

## Search Engines Series Pt. 2.b - Links, Content and Format

Posted At : December 19, 2006 7:19 PM | Posted By : Mark Kruger

Related Categories: Hosting and Networking

In this post, part 2 "b" in our search engine series, we will discuss how the content and structure of your page might influence how your site is viewed by search engines. In **part 1** we talked about having useful and valuable content. That lesson is the foundation on which all other legitimate techniques must be based. If your content is not useful you are part of the problem we are trying to solve. In **Part 2 "a"** we talked about stuff that goes into the header. Now it's time to talk about things that go into the actual page.

### Why Style is Important

One of the things you *may not* think about is the sheer number of characters that are a part of the page. Some research indicates that search engines simply *stop indexing* your page after a few kilobytes. I suppose that's one reason why the title is so important. It also means that one of the other things you should think about is how you style your page. If you are accustomed to using lots of nested tables and in-line styles you might consider using CSS instead.

There are a host of resources out there for CSS. You can almost always accomplish the same thing with CSS as you can with tables. It just requires a different, more abstract mindset. For me, tables always provided a visual "grid" in my head. This grid was a conceptual tool that my pea brain could take hold of and use to layout content. With CSS it's much more left-brained (not my forte). Still, using CSS can make your page more semantic and more readable. Consider this example:

```
<table style="font-size: 10pt; font-family: Arial, Tahoma; border: 1px solid #0000;
background-color: #FFFFFF; margin: 0px; padding: 0px;">
  <tr>
    <td style="background-color: #CCCCCC; border: 1px solid #CCCCCC;">Alaska is the
largest State.</td>
  </tr>
  <tr>
    <td style="border: 1px solid #FFFFFF;">Nebraska is the coolest state.</td>
  </tr>
  <tr>
    <td style="background-color: #CCCCCC; border: 1px solid #CCCCCC;">Rhode Island
is the smallest state.</td>
  </tr>
  <tr>
    <td style="border: 1px solid #FFFFFF;">Wisconsin is the cheesiest state.</td>
  </tr>
</table>
```

What is the relevant content in this little snippet? It's 4 facts about states - right? It identifies the largest, smallest, coolest and cheesiest. So out of the 500 or so characters above only a little more than 100 are significant characters - parts of words we want indexed. A search engine will need "tease out" this text for indexing. If you had say... 1500 facts about states some of your facts might be left out.

Furthermore, the HTML code is not "semantic". It is pure layout. HTML you might recall is intended to be *markup*. The tags are supposed to say more than just "put this here

and make it yea big". They are supposed to say something *about the content*. An "H1" tag is supposed to indicate content that is *more important* than "H2", not just bigger. List items are supposed to contain lists. Tags like "Strong" (for boldface) and "Em" (for italic) are supposed to mean "emphasis this" or "give this more weight". In the case of the example above the items in question are really list items. With a combination of CSS classes you could get exactly this effect, indicate it is a list and reduce the clutter in the HTML.

```
<ul id="state-facts">
  <li class="list-oddrow">Alaska is the largest State.</li>
  <li>Nebraska is the coolest state.</li>
  <li class="list-oddrow">Rhode Island is the smallest state.</li>
  <li>Wisconsin is the cheesiest state.</li>
</ul>
```

What is the result? We went from 500 characters to 200. We are using a list which matches the type of content (a list of facts about states) and we have added an id called "state-facts" that is also a semantic indicator.

Now before I get letters about how the semantic web is a perpetual lady-in-waiting, I'm well aware that the concept doesn't quite live up to it's promise. Practically speaking, 90% of the HTML found on the web is useless from this point of view. That doesn't mean we shouldn't adopt it. It makes your HTML code more readable and maintainable. Such content can be more easily consumed by text readers, PDAs, RSS generators and the like. CSS will give you a leaner *page weight* (total size of the page) and make it load faster and cleaner.

## JavaScript and Content

Here's a tip - don't use JavaScript to output content that you want indexed. Does that sound like a joke? One of the most visually appealing and functional sports related sites on the web is the [Major League Baseball](#) site. Three or Four years ago this site actually used in-line JavaScript that "wrote out" the content to the browser. It was a huge mess of "document.write( )" commands. Why did they do it? Who knows. Maybe they hired developers from a rival sport. Maybe they were trying to "protect" their source code by obscuring it (a futile endeavor). Maybe they were just all on drugs (drugs and baseball - what are the odds). In any case it was a foolish and asinine thing to do.

As you probably already know, search engines don't index JavaScript. If you put your content into JavaScript you can bet that it won't be indexed, including those cool little DHTML widgets that switch out the "innerHTML" values of a div tag. It also includes links that use JavaScript. Links are an important part of your strategy on the web. Your site should link to other sites and other sites should link to your site. Rating and tracking the inbound and outbound links has a lot to do with your popularity on the web. If you choose to put your links inside of a JavaScript function they will not be followed.

Let's take an ecommerce site as an example. Your product page lists your products, titles and prices. If a user clicks on the "more info" link you "pop up" a JavaScript controlled window that contains the product description. From a user perspective this may make for an ideal user experience. You can keep their eyeballs on your page of products enticing them to buy, while still providing useful and helpful information on the product in question. Unfortunately, the content that the search engine may be

most "interested in" (the useful and valuable content) is in the product description - but the search engine won't see it unless you provide an alternate link for it to index. So on a page with such links include a "noscript" block that includes all those missing links so they can be followed and indexed.

## Flash, Ajax and SEO

Do search engines index flash content? You can find hundreds of blogs and articles on the web that bolster one side or the other. You can also find many approaches to creating flash content that is easier to consume using by SDKs and tools. In fact, in your searching for all these blogs and articles there is one thing you will not find. You won't stumble onto any content that *comes from a Flash movie*. In fact, in my experience as someone who spends a good part of each week scouring the web for content on various topics of interest, I have never yet stumbled onto anything I was searching for that was obviously pulled from a flash movie. I'm not saying it has never happened - I'm just saying that the vast vast majority of indexed content on the web is text from HTML or other markup - not Flash.

As an experiment I did a search for *shark filetype:swf*. My searched returned links to no less than 14,200 flash objects including this one from the [shark research institute](#). If you go to the [home](#) page of the shark research institute you will indeed see this movie embedded as the home page. Does this mean that Google indexed the content of the swf file? Yes indeed. The swf file has been indexed and as long as I search by filetype it shows up. But if I just search for "shark" without the filetype I get 36 million results - and nothing on the first 20 pages is from a flash file.

More to the point, flash precludes the possibility of semantic markup. It's an "object" embedded in the page. By definition it is not "content" to the search engine any more than your fancy javascript is content. Links within the flash move are not indexable either - so a movie that loads other movies will fail to be "followed".

Which brings us to Ajax. The idea behind Ajax, like flex and flash remoting, is to use http requests to retrieve data in an XML format and make "on-the-fly" changes to a page without reloading the whole browser. A Search engine, landing on your page, will do what exactly? Will it be able to retrieve the XML that contains the data you are loading? Those links are usually embedded in JavaScript and will not be followed. In theory, the XML that is used to feed Ajax would be ideal for indexing. In practice however, I suspect that most search engines will not follow the links in question.

Let me say one final note on Flash, Flex and Ajax. There are certainly cases where content embedded in these technologies *should* be indexed. In particular, I have seen content embedded in flash movies that rightly belongs in HTML. But the purpose of these technologies is not always to create pretty web sites that are candidates for indexing. Often, Ajax and Flex are used to create interactivity - applications and user interfaces. Such things are not always appropriate for indexing anyway - so the use of these technologies need not be seen as a negative.

## How to Write Appropriate Content for Indexing

There's no magic to effective writing - but some things that work in the human world will become obstacles in the world of search engines. Here are a couple of tips on writing valuable content.

- **Make Judicious Use of Marketing** - I'm not saying to have *no* marketing copy on

your site (although if I could I would banish all marketing copy to the isle of misfit prose). I'm only saying that the words you use to embellish your product or services are not the words that folks will use when looking for them. No one does a search for a "best of breed integrator". Plus, you've just created a situation where you can show up in searches for "breeding" which may not be exactly what you were looking for (or maybe it is what your looking for ... who am I to judge). Instead, use concrete words in your descriptions. If your service is putting in new floors feel free to use words like sparkly and shiny, but don't forget to use the words "floor, flooring, wood floor, linoleum, tile floor, no-wax floor" ...and any other floor that floor people might be looking for. So saying something like "The shiniest, brightest surface upon which you've ever trod!" would have far less impact on your search engine traffic than saying something like "durable floor".

- **Make Your Information Valuable** - If you know something, share it. Are you a financial advisor? Make your site replete with financial advice. Are you a cooking site? Add recipes and product reviews. Selling craft supplies? Add how-to articles and step-by-step instructions. These items are things *people are scouring the web to find*. You have knowledge that they need - share it. Some people think that this is letting the cat out of the bag - that hanging on to proprietary knowledge makes your more valuable. These people have never heard of Wikipedia. Let me clue you in on 3 tidbits I've learned in 20 years of adult life. 1) You don't know half of what you think you know. B) Someone out there is smarter than you and is already publishing and 3) being generous with "knowledge sharing" will enhance your life both personally and professionally. Come on, lighten up! As Red Green is fond of saying - "I'm pulling for you... we are all in this together."
- **Blogging** - I'll cover "alternative content" a bit more in my next post, but for now let me say a word about blogging. Blogging will add a dimension to your site that search engines love - namely, fresh content regularly updated and relevant. If you have the capacity to write, start a blog and keep it up to date. You will be surprised at how it becomes a conduit to your services. At [CF Webtools](#) about 80% of our business comes to the door through this very blog.

## Conclusions

Hopefully you've gained a few tips from this post about how to style better and write better content. In our next post we will talk about creating strategies for better page ranking.