

Understanding Search-Engine Optimization

Venkat N. Gudivada and Dhana Rao, East Carolina University

Jordan Paris, CBS Interactive

Because users rarely click on links beyond the first search results page, boosting search-engine ranking has become essential to business success. With a deeper knowledge of search-engine optimization best practices, organizations can avoid unethical practices and effectively monitor strategies approved by popular search engines.

As more organizations make the Web central to their mission and product lines, search-engine rankings are becoming essential to strategic marketing and sales. Each search-engine results page (SERP) presented in response to a user's request contains a series of snippets—clickable links that often include preview text to establish the webpage's relevance to the search. The snippet's SERP ranking, which is based on a complex algorithm that considers more than 200 factors, can determine whether a user visits that page. A snippet is organic if its webpage warrants placement high in the SERP listing solely because of intrinsic merit; it is inorganic, or sponsored, if the organization has paid a fee to gain that placement.

Obviously, a high ranking in the first SERP can greatly boost an organization's visibility. According to HubSpot, for example, 61 percent of global Internet users research products online, and 44 percent of online shoppers begin their experience with a search engine. Of these online shoppers, 75 percent never scroll past the first SERP.¹ The study also found that snippet type is important: 70 percent of the links that users click on are organic, with 60 percent going to the top three links.

These statistics imply that securing a high spot in the first SERP not only increases sales but also helps build the target user's trust, which can lead to better brand development. To achieve a high SERP ranking, webmasters use an assortment of practices, collectively referred to as search-engine optimization (SEO).

To better understand SEO's relationship to business success, we surveyed current SEO practices and examined possible future directions. To aid organizations new to SEO, as well as those with practices in place, we compiled a list of the best tools and resources that can help organizations create and maintain SEO strategies approved by popular search engines.

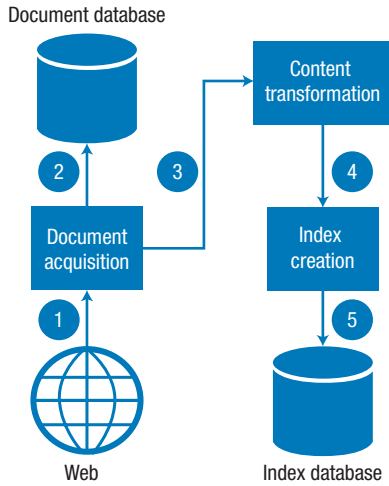


FIGURE 1. Indexing. Indexing involves fetching HTML documents, storing them in their original form, transforming the documents by processes such as stop-word removal and stemming, and generating indexes and storing them in a database.

SEO AND SEARCH ENGINES

Search engines are clearly foundational to SEO, but many organizations lack knowledge about how they work. Websites host a range of HTML documents, each with a unique Uniform Resource Locator (URL). A search engine enables Web searching by creating an index, a process transparent to the user, and responding to queries, a process that requires the user’s active participation.

Figure 1 shows a conceptual schematic of webpage (or document) indexing. The circled numbers correspond to the following steps:

1. A webcrawler, also known as a robot or bot, scours the Web to retrieve HTML pages.
2. The webcrawler stores these pages in their original form in its search engine’s document database.
3. The pages go through transformations, such as HTML tag and stop-word removal and stemming. The transformation, conducted by the search engine, extracts significant textual content and information about links for indexing.

4. The search engine creates indexes by generating direct and surrogate page representations, such as single words or phrases and their positional information on the page. It also notes information about incoming and outgoing links and generates a snippet.
5. The search engine stores indexes in its index database.

Figure 2 shows a conceptual schematic of the querying and document-retrieval process. The circled numbers correspond to the following steps:

1. The user employs a search engine’s browser to enter a search query—typically a single keyword or short phrase. As in step 3 of the indexing process, the search engine transforms the user’s query into a canonical representation.
2. The search engine’s query-ranking algorithm generates a ranked list of URLs for documents it deems relevant on the basis of the index database and contextual information in the user query. The search engine then shows the snippets corresponding to the ranked URLs to the user in SERPs.
3. The user browses the snippets and clicks on certain ones to retrieve the corresponding full documents in their original form from the document database.
4. The search engine’s retrieval-evaluation component helps the user further refine the search on the basis of feedback about the document’s relevance: the user

- explicitly indicates relevance (direct feedback) or clicks on relevant links (indirect feedback).
5. Using the relevance feedback, the search engine might reformulate the user query and re-execute it. This process repeats until the user is satisfied with the search results or ends the query session.
6. The search engine stores meta information such as user queries, relevance feedback, and clicked snippets in the log database, which it uses to improve its search performance.

SPAMDEXING

Not all companies follow approved SEO methods, and unethical practices, such as generating spam that subverts ranking veracity, have grown increasingly troublesome in the past five years. Indeed, these practices, combined with an escalating reliance on the Web as a marketing tool, have spawned companies and consultants whose specialty is to help organizations ensure that their websites reach the top of the first SERP.

Although most of these companies and consultants encourage approved, or *white-hat*, SEO practices, a fair number resort to deceptive and misleading, or *black-hat*, practices like spamdexing to get a high ranking. (See the “Two Cases, One Key Phrase” sidebar.) As a 2012 article on the consequences of Web spam noted, such search-engine spam is a serious problem, costing businesses with lowered rankings an estimated US\$130 billion annually.²

How it works

Spamdexing refers to an array of deceptive practices to secure top placement

in the first SERP by building webpages that trick search-engine algorithms and thus artificially boost the page's ranking. With spamdexing, even a page irrelevant to the search word or phrase can achieve a high ranking. Such practices affect both search-engine efficiency and results integrity, and have become sufficiently problematic that search-engine companies have included efforts to counter spamdexing in their strategic initiatives.

In SEO's early days, meta keywords were the basis for indexing a webpage. A webmaster could include several keywords with the meta keywords tag even though the page content might have little or no relevance to those keywords. To avoid the tag's abuse, all the major search engines stopped using it by 2009. However, webmasters still resort to *keyword stuffing*—the unnatural repeated use of a keyword or phrase simply to increase its frequency on the page.

Prevalent schemes

Current spamdexing schemes encompass automatically generated page content, deceitful redirects, cloaking, link spam, hidden text and links, doorway pages, spam from affiliate programs, embedded malicious behavior, and user-generated spam.^{3,4} The sidebar, "Google's Fight against Spamdexing," gives an idea of spamdexing's pervasiveness and the serious consequences of using this black-hat practice.

Automatic page generation. Scripts are used to generate webpages using algorithms that intersperse random text with desired keywords. The page might also include machine-translated text that has not undergone human

review and new text that obfuscates existing text through the use of synonyms. Some webmasters even use content taken verbatim from more reputable sites with no regard to relevance or copyright infringement. Webpages are also automatically generated from snippets of either search results or webpages that contain desired keywords. Such pages often contain just the generated snippets without any real content.

Redirecting. Some webpages send the user to a URL other than the one requested. URL redirecting has valid uses such as facilitating a website's move to a new address. However, some redirects are designed to show the user and a webcrawler different webpages for the same URL. Using this scheme, it is easy to index a page for children's stories but take the user to a page with pornographic images, for example.

Cloaking. Bots and users see the same page in different ways: bots see content

that has the desired keywords and follows approved guidelines, while users see content that is often malicious or undesirable. Some hackers use cloaking to keep the webmaster from detecting their work, for example.

Link spam. Any links in a webpage that exist solely to increase page rank are considered link spam, regardless of whether the links are outgoing or incoming. Link-spam schemes include buying and selling links, excessively exchanging links as in mutually linking partner pages, acquiring keyword-rich anchor text links through article marketing, and inserting comment links in blog postings.

Hidden text and links. Excessive keywords are hidden from a user but visible to search engines. Methods to hide text include placing text behind images, setting the text font size to zero, and using cascading style sheets to position text offscreen.

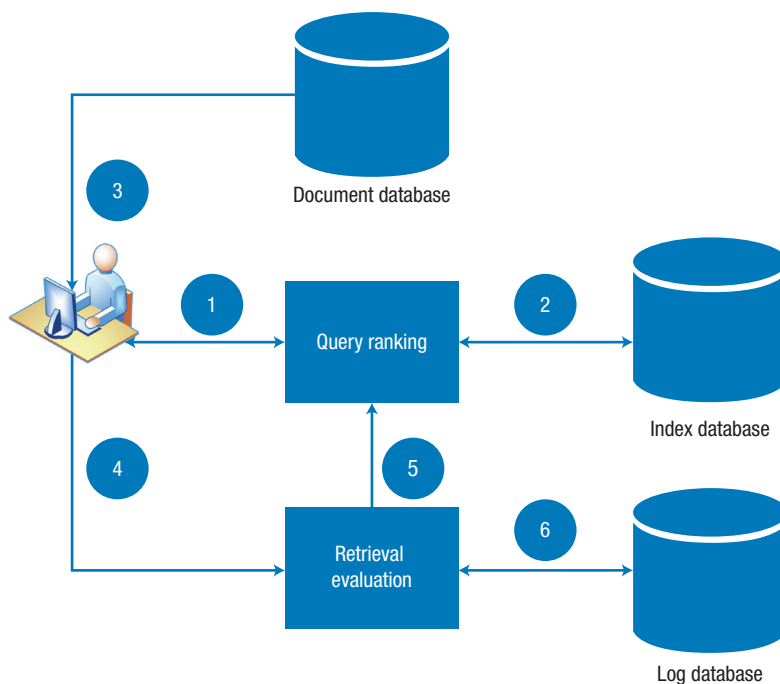


FIGURE 2. Querying and document retrieval. From a user query, an algorithm generates a ranked list of relevant documents, from which the user browses retrieved documents by clicking on the corresponding links. Queries can be refined and re-executed on the basis of user feedback, and the search engine stores meta information about the current search to improve its future performance

TWO CASES, ONE KEY PHRASE

Because competitive business advantage and confidentiality issues surround search-engine optimization (SEO) work, empirical data on SEO practices is rarely published. However, Seospoiler.com provided descriptions of two companies that ranked number one for the same key phrase but reached their ranking along different paths.

BLACK HAT

In late 2012, Greencowseo.com ranked number one for the coveted key phrase “SEO company.” However, it managed to have a disturbing number of incoming links from sites that were highly trusted but unrelated to SEO. Google removed Greencowseo.com from its index with the same alacrity that the SEO company used to secure the top spot.¹

WHITE HAT

In March 2014, Seoexplode.us (formerly Explodeseo.com) ranked number one for the same “SEO company” key phrase. However, although it used “SEO company” along with its other meta key phrases “SEO services” and “risk free SEO” as in the

page title (“SEO Company Providing Risk Free SEO Services”), its use of those key phrases was entirely ethical. Its meta description tag used “SEO company” and “SEO services” in a natural way, and of the 624 words on its homepage, the key phrases made up less than 3 percent each (2.56 percent for “SEO services” and 1.92 percent for “SEO company”). The keyword “SEO” was at 3 percent and hovered just at the high watermark. In 2014, the homepage had over 500 incoming links, all either directly or tangentially related to SEO.²

References

1. “SEO Company GreenCowSEO Exposed! Black Hat SEO Secret Techniques—Case Study,” *SEO Case Studies—Exposing Top SEO Companies & Secret Methods*, 2012; <http://seospoiler.com/seo-company-greencowseo-exposed-black-hat-seo-secret-techniques-case-study>.
2. “Explodeseo, SEO Company Reviews Case Study, How Top SEO Sites Rank,” *SEO Case Studies—Exposing Top SEO Companies & Secret Methods*, 2012; <http://seospoiler.com/explodeseo-seo-company-reviews-case-study-how-top-seo-sites-rank>.

Doorway pages. Doorway pages feature poor-quality content but are optimized to rank high for specific keywords. Their sole purpose is to funnel users to a single page, usually one they did not select.

Spam from affiliate programs. In affiliate marketing, a company rewards affiliates for applying various marketing schemes to bring customers to the company’s website. Schemes include automatically generating page content and using specific keywords to skew SEO. Such sites are penalized with lower search rankings.

Embedded malicious behavior. Here, the goal is to add a malicious action when the user clicks on a link, such as to install advertisements, viruses, malware, Trojans, and spyware on the user’s computer. The user might also click on a certain button and cause another, unwanted, link on the same page to activate.

User-generated spam. Not every website visitor has good intentions, and malicious users often generate comment spam—comments that include advertisements and links to unrelated pages. Some webmasters build poor-quality links to their competitors’ websites so that the latter get search-engine penalties, such as a lower ranking in the SERPs or removal from the index. This is enabled by *backlink blasting*—a software-driven link scheme to generate thousands of backlinks (external links pointing to the webpage). Both comment spam and backlink blasting are prevalent enough that Google provides a tool, Disavow, to deal with them.

IMPLEMENTING WHITE-HAT PRACTICES

To combat search-engine spam and help webmasters develop websites that adhere to white-hat practices,

search-engine companies provide starter guides and webmaster tools such as those listed in the sidebar “SEO Tools and Resources.” Starter guides contain conceptual details about how search engines index documents and process queries without revealing any strategic or proprietary information. Webmaster tools help assess the webpages’ conformance to guidelines and best practices. These resources are excellent starting points toward establishing or maintaining SEO practices that can help businesses achieve long-term strategic goals.

Organizations need to view SEO as a by-product of good website design that is part of an evolutionary organizational process, not an afterthought. SEO practices are constantly changing to fit both demand for high rankings and changing technology, so an organization must be vigilant about its SEO strategy’s fit with both business

GOOGLE'S FIGHT AGAINST SPAMDEXING

Google has penalized several prominent companies for employing black-hat SEO practices, including BMW, *Newsday*, JCPenney, Forbes, and Overstock.com.¹ Ironically, in late 2012, Google moved from litigator to litigant for similar issues after the European Commission maintained that Google was favoring its own products while pushing its competitors further down in the SERP rankings.²

BMW

In 2006, Google charged BMW with using doorway pages to funnel traffic to BMW's German website and briefly removed the site from Google Search results. BMW admitted to these charges and stated that it was not aware that doorway pages were search-engine spam. To reinstate its website in Google Search results, BMW had to remove the JavaScript that implemented doorway pages.

NEWSDAY

In 2007, Google charged *Newsday* with having outgoing links to unrelated websites, which is paramount to selling links to unrelated sites. However, it is considered an acceptable SEO practice if the links to unrelated websites include a `nofollow` attribute. The latter is a hint to search engines not to factor in the reputation of *Newsday.com* in calculating ranks for the pages targeted by the outgoing links.

JCPENNEY

During the 2010 holiday season, JCPenney used backlink blasting to boost its SERP position. A

subsequent *New York Times* investigation revealed that thousands of unrelated websites were linking to JCPenney.com. Google's corrective action resulted in a dramatic slide of JCPenney.com's SERP ranking.

FORBES

In 2011, Google charged Forbes with selling outgoing links on the Forbes website and penalized it with a lower search ranking. To lift the penalty, Forbes had to remove the offending links.

OVERSTOCK.COM

Google places enormous trust in links from education domain (.edu) webpages. In 2011, Overstock.com offered a 10 percent discount on their products to university students and faculty in exchange for the inclusion of its links on .edu sites. Because of the domain's high trust factor, Overstock.com's search ranking went up substantially. Google penalized the company by lowering its ranking in search results, which contributed to Overstock.com's 2011 fiscal year loss of US\$19.4 million. The company had to remove these links to reinstate its search results ranking.

References

1. A. Krush, "SEO Epic Fails: 5 Big Names Penalized by Google for Going Wrong with Their SEO," 2011; www.link-assistant.com/blog/seo-epic-fails5-big-names-penalized-by-google-for-going-wrong-with-their-seo.
2. S. Shankland, "Google Faces European Charge It Abused Search Dominance," *CNET*, 15 April 2015; www.cnet.com/news/google-faces-european-charge-it-abused-search-dominance.

requirements and current SEO best practices. Above all, organizations must avoid targeting one search engine and keep in mind that SEO's focus is on the website user, not the search-engine company.

Google Search is an example of how search engines evolve guidance and best practices. In 2011, Google Search added Panda, a quality-content filter whose primary goal was to ensure that low-quality sites ranked low and relevant sites appeared at the top of the SERPs. Penguin, the 2012 update to Panda, aimed to penalize sites that

employed link spam with lower rankings or even eliminate them from Penguin's index database. Hummingbird, released in 2013, aims to assign importance to pages according to the algorithm's semantic understanding of webpage content. Hummingbird incorporates intent and the keyword's contextual meaning in the user's search query. This algorithm has proven more effective in retrieving pages than an algorithm based on keyword frequency in the page. Pigeon, the latest update to Google Search's algorithm, incorporates user location and local

listings information in ranking search results.

On-page optimization

Broadly viewed, white-hat SEO has two major classes, the first of which is on-page optimization, which deals with website structure and content. Considerations range from word choice to the provision of a mechanism for restricted indexing.

Word choice. Creating precise, relevant, useful, and compelling content is a fundamental SEO requirement.

SEO TOOLS AND RESOURCES

Organizations must continually monitor search-engine optimization (SEO) practices, which requires more than a surface knowledge of available methods and their benefits, as well as an awareness of the pitfalls that can result from unethical strategies.

SEO BACKGROUND READING

- » Tips and information about Web searching, the search-engine industry, and SEO: <http://searchenginewatch.com>
- » Ranking of Internet marketing agencies and tools: www.topseos.com
- » Directories for listing websites: www.directorycritic.com
- » Article directories: <http://ezinearticles.com>, <http://goarticles.com>, www.articledashboard.com, www.galoor.com
- » Link exchange market: www.linkmarket.net
- » R. Fishkin et al., *The Beginners Guide to SEO* (comprehensive information for professional SEO), 8 Jan. 2015; <http://moz.com/beginners-guide-to-seo>
- » A. Williams, *SEO Checklist: A Step-by-Step Plan for Fixing SEO Problems with Your Website*, v1.5, 6 Aug. 2014; ezSEONews.com
- » Moz, "Search Engine Ranking Factors"

(characteristics of web pages that tend to rank higher): <http://moz.com/search-ranking-factors>

GENERAL TOOLS AND GUIDELINES

- » World Wide Web Consortium (W3C) markup validation service: <http://validator.w3.org>
- » W3C Cascading Style Sheets (CSS) validation service: <http://jigsaw.w3.org/css-validator>
- » SEO analyzer (to determine whether or not a webpage complies with 15 SEO best practices): www.bing.com/webmaster/help/seo-analyzer-97615e21
- » Bing webmaster guidelines (suggested practices for page structuring and content development to enable effective indexing by the Bing search engine): www.bing.com/webmaster/help/webmaster-guidelines-30fba23a
- » Bing webmaster tool suite for SEO: www.bing.com/toolbox/webmaster
- » Bing link explorer (for exploring backlinks to any site and gaining insight into which sites link to sites like yours): www.bing.com/webmaster/help/how-to-uselink-explorer-dddfa0a
- » Yahoo! webmaster resources: <https://help.yahoo.com/kb/SLN2248.html>

Content should naturally use keywords without concern for their frequency and contain natural and authentic word phrases, both short and long, that capture the page's topic. Content creators should bear in mind that users have different vocabularies and accommodate those differences through a mix of theme words and phrases. Content should look authoritative with relevant theme words and phrases spread throughout the page.

Anchor text. Relatively short but meaningful URLs are preferred for webpages, and natural keywords form the basis for URL text. Similar principles apply to anchor text. Ideally, anchor text should capture the topic of the page that the anchor link points to.

Semantic indexing. Search engines are starting to use more sophisticated indexing algorithms such as latent semantic indexing (LSI), which calculates a page's relevance not only on the basis of keywords but also on the page's overall topic. Consequently, strategies that focus exclusively on a select keyword can no longer guarantee the webpage's high ranking.

Non-HTML content. To ensure effective indexing, most content should be in HTML format, but content other than text requires supplemental information to enable indexing. Content creators should use the alt attribute to concisely describe image content and include a transcript for audio and video content.

Title and meta description tags. The title tag should reflect the page's topic, each page should have a distinct and appropriate title tag, and page content should be displayed in the snippet's first line. HTML5 semantic elements and heading tags (<h1> through <h6>) should reflect the page content's hierarchical organization. Although search engines do not use the meta description tag's keywords and phrases in calculating the page's ranking, they can use them to generate snippets. Thus, it is wise to have a precise and concise meta description tag for each page.

Navigational search. Users conduct a navigational search to find a particular webpage. Including text in the webpage title, body, and first four heading levels as well as the anchor text of

- » Google webmaster guidelines (best practices to help Googlebot find, crawl, and index websites): <https://support.google.com/webmasters/answer/35769?hl=en>
- » J. DeMers, "How to Use Google Webmaster Tools to Maximize Your SEO Campaign," <http://searchenginewatch.com/sew/how-to/2273660/how-to-use-google-webmaster-tools-to-maximize-your-seo-campaign#>
- » Google analytics for identifying SEO opportunities in websites: www.google.com/analytics
- » Generating a sitemap automatically: www.xml-sitemaps.com
- » Generating a robots.txt file: www.internetmarketingninjas.com/seo-tools/robots-txt-generator
- » Creating .htaccess and .htpasswd files for protecting special directories on web servers: www.yellowpipe.com/yis/tools/htaccess_generator/index.php
- » Analyzing websites for organic SEO: www.semrush.com, www.webseoanalytics.com
- » Webpage load speed test: <http://tools.pingdom.com/fpt>
- » Checking domain history: www.domaintools.com
- » Internal link analysis: <http://tools.seoachat.com/tools/page-link-analyzer-seo>
- » Building backlinks through press releases: www.free-press-release.com
- » Open Site Explorer: The Search Engine for Links (performing competitive link research, exploring backlinks, and evaluating anchor text): www.opensiteexplorer.org
- » Generating a website privacy policy: www.easyriver.com/myprivacy.htm, www.freeprivacypolicy.com

CONTENT AND COMPLIANCE EVALUATION TOOLS

- » Tools for checking website content for duplication and plagiarism: <http://copyscape.com>, www.dustball.com/cs/plagiarism.checker, <http://plagiarismdetect.com>, www.plagtracker.com, www.comparemyfiles.com
- » Analyzing keyword density: www.internetmarketingninjas.com/seo-tools/keyword-density
- » Fetch as Bingbot (see page exactly the way it appears to Bingbot): www.bing.com/webmaster/help/etch-as-bingbot-fe18fa0d

inbound links and their number can help achieve higher ranking in a navigational search.

Privacy policy. All websites should include a page that describes the site's privacy policy, such as what personal information the site collects and its use and distribution. The privacy policy can help further convey the image of a professionally managed website, and some search engines consider its inclusion as a measure of a website's trustworthiness.

Custom 404 page. Webservers return a 404 page when the search engine cannot find the requested webpage. Customizing the 404 page helps keep users on the site and can even enhance their search experience. Customization might involve adding a pointer to the

site's homepage or providing links to other site content related to the search.

Restricted page indexing. The webmaster might not want to index certain website pages—for example, when a page has too many links or privacy is a concern. Also, certain content might be intended for registered users. In such cases, the webmaster can use the robots.txt file to inform web crawlers which pages are accessible for indexing, but not all search engine indexers are guaranteed to respect the robots.txt protocol.

Off-page optimization

The second white-hat SEO class is off-page optimization, which addresses best practices in incorporating both inbound and outbound external links. A carefully designed directory

structure for website content not only helps with site maintenance, but also enables bots to traverse a website and index its content. Website navigation structures include breadcrumbs and sitemaps—both of which users and web crawlers should find natural and intuitive to traverse. Web crawlers might have difficulty navigating through drop-down menus created using JavaScript, for example. To enable comprehensive indexing, all links should be in text, not images.

Navigation structures also include robots meta tags, which control indexing at the page level; text links that web crawlers follow to retrieve the corresponding documents; and backlinks, which determine the page's reputation and contribute to its score assignment.

Breadcrumbs. Breadcrumbs show the user's navigation trail through the website and are intended primarily to improve site usability. If the breadcrumb information is available as HTML markup in the body of a webpage, some search engines include it in SERPs. To avoid penalties, usually reduced ranking, webmasters should ensure that navigation does not create distinct URLs for the same content.

Sitemaps. Sitemaps are an important element of well-designed websites because they depict the website's structure, which facilitates site navigation. A sitemap should be in both XML format for the search engine and plain-text format for the user.

The XML sitemap version should feature information about every website page, including the page's URL, last modified date, page-update frequency, and the URL's priority value relative to that of the site's other webpage URLs. Most important URLs will have priority 1 (highest), with lower values indicating decreased importance. Search engines use these values to determine the webpages' indexing order. Because a search engine might index only some pages, the values are also a way for the webmaster to promote the most important pages. Webmasters typically submit the XML sitemap version to the search engine because it is more likely to result in a complete site indexing.

The plain-text sitemap version is useful if the site visitor cannot get to the desired content using navigational structures. To increase ease of use, the plain-text version should have each URL on a separate line. If the site features many images, the webmaster should provide search engines with an

image sitemap with a structure similar to the XML version. Information should include a caption and title, geographic location (city, country), and a URL to the image's copyright and licensing terms.

Text links. If the site lacks a map, text links are critical in enabling webcrawlers to navigate among pages. Crawlers also index text links—indexing information that search engines use—although not all pages are indexed. For example, the webmaster might have failed to submit the required forms to the search engines, the pages could have links and other plug-ins buried in JavaScript code that the search engine cannot parse, or the page might contain an excessive number of links or a robots meta tag or robots.txt file that is blocking other pages.


Backlinks. Backlinks help secure a higher SERP ranking, but link quality is critical. Link spam and other black-hat methods, such as creating link farms (a website group in which each site links to all other sites in the group) and inserting links in a blog's comments section, can artificially boost backlinks and thus the webpage's importance. When artificial links are unavoidable, webmasters can use the `rel=nofollow` attribute in the HTML anchor element to inform search engines that they should discard the link when computing the rank of the page the link points to. If a page points to external links, those links should be pointing to trusted and authoritative sites. Otherwise, the external links risk incurring penalties for that page and, in some cases, even the entire site's removal from the search engine's index.

Robots meta tag. The robots meta tag specifies whether or not a webcrawler can index a page (`index/noinindex`) or whether or not it can traverse the links on a page (`follow/nofollow`). The `noarchive` directive prohibits search engines from saving a cached copy of the page; the `nosnippet` directive keeps them from displaying the page snippet in the SERPs. The `noodp` directive instructs search engines not to use descriptive text of the page from the Open Directory Project (DMOZ). With these bans, webmasters can use a page-specific approach to control how search results should index individual pages and show them to users. Although webmasters employ the robots meta tag link attribute `rel="nofollow"` to prevent link-injection spam, search engines also use it as a way to discount any link value that they would otherwise include in computing page importance. The X-robots-tag HTTP header can also be used to specify this indexing and traversal permission.


MONITORING SEO PERFORMANCE

SEO performance monitoring should be an ongoing activity. Webmasters need to respond to changing guidelines and suggested SEO strategies from search-engine companies to ensure that the practices they are following are in compliance. Also, webmasters must maintain constant vigilance to detect negative or subversive SEO from competitors and unsuspecting third parties, monitoring closely for undermining activities such as the insertion of undesirable backlinks.

Web analytics programs can help in performance monitoring, providing insights into how to ethically improve SERP ranking. They can inform the



SMALL VISUAL INTERFACES WILL CONTINUE TO POSE CHALLENGES FOR SPECIFYING QUERIES AND PRESENTING SEARCH RESULTS.



webmaster about how visitors reach the website, their navigation patterns, and how changes to the page title and meta description tags affect search-engine traffic. The programs enable simulated content and site organization experiments in a sandbox environment so that webmasters can build multiple page versions and evaluate them in what-if scenarios.

IMPLICATIONS OF EMERGING TRENDS

As organizations and individuals demand more ways to leverage Web content for commercial, political, and personal gain, efforts to promote ethical practices as well as thwart unethical attempts to raise search-engine rankings are likely to change the face of SEO.

Adversarial information retrieval

Adversarial information retrieval (AIR) is a new subfield of information retrieval whose focus is on subverting search engines to effect higher ranks for pages using dubious and often unethical techniques. Growing AIR sophistication is raising complex problems for search-engine companies as they attempt to develop effective AIR countermeasures.

Several AIRWeb workshops and other forums held at the annual WWW and ACM SIGIR conferences attest to the increasing significance of AIR research. Examples include employing user-behavior features for separating Web spam pages from ordinary pages⁵ and a method to combat new types of spam based on click-through data analysis.⁶

Social media content

Content from LinkedIn, ResearchGate, Twitter, Google+, Pinterest, Instagram and other social networking sites is

becoming more important for search engines. Indeed, as more people turn to social media to find relevant information, the distinction between Web search and social search is blurring. Facebook Search (<http://search.fb.com>) is an example of this company's foray into the social search space. Social Searcher (www.social-searcher.com) is a social media search engine that looks for publicly posted information in social networks.

Mobile device use

According to a 2014 Mobile Path to Purchase survey by Telemetrics and xAd (www.mobilepathtopurchase.com), 50 percent of respondents use mobile devices to start the search process and two out of three mobile shoppers ultimately make a purchase. Despite these statistics, some websites still do not work properly with mobile devices. Responsive Web design is emerging to address this issue.

A concomitant effect of increased use of mobile devices for search is the growing use of voice commands rather than text queries. Under this new paradigm, users employ multiword phrases and nearly complete sentences as queries rather than keywords.

Locale information will assume greater prominence as a factor in ranking search results. This will help search engines show information that is not only relevant but also has the potential for the user to take immediate action. Search engines will introduce the practice of user registration to receive higher-quality search results.

Information fusing

Some complex information needs are extremely difficult to fulfill using

search engines. Meeting such needs involves fusing information from multiple sources, which typically requires human expertise. Community-based question-answering systems are used for this purpose and will gain more visibility. For example, Stack Overflow (www.stackoverflow.com) and Slashdot (www.slashdot.org) are question-answering systems serving specialized vertical search spaces. A new search engine, Blekko (www.blekko.com), employs crowdsourcing for content selection and curation.

Related to community-based question answering is collaborative search, which involves a group of users with the same information need. Examples include students working together to search for information for a term paper, and team members combining efforts to assess the environmental impacts of a proposed river dam.

Web search engines will morph into metasearch engines. Behind a user-facing search engine, there will be several search engines behind the scenes, each targeting a different vertical search space.

Shrinking interfaces

Small visual interfaces will continue to pose challenges for specifying queries and presenting search results. For example, search results might be returned as info cards—condensed information that information-extraction algorithms generate from unstructured data—instead of as links, which requires using information-extraction techniques to generate condensed factual information. Also, personal and contextual information will play a greater role in tailoring search results to the limited number that a small interface can handle.

ABOUT THE AUTHORS

VENKAT N. GUDIVADA is a professor and chair of the Department of Computer Science in the College of Engineering and Technology at East Carolina University. His research interests include database management, information retrieval, high-performance computing, and personalized electronic learning. While conducting the research for this article, Gudivada was a professor of computer science at Marshall University. He received a PhD in computer science from the University of Louisiana at Lafayette. He is a member of the IEEE Computer Society. Contact him at gudivadav15@ecu.edu.

DHANA RAO is an assistant professor in the Department of Biology in the Thomas Harriot College of Arts and Sciences at East Carolina University. Her research interests include microbial ecology and bioinformatics. While conducting the research for this article, Rao was an assistant professor of biological sciences at Marshall University. Rao received a PhD from the University of New South Wales. She is a member of the American Society of Microbiology. Contact her at raodh15@ecu.edu.


JORDAN PARIS is an associate software engineer at CBS Interactive. His research interests include information retrieval and high-performance computing. While conducting the research for this article, Paris was an undergraduate in computer science at Marshall University, where he received a BS in computer science. Contact him at jordan.paris@cbsinteractive.com.

Long-term and sustainable high ranks for webpages will come from adhering to SEO guidelines and best practices coupled with an emphasis on high-quality authentic content authored naturally. The resulting Web domain trustworthiness and authority will reap dividends in an environment where search engines enforce zero tolerance of spam.

Webmasters should exercise due diligence when deciding to include third-party plug-ins and services in their webpages, particularly those recommended by affiliates. Hundreds of websites are

penalized daily for employing black-hat SEO practices. Webmasters should remove any undesirable and damaging backlinks that past SEO consultants have inserted.

Rapidly maturing natural-language translation and machine-learning technologies have fundamental implications for Web search. It is not far-fetched for a search-engine user to issue a query in one language, and expect to retrieve relevant documents authored in multiple other languages that are automatically translated into the language in which the query is issued.

As these and other technologies emerge and mature, SEO methods will evolve accordingly, but SEO will continue to be an integral part of an organization's long-term strategic business plan. 

REFERENCES

1. HubSpot, "All the Marketing Statistics You Need," 2014; www.hubspot.com/marketing-statistics#SEO.
2. N. Spirin and J. Han, "Survey on Web Spam Detection: Principles and Algorithms," *ACM SIGKDD Exploration Newsletter*, vol. 13, no. 2, 2012, pp. 50–64.
3. "Bing Webmaster Guidelines," www.bing.com/webmaster/help/webmaster-guidelines-30fba23a.
4. Google, "Webmaster Guidelines," <https://support.google.com/webmasters/answer/35769?hl=en>.
5. Y. Liu et al., "Identifying Web Spam with the Wisdom of the Crowds," *ACM Trans. Web*, vol. 6, no. 1, 2012, pp. 2:1–2:30.
6. C. Wei et al., "Fighting against Web Spam: A Novel Propagation Method Based on Click-through Data," *Proc. 35th Int'l ACM Conf. Research and Development in Information Retrieval (SIGIR 12)*, 2012, pp. 395–404.



Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.