



Trusting Google and Yahoo Search Engines & Information Literacy Jay Moonah

The Importance of Search

Web search engines such as Google and Yahoo have quickly become essential to our lives. We use search engines for business and academic research, for shopping, for informing critical decisions about health care and financial planning.

A Forrester Research survey conducted in the second quarter of 2008 found that 50% of North Americans consider the results they get from portals and search engines to be trustworthy. That's a higher level of trust than given to newspapers, magazines, radio and television, as well as most other online information sources. This may be because we have a certain understanding of traditional media. We understand that these mediums have human contributors and an editorial bias. Most of us have developed a healthy scepticism of media over decades of watching and listening.

But how do our critical thinking skills apply to what we find in our searches?

Because the results seem to appear like magic, many of us tend to think of search results as being unbiased. But in actual fact, there are many individuals and companies working hard every day to push their information to the top of the page in *your* Google search.

Given the frequency and significance of their use, one could argue that understanding how search engines work is as much a part of information literacy as understanding how a TV news department chooses which stories to cover. But how much does the average person really know about how search engines work, and how they retrieve information?

This manifesto will explain how those “magical” search results are generated. It will describe the basics of “search engine optimization” (or SEO) to give a peek behind the curtain of how content publishers, online marketers and Internet developers are constantly refining their own websites to push their rankings in Google, Yahoo and the other search sites ever higher.

This is not a how-to SEO manual, but rather a call for critical thinking from those who use search engines. Hopefully this manifesto will provide a starting point to making the most of your own search experience.

A Brief History of Web Search Engines

In the early-to-mid-nineties, back when websites could be counted in the tens or hundreds of thousands (as opposed to the hundreds of millions we have today), there were two major kinds of web search engines: directory searches and text index searches.

The original version of the Yahoo search was among the most popular of the directory searches. Actual human beings would submit sites to Yahoo, which employed other actual human beings to review and approve submissions into an increasingly massive hierarchical directory. The directory could then be keyword searched or browsed by category to find relevant information. The concept was great, but the methodology couldn't keep up with the explosive growth of the web later in the decade—too many new sites were coming online for even an army of human editors to review.

By contrast, the text index searches used automated systems to populate their databases. Computer programs called “crawlers” or “bots” (short for robot) would follow the links on pages to find new web pages, where they would record all the text appearing on the page. Some of the best known of these early text index search sites included Webcrawler, Lycos, Alta Vista and HotBot. But the text index approach also had problems scaling to huge numbers. One major issue was that there was no way to determine the relative importance of a particular web page other than pure evaluation of the contents. Unscrupulous site owners soon realized that they could game the system simply by packing their pages with key terms. If you were selling desks and your competitor had the number one result for “desks” because their web page mentioned the word desks 50 times, you just needed to make sure your site mentioned it 51 times, and soon you'd be on top. Enter Google.

Google's search engine, developed in 1997, uses the same basic method of the text index searches—it uses a bot to populate its database with the text of websites, and then allows users to search that database. But Google's savvy creators Larry Page and Sergey Brin added a new concept into their design, a system for determining the relative relevance of different pages. Within the Google data-

base, every page is given a rank based on the number of sites that link to it. This rank is then used to help order the results displayed to a user when they performed their search. The theory was that the more web links to a piece of content, the more relevant that content is likely to be. And it worked—Google’s model was so successful in returning relevant results that it supplanted its older competitors as the dominant search engine within a few years.

The Google model has since been adopted by most of the remaining search engines, including the re-developed Yahoo search, as well as Microsoft’s MSN Live Search. According to the site Hitwise which tracks website usage, these three accounted for nearly 95% of U.S. search traffic in 2008. If you are one of those 50% of Americans who use a search engine every day, chances are it is one based on the relevance model pioneered by Page and Brin.

Because the results seem to appear like magic, many of us tend to think of search results as being unbiased.

An Economy of Links

The model of using links to determine relevance does not value all links equally. If two pages have exactly the same content but one has a thousand links from other sites and the other only has 10, the page with a thousand links will ordinarily appear higher in search results. However, if the site with only 10 links sees all those links coming from sites that are of particularly high relevance (determined, in turn, by where THEIR links are coming from) that will have a positive impact on how high it appears in those search results. A page with only a couple of links can actually beat out one with many if the quality of those links is high enough.

This creates an economy of links, where some links are more valuable than others. Essentially, if your site is popular and attracts more links that are of high quality, links *from* your site will also be considered of higher quality.

Google actually makes the relative value of different pages of content visible to everyone. If you have the Google toolbar installed on your web browser, you can view what Google calls the “PageRank” for any given page of content. PageRank is a rating from one to ten, denoting the relative importance of that particular piece of content. For example, a small personal website with few inbound links might only have a PageRank of one or two, a medium-sized company site might rank four or five, and a very popular blog might rate a seven or eight ranking. Nine and ten rankings are generally limited to the homepages of the most popular sites on the web such as YouTube, Facebook, MySpace and the like. This means that a direct link from one of these sites is very valuable indeed!

Words Matter

But number and quality of links is only half the story. The other half is what those links say, and by extension say about a particular website.

Picture a piece of linked text on a web page. Very often, the words you are meant to click on are coloured blue and underlined, so let's assume that convention for our example. In this case, the blue underlined words are "John Smith." Now, there are lots of John Smiths from history and fiction—there's the sea captain who met Pocahontas, the former leader of the British Labour Party, the character from the Stephen King novel *The Dead Zone*... and of course that's forgetting all the real live John Smiths in the world! Each time the words John Smith are used as a link that is, in essence, a "vote" for what is most relevant when someone searches for "John Smith". The value of that vote will depend on the rank of the page it is coming from. The top result when someone types John Smith into Google will be the one with most links of the greatest rank.

Essentially, if your site is popular and attracts more links that are of high quality, links *from* your site will also be considered of higher quality.

Watch Out For That Google Bomb!

One of the best ways to demonstrate how search engine results are generated is by using the example of a “Google bomb,” also known as a link bomb.

Google bombing is the practice of getting a number of sites to link to a page with similar link text to artificially push it up in the search results. Possibly the best-known example of this practice is from 2006, when a number of website contributors displeased with the policies of George W. Bush posted the words “miserable failure” as a link to his official White House biography page. This was done as a protest and a prank, and it caused a search for “miserable failure” to return his bio as the number one result, much to the chagrin of his supporters!

Google and other searches have since cracked down on politically-motivated Google bombing, but unintentional Google bombs are still common. For example, as of this writing a search for the term “click here” in most search engines will return the download pages for various common software programs such as Adobe Acrobat, Flash Player and Apple iTunes. If you think about all the pages on the web with links reading something like “click here to download Acrobat” you’ll understand why this is the case. Most of these pages don’t even have the words “click here” in their text, but Google and other searches consider them highly relevant for this search because of all those links. This has not been done on purpose like the Bush Google bomb, but both demonstrate how those otherwise trustworthy search engines might generate unexpected results.

Search Engine Optimization

The rise to prominence of relevance models for ranking pages helped eliminate some of the dirty tricks that webmasters used to use, such as keyword packing. But the new model has spawned the creation of a new industry aimed at helping site owners improve their search engine rankings. This is the art and science of search engine optimization (SEO).

SEO specialists essentially perform two major tasks: (1) fine tuning web content to maximize the things that users are searching for and (2) generating links to that content. Now, there are many ways of going about these broad tasks, and like most areas of business there are those who perform with high ethical standards and those who don't. In SEO parlance, these groups are often known as the "white hats" and the "black hats."

White hat SEO practitioners generate links by helping to promote genuine interest in the content of the sites they are working on. They work with the owners of like-minded websites and appeal to those who might post links in forums, on blogs and the like. By contrast, black hats use things like "link farms," networks of fake sites created for the sole purpose of pushing up link numbers. The search engines do try to fight such unscrupulous practices, but like many professional rule-breakers (such as career criminals or unscrupulous computer hackers) the black hats are usually several steps ahead.

What SEO Means to You

The fact that people and organizations work to optimize their position in search results don't make them untrustworthy, but understanding how search results are generated and sometimes manipulated can help you make the most of your searches. Here are a few tips:

- 1. Look beyond the first results.** The SEO practitioner's ultimate goal is to get a given site to be the top result for as many searches as possible. Looking beyond the first couple of results can give you a sense of whether those top few are anomalies, which may indicate shady practices are at work.
- 2. Use variations on your search words.** Using different combinations of search words can help you hone in on the best information. Try adding things like location or industry-specific terminology if your results are too broad. Words like "review" or "article" can help if you are looking for independent third-party results, rather than information posted by an interested party.
- 3. Link to good things.** You may not think of yourself as a web content publisher, but if you post messages in public forums or use certain social network tools like Myspace or Twitter, the links you share can be picked up by search bots. It's added incentive to point to those sites you find useful. Also remember the unintentional "click here" Google bomb—if possible, use relevant terms as your link text.

You may not think of yourself as a web content publisher, but if you post messages in public forums or use certain social network tools ... the links you share can be picked up by search bots.

Trust and Search

Search engines like Google and Yahoo provide instant access to a vast amount of information, but they should not be trusted blindly. Hopefully, armed with information from this manifesto, you can become a little savvier about how these powerful tools work, and how to use them to find the best information you can.

Search engines like Google and Yahoo provide instant access to a vast amount of information, but they should not be trusted blindly.

Happy searching! 

ABOUT THE AUTHOR

Jay Moonah is the Vice President of Marketing for [Wild Apricot](#), which provides web software that helps associations and nonprofit organizations manage their websites, membership records, events and donations. Prior to Wild Apricot, Jay worked with many clients, including Molson Canada, CTVglobemedia, the Canadian Broadcasting Corporation (CBC) and Canoe/Sun Media. He has also taught at a number of institutions, including University of Toronto and Ryerson University. He is co-organizer of Podcamp Toronto, an annual new media “unconference.” He is also an independent musician and song writer. Jay’s blog and podcast “Media Driving” can be found at [mediadriving.com](#).

SEND THIS

[Pass along a copy](#) of this manifesto to others.

SUBSCRIBE

[Sign up for our free e-newsletter](#) to learn about our latest manifestos as soon as they are available.

BORN ON DATE

This document was created on June 3, 2009 and is based on the best information available at that time. Check [here](#) for updates.

ABOUT CHANGETHIS

[ChangeThis](#) is a vehicle, not a publisher. We make it easy for big ideas to spread. While the authors we work with are responsible for their own work, they don’t necessarily agree with everything available in ChangeThis format. But you knew that already.

ChangeThis is supported by the love and tender care of 800-CEO-READ. Visit us at [800-CEO-READ](#) or at our daily [blog](#).

COPYRIGHT INFO

The copyright of this work belongs to the author, who is solely responsible for the content.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License. To view a copy of this license, visit [Creative Commons](#) or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

WHAT YOU CAN DO

You are given the unlimited right to print this manifesto and to distribute it electronically (via email, your website, or any other means). You can print out pages and put them in your favorite coffee shop’s windows or your doctor’s waiting room. You can transcribe the author’s words onto the sidewalk, or you can hand out copies to everyone you meet. You may not alter this manifesto in any way, though, and you may not charge for it.