BY ROSS A. MALAGA

Worst Practices in Search Engine Optimization

MANY ONLINE COMPANIES HAVE BECOME AWARE of the importance of ranking well in the search engines. A recent article reveals that 62% of search engine users click only on results that appear on the first search engine results page (SERP) and less than 10% of users click on results that appear after the third page.[3]

In order to place well in the SERPs companies have begun to use search engine optimization techniques (SEO). That is they manipulate the site's content and meta tags, as well as attempt to attract incoming links from other sites. However, certain SEO techniques directly violate the guidelines published by the search engines. While the specific guidelines vary a bit, they can all be summed up as: show the same content to search engines as you show to users.

Failure to conform to search engine guidelines can lead to penalties, such as worse placement in the SERPs or an outright ban from the search engine. Consider the case of BMW's German Web site (www.bmw.de). On February 7, 2006, Google banned this site for using a "doorway" page, essentially showing one page to the search engines and a different page to humans. According to a Google spokesperson,

"Google may temporarily or permanently ban any site or site authors that engage in tactics designed to distort their rankings or mislead users in order to preserve the accuracy and quality of our search results."[1]

This article examines some of the techniques that can lead the search engines to ban a site—so called "black hat" techniques. It is important for all webmasters and those that outsource their search engine optimization programs to understand these techniques and the impact they can have on search engine placement. One problem faced by legitimate sites is that black hat sites may rank well for short periods of time (before they are banned). High-ranking black hat sites will push legitimate sites down in the SERPs. In fact, many black hats make a living by automatically generating thousands of sites that rank well for a short period of time. Many of these sites make only a few cents a day, but multiplied by thousands or tens of thousands of sites, it adds up to a lucrative business.

Another problem is that many black hat optimizers openly steal content from legitimate sites. A thriving consulting business has sprung up to provide search engine optimization services. While many consultants use "white hat" methods (those that are not likely to lead to a penalty or ban), some use black hat techniques. For example, according to Google insider Matt Cutts' Blog[2] the SEO consulting company Traffic Power was banned from the Google index. In addition, Google also banned Traffic Powers' clients. The worst black hat optimizers use techniques aimed at having their competition penalized or banned by the search engines.

We discuss search engine optimization, then examine black hat indexing techniques, followed by on-page and off-page methods. We also discuss how black hat optimizers manipulate the rankings of their competitors.

Search Engine Optimization
A search engine is simply a database of Web pages, a method for finding Web

pages and indexing them, and a way to search the database. Search engines rely on spiders — software that follows hyperlinks — to find new Web pages to index and insure that pages that have already been indexed are kept up to date.

According to Wikipedia,[6] "Search engine optimization (SEO) is a set of methods aimed at improving the ranking of a Web site in search engine listings..." These methods include manipulation of dozens or even hundreds of Web site elements. SEO can be broken into four major categories: key word/ key phrase research and selection, getting the search engines to index the site, on-page optimization, and off-page optimization.

During the first phase a list of key words and/or phrases are developed. These are the terms a user would type into the search engine that would lead to the site appearing in the SERPs. In addition to developing a list of words and phrases, the SEO professional will usually determine how competitive each term is and how often each terms is used in a search.

Phase two is concerned with quickly getting the search engines to index the site. This is usually accomplished by submitting sites directly to the search engines, having a site that is already indexed include a link to the target site, or the use of black hat methods that are described below.

During the third step, the Webmaster or SEO professional will manipulate various on-page components, such as Meta tags, page content, and site navigation in order to improve the site in the SERPs. For example, a number of researchers, including Malaga,[4] Raisinghani,[5] and Zhang and Dimitroff[8] have found that sites that make proper use of Meta tags achieve better search engine results. Zhang and Dimitroff[8] also found that sites with key words that appear in both the site title and throughout the site's text achieve better search engine rankings than sites that only optimize the title.

Finally, the major search engines all consider the number and relevance of links from external sites to the target site. Therefore, SEO projects usually include a link building phase (also called off-page optimization). During this phase optimizers request links from Webmasters and may use link building programs.

Black Hat Indexing Tricks

One of the primary tricks black hat SEOs use to attract search engine spiders is called Blog-ping (BP). This technique consists of establishing hundreds or even thousands of Blogs. The optimizer then posts a link to the new site on each Blog. The final step is to continually ping the Blogs. Pinging automatically sends a message to a number of Blog servers that the Blog has been updated. The number of Blogs and continuous pinging attracts the search engine spiders that then follow the link.

It should be noted that many white hat SEOs use the BP technique in an ethical manner. That is, they post a link to the new site on one (or a few) Blogs and then ping it only after an update. This method has been shown to attract search engine spiders in a few days.[4]

On-Page Black Hat Techniques

Black Hat optimizers use a variety of on-page methods. Most of these are aimed at providing certain content only to the spiders, while actual users see completely different content. The reason for this is that the content used to achieve high rankings, may not be conducive to good site design or a high conversion rate (the rate at which site visitors perform a monetizing action, such as make a purchase). The three main methods that fall into this category are cloaking, doorway pages, and invisible elements.

The purpose of cloaking is to achieve high rankings on all of the major search engines. Since each search engine uses a different ranking algorithm, a page that ranks well on one may not necessarily rank well on the others. Since users will not see a cloaked page, it can contain only optimized text — no design elements are needed. So the black hat optimizer will set up a normal Web site and individual, text only, pages for each of the search engines. The final step is to monitor requesting IP addresses. Since the IP addresses for most of the major search engine spiders are well known, the optimizer can serve the appropriate page to the correct spider (see Figure 1).
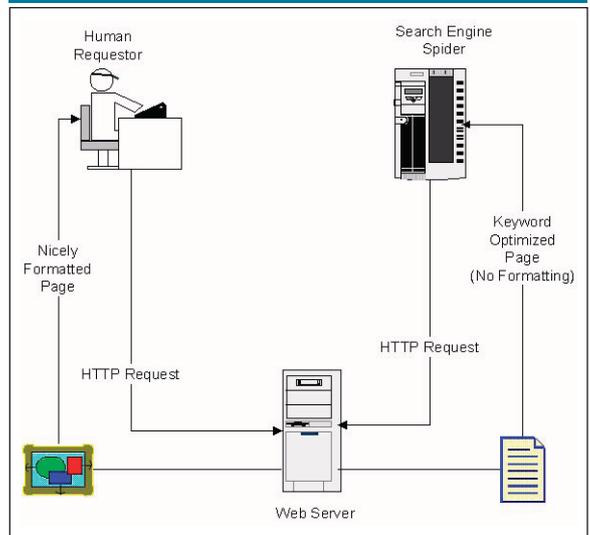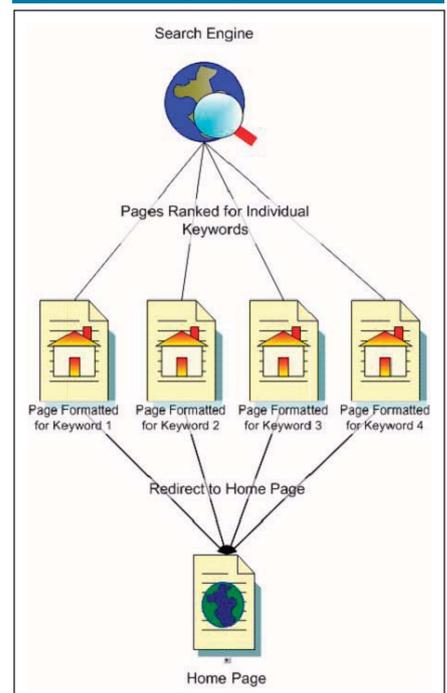

Figure 1. Cloaking

Figure 2. Doorway_Pages

If the requestor is not a spider the normal Web page is served.

The goal of doorway pages is to achieve high rankings for multiple keywords or terms. The optimizer will create a separate page for each keyword or term. Some optimizers use hundreds of these pages. Doorway pages typically use a fast meta refresh to redirect users to the main page (see Figure 2). A meta refresh is an HTML command that automatically switches users to another pages after a specified period of time. Meta refresh is typically used on out of date Web pages – for example you might see a page that states "you will be taken

to the new page after 5 seconds" A fast meta refresh occurs almost instantly, so the user is not aware of it. All of the major search engines now remove pages that contain meta refresh. Of course, the black hats have fought back with a variety of other techniques, including the use of Javascript, PHP, and other Web programming tools. This is the specific technique that caused Google to ban bmw.de and ricoh.de.

Invisible content is not new, but has been revived recently. Early optimizers used HTML elements such as the same foreground and background colors, or very small fonts to add invisible content to their sites. However, the search engines quickly caught on to these techniques and began to penalize sites that used them. More recently optimizers have taken to using cascading style sheets (CSS) to hide elements. The elements the optimizer wants to hide are placed within hidden div tags. Google, for one, has begun removing content contained within hidden div tags from its index. However, this may cause a problem for legitimate Web site developers who use hidden divisions for design purposes.

Black hat optimizers also make use of tools that allow them to automatically generate thousands of Web pages very quickly. These so-called content generators actually search the Web for keywords and terms specified by the optimizer. The software then basically copies content from other sites and includes it in the new one. Content generators represent a problem for legitimate Web owners as their original content may be copied extensively. Since some search engines penalize duplicate content, legitimate sites may also be penalized.

Off-page Black Hat Techniques

All of the major search engines consider the number and quality of incoming links to a site as part of their algorithm. Links are especially important for ranking well on Google. Therefore, black hat optimizers use a variety of methods to increase their site's back links (links from other sites).

One of the simplest black hat linking techniques is guest book spamming. Optimizers simply look for guest book programs running on authority (usually .edu or .gov) sites. They then simply add a new entry with their link in the comments area.

Black hat optimizers might also create or make use of existing link farms. A link farm is a group of pages created for the sole purpose of containing links to each other. Link farms are usually created using automated tools.

One popular off-site black hat method is HTML injection, which allows optimizers to insert a link in search programs that run on another site. For example, WebGlimpse is a Web site search program widely used on academia and government Web sites. The Stanford Encyclopedia of Philosophy Web site located at plato.stanford.edu, which has a Google page rank of 8 (links from sites with a high page rank are highly valued), uses the WebGlimpse package. So an optimizer that would like a link from this authority site could simply navigate to

http://plato.stanford.edu/cgi-bin/webglimpse.cgi?nonascii=on&query=%22%3E%3Ca+href%3Dhttp%3A%2F%2F##site##%3E##word##%3C%2Fa%3E&rankby=DEFAULT&errors=0&maxfiles=50&maxlines=30&maxchars=10000&ID=1.

The optimizer then replaces ##site## with the target site's URL and ##word## with the anchor text.

SEO Glossary

Blog-ping (BP) – a method for attracting search engine spiders that involves the creation of hundreds (or even thousands) of Blogs and then continuously pinging Blog servers (telling the servers that the Blog has been updated).

Bowling – techniques used to cause competitors to receive a search engine penalty or ban.

Cloaking – a technique that shows one set of Web pages to search engines and another set to humans.

Content generator – software that searches the Web for specified content and then copies that content into a new Web page.

Doorway pages – a set of Web pages, each of which is optimized for a particular keyword. Each page redirects users to a Web page designed for humans.

HTML injection – occurs when a user exploits a security vulnerability in Web site search programs, by sending the program a search string which contains special HTML characters. These characters cause the insertion of data specified by the user into the site.

Meta refresh – an HTML meta tag used to redirect users to a different Web page after a certain amount of time. A fast meta refresh redirects the user with no time delay.

Meta tag – HTML tags placed in the header section of a Web page that provides that provides metadata about the page.

Search engine optimization (SEO) – methods aimed at manipulating a Web site and links to a Web site for the purpose of improving the site's ranking on the search engine results pages.

Search engine results page (SERP) – the listing of Web pages returned by a specific query on a search engine

Bowling the Competition

One of the most insidious black hat methods is manipulating competitors' search engine results. The result is a search engine penalty or outright ban (a term black hatters call bowling). The incentive for this type of behavior is fairly obvious. If a black hat site is ranked third for a key term, the optimizer who can get the top two sites banned will be ranked first.

There are a number of techniques that can be used for bowling. For instance, the HTML injection approach discussed above can be used to change the content that appears on a competitor's site. If a black hat optimizer is targeting a site that sells computers, for example, the HTML injected might be <h1>computer, computer, computer... The extensive use of keywords over and over again is almost guaranteed to lead to a penalty or outright ban in all the major search engines.

Since the major search engines, and Google in particular, use the quality of the links coming into a site to determine rankings, black hat optimizers manipulate these links in order to negatively impact competitors. For instance, a black hat might request links to the competitor's site from link farms, gambling sites, or adult oriented sites. Links from these bad neighborhoods result in penalties and bans.

Conclusion

Clearly the growth and popularity of Web search is an indication that Webmasters and online marketers must consider search engine optimization as part of their overall marketing plans. However, those that pursue SEO are up against an arsenal of black hat techniques. In addition, even those optimizers who try to stay on the white hat side may find that they have inadvertently crossed the line leading to penalties or even a ban.

So, how should one proceed with SEO? When hiring an SEO consultant there are a number of factors to consider and questions to ask. First, see how the consultant or company ranks for the term "search engine optimization" on the major search engines. Obviously, if a consultant cannot rank well for their own site, it is not likely that they will succeed with your site. Second, you choose the key words and terms you want to

rank well for. While a good SEO consultant may make recommendations, you must make the final decisions. Many unscrupulous consultants guarantee a high ranking, but you may find that the key words are not very competitive or searched for often. Third, do not turn over your site to a consultant. The consultant should recommend changes, but black hats have been known to insert too many key words and even add automated software to sites to connect them to link farms. Finally, get references. Be sure to ask about specific results and return on investment.

Of course there is no requirement to hire an outside consultant. Many sites have done very well handling their SEO in-house. There are numerous online resources, such as www.seochat.com, www.searchenginewatch.com, and forums.digitalpoint.com that provide a wealth of excellent information on SEO.

While they might like to, Webmasters cannot simply ignore black hat optimization. Black hat methods may lead to worse rankings for white hat sites – through black hat sites that rank well temporarily and techniques aimed at bowling sites. In addition, white hats should not ignore black hat approaches as they can learn or adapt new SEO methods from them. For example, many white hat optimizers have successfully used the Blog and ping approach, in a more moderate manner, to achieve quick search engine indexing. ▣

References
1. CNN. Google blacklists BMW Web site. (Feb. 7, 2006); www.cnn.com/2006/BUSINESS/02/07/google/
2. Cutts, M. Confirming a penalty. (Feb. 11, 2006); www.mattcutts.com/blog/confirming-a-penalty/
3. iProspect. iProspect Search Engine User Behavior Study. www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior/
4. Malaga, R.A. The value of search engine optimization – An action research project at a new e-commerce site, *Electronic Commerce in Organizations*, 5, 3, (2007), 68-82.
5. Raisinghani, M. Future trends in search engines, *Electronic Commerce in Organizations*. (Jul-Sep 2005) 3, 3.
6. Wikipedia. Search Engine Optimization. en.wikipedia.org/wiki/Search_engine_optimization/
7. Zhang, J. and Dimitroff, A. The impact of metadata implementation on webpage visibility in search engine results (Part II). *Information Processing and Management* 41 (2005). 691-715.
8. Zhang, J. and Dimitroff, A. The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing and Management* 41 (2005) 665-690.

Ross A. Malaga (malagar@mail.montclair.edu) is an Associate Professor in the School of Business at Montclair State University in Montclair, NJ.