The Economics of Internet Search

Hal R. Varian

November 2006

Abstract

This lecture examines some of the economic issues related to Internet search engines. It was delivered as the 2007 Angelo Costa Lecture in Rome.

1 Introduction

Search engines are one of the most widely used Internet applications. According to Fallows [2005] "Search engines are highly popular among Internet users. Searching the Internet is one of the earliest activities people try when they first start using the Internet, and most users quickly feel comfortable with the act of searching." The 2005 report indicates that 84% of internet users have used search engines and, on a given day, 56% of those online use a search engine.

Not only are search engines widely used, they are also highly profitable. Their primary source of revenue comes from selling advertisements that are related to the search queries. Since users tend to find these ads to be highly relevant to their interests, advertisers will pay well to place them. Since marginal costs are very low for search engines, profit margins tend to be high.

Online advertising is, by its very nature a scale intensive business. A good ad clickthrough rate might be 3% and a typical conversion (purchase) rate might also be around 3%. This implies that fewer than one out of a thousand people who see the ad actually buy the product being advertised. Despite this seemingly low yield, search engine ads are one of the most effective forms of advertising. TV ads or newspaper ads are significantly less effective since a much smaller fraction of those who see an ad actually purchase the product being advertised.

Since the probability of purchase is low, even when ads are relevant, one has to reach a large audience to have any hope of selling a product. Hence new search engines who hope to become economically successful have to pay large fixed costs to build the scale necessary to serve enough ads to cover those entry costs.

On the demand side, user switching costs for search engine users are very low: the competition is just a click away. Fallows [2005] indicates that 56% of search engine users use more than one search engine. Hence, we can expect to see robust competition for users among the incumbent search engines.

Not only are users not exclusively tied to a single search engine; neither are advertisers. Typically advertisers will "follow the eyeballs" and advertise wherever there are enough potential customers to warrant investment in the industry.

These characteristics—high fixed costs, low marginal costs, the requirement of a mass market, low switching costs, and an advertiser supported business model—means that the likely market structure will be one with a few large competitors in in a given country or language group.

The equilibrium market structure might be similar to that of national newspapers or news magazines: a few large providers, supported mainly by advertising with continuous competition for new readers. There are no significant network effects or demand-side economies of scale that would drive the market to a single supplier.

I will argue later that the most important economic factor determining search engine success is learning-by-doing (Arrow [1962]). Because of the low user switching costs, search engines have to continually invest in improving both their search and their monetization. Though this could be said to be true of virtually any product, continuous improvement is particularly important in online products since pace of experimentation and implementation is particularly rapid.

Though there are dozens of search engines available, the big three in terms of market share are Google, Yahoo and MSN. I will mostly discuss Google, since I am most familiar with its practices, but the other search engines tend to use similar business models.

2 Two-sided matching

First, what does Google do? The answer, I claim is that Google is a "yenta" — a traditional Yiddish word for "matchmaker". On the search side, it matches people who are seeking information to people who provide information. On the ad side, it matches people who want to buy things to those who want to sell things.

From an economics perspective, Google runs a "two sided matching" mechanism. This subject has a long history in economics, starting with the classical *linear assignment problem* which seeks to find a matching of partners that maximizes some value function. Not surprisingly, the mathematical theory of the assignment problem turns out to be closely related to the Google ad auction.

The need for efficient matching of users and content is apparent: the growth of content on the Internet has been phenomenal. According to the `netcraft.com` there are about 100 million web servers. Obviously, the more content that is on the web, the more important it is to have good search engines. The web without search engines would be like Borges's universal library with no card catalog.

In this talk I will briefly discuss the history of information retrieval, emphasizing some of the points of interest to economics. I will then describe the evolution of the business model to support online search engines, and conclude by sketching some of economic aspects of the Google ad auction.

3 A brief history of information retrieval

Almost as soon as textual information was stored on computers researchers began to investigate how it could be easily retrieved. Significant progress was made in the 1960s and operational systems were widely available by the 1970s. The field was reasonably mature by the 1990s, with the primary users being professional librarians and researchers.[1]

By the early 1990s most of the low-hanging fruit had been harvested and intensive users of information retrieval technology were worried that technological progress was grinding to a halt. This concern led to the creation in 1992 of TREC (Text Retrieval and Extraction Conference) by DARPA.

DARPA compiled training data consisting of many queries and many documents along with a 0-1 indicator of whether or not the document was relevant to the query. These relevance indicators were determined by human judges. Research teams then trained their systems on the TREC data. Subsequently, TREC provided a second set of data for which the research teams tried to forecast relevance using their trained systems.

Hence TREC provided a test collection and forum for exchange of ideas and most groups working in information retrieval participated in TREC. (See TREC8 [2000].) Having a standard base for comparing different algorithms was very helpful in evaluating different approaches to the task.

Though search engines use a variety of techniques, one that will be very familiar to economists is logistic regression. One chooses characteristics of

[1] See Lesk [1995].

the document and the query and then tries to predict the probability of relevance using simple logistic regression. As an example of this approach, Cooper et al. [1993, 1994] used the following variables:

- The number of terms in common between the document and the query.
- Log of the absolute frequency of occurrence of a query term in the document averaged over all terms that co-occur in the query and document.
- Square root of the query length.
- Frequency of occurrence of a query term in the collection.
- Square root of the collection size.
- The inverse collection frequency, which is a measure of how rare the term is in the collection.

Other systems use different variables and different forms for predicting relevance, but this list is representative.

By the mid 1990s it was widely felt that search had become commoditized. There were several algorithms that had roughly similar performance and improvements tended to be incremental.

When the web came along in 1995, the need for better Internet search engines became apparent and many of the algorithms developed by the TREC community were used to address this need. However, the challenge of indexing the web wasn't as compelling to the IR community as one might have thought. The problem was that the Web wasn't TREC. TREC had become so successful in defining the information retrieval problem that most attention was focused on that particular research challenge, to the exclusion of other applications.

The computer scientists, on the other hand, saw the web as the problem du jour. The NSF Digital Library project and other similar initiatives provided funding for research on wide scale information retrieval.

The Stanford computer science department received one of these Digital Library grants and two students there, Larry Page and Sergey Brin, became interested in the web search problem. They developed the PageRank algorithm—an approach to information retrieval that used the link structure of the web. The basic idea (to oversimplify somewhat) was that sites that had a lot of links from important sites pointing to them were likely to contain relevant information.[2]

PageRank was a big improvement on existing algorithms and Page and Brin dropped out of school in 1998 to build a commercial search engine: Google.

The algorithm that Google now uses for search is proprietary, of course. It is also very complex. The basic design combines PageRank score with and information retrieval score. The real secret to Google's success is that they are constantly experimenting with the algorithm, adjusting, tuning and tweaking virtually continuously.

One of the tenets of the Japanese approach to quality control is *kaizen* which is commonly translated as "continuous improvement." One reason for the rapid pace of technological progress on the web is that it is very easy to experiment—to use a new search algorithm for one query out of a thousand. If the new algorithm outperforms the old one, it can quickly be deployed. Using this sort of simple experimentation, Google has refined its search engine over the years to offer a highly refined product with many specialized features.

Google is hardly the only online business that engages in kaizen; Amazon, eBay, Yahoo and others are constantly refining their web sites. Such

[2]See Langville and Meyer [2006] for a detailed description of the mathematics behind PageRank.

refinements are typically based on systematic experimentation and statistical analysis, as in the traditional quality control practice.

4 Development of a business model

When Brin and Page started Google they did not have a business model in mind. At one point they offered to sell the PageRank algorithm they used to Yahoo for $1 million. When Yahoo turned them down, they thought about selling intranet search services to companies.

Meanwhile, a company in Pasadena named GoTo.com was starting to auction off search results. In 1999 they filed U.S. Patent 6,296,361 (granted July 31, 2001) which described the idea of auctioning search results.[3]

Auctioning search results didn't work very well, since willingness to pay for placement is not a very good indication of relevance to users, so GoTo eventually adopted a new business model in which they auctioned off advertisements to accompany what they referred to as the "algorithmic" search results. At about the same time they changed their name to Overture.

Two Google employees, Salar Kamangar and Eric Veach, watched what Overture was doing and decided they could improve upon it. During the Fall of 2001 they developed the Google Ad Auction.

In their model ads were ranked by a combination of bids and estimated clickthrough rate. Since bids are expressed in units of cost/click and the clickthrough rate is clicks/impressions, this means that ads are ranked by cost per impression. The idea was to put the ads that have the highest expected revenue in the best positions—i.e., the positions where they would be most likely to receive clicks.

Just as a a firm cares about price *times* quantity sold, a search engine should care about the price per click *times* the number of clicks expected to

[3] I am told that this idea may have been stimulated by a student who took Charlie Plott's course in experimental economics at Cal Tech. So economists seemed to have played a role in this auction design from an early stage!

be received—since that is the total revenue from showing the ad. Of course, this requires a way to estimate the probability of a click, a nontrivial task. I will discuss how this is done below.

Google soon realized that a first-price auction (where advertisers paid their bid amount) was not attractive since they would want reduce their bid to the lowest amount that would retain their position. This constant monitoring of the system would put a significant load on the servers, so Google decided to automatically set the price paid to be equal to the second highest bid—since that's what the advertisers would want to do anyway. This choice had nothing to do with Vickrey auctions—it was primarily an engineering design decision.[4]

Initially the Google ad auction only applied to the ads appearing on the right-hand side of the page, with the top ads (the best performing area) reserved for negotiated pricing by a sales force. Eventually it became clear that the prices generated by the auction were more appropriate than those generated by negotiation, so Google switched to using an auction for all ads displayed.

5 The Google ad auction

The Google ad auction is probably the largest auction in the world, with billions of auctions being run per week. It turns out also to have a very nice theoretical structure as described in Edelman et al. [2005] and Varian [2006].

There are several slots where advertisements can go, but some receive more clicks than others. In equilibrium, each bidder must prefer the slot it is in to any other slot. This leads to a series of "revealed preference" relations which can be solved for equilibrium bidding rules. Conversely, given some observed bids, one can invert the bidding rules to find out what values the

[4]GoTo.com experimented with a first price auction for some time and found it to lead to unstable behavior. Zhang and Feng [2005] and Zhang [2005] document and model this phenomenon.

advertisers place on clicks.

To see how this works, consider an bidder who is contemplating entering a keyword auction. The current participants are each bidding some amounts. Hence the new bidder thus faces a "supply curve of clicks." As it bids higher it will displace more of the incumbent bidders, leading to a higher position and more clicks.

In choosing its bid, the advertiser should consider the *incremental cost per click:* how much more money it will have to spend to get additional clicks. If the incremental cost per click is less than the value per click, the advertiser should increase its bid; if the incremental cost per click is less than the value per click, it should decrease its bid. In equilibrium the incremental cost of moving up one position should exceed the bidder's value per click, but the incremental savings from moving down one position should be less than the bidder's value per click.

This has the implication that in equilibrium the incremental cost per click should be increasing in the click-through-rate. Why? Suppose it decreased in moving from one position to the next. Then there was some bidder who purchased expensive clicks but passed up cheap ones, contradicting the assumption of equilibrium.

Furthermore, since the value per click should be bounded by the incremental cost per click in equilibrium, the observed incremental costs allow us to infer valuable information about the bidders' values. In practice, incremental cost per click seems to give a plausible estimate of click value.

However, it is important to note that there is still a certain indeterminacy of equilibrium. The requirement that each agent prefers its position to other possible positions does not pin down a unique outcome. Rather it determines a range of equilibrium bids. Two particularly interesting equilibria are the ones that yield the maximum and the minimum revenue for the search engine.

6 VCG pricing

The Google ad auction is one way to auction off ad positions, but there are other ways that can be considered. One defect of the current auction is that each advertiser has to compare its incremental costs to its value, and those incremental costs depend on other bidders' choices.

As it happens there is another auction-like mechanism that does not have this defect: the Vickrey-Clarke-Groves mechanism (VCG). In the VCG mechanism: 1) each agent reports a value; 2) the search engine assigns agents to slots to maximize total value of the assignment; 3) each agent a then pays a charge equal to the total value accruing to the other agents if a is present minus the total value accruing to the other agents if a is absent. Thus each agent pays an amount equal to the cost that it imposes on the other agents.

It can be shown that for this mechanism, each agent should report its true value, regardless of the reports of the other agents. Leonard [1983] was the first to apply this mechanism to the classic assignment problem. A few years later Demange and Gale [1985] showed that this mechanism results in the same payments as the minimum-revenue Nash equilibrium of a market equivalent to the second-price ad auction.[5]

There other nice properties of the VCG auction. For example, Krishna and Perry [1998] show that the VCG mechanism maximizes the search engine's revenue across all efficient mechanisms. Despite the apparent advantages of VCG, it has not as yet been deployed by any of the major search engines.

7 The importance of competition

It is widely recognized that revenue realized in an auction depends critically on how much competition there is in that auction. Klemperer [2002] de-

[5] I am simplifying the actual result for ease of exposition; see Varian [2006] for the details.

scribes the case of the June 2000 auctions for mobile phone licenses in the Netherlands where there were 5 licenses and 6 bidders. One bidder threatened another with legal action if it continued bidding inducing it to drop out, leaving 5 bidders for 5 licenses—not much competition! In fact the auction raised less than 30% of what the Dutch government had forecast.

The same principle holds true for the position auction: revenue doesn't really take off until there is competition.

In the Google auction there are 8 slots for ads on the right-hand side of the page and up to 3 slots on the top of the page. As mentioned earlier, the ordering of ads is determined by bids and click-through rates , but the ads that are "promoted" (moved the the top of the page) have to satisfy some additional criteria involving ad quality.

To simplify a bit, if an auction has fewer bidders than available slots, or just enough bidders to fill the available slots, we say it is "undersold." If it has more bidders than slots we say it is "oversold." If an auction is undersold, the price paid by the last bidder on the page is the reservation price, which we will take to be 5 cents.[6] If the page is over sold, the price paid by the last bidder on the page is determined by the bid of the first excluded agent, which can easily be at least 10 times higher than the reserve price.

Consider a simple example where all bidders have the same value v and the reserve price is r. Let p_s be the price paid for slot s and let x_s be the number of clicks that slot s receives. If the page is undersold, each bidder has to be indifferent between paying p_s and receiving x_s clicks versus paying r and receiving x_m clicks, where m is the last ad shown on the page. This implies

$$(v - p_s)x_s = (v - r)x_m$$

or

$$p_s x_s = v(x_s - x_m) + r x_m.$$

[6]The reservation price actually depends on ad quality as well.

This equations says that the expenditure on slot s has to be the expenditure on the last slot plus the *incremental* value of the clicks in position s.

On the other hand, suppose the page is oversold so that there is at least one excluded bidder with value v. Then each bidder has to be indifferent between what it is paying and the profit from being excluded — which is zero. This gives us

$$(v - p_s)x_s = 0,$$

which implies $p_s = v$.

Note the big leap in revenue in going from a partially sold page to a over sold page. In the first case, everybody is indifferent between being in the slot they are in and being in the worst slot. In the second case, everybody is indifferent between being shown and not shown at all, which means prices are competed up to the equal value.

To drive this point home, consider a simple example.

Suppose that there are 2 slots. The top one gets 100 clicks per day, the second one 80 clicks per day. There are two advertisers, each of whom values a click at 50 cents.

In this model, one advertiser occupies slot 2 and gets 80 clicks per day, for which he pays 5 cents per click = \$4.00 in total spend. The second advertiser occupies the top slot getting 20 additional clicks per day. Competition forces him to pay \$10 more for those clicks than the advertiser in slot 2. Thus he spends \$14 = \$4 + \$10. in total. Total revenue from the two advertisers is \$18.

Now suppose there are 3 advertisers who value clicks at 50 cents each, but there are still only two slots. The Equilibrium bid is now 50 cents per click, there are 180 clicks in total, sot he total revenue from the two advertiser is \$90. The addition of one more advertiser increases revenue from \$18 to \$90!

This example illustrates the important point that over sold pages are far more profitable than partially sold pages not just because there are more bidders, but also because the forces of competition are much stronger.

This point also illustrates the importance of the matching algorithm used for displaying ads. The user enters a "query" and the advertiser buys "keywords." The advertiser can specify "exact match," which means that the ad is only shown if the user's query exactly matches the advertiser's keywords. But it is more common for advertisers to specify "broad match" which means that the query will match various expansions of the keyword such as synonyms and substrings.

The additional ads due to broad match benefits the user and the advertiser, since they make it more likely that the user will click. But they also increase the competition in the auction, raising prices.

8 Ad quality

I have indicated earlier that the ranking used by both Google and Yahoo is based not only on bids, but also on a measure of ad quality. In the simplest case, we can think of ad quality as the predicted clickthrough rate. Google ranks ads by bid times expected clickthrough rate, but where does the estimate of expected clickthrough come from?

Think of a model where the actual clickthrough rate that an ad receives depends on both a position-specific effect (x_p) and an ad-specific effect (e_a). The simplest specification that the clickthrough rate for ad a in position p is given by $e_a x_p$.

Given this multiplicative form, it is relatively easy to estimate the relevant values: simply put random ads in position p to estimate the position-specific effect. Once this is known, you can use the history of clicks on a given ad to estimate the ad-specific effect. One can also use various other predictors to supplement the historical data.

The ranking of ads is based on bids times ad-specific effects: $b_a e_a$. The bid is dollars per click and the ad-specific effect is clicks per impression. Hence $b_a e_a$ is bid per impression: how much the advertiser is willing to pay for its ad to be shown to a user. The advertiser with the highest value for an impression is given the best position: to position most likely to receive a click. The advertiser with the second highest value per impression gets the next best position, and so on.

Hence an ad with a high bid per click could be displaced by an ad with a lower bid if the high-bid ad had a low clickthrough rate. Assigning ads on the basis on $b_a e_a$ maximizes the value of the impressions on the page, leading to an increase in expected revenue.

Just as it is imported to determine which ads to show, it is equally important to determine which ads *not* to show. The reason is that the likelihood of a user clicking on an ad depends on how relevant he or she expects that ad to be. And this expectation depends, at least in part, on what the user's previous experience has been.

Thus showing a "bad ad" can affect users' future propensity to click. Offering a bad ad in a particularly prominent position can be especially costly.

The decision of whether and where to show an ad should depend not just on current ad revenue, but on an estimate of how the ad's relevance will affect future propensities to click. It is possible to model these choices analytically. Showing an ad today brings in a known amount of revenue but also has a probabilistic effect on future revenue by influencing the propensity to click in the future. Modeling these effects leads to a stochastic dynamic programming problem that offers a rationale for current practices and a guide to how they might be refined.

9 Conclusion

Search engines are an example of a two-sided matching model supported by advertising. Not only are they interesting in their own right, but they offer a fertile ground for economic analysis.

During the 1960s and 70s the scientific study of financial markets flourished due to the availability of massive amounts of data and the application of quantitative methods. I think that marketing is at the same position finance was in the early 1960s. Large amounts of computer readable data on marketing performance are just now becoming available via search engines, supermarket scanners, and other sorts of information technology. Such data provides the raw material for scientific studies of consumer behavior and I expect that there will much progress in this area in the coming decade.

References

Kenneth J. Arrow. The economic implications of learning by doing. *Review of Economic Studies*, 29:155–173, 1962.

William Cooper, Aitao Chen, and Fredric C. Gey. Full text retrieval based on probalistic equations with coefficients fitted by logistic regression. *TREC93*, 1993.

William Cooper, Aitao Chen, and Fredric C. Gey. Experiments in the probabilistic retreival of full text dcouments. *TREC 1994*, 1994.

Gabrielle Demange and David Gale. The strategy structure of two-sided matching markets. *Econometrica*, 53:873–878, 1985.

Benjamin Edelman, Michael Ostrovsky, and Michael Schwartz. Internet advertising and the generalized second price auction. *NBER Working Paper*, 11765, November 2005.

Deborah Fallows. Search engine users. *Pew Foundation*, 2005. URL `http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf`.

Paul Klemperer. What really matters in auction design. *Journal of Economic Perspectives*, 16(1):169–189, 2002. URL `http://www.nuff.ox.ac.uk/users/klemperer/design3aweb.pdf`.

Vijay Krishna and Motty Perry. Efficient mechanism design. Technical report, SSRN, 1998. URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=64934`.

Amy N Langville and Carl D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ, 2006.

Herman B. Leonard. Elicitation of honest preferences for the assignment of individuals to positions. *Journal of Political Economy*, 91:461–79, 1983.

Michael Lesk. The seven ages of information retrieval. Technical report, International Federation of Library Associations and Institutions, 1995. URL `http://www.ifla.org/VI/5/op/udtop5/udtop5.htm`.

TREC8. The TREC conferences: An introduction. *NIST*, 2000. URL `http://trec.nist.gov/presentations/TREC8/intro/`.

Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 24(7):1–10, 2006. URL `http://www.sims.berkeley.edu/~hal`.

X. M. Zhang. Finding Edgeworth cycles in online advertising. Technical report, MIT Sloan School, 2005. URL `marketing.wharton.upenn.edu/news/colloquia/michael_zhang.pdf`.

X. M. Zhang and J. Feng. Price cycles in online advertising auctions. In *Proceedings of the 26th International Converence on Information Systems*, Las Vegas, NV, 2005. ICIS.