

Sitemaps

The Missing Link of Search Engine Optimization

By

Dr. A J Williams
ezSEO

ez-search-engine-optimization.com

Overview of sitemaps

A site map is a web page (or multiple pages) that lists all of the pages on your web site – or at least the pages that you want the search engines to find and include in their index.

Without a doubt, it is one of the most under-rated weapons in your search engine optimization toolbox.

Before we explore the value and use of sitemaps, let us just consider how new web pages get into the search engines. We will place emphasis on Google, being the most important search engine for free traffic.

There are two main ways:

1. Submitting new pages to the search engines.

This means that you visit a page that the search engines have provided, and fill in information about your new page. When you click the submit button, that information is stored in a database, and eventually the search engine will send out a spider to examine the pages. It is probably the case that pages that are submitted in this way will not rank as well as pages that the search engines find for themselves.

Google has its “Add URL” page at:
<http://www.google.com/addurl.html>

Altavista has theirs at:
<http://addurl.altavista.com/addurl/new>

Some search engines do not have a submit page, or at least charge money for submitting pages. An example is Inktomi, which does not actually have a search engine itself. Inktomi does have a huge database of web sites that it has collected. It then rents or sells the website database to other search engines. At the time of writing this document, MSN uses Inktomi search results in its engine. I doubt that this will last for too much longer as MSN is now sending out its own spiders around the web. Are they going to provide their own results? Probably.

2. Letting the search engines find the pages themselves (preferred method)

Search engine spiders are software programs that scour the Internet looking for new and changed web pages. These search engine spiders will follow links on every web page they encounter in their unending quest for more data. When a new page is found (or changes are detected in known web pages), these busy creatures take notes and eventually return to the mother ship. These spiders will then unload their new data into the search engine database. Eventually these pages should end up in the search engine index, and be found in the search engine results pages.

Suppose you create a new web page at:
<http://mydomain.com/findme.html>

How is Google going to find that page?

The answer is by one of the methods above. Either you submit it, or you let Google find it by itself. I do not recommend submitting any pages manually. Let Google find them itself.

Now from what we discussed above, you know that for Google to find the page, there must be another page that links to it.

Consider this:

Your homepage is:

<http://mydomain.com/index.html>

On your homepage you place a link to your new page at:

<http://mydomain.com/findme.html>

Now, when the spider from Google visits your homepage, it will find your new page by following the link. It can then take back the information about the new page and include it in the index.

But. How is Google going to find out about your homepage?

The answer, of course, is that you need to either submit the homepage itself, or have a link pointing to it from another page so the search engine spiders can find it. Vicious circle isn't it?

The very best way to get a new site found and included in the search engine, is to link to it from a web page that Google ALREADY KNOWS ABOUT. Google sends out its spiders every month looking for new pages. These spiders will start at pages they already know about and look for new links to new pages.

Suppose you have a web site with 100 pages. You have just managed to get your site listed in Google, but for some reason Google has only found 10 pages. What next?
It often takes time for Google to find all of your pages. The process of discovering your pages is related to how deep into your site the spiders have to go, in order to find all of your pages.

Example 1. Consider this:

Page 1 links to page 2
Page 2 links to page 3
Page 3 links to page 4
Page 4 links to page 5
Page 5 links to page 6
Page 6 links to page 7
Page 7 links to page 8

For the spider to be able to find all 8 pages, it would need to follow the links on 7 different pages – this site is 8 levels deep.
A spider may only go 2 levels deep on any one visit to your site, but remember, it can do so starting from any page that it knows about.

So on the:

1st visit to your site, the spider starts on page 1 and finds page 2 & then page 3

2nd visit, the spider can start on page 3 and finds page 4 & page 5

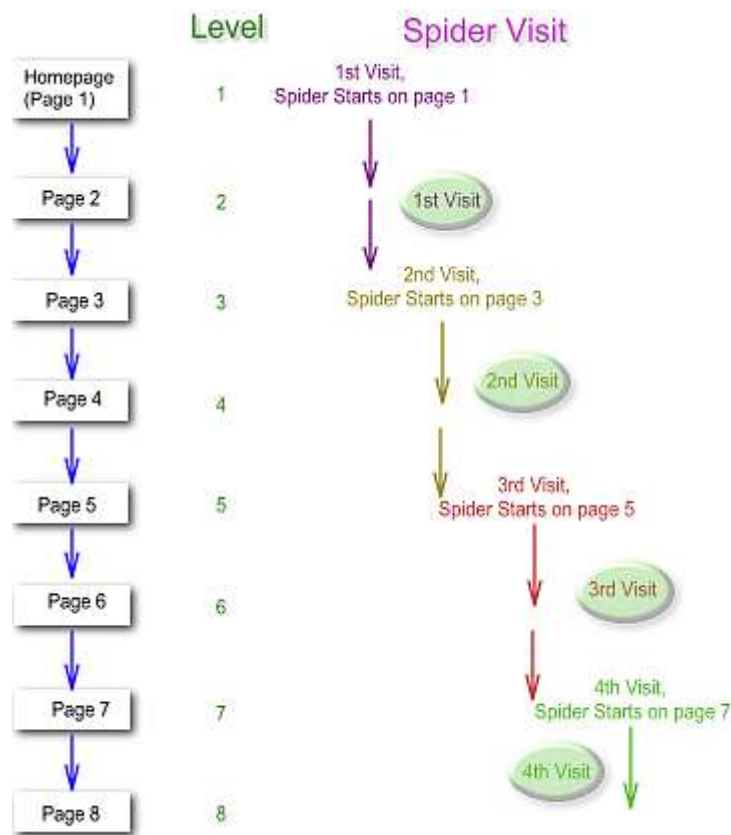
3rd visit, the spider can start on page 5 and finds page 6 & page 7

4th visit, the spider can start on page 7 and finds page 8.

See diagram below.

If the spider only visits your site once a month, then it will take 4 months to collect information about all 8 pages of your site.

Figure 1. An 8 level deep site



Example 2 -

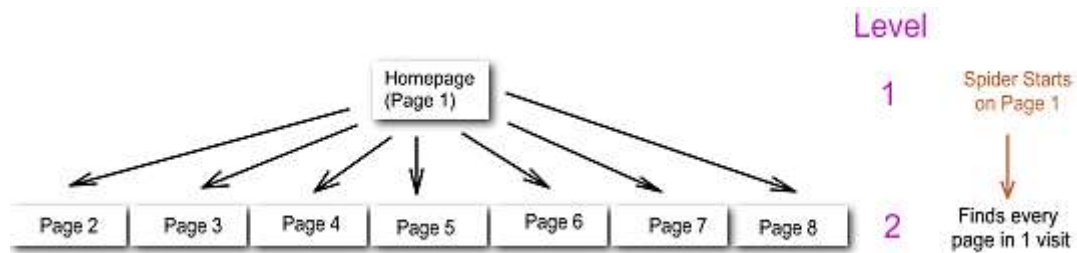
Now consider this site:

- Page 1 links to page 2
- Page 1 links to page 3
- Page 1 links to page 4
- Page 1 links to page 5
- Page 1 links to page 6
- Page 1 links to page 7
- Page 1 links to page 8

This site has links to all pages of the site on page 1. The spider comes along and can follow all links on page 1. The spider can collect information about every page in your site in just one visit.

Which would you prefer? 4 months to get all of your pages included in a search engines database, or 1 month. OK, that is a rhetorical question.

Figure 2 – A 2 level deep web site



So, is it best to link to all of your sites pages from the homepage? The answer is a little more complicated than just saying yes or no. It depends on other issues like Google Page Rank (PR) and how you want to spread PR around your site. These issues are beyond the scope of this article, but see Appendix 1 on site linking strategies for a brief explanation.

In terms of getting your site found, and included in a search engine, it is definitely better to have links to all of your pages from a homepage, or no more than 2 or 3 levels deep.

Can you see a problem with this?

As your site grows, you will have more and more links on the homepage. Eventually there will be no room for content on the homepage. Is this what you want? I have web sites in excess of 3000 pages. I definitely cannot include links to all of these pages on a home page.

Is there a Solution?

Of course. The solution is to use a site map. The sitemap will be a page (or collection of pages all linked together) that includes links to every other page on your site. From the sitemap, the search engine spiders can find every single page of your site. There is no need to include all of these links on your homepage.

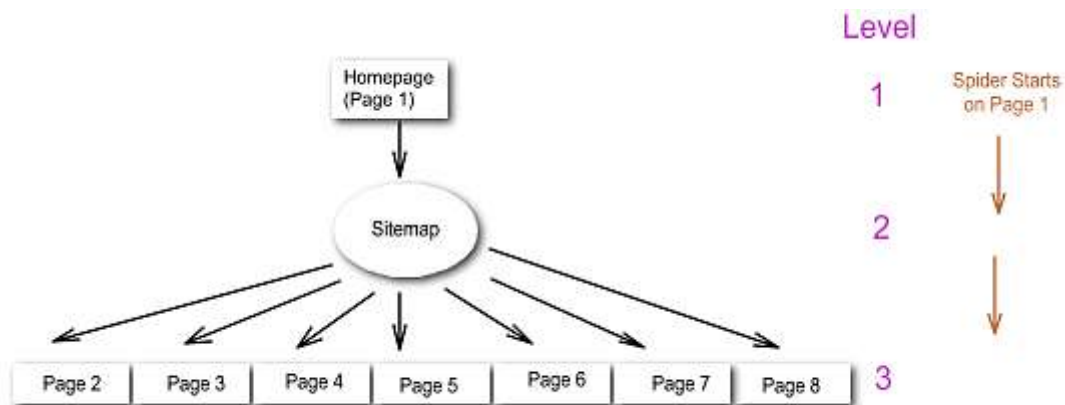
You just link to the sitemap from your homepage. That means one single link on the homepage, leading to a level 2 pages that has links to all of your pages.

Consider this scenario
Page 1 links to sitemap page
Sitemap page links to page 2
Sitemap page links to page 3
Sitemap page links to page 4
Sitemap page links to page 5
Sitemap page links to page 6
Sitemap page links to page 7
Sitemap page links to page 8

This site is 3 levels deep. So using our earlier examples of a spider visit going through 2 levels each visit, how long will it take these 8 pages to be found?

1st visit the spider starts on the homepage and finds the sitemap page and then the pages linked from the sitemap page. In other words, it finds all 8 pages on the first visit.

Figure 3 – Site using a sitemap



So what have we accomplished by using this model where the homepage links to a sitemap, and sitemap link to our pages?

Hopefully the answer is fairly obvious. We get our site spidered quickly, while maintaining a clean homepage. There are in fact other PR related benefits (see appendix 1).

What about big sites?

Now, suppose that you have a site with 1000 pages. Does that mean you need a sitemap page with 1000 links on it?

The answer to this one has to be NO. A spider will only read a certain number of links on a page, probably governed largely by the length of the page. It has been suggested on forums that the maximum number of links that will be spidered on a page will be around the 100 mark. So where do we go with our idea of a sitemap?

My solution is to have a mini-site sitemap. A group of sitemap pages all linked together. The homepage will link to the first sitemap page, which links to every other sitemap page.

For a 1000 page site, you could have 10 sitemap pages, each with links to 100 site pages. I personally go for a smaller number of links per page, say 20 pages of 50 links each, but that is personal preference.

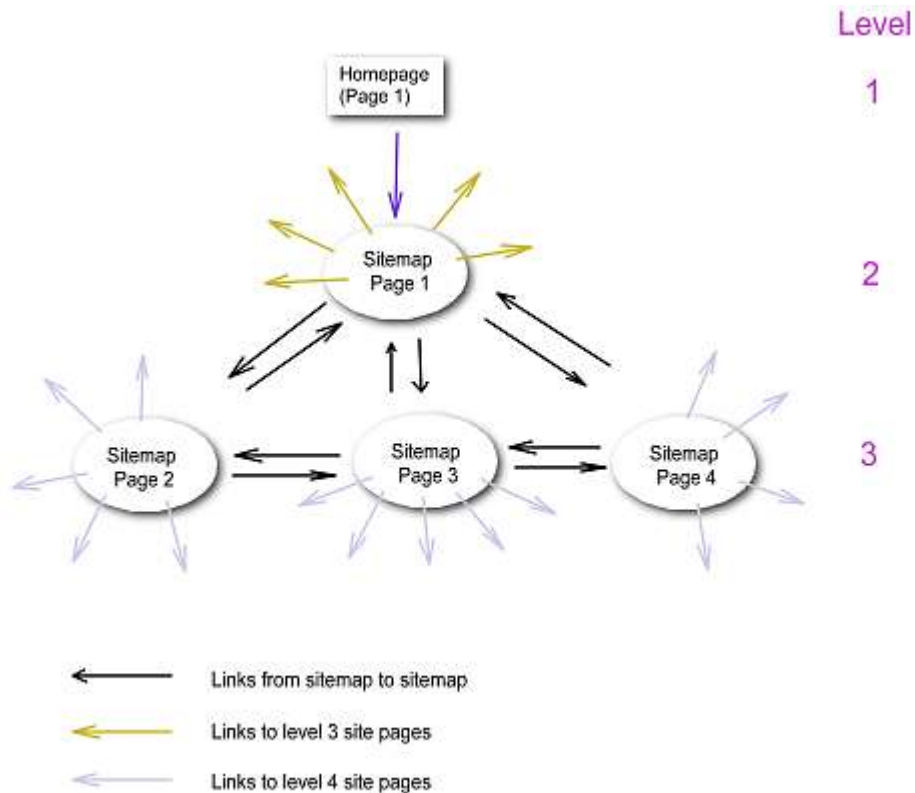
Each of my 20 site map pages will have links to 50 pages on my site PLUS, links to the other 19 sitemap pages. That way, every sitemap page is linked to every other. For a 1000 page site map, using this strategy, how long would it take to get all pages found by the spider?

Example of a mini-site sitemap

The homepage links to sitemap page 1

Sitemap page 1 is linked to all other sitemap pages as well as to a number of normal site pages.

Figure 4 –A Site Map Mini-site



Can you see what the spider will do?:

On the 1st visit, the spider starts on the homepage and finds sitemap page 1. The spider will then find every page that sitemap page 1 links to. I.E. The spider will find all other site map pages, plus all standard site pages that are linked to from site map page 1.

On the 2nd visit (month 2) the spider can start on each and every sitemap page, as it found them all during its last visit. Remember the spider can start on any page it knows about, and will visit your site several times during a month.

From the sitemap pages, the spider can find all other pages of the site by following the page links.

So, for even large 1000+ page web sites, the spider takes only 2 months to collect information about every page. That is not bad.

Submitting the sitemap to the search engines

Something that can speed up the spidering of a new site that is not yet in the search engines is if you submit the sitemap page 1 to the search engine manually. Then the spider can start on sitemap page 1 and follow links to every single page in the same visit (the site is 2 levels deep from sitemap page 1).

Alternatively, if you have a page already in Google – perhaps on a different site, just link to the sitemap from it. This can get you included very quickly.

1000 page site map? You have to be kidding, it will take forever to build!

Unfortunately, sitemaps do take time to build. On a 1000 page site, this could take many hours or even days to do properly. Even relatively small sitemaps can take several hours to build and check. There is, however, a software solution that can cut those hours & days into minutes. Before we look at it, we need to consider what makes a good sitemap page.

Anatomy of a good sitemap page

A sitemap page is a list of links to other pages on your site. If you just provide a list of links, the page will look very skeletal and of little value to your visitors. Yes, a spider can find your pages, but why not design a sitemap page that adds value to your site for the visitors too? With that in mind, each link should probably have a description of what the target page is about for each link. This will ensure that human visitors can use your sitemap as a navigation system.

BONUS: The added bonus of a description for each link is that it gives the search engine spiders more keyword rich text so that sitemap pages themselves can get good ranking in the search engines, not just your target pages.

Look at these two example links. Which would you prefer your site map to have?

a) Link Type I – with description

[Collectible Barbie dolls](#)

Collector's edition Barbie dolls, Barbie books & early edition Ken dolls. Buy collectible Barbie memorabilia online at great prices.

[Barbie Doll Collectible Books](#)

Buy early first edition Barbie Doll books. Barbie doll collector's guides, magazines & Barbie memorabilia.

b) Link Type II – no description

[Collectible Barbie dolls](#)

[Barbie Doll Collectible Books](#)

OK, which would you rather have?

From a human visitors perspective, Type I is obviously far superior. It looks more like a search engine listing style that people have become comfortable with, and could be easily used to navigate your site.

Most humans would not be able to, or want to use a Type II site map.

The fact that both the search engine spider and the human visitor can use your sitemap, make Type I site maps, the site map style of choice for the professional webmaster.

But there is more..

The added bonus of Type I is that it can be used to add even more value to your site for the visitor, and also increase your search engine rankings. Let me explain both of these points.

Themed sitemap pages - added value for the visitor

If you build your sitemap pages by only putting links to closely related pages on each sitemap page, you make what I call a themed sitemap page. This gives your visitor real added value. Your visitor will have a list of related pages of a topic they are interested in.

Looking at the last example:

[Collectible Barbie dolls](#)

Collector's edition Barbie dolls, Barbie books & early edition Ken dolls. Buy collectible Barbie memorabilia online at great prices.

[Barbie Doll Collectible Books](#)

Buy early first edition Barbie Doll books. Barbie doll collector's guides, magazines & Barbie memorabilia.

Suppose your site has 10 pages all related to collecting Barbie dolls. If you included links to all 10 pages on the same site map page, you create a Barbie Doll themed sitemap page.

This gives your visitors an excellent navigation system to all pages that they might be interested in. Suppose your visitor found this page by searching Google for "barbie collectibles". How useful do you think this page would be to your visitor?

Themed sitemap pages – increase your search engine rankings

Imagine the same Barbie Doll Themed sitemap page we discussed above. By its very nature, it will have everything that a search engine looks for in a high-ranking web page. It has keyword rich text. It has keyword rich links to highly related pages. All you need to do to make this great spider food is add title tag, meta description, a header and some introductory text and bingo – search engines will love the page and rank it well for a range of keywords.

So, a sitemap can be more than a way for you to tell a search engine about your pages. It can be used as a navigation system by visitors, as well as spider food for the hungry spiders. This can get your sitemap pages great rankings in the search engines.

Building Site Maps

There is nothing to stop you from building a sitemap by hand. Especially if your site is quite small.

1. Create a web page for your sitemap.
2. Add keyword rich links, and keyword rich descriptions for each web page you want to link to.
3. Upload it to your server.
4. Add a link to your sitemap from your homepage.
5. As you add more pages, remember to add these pages to the sitemap.
6. As the sitemap gets bigger, you will need to split the sitemap into separate pages.
7. Consider themeing your sitemap pages.

These sitemaps will take up more and more time as your site gets bigger and bigger. Fortunately, a software solution is available.

[Sitemap Creator](#) takes the work out of building Type I sitemaps.

Imagine entering a few details about your site, clicking a few buttons, and having a type I sitemap on your server and ready for visitors and spiders within minutes of starting the software.

Sitemap Creator has many benefits:

- It builds pages for you and uploads them to your server using the built in FTP feature.
- It automatically uses the web page titles for the link text. Lets face it, we all know that our titles should be keyword rich, so if we are designing for good positioning in the search engines anyway, what better than the web page title for the link text?.
- It automatically creates the description for each link. If a META Description tag is present, it can use that. If not, it will use the first x number (x is defined by webmaster) of characters from the visible web page for the description. You can default to using visible text instead of META description if you prefer.
- You define where your files are located, Sitemap Creator will rebuild sites for you as the site grows.
- You define how many links per page, Sitemap Creator will follow your instructions.
- You can easily select individual files to be included in a site map – ideal for themed sitemaps.
- New in the latest version – Re-order the file list before creating your sitemap. For large sitemaps (more than one page), this

allows pages with similar themes to be grouped together ensuring they appear on the same sitemap page. No easier way to build themed sitemap pages.

- It provides you with the link you need to place on your homepage
- It puts links to every other sitemap page on every sitemap page.
- Sitemap Creator uses your templates. You make sure your templates look like your site, with your logo etc, Sitemap Creator will make your sitemap pages using that template.
- Use of variable allow quick and easy insertion of details like homepage links, page numbers etc., where you want them.
- You can name your sitemap pages with the .shtml extension if you wish. That way you can use server side includes (SSI) on your pages.
- Probably the best benefit (besides producing site maps that you can be proud of, that your visitors and search engines will love), is the fact that this software will save you hours, if not days for every sitemap you create. How much is that worth to you?

Visit the [Sitemap Creator homepage](#)

Appendix 1 – Linking strategies.

This report is about sitemaps, not linking strategies. This appendix serves only as a brief introduction to linking strategy. For more information, I would recommend reading from the masters of link strategy – Michael Campbell & Leslie Rohde. You can see my review of their work at:

[Revenge of the Mininet](#)

Overview of link importance

Links to pages can be done with standard text links, or image links. While many people use image links, it is often not the best choice. Links are more than just methods of navigating from one page to another. They are also used by search engines to rank pages. An image link that has nice graphics and human appeal, does not tell the search engine spider anything about the content of the page it links to. True, “alt text” can be used for this, but search engines often don’t give this as much importance as perhaps they should.

A text link not only provides a route to another page, it also uses text to define what that page is about.

Consider this link:

[Barbie Doll Collectible Books](#)

Buy early first edition Barbie Doll books. Barbie doll collector’s guides, magazines & Barbie memorabilia.

A search engine spider can follow it and find the page about “Barbie Doll Collectible Books” and therefore index that page. However, the spider also uses the text in the link to help decide what that page is about, even before it visits. In this case the spider expects a page about “Barbie Doll Collectible Books”. If it finds that the page has title tag, headers and body text about “Barbie Doll Collectible Books”, then the fact that the text link also had those keywords in, gives that page a boost in what Leslie Rohde calls “Link Reputation” (See Revenge of the Mininet resource at the end of this report for more info).

If 100 pages all link to the Barbie Doll Collectible Books page using this same link text, Google is going to think – great, this page is about Barbie Doll Collectible Books. If someone searches Google with this as the search text, there is a good chance this page will rank highly because of a high Link Reputation for this term.

Where does Page Rank (PR) come in?

Well, I am glad you asked. PR is something Google dreamed up. Basically a page is ranked according to how important Google thinks it is. Calculation of PR is complicated, but in simple terms, the more pages that link to your page, the higher the PR of your page.

Each link to your page is a vote for your page.

But, “votes” from pages that Google deem important (high PR), will count for more than “votes” from low PR pages.

e.g. It is probably better to have one link pointing to your site from a PR 7 page, than 20 links pointing to your site from PR 1 or 2 pages. So, the quality of the links are taken into account when PR is calculated for your pages.

The size of the vote comes from the PR of the page that votes for you. If you have one link on a PR7 page to another page, the full PR 7 is passed on to the other page (not as simple as this, but it is easy to think this way for this example).

If you have 2 links on the PR 7 page, then half of the PR is passed to both pages. Each page will get a 3.5 PR vote. If you have 10 links on a page, each of the linked pages get 1/10 of the PR7 i.e. 0.7 passed to them.

How can we use this knowledge?

As discussed above, when a page A links to page B, page A is in fact voting for page B.

Do internal links in my site count as votes?

Yes. And this is where linking becomes important. Suppose you have a 100 page site about “haemorrhoid cream”. Each page should link back to the homepage so visitors can find your main site page. What text should you use in the link to your homepage from all of your pages?

What if you put this – we do see it a lot on websites:

[Homepage](#)

99 pages linking to your homepage with link text “Homepage”.

Google is going to think that your homepage is about....

“homepage”.

This is obviously a wasted opportunity. Why not use:

[haemorrhoid cream](#) homepage

You will have 99 pages linking to your homepage, all telling Google that your homepage is about haemorrhoid cream. Google thinks – “Hmmm, this page must be about “haemorrhoid cream” as there are 99 votes telling me this”. When Google visits the page, it sees that yes indeed, the page is about “haemorrhoid cream”.

These internal votes **increase the link reputation** of that page.

Taking this a step further

Using these techniques, we can obviously channel votes to important pages making them more important in the eyes of Google.

Also, by linking from high PR pages, we can boost PR and Link Reputation of important pages on our site.

Warning – some people who request a link exchange ask for their link to be on your homepage. Can you see why this is a bad idea? Also some ask for a link where their description actually has several links in it, like:

[Buy generic viagra online](#). Also we stock [natural alternatives to viagra](#), all at the cheapest prices.

This is also a bad idea. Can you see why?

Taking these ideas a step further, you can design your own network of related sites, link them together so that PR and reputation is channelled for maximum benefit.

Michael Campbell does a great job of explaining this in his book "Revenge of the Mininet". His clear diagrams make it possible for anyone to build high ranking sites. Michael has bundled a report by Leslie Rohde on Linking strategies with his mininet book. They are essential reading for anyone who wants to use the power of linking to bring free traffic. See the resources at the end of this report for more details.

Conclusions

Well, that brings this report to an end. It should be obvious that linking is very important in getting good search engine rankings. Done properly, it will provide a steady stream of free search engine traffic. Done badly – your site may not be found by searchers at the search engines. It really is that important.

Of course, your sites pages need to be spidered. Sitemap Creator can help you achieve this goal. It creates the otherwise time-consuming, search engine friendly sitemap pages, freeing you to concentrate on your linking strategy of the main site. Used on its own, Sitemap Creator can help you achieve fanatastic results. Used together with Michael Campbell's Mininet book & Leslie Rohde's Linking Strategy guide, Sitemap Creator becomes an even more powerful tool in your search engine optimization toolbox.

Resources:

[Revenge of the Mininet](#) – by Michael Campbell. This book explains how to link your pages & sites to get free traffic. Comes with a linking strategy guide by link-guru Leslie Rohde.

[Sitemap Creator](#) – build sitemaps that visitors & spiders love.