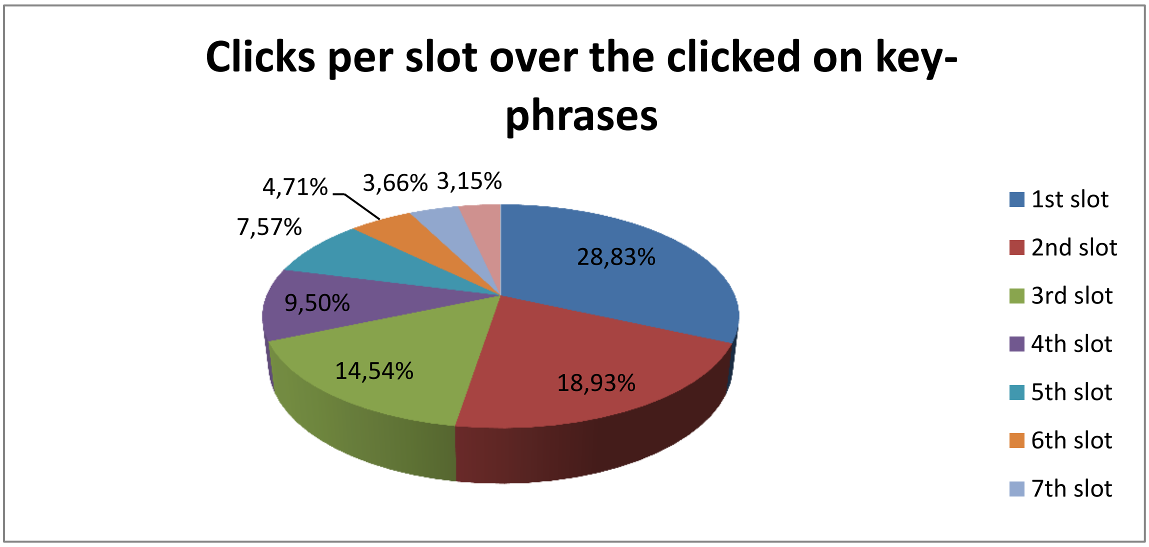

Chart 3: Percentage of clicks per slot for clicked on advertisements

We perform the same analysis over the total data, including the advertisements which were not clicked on. The results are verified to be proportionally correlated as expected.

As with search results, the probability that a user clicks on an advertisement declines rapidly, as much as 90%, with display position. For the first 8 slots the CTR is decreased from 0.01378 for the 1st slot to 0.00151 for the 8th slot. The following chart 4, depicts the reduction of CTR: *0.01378* for slot 1, *0.00905* for slot 2, *0.00695* for slot 3, *0.00454* for slot 4, *0.00362* for slot 5, *0.00225* for slot 6, *0.00175* for slot 7 and *0.00151* for slot 8, which totally gives a coverage of 91.862%. This slight difference between the CTR coverage percentage, as seen in chart 5, is justified by the fact that there were clicked on advertisements in more than the 8 first slots which we are studying.

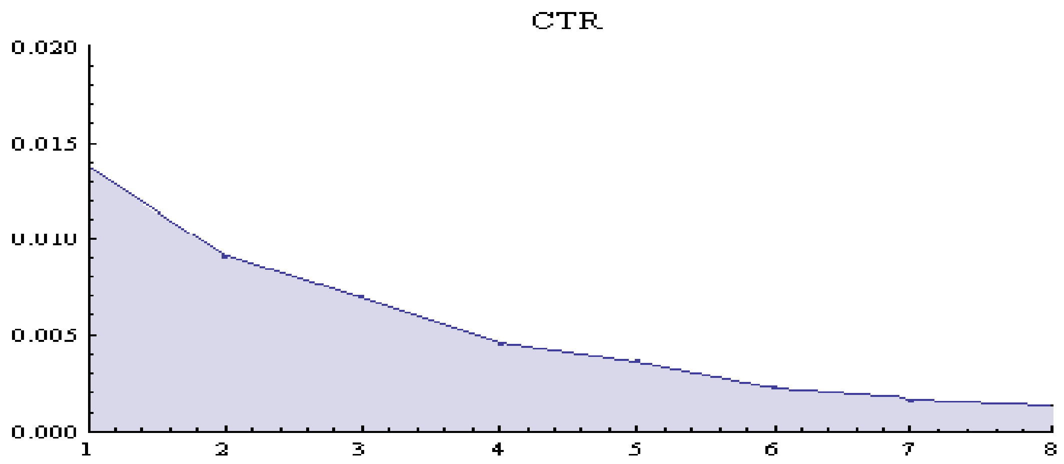

Chart 4: Reduction of CTR throughout slots 1 to 8

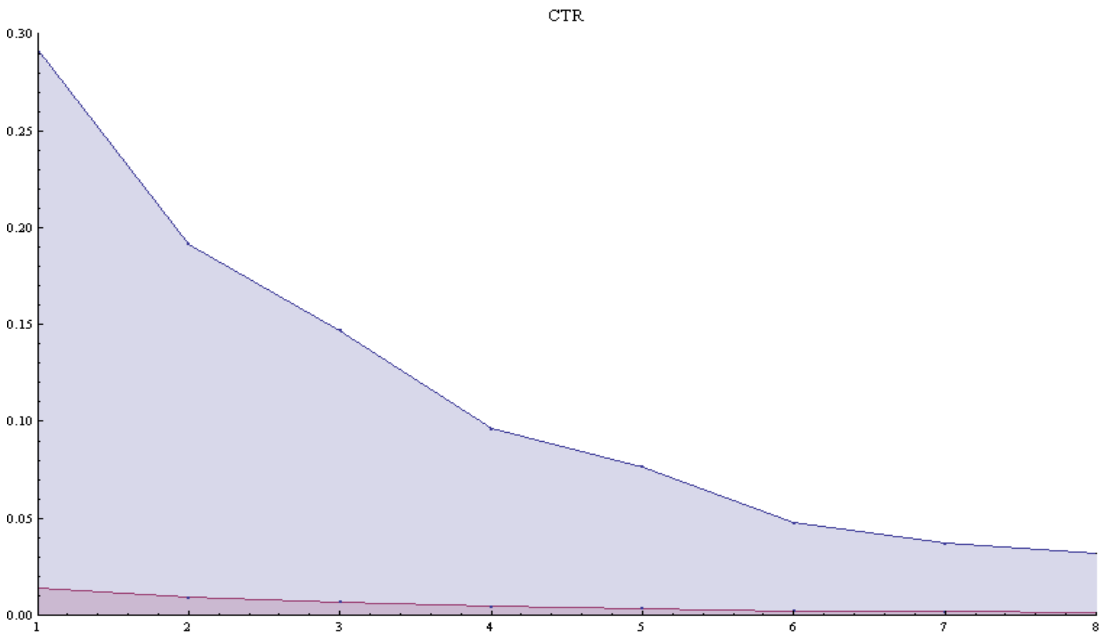

Chart 5: Reduction of the CTR (comparison)

Click Through Rate Exponential Decay Model

The click-throughs that a listing generates depends on both its relevance and its rank within the sponsor-search section, because users are inherently more likely to click on higher-ranked items. Our analysis of the Yahoo! dataset intents to verify the expected number of click-throughs for an item x at position j, as it was computed by (Feng J., Bhargava H. K., Pennock D. M., 2007).

In their study, Feng, Bhargava and Pennock used the exponentially decaying attention model with factor $\delta > 1$, computing the average click-through as $\frac{a_x}{\delta^{j-1}}$. According to them, exponential decay of attention is a fairly standard assumption, which is borne out in practice. Analysis of the actual click-through data obtained from Overture during 2003 for the top five positions across all affiliates -including Yahoo!, MSN, and AltaVista- were fitted extremely well by an exponential decay model with $\delta = 1.428$.

In our analysis we estimated click through rates under the exponentially decaying model as well. Our analysis included, though, the top 8 slots and only for Yahoo!. Let A be the CTR of the 1st slot.

$$CTR_j = \frac{A}{\delta^{j-1}}$$

Subsequently, under the exponential decaying model, the CTRs of the following slots are $\frac{A}{\delta}$ for the 2nd slot, $\frac{A}{\delta^2}$ for the 3rd slot and so on.

Parameter δ is defined as following:

$$\frac{CTR_j}{CTR_{j+1}} = \delta \qquad (4)$$

$$(4) \Rightarrow logCTR_j - logCTR_{j+1} = log\delta, \delta > 1 \qquad (5)$$

The following tables contain the logδ and δ (delta) values, as they were computed according to (5) and (6). The average delta for both clicked-on and not clicked-on advertisements resulted as the average of individuals deltas.

logδ	
Ads with Clicks>0	**All Ads**
0.1825	0.1826
0.1147	0.1146
0.1846	0.1849
0.0990	0.0983
0.2062	0.2065
0.1089	0.1091
0.0651	0.0640
(average = 0.1373)	(average = 0.1371)

delta	
Ads with Clicks>0	**All Ads**
1.5225	1.5226
1.3022	1.3021
1.5229	1.5308
1.2561	1.2541
1.6077	1.6088
1.2850	1.2857
1.1618	1.1589
(average = 1.3808)	(average = 1.3804)

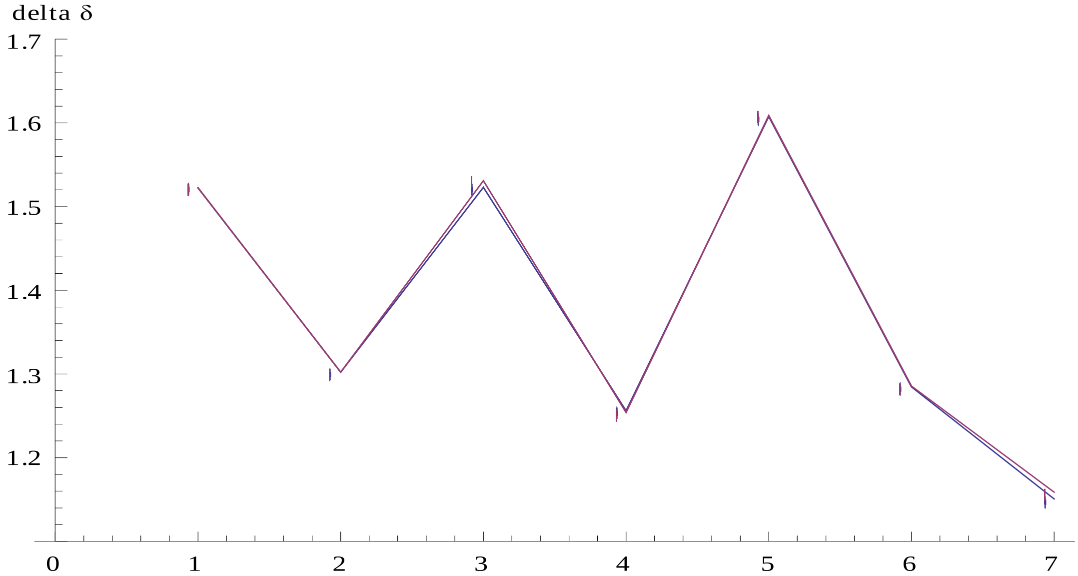

Chart 6: Graphic representation of δ (delta) values. The blue line corresponds to advertisements at least once clicked on, while the purple one to the total advertisements.

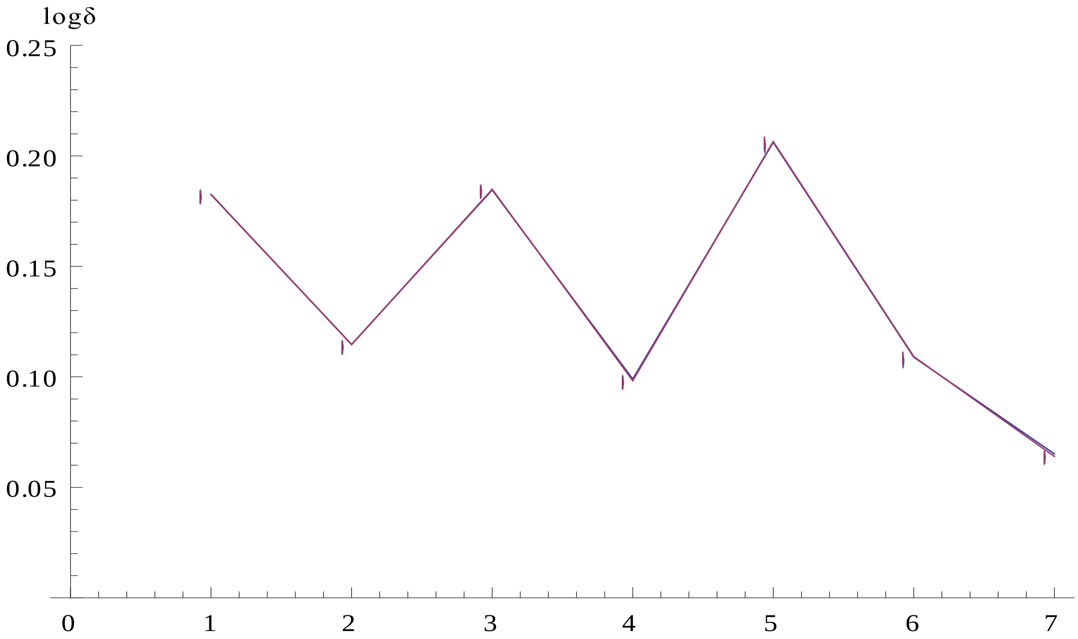

Chart 7: Graphic representation of logδ values. The blue line corresponds to advertisements at least once clicked on, while the purple one to the total advertisements.

As we can see from the table and the following graphic representation, delta values are almost identical and very close to $\delta = 1.428$, suggested by Feng, Bhargava and Pennock. The slight deviation between clicked-on and not clicked on advertisements is justified by the fact that in our analysis there were clicked on advertisements in more than the 8 first slots which we are studying. As far as the deviation from the Feng, Bhargava and Pennock research is concerned, this might be justified by the fact that they used data from other affiliates as well (MSN and AltaVista) and the fact that they were examining top five positions (instead of 8).

Because of the variation of the delta values, and the possible decay of importance of delta as we move down from slot 1 to slot 8, we decided to take the running average of the delta values[2]. Through this we observed a convergence of the values to both our average and the delta derived by Feng, Bhargava and Pennock's research. In the following charts we can see an overview of delta values and the corresponding averages, for both the ads that were clicked on and the total population of advertisements.

[2] As we go down from slot 1 to slot 8, we have fewer impressions and fewer clicks. This fluctuation may impose undesirable deviations and lead to controversial results.

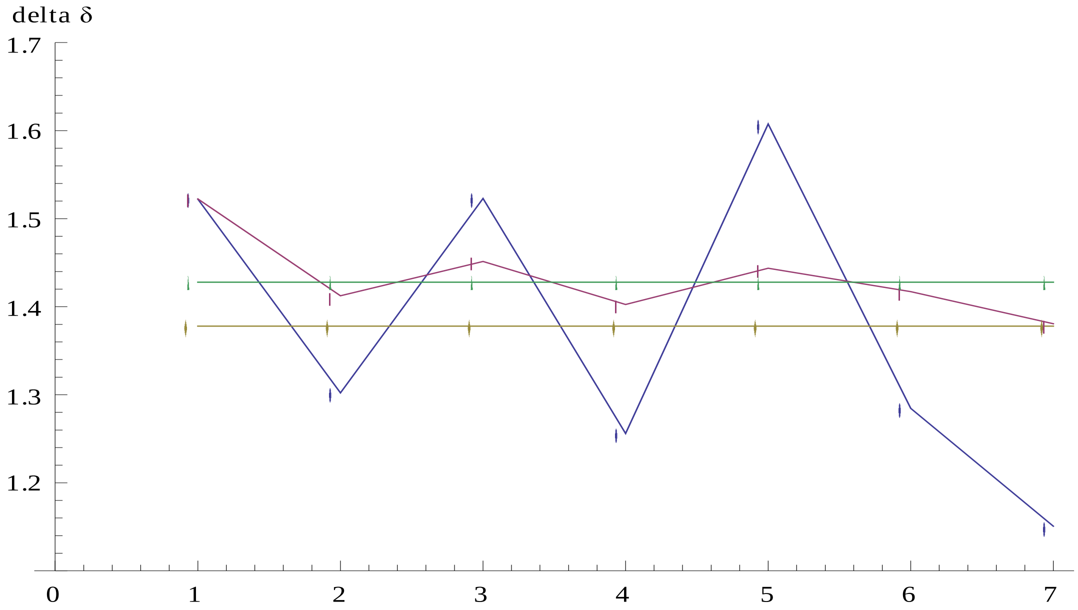

Chart 8: Graphic representation of δ (delta) values of advertisements at least once clicked. The blue line corresponds to the delta values, the purple one to the running average, the olive one to the average and finally the green one to Feng, Bhargava and Pennock's delta.

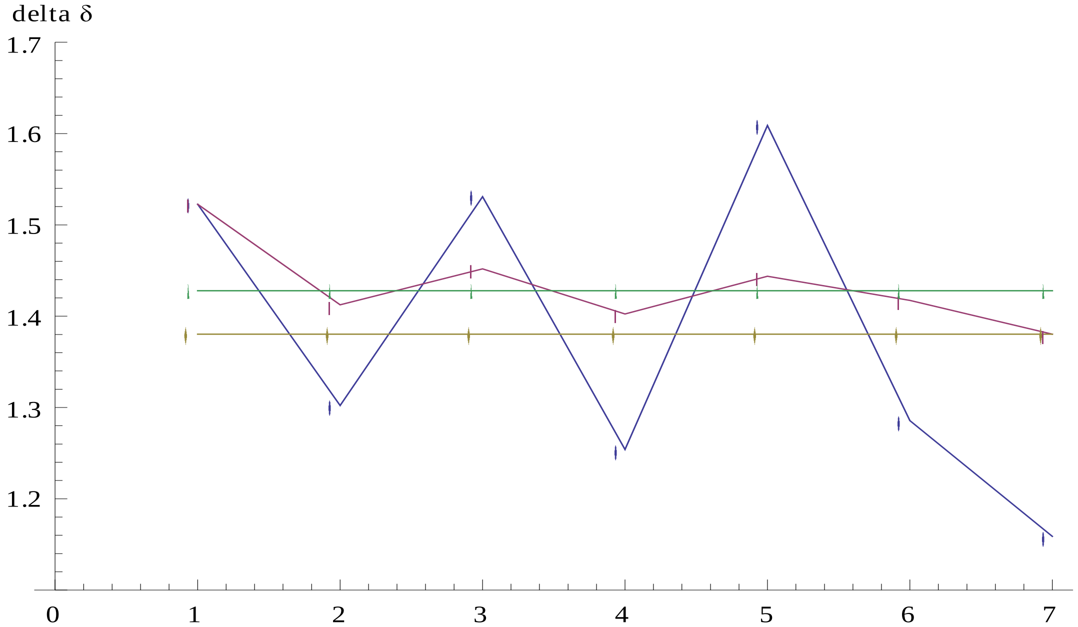

Chart 9: Graphic representation of δ (delta) values of all advertisements. The blue line corresponds to the delta values, the purple one to the running average, the olive one to the average and finally the green one to Feng, Bhargava and Pennock's delta.

Click Through Rate Power Law Model

During our analysis we decided to test other distribution models, concerning the CTR. We opined that power law distributions would most likely fit to our purposes. In this context we selected *Zipf Law* distribution. The exact model form that was used is described hereupon:

$$CTR_j = \frac{A}{j^r}$$

$$\log CTR_j = \log \frac{A}{j^r} = \log A - \mathrm{r} * \log j \qquad (6)$$

$$\log CTR_{j+1} = \log \frac{A}{(j+1)^r} = \log A - \mathrm{r} * \log (j+1) \qquad (7)$$

$$(6)(7) \Longrightarrow \qquad \log CTR_j - \log CTR_{j+1} = \mathrm{r} * \log (j+1) - \mathrm{r} * \log j$$

$$\log \frac{j}{CTR_{j+1}} = \mathrm{r} * \log \frac{j+1}{j} \rightarrow r = \frac{log \frac{CTR_j}{CTR_{j+1}}}{log \frac{j+1}{j}} \qquad (8)$$

According to the values of CTR calculated previously we take the following:

r	$\mathbf{log \frac{CTR_j}{CTR_{j+1}}}$	$\mathbf{log \frac{j+1}{j}}$
0.6063	0.1825	0.301
0.6517	0.1147	0.176
1.4779	0.1846	0.1249
1.0216	0.099	0.0969
2.6068	0.2062	0.0791
1.6263	0.1088	0.0669
1.1255	0.06517	0.0579

As we can notice in the value table above and the following chart, there is a correlation between $\log \frac{CTR_j}{CTR_{j+1}}$ and r, which present quite similar behavior (with fluctuations). The amount $\log \frac{j+1}{j}$ functions as a regulatory factor. However, it is obvious that there is no convergence of the values, as they seem to introduce considerable deviations. Therefore, Zipf Law distribution is rated as inappropriate and inacceptable for our purposes.

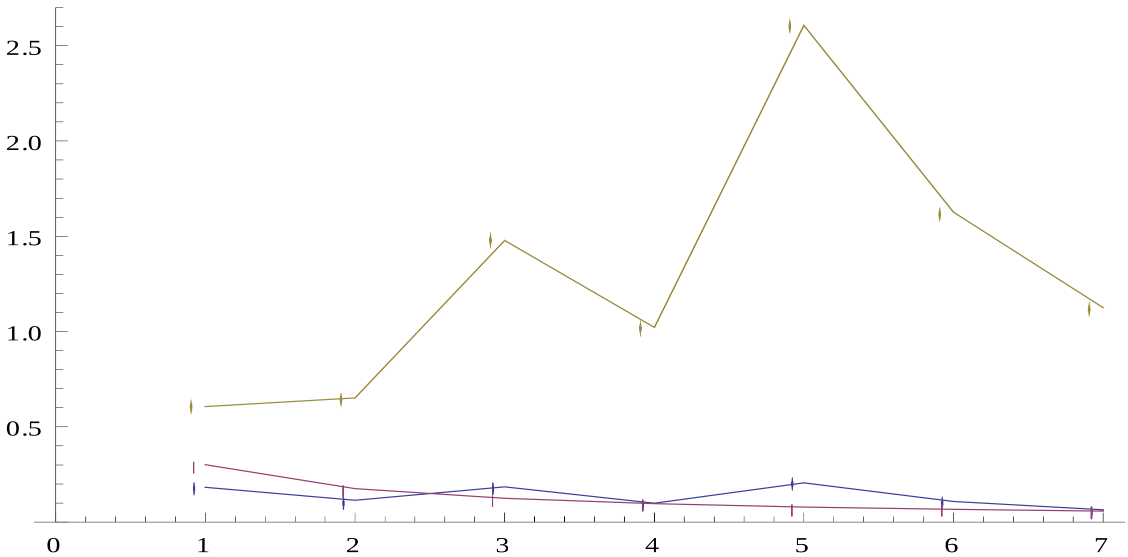

Chart 10: Zipf Law distribution of CTR. The olive line represenets r, blue line $\mathbf{log}\frac{CTR_j}{CTR_{j+1}}$ and purple line $\mathbf{log}\frac{j+1}{j}$

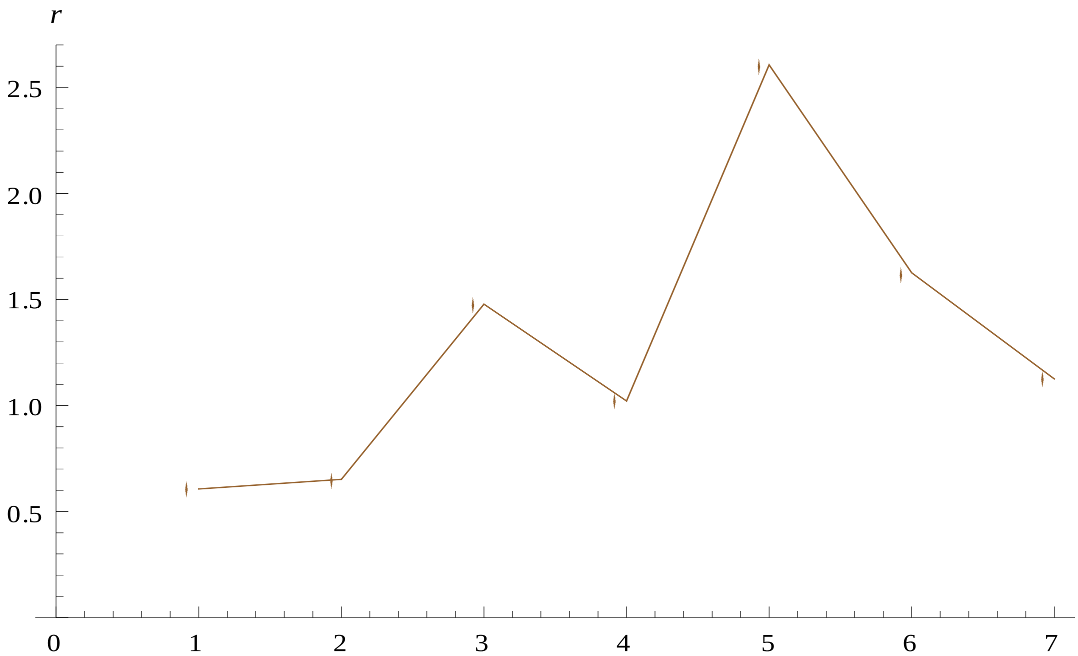

Chart 11: Graphic representation of r**.**

Popular Keywords' / Popular Categories' Analysis

The analysis of the dataset has provided us with some interesting information concerning the keywords and their categories. As it's mentioned earlier, the data is anonymized insofar there is no certified, documented or secure way to conclude of the nature of both keywords and keywords categories or the market "power" they may hold.

Unfortunately the huge bulk of data made the processing and analysis of the dataset particularly difficult and problematic and the existence of a great number of different keyphrases, very similar to each other and often with correlation to more than one keyword categories, contributed to this too. However, after exhaustive processing we managed to identify the most popular keyword categories and the most popular keywords. The popularity of the keywords and the categories was measured by means of total appearances, independently of receiving or not any clicks.

As far as the popular keywords categories are concerned, the most popular is identified under the anonymized keyphrase "*aef4ee042bea9c6b fc4f04e287746c48*". This indicates that advertisements under this keyphrase correspond to two different keyword categories, the "*aef4ee042bea9c6b*" and "*fc4f04e287746c48*". The occurrence of this combined category to the total number of advertisement which received at least one click is equal to 4.63%, while to the total number of advertisements processed, 0.222%. As far as the popular keywords are concerned, the most popular of them is identified under the anonymized keyphrase "*cd74a8342d25d090 9098047d8e656cfc*" which represents 0.165% of the total number of advertisements which were at least once clicked on, while 0.0079% of the total number of advertisements. This indicates that advertisements under this keyphrase correspond to the keyword category "*cd74a8342d25d090*" as well as to the keyword under the identifier "*9098047d8e656cfc*".

	aef4ee042bea9c6b fc4f04e287746c48	***cd74a8342d25d090 9098047d8e656cfc***
Impressions	107058568	3748187
Clicks	404957	170217

The interesting fact about this keyword and keyword category is the results of the analysis under the exponential decay model. The analysis showed that they do not follow the "typical", expected exponential decay model as this was described previously in this chapter. Analytically, the first 5 slots do present a descending CTR; however for both "*aef4ee042bea9c6b fc4f04e287746c48*" and "*cd74a8342d25d090 9098047d8e656cfc*" the 6th slot appears to have a CTR lower than the corresponding CTR for the 7th slot.

As far as the keyword category is concerned, despite the small variations, the average δ (delta) is 1.438, which is very close to $\delta = 1.428$, suggested by Feng, Bhargava and Pennock. Nevertheless, this does not apply to keyword "*cd74a8342d25d090 9098047d8e656cfc*", where the average $\delta = 2.061$. This deviation may be explained by the comparatively small portion of it to the total number of advertisements, which were at least once clicked on as well as the total number of advertisements (0.165% and 0.0079% respectively).

The following tables depict the variation of the logδ and delta values correspondingly. The blue line in the chart corresponds to "*aef4ee042bea9c6b fc4f04e287746c48*", while the purple one to "*cd74a8342d25d090 9098047d8e656cfc*".

logδ	
aef4ee042bea9c6b fc4f04e287746c48	***cd74a8342d25d090 9098047d8e656cfc***
0.401154	0.445670
0.174679	0.142227
0.256642	0.514530
0.102041	0.333201
-0.285400	0.182651
0.061253	-0.050389
0.118290	0.382998

delta	
aef4ee042bea9c6b fc4f04e287746c48	***cd74a8342d25d090 9098047d8e656cfc***
2.518573	2.790423
1.495131	1.387481
1.805688	3.269866
1.264854	2.153782
0.518324	1.522828
1.151472	0.890460
1.313075	2.415450
(average: 1.438160)	(average: 2.061470)

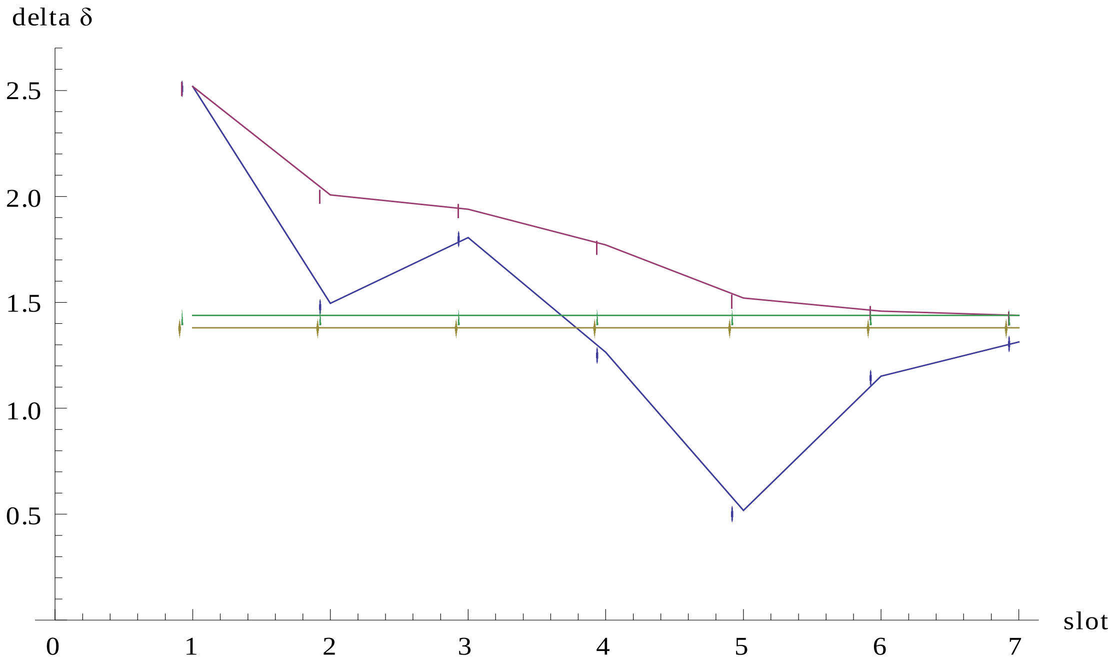

Chart 12: Graphic representation of δ (delta) values for "*aef4ee042bea9c6b fc4f04e287746c48*". The blue line correspond to delta values, the purple line to running average, the green one to average, and finally the olive one to total average delta as we computed it.

5

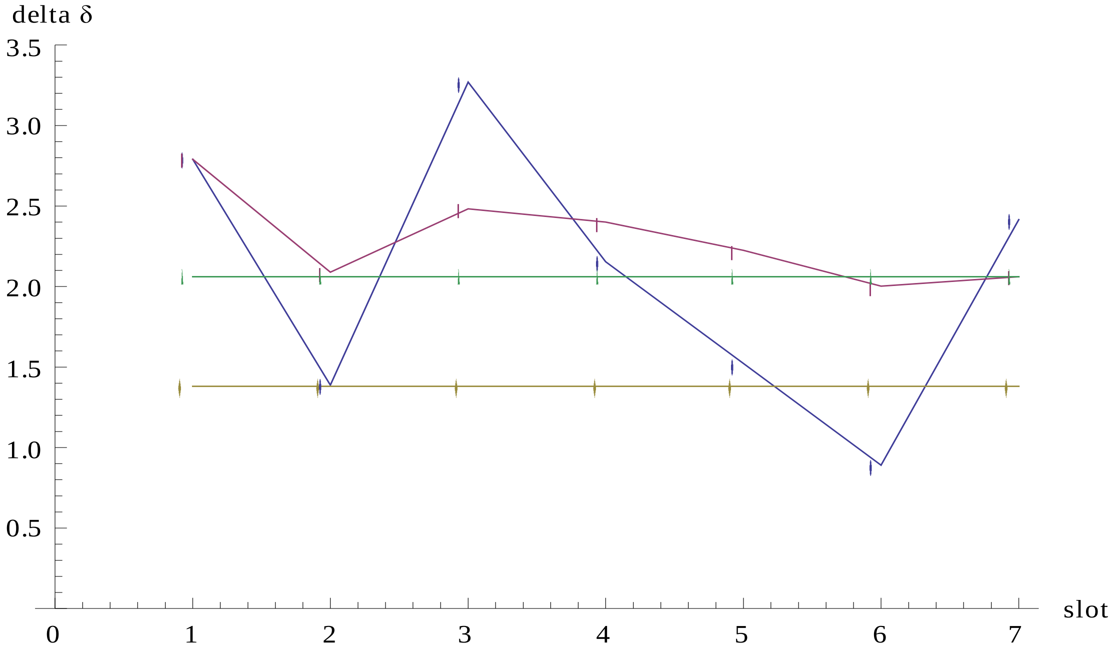

Chart 13: Graphic representation of δ (delta) values for "*cd74a8342d25d090 9098047d8e656cfc*". The blue line correspond to delta values, the purple line to running average, the green one to average, and finally the olive one to total average delta as we computed it.

We are going to apply Zipf's power law model as well. According to equation (8) we have:

r	
aef4ee042bea9c6b fc4f04e287746c48	***cd74a8342d25d090 9098047d8e656cfc***
1.332739	1.480631
0.992496	0.808109
2.054785	4.11953
1.053051	3.438611
-3.608068	2.309115
0.915597	-0.753147
2.042998	6.614821

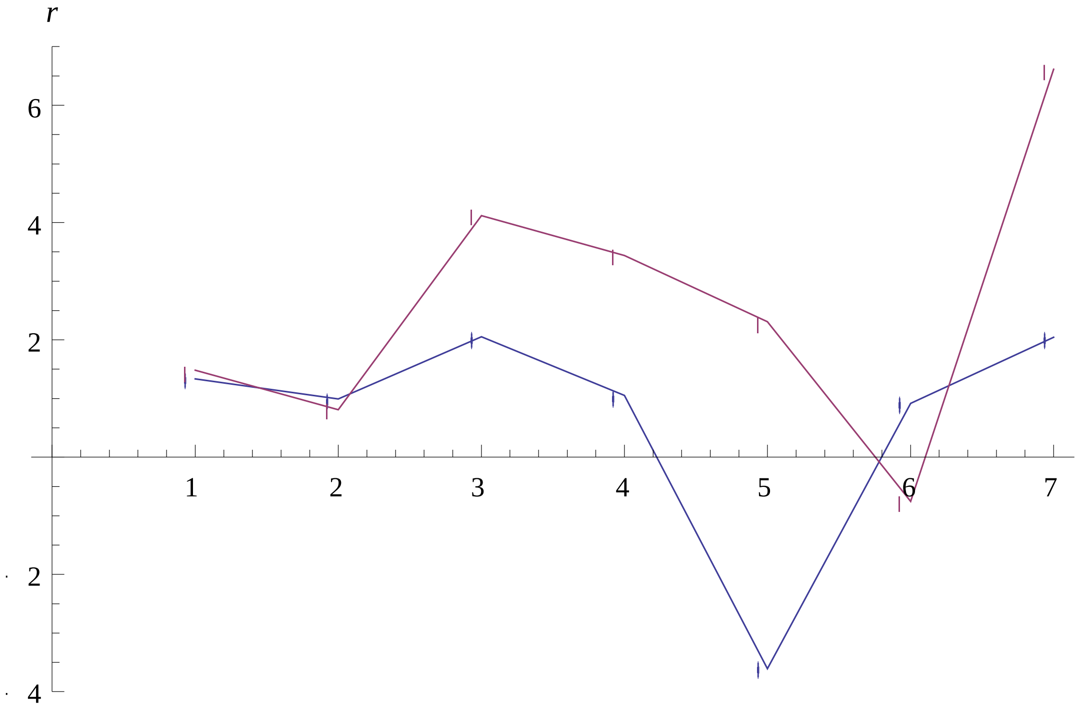

Chart 14: Graphic representation of r values. The blue line corresponds to "*aef4ee042bea9c6b fc4f04e287746c48*", while the purple one to "*cd74a8342d25d090 9098047d8e656cfc*".

As stated before, it is obvious that there is no convergence of the values for none of the "*aef4ee042bea9c6b fc4f04e287746c48*", "*cd74a8342d25d090 9098047d8e656cfc*", as they seem to introduce considerable deviations. Therefore, Zipf Law distribution is rated once more as inappropriate and inacceptable for our purposes.

Adjustment of the data

During the processing of popular keyword category "*aef4ee042bea9c6b fc4f04e287746c48*" we observed an inconsistency. The impressions of the 3rd, 4th and 5th slot outnumbered those of the 1st and 2nd slot, as it is seen in the following table.

Slot	Impressions
1	15377054
2	17427643
3	20840656
4	19775056
5	19218087
6	5174398
7	4799064
8	4604555

This imposes serious questions about the dataset itself. Since the auctions concern 8 slots, the expected result of the analysis should be that the 1st slot would receive the most impressions and the impressions would follow a descending distribution for the 2nd to 8th slot. This assumption is satisfied for the dataset in general, as seen before, however when it comes to keyword category "*aef4ee042bea9c6b fc4f04e287746c48*" there is a deviation.

Considering that if an advertisement is displayed X times at slot i, then it should have been displayed at least X times plus a small positive quantity at slots $i - z$, where $z > 0$, the previous table becomes:

Slot	Impressions
1	20840656+ε'
2	20840656+ε
3	20840656
4	19775056
5	19218087
6	5174398
7	4799064
8	4604555

Where $\varepsilon > 0$ and $\varepsilon' > \varepsilon$. To simplify, we consider $\varepsilon = \varepsilon' = 0$.

logδ	delta
0.346789	2.222230
0.097006	1.250278
0.256642	1.805688
0.102041	1.264854
-0.285400	0.518324
0.061253	1.151472
0.118290	1.313075
(average: 0.099517)	(average: 1.360846)

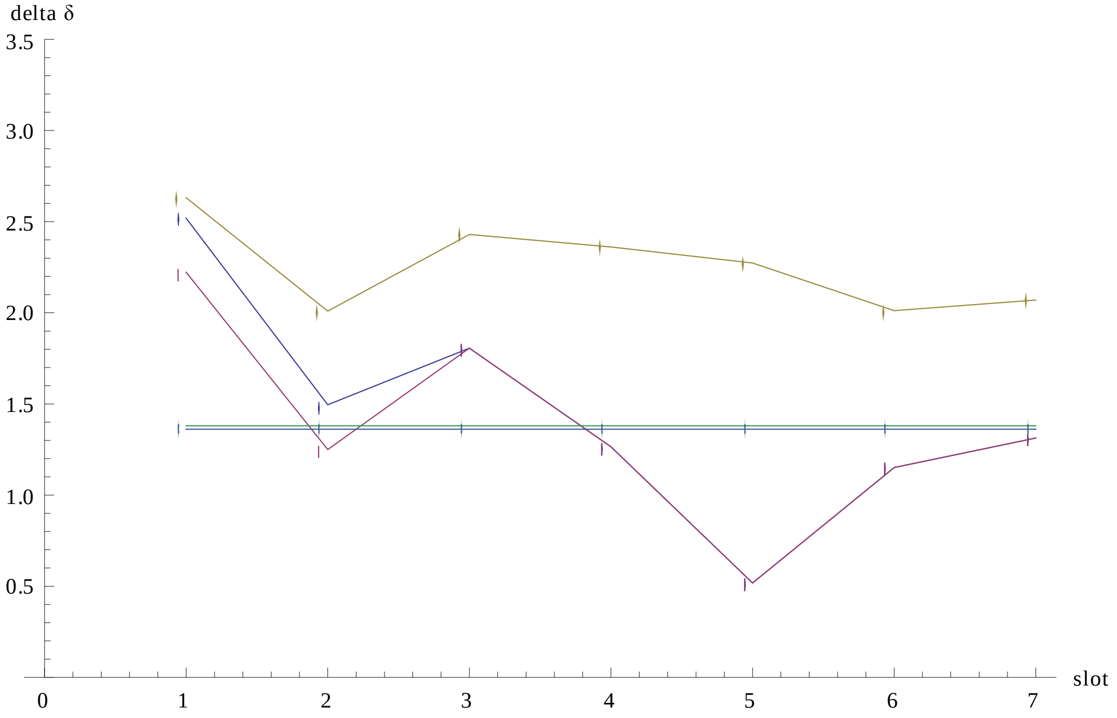

Chart 15: Graphic representation of δ (delta) values for "*aef4ee042bea9c6b fc4f04e287746c48"* before (purple line) and after the adjustment (dark blue –circle- line) as well as average (light blue –triangle- line) and running average (olive line) of adjusted values and total average delta (green line).

We can see there is a convergence to the average delta value, as it was calculated for the entire dataset (delta=1.3804). However the values still insert a small variation (between the 1st and the 2nd slot and the 5th and the 6th.

As far as the Zipf distribution is concerned, this adjustment introduced no significant improvement.

A similar situation occurs with popular keyword "*cd74a8342d25d090 9098047d8e656cfc*". The impressions of the 2nd and 3rd slot outnumbered those of the

1^{st} as well as the impressions of the 7^{th} slot outnumber those of the 6^{th}, as it is seen in the following table.

Slot	Impressions
1	621354
2	658820
3	635205
4	595371
5	495759
6	249770
7	315540
8	176375

After the adjustment we have the following values:

Slot	Impressions
1	658820+ε
2	658820
3	635205
4	595371
5	495759
6	315540+ε
7	315540
8	176375

Once more we consider ε=0. The logδ and delta values are the following:

logδ	delta
0.445670	2.631736
0.142227	1.387481
0.514530	3.269866
0.333201	2.153782
0.182651	1.923823
-0.050389	0.704855
0.382998	2.415450
(average: 0.278698)	(average: 2.069571)

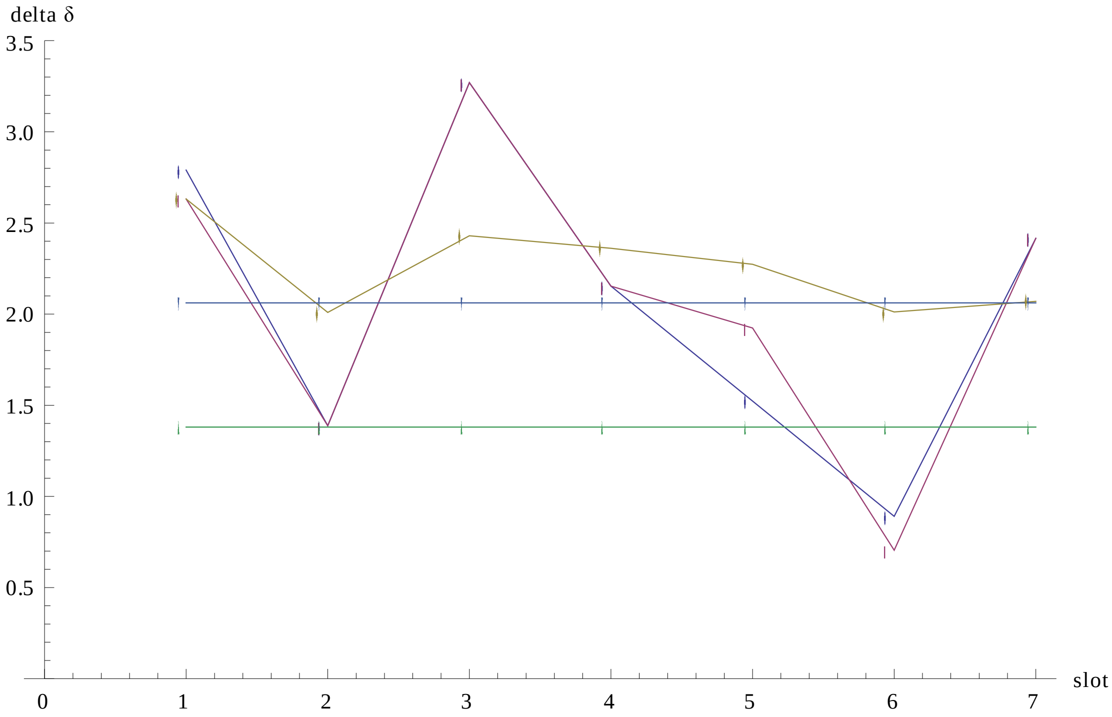

Chart 17: Graphic representation of δ (delta) values for "*cd74a8342d25d090 9098047d8e656cfc*" before (purple line) and after the adjustment (dark blue –circle- line) as well as average (light blue –triangle- line) and running average (olive line) of adjusted values and total average delta (green line).

Similarly to popular category keyword adjustment, as far as the Zipf distribution is concerned, this adjustment introduced no significant improvement.

Chapter Overview

In this chapter we performed an analysis of Overture's dataset. As the dataset was anonymized, it offered no possibility to retrieve valuable information about data's nature. As a result our analysis was in base of Click Through Rate estimation, processing rank (slot awarded), impressions and clicks. We examined slots 1 to 8. However, as we consider that the first 5 slots may carry more information due to their, comparatively to slots 6-8, larger volume of impressions and clicks, we applied running averages which were proved to be close to the delta value proposed by Feng, Bhargava and Pennock. We applied different distribution models and our findings indicated that the geometric distribution may fit real data very well in some cases, i.e. when referring to the entire dataset, while in other cases smaller or larger deviations may occur (popular keyword and category analysis). On the other hand Zipf's Power Law, which was tested as well, was proved to be an inappropriate and inacceptable model, as it introduced considerable deviations and resulted to no convergence of r values. It is still an interesting direction to test whether the Zipf distribution can fit real datasets using more elaborate techniques than regression.

Finally, we segregated key-phrases, responding to keyword and category, which presented a relatively high popularity and we checked geometric and Zipf's Law distributions. During the analysis of popular keyword and popular keyword category, we observed an inconsistency of the expected, according to the distribution model, behavior of the dataset. CTR appeared not follow a monotonically descending rate; there were fluctuations instead. Due to that we proceeded to adjustment of the data, which, however, did not improve significantly the outcome of our analysis.

4

Bibbing Strategies in the Generalized Second Price Auctions

Bidding Strategies

Once the search engine has chosen a mechanism, most commonly GSP, it is then the advertisers' turn to establish the principles of the game. The main question which arises for the advertisers is the difficulty to decide the way they should place a bid and its price. As we have already mentioned, truthful bidding is not a dominant strategy under the GSP mechanism, which gives rise to strategic behavior by the bidders in order to increase their utility as well as their revenue (Edelman B., Ostrovsky M., 2007). Furthermore, even when bidders try to profit by lying, viewing the process as a repeated game, it is not obvious whether the game will converge to a better state for them.

We examine sponsored search auctions as one-shot games of complete information, in which the players’ valuations per click and click-through-rates are of common knowledge. A typical justification for such an approach is the abundance of information in the system, since the advertisers have ample opportunity to explore, submit and resubmit bids at will (Vorobeychik Y., Reeves D. M., 2007).

Defining one’s bidding strategy is rather complex and challenging given the fact that even though the existence of a plethora of Nash equilibria is scientifically verified, it is not a priori clear whether any of these equilibria are actually reached in real keyword auctions. Hence, advertisers often end up assigning their bidding campaign to consultants or other companies, specializing in such campaigns (Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B., 2010).

In this section we will review some of the existing bidding schemes and study their main properties.

Greedy strategies

It is a frequent phenomenon that there is a repetitive model for auctions for the same keyword. Thus, a natural approach to bidding is to use the past as a prediction for the future. Hence, if an advertiser, according to statistical evidence, assumes that other players' bids will not change in the next round, the best choice for him is to bid in a way that will make him win the slot which maximizes his own utility (or, in case winning a slot leads to negative utility, to refrain from bidding). Therefore, we can define the class of greedy bidding strategies as the group of strategies in which an advertiser n chooses his bid for the next round so as to maximize his utility, assuming that the bids of other bidders remain fixed at their values, which were given to them in the previous round.

In the majority of auctions, a range of bids which maximize utility, given the bids of the other players, may be available. Specifying further how to choose the best bid from this range gives rise to various greedy strategies. Supposing that advertiser n, maximizes his utility by acquiring slot s. He can achieve this either by submitting the smallest possible value or try and push the other bidders' payments as high as possible and hence submit the maximum bid that will guarantee slot s and not slot $s-1$. The first strategy is usually referred to as *altruistic bidding,* as in that case the bidder above slot s will pay the smallest possible amount, while the second one is referred to as *competitor busting*. Finally, a more balanced approach is to bid somewhere in the middle so as to still push prices up but without running the risk of paying more than expected if one of the other bidders changes his bid. To become more formal, let $p_s(n)$ be the price that player n has to pay when he bids so as to win slot s, given other players bids. (Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B., 2010)

Cary et al (Cary M., Das A., Edelman B., Giotis I., Heimerl K., Karlin A. R., Mathieu C., Schwarz M., 2007) introduce and study the following greedy bidding strategies: Balanced Bidding (BB), Restricted Balanced Bidding (RBB), Altruistic Bidding (AB), Competitor Busting (CB).

More specifically, in *Balanced Bidding (BB),* bidder n first targets the slot s_n^* that maximizes his utility, i.e., $s_n^* \in arg\, max_s \{\theta_s(u_n - p_s(j))\}$. Given the desired slot, he then chooses his bid b for the next round so as to satisfy:

$$\theta_{s_n^*}\left(u_n - p_{s_n^*}(j)\right) = \theta_{s_n^* - 1} \left(u_n - b\right) \quad (9)$$

The notion behind this is that player n should bid high enough so as to push the prices paid by his competitors up but at the same time it should not be the case that his utility decreases if a competitor bids just below b and n ends up at the higher slot $s_n^* - 1$.

Restricted Balanced Bidding (RBB) is based on the same intuition as BB. However bidder n looks only at slots which have lower CTR than his current slot. If his current slot is s_n, then he first targets the slot $s_n^* \in arg\,max_s \{\theta_s(u_n - p_s(j)) : s \geq s_n\}$. He then chooses his bid b, according to s_n^*, in order to satisfy equation (10).

In *Altruistic Bidding (AB)*, bidders are trying to bid in such a way that will not lead to overcharging other players. In this context, they bid just what is necessary to get the desired slot. Hence the slot s_n^* is selected according to BB principles, but the bid b is chosen equal to $\min\{u_n,\ p_{s_n^*}(j) + \varepsilon\}$, where ε is a small positive quantity.

Finally, *Competitor Busting (CB)*, can be considered as the "opposite" strategy of AB. Here, bidders are simply trying to push prices as high up as possible in order to render their competitors busted. Again s_n^* is selected as in BB but then the bid b is set to $\min\{u_n,\ p_{s_n^* - 1}(j) - \varepsilon\}$. This strategy has been observed in practice and is also referred to as anti-social or vindictive bidding (Brandt F.,Weiß G., 2001), (Y. Zhouand, R. Lukose, August 2007), (Markakis E., Telelis O., 2010) .

Unfortunately, not all of the above strategies converge to some steady state and bidding cycles may appear. BB strategy always converges in an asynchronous setting, however there is no guaranteed convergence in synchronous bidding. The RBB strategy is designed so that it has the same unique fixed point as BB (at which players bid according to the VCG equilibrium), so it is preferable than BB when it comes to asynchronous setting. Convergence issues for AB and CB are much more serious than for BB as, in general, AB and CB strategy do not have a fixed point! The only fixed point for CB is when all players are bidding their values and those bids happen to be a Nash equilibrium for the GSP strategy. For more on greedy bidding strategies and their convergence see (Cary M., Das A., Edelman B., Giotis I., Heimerl K., Karlin A. R., Mathieu C., Schwarz M., 2007), (Markakis E., Telelis O., 2010).

The properties of AB and CB have also been studied further, especially since CB is often encountered in practice. The pricing mechanism embedded in GSP appears to be prone to the Competitor Busting phenomenon, since an advertiser may raise his bid, thereby increasing the price paid by his competitor for the next higher slot, while suffering no consequences as to the price he is paying. There has been a proposal of a new pricing rule, under the name Penalized Second Price (PSP), in order to alleviate CB. According to this rule, the price paid by each advertiser is a linear combination of its own bid as well as of the next lower bid. With PSP an advertiser pays the consequences of its own aggressive strategy. (Grillo A., Lentini A., Naldi M., Italiano F. G, May 2010).

Experimental Analysis of Bidding Strategies

In the context of our experimental analysis of bidding strategies we decided to implement a new bidding strategy. As a result we co-operated with dr. Orestis Telelis, who provided us with a C++ code he had already developed. The code was composed of two individual parts. The first one (GEN.cc) was a bidding data generator, while the second one (ADSTRAT.cc) implemented some of the most popular bidding strategies.

Code Received

Concerning the C++ files we received, as far as the GEN.cc is concerned, the user could give the desired number of bidders (players) and available slots as well as the minimum starting valuation and minimum starting CTR. The outcome, which included a series of valuations, click-through-rates and other information regarding an auction, was used as an input file for the ADSTRAT.cc.

As far as the ADSTRAT.cc file is concerned, the user has a plethora of options, including the selection of a bidding strategy, between Competitor Busting and Mixed Bidding, bidding direction, overbidding option and the number of experiments to be performed. The outcome provided the users with a list of the slots and the corresponding CTR, as well as a list of the players (bidders), the slot each one was awarded, if awarded, their valuations, bids and utility. Finally the social welfare and the search engine’s revenue were computed too.

New Bidding Strategy Implementation

The existence of strategic behavior in the majority of sponsored search auctions and the problems related to it, lead us to consider the possible format of the new bidding strategy. The alternative implementation of Competitor Busting, proposed by Markakis and Telelis in their recent paper (Markakis E., Telelis O., 2010) was considered to fit the goals of the experiment. However, the strategy we finally decided to implement involved the following:

We have 8 slots, 16 competing bidders and click-through-rates following geometric distribution with δ=1.4. The valuation could be defined according to two different scenarios. In the first one, minimum starting valuation is uniformly at random assigned in the interval [10, 100]. In the second one, we have a "rich-poor" mode were valuations are assigned to every individual bidder according to the following rule:

$$v_i = \begin{cases} uniformly\ at\ random\ [50, 100], \text{for the first 5 (out of 16) players} \\ uniformly\ at\ random\ [10, 60], \text{for the 11 remaining players} \end{cases}$$

The auction is organized in periods, where each period is consisted of 4 rounds. In every round, 4 bidders submit their bids, so that in every round we have a different four-bidder group. This means that in the first round we have 4 bidders out of the total 16, in the second round 4 bidders out of the rest 12, in the third round another 4 bidders out of the rest 8 and finally in the fourth round the last 4 bidders.

After the first assignment of the slots, the first period starts. Each one of the 4 bidders selected, checks the directly "neighboring" slots for a possible improvement in his utility. If there is no improvement of the utility, then he checks his neighbors' neighboring slots and so on, until no other slot is available, or until an improvement of the utility occurs. For example, given a bidder at slot 4, he first looks at slots 3 and 5, then slots 2 and 6, following slots 1 and 7, and then only slot 8. If a slot that improves one's utility is found, then the "slot search" stops and the bidder increases his bid, so that he wins this slot[3]. When a bid changes, we compute the convergence between previous and current bids. This applies to every individual bidder. At the end of the period, when all players have reassessed their utilities, we check if there is a

---

[3] We consider the minimum increase equal to 0.01.

convergence of bids. Moreover we compare the search engine's revenue in comparison to the previous period.

Simulations

We produced 100 instances of input files with the following parameters: 8 slots, 16 bidders, click-through-rates as described above, and minimum starting valuation uniformly at random [10, 100] as well as 100 instances of input files with the following parameters: 8 slots, 16 competitor bidders, click-through-rate equal to 1.4 and valuations according to rich-poor mode.

We run our new strategy as following. 50 instances of each of the two "valuation selection" groups run under the Competitor Busting mode, while the other 50 plus 50 instances run under the Mixed mode. We consider 100 periods for each test. At the end of these periods, we evaluate the convergence of bids as well as the total revenue of the search engine.

Findings

Simulations suggested the following:

There is no significant difference between experiments run under competitor busting and mixed mode, or between randomly selected minimum valuation and rich poor valuation mode.

Bidders did not tend to change their slot. At the end of each period (4 rounds) the average number of players changing their slots is 4. The players tended mostly to switch to slots which were above their own at a percentage of 68.29%, and in the great majority (82.07%) it is directly neighboring slots switches that took place. However the final distribution of slots was often identical to the original one, as mutual slot switches have been observed.

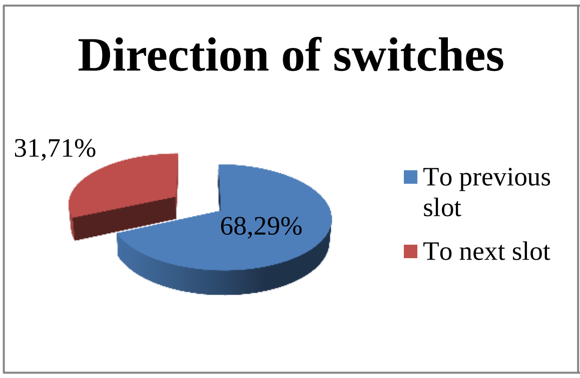

Chart 18: Slot switching distribution, regarding direction of switch

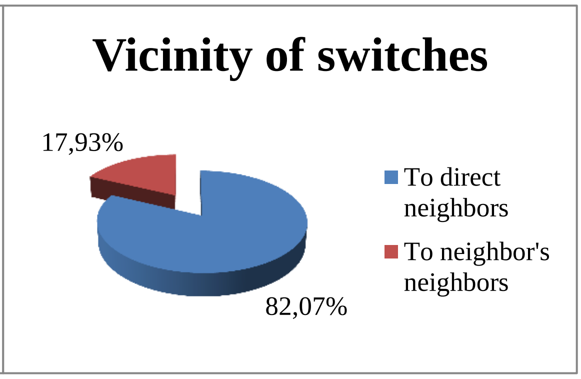

Chart 19: Slot switching distribution, regarding vicinity of slots

As seen in the following chart, slot switches were mostly concentrated at slots 3, 4, 5 and 6. No player, who has not been awarded a slot, apart from the one at slot 9[4], switched to a winning slot[5]. An interesting observation was that many players, in a considerably high percentage of the experiments, mutually switched slots, in a way that, at the end of a period they had the same slots they were originally awarded.

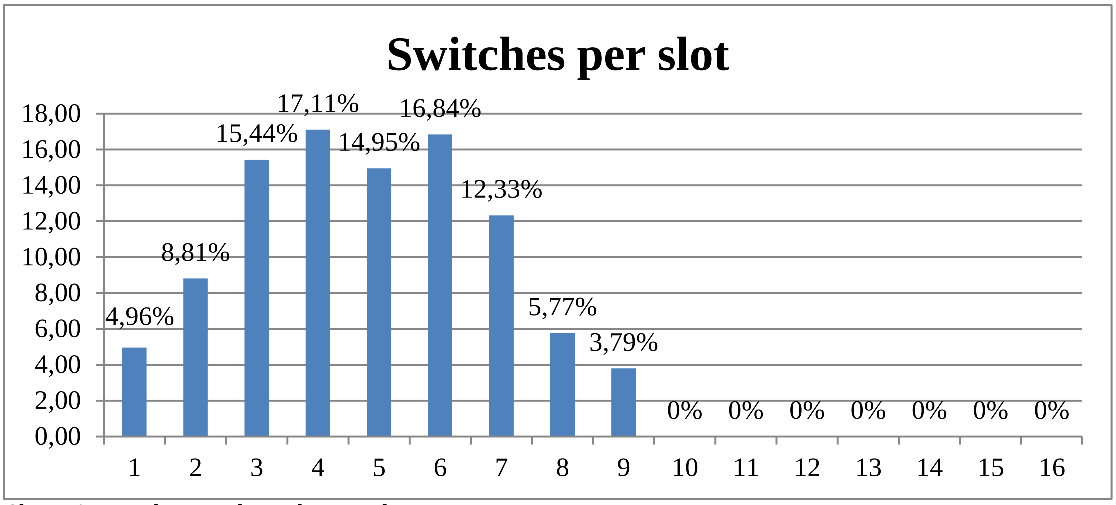

Chart 20: Distribution of switches per slot

As far as convergence of bids is concerned, we have observed that it is quickly reached, after an average number of 8 periods. The first slot switches, introduce significant bid differences, as expected, however after a few periods these differences tend to be eliminated, and convergence is succeeded. The experiments have shown that as far as search engine's revenue is concerned, there are no significant deviations observed from period to period, despite slight increases or decreases. Social welfare remains mostly fixed.

To sum up, our experiments show than convergence of bids is succeeded as after a number of periods no bidder has an incentive to switch his slot.

[4] As bidder at slot 9, we denote the bidder who has been ranked 9^{th}, and as a result has not been awarded a slot.
[5] Slots 1 to 8 are considered to be winning slot, as they "win" a place at the auction

5

Conclusions – Further Research

In this Thesis, we have presented an overview of sponsored search auctions, the ranking and pricing mechanisms used. We focused on the fundamental notion of the Click-Through-Rate, we examined its distribution, and we tested the fit of proposed CTR distribution models in real data, which were acquired as part of the Yahoo! Webscope program. We examined the dataset as a whole, as well as keywords, which we identified as popular, in terms of impressions. The CTR Exponential Decay model was proved to fit real data satisfyingly well, while Zipf's Power Law Distribution model introduced considerable deviations. Finally, we presented a new bidding strategy, were bidders pursue to improve their utility and we examined the convergence of bids. Simulations suggested that bidders tend not to change their acquired slots and in case they do, convergence επέρχεται after a short period of iterations.

We believe sponsored search advertising is a promising area for future research. Thus, a series of other CTR Distribution models should also be tested, with emphasis to real datasets, in order to promote the most elaborate mechanisms for both the bidders and the search engines. In this context, more empirical datasets, which would ideally provide information about the nature of keywords too, should be publicly available. Last but not least, theoretical analysis should be accompanied by further experimental analysis, which would provide a complete overview and lead to the progress of the field.

References

Microsoft and Yahoo seal web deal. (2009, July 29). Retrieved December 02, 2010, from BBC News UK: http://news.bbc.co.uk/2/hi/business/8174763.stm

Aggarwal G., Feldman J., Motwan R. (2006). Truthful Auctions for Pricing Search Keywords. *ACM Conference on Electronic Commerce,* (pp. 1-7).

Beest, M. V. (n.d.). *The History of Web Advertising.* Retrieved December 02, 2010, from eHow.com: http://www.ehow.com/about_5290228_history-advertising.html#ixzz17Wc23fpq

Blumrosen L., Hartline J. D., Nong S. (2008). Position Auctions and Non-uniform Conversion Rates. *Workshop on Ad Auctions*.

Borgers T., Cox I., Pesendorfer M. (2007). Equilibrium Bids in Sponsored Search Auctions: Theory and Evidence.

Brandt F.,Weiß G. (2001). Antisocial Agents and Vickrey Auctions. *Intelligent Agents 8th International Workshop (ATAL),* (p. 335{347).

Cary M., Das A., Edelman B., Giotis I., Heimerl K., Karlin A. R., Mathieu C., Schwarz M. (2007, June 11-15). Greedy Bidding Strategies for Keyword Auctions. *8th ACM Conference on Electronic Commerce.* San Diego, CA: ACM.

Drosos D., Markakis E., Stamoulis G. D. (2010). Budget constrained bidding in sponsored search auctions. Under preparation

Easley D., Kleinberg J. . (2010). Sponsored Search Markets. In *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press.

Edelman B., Ostrovsky M. (2007, February). Strategic Bidder Behavior in Sponsored Search Auctions. *Decision Support Systems, 43*(1), 192-198.

Edelman B., Ostrovsky M., Schwarz M. (2007, March). Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords. *The American Economic View, 97*(1).

Edelman B., Schwarz M. (2010). Optimal Auction Design and Equilibrium Selection in Sponsored Search Auctions.

Feng J., Bhargava H. K., Pennock D. M. (2007). Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms. *INFORMS Journal on Computing, 19*(1), 137-148.

Grillo A., Lentini A., Naldi M., Italiano F. G. (May 2010, May 11-14). Penalized Second Price: A New Pricing Algorithm for Advertising in Search Engines., (pp. 207-214). Montreal, QC, Canada.

HitBox. (n.d.). Retrieved from http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4

Jehiel P., Moldovanu B. (2005, October). Allocative and Informational Externalities in Auctions and Related Mechanisms. (W. N. Richard Blundell, Ed.) *In The Proceedings of the 9th World Congress of the Econometric Society*, 102-135.

Jordan P. R., Wellman M. P. (2010). Designing an Ad Auctions Game for the Trading. 146-162.

Lahaie S. (2006). An Analysis of Alternative Slot Auction Designs for Sponsored Search. In C. J. Feigenbaum J. (Ed.), *In the proceedings of the 7th ACM Conference on Electrinic Commerce* (pp. 218-227). Ann Arbor: Association for Computer Machinery.

Linden G., Meek C., Chickering M. (2006, July 6). The Pollution Effect: Optimizing Keyword Auctions by Favoring Relevant Advertising. *Fifth Workshop on Ad Auctions*.

Maille P., Markakis E., Naldi M., Stamoulis G. D., Tuffin B. (2010, October). An Overview of Research on Sponsored Search Auctions. Under review.

Markakis E., Telelis O. (2010). Discrete Strategies in Keyword Auctions and their Inefficiency for Locally Aware Bidders. *In Proceedings of the 6th international Workshop on Internet and Network Economics (WINE).*

Richardson M., Dominowska E., Ragno R. (2007, May 8-12). Predicting Clicks: Estimating the Click-Through Rate for New Ads. *In the proceedings of 16th*

International World Wide Web Conference Committee (IW3C2), (pp. 521-530). Banff, Alberta, Canada.

SEMPO, E. i. (2010). *State of Search Engine Marketing Report 2010.*

Varian H. R. (2007). Position auctions. *International Journal of Industrial Organization*, 1163–1178.

Vorobeychik Y., Reeves D. M. (2007). Equilibrium Analysis of Dynamic Bidding in Sponsored Search Auctions. *3rd International Workshop on Internet and Network Economics (WINE).*

Wikipedia "Online Advertising". (n.d.). Retrieved 11 29, 2010, from Wikipedia - the Free Encyclopedia: http://en.wikipedia.org/wiki/Online_advertising

Y. Zhouand, R. Lukose. (August 2007). Vindictive Bidding In Keyword Auctions. *In the Proccedings of 9th International Conference on Electronic Commerce* (pp. 141-146). Minneapolis: ACM.

Yahoo! Webscope dataset ydata-ymusic-user-artist-ratings-v1_0. (n.d.). Retrieved from http://research.yahoo.com/Academic_Relations