

## **Arachnophilia, the Joy of Playing with Spiders**

By **Jim Hedger**, StepForth News Editor, [StepForth Placement Inc.](#)

Spiders make great geek pets, at least virtual ones do. Here at StepForth, we keep a couple spiders on our system to test sites, pages and documents in the hopes of learning more about the behaviours of common search engine spiders such as GoogleBot, Yahoo's Slurp and MSNBot. Recently, we learned that virtual pets share a similar problem with live pets; they grow old and eventually die. While our mock-spiders are still very much alive, the information we glean from their behaviours is increasingly irrelevant to predicting how a spider from a major search engine will behave. Our pet-spiders have grown too old to shower us with the informative affection they once did.

It used to be easy to predict the behaviour of common search engine spiders. Today, predicting search spiders is not so easy and with a growing number of spiders and search databases to consider, trying to get a leg-up on where the spiders are going is rather tricky. In previous years, Google, Inktomi and other electronic 'bots could be relied on to visit a site on a regular basis. The working environment was a bit simpler a few years ago, easily summed up with nine letters, G-O-O-G-L-E-B-O-T. GoogleBot was at one time the only important search spider around. While others existed, even as recently as two years ago, Google fed search results to most of its competitors.

Visiting on a somewhat regular monthly schedule, Googlebot would compile information on all the documents in its database, a process that took about one week and then rearrange their listings during the eagerly anticipated GoogleDance. Search engine optimization firms were often able to anticipate the unscheduled start dates of the GoogleDance by examining spidering activities in their weblogs and noting PageRank and back-link updates that generally preceded a shift in Google's rankings. When the shift actually happened, changes stemming from it were fairly significant as many of the search results would be altered based on new data found during the monthly spider-cycle.

What a difference a couple of years can make. Today there are four major general search engines and several vertical search tools, each with a unique algorithm and spidering schedule. So just how important is it to know the spidering schedule of the various search engines?

In previous years, most SEOs would say it was extremely important to know when a spider was going to visit a client's site. SEOs worked with fairly fixed deadlines, hoping to have clients' optimized content uploaded about a week before the expected GoogleDance began. Even then one was not entirely sure that the date they predicted for the Dance was correct but with a somewhat regular spider/update cycle, SEOs had fixed windows of opportunity with subsequent weeks to tweak and rework content if rankings didn't materialize during the last update.

Today's spiders have become almost intuitive and it is less important to know when a spider will visit as it is to know where a spider will visit. Most spiders visit an active website very frequently. According to three months worth of stats compiled by Click Tracks, spiders from Ask Jeeves visits at least once a day while MSN and Yahoo spider the index page of the StepForth site several times a day. Google only visits our index page, every four days on average. Compared to previous years, even the least frequent visitor, GoogleBot is gobbling up content. With daily or even weekly visits, the increased number of visits gives

SEOs a much faster turn around time from completing optimization on a site to seeing results in the Search Engine Results pages.

A major shift in the way search engines think about content is seen in where spiders will visit, the frequency of visits, and what drives them there. Previously, search engine spiders would consider a domain or URL as the top level source of information. It would go to the index page and spider its way through the site from that point. That is no longer the case as search engine spiders are now better able to contextualize content found on unique documents within a domain and schedule spider frequencies accordingly. For example, on a site dedicated to the sale of Widgets, the document that refers to the highly popular Blue Widgets will see more spider traffic than a document referring to the less popular Red Widgets. Similarly, a document that changes regularly will see more visits as the search engines tend to know when changes are made on documents in their database. In other words, search engine spiders tend to know your website as a collection of unique documents contained under a single URL or domain, as opposed to a collection of topically themed documents under a single URL or domain. Based on the number of searches for relevant keywords performed by search engine users, the number of incoming links, the frequency of change, and the frequency of live-human visits to a document, the 4 major search spiders are now setting their own schedules.

While the timing of spider visits has changed radically, many standard behaviours remain the same. Spiders still travel where links, both internal and external, take them. The difference today is those links often lead to internal pages. In previous years, most links lead to the index or home page of a site. With the advent of PPC programs such as AdWords and Yahoo Search Marketing, webmasters and search engine marketers are creating product specific landing pages, each of which might be relevant to organic searches. This has allowed savvy SEOs to optimize landing pages for organic rankings as well as PPC conversions. Search engine results now tend to be more relevant to the specifics of any given topic as opposed to a general overview of that topic.

Of all the spiders, the most active by far is MSNBot. Visiting each document in its index at least once per day and often more frequently, MSNBot has been known to crash servers housing sites with dynamically generated content as the 'bot sometimes doesn't know when to quit. After MSNBot, Ask Jeeves and Yahoo are the busiest of the major bots. Oddly enough, the quietest is GoogleBot, which visits each document in our site at least once per month but with little or no discernable pattern.

In order to prompt spiders through the site, we suggest creating a basic, text based sitemap appended to the back of your website. The sitemap should list every document in your website. To jazz it up, add a short description of the content of the document linked to below the link. Add a link to the sitemap to the footer of each page in your site. That will help with Ask, MSN and Yahoo. For Google, a slightly more complex solution is available through the creation of an [XML based sitemap](#).

About two weeks after implementing the HTML sitemap on your site and uploading your XML sitemap to Google, start to watch your server logs for increased spider visits. Be sure to watch for where the spiders are going and which documents receive the most frequent visits. You may be pleasantly surprised at how friendly modern spiders can be.