# Domain Changes, Redesigns, and Troubleshooting

Whenever you make structural changes to your website, there is a risk that you will confuse the search engines and harm your search rankings and organic traffic. The types of site changes that can influence your site's organic search performance include changing your domain name, changing your content management system (CMS), redesigning your site, adding ecommerce functionality, changing your blog platform, and many others—basically, anything that fundamentally alters your site's frontend and/or backend visual or functional elements can potentially influence your organic search performance.

In this chapter, we will review the various scenarios from the top down. Be sure to refer back to "Content Delivery and Search Spider Control" on page 334 to review the technical specifics of options for moving content from one location to another.

Google groups site moves into one of two categories (*http://bit.ly/site_moves*): moves "with" URL changes, and moves "without" URL changes. It provides specific guidelines for handling moves within each category.

## The Basics of Moving Content

"Moving content" refers to a situation in which content that used to be located and accessed at one URL (e.g., *http://www.yourdomain.com/pageA*) is moved to another URL (e.g., *http://www.yourdomain.com/products/pageA*). One of your goals when you move content is to make sure users and search engine crawlers that attempt to visit the old URL (*/pageA*) are presented with the content from the new location (*/products/pageA*). In addition, when you move content from one URL to another, the links to the old URL will stop providing value to your rankings in the search engines for that content unless you properly implement a 301 (permanently moved) redirect. There are very

**795**

important reasons to move content properly—both of which can be easily overlooked by inexperienced or hurried webmasters and development teams.

In "Duplicate Content Issues" on page 320, we covered the technical specifics of how to do this in detail, including the golden rule of moving content: the search engine needs to see a 301 HTTP status code whenever you redirect the content's URL to a new location.

The 301 HTTP status code causes the search engine to pass most of the value of any links for the original page over to the new page, and should result in the rapid deindexation of the old URL. Because link authority is a precious asset, you want to make sure you use a 301 redirect every time.

## Large-Scale Content Moves

Setting up the redirects can become difficult when changes result in movement of large quantities of content. For example, when you change your domain name, every single piece of content on your site will move to a new URL, even if the site architecture is identical (*http://www.olddomain.com/*... moves to *http://www.newdomain.com/*...).

This is challenging because you might have to set up individual 301 redirects for every single page on the site, as in this example:

- *http://www.olddomain.com/page1* to *http://www.newdomain.com/page1*

- *http://www.olddomain.com/page2* to *http://www.newdomain.com/page2*

- *http://www.olddomain.com/page3* to *http://www.newdomain.com/page3*

- ...

- ...

- *http://www.olddomain.com/page1000* to *http://www.newdomain.com/page1000*

Unfortunately, some systems still require that these redirects be set up one at a time, so this could be quite a painful process. Imagine a site with 1 million pages!

Fortunately, publishers who use an Apache web server (Unix and Linux servers) can take advantage of the power of Apache's mod_rewrite module, which can perform the redirect of every URL on the old domain to the same URL on the new domain in two lines of code:

```
RewriteCond %{HTTP_HOST} ^olddomain\.com [NC]
RewriteRule ^/(.*) http://www.newdomain.com/$1 [R=301,L]
```

The preceding code presumes that you prefer the "www" version as the canonical URL. You can also use two similar lines of code to specify the "non-www" version as the canonical URL (see "URL Rewrites & Redirects: The Gory Details (Part 1 of 2)" for examples without "www" and other alternative approaches).

Another highly popular web server is Microsoft's IIS. In many installations of IIS, you will find yourself in a situation where you have to implement a separate redirect instruction for each page, one at a time. Fortunately, you can utilize an ISAPI plug-in such as ISAPI_Rewrite, which enables you to perform large, scalable rewrites in a language similar to that used by Apache's mod_rewrite. You can learn more about mod_rewrite, ISAPI_Rewrite, and regular expressions in Chapter 6.

## Mapping Content Moves

Sometimes a site redesign is simply a "reskinning" of the visual elements of the old site with a new look and feel, retaining the same technical elements of information architecture, URL file and directory names, and user navigation. Other times, a redesign changes both the visual design and the technical elements. And sometimes it's a combination of the two approaches. For sites changing both design and function, the first stage of planning is to figure out which content will be moved where and which content will be removed altogether. You will need this information to tell you which URLs you will need to redirect and to which new locations.

The best way to start this process is by getting a complete map of your information architecture with full URLs. For many websites this is not as simple as it sounds. Fortunately, tools are available to make the job easier. Here are some ways to tackle this problem:

- Extract a list of URLs from your web server's logfiles and site architecture documentation.
- Pull the list from your XML sitemap file, provided you believe it is reasonably complete.
- Use a free crawling tool, such as Screaming Frog SEO Spider Tool.
- Use tools such as Moz.com's Open Site Explorer and Google Search Console to pull a list of the external links to your site, and make sure all pages that have received links on your site are included.
- Check Bing Webmaster Tools' Index Explorer to find all of the crawlable URLs that you may not know still exist on the site.

These tools should help you assemble a decent list of all your URLs. After determining which URLs have content that will remain on the site, you must then map out the pages that you want to redirect the "migrating" content to. Additionally, for content that is being "retired," you need to determine whether to redirect them at all (a definite yes if the URLs for these pages have many internal and external links), and if so, what new URLs to redirect them to.

One way to do this is to lay it out in a spreadsheet, which might end up looking like Table 12-1.

**Table 12-1.** *Planning your content moves in advance*

| Old URL | New URL |
| --- | --- |
| http://www.olddomain.com/page1 | http://www.newdomain.com/page1 |
| http://www.olddomain.com/page2 | http://www.newdomain.com/page2 |
| http://www.olddomain.com/page3 | http://www.newdomain.com/page3 |
| http://www.olddomain.com/page4 | http://www.newdomain.com/page4 |
| http://www.olddomain.com/page5 | http://www.newdomain.com/page5 |
| http://www.olddomain.com/page6 | http://www.newdomain.com/page6 |
| http://www.olddomain.com/page7 | http://www.newdomain.com/page7 |
| http://www.olddomain.com/page8 | http://www.newdomain.com/page8 |
| http://www.olddomain.com/page9 | http://www.newdomain.com/page9 |
| http://www.olddomain.com/page10 | http://www.newdomain.com/page10 |

If you are redirecting a massive number of URLs, you should look for ways to simplify this process, such as writing rules that communicate what you need to know. For example, you could abbreviate the list in Table 12-1 to the short list in Table 12-2.

**Table 12-2.** *Simplifying content move planning with wildcards*

| Old URLs | New URLs |
| --- | --- |
| http://www.olddomain.com/page* | http://www.newdomain.com/page* |

Then you can save the individual lines for the more complicated moves, so your resulting spreadsheet would look like Table 12-3.

**Table 12-3.** *Mapping all your content moves completely*

**Individual page moves**

| Old URL | New URL |
| --- | --- |
| http://www.olddomain.com/about-us | http://www.newdomain.com/about-us |
| http://www.olddomain.com/contact-us | http://www.newdomain.com/contact-us |
| http://www.olddomain.com/press-relations | http://www.newdomain.com/press |

**Large-scale page moves**

| Old URL | New URL |
| --- | --- |
| http://www.olddomain.com/content/* | http://www.newdomain.com/content/* |
| http://www.olddomain.com/page* | http://www.newdomain.com/page* |

The purpose of this is to efficiently give your developers a map for how the content movement should take place. Note that the spreadsheet should contain a map of all changed URLs, which may include downloadable content such as PDF files, Power-Point presentations, Flash files, multimedia, or any other types of content being moved.

You can also note retiring content via additional entries in the left column, with the entries in the right column indicating where users looking for that old content should be sent. Now your spreadsheet might look like Table 12-4.

**Table 12-4.** *Identifying pages that have been removed*

**Individual page moves**

| Old URL | New URL |
| --- | --- |
| http://www.olddomain.com/about-us | http://www.newdomain.com/about-us |
| http://www.olddomain.com/contact-us | http://www.newdomain.com/contact-us |
| http://www.olddomain.com/press-relations | http://www.newdomain.com/press |

**Large-scale page moves**

| Old URL | New URL |
| --- | --- |
| http://www.olddomain.com/content/* | http://www.newdomain.com/content/* |
| http://www.olddomain.com/page* | http://www.newdomain.com/page* |

**Eliminated pages**

| Old URL | Redirect to |
| --- | --- |
| http://www.olddomain.com/widgets/azure | http://www.newdomain.com/widgets/blue |
| http://www.olddomain.com/widgets/teal | http://www.newdomain.com/widgets/green |
| http://www.olddomain.com/widgets/puce | http://www.newdomain.com/widgets/ |

The new entries show what should happen to retired pages. The first two retired pages may represent products that you no longer carry, so you would likely want to redirect them to the closest existing products you have. The third retired page represents a URL where there is no sufficient replacement, so you may choose to redirect that one to the parent page for that topic area.

As you can see, a major SEO objective during content migration is to preserve as much link authority and traffic from the old URLs as possible, while providing the best possible user experience for people who arrive at the old URLs.

## Expectations for Content Moves

The big downside to content migration is that the search engines won't necessarily adapt to the URL changes immediately. Many sites temporarily lose rankings after

making a large-scale content move, then recover after a period of time. So naturally, the question is, how long will it take to get your organic rankings and traffic back?

The reality is that a number of factors are involved, depending on your particular situation. Some examples of these factors might include:

*The size and complexity of your site*
> Bigger, more complex sites may take longer to process.

*The complexity of the move*
> If the site has been fundamentally restructured, it is likely to take more time for the search engines to adapt to the new structure.

*The perceived authority of the site*
> Sites that have a higher (search engine) perceived authority may be processed faster. Related to this is the rate at which the site is typically crawled.

*The addition of new links to the new pages*
> Obtaining new links to the new URLs, or changing old links that used to point to the old URLs so that they point to the new URLs, can help speed up the process.

If you are moving to an entirely new domain, you can aid the process in Google by using the Change of Address tool inside Google Search Console. Before using this tool, make sure that both your old domain and your new domain are verified in Search Console. Then, on the Search Console home page, click on the old domain. Under Site Configuration, click "Change of Address," and then select the new site. You can also use the Bing Change of Address tool to let Bing know about your site move.

When all is said and done, a reasonable estimate is that a significant traffic dip from the search engines after a move should rarely last longer than 60 to 90 days, and many sites recover in a shorter timespan.

Another approach to content moves (especially when you're updating and redirecting an entire site's URLs) is to perform the URL migration in a phased manner as opposed to "wiping out" and redirecting all of a site's URLs at once. You'd do this for a few reasons, such as wanting to test the search engines' handling of such a migration on your site before committing to the sitewide change and its resulting impact. Another reason is to mitigate potential organic traffic dips that will occur during the updating period; it is often easier to tolerate 10% traffic loss at a time, versus a 30%–40% traffic loss all at once. This can be especially true for websites that rely upon traffic-based advertising revenue. Google's Matt Cutts describes the benefits of this phased approach in a quote we cite in "Pre-Move Preparations" on page 803.

# Maintaining Search Engine Visibility During and After a Site Redesign

Companies may decide to launch a site redesign as part of a rebranding of their business, a shift in their product lines, a marketing makeover, or for a variety of other business reasons. As discussed, any number of things may change during a site redesign. For example:

- Content may move to new URLs.

- Content might be eliminated.

- Content may be changed.

- Content could be moved behind a login or paywall.

- New site sections may be added.

- New site functionality may be added.

- Navigation/internal linking structure may be changed significantly.

Of course, the move may involve moving everything to a new domain as well, but we will cover that in the next section, "Maintaining Search Engine Visibility During and After Domain Name Changes" on page 802. Here are some best practices for handling a site redesign that involves these technical elements:

- Create 301 redirects for all URLs from the original version of the site pointing to the new URLs on the redesigned site. This should cover scenarios such as any remapping of locations of content and any content that has been eliminated. Use a spreadsheet similar to the ones we outlined at the beginning of this chapter to map out the moves to make sure you cover all of them.

- Review your analytics for the top 100 or so domains sending traffic to the moved and/or eliminated pages and contact as many of these webmasters as possible about changing their links. This can help the search engines understand the new layout of your site more rapidly and provides both better branding and a better user experience.

- Review a backlink report (using your favorite backlink analysis tool) for your site and repeat the process in the preceding bulleted item with the top 200 to 300 or so results returned. Consider using more advanced tools, such as Open Site Explorer or Majestic SEO, which allow you to filter your links to more easily identify the most important ones. Ahrefs' Site Explorer and Cemper's LinkResearchTools are good resources too.

- Make sure you update your sitemap and submit it to Google Search Console and Bing Webmaster Tools. Consider using multiple sitemaps, one for each content

type and/or content area, to submit and monitor the indexing of your new site URLs.

- Monitor your rankings for the content, comparing old to new over time—if the rankings fall, post in the Google Webmaster Central Help Forum detailing what you did, what happened, and any information that might help someone help you. Google employees do monitor these forums and sometimes comment in situations where they think they can help. Don't use the forums to complain; state what has happened and ask for help, as this gives you the best chance of getting feedback.

- Monitor your Search Console account and your analytics for 404 errors and to see how well Google is handling your 301s. When you see 404 errors occurring, make sure you have a properly implemented 301 redirect in place. Don't limit this checking just to 404 errors. Also be on the lookout for HTTP status codes such as 500 and 302. Maintain the XML sitemap of old URLs until search engines discover the 301s.

# Maintaining Search Engine Visibility During and After Domain Name Changes

There may come a time when you have a strong business need—such as a rebranding, renaming, or merger/acquisition—to change your site's domain name. This section will cover some of the considerations and challenges involved in a domain name change.

## Unique Challenges of Domain Name Changes

One of the more challenging aspects of a domain name change is potentially losing the trust the search engines have associated with your old domain. Another issue is that if there were business-specific keywords present in your old domain name that are not in your new domain name, you may see a decline in organic traffic, even if you maintain or recover placement after migration. This decline is a result of "domain bias"— the propensity for searchers to click on domains in search results that include keywords they used in their search query. You may also see a slightly negative impact in organic search terms related to the keywords in your previous domain; although Google is cracking down on exact-match-domain (EMD) websites with low-quality content that were ranking well in search, it still places weight on the words in a domain.

Another unique challenge is the "youth" of the new domain, especially if it was recently purchased and/or has no historical backlink profile. Because of its age, the new domain may be slow to rank. Although the site's relevance and inbound link profile (including the links 301-redirected from the old domain) may suggest a high ranking for some search queries, because the new domain is not yet trusted, the rankings are suppressed and traffic is much lower than it would otherwise be. Domain youth is

another reason why updating valuable third-party links to reflect your new domain is important.

If the prospect of taking a "young domain" hit is too unappealing, another tactic you can try is to make the move to a different domain that has a backlink history associated with it—just make sure that history is a positive one! You don't want to move to an old domain that had any historical spam, manual reviews, or other negative associations, so be sure to perform a thorough backlink audit with your preferred link auditing tools. And while you're at it, see if you can get Google Search Console access to research whether there were any manual spam actions reported against the domain.

## Pre-Move Preparations

Unfortunately, lost traffic is common when you make domain name changes, though the traffic loss is usually temporary. If you do things properly, you can and should recover from any negative impact, and hopefully quickly—but you should be prepared for the potential traffic impact of a domain switch.

If you are planning a domain migration, buy the new domain as early as you can, get some initial content on it, and acquire some links. The purpose of this exercise is to get the domain indexed and recognized by the engines ahead of time.

Then, register the new domain with Google Search Console and Bing Webmaster Tools. This is just another part of making sure Google and Bing know about your new domain as early as possible and in as many ways as possible.

Once you've done this, follow these best practices for handling a domain name change:

- Create 301 redirects for all URLs from the old site pointing to the proper URLs on the new site. Hopefully you will be able to use mod_rewrite or ISAPI_Rewrite to handle the bulk of the work. Use individual rewrite rules to cover any exceptions. Have this in place at launch.

- Review your analytics for the top 100 or so domains sending traffic to the old pages, and contact as many of these webmasters as possible about changing their links.

- Make sure that both the old site and the new site have been verified and have sitemaps submitted at Google Search Console and Bing Webmaster Tools.

- Launch with a media and online marketing blitz—your goals are to get as many new inbound links as possible pointing to the new site as quickly as possible, and to attract a high number of branded searches for the redesigned site.

- Monitor your rankings for the content, comparing old to new over time.

- Monitor your Search Console account for 404 errors and to see how well Google is handling your 301s. When you see some 404s pop up, make sure you have a properly implemented 301 redirect in place. If not, fix it.

- Monitor the search engine spidering activity on the new domain. This can provide a crude measurement of search engine trust. Search engines spend more time crawling sites they trust. When the crawl level at the new site starts to get close to where it was with the old site, you are probably most of the way there.

- Watch your search traffic referrals as well. This should provide you some guidance as to how far along in the process you have come.

- You can also check your server logs for 404 and 500 errors. These will sometimes flag problems that your other checks have not revealed.

As mentioned earlier, Google's Matt Cutts discussed the benefits of a phased approach at a PubCon event:

> So here's the extra step. Don't just move the entire domain from the old domain to the new domain. Start out and then move a subdirectory or a subdomain. Move that first; if you've got a forum, move one part of your forum. Move that over to the new domain, and make sure that the rankings for that one part of your site don't crash. Sometimes it takes a week or so for them to sort of equalize out, because we have to crawl that page to see that it's moved. So if you move a part of your site first, and it goes fine, then you know that you're pretty safe. So instead of doing one huge move, if you can break it down into smaller chunks and start out by moving a small part of your site first, you'll know that you'll be gold.

The value of this approach is that it reduces the risk associated with the move by breaking the migration process down into more manageable chunks. Even if you use this approach, however, you should still follow the guidelines outlined in this section to implement the move of each site section, and check on its progress.

## Changing Servers

You might decide you want to move servers without changing your domain name or any of your URLs. A common reason for this change is that the growth of your traffic requires you to upgrade your hosting environment to a faster server. If you are using third-party hosting, perhaps you are changing your hosting company, or if you have your own data center, you may need to move or expand your facilities, resulting in a change in the IP addresses of your servers.

This is normally a straightforward process, as you can simply go to the registrar where you registered the domain name and update the domain name system (DNS) records

to point to the new server location. You can also temporarily decrease the site's DNS Time to Live (TTL) to five minutes (or something similar) to make the move take place faster. This is really the bulk of what you need to do, though you should follow the monitoring recommendations we will outline shortly.

Even if you follow this process, certain types of problems can arise. Here are the most common:

- You may have content that can't function on the new platform—for example, if you use Perl in implementing your site and Perl is not installed on the new server. This can happen for various other reasons as well, and the result can be pages that return 404 or 500 errors instead of the content you intended.

- Unfortunately, publishers commonly forget to move key content or files over, such as *robots.txt*, analytics files, *sitemaps.xml*, or the *.htaccess* file. It is imperative that these important files are migrated to your new server.

- Server configuration differences can also lead to mishandling of certain types of requests. For example, even if both your old server and your new server are running IIS, it is possible that the new server is configured in such a way that it will transform any 301 redirects you have in place into 302 redirects. Be sure to double- and triple-check that all server directives are properly migrated from the old server to the new.

The best advice for dealing with these concerns is to make a list of special files and configuration requirements and verify that everything is in place prior to flipping the switch on any server moves.

In addition, you should test the new site in its new location before making the move. You will need to access the content on the new site using its physical IP address. So, the page at *http://www.yourdomain.com/pageA* will be found at an address similar to *http://206.130.117.215/pageA*. To access the site, add that IP address to your test machine's hosts file (this assumes you are running Windows) with a corresponding hostname of *http://www.yourdomain.com*, which will allow you to surf the site at the new IP address seamlessly. This advance testing should allow you to check for any unexpected errors. Note that the location of the hosts file varies across different versions of Windows, so you may need to search online to get information on where to find it on your machine.

## Monitoring After Your Server Move

As with our other scenarios, post-launch monitoring is important. Here are the basic monitoring steps you should take:

- Monitor your Google Search Console and Bing Webmaster Tools accounts for 404 errors and to see how well the search engines are handling your 301s. When you see 404 errors, make sure you have a properly implemented 301 redirect in place.

- Monitor the spidering activity on the new domain to make sure no unexpected drops occur.

- Watch your search traffic referrals for unexpected changes.

- You can also check your server logs for 404 and 500 errors, which will sometimes expose problems that your other checks have not revealed.

# Hidden Content

In "Content Delivery and Search Spider Control" on page 334, we discussed ways that you can hide content from the search engines when you want to. However, at times this is done unintentionally—that is, sometimes publishers produce great content and then, for one reason or another, fail to expose that content to search engines.

Valuable content can be inadvertently hidden from the search engines, and occasionally, the engines can find hidden content and construe it as spam, whether that was your intent or not.

## Identifying Content That Search Engines Don't See

How do you determine when you have unintended hidden content? Sometimes the situation is readily apparent; for example, if you have a site that receives high traffic volume and then your developer accidentally `noindex`es every page on the site, you will see a catastrophic drop in traffic. Most likely this will set off a panicked investigation, during which you'll quickly identify the `noindex` issue as the culprit.

Does this really happen? Unfortunately, it does. Here is an example scenario. Suppose you work on site updates on a staging server. Because you don't want the search engines to discover this duplicate version of your site, you keep the pages on the staging server `noindex`ed. Normally, when you move the site from the staging server to the live server, you remove the `noindex` tags, but unfortunately, many site owners forget to do this.

This type of problem can also emerge in another scenario. Some webmasters implement a *robots.txt* file that prohibits the crawling of their staging server website. If this file gets copied over when the site on the staging server is switched to the live server, the consequences will be just as bad as in the `noindex` scenario just outlined. The best way to prevent this type of situation is to implement a series of safety checks on the site that take place immediately after any update of the live server.

There are potential problems, however, that are much more difficult to detect. First, with a new site launch, you won't have any preexisting traffic, so there will be no drop in traffic levels to alert you that something is wrong. In another scenario, you may have an established site where you accidentally do something to hide only a portion of the site from the engines, so the issue is less obvious.

Regardless of your situation, web analytics can help you in the detection process. Use your analytics software to find pages on your site that get page views but no referring search traffic. By itself, this is not conclusive, but it provides a good clue as to where to start. Note that the converse of this is interesting for another situation—if you see content that is getting search referrals even though you don't want or expect it to, you may want to hide that content.

Another data point you can examine is the number of pages the search engines report as indexed for your site. In a new site scenario, you can look at this to see whether the search engines appear to be picking up your content. For example, if you have a site with 1,000 pages with a good inbound link profile, and after three months only 10 pages are indexed, that could be a clue that there is a technical problem. Using multiple sitemap files, one for each site content area covering a specific segment of URLs, can be helpful in diagnosing such problems.

You do need to be careful not to overreact to the count of indexed pages, because the numbers that the search engines report will naturally fluctuate quite a bit. But if you are aware of the types of numbers typically reported for your site, and they drop to an unusually low level and stay there (or keep dropping), you probably have a problem.

## Identifying the Cause of Non-Spidering

Once you realize you have a spidering problem, there are a series of checks you can perform to identify the potential causes.

### Blocked by robots.txt

Check your *robots.txt* file (*http://www.yourdomain.com/robots.txt*) to see whether you are preventing the crawlers from accessing parts of the site that you actually want them to see. This mistake is quite common.

Both Google Search Console (see Figure 12-1 and Figure 12-2) and Bing Webmaster Tools provide simple ways for you to see whether they are aware of content that *robots.txt* is blocking them from crawling.
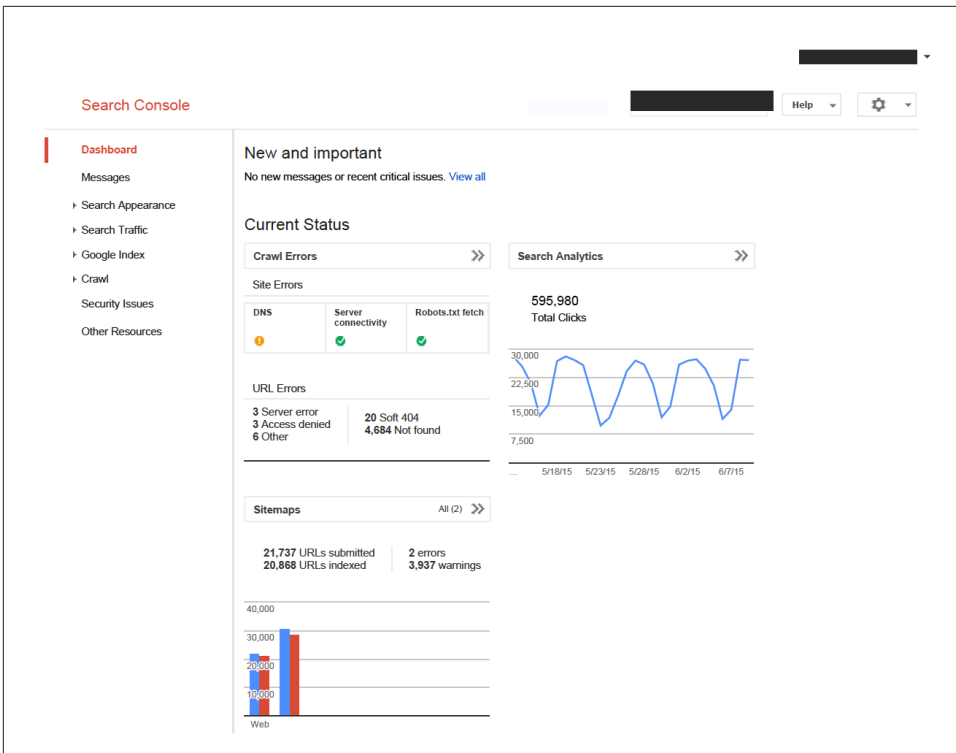
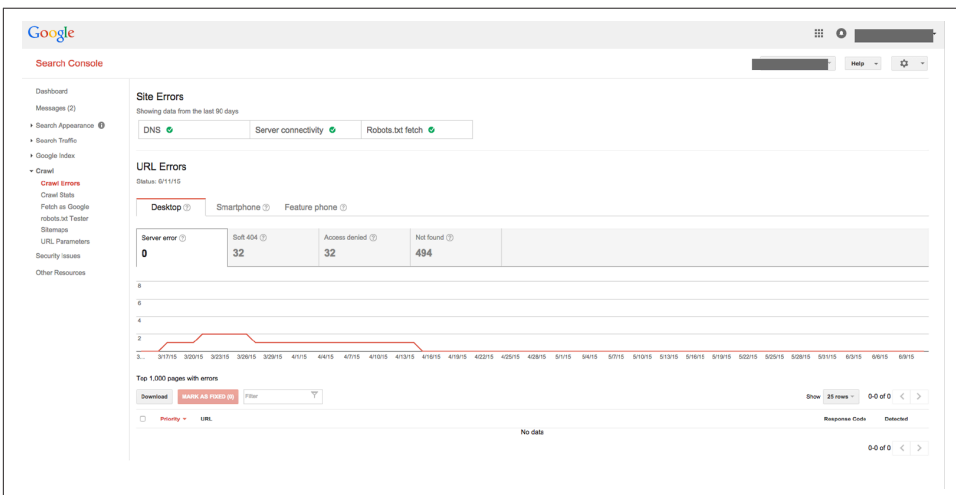**Figure 12-1.** *Google Search Console: restricted by robots.txt*



**Figure 12-2.** *Google Search Console: crawl errors*

These reports are helpful when you have content on the site that has links to it (either internal or external) but that the search engines don't crawl because they are excluded from it in *robots.txt*.

The solution is simple: figure out what line in your *robots.txt* file is blocking the search engines and remove it, or update it so that the content is no longer being blocked.

### Blocked by the robots meta tag

The `robots` meta tag in a page's header might look something like this:

```
<meta name="robots" content="noindex, nofollow">
```

As we discussed in "Content Delivery and Search Spider Control" on page 334, a setting of `noindex` will tell the search engine that it is not allowed to include the page in its index. Clearly, you should check to see whether you have made this error if you find that the engines are not crawling certain pages that you want crawled.

In addition, the `nofollow` tag will tell search engines not to pass any link authority to the pages linked to on that page. If all the links on your site to a particular piece of content are `nofollow`ed, you are passing no link authority to the page. This tells the search engines that you don't value the page, and as a result they won't treat the links as endorsements for it.

Solving this problem requires locating the places where these `robots` meta tags are on your site and removing them. Note that the default setting for the `robots` meta tag is `"index,follow"`, so there is no need to implement the tag if that is your desired setting. Just make sure you don't have `robots` meta tags in place that change the default in places where that is not the desired behavior.

### No direct links

You may find that a particular piece of content has no links to it, or that all of the links to that content are `nofollow`ed. You can also make links invisible to the search engines (possibly unintentionally) by showing them only to visitors that support cookies, or by encrypting the links to the content in some fashion. The solution here is to make sure you implement plain-text (or image) links to the content. Better still, get some third-party websites to link to the content as well.

### Form submission requirement

Requiring a login or some other type of form submission to see content is another common cause of non-spidering. Search engines will not attempt to fill out forms to see what is behind them. The simplest solution is often to remove the requirement for the form if you want the search engines to index this content.

However, some sites sell content on a subscription basis (also referred to as being behind a "paywall"), and they will not want to offer their content for free. In October 2008, Google announced the First Click Free program (*http://bit.ly/reg_sub_sites*), which allows subscription-based sites to have their content crawled and indexed by Google, but still allows the publisher to require human visitors to subscribe to access the content. As of 2014, Google requires that any user coming from a Google domain must be able to see a minimum of five free articles per day, which it refers to as "metering." You can read more about First Click Free in "Content Delivery and Search Spider Control" on page 334.

## Session IDs

Session IDs confuse search engine crawlers: every time the engines come to your site, they see a different page. For example, they may see *http://www.yourdomain.com?SessID=2143789* one time and *http://www.yourdomain.com?SessID=2145394* the next. Even though your intent is to track the session of a particular user, and you think of these URLs as the same page, the search engine does not. You can read more about session IDs in "Controlling Content with Cookies and Session IDs" on page 330.

## Not enough link authority to remain in main indexes

Sometimes the non-spidering problem has nothing to do with the issues we just discussed. The search engines may see the page just fine, but there may not be enough link juice going to it to merit inclusion in their main indexes. This is more common than people think, and it happens because the search engines do not attempt to index all the world's web pages.

For example, content that Google perceives to be of low importance (i.e., content that doesn't have enough link authority, or is perceived to be duplicate content) will be excluded from the main index. In previous years, this content may have been relegated to what Google called its "supplemental index," but in 2014 Google's John Mueller confirmed that for the purposes of treating pages differently, Google no longer has a supplemental index.

Google wants to emphasize the more important pages on the Web, and doesn't want the rate at which it delivers search results to be slowed down by pages that most people probably don't want to see.

# Identifying Hidden Content That May Be Viewed as Spam

Hidden text is one of the challenges that webmasters and search engines still face. Spammers continue to use hidden text to stuff keywords into their pages, for the purposes of artificially boosting their rankings. Search engines seek to figure out when spammers are doing this and then take appropriate action. There are many ways to

create hidden text unintentionally, though, and no one wants to be penalized for something they did not intend to do. Google's Webmaster Guidelines for hidden text state the following:

> Hiding text or links in your content to manipulate Google's search rankings can be seen as deceptive and is a violation of Google's Webmaster Guidelines.

In a Google Groups thread (*http://groups.google.com/group/Google_Webmaster_Help-Indexing/browse_thread/thread/928aa76a1226cf89/32b089e3248cef78?pli=1*), Googler Susan Moskwa had this to say:

> Of course, as with many techniques, there are shades of gray between "this is clearly deceptive and wrong" and "this is perfectly acceptable." Matt [Cutts, former head of Google's webspam team] did say that hiding text moves you a step further towards the gray area. But if you're running a perfectly legitimate site, you don't need to worry about it. If, on the other hand, your site already exhibits a bunch of other semi-shady techniques, hidden text starts to look like one more item on that list. It is like how 1 grain of sand isn't noticeable, but many grains together start to look like a beach.

Related to this is a posting by Matt Cutts on Threadwatch (*http://www.threadwatch.org/node/4313#comment-26883*):

> If you're straight-out using CSS to hide text, don't be surprised if that is called spam. I'm not saying that mouseovers or DHTML text or have-a-logo-but-also-have-text is spam; I answered that last one at a conference when I said, "imagine how it would look to a visitor, a competitor, or someone checking out a spam report. If you show your company's name and it is Expo Markers instead of an Expo Markers logo, you should be fine. If the text you decide to show is "Expo Markers cheap online discount buy online Expo Markers sale..." then I would be more cautious, because that can look bad.

Obviously, this is a fate you want to avoid. Note the use of the word *perceived* in the Google Webmaster Guidelines snippet. It doesn't sound like a simple black-and-white problem, does it? In fact, it is not, as there are many ways to create hidden text.

## Unintentionally creating hidden text

There are a few ways to create hidden text without intending to do so. One of the most common ways is via your CMS, which has some CSS-based methods built into it. For example, many content management systems use the `display:none` technique to implement drop-down menus or other widgets that "expand" to display more text when clicked. Tab folders are a great example of this. Sometimes the `display:none` technique is used in user-generated content systems where the page normally shows

the number of comments on a post, but suppresses the text "0 Comments" in the event that no comments have been made.

People can also commonly create hidden text when they provide enhancements for the visually impaired. For example, you may have a Flash object on your web page and want to provide users with a text description of the content. You may not want to place the text on the page, as it might make the page look cluttered to a user with normal vision. The solution some people use to serve both audiences is to hide the text from the sighted users.

Many of these scenarios have no SEO value, even when manipulated by spammers. These types of techniques generally do not carry a risk of being penalized, because there is no reason to suspect negative intent.

## Spam Filtering and Penalties

Over time, it has become a lot more difficult to "game" the search engines and a lot easier to fall victim to a search engine penalty or outright ban. It is hard to recover from these.

Consequences can include ranking penalties, removal of the site's "voting" power (i.e., ability to pass PageRank), incomplete indexation (i.e., a partial site ban), or, worst of all, a total site ban.

Not even the largest corporations spending big dollars on Google AdWords are immune. In 2013, both BBC News and the popular music lyrics website Rap Genius were hit with Google penalties. In the case of BBC News, Google Webmaster Tools (what Search Console was called at the time) sent an "unnatural link" notification (which represents 1%–2% of Google Webmaster Tools/Search Console notifications.[1] While the case turned out to be targeted to only one article, it caused a stir.[2]

The Rap Genius case, however, was a much more serious matter that involved the company's invitation to bloggers to add links to Rap Genius lyric URLs in exchange for Rap Genius subsequently tweeting the bloggers' posts to its Twitter followers. This link scheme came to the attention of Google's Matt Cutts, who vowed Google would investigate, and it did—resulting in Rap Genius incurring an official Google penalty for link schemes.[3]

---

1 Barry Schwartz, "Google Sends Hundreds of Thousands of Webmaster Notifications Each Month & 90% Are Black Hat Related," Search Engine Land, February 14, 2013, *http://bit.ly/black_hat_notifications*.

2 Barry Schwartz, "Google Penalized One Article On BBC's Web Site," Search Engine Land, March 18, 2013, *http://bit.ly/bbc_penalized*.

3 Barry Schwartz, "Google Has Officially Penalized Rap Genius for Link Schemes," Search Engine Land, December 25, 2013, *http://bit.ly/rap_genius_penalized*.

Search engines rely primarily on automated means for detecting spam, with some auxiliary assistance from paid evaluators, spam vigilantes, and even your competitors. Search engineers at Google and Microsoft write sophisticated algorithms to look for abnormalities in inbound and outbound linking, in sentence structure, in HTML coding, and so on.

As far as the search engines are concerned, SEO has both an acceptable side and an unacceptable side—and in general terms, many efforts intended to boost a site's search engine ranking without improving the measurable value of a page can be considered spamming.

Each search engine has different published guidelines. Here is where you can find them:

- Google's Webmaster Guidelines
- Bing Webmaster Guidelines

The search engines have varying degrees of tolerance for SEO tactics. Anything that violates these guidelines, pollutes the search results with irrelevant or useless information, or would embarrass you if your Google AdWords or Bing rep discovered it is unsustainable, and should generally be avoided.

There is a difference between "search engine friendly," and crossing the line into spam territory. Search engine friendly can mean, for example, that the site is easily accessible to spiders, even if it is database-driven; that HTML code is streamlined to minimize the amount of superfluous code; that important headings, such as product names, are set apart from the rest of the text (e.g., with <h1> tags) and contain relevant keywords; or that link text is contextual, instead of comprising just "click here" or "more info" references.

Contrast these basic SEO practices with the following manipulative search engine spam tactics:

- Serving pages to the search engines that are useless, incomprehensible, unsuitable for human viewing, or otherwise devoid of valuable content—such as *doorway pages*, which SEO vendors may refer to by more innocuous names, including *gateway pages*, *bridge pages*, *jump pages*, *attraction pages*, *advertising pages*, *channel pages*, *directory information pages*, *search engine entry pages*, *satellite sites*, *mini sites*, *magnet sites*, or *shadow domains*. Whatever you call them, by definition they are created for the sole purpose of boosting search engine rankings.
- Creating sites with low-quality content. There are many techniques for doing this, including:

- —Duplicating pages with minimal or no changes and exposing them to the same search engines under new URLs or domains.

- —Machine-generating content to chosen keyword densities (e.g., using a technique such as Markov chains, which are not recommended).

- —Incorporating keyword-rich but nonsensical gibberish (also known as *spamglish*) into site content.

- —Creating a low-value site solely for affiliate marketing purposes (see "Duplicate Content" on page 788 for a more complete definition of *thin affiliate*).

- Repeating the same keyword phrase in the `<title>` tag, the `<h1>` tag, the first `alt` attribute on the page, the meta description, the first sentence of body copy, and the anchor text in links pointing to the page.

- Targeting topically irrelevant keywords.

- Concealing or obscuring keyword-rich text or links within the HTML of a page so that it is not visible to or accessible by human users (i.e., by placing it within comment tags, `<noscript>` tags, or `<noframe>` tags; or by using colored text on a similarly colored background, tiny font sizes, layers, or links that don't show as links to users because they are not highlighted in some manner, such as with an underline).

- Hijacking or stealing content from other sites and using it as content fodder for search engines. This is a practice normally implemented through scrapers.

- Purchasing links for the sole purpose of influencing search rankings.

- Participating in *link farms* (which can be distinguished from directories in that they are less organized and have more links per page) or *reciprocal linking schemes* (link exchanges) with irrelevant sites for the purpose of artificially boosting your site's importance.

- Peppering websites' guest books, blogs, or forums in bulk with keyword-rich text links for the purpose of artificially boosting your site's importance.

- Conducting *sneaky redirects* (immediately redirecting searchers entering your site from a keyword-rich page that ranks in the search engine to some other page that would not rank as well).

- *Cloaking*, or detecting search engine spiders when they visit and modifying the page content specifically for the spiders to improve rankings.

- *Negative SEO*, or submitting your competitors to link farms and/or otherwise creating third-party associations on the competitors' behalf so that they will be penalized.

These tactics are questionable in terms of effectiveness and dubious in the eyes of the search engines, often resulting in the offending site being penalized by or banned from the search engines—a risk that's only going to increase as the engines become more aggressive and sophisticated at identifying and removing offenders from their indexes. We do not advocate implementing these tactics to those interested in achieving the long-term benefits of SEO.

The search engines detect these tactics not just through sophisticated automated spam-catching algorithms, but also through spam reports submitted by searchers—and yes, by your competitors.

Oftentimes search marketers don't even know they're in the wrong and running afoul of search engine guidelines. For example, search engines place more scrutiny on pages that show signs of potential deception, such as `<no-archive>` tags, `<noscript>` tags, `<noframe>` tags, and cloaking, even though all of these can be, and often are, used ethically.

Seeing SEO strictly as a chess game between yourself and the search engines is a short-sighted view, as the goal of the search engines is to provide relevant search results to their users, and your goal as a search marketer is to provide people with valuable content and a positive website experience. Trying to fool the search engines and take unfair advantage by exploiting how they work is simply not a sustainable approach.

Chapter 9 addresses the topic of Google's manually applied penalties, as well as its Panda and Penguin algorithms, in detail.

## Low-Quality Domains and Spam Sites

You can spot a poor-quality website in many ways, not the least of which is the "common sense" check. Would you hire a company with a website named *www.seoseosan-franciscoseo.com* or *www.sonysonyelectronicssonyforsale.com*? The domain, of course, is only one signal, and search engines rely on a wide range of signals as indicators of quality. Some of the most obvious signals are site owners who are actively spamming the search engines with their offsite activities—for example, if the site is actively buying links, or text-link-spamming blog posts, forums, and article comments.

However, there are also less obvious signals. Many such signals mean nothing by themselves and gain significance only when they are combined with a variety of other signals. When a number of these factors appear in combination on a site, the likelihood of it being seen as a low-quality or spam site increases.

Here is a long list of some of these types of signals:

- Short registration period
- High ratio of ad blocks to content

- JavaScript redirects from initial landing pages
- Use of common, high-commercial-value spam keywords such as *mortgage*, *poker*, *texas hold 'em*, *porn*, *student credit cards*, and related terms
- Many links to other low-quality spam sites
- Few links to high-quality, trusted sites
- High keyword frequencies and keyword densities
- Zero, or very little, unique content
- Very few direct visits
- Registered to people/entities previously associated with untrusted sites
- Not registered with services such as Google Search Console or Bing Webmaster Tools
- Rarely have short, high-value domain names
- Often contain many keyword-stuffed subdomains
- More likely to have longer domain names (as above)
- More likely to contain multiple hyphens in the domain name
- Less likely to have links from trusted sources
- Less likely to have SSL security certificates
- Less likely to be in high-quality directories such as DMOZ, Yahoo!, and Best of the Web
- Unlikely to have any significant quantity of branded searches
- Unlikely to be bookmarked in services such as Delicious
- Unlikely to get featured in social voting sites such as Digg, reddit, StumbleUpon, and so forth
- Unlikely to have channels on YouTube, communities on Google+ or Facebook, or links from Wikipedia
- Unlikely to be mentioned on major news sites (either with or without link attribution)
- Unlikely to be registered with Google and Bing local listings
- Unlikely to have a legitimate physical address/phone number on the website
- Likely to have the domain associated with emails on blacklists
- Often contain a large number of snippets of "duplicate" content found elsewhere on the Web
- Frequently feature commercially focused content

- Many levels of links away from highly trusted websites

- Rarely contain privacy policy and copyright notice pages

- Rarely listed in the Better Business Bureau's Online Directory

- Rarely contain high-grade-level text content (as measured by metrics such as the Flesch-Kincaid Reading Level)

- Rarely have small snippets of text quoted on other websites and pages

- Commonly employ cloaking based on user agent or IP address

- Rarely have online or offline marketing campaigns

- Rarely have affiliate link programs pointing to them

- May have links to a significant portion of the sites and pages that link to them

- Extremely unlikely to be mentioned or linked to in scientific research papers

- Unlikely to use expensive web technologies (Microsoft Server and coding products that require a licensing fee)

- More likely to contain malware, viruses, or spyware (or any automated downloads)

- Likely to have privacy protection on the whois information for their domain

It is important to note that while many of these signals can be viewed negatively in aggregate, having one, such as a private domain registration, in and of itself is not going to be interpreted as a spam signal. Many legitimate sites will have one or more of these signals associated with them. For example, there are many good sites with a *.info* TLD.

There are also some signals that require data from a web analytics tool (which Google may be able to obtain from the Google Analytics account of the site in question):

- Rarely receive high quantities of monthly visits

- Rarely have visits lasting longer than 30 seconds

- Rarely have visitors bookmarking their domains in the browser

- Unlikely to buy significant quantities of PPC ad traffic

- Rarely have banner ad media buys

- Unlikely to attract significant return traffic

There can be legitimate reasons for the occurrence of many (possibly even most) of the aforementioned signals. For instance:

- Businesses outside the United States will not be in the Better Business Bureau directory.

- The site may not be relevant to scientific research papers.

- The publisher may not be aware of Google Search Console or Bing Webmaster Tools.

As these three examples illustrate, it's important to put all of these signals into proper context before making an assessment of site quality.

## Spam Reports

Search engines supplement their internal spam fighting efforts by allowing users—including your competitors—to submit spam reports. For example, Google provides a form for reporting spam at *https://www.google.com/webmasters/tools/spamreport*.

In addition, the search engines can and do make use of human reviewers who conduct quality reviews.

In late 2012, Google published "Search Quality Rating Guidelines", a "CliffsNotes" version of the criteria it uses for search quality rating.

## Duplicate Content

As we discuss in "Content Management System Issues" on page 362, there are many ways to unintentionally create duplicate content. For this reason, search engines' normal response to duplicate content is to filter it out, but otherwise not penalize the publisher for it. They filter it out because they don't want to show multiple copies of the same piece of content in their search results, as this does not really bring any value to users. They don't punish the publisher because the great majority of these situations are unintentional.

In fact, in late 2013, Matt Cutts declared that 25%–30% of content on the Web is duplicative and that unless the content duplication was egregious and malicious, you don't need to worry too much about it.[4]

However, there are three notable exceptions that can result in your site being impacted by Google's Panda algorithm (*http://bit.ly/google_panda_update*):

---

[4] For more information, check out "How Does Google Handle Duplicate Content?" on the Google Webmasters YouTube channel.

*Copyright violations*

In this scenario, there is an actual copyright violation where a publisher is showing a copy of another publisher's content, usually via scraping, without permission.

In 2014, Google announced its new Scraper Report), with Matt Cutts tweeting: "If you see a scraper URL outranking the original source of content in Google, please tell us about it."

*Thin affiliate sites*

In this scenario, the publisher has permission from another publisher (the "Content Owner") to display its content, but the Content Owner also shares that same content with many other sites. The common scenario is a Content Owner running an affiliate network and generating leads or sales for its products or clients, largely by offering an affiliate program to other publishers.

All of the affiliate sites publish the same (or very similar) content. The problem the search engines have with these sites is that they offer very little value, as there is nothing truly unique about their content.

The affiliate site may also create hundreds or thousands of pages to target vertical search terms with little change in content. The classic example of this is creating hundreds of web pages that are identical except for the city name referred to on each page ("Phoenix Oil Changes," "Austin Oil Changes," "Orlando Oil Changes," etc.).

*Massive duplication*

This scenario applies to a site that is not a thin affiliate but has a very large amount of duplicate content (even if it has duplicated that content with permission). The threshold is unknown, and it probably changes over time, but our experience suggests that sites in which 70% or more of the pages are duplicates of other pages on the Web are likely to be subject to a penalty.

There may be other spammy forms of duplicate content not identified here. It is likely to be spam if it is implemented intentionally (e.g., the thin affiliate site example qualifies here), if it is intended to manipulate search results, and if the content adds no value to the Web. On the other hand, syndicated content (content published on the original source's website and then copied with express permission on third-party websites), when published properly, presents no duplicate content issues. Coauthor Eric Enge's Search Engine Land article "Syndicated Content: Why, When, and How" provides guidelines for how to properly and safely syndicate your site content.

## Basic Rules for Spam-Free SEO

Especially if you are new to SEO, the first and most important rule is to be familiar with the guidelines of the search engines (see the beginning of this section for their locations).

Second, it is essential that you learn to apply a basic personal filter to your SEO activities. Generally speaking, if you engage in an activity for the sole purpose of influencing search rankings that has no benefit to users or search engine crawlers, you are putting your site's rankings at risk.

For example, if you start buying keyword-rich text links from a bunch of sites across the Web and you do not expect to get significant traffic from these links (just enough to justify placing the advertisements), you are headed for trouble and likely will eventually be hit by Google's Penguin algorithm.

Of course, there are more ways you can get into trouble than through purchasing or aggressively spamming other websites with keyword-rich text links. Earlier in this chapter, we listed a number of spammer techniques. Most publishers/SEO practitioners won't run into the majority of these, as they involve extremely manipulative behavior, as we outlined previously in this section. However, newcomers to SEO do tend to make certain mistakes. Here are some of the more common ones:

- Stuffing keywords into your web page so it is unnaturally rich in those words.
- Overoptimizing internal links (links on your website to other pages on your website). Generally speaking, you do this by overstuffing keywords into the anchor text of those links.
- Cloaking, or showing different content to the search engine crawlers than you show to users.
- Creating websites with lots of very thin content pages, such as the thin affiliate sites we discussed previously.
- Implementing pages that have search engine–readable text that is invisible to users (a.k.a. hidden text).
- Participating in link schemes, link farms, or aggressive guest blog posting, or using other tactics to artificially boost link popularity.

As we discussed previously, there are many other ways to end up in the crosshairs of the search engines, but most of those fall into the category of highly manipulative SEO tactics (generally referred to in the industry as *black hat SEO*).

Ultimately, you want to look at your intent in pursuing a particular SEO practice. Is it something you would have done if the search engines did not exist? Would it have been part of a publishing and promotional plan for your website in such a world?

The notion of intent is something that the search engines look at very closely when evaluating a site to see whether it is engaging in spam tactics. In fact, search engine representatives speak about *intent* (why you did it) and *extent* (how much you did it) as being key things they evaluate when conducting a manual review of a website.

A perceived high degree of intent and pursuing something to a significant extent can lead to more severe penalties. But even if your intent is pure and you don't pursue a proscribed practice to a significant degree, the search engines will want to discount the possible benefits of such behavior from the rankings equation. For example, if you buy a few dozen links and the search engines figure out that the links were paid for, they will discount those links. This behavior may not be egregious enough for them to penalize you, but they don't want you to benefit from it either.

### Search Engine Penalties and Reconsideration Requests

The reality is that penalties are imposed on websites. It is often hard to know what has caused a penalty; although the search engines might describe the general reason for it (such as "unnatural links"), they will not identify the behaviors specifically. You will need to review your site to see where you have deviated from best practices and address all such issues. For a thorough discussion on how to identify search engine penalties and how to recover from them, see Chapter 9.

## Content Theft

The Web is sometimes a little bit of a Wild West environment, and one of the bad things that can occur is that another publisher (a spammer) can copy your content and republish it on one of his own sites. This is very easy for infringers to do. They simply "View Source" in their browsers and take whatever they like without regard to copyright, or send a spider to collect all of your pages in one fell swoop.

You can locate copyright infringers pretty easily using various online tools, including Plagium and Copyscape, if they've lifted some of your page copy. It's much more difficult if they've limited their sticky fingers to just your design.

There are six major options for dealing with an infringer:

- File a DMCA (Digital Millennium Copyright Act) infringement notification.
- Contact the infringing company's CEO, head of marketing, legal counsel, or whomever you feel is most likely to respond.
- Notify the site's hosting company, if it is not self-hosted.
- Report the activity on your blog, or if your blog is not that popular, get someone with a popular blog to report it.

• Have your lawyer send the infringer a cease and desist letter.

Doing nothing is not an option if you're serious about protecting your intellectual property rights. At the very least, you want to show you have made an effort to protect your IP in the event that ownership comes into question in other scenarios.

A phone call to the infringing company is inexpensive, and it can work. Sometimes publishers who use contract writers are not even aware that they are infringing. But this is often not a viable option if you're dealing with an entity that has masked its domain contact information and is hosting its site with a company that will not respond to your requests, and where a copyright complaint will not be effective (i.e., in countries without copyright laws).

In the United States, the DMCA provides a useful hammer to beat on copyright infringers without the need to contact them yourself, in the form of a DMCA infringement notification. In DMCA legal speak, this is also known as a "takedown notice." Other countries with strong copyright laws have similar processes.

As a content producer, you have the right to enforce your copyright. When your content gets "repurposed" on others' websites without your permission, you can simply file a DMCA infringement notification with the infringer's web hosting provider to get that infringer's website shut down.

You can also get the infringing content delisted from the search engines by filing DMCA notices with Google and the other major engines. It is not a daunting procedure. It might take an hour of your time, and it is well worth it. The process is as follows:

1. First, look up the web host and the domain registrar of the offending site using lookup tools such as DomainTools. You can usually ascertain who the web host is from the name servers and/or the netblock owner.

2. Check the official directory of designated DMCA agents for the host and the registrar. (Hopefully, they're listed.)

3. Prepare a letter to send to the designated agent of the web host. The notice you write should include your contact information, a description of and a link to the content that was copied, the web address of the copied content, a statement that you have a good faith belief that the duplication of the material is not legal, a statement that under penalty of perjury you are the copyright holder, and your signature. Some web hosts will allow you to email your notice to them, making this process even easier.

4. Send a similar notification to the search engines. That will cut off the site's air supply (its traffic) in case the content doesn't get taken down right away. Instructions and contact details for each engine are available at the following URLs:

- Google: *http://bit.ly/removing_content*

- Bing: *https://www.microsoft.com/info/cpyrtInfrg.aspx*

5. If the web host doesn't take the site down promptly, submit a DMCA notice to the infringer's domain registrar. It might be worth sending a notice to the data center that the web host uses before you try the registrar.

---

**NOTE**

DMCA letter templates specific to each search engine can be found at *http://www.mcanerin.com/EN/articles/copyright-03.asp*.

---

Be aware that the DMCA takedown procedure is a double-edged sword. Like any other tool, a DMCA notice can be used for good or evil. Specifically, someone could use one unfairly against you! It happens—sites do sometimes use the DMCA to silence their competitors.

Therefore, you as a website owner need to protect yourself from unwarranted (or at least unwelcome) prosecution. If the potential exists for you to inadvertently host infringing material on your website—for example, if you are hosting online forums, group blogs, blog comments, or other types of content that can be submitted by others —here are some actions you can take to help protect yourself:

- It's helpful if you can qualify as a service provider that can be covered under the Safe Harbor provision. For example, you may qualify if you offer a search engine or a bulletin board system.

  — If so, notify your customers of your policies regarding copyright infringement and the consequences of repeated infringing activity. One way is to make this part of your Terms of Use.

- Publish a page on your website with DMCA filing instructions and state that, if and when you get a DMCA notification, you will act on it.

- Most importantly, check the directory of designated agents; if your company isn't listed there, complete the required form for inclusion in the directory and file it with the Copyright Office (*http://www.copyright.gov/onlinesp/list/a_agents.html*).

Regardless of the path you choose, it is helpful to be able to simply and rapidly prove that you are the original author of the content. One tool that can help with that is the Internet Archive's Wayback Machine. This site keeps copies of websites over time. You enter a website name, click Search, and get a screen that may look a bit like Figure 12-3.
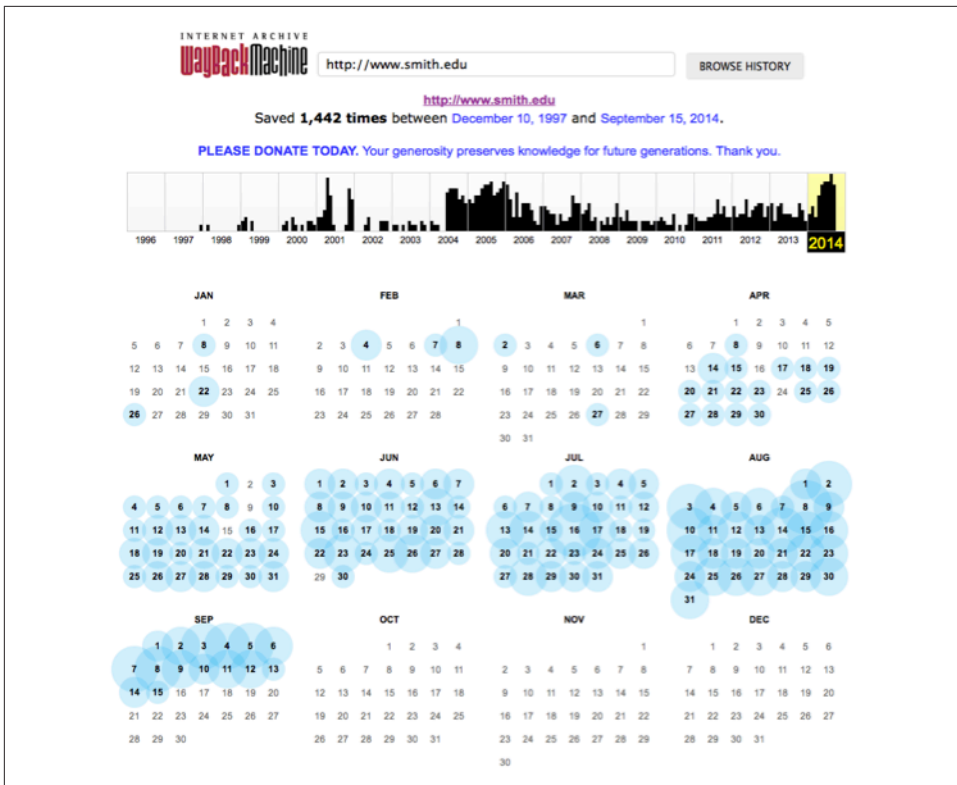
**Figure 12-3.** *Sample archived website data from the Wayback Machine*

You can then click on a date to see the state of the website at that time. The Wayback Machine does not always keep complete site copies, but a lot of data is available there, and it can often be used to provide clear proof that you were the first person to publish a given piece of content.

# Changing SEO Vendors or Staff Members

Another major transition that can happen is when the publisher changes its SEO staff. This can happen both with in-house SEO staff and when you use an external SEO agency (e.g., the publisher can decide to switch agencies, or the agency can have internal changes). Companies make changes in their SEO staffs for many reasons, and issues can arise when this happens.

## Potential Problems with SEO Staff Changes

One potential issue is a loss of momentum. The prior SEO team probably had some programs under way, and the new team will inherit those programs and have to sus-

tain them, or there will be a delay before any results are seen from the new team's SEO efforts.

Even if the new team does its best to sustain an existing SEO strategy, it will take them some time to become familiar with it, and the specific tactics the old team used may not fit the new team's skill set that well. It is best to be prepared for this lost momentum when you're making significant changes in your SEO team. One way to try to deal with this is to hire the new team and have them work for a while in parallel with the old team. This can work in some situations, but it will not be an option in others.

Another situation that can occur is that the new team may have a different philosophy than the old team (this may be why the publisher made the switch). This will invariably mean a restart of efforts, and it can potentially take many weeks to regain the lost momentum.

## SEO Documentation for Actions and Progress

One way to limit exposure to changes in the SEO team is through detailed documentation. Keeping a detailed record of what has been done, when it was done, and why it was done can be invaluable in these scenarios. Ideally, this should be combined with your website analytics data.

Traditional web analytic tools make it easy to go back in time and see traffic data over time, but tools such as Google Search Console and Bing Webmaster Tools don't; nor do the search engines provide historical data on metrics such as indexed pages, link counts, and the like. You should also keep a record of this data over time. If your site meets the data inclusion criteria for services like SEMrush, you may be able to obtain historical ranking data going back up to two years for your site, which can help you understand past organic trends. Archive Google Search Console data every 90 days.

Make sure you have ownership of all relevant website accounts, usernames, and passwords. You don't want any transition to leave you in a situation where you can't access your critical data.

## SEO Documentation for Rapid Training

Make sure that you have documented company processes regarding SEO, and that people associated with the SEO team and/or the management of the SEO team are all familiar with those processes. This can help speed up transitions as well, as the new team can be trained on what those processes are. Part of this includes having a detailed outline of the tasks and opportunities that are the first priorities for the new team.

With this information in hand, you can get things running smoothly much faster after a team reorganization. You can start the new team with one or more training sessions

to get them up to speed on campaign history, login information, and what needs to be done going forward.

The better prepared you are to get new SEO staff up to speed and moving forward, the better off you will be. Bear in mind that you have hired the new team for their expertise; while the initial training sessions may result in the creation of a modified plan, you will get there much faster if you bring all these issues to the table on day one.

### Cleanup and Auditing

Once your migration to new SEO staff (or a new SEO vendor) is complete, you should revisit and remove the old team's access to your analytics and search engine tools systems. This is one good reason to have universal, or "anonymous" business user and email accounts associated with important analytics and SEO logins. For example, tie your Google Analytics to a Google account with an email address like *mybusinessanalytics@gmail.com* or *analytics@mywebsite.com*, as opposed to the personal email account of someone who may or may not be with your company a year or two from now. Each user can be verified and removed individually. Audit this list on a regular basis.

# Conclusion

Large-scale website changes to already established site properties are a fact of life, and they should be handled with as much foresight and planning as possible. In SEO, as in life, the only constant is change—and at the end of the day, we are grateful for this, as it means there is always opportunity for increased search exposure, and for new businesses to enter the playing field and succeed.