10

Free Search Engine Tools and Services

You can communicate information about your site to search engines and see your site from their perspective using some free services and utilities from Yahoo! and Google.

Yahoo! and Google provide some free and exceptionally useful tools and services to help you communicate the structure of your site, and get a clearer understanding of how well it's being indexed. These tools also provide valuable insight about inbound links to your site, keywords generating search referrals, and various other data that can help you assess the success of your findability efforts.

Before evaluating statistics you'll need to make sure search engines are indexing all of the content on your site. You can make it easier for search engine spiders to crawl your site by drawing them a map.

Building and Submitting sitemap.xml

Historically, the communication between webmasters and search engines has been very limited. In the past, once you'd built your site you would submit the home page URL to all of the major search engines to let them know you'd like their spiders to begin indexing your content. With nothing but the home page URL, spiders can potentially overlook some pages in your site, especially those that may only be accessible via your search system. A few years ago Google recognized that this problem could adversely affect the comprehensiveness of its search index, and created a simple solution called `sitemap.xml`.

In June 2005 Google introduced a standardized XML sitemap protocol that allows webmasters to communicate the structure of their site to search engines for more accurate indexing (**http://sitemaps.org**). Today, because the `sitemap.xml` protocol is supported by Yahoo!, Ask, and MSN Live Search as well, the same XML file can let all major search engines know which pages they should index in your site.

The `sitemap.xml` protocol also lets webmasters include information about each page, including the date it was updated, the frequency of change, and how important it is in the site. This type of additional information can help search engines crawl your site more intelligently.

The structure of a `sitemap.xml` file is relatively simple. Here's an abbreviated example that illustrates the tags common to the protocol:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">

<url>
  <loc>http://aarronwalter.com/</loc>
  <priority>1.0</priority>
  <changefreq>daily</changefreq>
</url>

<url>
  <loc>http://aarronwalter.com/about/</loc>
  <priority>0.5</priority>
  <changefreq>monthly</changefreq>
</url>

</urlset>
```

Although this example provides URLs for just two pages on my website, the `sitemap.xml` protocol can describe sites of any scale. The file begins with the XML prologue and a definition of the schema being used. That's not the important part, though. Notice there are two open and close `<url>` tags, each containing different information. Each `<url>` tag defines a different page's location, priority, and change frequency. Because the home page is the most important page in the site it has a higher priority value than the interior page in this example.

You can optionally define the priority and change frequency of your pages from within each `<url>` tag. The `<priority>` tag contains a floating point number from 0.0 to 1.0, where 1.0 is the highest priority. It lets search engines know which pages you deem most important so spiders can prioritize as they index your site.

The `<changefreq>` tag provides search engines general information about how often your pages will change, but don't take this too seriously as it may not correlate to how often your page gets crawled. You'll find a list of possible values for the `<changefreq>` tag and further information about `sitemap.xml` tags at **http://www.sitemaps.org/protocol.php**.

This file is quite short and it is relatively simple so it would be easy to build. If your site had hundreds or thousands of pages it would be far too time consuming and tedious to write it by hand.

Luckily there are scores of desktop and Web applications that will crawl your site and create a `sitemap.xml` file automatically so you can spend more time in your hammock than writing repetitive XML documents. There are even scripts in a variety of languages that you can freely integrate into your sites, Content Management Systems, or Web applications to automate the process even further.

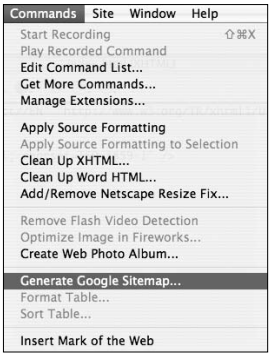


FIGURE 10.1 *George Petrov's Google Sitemap Generator is a free extension for Dreamweaver that greatly simplifies the creation of* `sitemap.xml` *files.*

You'll find an exhaustive list of options at **http://code.google.com/sm_thirdparty.html**. Some options are free while others may cost you a few bucks. If you're a Dreamweaver user you may want to try George Petrov's free Google Sitemap Generator extension available on DMXZone.com (**http://www.dmxzone.com/showDetail.asp?TypeId=3&NewsId=10538**). See **FIGURE 10.1**. Once you've downloaded and installed the extension you'll need to restart Dreamweaver and then define a site in the site manger so the extension can crawl all of the HTML pages in your site. To define a site simply choose Site, select New Site, then enter the information requested.

Once your site is defined, select Commands, choose Create Google Sitemap and the extension will work its magic. The final `sitemap.xml` file will be placed in the root folder of your site.

If you are using a server-side scripting language to create a page template system, you'll find that the extension will have trouble crawling your site's files. Instead, you will need to use a sitemap generation tool that crawls the live site in order to create the `sitemap.xml` file.

One such option is XML-Sitemaps.com (**http://www.xml-sitemaps.com/**). If your site is 500 pages or less, XML-Simaps.com will crawl your site and generate a `sitemap.xml` file for free. For sites that are larger you would need to purchase their PHP script, which you could integrate into your own projects. The script also generates sitemaps in RSS and HTML formats. HTML sitemaps are especially useful to users who've lost their way on your site, or just want a little help understanding the scope of your information. Any search engine that doesn't support the `sitemap.xml` protocol could use the HTML sitemap to navigate your site.

When the script runs, it automatically notifies Google of the file update so it can revisit your site to index recent content. The XML-Sitemaps.com PHP

script will also track down broken links in your site so you can repair them.

If you are looking for an equally feature-rich PHP script to integrate into your projects for free, check out phpSitemapNG (**http://enarion.net/google/**). You can schedule a regular refresh of your `sitemap.xml` file with phpSitemapNG using Cron, a Unix and Linux operating system utility that schedules tasks to run automatically. If you're on a Windows server, use the built-in task scheduler in the Control Panel to run the PHP script instead.

If you're new to Cron, check out Aaron Brazell's helpful article at http://www.sitepoint.com/article/introducing-cron. Windows server users looking for an introduction to Task Scheduler can check out http://www.iopus.com/guides/winscheduler.htm.

There are also good desktop applications that can make short work of `sitemap.xml` file development. RAGE makes Google Sitemap Automator for the Mac (**http://www.ragesw.com/products/googlesitemap.html**) that will crawl your site, generate the `sitemap.xml` file, upload it to your server, and tell Google once it's posted (see **FIGURE 10.2**). Its intuitive interface makes the process very easy. Although you can use the demo version of Google Sitemap Automator to generate and upload your `sitemap.xml` file, you'll have to purchase a license to use the Google notification feature.

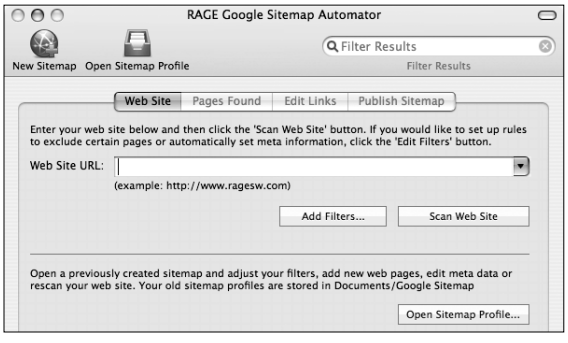


FIGURE 10.2 *RAGE's Google Sitemap Automator (**http://www.ragesw.com/products/googlesitemap.html**) quickly builds a `sitemap.xml` file, uploads it to your server and notifies Google of its location.*

iArchitect makes a great sitemap utility for Windows called Sitemap Generator (**http://www.iarchitect.net/Products/Sitemap-Generator/**). See **FIGURE 10.3**. In addition to creating, uploading, and notifying search engines of your sitemap, Sitemap Generator creates HTML sitemaps and provides scheduling options so you can further automate the building and updating of your `sitemap.xml` file.

If you're not using a sitemap generation program that will automatically upload your `sitemap.xml` file you'll need to do it yourself. Generally the file is placed in the Web root folder on a server, which is typically called `public_html` or `www`. Then you'll need to inform search engines of its location so they can read the sitemap and begin crawling your site.

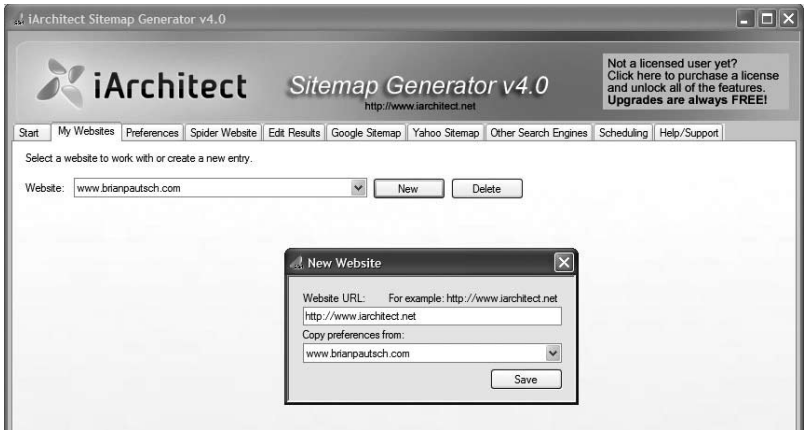


FIGURE 10.3 *Sitemap Generator for Windows offers features beyond most of its competitors, including HTML sitemaps and scheduled updating of your* `sitemap.xml` *file.*

Informing Search Engines About Your sitemap.xml File

There are three ways you can let search engines know about your `sitemap.xml` file once it's uploaded or updated:

- `robots.txt`
- Ping
- Manual submission

For the best results you'll want to use a combination of these three to ensure all search engines are aware of your sitemap.

robots.txt Chapter 3, "Server-Side Strategies," introduced the `robots.txt` protocol used to tell search engine spiders which files or directories on a server should be excluded from indexing. The file is placed in the Web root folder of your server, and is automatically read by all search engine spiders.

A `robots.txt` file can also be used to communicate the location of your `sitemap.xml` file to all search spiders that visit your site. Simply add the following to your `robots.txt` file to define the location:

```
Sitemap: sitemap.xml
```

The benefit of using `robots.txt` to communicate your sitemap's location is that any search engine that supports the protocol will automatically find the file when it visits your site. This approach does not send a message out to search

engines inviting them to index your site. If you're launching a new site, it's a good idea to notify search engines directly as well using one of the next two methods.

Ping A *ping* is a short message from one computer to another. When you publish a new site and want to request a full indexing you can ping search engines individually to let them know where your `sitemap.xml` file is located.

Google, Ask, and Yahoo! all offer ping notification services, which can be used by simply navigating to each in your browser. Replace the highlighted sample URL in these examples with the absolute path to your `sitemap.xml` file.

- **Google:** http://www.google.com/webmasters/sitemaps/ping?sitemap= `http://example.com/sitemap.xml`
- **Ask:** http://submissions.ask.com/ping?sitemap= `http://example.com/sitemap.xml`
- **Yahoo!:** http://search.yahooapis.com/SiteExplorerService/V1/ping?sitemap= `http://example.com/sitemap.xml`

MSN Live Search is conspicuously absent from this list. Although it supports the `sitemap.xml` protocol, at the time of writing it didn't offer a sitemap submission tool of any kind. The only way to let MSN Live Seach know about your sitemap is via your `robots.txt` file.

These ping URLs are also very useful if you're building a content management system or Web application that needs to continually update search engines with new pages. Your application could automatically update your `sitemap.xml` file, then connect to these ping services for you.

To quickly ping all search engines with `sitemap.xml` ping services, create a bookmark folder in your browser, then add bookmarks for each ping service with your `sitemap.xml` URL trailing. Anytime you update your `sitemap.xml` file simply launch these bookmarks to instantly send word of your new content to search engines.

Manual Submission You can also let search engines know about your `sitemap.xml` file by visiting each one and manually submitting your sitemap URL. Unfortunately, only Google and Yahoo! currently provide manual `sitemap.xml` submission tools. For sitemap updates this is certainly the most tedious of the three options. If you're launching a new site, a manual submission is probably a good idea.

As we'll discover in the rest of this chapter, when you manually submit your sitemap to Google and Yahoo! they will parse your file and notify you of any errors they've encountered. They also provide information about the last time they read your sitemap so you'll know if your newest content is in their search index.

NOTE If you need to generate a sitemap for a WordPress blog you'll want to read the section in Chapter 5 entitled "Automatically Generating an XML Sitemap."

Google Webmaster Central Services

Google provides an amazing array of free statistics, diagnostics, and management utilities in Webmaster Central (**http://www.google.com/webmasters/**). See **FIGURE 10.4**. From here you can submit your URL to Google, and check the status of your site to see if it has been indexed. But the utility of Webmaster Central goes far beyond URL submission.

The bulk of Google's free utilities in Webmaster Central can be found in Webmaster Tools (**https://www.google.com/webmasters/tools/**). Before you can use Google's Webmaster Tools, you'll need to create a Google account if

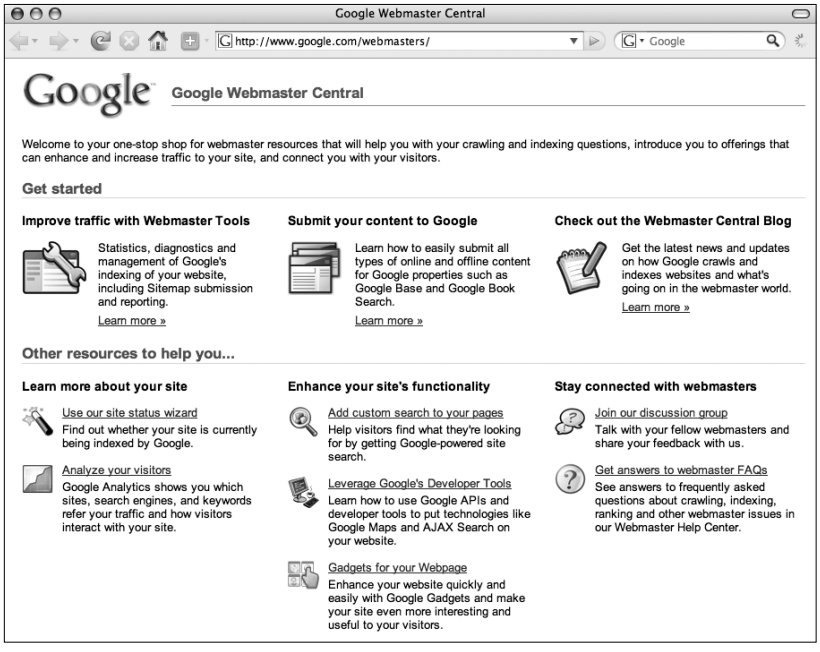


FIGURE 10.4 *Google's Webmaster Central contains a wide variety of free and very useful utilities that provide statistics, diagnostics, and management features.*

you don't already have one. You can manage sitemaps and view statistics for multiple sites from a single account, which is great if you have to keep tabs on many client websites.

Once you've logged in you'll be taken to the dashboard, where you can add your site's URL in order to view data about it and manage some preferences. You'll be prompted to verify that you have the authority to manage the site in one of two ways. You can either add a special meta tag to your home page with a value unique to your site, or upload a special file Google provides to your server's Web root folder. When you've completed one of these tasks you simply click the "verify" button, and Google will visit your site immediately to confirm that the meta tag or file is present. Once you've proven you are the webmaster of the site, you can start using Google's tools.

Webmaster Tools

Google's Webmaster Tools are divided into five key sections:

- Diagnostics
- Statistics
- Links
- Sitemaps
- Tools

When you log in to Webmaster Tools and have clicked on a URL you want to manage you'll be taken to an overview page that provides some quick information (see **FIGURE 10.5**).

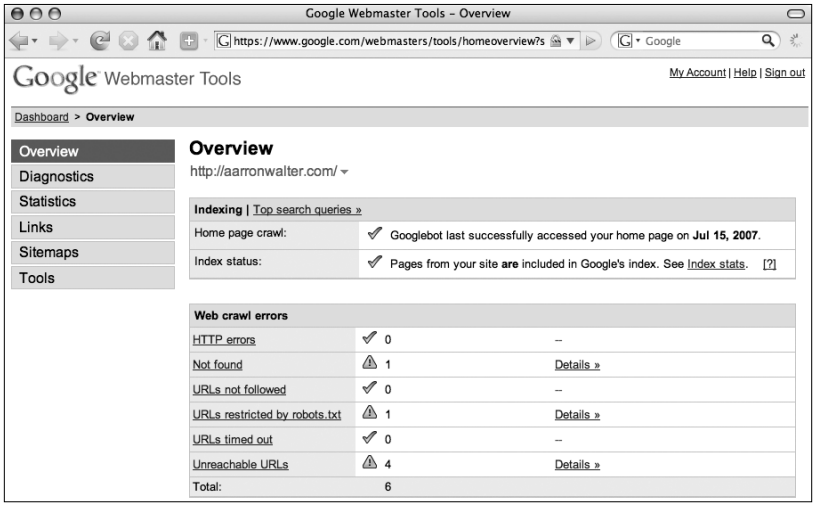


FIGURE 10.5 *The overview page in Google's Webmaster Tools provides quick access to statistics and information that is available within subsections.*

Essential information like 404 errors, pages that were requested but timed out, HTTP errors, or pages blocked by your `robots.txt` file can be viewed at a glance on the overview page. If any errors or issues of concern were detected a link is provided to learn more about the problem.

Also on the overview page you'll find the date of the last index of your home page and indication of whether your site's content is in Google's index. This quick view makes spotting serious problems easier. Let's examine some of the more detailed features of Google's Webmaster Tools.

Diagnostics

In the Diagnostics section you'll find further detail about trouble Google encountered while indexing your site (see **FIGURE 10.6**). There are six types of errors and issues that Google logs:

- **HTTP errors:** server configuration errors, forbidden directories, etc.
- **Not found:** 404 errors
- **URLs not followed:** any pages you may have indicated that were not to be indexed
- **URLs restricted by robots.txt**
- **URLs timed out:** pages that were requested but couldn't be indexed because of a slow network connection, defective code, or some other reason
- **Unreachable URLs:** pages listed in your `sitemap.xml` file but were not reachable

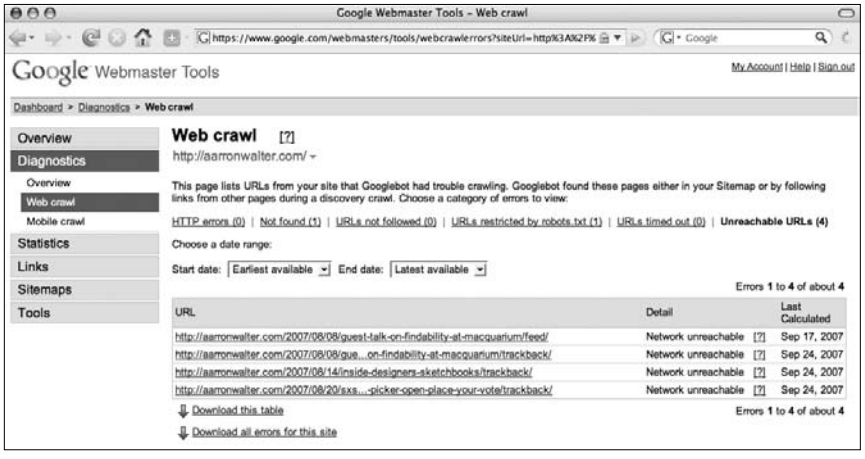


FIGURE 10.6 *You can pinpoint any problems Google encountered when crawling your website from the Diagnostics section of Webmaster Tools.*

You'll notice in the Diagnostics section a sub-section entitled "Mobile Crawl." If your site is delivered in a format specific to mobile devices such as CHTML or WML, Google will still index your content. Look for issues indexing your mobile content here.

Statistics

Statistics is perhaps the most interesting of the Webmaster Tools (see **FIGURE 10.7**). It contains a series of subsections that show very interesting data about what keywords are sending people your way, what Google sees when it crawls your site, index stats, crawl stats, and stats on how many people are subscribing to your RSS feeds. It's a lot of information, all of which can really open your eyes to Google's perspective of your site.

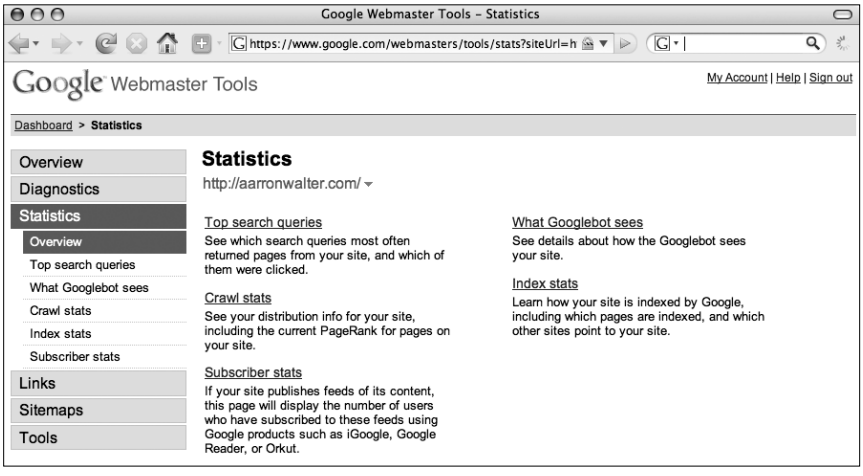


FIGURE 10.7 *The Statistics section of Google's Webmaster Tools provides plenty of detailed information about your site. You can learn what content Google perceives as important on your pages, what keywords or phrases are generating traffic to your site, and more.*

Top Search Queries Top Search Queries is an especially interesting area to explore. Not only can you learn what keywords and phrases are generating traffic to your site, but you can also see what your ranking is for each query. Although it's interesting to discover in this data the keywords you've targeted in your site, it's even more interesting to discover those that you'd never have thought would generate any traffic. You're likely to find some pretty bizarre search phrases that are directing people to your site! Watch this area closely for cues on what content you should expand or enhance on your site.

You can isolate and view segments of the data by geographic location, or search type including blog, image, mobile, or Web.

What Googlebot Sees In the What Googlebot Sees section you can view Google's ranking of keywords and phrases in the content of your site. If you

don't see the keywords you originally targeted at the top of the list, then you'll need to revisit the keyword strategies outlined in Chapter 2, "Markup Strategies," and Chapter 4, "Creating Content that Drives Traffic," to try to improve their prominence in your pages.

Even more intriguing in this section is the listing of keywords other people have included in their links to your site. It's rumored that this is one of the most significant factors that Google weighs when attempting to understand the content of a site and assigning their proprietary PageRank. Unfortunately they don't provide links here to the sites that have created inbound links to your site, but they do in the next section of the Webmaster Tools. We'll explore additional methods of attaining information about inbound links later in this chapter.

The more correlation you can create between the top keywords in your site and the top keywords in inbound link labels, the more success you'll have achieving top rankings for these words. Controlling text in your site is easy, but controlling it on other people's sites is pretty tough. If you know the people running sites that link to yours, you can always make a friendly request that they rewrite their link label to include your target keywords.

Further down the page in this section you'll see a breakdown of the types of content that Google has crawled on your site, including HTML, XML, PDF, Flash, plain text, and other formats.

Crawl Stats The Crawl Stats subsection has less information than others, but the data is equally informative. Here you'll find the average Google PageRank of pages in your site, and your top-rated page for the past three months. Evaluate the page that ranks highest in your site to determine what qualities make it stand out over the others. Usually the quantity and quality of inbound links from reputable sites will determine which page ranks highest. As discussed in Chapter 4, valuable content will usually elicit plenty of inbound links. This is another important bit of information to monitor for ideas on what content is most valuable to your audience.

Index Stats The Index Stats section provides some links to run Google search queries using search operators that will reveal information about your site. You actually don't need to visit this section of Google's Webmaster Tools to view this data. You can simply search for your URL preceded by any of the following search operators in Google's search box to learn a little about how they've indexed your site:

- site: example.com—indexed pages in your site
- allinurl: example.com—pages that refer to your site's URL

- link: example.com–pages that link to your site
- cache: example.com–the current cache of your site
- info: example.com–information Google has about your site
- related: example.com–pages that are similar to your site

A word of warning: The `link` operator provides a pretty incomplete listing of all inbound links to a site. Don't panic if you see a surprisingly short list of results when using this operator. The Links tool, which we'll take a look at shortly, provides a comprehensive list with very useful extended data. You can also use the `allinurl` operator to see a more comprehensive list of inbound links.

Subscriber Stats The information provided in the Subscriber Stats section is nice, but provide a very limited snapshot of the number of people who have subscribed to your site's RSS feeds. Google only tracks subscriptions within its own RSS aggregators. If users are subscribing to your feeds in Bloglines (**http://bloglines.com**), Netvibes (**http://Netvibes.com**), or any other RSS aggregator besides those created by Google, you won't see this data reflected here.

Take this information with a big grain of salt. In Chapter 13, "Analyzing Your Traffic," we'll take a look at FeedBurner's (**http://feedburner.com**) subscription statistics, which provide a more complete snapshot of how many people are subscribing to and reading your feeds.

Links

The Links section provides really interesting data about what sites are linking to yours, and which pages in your site are garnering the most inbound links (see **FIGURE 10.8**).

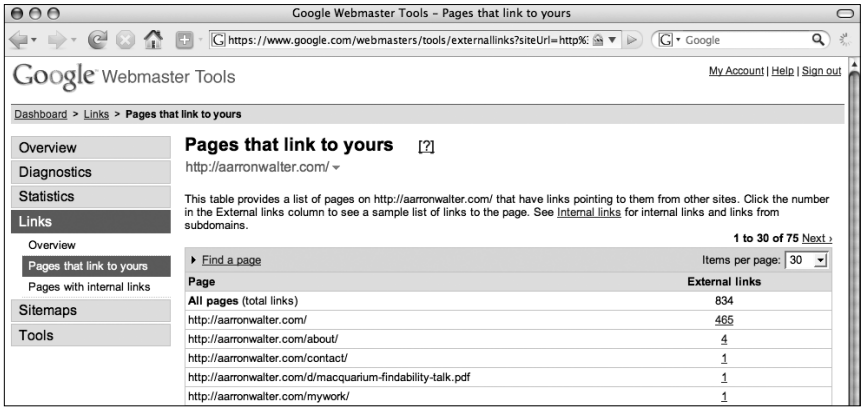


FIGURE 10.8 *Using the Links tool you can identify which pages in your site are receiving the most inbound links, and what sites are linking to them.*

Pages that are receiving many inbound links are obviously providing your audience with the type of content they find interesting. Watch this information closely so you can determine which pages are popular, and worth expanding to generate even more traffic.

The Links section also lists all of the pages in your site that have internal links to other pages in your site. This is primarily useful to determine if you are cross-linking enough to help circulate traffic through the site. The more traffic circulation you can create, the longer your users will stay on your site, and perhaps complete business objectives, like make a purchase or sign up for the mailing list.

Sitemaps

Earlier in this chapter you learned how to create a `sitemap.xml` file and submit it to search engines. Providing search engines with a sitemap helps them more intelligently index your content.

In the Sitemaps section of Google's Webmaster Tools you can post the URL for your `sitemap.xml` file and monitor its status (see **FIGURE 10.9**). If you're launching a new site it's a good idea to post your sitemap here so you can observe any parsing problems that Google might encounter.

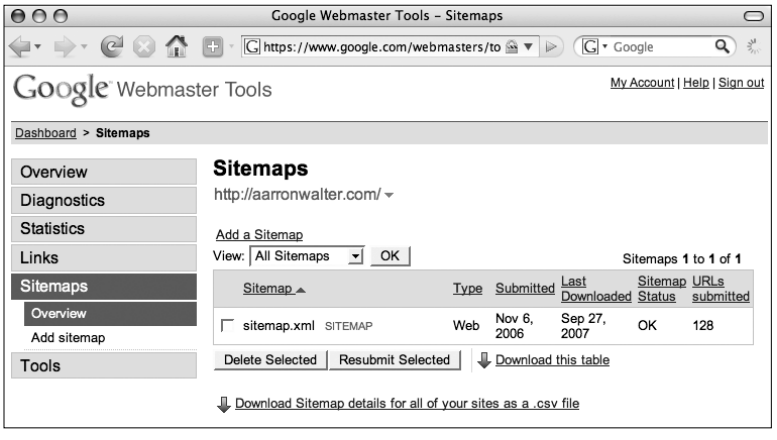


FIGURE 10.9 *With the Sitemaps tool you can post the URL for your sitemap file and monitor its status. Google will let you know if it runs into trouble reading your file.*

Besides letting you know if it encounters errors parsing your sitemap file, the Sitemaps tool also tells you when the file was last read, and how many URLs were included. You can also provide a separate `sitemap.xml` file if you have a mobile version of your site.

Tools

In the Tools section (see **FIGURE 10.10**), you can analyze your `robots.txt` file, manage the site verification process you selected, set the rate at which Google will crawl your site, define the preferred domain name format for your site, and remove certain URLs from Google's index.

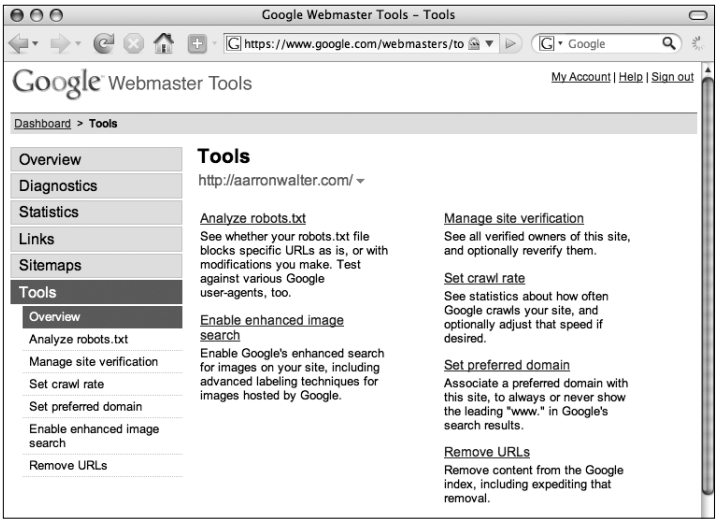


FIGURE 10.10 *The Tools section provides a host of useful preferences and utilities.*

Analyze robots.txt You can identify parse errors in your `robots.txt` file using the Analyze `robots.txt` tool. If Google has been to your site and found a `robots.txt` file, you'll see its content on the page here. This tool also lets you enter URLs to pages on your site to test whether your `robots.txt` file will prevent Google from indexing them.

Manage Site Verification If for some reason you need to get another look at the meta tag or file name Google is using to verify that you are the owner of the site being managed, you'll find that information in the Manage Site Verification section. Unfortunately, there's no way to switch verification methods.

Set Crawl Rate You can keep tabs on the frequency and speed at which Google is crawling your site in the Set Crawl Rate section. Information such as the average number of pages on your site Google indexes per day, the average number of kilobytes downloaded per day, and the time it takes Google to load your pages can be found in this section.

You'll notice a dramatic drop in load times if you optimize your site's performance as outlined in Chapter 3, "Server-Side Strategies." **FIGURE 10.11** shows a big change in the time it took Google to crawl my site before optimization and after.

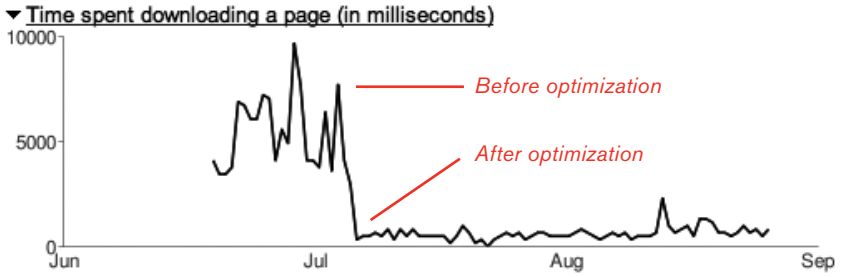


FIGURE 10.11 *After optimizing my site, I saw an approximately 90 percent speed increase in Google's indexing of my content.*

If your site suddenly becomes extremely popular, and bandwidth is more of a concern than keeping Google's index current, you can throttle back the indexing frequency of your site in this section.

Set Preferred Domain In Chapter 3 you learned that when Google indexes sites, it sees URLs with and without the preceding www as entirely different sites. Because the URL http://www.mysite.com might have more inbound links to it than http://mysite.com, Google might assign it a higher PageRank even though these URLs go to the same site. This is called the Google canonical problem.

In the Set Preferred Domain section, you can tell Google to choose one URL format so it doesn't split your PageRank. Chapter 3 provides another solution to the problem that uses Apache's `mod_rewrite` module to remap all page requests to a single, consolidated format. It's not a bad idea to do both to cover your bases with Google and other search engines as well.

Enable Enhanced Image Search If you're OK with your images being discovered via Google's image search, you can choose to enable enhanced image search. Images can be located via their file names or `alt` text, but Google has an even more ingenious approach that makes their image search more accurate.

Google Image Labeler (**http://images.google.com/imagelabeler/**) is an image identification project that presents volunteers with a random series of images gleaned from sites that have enabled the enhanced image search option. As volunteers view each image, they provide descriptive meta data that Google's image search uses to generate exceptionally accurate results to queries. Because image search is hugely popular, enabling this option can generate a lot of traffic to your site.

Remove URLs If Google has indexed content on your site you'd rather it didn't, you can submit a request to have it removed in the Remove URLs

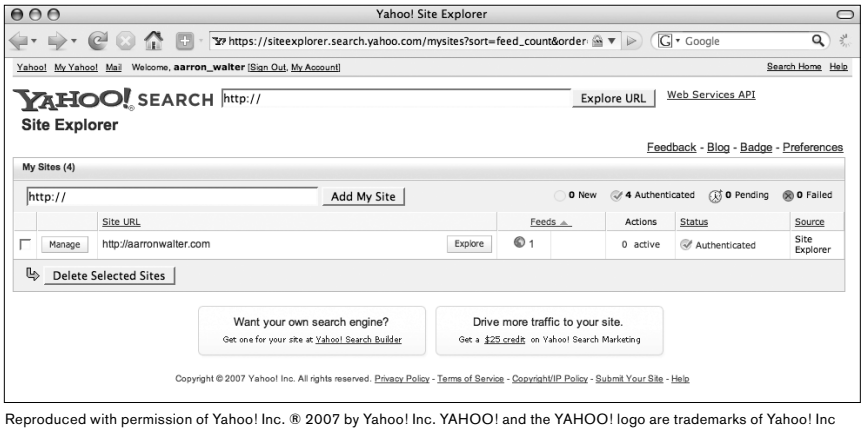
section. Of course you can block indexing using `robots.txt` (see Chapter 3, "Server-Side Strategies") or the `noindex` meta tag (see Chapter 2, "Markup Strategies"). But neither method will immediately remove a page from Google's index. The content would only be removed from the index once Google returns to re-index your site. If you need something removed immediately, the Remove URLs section is the place to do it.

Getting Info About Your Site with Yahoo! Site Explorer

Site Explorer is a free set of tools that provide insight into how Yahoo! is indexing your site (**https://siteexplorer.search.yahoo.com/**), and who is linking to your pages. Site Explorer is also where you would submit a `sitemap.xml` file to Yahoo! when you launch a new website.

To use Site Explorer you'll need to create a Yahoo! account, if you don't already have one, then authenticate your site following a similar process as Google's Webmaster Tools. You can prove a site is yours by adding a special meta tag to your home page or by uploading a file Yahoo! provides to your server's Web root folder for authentication. Once you've completed one of these tasks you can initiate Yahoo! to verify that you are the owner of the site.

From the main control panel area called My Sites you can add URLs you'd like to explore, and keep tabs on your site's authentication status (see **FIGURE 10.12**). You can let Yahoo! know the location of your `sitemap.xml` file in the Feeds section of Site Explorer, which can be found by clicking the Manage button next to your URL.



Reproduced with permission of Yahoo! Inc. ® 2007 by Yahoo! Inc. YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc

FIGURE 10.12 *You can manage and explore any number of sites in Yahoo!'s Site Explorer.*

In the Feeds section, simply enter the path to your sitemap file on your server and Yahoo! will read it, then crawl your site (see **FIGURE 10.13**). Here you can also define a sitemap for a mobile version of your site, if you have one. The date and time Yahoo! last read and processed your sitemap file will be displayed here as well.

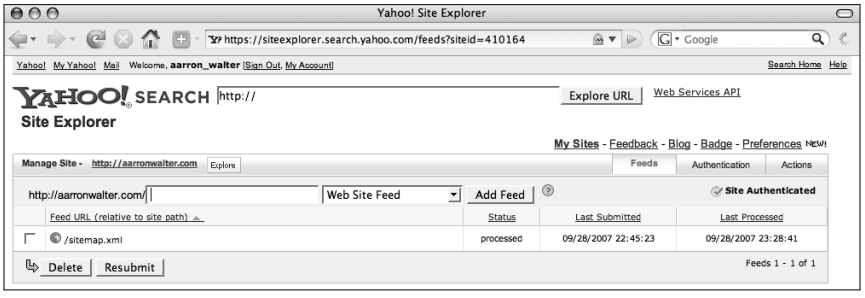


FIGURE 10.13 *You can let Yahoo! know the location of your Web and mobile* `sitemap.xml` *files in the Feeds section of Site Explorer.*

The real heart of Site Explorer is the Explore tool, which you can access by returning to the My Sites page, then clicking the Explore button to the right of your URL. Site Explorer lists all of the pages that Yahoo! has indexed in your site (see **FIGURE 10.14**). This is especially useful when you first launch your site, as you can keep a close eye on what content has officially made it into the Yahoo! index. You can also remove any page from Yahoo!'s index by clicking the Delete URL/Path button, which is visible when you hover over any page record.

You can view a comprehensive list of sites that link to your pages by clicking the link labeled Inlinks at the top of the page. This is important information, as the more links from reputable sources your site receives the higher your page

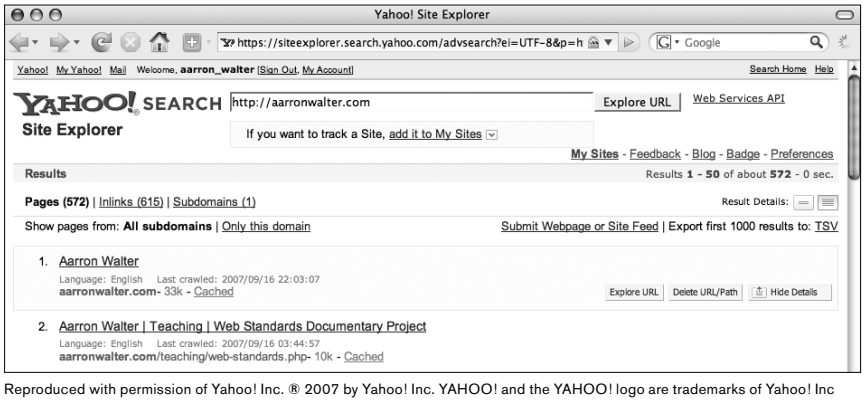


FIGURE 10.14 *Site Explorer lists all pages in your site that Yahoo! has indexed. Here you can also view a complete listing of sites that link to yours.*

rank will be. If you've asked friends, colleagues, and affiliates to create a link on their site to yours, you can watch this section of Site Explorer to see when Yahoo! has noticed the new inbound links.

Some sites try to dishonestly create inbound links by blogging about and linking to blog posts on high-ranking websites in order to create a trackback. As explained in Chapter 5, "Building a Findable WordPress Blog," a trackback is an automatically generated post excerpt that will be displayed under a blog post when another blog links to that post. The trackback will usually include a link to the site that generated it.

Some bloggers may write meaningless or unrelated posts with links to your site so they can build their inbound links and search rankings dishonestly. If you see inbound links to your site that look like spam when browsing Site Explorer, you can report them to Yahoo! by hovering over the record and clicking the "Report Spam" button.

There's a lot to be learned about your site from Site Explorer. It's a good idea to monitor the pages indexed and inbound links to your site on a regular basis to ensure your site is visible to your audience via Yahoo! search.