

This Free E-Book is brought to you by [Natural-Aging.com](http://Natural-Aging.com).

**[100% Effective Natural Hormone Treatment](#)  
Menopause, Andropause And Other Hormone Imbalances  
Impair Healthy Healing In People Over The Age Of 30!**

## Using Spider Traps to Discourage Site Scraping

By Rob Sullivan

Sometimes your competitors will do almost anything to compete with you including stealing your content.

To do this they sometimes employ automated software much like a search engine crawler to make the process quicker and easier than manually copying your site. This can cause many problems for you.

In this article we look at ways to stop this from happening.

Stealing on the web is rampant. I don't mean stealing people's user id's and passwords. I mean the stealing that goes on to a website.

Webmasters and designers steal images they like, or find a cool JavaScript they like so they steal that as well.

But what really causes problems is when your competitors steal your content.

As we all know, content is king on the web. Whoever has the most content wins. So if a competitor of yours needs to grow quickly, one of the easiest ways to do it is through the use of a website harvester.

A website harvester is no different than any other search engine crawler. It goes and requests all the URLs it can find and then proceeds to download all the content associated with those URLs.

So how does one protect themselves from malicious scrapers?

Simple really. You build a spider trap.

As the name implies, you create a section of your site devoted to luring in the spiders that are not friendly, and then you proceed to either trap them or ban them from accessing your site.

What's involved in making a spider trap?

## Using Spider Traps to Discourage Site Scraping

Usually a bit of PHP code combined with a database and a URL rewriter.

The first thing you need to do is create the space on the site dedicated to capturing those bad bots. You then use robots.txt to exclude that section from crawling.

You do this because you want to ensure Googlebot, Yahoo! Slurp, MSNbot and the others don't also get trapped. Since most good spiders will follow the robots.txt exclusion protocol you are going to politely deny them access to this location.

From here there are various options. One of my favorite involves logging to a database or text file and then dynamically denying access to the bad bot.

How does it work?

Let me give you a practical example.

I once had a client that was getting harvested many times per day by many different bad spiders. It was so bad at one point that the bad bots were doubling his bandwidth usage.

So we devised a plan whereby we'd create this trap as mentioned above and when we captured the user agent and IP info we immediately banned them from the site.

This is how it worked:

The bad bot would come to the site and find a link on an image. The link would point to the trap directory.

Normally, a regular spider would first check the robots.txt file to ensure they could in fact index the content in that directory. Since the file excluded this directory, the "good" spiders wouldn't go in.

However the bad spiders ignored robots.txt and went into the directory.

From here there was a PHP script which would run and capture the user agent and IP address.

Another script would take that information and rewrite the .htaccess file with the bad spider information as soon as it was received.

Then the .htaccess file was reloaded by the server and that spider was then not allowed to visit the site anymore.

In another incarnation this system would right to a database or text file that is then referenced by the site pages through a small php script that would allow or deny access based on that list.

Keep in mind that this is very advanced stuff. You don't want to take it on lightly. Doing so can (and likely will) get your site removed from any search indexes.

## Using Spider Traps to Discourage Site Scraping

Not because you are doing something you shouldn't but because there's always the chance that somehow a good spider like Googlebot would end up on your blacklist.

Therefore, before you get into such advanced things, I'd make sure you are intimately familiar with what they are and how they work.

A good place to start is to read this page. I wouldn't advise using this code just yet, but take a look at it to see what it can do.

Also, do some searches on the engines for "bot trap" and "spider trap" to see what other options there are out there. Then, pick the one that works best for you.

In the end, the best bot trap is the one that does what it is supposed to - block harvesters from scraping your site while allowing legitimate search engines to effectively and efficiently index your site.

And if you are at all concerned with this tactic, don't use it. It's better to use the manual approach - scour your server logs looking for high activity from unknown user agents then manually ban them using .htaccess.

Rob Sullivan – SEO Specialist and Internet Marketing Consultant. Reproduction of this article is allowed with an html link pointing to

<http://www.textlinkbrokers.com>

## **How to Increase Web Site Visitors For Relevant Terms Using Your Less Relevant Listings**

**By Tinu AbayomiPaul**

Completely by accident and through no effort of my own, besides the effort to get my site spidered by Yahoo's Slurp Search Engine Spider and Google's Googlebot Search Engine spider, I've seen an accidental increase in the visitors to my web site.

Suddenly, I was ranking first page for Google, Yahoo and AOL for the phrases like "picture of William Hung", and "william hung she bangs".

The Good news is, this means that when the Yahoo Slurp spider came to the site and indexed my front page last Saturday, (and no, I didn't submit my site OR pay for inclusion), my results were in the search engine by Sunday apparently, which is when I started getting clicks from Yahoo for this unusual term.

The Bad News is, William Hung, except for being an example of the power of Free Marketing, has NOTHING to do with my site!

So how does one fix this situation? How do you turn an accidental good ranking into several good

## Using Spider Traps to Discourage Site Scraping

rankings for other terms? Or say you've got great rankings for a relevant, but fairly insignificant term? The solution is the same.

Just give the spider more of what it wants.

Google, Yahoo – and apparently their search partners – love that page. So, first thing I did was improve the page in question. At the time there was no picture of William Hung on the page that was ranking so well. It's important that you satisfy web site visitors who accidentally happen up on your site.

It only happens once in a while, but sometimes the person who come by is also interested in your regular content.

The next thing I did, was create other pages I believed the two sites would like. Since the articles here primarily offer resources to webmasters who want to learn how to increase the number of visitors to their sites, all I had to do was tweak some of the wording to the liking of both those visitors and the search engine spiders, and write one more article.

How did I know what the search engines would like, and how that would increase the number of people who would come to the site? Well, I found out that some guarded information about the latest optimization tips for Yahoo and Google was available on the Web FREE. (If you want to know what that resource is, just download the updated Free Google Optimization Guide.)

Once I had the pages optimized to my satisfaction, all I had to do was link some of those pages to the front of my site (where the article was then located) and wait for the increase in traffic to my site. Google spiders me every day like clockwork ever since I made this tiny change to my website that

increased my site rank from zero to five in January.

Of course, I then had to figure out if my secret technique to baiting the Google spider was going to work with Yahoo's new search engine spider.

Not only did it work, but it worked faster than leading the Google spider to specific page of my site! Yahoo's spider is not back at my site on a daily basis yet, but it does come regularly, finding new pages, and reindexing the area with the most content.

So there you have it. If your site has been indexed well for a term you didn't really want, first, make the page as relevant as possible to the visitors who end up at your site.

Next, give the search engine spider "food" to eat that is more relevant to your site. It doesn't hurt to optimize this page. You can now download the new optimization guide for FREE here– IF you're one of the next 100 subscribers. Only 1000 copies will be available for download so get yours now.

Third and last, get the spider to come on back to your site. If you know how to bait a search engine spider, great! You'll get your results faster, and since the spider finds the link on its' own, you may get better placement. Or, you can submit your site to one of Inktomi's partners, and/or Google and wait the

6–8 weeks it takes to be included.

Copyright 2004 Tinu AbayomiPaul

Tinu is a free traffic FANATIC> She's been offered treatment for this condition, but says it just saves too much money. Her parting words "If you don't know how to bait the Google spider, subscribe to my newsletter. I give that tip away free in the second issue. Email [ftdsecrets-subscribe@topica.com](mailto:ftdsecrets-subscribe@topica.com) or join at

<http://www.freetrafficdirectory.com>

, where you can also read more articles about getting quality

visitors for free."

How to Increase Web Site Visitors For Relevant Terms Using Your Less Relevant Listings

How To Select Scraped//Distressed Hardwood Flooring

Submit Your Site To Yahoo! For Free!

Big Site? Make the most of it on Google

Blog And Ping Does It Work?

Email Spider Software

Fax Reaper Pro Software

Web Biz for Beginners

Scams Exposed

Google Site Map Maker



This Free E-Book has been brought to you by [Natural-Aging.com](http://Natural-Aging.com).

**[100% Effective Natural Hormone Treatment](#)**  
**Menopause, Andropause And Other Hormone Imbalances**  
**Impair Healthy Healing In People Over The Age Of 30!**