



Search Engine Optimization and Non-HTML Sites

“If you build it, they will come” might be the line from *Field of Dreams* but the same theory doesn’t apply to Web sites. What’s the good of building a great-looking Web site that no one knows about, or can find? Unfortunately, *that* is the reality of building Web sites entirely in Flash.

It’s not just Web sites that are built entirely in Flash that have this problem. Many Web sites contain an enormous amount of content that can only be seen with the human eye. Think about all the PDF, sound and other rich media files out there. Can this content be indexed by search engines? The answer is yes and no.

Why should you care about search engines? Since the launch of Yahoo (merely a directory of sites) many years ago, and the very first search engines (WebCrawler, Infoseek, Altavista and many more), people have used them to hunt down specific Web sites by searching for relevant content. If your content isn’t optimized for search engines, then it is almost impossible for people to find your site. And without this traffic, all your work in building that site is wasted.

Flash

So what can you do? Don’t build a site entirely in Flash.

The Problem

While Flash may be a favorite design tool for graphic designers—because of its pinpoint accuracy of design, the ease of animation and its cross-browser and cross-platform compatibility (debatable)—it is merely treated as a graphic by search engines. Search engines read text and ignore graphics. The other problem is not just that the content of a Flash page will be ignored, but that links contained in the Flash animation will also be ignored, and the indexing of the site will stop at the first page.

The mistake of building entire sites in Flash is not just an amateur’s mistake—many leading Web designers, who are paid copious amounts of money, do the same thing. Sometimes the use of Flash is the only way to achieve a specific function (e.g. Web-based games), so you need Flash for that feature—but do you need it for the whole site?



407.681.2263

www.tolsmultimedia.com

According to Gregory Markel of Infuse Creative, one needs only to look at the movie studio that released the “I, Robot” Web site (in preparation for the movie launch in the summer of 2004) to see the problem with using only Flash. Markel, a Flash enthusiast, pointed out that this site, built entirely in Flash, could not be found in any of the leading search engines. While the movie studio may have deep pockets to market the site—and the movie should generate enough loyal fans who’ll link to the site, eventually, thus getting it listed appropriately in the search engines—don’t you think it would have been better if those wanting to find the movie Web site could simply type in the name of the movie in their favorite search engine?

We alluded to one of the benefits of Flash’s cross-browser compatibility. This is only true if the user has the Flash plug-in. While Macromedia and others estimate the penetration rate for Flash may be as high as 98%, it still doesn’t guarantee that everyone has it. If the user is blind or visually impaired, and using a brail reader or Web page voice reader, they see what the search engines see, something that simply says “Flash”.

If you want to see what these people (and the search engines) see, try accessing your site using the old-but-reliable Lynx browser. This original browser dates back to a time when graphics and Flash were not yet part of the Web. Look at your site, do you see content? Can you navigate your site? Ah, the problem of Flash-only sites.

The Solution

First, for the benefit of the search engines, if you’re going to build your entire site in Flash, be sure to add effective title and description meta tags.

Second, when it comes to links, make sure there are some standard links (<a href>) on the page. This way, the search engines will be able to find more than one page on the site.

Finally, where possible, remove text content that doesn’t need to be in Flash and put it into good old HTML. If you want the pinpoint accuracy of Flash, consider using XHTML and CSS. This method works on all modern browsers (IE 5+, Netscape 6+, Mozilla, Opera etc.)

PDF Content

Just about everyone should be familiar with Adobe’s PDF (Portable Document Format). This is a great way to ensure your content always looks the same and prints properly. The problem with HTML is that it was invented for viewing with a browser, not for printing. PDF files can be viewed in most browsers (plug-in required) and the content prints perfectly. PDF also provides an excellent way to publish large documents (e.g. white papers) originally prepared using a word processor.



407.681.2263

www.tolsmultimedia.com

Other benefits of PDF include font preservation and the fact that it can contain graphics at a much higher resolution than the standard JPG and GIF used on the Web. This is vital for technical specifications and other technical material.

The Problem

While most leading search engines can now read and index the content of a PDF, they have certain restrictions and may only index the first N hundred or thousand characters. Further, the file size of a PDF document frequently exceeds 100K and may take a long time to download.

The Solution

First, make sure your PDF contains text. There is no point in worrying about indexing the content if there are no words.

Second, just like optimizing any Web page, make sure your PDF contains your keywords and phrases. Use the keywords prominently (table of contents, page titles, etc.). If you think the words are important, so will the search engines.

Third, if you have a very large PDF, consider breaking it up into several documents. This will ensure that the maximum amount of content gets included in the search engine's index. You can also try creating an HTML-formatted abstract of the PDF and linking it to the PDF.

Another way to reduce the size of your PDF is to limit the number of different typefaces. Beyond making the file smaller, this also makes good design sense.

Finally, PDF can be a great lead generator. If you have a very large PDF, simply offer the first section as a free download and then require users to register (create a lead) for the rest of the document. By following this last step, you've now turned a potential problem into an opportunity for a measurable Web transaction.

Audio and Rich Media Files

Everyone loves all the MP3s out there on the Web. What about Real Audio or Real Video files and other rich media files? Do you have them on your Web site? Is it possible someone might be searching for them? If you thought you couldn't search for them as content through search engines, think again.

You can use Singfish.com to locate various rich media files. If you haven't heard of Singfish don't worry about it, they'll soon be serving you search results from various search engines. Other search engines, like Alltheweb.com, have also been indexing rich media files for the past several years.



407.681.2263

www.tolsmultimedia.com

The Problem

The only real problem with rich media files like Flash is that there is no simple text for search engines to index.

The Solution

When creating a rich media file, be sure to complete all metadata with effective and well-structured content. According to Karen How (General Manager, Singingfish, AOL), missing or poorly-structured metadata is the biggest reason content gets ignored by Singingfish.

Common metadata fields, which should always be used, are:

- Title
- Author
- Copyright
- Description

So what does bad metadata vs. good metadata look like?

Metadata	Bad	Good
Title	my.mov	Widgets and their use in HTML editing
Copyright	Mine	K'nechtology Inc. © 2004
Description	Quicktime file on widgets	A visual guide to using widgets to simplify HTML editing

Beyond completing the metadata information, be sure that the HTML page that links to it also:

- has accurate anchor text (on the link to the file)
- can be spidered by the search engines
- has descriptive page titles

When linking your rich media file to your Web page, use:

```
<embed ... title="A visual guide ..."></embed>
```

or:

```
<object ... title="A visual guide ...">A visual guide  
...</object>
```

Finally, name your files appropriately. Search engines *do* value file names so give them names like “widgets-html-editing.mov” and try to stay away from names like “wig1234.mov” which hold no meaning.

Since search engines—for now and the foreseeable future—will continue to prefer plain text for indexing, you should follow these guidelines to ensure that your content is spidered by the search engines, and, as a result, will enjoy high rankings in search results. By anticipating the problems and implementing these solutions, you can make your Web site more search engine-friendly and start attracting an audience you might otherwise have never found.



407.681.2263

www.tolsmultimedia.com