

WC JOURNAL

Site Design for Search Engines

Will Weidman
Weidman Consulting, Inc.
will@weidmanconsulting.com
(585) 787-7860

December 17, 2002 Issue IV

This month's article looks at how the design of your Web site affects its search engine compatibility. Compatibility in this sense indicates how easy it is for the search engine crawlers to find and extract information from your Web site to populate their databases. If you're considering a new site or a significant redesign, make sure that search engine "friendliness" is considered during the initial design phase.



Crawlers, Spiders, and Bots

It's not science fiction. It's not horror. It's simply a set of names that have been given to the programs developed by search engine companies to automatically visit your Web site and gather information from it. The task of indexing data from millions of Web sites is significant, so it was out of necessity that software programs were designed to visit sites and populate the large databases that drive the search engines. During their periodic trips to your site, they pay attention to two things: **text** from the pages and **links** to other pages. The links show the crawlers where to find other pages to index. The text feeds the database so that search terms can be matched. Crawlers ignore images, and they frankly don't care what your site looks like. If you've ever wondered how a horrible-looking site placed higher in the search results than your graphically appealing site, that's the reason. It's all about the text.

Now that you understand what they're looking for, it's easier to see how the design of your site impacts the crawler's ability to properly do its job. First we will look at the things that make indexing the site difficult.

Inhibitors

Text Embedded In Images – Since crawlers ignore images, any text embedded into your images is crawler "invisible." From the designer's perspective, image-based text gives them ultimate control over the appearance of the text on the page. HTML text is not as easy to control, and therefore is less desirable if a precise look and feel is needed. It's not necessary to eliminate all

embedded text, just make sure that the bulk of your content, including your important keywords, is in HTML.

Frames – HTML allows the use of frames to build Web pages. When frames are employed, your browser window is visibly (or invisibly) divided into smaller browser windows, each presenting its own separate Web page. When a search engine crawler encounters a framed page, it indexes only the *main* frame, not all of the framed pages. Generally, there is little or no content in the main frame since it acts as a “container” for the secondary ones. If you have ever seen a page in your search results with an odd description like, “Your browser does not support frames,” this is the reason. The only content that the crawler found to index was the short text message intended for users whose older browser does not support a framed display.

Flash – Macromedia Flash is an animation technology widely used in Web designs. While Flash is excellent for adding dynamic content to Web sites, search engine crawlers cannot see any text contained in the animation. The guidelines for Flash are the same as they are for images: keep the critical text in HTML.

Image Maps – Image maps extend the usefulness of images. The map is a set of “hot spots” defined for the image so that each area can be assigned its own link. A common example is an image of the United States that allows you to select an individual state to browse into. The image map defines the pixel locations of the state borders. Search engine crawlers will not follow the links defined in image maps. If your site’s navigation scheme utilizes them make sure that there’s an alternate text-only set of links placed elsewhere on the page (commonly in the footer).

JavaScript – JavaScript is one of many technologies that help make your site more dynamic. Your browser interprets the JavaScript programming instructions, which is why crawlers have an issue with it. Think of the crawler as a bare bones Web browser, one that needs to process your page as quickly as possible. Your Web browser (Internet Explorer or Netscape Navigator, most likely) has lots of features built in, including a JavaScript interpreter. The time to interpret programming instructions on the fly is a luxury the crawler can’t afford. If you have text or links that are JavaScript generated, they will be missed. As with all of our inhibiting design factors, judicious use is advised.

Invalid HTML – The most popular Web browsers are very forgiving about sloppy programming. If a page does not strictly adhere to the HTML specification it may still display properly, and the careless coding might never be noticed. As we learned above, the crawlers aren’t programmed with all of the features of a full Web browser, and it turns out that they’re not as lenient with coding errors either. If you have mistakes in your pages, critical content might be missed when the crawler encounters the errors. There are numerous HTML validators on the Web (e.g., <http://validator.w3.org/detailed.html>). Have your developer make sure that your pages pass the test.

Enablers

On to the list of design enablers – those things you can do to make your site easier to index (or be found).

Web Site Domain Name – Having a domain name that includes your keywords can be a significant advantage in obtaining higher rankings. If you are lucky enough to own one of these valuable domain names make sure you keep paying the renewal fee.

Web Page Title Tags – Each page in your site is treated separately by search engines, and the title tag (whose contents are visible to the page visitor at the top of their browser window) for each page should be unique and include relevant keywords. Avoid the common error of including your company name in every title tag. More often than not, the prospects you’re trying to reach via search engines won’t be searching for your company by name. Keep your title tags concentrated on your keywords and phrases.

Text – Not that it hasn’t already been highlighted enough to this point, but here’s one more reminder to use keyword rich HTML text on all of your critical pages.

Text Navigation – There are many ways to implement the navigation links for your site, and we described some above that will cause problems. Text-based links for navigation are the best from a crawler’s perspective. It’s wise to include them for site usability as well. Since they load quickly, users can click ahead to their desired destination without having to wait for all of the page images to load.

Site Map – The ideal scenario is that there is a link to every page on your site from the home page. This sets up the crawlers to be only one “click” away from any spot in your site (crawlers do limit how deep they go). In many cases, it’s impractical to have so many links on the home page. A good alternative is a site map page that includes text links to all of your site pages. By placing a link to the site map on your home page, the crawler is two “clicks” away from any page. Not as good as being one away, but still pretty good.

KEYWORDS and DESCRIPTION Tags – Although not required, many pages include the KEYWORDS and DESCRIPTION tags. These are embedded in the Web page code, and are not immediately visible in the Web browser. Originally, these were defined to enable the programmer to describe the contents of the page, and, understandably, search engines read and indexed the contents of these tags. But these tags became easy to abuse by stuffing them with repeated or irrelevant keywords. For that reason, most search engines now ignore one or both tags. But because some search engines still pay attention to them, they should be populated. There is no penalty for having them there.

Internal Links – An extension of the site map guideline, make sure that you provide lots of links from the pages in your site to the other pages in your site. Give the crawlers as many ways as possible to find your pages to index.

Image ALT Tags – HTML provides a mechanism to describe the contents of each image on a Web page for impaired users and for browsers that do not support graphics. The image ALT tag is designed for this purpose and should be used to accurately describe the image. Include keywords and phrases where appropriate.

Headings – HTML uses a set of heading tags to define “more important” text. The intent was that they would be used for paragraph and section headings. For that reason, text in heading tags is weighted more significantly than normal body text when calculating how well a page matches a search phrase. Their increased weighting makes it desirable to include keywords and phrases in headings.



Plan Ahead

The design choices made early in development can significantly change the search engine friendliness of your site. It is important to work with your Web developer to make sure that search engine optimization is taken into consideration during the initial stages of the design.

From that perspective, good Web site design begins with understanding how you expect your prospective visitors to search for your products and services. The keywords and phrases that you expect they will use (or know that they are using based on data from your current site) should guide the creation of the text copy for your new site. Repeat the keywords and phrases often, but not in such a way that the page stops making sense. Pages with too many keywords can be penalized by the search engines, whose algorithms perceive the page to be an attempt at cheating the system.



On a related note, Google has published a new page on their site that provides guidelines for identifying unethical Search Engine Optimization (SEO) firms. It’s a must-read, and can be found at:

<http://www.google.com/webmasters/seo.html>



We hope that you continue to find insight and value in the WC Journal, and we welcome your comments, which can be addressed to Will Weidman at will@weidmanconsulting.com.

You can subscribe to the WC Journal at <http://www.weidmanconsulting.com/subscribe.html>.