"Don't be evil": Uncovering the Implications of Google Search

Stephanie Winston

Abstract: Google has taken the world by storm, and with its fast and seemingly accurate search results, internet users are increasingly using this search engine to obtain the majority of their information. This paper studies the implications of internet users' increasing reliance on Google for retrieving information. It specifically examines Google search because it is the world's most prominent search engine, and studies the potential flaws in its algorithm, such as search engine bias and the rich-gets-richer phenomenon, as well as issues in the overall website retrieval process. It concludes that due to the sheer number of users relying on the first page of Google results for information, the seemingly indiscernible flaws in the system have a significant effect on the formation of knowledge.

About the Author(s): Stephanie Winston is currently working towards a Masters of Library and Information Studies degree from Dalhousie University and can honestly say that she loves every minute of it. She is a recent graduate of Queen's University where she obtained a Masters of Arts in English Literature, and holds an honours Bachelor degree in English and history from Carleton University. Originally from Toronto, Stephanie currently lives in Halifax with her fiancé Michael and their cat Max. This paper was originally written for an Information in Society MLIS course.

Introduction: The Search Habits of Internet Users

Currently, the number of internet users using search engines is at an all-time high, and this is not entirely good news. On a typical day, 49% of all American internet users are using search engines. From 2002, this number has increased to 62%, and there is no expectation that this figure will decrease anytime soon (Fallows, 2008). To put these figures into perspective, in 2005 researchers found that "on any given day more than 60 million adults are sending over 200 million requests to web search engines" (Spink & Zimmer, 2008, p. 3), and, in 2010, users are now doing more than 34,000 searches per second with a total of 88 billion searches per month (Cowan, 2010). This is a remarkable increase. Furthermore, with all these search queries, 73% of these information seekers never even venture past the first page (Evans, 2007). This means that a large majority of users' information is solely coming from online sources, and search engines are principally the ones mediating and retrieving user information. This also indicates that most of the information users access comes from very few sources, since the first page of the search results usually only has ten hits. There is one more significant statistic to address: Google is now the most popular search engine on the internet and, in 2005, got 46.2% of all internet search queries (Evans, 2007). This number is especially significant because since that time the number of widely used search engines has dropped from around ten to an oligarchy of just three (Yahoo!, Google, and Microsoft), with Google fiercely dominating the pack.

According to Hinman (2008), users must recognize that never before will so few have controlled so much information and be aware that search engines' websites are not always authoritative. While Google is not the only search engine on the internet, users are increasingly relying solely on Google for the majority of their information. As a result, it is essential to discern whether or not Google restricts the public's access to information in their search results and understand what some of the long term implications of this are on users' knowledge. This paper finds that the way Google formulates its search hits for the user and the different components and biases in its algorithm all influence and control where users are getting their information. It finds that the role money and marketing play in its results affect the variety of information users are receiving. Furthermore, this paper shows that users' knowledge is both mediated and restricted by Google due to all the aforementioned flaws in Google's system and the sheer volume of people using Google as their primary resource.

How Does Google Control Its Search Results?

Google's algorithm controls what information users retrieve, and it also decides what order to rank the search hits in the search results page. Understanding why certain webpages are on the first results page is especially important because, even though Google returns hundreds of thousands of pages of search results, the top ranked sites are the ones that users look at first

and, on most occasions, the first page is the only one users use to get their information (Evans, 2007). Google's search result strategy is strictly protected by trade secret laws since it is a business, after all, that is trying to remain competitive and protect its interests. As a result, it is impossible to know the exact formula a Google search uses to decide the placement of each hit. However, even without the exact equation, we can still study the similar components in pages found in Google's search results to try to determine how the algorithm works. Since Google uses over 200 different factors, or signals, to come up with its search results, with each signal having a different weight, it is important to study the signals that most likely have the greatest influence on a page's rank.

Since no one outside of Google knows exactly how its page ranking works, marketers use a trial and error method to determine the most important algorithmic factors that are needed on a webpage in order to get it a higher ranking. This process is called Search Engine Optimization (SEO), and marketers (also known as search engine optimizers) use SEO to try to advance businesses by making their webpages more accessible on Google (Fortunato, Flammini, Menczer, & Vespignani, 2006). Even though different query terms may employ different ranking factors and weight, and different Google data-centres use subtly different ranking algorithms (Evans, 2007), there are still a few basic factors that play an influential role in determining a page's ranking.

Google's PageRank, "the 'heart' of its software" (Diaz, 2008, p. 15), is the most important and controversial factor for deciding the accessibility of a webpage through a search. This is because, in all cases, the "higher a site's PageRanking, the higher its [search result] ranking" (Evans, 2007, p. 25). This algorithm judges a webpage's importance in terms of how many other sites have linked to it and the authority (or PageRanking) of those websites that link to the original webpage. By doing so, PageRank decides a webpage's rank based on its popularity, prominence, and authority on the web. This makes PageRank seemingly democratic, since Google itself is not the one that decides the webpage's PageRank; rather, other sites on the internet give it its value. Google (2009) contends that "democracy on the web works . . . because millions of individuals . . . determine which other sites offer content of value" (para. 4). I argue that following a democratic approach is not the best method of retrieving comprehensive information. This democratic system, one that Google calls a "Googlocracy" (Diaz, 2008), is problematic because it is based on quantity and popularity, rather than quality and credibility of information. With Googlocracy, a great many pages link to a very small number of sites, called hubs, while the vast majority of pages are linked to few or no external sites at all. As a result, the pages with the most links are more easily discoverable, and they remain popular because their accessibility increases the likelihood of them continually getting linked. At the same time, the pages without many links remain almost impossible to find through Google (Diaz, 2008). This instigates a "rich-get-richer" vicious cycle (also called Googlearchy) that prevents users from getting a variety of information from diverse and underrepresented websites (Diaz, 2008; Goldman, 2008).

It is important to remember that, by using PageRank, Google is trying to be egalitarian. This is because PageRank lets website owners decide for themselves which other webpages they should promote for internet users, and, ultimately, this gives people the power and agency to decide what information they receive. According to Google proponents, "web authors vote by creating hyperlinks, and Google counts the votes. If we want to understand Google we need to see democracy as Google's very nature" (Diaz, 2008, p.16). It seems that Google is mirroring the common wishes of the user, rather than influencing them. At the same time, the effect of PageRank is that it promotes popular, commercial pages over non-commercial sources of information. As a result, minority interests, and the websites that cater to them, often receive marginal exposure in search results (Goldman, 2008). This is problematic because, instead of offering pages that users *need* to see, PageRank makes it so Google provides only the information that other users *want* to see (Diaz, 2008). Through PageRank's seemingly egalitarian algorithm, users are unlikely to visit the variety of less traveled, yet equally important sites, and this limits the information that users receive and the overall knowledge they gain.

Although a website's PageRank is fundamental in determining its place in the search results, Google is quick to respond that PageRank is just one signal out of over 200 other important factors that all play a role in deciding the position of web pages in Google search results (Google Sites, 2010). Along with PageRank, Evans (2007) discusses the other signals that exert the greatest influence in defining a page's rank. He explains that the website's domain name can impact its ranking since older domain names often convey more trust. These sites with older domain names are given more authority because they have been linked to, or cited, more often by other websites on the internet. As well, the number of web pages indexed in a site—basically, the bigger the website is in magnitude—usually increases the website's ranking. Finally, the number of listings the webpage has in human-edited directories such as Yahoo! and DMoz directories, and the number of times the website is bookmarked by Del.icio.us (a social bookmarking site) also have implications on the site's ranking. Google deems these last two signals to be especially authoritative because they are based on human opinion, rather than mathematical formula, and they are also democratic in nature since all internet users have an equal say in deciding the importance and popularity of a webpage through bookmarking a page (Evans, 2007). At the same time, these four additional factors all follow the same rich-get-richer paradigm because they prioritize web pages based on their popularity and pre-existing visibility. Even though Google uses a variety of signals in its ranking system, many of the prominent factors have the same flaw, which prevents a true egalitarian display of information.

Google has also started to use personalized algorithms alongside its standardized one-size-fits-all model. Goldman (2008) explains that personalized algorithms "produce search results that are custom-tailored to each searcher's interests, so searchers will see different results in response to the same search query" (p. 130). This guarantees that Google's top search results

are centred on information catered to each specific user, which means that the search rankings are based on more than just the general consensus since searcher-specific factors also play a role in the ranking system. However, there is a significant bias in this algorithm, and that is the users themselves. As previously mentioned, the results from a personalized algorithm are catered specifically to the users' interests. This means that the results are what the user wants and expects to see; the information users get is not comprehensive, but limited to what Google believes is the user's preference. Of course, Google is trying to please users by giving them websites that will best suit their needs, and the majority of users are probably happy that Google always fulfils their expectations. At the same time, this restricts the information that users get, and the pages that they are not normally interested in receiving will not show up unless they specifically request it. As a result, it is hard for users to notice that they are missing key websites from the search results and, consequentially, it is almost impossible for them to get that missing information through follow-up searches unless they are looking for it specifically or Google recognizes subtle changes in their interests.

Google Instant takes the personalized algorithm to a new, almost mind-reading, level. Google Instant, launched on September 8, 2010, is a search enhancement that shows the search results as, and even before, users type in their query. It is able to do this by predicting the query before the user is finished typing it based on both personalized and standardized algorithms. Google Instant's production manager, Jonathan Effrat, explains that "seeing results instantly lets you refine your search as you type" (Google Blog, 2010). By doing so, Google influences more than just the information users retrieve; it also controls each specific search before the user even types it. This has serious consequences for users because it leads them to become passive information seekers. The problem with becoming a passive information seeker is that users gradually become comfortable with letting another entity take control of the information they are retrieving and have an underlying trust in the overall knowledge that they are gaining. In fact, a study by Fallows (2005) indicates that most users believe that search engines are a fair and unbiased source of information. This belief is particularly ironic because 62% of users are unaware that there is a distinction between paid and unpaid search results. As a result of users' increasing passivity, they lose agency because they are no longer making their own decisions, since Google's terms are now more accessible and, searchably-speaking, more acceptable than the users' search terms. Passive information seekers also give Google an extensive degree of control over information, and the consequence of Google abusing its authority is immense. This is an issue that will be discussed in more detail below. As Hinman (2008) fittingly remarks, "never before will so few have controlled so much with so little public oversight or regulation. . . . [P]ublic mistrust may be the more appropriate attitude" (p. 74-5).

Search Engine Marketing and Management

Google has two types of customers who both sustain the search engine and influence each other: the average user and companies who pay for advertising. The average user merely

uses the search engine to find information and buy products, while advertisers give money to Google, which enables it to remain free for the user, and inadvertently receive money from the average user through purchases made from paid advertisements. It is difficult, if not impossible, to see how Google distinguishes between its two customers because all three are entirely reliant on each other for their success. For instance, Google tracks, collects, and sells information about its users' activities to marketing companies, and many users, who are aware that this is occurring, are uncomfortable with this (Hinman, 2008). At the same time, tracking is also beneficial for users, specifically smartphone users, because it enables them to obtain wireless internet access, better service, and allows them to "find exactly the goods and services that they are looking for because, using profile data, search engines can better predict the kinds of items a particular user will find interesting and worth buying" (Hinman, 2008, p. 72). Another example is when Google offered to build a free wireless network in San Francisco because it could attain such valuable information about individual users and collective user groups that the value of the information, for advertising companies, outweighed the cost of building the network. Users are pleased with the opportunity to get free internet and marketers get exactly the information they desire (Hinman, 2008). In this way, Google's effective marketing strategy is also a successful user satisfaction strategy.

Even with its intricate stakeholder relationships, Google's monetary success is still the only factor that matters because it guarantees Google's continuance. It is also important to remember that its monetary success solely depends on being able to adequately please its advertisers. Marketing makes up 99% of Google's revenue and, as a result, Google has a huge financial incentive to direct users towards the sites of its sponsors (Diaz, 2008; Hinman, 2008). Although Google visibly limits its sponsored advertisements to the top of the page and the right side of the screen in its search results, Hinman (2008) notes that "subtle shifts in tightly-concealed algorithms could result in subtle [marketing] biases that would be impossible to prove" (71). Indeed, the beauty of having a trade secret algorithm is that no one knows what the exact marketing biases are on Google. Although it is unknown whether Google is giving certain sites an unjustifiably prominent position in the search results, marketing biases are still a possibility, which, if they did exist, would play a considerable role in the kind of information users are directed to and ultimately accessing.

While it is unknown whether Google's algorithm has any biases towards its sponsors, it is clear that the system for ranking search results is not wholly standardized. Goldman (2008) found that in search engines, such as Google, users are under the impression that

> machines, not humans, appear to make the crucial judgements, creating the impression that search engines bypass the structural biases and skewed data presentations inherent in any human-edited media. . . . Unfortunately, this romanticized view of search engines does not match reality. Search engines are media companies . . . [and they] make editorial choices designed to satisfy their audience. . . . [T]hese choices

systematically favour certain types of content over others, producing a phenomenon called search engine bias. (p. 122)

In theory, Google's editorial control is actually beneficial for users trying to gain access to more information because it prevents the results from being overtaken by spammers and enables Google to provide the most relevant search results for the user, which ultimately gives the best possible user experience. At the same time, it is important to remember that editorial control, including personalized algorithms, omits certain sites and, in Google's case, has led to situations where certain sites become censored. In North America, for example, anti-Semitic sites are easy to find on Google, but they do not show up on French or German versions of Google. While this seems like no great loss, the situation in China is another story. When searching for websites about the Dali Lama and the Tiananmen Square Massacre, Chinese users are led only to government-approved sites (Hinman, 2008). In North America, previously published photos of prisoner abuse at Abu Ghraib were not showing up in the Google Image search, and Google "spokespersons were unable to provide a convincing explanation for their disappearance" (Hinman, 2008, p.74). These examples illustrate that editorial control has the potential to be abused, and there is a fine line between facilitating and controlling user access to information.

While Google has the capacity to restrict public access to information, it is also very evident that individual websites strongly market their pages to try to appear in the first page of Google's search results, and this impacts the quality of information that the user gets. Websites do this because the closer a webpage is to the first search result page the more likely it will be perused by users and, depending on the site, the likelihood that users will buy its products is increased. Of course, knowing and understanding Google's algorithm would facilitate webpages trying to get a high ranking, but since this is not possible they instead hire search engine optimizers and marketers to try to improve their ranking. Through years of trial and error, these optimizers try and, in many cases, come close to

crack[ing] the code of being visible to your customers when they are searching on the internet and when they [users] are ready to buy . . . [marketers] teach you how to appropriately connect with customers throughout their search process, especially when they are ready to take action. (Lutze, 2009, xxiii)

The possibility that these optimizers are able to increase websites' rankings and make themselves accessible for the average user indicates that the rich (those who can afford to increase their pages' ranking) control a large portion of the information users get. This is because the more money a website spends on optimizing its page, then the more likely it is to get significant ranking results. As a result, some of the pages on Google's results page have no business being there and, consequentially, some of the information users get, which contributes to the formation of knowledge, is probably subpar.

Implications of Search Result Biases on User Knowledge

As previously mentioned, there are only three predominant search engines on the internet today and, in 2005, while other search engines were still being used, these same three engines handled over four-fifths of all the search traffic (Diaz, 2008). Furthermore, even when there was a variety of search engines to choose from, Fallows and Rainie (2004) explain that almost half of all web users frequent only a single search site (Diaz, 2008). Since that time, Google has become, by far, the most widely used engine of the group. Due to its bundled services, such as e-mail (gmail) and GPS mapping (Google Maps), users build even more brand loyalty than in the past and unconsciously get locked into solely using Google for all their searching. The inception of Chrome, the Google-run browser that searches Google in the URL box, takes this point even further because it compels users to forgo typing specific URLs altogether in favour of letting Google automatically retrieve all their sites. This further illustrates how users are becoming passive information seekers and shows how users increasingly let Google make their search decisions for them. Diaz (2008) provides a very appropriate statement for this situation: "so powerful has Google become that many . . . view it as the web itself: it you're not listed on its indexes, they say, you might as well not exist" (p. 27). The issue here is that there is very little competition, and this has the same problematic effect in a democratic society as the consolidation of news media (Mowshowitz & Kawaguchi, 2005). This is seemingly ironic because Google prides itself on its search results being democratic; however, its domination in the search engine world makes it closer to dictatorship. At the same time, even though there is only really one source from which users get their online information, based on Fallow and Rainie's statistic, even if there were more competition, users would probably continue to get their information from the one source.

As a result of Google's monopolization, the aforementioned biases found in Google's ranking of search results are all the more concentrated and have a much greater impact than if there were a greater variety of search engines in common use. When Google tried to explain the implications that Google Instant would have on the individual user, it used the sheer quantity of users to better illustrate its overarching global significance. Google Blog (2010) explains, "Google Instant saves the average searcher two to five seconds per search. That may not seem like a lot at first, but it adds up. With Google Instant, we estimate that we'll save our users 11 hours with each passing second!" For the individual user, the implication of seemingly slight biases found in Google's search rankings and algorithm may not seem particularly significant, but when millions of users are exposed to the same biases over the course of billions of monthly searches, the ultimate effect on the kind of information that users retrieve and utilize is huge. Of course, there is no way to quantify the specific impact that these biases have on users' overall world knowledge; however, understanding the biases themselves and the quantity of users who are constantly exposed to them surely helps one make sense of the seriousness of the situation.

Due to the sheer number of individuals using Google and the massive number of searches occurring on a daily basis, one can safely say that much of users' knowledge is a result of Google and the way it disseminates information. Hinman (2008) further explains that "search engines have a tremendous influence on access to knowledge; indeed, search engines will increasingly be in a position to construct knowledge through control of access" (p. 72-3). Since Google controls the information users see and utilize, it also shapes the users' development of knowledge on that issue. One counterargument to this point is that knowledge is shaped from a variety of both online and offline sources: academic and scholarly publications, movies, newspapers, books, and other sources. However, with the ownership and control of Google Books, Google Scholar, Youtube (Google purchased the company in 2006), Google News, and Google Maps, it seems Google is already controlling the majority of online resources that contribute to the formation of knowledge. Although it seems obvious, one cannot really distinguish between the knowledge attained through offline and online sources anymore because knowledge comes from such a wide range of channels that, at some point, Google will have probably played a role. Even if two laggards are taking part in conversation and sharing their knowledge, it is still probable that some of their shared knowledge was mediated to them by another source that used Google to attain information.

With millions of people using Google to find information, there are still specific demographics that use Google more than others, and this shows further the implications of Google, and its biases, on North American society. According to Fallows' (2008) most recent study on search engine users, it seems that men making over $50,000 a year, with a college education, high speed internet, and more than six years of online experience, are using search engines the most. Interestingly, according to theorists such as Karl Marx, these individuals, more specifically men with money and education (he was not theorizing about internet access), play the biggest role in initiating social changes (Marxism, 2008). As a result, it seems Google, including its biases, contributes to the formation of knowledge and, over time, it is clear that it will continue to influence our ever-changing society.

Final Remarks

With Google's prominence in society, its ability to disseminate information and ultimately facilitate the formation of knowledge, it is essential to question if there are any flaws in the information it provides its users. As I have shown, there are indeed many factors that play a significant role in limiting and, ultimately, controlling users' knowledge. For instance, Google's search signals, such as PageRank, fall prey to a rich-get-richer phenomenon that affects the information available to users in the first page of Google's search results. As well, there are also Google marketing strategists who try to increase a website's ranking to help them make a profit, and this has implications on the credibility of the information that users access. Overall, while the user may never notice the flaws in the quality and variety of Google's search results, the sheer number of users who develop their knowledge through the first page of their search

results illustrates that slight biases have a significant effect on users' knowledge. Can Google change its system to compensate for these faults? Should it have to? Perhaps it is better that users are aware of these biases and remain ever faithful that Google will adhere to its motto: Don't be evil (Google, 2009).

References

Cowan, T. (2010, November 16). Infographic: Human knowledge, by the numbers. *The Globe and Mail.* Retrieved December 1 2010, from http://www.theglobeandmail.com/news/national/time-to-lead/internet/human-knowledge-by-the-numbers/article1801253/page6/

Marxism. (2008). In *The Columbia Encyclopedia.* Retrieved from http://www.credoreference.com/entry/columency/marxism

Diaz, A. (2008). Through Google goggles: Sociopolitical bias in search engine design. *Information Science and Knowledge Management, 14*(2), 11-34. doi: 10.1007/978-3-540-75829-7_2

Evans, M.P. (2007). Analysing Google rankings through search engine optimization data. *Internet research, 17*(1), 21-37. doi:10.1108/10662240710730470

Fallows, D. (2005). Search engine users: Internet searchers are confident, satisfied and trusting—but they are also unaware and naïve. *Pew Internet & American Life Project.* Retrieved from http://www.pewinternet.org/~/media/Files/Reports/2005/PIP_Searchengine_users.pdf.pdf

Fallows, D. (2008, Aug). Data memorandum: Search engine use. *Pew Internet & American Life Project.* Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2008/PIP_Search_Aug08.pdf.pdf

Fortunato, S. Flammini, A., Menczer, F., & Vespignani, A. (2006, May). The egalitarian effect of search engines. Proceedings from WWW2006 Conference, Edinburgh, United Kingdom. Retrieved from http://arxiv.org/pdf/cs.CY/0511005

Goldman, E. (2008). Search engine bias and the demise of search engine utopianism. *Information Science and Knowledge Management, 14*(3), 121-133. doi: 10.1007/978-3-540-75829-7_8.

Google. (2009). *Our philosophy: Ten things we know to be true.* Retrieved from http://www.google.com/corporate/tenthings.html

Google Blog. (2010, September 8). Search: Now faster than the speed of type. [Web log comment]. Retrieved from http://googleblog.blogspot.com/2010/09/search-now-faster-than-speed-of-type.html

Google Sites. (2010). *FAQ: Crawling, indexing, & ranking*. Retrieved from http://sites.google.com/site/webmasterhelpforum/en/faq--crawling--indexing---ranking#pagerank

Hinman, L.M. (2008). Searching ethics: The role of search engines in the construction and distribution of knowledge. In A. Spink & M. Zimmer (Eds.), *Web Search: Multidisciplinary Perspectives* (67-76). Berlin: Springer-Verlag Berlin Heidelberg.

Lutze, H. (2009). *The findability formula: The easy, non-technical approach to search engine marketing.* New Jersey: John Wiley & Sons, Inc.

Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. *Information Processing and Management, 41*, 1193-1205. doi: 10.1016/j.ipm.2004.05.005

Spink, A., & Zimmer, M., (Eds.). (2008). *Web search: Multidisciplinary perspectives*. Berlin: Springer-Verlag Berlin Heidelberg.

Zimmer, M. (2006). The value implications of the practice of paid search. *Bulletin of the American Society for Information Science and Technology, 32*(2), 23-25. doi: 10.1002/bult.1720320211