Findability and Search Engine Optimisation
Satyadarshin, version 2.2: January 2006

Introduction

You probably know that just posting a website doesn't mean it will suddenly appear as a result in a search through Google; you need to be found and indexed. However, unlike printed matter, the quality of site construction has a direct bearing on whether a casual visitor will ever see your site. Unlike a poster or flyer, you cannot catch anyone's eye by producing a beautiful homepage - although this might induce someone to stay at your site when they get there - you can only really attract visitors by the quality of information on it.

Improving the quality of individual HTML documents would make a huge difference to many sites. HTML that is 'well formed' and 'standards compliant' produces a number of benefits: altruistically there are benefits for people with physical disabilities; for yourself these enhancements are good for you in terms in terms of 'findability', the likelihood that your site will be ranked well for key search terms.

Much of the advice below is based on the willingness to 'tinker under the hood'. However, for many of us, editing HTML directly is a prohibitively daunting task. Many site designers know little if any HTML and are bound by the results generated by their webpage editor (the program in which they compose their website, e.g. Macromedia Dreamweaver).

That said, there are a number of things we can do to help improve our returns.

The following article assumes a modest understanding of Internet computing. If you find much of this piece 'over your head', don't worry too much: do what you can. See the article on writing *Web Friendly Text* and apply that; many of its recommendations are a distillation of some of these points.

How are we found at all?

If your site is found on the Internet it was reached by one of two routes: direct or casual.
* A direct visitor goes to a particular website by typing the site's URL into their browser's address bar*. They know exactly where they are going, because they read the URL in printed literature or received it by recommendation from someone else. They may already have your site 'bookmarked'.
* A casual visitor find a site by some form of keyword search or following a chain of hyperlinks.

There are two sources of keyword information on the Internet: *directories* and *search engines*.

A *directory* is compiled by humans. It is a set of hand picked lists, sorted by category. The main advantage of a directory is its relevance; there is an improved chance of getting a quality result for your search term. The disadvantage is that you ultimately have to get yourself added to a list. Entering a directory is usually done by submission, that is, you must let the directory editors know that you are there.

A *search engine* is a sophisticated piece of content analysis software that sends mini programs ('webcrawlers') out into the Web logging phrases and terms on the sites they visit. You will therefore get millions of returns for some search criteria, but of various - often even dubious - quality: it's very much the law of the Jungle.

The Internet is a very new, and very developing, medium. People's usage of the Internet may not be very sophisticated. A *casual* search of the web can be more or less directed (using multiple search engines, refining keywords), but it seems that a large percentage of users use the Google search engine more or less by default (and possibly without really understanding what it actually is*). Even if this is actually true, for the rest of the article I will refrain from talking about optimising for Google specifically. Instead I will talk more about crawler-based search engines and human-powered directories.

* In many cases web users I know enter the domain name of the site they want to visit into a Google search window, clicking on the URL when it appears at the top of a returns page. This is a bit like knowing a telephone number but always phoning directory enquiries to put you through. However, users occasionally 'can't find' a site that has been deliberately hidden from webcrawlers. The root problem is not understanding what a search engine actually does. Some users I know don't really understand the difference between a browser and a search engine at all, assuming that 'Google' and 'browser' are synonymous terms.

Why you should take directories seriously

Many search engines base their results *partly* on searches through directories. It's one very good way of getting indexed by a search engine - they know where to find you! But of course actually being categorised under your primary key words will give you a major boost in a search engine returns too.

Submitting to a directory
Do some preparation. Go and look at the directory. Figure out what category your site would naturally seem to fall under. Then write about 25 words - in natural language - on what your site is about. Follow the site's submission guidelines.

There are two important directories.

The Open Directory Project (also known as DMOZ)is a free service and has a clear submission page:

<www.dmoz.org/add.html>

Yahoo! Has a free, non-commercial submission section. You are required to login as a Yahoo! member. This may seem slightly tedious, but is worth the effort and your login details allow you to join other Yahoo! services like the FWBO Webmaster's Forum. Follow links to the 'Yahoo! Standard' service (commercial directory listing costs!)
<docs.yahoo.com/info/suggest/>

Bear in mind that you have no automatic 'right' to be included in a directory and no right to an explanation about not being included.

Submitting to crawler based search engines
Letting a crawler-based search engine know that you are there is no guarantee that you will get any particular result, but it is a proactive step towards getting yourself into the running.

Resubmit your site whenever you make major revisions.

<www.google.com/addurl/?continue=/addurl>

<search.yahoo.com/info/submit.html>

Your website: the public face
Read this section in conjunction with the article on writing *Web Friendly Text*. None of the advice you find here (or elsewhere) will be of any use if your site content - buried at the bottom of the HTML document, almost invisible under excess HTML tags - is itself poorly considered.

Keep your content fresh and up-to-date. This makes a difference to search engine evaluations; fresh text is attractive and will keep your current ranking lively.

You may want to consider having a strong internal page that relates very closely to an important search term, e.g. "meditation Yourtown".

Links should be well considered and natural. Make sure that there are plenty of cross-links to the pages on your own site. Don't just mention a particular retreat centre; add a hyperlink to it. Don't just make a passing reference to a book published by Windhorse, add a link to their site. For example:

Windhorse Publications celebrated the first anniversary of moving into their new premises behind the Birmingham Buddhist Centre by holding a dedication day in the middle of November, to which Sangharakshita was invited.

This will benefit both you and all of the sites you link to. Apply some discretion: there's no need to turn *every* reference to Sangharakshita into a hyperlink to his own site's homepage. But do look for opportunities to make 'deep links', which is where you link on to a page which naturally extends the visitors knowledge on that topic. A wholesale links page to anyone and everyone in the FWBO is a very good thing, but that's simply painting with a broad brush. Links aren't about shear quantity, they just get your site found. 'Good' links flag relevancy.

Many large sites would benefit from a site-map - essentially a contents page - made up of text links. A small site can often afford to have a simple index of its entire contents on every page. For bigger sites this would be unfeasible, so a link on every page to the site map ensures that every page will get visited and indexed by your webcrawler. Obviously you should never create dead end pages (no hyperlinks out; even worse, a page that also forces a new window to open so even the 'back' button won't work).

Your website: behind the scenes
Keep your text content high up in your document structure. The upper layers of a document are where many webcrawlers look for the principle information about a site. A document that is top heavy with redundant Javascript, HTML table tags, styling attributes (many of which are themselves redundant) is poorly optimised.

(N.b. stripping out excess code reduces file sizes, making them easier for the visitor to download and, if your Web host imposes limits on server and bandwidth usage, getting rid of that flab will let you have more for you money).

It follows that your most important keywords should appear in the first paragraph of text and that under no circumstances should you bury keywords in images, image maps or anything that will render them invisible to a text reader.

The more you can separate styling instructions from the HTML tags the easier it will be for search engine evaluation software to index your site. This is particularly important for headings, which can easily get lost in burdensome HTML:

 Buddhism, Yourtown <font face="Arial, Helvetica, sans-serif">

The example above would have much more impact if it was simply

<h1>Buddhism, Yourtown</h1>

This is because the HTML heading tag itself flags the relative importance of the information enclosed within it (because it is a *semantic* tag). In the first example above, if the phrase is only mentioned once on your page, the webcrawler has no reason to think it has any particular importance. In the second, it knows that 'Buddhism, Yourtown' is important: the tag itself says so.

Tabular layouts can break up the meaning and relevancy of terms and phrases; remember they will be read by a webcrawler as a list of tagged entries, in the order that they appear in the HTML document. The webcrawler never 'sees' the page in a browser and could not 'read' it, even if it did.

Even if you don't understand HTML, read through your document and see if you can still make sense of the text content in the order that you read it. If you can't, try to re-organise the structure of the document.

Don't allow *any* links to be activated by Javascript. Webcrawlers can't activate the script, so can't follow the links. Links buried in image maps will also not be followed and for the same reasons, with the targeted pages not indexed (this is also a nuisance for human visitors who can't run the script).

HTML meta tags
Search engines put different emphasis on the contents of the various meta tags in the head section of a HTML document. Most search engines apparently place little value on the keywords attributed meta tag, but since you've gone to all that trouble of writing web friendly text and listing your keywords, why not add them just in case - search engines do change and develop their algorithms after all.

If you look, you will notice that in a search return there are usually three levels of information per return (a fourth appears in returns in queries to a directory: the directory path)

The HTML page title
A brief content description. which amounts to about thirty words
The page URL

That brief 'content description' may well be the few words you submitted with your entry to a directory, if that's where the return was extracted from (remember that many crawler based search engines base their

returns on a directory). If not, it may well be the content of your description meta tag. Failing that - if neither is available - it may well be the first 30 words of text content that appear in your HTML.

On this latter point it emphasises that the structure of your document can really make a difference. Have a look at a selection of your HTML documents. Amongst all the code, what are the first thirty words of text that will be displayed in the browser? Would you be happy for those to summarise the content of your site? If not, can you re-write the text to be punchier? Can you re-organise the HTML?

In any case, this illustrates three things:

* that your HTML page title *really* needs to count,
* that you should submit your site to directories,
* that you should make good use of the description meta tag and that the first 30 words that appear in your HTML *document* (which may not be the same as they are displayed in a browser) should also be *very* well considered.

title attributes
An 'attribute' is a syntactical statement placed insie an opening HTML tag. There are a number of valid attributes, but the most important one in terms of findability is the title attribute: title="Insert your text here". It can be inserted into quite a lot of HTML tags, and is very useful in *inline* tags, i.e. tags that enclose a term *inside* a paragraph. For example:

        <abbr title="Friends of the Western Buddhist Order">FWBO</abbr>

        <cite title="Birmingham: Windhorse Publications, 2005">The Yogi's Joy</cite>

        <em title="Pali: Loving-kindness">metta

In a Web page, when the pointer hovers over the tagged text, the phrase between the double quotation marks will appear in a little box offset from the inline text, rather like a 'tooltip'. The title attribute can be used to expand on the keywords emphasised by being a hyperlink. So:

    [Ö] signing copies of his new Dharma book about Milarepa, <a href="
    http://windhorsepublications.com/CartV2/Details.asp?ProductID=708" title="More books on
    Buddhism and meditation"><cite>The Yogi's Joy</cite>.

This little addition adds up to a *string* of benefits.

The anchor tag (the tag that creates a hyperlink) flags the relevance of the Windhorse Publications site by its context (in proximity to phrases like 'Dharma book' for example). Taking the visitor directly to a payoff (i.e. directly to the page where they can view and buy the book) is not only a courteous service, but any keywords in the target page will be emphasised because you 'deep linked' to them from a keyword rich source.

The link text - the words that you actually click on in the resulting Web page - strongly emphasise relevancy, and the title attribute elaborates and emphasises that.

As far as the FWBO is concerned, not only do we need to add constructive content about the movement to all our respective sites, we can double the potential value of the term 'FWBO' every time we use it by attribution, because 'FWBO' (and 'WBO') is an abbreviation.

```
<abbr title="Friends of the Western Buddhist Order">FWBO</abbr>
```

In a standards compliant web browser, the attributed <abbr> tag will add a dotted line under the enclosed term. This flags to the visitor that they can see a definition (a 'tooltip') by placing their pointer over the abbreviation. However, in finability terms, this means that you've managed to work both 'FWBO' *and* 'Friends of the Western Buddhist Order' into the document in a genuinely meaningful way.

It is true that Internet Explorer doesn't currently support this tag (although it does support <acronym>, but stick to <abbr> because future editions of IE may well do so.

Keep it fresh

New pages appearing at a regular rate make webcrawlers feel that the site is alive and flourishing. Some sites of archive material do stay high in the rankings for quite some time, but they slip eventually. It's therefore important to add fresh content periodically.

If your pages haven't changed that much, or you haven't added any recently, think about adding some content. Remember, however, never to simply duplicate text from other sites: this kind of plagiarism may well be interpreted as a 'black hat' attempt to influence rankings and you may well find your site being penalised.

Web Standards and accessibility

Web standards are proposed by a body called the World Wide Web Consortium (W3C). A Web standard is a recommendation from them about the 'legal' functioning of a software component of the Internet, including computer languages like HTML, XHTML, XML and CSS. Applying a web standard as rigorously as you are able has three principle benefits:

* It ensures your website will work - in near enough the same way - in the maximum number of web browsers for years to come.
* It makes the site much more accessible to people with disabilities
* It makes our site much more findable.

The most important thing you can do to help yourself is become user centred. Many practices that are aimed at improving accessibility for disabled users have an equal benefit in terms of search engine optimisation. Consider your site as a valuable service to your visitor. If you do that, everything you do will improve your site's search optimisation.

Satyadarshin