

Search Engine Optimization

Erin S. McIntyre

Fort Hays State University

October 11, 2015

Search Engine Optimization

Erin S. McIntyre

Fort Hays State University

Introduction

In today's digital landscape, many businesses and individuals find it advantageous to create and maintain a web presence. Ensuring that this web presence can be found by visitors who are looking for what's being offered (goods, services, or information) is an ever-changing challenge. This paper explores the evolution of search engines and search engine ranking algorithms, and presents various search engine optimization (SEO) techniques.

A Brief History of Search Engines

It's helpful to understand the history of search engines in order to understand SEO methodology. Computerized information retrieval systems began to appear as early as the 1960s with Gerard Salton's SMART (Salton's Magic Automatic Retriever of Text) and Ted Nelson's computer networking initiative, Project Xanadu (Wall, 2015). The early 1990s saw a surge of precursors to today's search engines, with McGill University student Alan Emtage's Archie, the University of Nevada's Veronica, and Jughead, attributed to the University of Utah's Rhett Jones (The Search Engine Archive). Tim Berners-Lee (credited with the creation of the World Wide Web, with Robert Cailliau) built the first web browser prototype and created the first web catalogue, the Virtual Library (Wall, 2015).

As the World Wide Web grew and evolved, so did our ability to search it. Search engines Excite, World Wide Web Wanderer, and Aliweb emerged in 1993. 1994 brought Yahoo, AltaVista, Lycos, Infoseek, and Webcrawler onto the scene. Inktomi: Hotbot, Google, Ask

Jeeves, MSN, the Open Directory Project, and AllTheWeb surfaced in the second half of the decade. The creation of new search engines slowed after the 1990s, with just two notable additions: Baidu in 2000 and Bing in 2009 (Wall, 2015).

As of this writing, Google, Baidu, Bing, and Yahoo are the four biggest players in global search engine traffic for desktop users, with Google capturing 65.76%, Baidu capturing 13.64%, Bing capturing 9.37%, and Yahoo capturing 8.9% (NetMarketShare, 2015a). Google nets an astounding 92.22% of mobile and tablet search engine traffic (NetMarketShare, 2015b).

Because Google searches represent such a large portion of today's search engine traffic, it's important for webmasters to understand how Google ranks websites. Webmasters must also be able to differentiate between search engine optimization (SEO) methods that help a site's Google ranking, and methods that cause a site to rank poorly on Google (or be banned altogether). Given that Google represents the huge majority of search engine traffic, this paper focuses on Google-specific SEO, and the reader should understand that sites that are optimized for Google generally also rank well in other search engine results.

The PageRank Algorithm

The mid-90s were the "wild west era" of search engine evolution, and search engine results were often irrelevant or a mixture of high- and low-quality sites. The PageRank algorithm was developed by Larry Page and Sergey Brin in an effort to rank websites in a more relevant way (Page, Brin, Motwani, & Winograd, 1998). Page and Brin realized that the number of backlinks (links pointing to a given site from other sites) was not, in itself, a good measure of site quality. Additional factors were needed, and the pair proposed assigning a "PageRank" to each site to improve the relevance of search results. At PageRank's inception, PageRank was, broadly speaking, a number calculated using not just the quantity of links pointing to a site, but also by

weighting some backlinks more heavily than others in the equation. For example, a backlink from a respected, frequently-linked-to news website was more valuable than a backlink from an obscure individual's website of family photos. Page and Brin created the Google search engine to test their PageRank algorithm (Page et al., 1998). Google continues to use a version of that PageRank algorithm, along with a number of other metrics, today (Manber, 2008).

Black Hat SEO Tactics

As Google's search algorithm has evolved, so have the tactics employed by webmasters who attempt to manipulate their sites' positions in search engine results pages (SERPs). Black Hat SEO tactics are methods for boosting search engine ranking that detract from a positive user experience. Examples include keyword stuffing, hidden text, link schemes, cloaking, "sneaky" redirects, doorway pages, scraped content, and thin affiliate sites (Lian, 2014). Google punishes sites that use these tactics by placing them lower in its SERP.

Keyword Stuffing

Keyword stuffing refers to the practice of over-using a keyword for which the site is trying to rank well. Examples include writing content with the keyword used so often that the text sounds unnatural, and including lists of words or numbers that don't add value to the page's content (Google, 2015a).

Hidden Text

Hidden text is considered a Black Hat SEO tactic when it's aimed at manipulating search engine rank and doesn't improve user experience. For example, including text irrelevant to the site's content in a white font on a white background is a strategy that attempts to boost the page's rank for search terms that don't actually relate to the site (Google, 2015b). It should be noted that there are some circumstances where hidden text is acceptable (and even desirable), such as when

the site includes Flash, JavaScript, or images. It's useful to have accurate alternate text for these technologies to assist both robot crawlers and visitors who use assistive technology (Google, 2015b).

Link Schemes

Google's Webmaster Guidelines define a link scheme as "any links intended to manipulate PageRank or a site's ranking in Google search results" (Google, 2015c). This seems ambiguous at first glance. How can we determine what a link is intended to do? Furthermore, don't most site owners intend to make their sites rank as well as possible? The real goal of this particular guideline seems to be to discourage the buying and selling of PageRank-conferring links, rampant link exchanges, and "spammy" backlink comments on other sites. When in doubt, site owners should use the `rel="nofollow"` attribute in any links that could be considered part of a link scheme under this guideline.

Sneaky Redirects

Redirecting a user from one URL to another is not, in itself, a Black Hat SEO practice. Google considers a sneaky redirect one that redirects the user to a page with content significantly different from the page information that appears in the search results (Google, 2015d).

Cloaking

Cloaking is similar to sneaky redirecting in that both methods intend to serve the user content that doesn't match the search results page information. A site that engages in cloaking detects whether a visitor is a human or a robot, and serves significantly different content to each type of visitor (Google, 2015e).

Doorway Pages

A doorway page is a page created specifically to help a site rank well for a given keyword or search term. Google discourages the use of multiple keyword-targeted pages or domains that all funnel users to the same site (Google, 2015f).

Scraped Content

Scraped content refers to content that is duplicated from another site verbatim, or with slight modifications that don't add significant value (Google, 2015g). Google rewards original content and punishes excessive duplicate content.

Thin Affiliate Sites

A thin affiliate site is one that is either a duplicate of the original merchant site, or one that does not add significant value to the affiliate items. While Google does not punish sites for containing affiliate content per se, sites will rank better in Google search results when the majority of their content is fresh and unique (Google, 2015h).

Google's Changing Algorithm

Google frequently updates its algorithm to combat Black Hat SEO tactics and other issues deemed to detract from user experience. The first Panda update, on 2/23/2011, punished thin content and sites with high advertisement-to-content ratios, while the 11/3/2011 "freshness update" rewarded sites that kept their content recent. On 1/19/2012, Google introduced an update that punished sites with too much advertising space "above the fold". The 8/10/2012 DMCA Penalty/Pirate update penalized sites with multiple copyright violations. The first Penguin update, on 4/24/2012, punished sites for keyword stuffing, and the 6/11/2013 "payday loan" update targeted sites with "spammy" reputations. (Google's announcement of this update specifically mentioned payday loan sites and pornography as two of the intended targets.) The

HTTPS/SSL update of 8/6/2014 provided a slight boost to secure sites, and on 4/22/2015, Google rolled out an update that rewarded sites optimized for mobile devices (Moz, 2015).

White Hat SEO Techniques

Google maintains that webmasters should build sites with the user in mind, not the search engine (Google, 2015i). While it's true that fresh, original content is necessary for a high-quality website, there are additional measures that can be taken to make sure search engines understand and index a site's content. White Hat SEO techniques are ones that both improve the human experience of a website and make the site easily accessed and indexed by robots.

Tags

The `<title>` tag is crucial. The contents of this tag identify the page in search results, and each page of a site should have this tag (Connolly & Hoar, 2015). Title tags should be unique on each page, and should contain a keyword when reasonable. The "description" meta tag name attribute (`<meta name="description" content="" />`) should be present, and "content" should contain an accurate summary of the page, as this text often appears under the page title in search results (Connolly & Hoar, 2015). Include a relevant keyword or two, being careful to avoid keyword stuffing. Where reasonable, include a keyword in an `<h1>` tag (Dean, 2015). The `<nav>` tag, which identifies a site's navigation links, should be used for a site's main navigation menu. This assists Google crawlers as they attempt to determine a site's size and structure (Connolly & Hoar, 2015).

Help the Robots!

Sites should include a robots.txt file and make use of Google's robots.txt Tester Tool to make sure the file is functioning in the desired manner. To assist Google's crawlers, webmasters should submit a sitemap using Google Search Console (Google, 2015i).

URLs, Anchor Text, and Images

URLs should be optimized for search engines by including descriptive words in the URL paths instead of numbers or codes (i.e. `/products/bedding/sheets/pink/` gives a search engine more information than `/products/category004/item354/color250/`) (Connolly & Hoar, 2015). Anchor text should be descriptive (instead of the generic "Click Here"), and images should have descriptive file names and accurate alt attributes in the `` tag (Connolly & Hoar, 2015).

Links

Relevant outbound links to trusted sites may send a trust signal to Google, while too many outbound links may hurt PageRank. Most outbound links should use the `rel="nofollow"` attribute (Dean, 2015).

Video Hosting

Site owners who use video content on their websites may want to consider using YouTube as their video host. YouTube is owned by Google, and it has been argued that the presence of a YouTube video on a site may be a favorable factor in the site's Google SERP ranking (Dean, 2015).

Conclusion

As search engine technology continues to evolve, so must the strategies employed by site owners to make sure their websites appear in relevant search results. Webmasters benefit from understanding the factors that make up the ranking algorithm of Google, the search engine with the largest market share of global search engine traffic. By avoiding Black Hat SEO tactics and using White Hat SEO techniques, site owners can make sure their sites take advantage of their full Google ranking potential.

References

- Connolly, R., & Hoar, R. (2015). *Fundamentals of web development*. New York, NY: Pearson Education.
- Dean, B. (2015). Google's 200 ranking factors: The complete list. Retrieved from <http://backlinko.com/google-ranking-factors>
- Google. (2015a). Irrelevant keywords. Retrieved from <https://support.google.com/webmasters/answer/66358>
- Google. (2015b). Hidden text and links. Retrieved from <https://support.google.com/webmasters/answer/66353>
- Google. (2015c). Link schemes. Retrieved from <https://support.google.com/webmasters/answer/66356>
- Google. (2015d). Sneaky redirects. Retrieved from <https://support.google.com/webmasters/answer/2721217>
- Google. (2015e). Cloaking. Retrieved from <https://support.google.com/webmasters/answer/66355>
- Google. (2015f). Doorway pages. Retrieved from <https://support.google.com/webmasters/answer/2721311>
- Google. (2015g). Scraped content. Retrieved from <https://support.google.com/webmasters/answer/2721312>
- Google. (2015h). Affiliate programs. Retrieved from <https://support.google.com/webmasters/answer/76465>
- Google. (2015i). Webmaster guidelines. Retrieved from <https://support.google.com/webmasters/answer/35769>
- Lian, H. (2014). *Delivering user-friendly experiences by search engine optimization*. Unpublished manuscript, School of Business and Culture, Lapland University of Applied Sciences, Tornio, Finland.
- Manber, U. (2008). Introduction to Google search quality. Retrieved from <https://googleblog.blogspot.com/2008/05/introduction-to-google-search-quality.html>
- Moz. (2015). Google algorithm change history. Retrieved from <https://moz.com/google-algorithm-change>

- NetMarketShare. (2015a). Desktop search engine market share. Retrieved from <https://www.netmarketshare.com/search-engine-market-share>
- NetMarketShare. (2015b). Mobile/tablet search engine market share. Retrieved from <https://www.netmarketshare.com/search-engine-market-share>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: bringing order to the web*. Retrieved from Stanford University website: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
- The Search Engine Archive. Retrieved from <http://search-engine-archive.blogspot.com/2010/02/jughead.html>
- Wall, A. (2015). History of search engines: From 1945 to Google today. Retrieved from <http://www.searchenginehistory.com/>