Social Impact of Text Information Systems

Will Hennessy Stephen Herring

Spring 2016

1 The anatomy of a large-scale hypertextual web search engine

1.1 Citation

Brin, Sergey, and Lawrence Page. "Reprint of: The anatomy of a large-scale hypertextual web search engine." Computer networks 56.18 (2012): 3825-3833.

1.2 Summary

In this paper, Larry Page and Sergey Brin introduce their seminal project–the Google search engine. Despite Google's deep penetration into modern vernacular, it helps to first define what it actually is. In the words of Page and Brin, Google is "a large-scale search engine which makes use of the structure present in hypertext...designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems." The last part of this statement is probably the defining characteristic of Google: at the time, many search engines were already running in production, but most of them were highly ineffective. In fact, in November 1997, Page and Brin discovered that only one of the top four commercial search engines was able to find itself.

So what was going wrong? It turned out that many of the search engines already in production were not made to handle the Web of 1997. The number of documents on the Web (and therefore, search indices) was increasing at an astronomical rate, causing search engines to return lots of garbage results that would wash out relevant documents from a search. Diluting relevant results severely impacts a search engine?s effectiveness, since a user?s ability to read results is relatively static–most of the time a user only browses the first ten results of a search before giving up.

To resolve this issue, Page and Brin introduced PageRank: a citation (hyperlink) graph of the Web. Citations are a great way to measure the subjective importance a page–the more something is referenced online, the more likely it is to be important to a user. The reasoning behind this simple: citations are created by people.

Along with PageRank, Page and Brin introduced the idea of anchor text, which uses the text associated with a hyperlink to describe the page linked. This helps Google define pages effectively for two reasons. First, anchor text often defines a page better than the page itself. Second, anchor text helps give relevance to pages which cannot be defined otherwise. These are pages containing content such as videos or images.

2 Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality

2.1 Citation

Van Couvering, Elizabeth. "Is relevance relevant? Market, science, and war: Discourses of search engine quality." Journal of Computer-Mediated Communication 12.3 (2007): 866-887.

2.2 Summary

The author conducted a study to understand the motivating factors that drive search engine producers (SEP) to make changes to the engine. Between November 2002 and May 2004, the author interviewed eleven SEPs, including senior engineers and technical executives who direct future code development. These interviewees worked at every major and minor search engine on the market: Google, Yahoo!, MSN, Ask Jeeves, AOL, and more. Each individual was interviewed over the phone for 1-2 hours in a semistructured, in-depth format. Questions probed for specific instances of change to the search engine and inquired for the motivation behind that change. The text transcript of each interview was categorized to identify themes, from which the author determined two major schemas that motivate the development of search engine technology:

1) The Market Schema

Throughout these interviews, the most common category of motivator was the market schema, which includes revenues, costs, competition, and other business issues. In explaining this motivation, interviewees regarded their search engine as a commercial service competing for users in the marketplace. Thus, the primary motivator for SEPs is financial profit and that metric is linked to search engine quality via its direct correlation with customer satisfaction. Many of the technical changes were developed to increase profit.

2) The Science/Technology Schema

The second most common category of motivator was the science schema, which includes experimentation, measurement, feasibility, and objectivity. This motivator defined quality as relevance, or the ability to answer a user's question, and was defined by data-driven metrics. Many of the changes were developed to improve search result relevance, recall, or precision.

These interviews also revealed a subjective component of the search engines: blacklists, whitelists, and topic-specific weights. This censoring was often dictated by executives to respond to current events, but it is arbitrary and non-scientific. Still, these practices are accepted by the SEP because they strive to boost relevance.

This paper impacts all modern businesses because it describes the convoluted environment in which they compete for clicks. For many companies, the search engine is the primary portal through which consumers are reached. However, this paper shows that the portal is controlled by employees who seek to maximize profit or relevance for the search engine company, not for the online businesses who depend on the search engine. Thus, an online business must actively monitor and effect their position in the search engine rankings to maintain a steady flow of customers.

3 The Online Advertising Industry: Economics, Evolution, and Privacy

3.1 Citation

Evans, David S. "The online advertising industry: Economics, evolution, and privacy." Journal of Economic Perspectives (2009).

3.2 Summary

In 2015, media companies around the globe spent $545 billion dollars on advertisements. Historically, these ads have found a home in newspapers and TV commercials, but this model is under attack by the new era of online advertisements. Although Internet ads were born in the 1994, they have grown at an astonishing rate; ZenithOptimedia expects the Internet to account for more than one-third of U.S. ad spending in 2017, representing 400% growth since 2007. Indeed, many major newspaper businesses have gone out of business. But while this revolution has crippled some businesses, it has created new efficiencies and opportunities for the global economy by significantly reducing the transaction cost for merchants to find consumers.

Advertising is a matching game between merchants and consumers. The old model required merchants to buy a million newspaper ads with the hope that a small fraction of the population is interested. The search engine has changed this game entirely. Merchants are now able to identify individual consumers who are interested in the product by matching their search queries or registered account information. This ability to detect a consumer's interest and intent to purchase a product has transformed advertising campaigns from mass market tools into focused, personalized ads. The search engines are an intelligent intermediary in this lucrative matching game.

So how do these businesses play the game? Typically, a company will start with an objective (like "increase sales of product X" or "make our brand more friendly") and set a budget to achieve this goal. Advertisers will then divide this budget amongst the various forms of media: online, television, radio, magazines, newspapers, etc. based primarily on the expected rate of return for each medium. In a competitive marketplace, modern businesses have no choice but to take advantage the cost-efficient advertising offered by the online intermediary search engines. These online ads further increase their rate of return by directly linking the consumer to the merchants online portal to purchase goods.

In summary, the Internet's share of ad spending has grown over the past decade and will continue to grow because search engines provide a more cost-efficient, targeted method of matching consumers with merchants. Modern businesses must adopt this technology or risk becoming obsolete.

4 Optimal search engine marketing strategy

4.1 Citation

Sen, Ravi. "Optimal search engine marketing strategy." International Journal of Electronic Commerce 10.1 (2005): 9-25.

4.2 Summary

This study was written in 2005–a time when businesses were still trying to figure out how to leverage search engine technology as a marketing tool. The paper starts by listing several different kinds of search engine marketing (SEM) strategies:

1. Keyword-related banner advertisements
2. Paid submission for regular updates: Businesses can pay a one-time fee to a search engine provider to guarantee that their site is reviewed within a short time frame, so it can show up in search results as soon as possible.
3. Search engine optimization (SEO): Businesses can influence their site's visibility in a search engine's natural results by modifying their site code.
4. Paid placements: Businesses can pay a search engine provider in order to be listed in their "sponsored" section in search result pages.

At the time of publishing, the majority of costs associated with SEM (around 82 percent) went towards paid placement campaigns. Despite this, buyers using search engines to find information generally only follow links that are found in the editorial section of the results page (where the natural results are found). This implies a couple of things. First, that SEO is probably the most effective marketing strategy. If a business can modify its site such that it naturally appears at the top of results, the conversion rate would be significantly higher than what could be achieved from any of the other SEM strategies. Second, it implies that the quality of SEO solutions in 2005 was insufficient for producing quality results.

The study ran an analysis on various SEM strategies, and came to a few conclusions:

Even though SEM was a high-growth part of online marketing at the time, it was not the dominant form. This was due to advertisers having doubts about its effectiveness. The analysis discovered that SEM was worth investing in if a business's market is characterized by: buyers who have low search intensity (e.g., buyers with opportunity cost of time), a product sold by many other providers (e.g., computers), or if the product is of low value (e.g., books). For businesses involved in niche markets (e.g. vintage cars), SEM was not worth investing in. This is because the number of competing businesses online was never high enough for SEM to be worth the cost.

Economically speaking, the analysis found that SEO was never a part of a business's equilibrium strategy. When there were a lot of competing companies, the probability of appearing in top results of a search engine was so low, that SEO couldn't do much to improve the situation. In this case it made more sense to pay for sponsored links, which were guaranteed to appear to users. When there were not many competing companies, investment in SEO was redundant, since the probability of appearing in top results was high enough to begin with.

Concerns were raised about paid placement strategies. Their effectiveness is determined by click-through rates, which can be artificially boosted by either people or software.

5 So You Want to Get Involved in E-commerce

5.1 Citation

Wilson, Susan G., and Ivan Abel. "So you want to get involved in e-commerce." Industrial marketing management 31.2 (2002): 85-94.

5.2 Summary

The advent of the Internet and search engines has fundamentally altered the advertising game. While many papers discuss advertising within the search engine (i.e. sponsored search), businesses can also advertise a brand on their company website. However, that website must stand out from search engine results in order to attract users. How can a business boost the rank of their website in major search engines? The answer may save or sink the entire business.

As discussed in the first paper on PageRank [Brin, Page], search engines crawl the web and rank sites based on multiple factors. Recall also that pages with a low rank (top of page) are significantly more likely to be viewed, and those beyond the top ten are virtually dead. Businesses can craft their web page to earn a high ranking in a few different ways:

1. Register your site with as many search engines as possible. While the crawlers try to find new pages, manually submitting the URL to the engine's index ensures a swift discovery.

2. Use keywords in the Title of your webpage and make it as descriptive as possible. This will increase the likelihood of matching a keyword query.

3. Use META tags to provide extra information about your page. These tags are invisible to the reader.

4. Place important information near the top of the page because search engines assume the top portion is indicative of the entire page.

5. Give reputable sites (i.e. product review sites, news sites) a compelling reason to link to your website. This link will tell the search engine that your website is high quality and thus boost your rank.

If your business site can achieve a high rank in search results, then it will receive more visitors. But that is only half the battle. You must design a compelling website that delights the user and it must change dynamically over time. If your site is a simple static page, users have no reason to visit again. Some successful commerce sites show new products upon each visit or a new sale or review of a product. They also implement digital catalogs where users can browse products from the comfort of home. An interactive website that appears in the top search results will serve as an excellent advertisement for the business.

6 An Empirical Analysis of Sponsored Search Performance in Search Engine Advertising

6.1 Citation

Ghose, Anindya, and Sha Yang. "An empirical analysis of sponsored search performance in search engine advertising." Proceedings of the 2008 International Conference on Web Search and Data Mining. ACM, 2008.

6.2 Summary

Businesses have many choices for online advertising. In search engines, the predominant ad form is "sponsored searches" in which the advertiser pays a fee to appear next to organic search results. For example, a consumer who searches "digital camera" is likely to see advertisements from Kodak in the search results because the company paid for ad space next to the "digital camera" keyword. A company can purchase any number of different keywords, so choosing which keywords to purchase and how much to pay for them will dramatically impact the success of a marketing campaign.

This paper empirically models the relationship between keywords, click-through rates, and conversion rates by analyzing a real world dataset from a large retail company. The data is fit into a hierarchical Bayesian framework and then run through Markov Chain Monte Carlo methods to estimate the results. The goal is to examine how click-through rates and conversion rates are affected by the following three characteristics of a keyword:

1. **Brand:** does the keyword contain a brand name? (i.e. Kodak, Nikon)
2. **Retailer:** does the keyword contain a retailer name? (i.e. Best Buy, Amazon)
3. **Length:** how many words are in this keyword? (i.e. "camera" vs. "black dslr HD camera")

Impact on Click-Through Rates

The first interesting result is that keyword advertisements with retailer-specific terms cause a 28.31% increase in click-through rates. This confirms the belief that users actually enjoy advertisements when they search for a specific retailer. In a surprising contrast, keywords containing a specific brand have no statistically significant effect on click-through rates. Finally, the length of the keyword is inversely related to click-through rates; an increase in length by one word decreases the click-through rates by 6.6%.

Impact on Conversion Rates

Keywords containing a brand-specific term experience an increase in conversion rates by 21.35%. This finding is important to businesses because it confirms that branded keywords are extremely valuable to the advertiser. Conversion rates are also impact by click-through rates. If the click-through rate is increased from the minimum (0.0) to the maximum (1.0), the conversion rate increases by 63.31%. However, this increase is dwarfed by the impact of rank: an ad that moves from the worst rank (bottom of page) to the best rank (top of page) yields an increased conversion rate of 99.97%. Finally, the quality of a business's landing page (as scored by Google on a 1-10 point scale) corresponds with the conversion rate. Using this metric, the authors found that a one point increase in landing page quality corresponded to a 22.5% increase in conversion rate.

Businesses can use these statistics to choose the optimal budget for keywords based on brand, retailer, and length.