Online Gatekeepers: Myth or Reality?

Eszter Hargittai
tprc@eszter.com
Sociology Department
Princeton University

In this paper, I look at data on people's online actions to determine the extent to which users rely on and are dependent upon recommendations from commercial sites – especially search engines – for their content selection and destination online. In 2001, Commercial Alert, a non-profit consumer organization filed a formal complaint with the Federal Trade Commission charging that several search engines were "inserting advertisements in search engine results without clear and conspicuous disclosure that the ads are ads." In response, the FTC has recommended clear and conspicuous disclosures on Web sites. Here, I consider how different search engines present their results to users and how users make sense of search engine results. The data are based on in-person observations and interviews with a random sample of Internet users.

Introduction

In 2001, Commercial Alert, a non-profit consumer organization founded by Ralph Nader, filed a formal complaint with the Federal Trade Commission charging that several search engines were "inserting advertisements in search engine results without clear and conspicuous disclosure that the ads are ads." The FTC published its response in July, 2002.

> [T]he staff is recommending that all search engine companies* review their Web sites and make any changes necessary to ensure that:
>
> - any paid ranking search results are distinguished from non-paid results with clear and conspicuous disclosures;
>
> - the use of paid inclusion is clearly and conspicuously explained and disclosed; and
>
> - no affirmative statement is made that might mislead consumers as to the basis on which a search result is generated.
>
> (from http://www.ftc.gov/os/closings/staff/commercialalertletter.htm)

Is the above concern by Commercial Alert a legitimate concern? Is online advertising really deceiving to people? Do users not understand how search engines work and why they are pointed to some content and not other material online? Does it matter which search engine someone uses and who is their Internet service provider (e.g. whether they are AOL users)? How does the choice of a search engine or ISP affect one's online actions and exposure to certain content over other types of online information? I analyze data on the search strategies of randomly selected Internet users to look at the power of portal sites and search engines in allocating user attention to online content. I also draw on interviews with these people to examine their level of understanding of these sites.

The organization of content online

A tremendous amount of information is available on the Web given how easily any person with access to the Internet can post data on his or her own initiative. Attention scarcity leads individual creators of content to rely on online gatekeepers

* 8. This would include the named search engine companies, and other companies providing similar Internet search services to consumers, as well as meta search engines that submit simultaneous search queries to (and display results from) numerous third-party search engines. (footnote in original)

to channel their material toward users. Web services that categorize online information can be considered gatekeepers on the World Wide Web.

Because the barriers to entry are so low on the Web, a huge amount of information floods the market leaving the user overwhelmed with the task of navigating through the vast amount of options offered online. Although there may be numerous high quality and important sites on the Web, there is no guarantee that people will find their way to them. On the Web, the central concern is no longer what is produced, but what consumers hear and know about. Accordingly, gatekeeping activity still occurs, but its location has shifted from the decision about what should be produced to control of what materials get to consumers and what they become aware of. In the online world, the most popular navigational sites are the most prominent gatekeepers. This paper is an empirical study of whether portal sites truly function as gatekeepers to the content of the Web.

Billions of Web pages are available on the Web for public use (Bergmann 2002; Lake 2000). Any individual or organization with the know-how to create a site can contribute content to the public Web. The technicalities of making such content as available to users as the most popular Web sites are more or less the same. However, information abundance still leaves the problem of attention scarcity. Ironically, even people who have recognized the importance of attention scarcity have suggested that any individual will be able to sidestep organizations and corporate packaging in an attempt to receive attention (Goldhaber 1997). In contrast, I emphasize that attention scarcity leads individual creators of content to rely on online gatekeepers to channel their material toward users and leads users to rely on such services to find their way to content on the Web. Web services that categorize online information can be considered gatekeepers on the World Wide Web.

The term 'gatekeeper' refers to points that function as gates blocking the flow of some material while allowing other information to pass through (White 1950). Studies on industries that make cultural products offer a good basis for understanding where the important decisions are made in the process of bringing the products to an audience. Research on books (Powell 1985), news publications (Gieber 1964; Tuchman 1978) and popular music records (Lopes; Peterson and Berger 1975) has explored the role of gatekeepers in influencing the type of cultural

products that are produced and distributed on the market. With previous media, the costs of production were so high that a vitally important gatekeeping step concerned the decision about what products should be produced. Studies have documented the coping mechanisms that firms adopted to deal with the uncertainty of large investments in cultural products (Hirsch 1972). The common theme in all these analyses is that individual creators of cultural products have to go through both producers *and* distributors of their products to get attention on the market. The final link in the distribution chain – supermarket rack jobbers, disk jockeys, movie critics, book review editors – can be a key figure in allocating people's attention to material.

Although there may be numerous high quality sites on the Web, there is no guarantee that anyone will find their way to them. The central concern is no longer what is produced, but what consumers hear and know about. Accordingly, gatekeeping activity still occurs, but now takes place at the level of information exposure. Its location has shifted from the decision about what should be produced to control of what materials get to consumers and what they become aware of. Users with more advanced Web use skills will be less dependent on such gatekeepers and can more easily sidestep them to find information of interest to them.

In order to understand the implications of gatekeeping for the reachability of online content – whether commercial or not-for-profit content, individual or governmental materials – it is important to distinguish between content that is merely present on the Web in contrast to content that users are readily exposed to. To make this distinction, I use the word 'available' to refer to material that exists online and use 'accessible' to denote content that is easily within the reach of Web users. Whereas 'availability' means mere existence, 'accessibility' implies relative ease of reachability.

As the amount of Web content grew exponentially, search engines became increasingly important in sifting through online material. According to one survey, 85 percent of users have ever used a search engine (Pew 2002). Although seemingly neutral, search engines systematically exclude certain sites in favor of others either by design or by accident (Introna and Nissenbaum 2000). Search engines index no more than a small portion of all Web pages and even collectively the largest engines only account for a combined coverage of just a fraction of all information online

(Lawrence and Giles 1999). This suggests that there is great discrepancy between what is physically available on the Web and what information is realistically accessible to users.

Finding one's way online

Undoubtedly, the entry of the private sector into the Internet world encouraged its wide spread and the growth in online content. Search engines and portal sites assist millions of users every day in finding information online. So why is it a potential problem that commercial interests sometimes guide the content selection on popular sites? The concern is that search engines that are guided by profit motives will point people away from the most relevant and best quality sites in favor of those that have paid the highest bids for placement on the results page.

Analyses of large-scale search engine usage data suggest that users mainly rely on the first page of results to a search query. A study analyzing almost one billion queries on the AltaVista search engine showed that in 85 percent of the cases users only viewed the first screen of results (Silverstein et al. 1999). Web users' habits haven note changed much over the years. Another study (Spink et al. 2002) compared data on the use of the Excite search engine from 1997, 1999, and 2001 and found that the mean number of results pages users looked at had decreased over time. The data in this study also showed that the majority of users rely on simple queries without the use of advanced search features outlined above.

These findings suggest that users heavily rely on sites for presenting them with information rather than using sophisticated search strategies to fine-tune their queries. This implies that information prominently displayed on portal sites – whether selected because of high content value or for commercial reasons – has a good chance of being the destination of visitors. If users do not possess advanced know-how about how content is organized and presented to them online then they are especially at the mercy of what content sites decide to feature prominently and make easily accessible to them.

Sites spend significant resources on optimizing their content to show up as results. In fact, an entire industry has sprung up around "search engine optimization" offering advice on how companies and others can best assure that their Web sites climb to the top of search engine results. In contrast, the sites with the most

relevant content may be posted by a non-profit or an individual on his or her own initiative and only appears far down the results list because the owners of such sites do not have the resources to optimize for search engine positioning. So the overall concern due to the prominence of commercial interests on the Web is not that users will unknowingly be roped into purchasing information they could otherwise obtain for free – although this may happen as well – but that they may not find what they are looking for or may miss the best available information because those resources are crowded out by the profit-seeking ventures.

Commercial sites will often rise to the top of result lists despite not having the relevant information. A search on Overture – which is an openly pay-for-placement search engine – for something as specific and non-commercial as the `"museum of modern art"` will yield five commercial results before listing `http://www.moma.org` which is the Museum's own site. And although Overture may not be a widely used search engine, it has deals with several of the most popular search engines to feature its results on their pages (e.g. Yahoo!, MSN, Altavista, Dogpile and Lycos all feature Overture results prominently on their results pages).

How else might users be affected by search engine results? Consider the following scenario. A user is looking for information about political candidates, in particular, she is interested in comparing the views of two presidential candidates about a controversial issue, say abortion. There are thousands of Web sites that describe, critique, and compare political actors. However, a simple search on the candidate's name or using the word `abortion` will not yield any obvious results, rather, it will present the user with hundreds if not thousands of possible links to pages with only one of the two topics.

In this particular case, a user who understands how search queries can be refined through the use of quotation marks (to signal proximity of terms), the use of Boolean operators (to suggest whether terms should all be included in a search or whether some terms should be explicitly excluded) and through the use of multiple terms in a query will likely turn up helpful results almost regardless of the search engine used. A knowledgeable user may type the following into a search box: `bush gore abortion` and quickly find relevant results. A 26-year-old woman whizzed through the tasks by using Google in an informed manner with multiple term queries

(e.g. >`Al Gore views on abortion`< for the political comparison task). An 18-year-old college female also used elaborate queries in some cases (e.g. >`Bush and Gore and abortion`< for the same comparison question), but turned to directory listings on Yahoo! for other tasks such as finding local movie listings.

Nonetheless, even the use of such refined search queries requires additional know-how on the part of the user. Many sites come cluttered with images and text – often in an attempt to make a commercial venture viable – and it sometimes becomes quite challenging to find specific information on a page. Among the one hundred participants in the Web Use Project at Princeton University (http://www.webuse.org), only one ever used the Find function (available in all browsers and on all platforms) to search for a term on a Web page. In the case of this task, looking for the word "abortion" through use of the Find function would have aided many participants. This action can significantly reduce the effort it takes to find specific content on a page yet almost no one uses it.

Even knowledgeable users can get confused and sidetracked. One 56-year-old man who works in business development clicked on the 'Advanced search' feature of Google and then typed >`abortion comparison Bush and Gore`< in the search field labeled "with the exact phrase" which proved to be too nuanced and yielded no results. Confused by this, he clicked on the search button a few more times – without changing anything in the search field – and continued to get no results. After some additional attempts at using that feature of the search engine, he moved on to CNN's site and found the information via their search engine.

This type of task is especially challenging to those people who do not know how to use refined searches. Simply typing in `abortion` will lead to over a million sites most of which are not focused on comparing different presidential candidates' views on the issue. Moreover, those who do not use a search engine at all are completely at the mercy of how directory sites present content and end up taking quite a bit of time clicking through endless list of links refining their topic.

An action typical of those who understand little about the Web and searching is to continuously click on "related search" items after a search result. This leads to yet more search results. If the user simply keeps clicking on these links, she never gets to an actual site. A 44-year-old woman who works in retail repeatedly used this

method to find sites and took many iterations of searches to finally arrive at a page with content of interest. There are also those who never use search engines. These users rely on site recommendations from their Internet service providers.

[In the talk at TPRC, I will present more examples, show some video clips of people's searches to exemplify how users are influenced by what sites present to them and will offer some more aggregated results.]

Conclusion

Data from studies cited above attest to the fact that the majority of Web users regularly depends on recommendations from search engines for their destination. Data from observing people's online actions and collected via interviews suggest that many users do not understand the underlying decisions that go into particular search engine results. Although some users know that there may be commercial decisions underlying the results of search engines, they do not know how to distinguish among them or pinpoint said results. Underlying such confusion is users' lack of knowledge with respect to how content is organized, distributed and presented online more generally. More public understanding of how the Web is organized and how search engines work in general would benefit users' understanding of why they are shown and channeled toward some content and away from other content.