

New InSites into the FAST Search Engine . . . an interview with

Stephen Baker of Fast

By Robin Nobles

New InSites into the FAST Search Engine . . . an interview with Stephen Baker of Fast

Robin Nobles
RobinN@acws.com

Academy of Web Specialists <http://www.academywebspecialists.com>

Of the major engines in the search engine universe, FAST -- aka, alltheweb.com -- is arguably the most unique. In many ways, it operates differently than typical engines and, in light of recent developments, it's timely that we probe a little deeper into the inner workings of this strategically important engine.

To assist us with this article we invited Stephen Baker, Fast's Director of Business Development and Marketing, to lend his perspective. As you will see, Mr. Baker was quite upfront with his answers.

FAST powers other engines!

Like Inktomi, FAST powers other search engines. For example, you're probably already aware that FAST provides primary results for the Lycos search engine. But what other engines does FAST power? Mr. Baker explained,

"In North America, we power all of the Lycos and Lycos Latin properties as well as the Telus portals and AllTheWeb. Our focus over 2002 will be to establish broader reach in the US (currently 35%). 2001 was focused on European expansion, where we power over 75% of searches. This includes Lycos Europe and all of the Terra sites, T-Online, Tiscali, including Excite.it, Scandinavia Online and Web.de. Overall, we appear on over 70 portals in the European market."

...obviously an important consideration for companies doing business in Latin or European markets.

FAST also has its own engine.

Unlike Inktomi, FAST has its own public search engine. As you may know, you can't submit directly to the Inktomi engine itself. Instead you must submit through an Inktomi partner as well as pay a fee for inclusion. And, only recently did Position Technologies make available the pure Inktomi results for searching (<http://169.207.238.189/search/>) although certainly not in mainstream usage since the public is generally unaware this specific search option even exists.

FAST, however, provides the means to submit your site directly to them via [alltheweb.com](http://www.alltheweb.com) -- http://www.alltheweb.com/add_url.php -- which they use basically as a "testing" engine. AlltheWeb is where they first consider new features such as spam detection or the use of ODP categories as a training set or even linguistic analysis of non-ODP categories.

According to Mr. Baker, AlltheWeb has also begun indexing dynamic content and supports 48 languages with over 700 million full-text HTML pages in the index. Using a technique called query analysis, the new version of AlltheWeb will remember your choices and give you a more personal search experience. An interesting possibility is that query analysis could potentially supplant the importance of link popularity sometime in the future.

I asked Mr. Baker if the results at one FAST powered engine will match results at another? He replied that, like Inktomi, the partner engines powered by FAST use their own algorithms. He said,

"Our engine is entirely tunable, allowing customers to pre-establish offensive content, language, etc., settings. In addition, most of our customers blend FAST results with directories, PFP, and licensed results."

Fast's new pay inclusion program.

On February 25, FAST officially launched their paid inclusion program. However, typical of FAST, their paid inclusion program is unique.

Administered by partners such as Terra Lycos, Fast's PartnerSite program has four versions designed to benefit a variety of online businesses based on size. Partner Lycos calls the PartnerSite program Lycos InSite Select. The first level of the program provides a means by which you can pay per page and offers the following benefits:

- * Guaranteed inclusion in the Lycos index within 48 hours of submission.
- * Full refresh of your site content every 48 hours or less.
- * Automatic notification that your site has been submitted and indexed.
- * Guaranteed inclusion for a full year.

* Personalized submission reports to see exactly what has been added to the index and when.

* Access to Search Services Central, which is an online account where you can make changes and request support.

Cost?

For a "limited time" (they don't say when the offer expires), the annual membership is \$18 plus \$12 per URL. This is assuming you use Lycos InSite Select to initiate your service. Prices may vary according to reseller.

When you actually sign up for the service, you'll be asked to choose between Standard Submission where the spider begins at a starting point that you determine, and then indexes as many pages deep as you specify. Advanced Submission allows you to list all pages you specifically want indexed after which the FAST spider will continue to "discover" URLs within your site up to the total number you've specified. This option assures that your most important pages will make it into the FAST index rather than leave the indexing to chance.

Here's how Lycos InSite's spider works:

* The spider goes to a Web site and follows hyperlinks throughout the site.

* The spider parses and downloads certain pages as it moves through the site.

* The spider has reached the limit of URLs determined by the user's subscription service, it stops spidering the Web site.

* The file of spidered URLs are then built into a search catalog.

* If your account is missing pages and you use JavaScript heavily, you can submit individual pages using "Advanced Submission."

* If your pages use META refresh tags, enter the target destination pages instead.

* And, further details can be found at the web site --

<http://searchservices.lycos.com/searchservices/>

Besides the basic version of PartnerSite I (explained in the list above), FAST offers three additional versions of the program, all geared toward larger sites with different needs. Lycos calls these versions Lycos InSite Pro and within this category they have their PartnerSite II, III, and IV classifications.

Mr. Baker explained, "Recent studies indicate that 80% of shopping carts are abandoned because of poor site search." So, FAST is providing sites with a search-ability solution through its PartnerSite packages.

PartnerSite II: Designed for sites of up to 250 pages and includes a FAST-hosted onsite search engine. Cost: \$189 month.

PartnerSite III: Designed for sites up to 500 pages and also includes a FAST-hosted onsite search engine. Cost: \$279 month. Mr. Baker elaborated on PartnerSite III:

"When you register for version III, the FAST spider crawls your entire site, up to 500 pages. In addition to including those pages in the FAST index, we host an index of your site's pages and provide you with a Search My Site tool bar that you can paste into your HTML. This allows site visitors to search only the contents of your site. In the next version, we will provide more control to the subscriber over the site search rankings, but Web search rankings will always be determined by FAST."

PartnerSite IV: Designed for sites up to 500 pages and provides bulk inclusion via XML but no site search.

Baker added, "We have developed our own DTD (Document Type Definition) to integrate XML feeds. XML indexing is actually how we currently index database pages. We have the partner create an XML feed according to our DTD that has the 'page to fetch.' This essentially tells our crawler not to explicitly try to crawl the page, but rather, just 'fetch,' the page's contents."

Editor's Note: Document Type Definition defines the legal building blocks of an XML document. See the DTD School for more information: <http://www.w3schools.com/dtd/default.asp>

For the latest information on the different versions of PartnerSite, visit:

North American Customers:

<http://searchservices.lycos.com/searchservices/>

International Customers:

<http://www.lycos.de/>

Lycos is only one of Fast's pay inclusion program partners in a list that is still growing. For the complete partner list. go to:

http://www.fastsearch.com/products/internet/partnersite_partners.asp

Next subject . . .

What about free submissions to FAST? Are they a thing of the past?

According to Baker, "Free submit will continue to operate. There will be no effect on pages (already) in the index. The inclusion service is simply a guarantee that your pages will be indexed regularly and are guaranteed a reservation in the FAST index. The trick with free submit is the amount of spam we receive through that channel. It makes it difficult to keep up with all

of the submissions. PartnerSite provides an economic incentive to NOT submit spam."

How quickly does FAST index free submissions?

Mr. Baker answered, "Usually within 2-6 weeks. However, 95% of what comes through free submission is spam, and that's what causes the bottleneck. Also, there's no guarantee on the refresh rate of those sites that we pick up through free submit. Unfortunately, it's the 5% of the submitters that are spammers that ruin it for the rest of us. The extent that people go through to spam the index is truly amazing."

Is there a penalty for submitting your site through their free Add URL?

Baker replied, "Not really. Freshness does effect rank, but only marginally. I have never seen freshness boost ranking more than 1 or 2 spaces."

What does FAST consider spam? As stated earlier, FAST estimates that an amazing 95% of the submissions through their free Add URL page are spam.

Baker further explained, "Unfortunately this is the case. We believe there are approximately 30 million crawlable servers globally, two-thirds of which have been blacklisted as spam servers."

Whew! Think about it: 20 MILLION crawlable servers globally are blacklisted as spam servers. If this is really true, it explains why the engines have collectively gone to such great lengths in their efforts to curtail spam.

At the Dallas Search Engine Strategies Conference in November, FAST introduced their new spam policies. Baker explained that according to FAST, spam comes in three different categories:

- * Page Spam, which consists of any measures to boost ranking, such as link farms, etc.
- * Spam stuffing, such as keyword stuffing, invisible text, etc.
- * Offensive content, which is not so much spam, but is something that we detect and flag as such.

You can access and read Fast's entire spam policy at:

<http://www.alltheweb.com/info/spampolicy.html>

And, if you're so inclined, you can report spam to FAST by emailing spam@fastsearch.com.

The ultimate fear of a search engine marketer: your site has been banned. What recourse is there for a site that gets on Fast's spam list and becomes banished from the engine? Is this forever?

Baker replied, "This really is handled on a case by case basis. We have worked with sites that have contacted us and informed us that they have cleaned up their act. Obviously, this is very time consuming, so the combination of PartnerSite, the spam guidelines, and not having your site hosted on a banned server should enable a site to get in the index through a variety of means if they think they have cleaned up their act."

Translated: If your site is hosted on one of the 20 million servers tagged for spamming, you'd best move it to a different server before you initiate the kiss-and-makeup ritual with FAST.

Baker further explained, "What really matters is the server where the submission is coming from. So many servers have been completely blacklisted due to the proliferation of spam. I suggest 'know thy neighbor.' The crawler will take care of the rest."

This again underscores the importance of your site having its own unique IP address to insure against problems caused by an unruly site sharing the same IP.

How does FAST feel about cloaking? In Dallas, Mr. Baker and I had a long discussion about cloaking and how the FAST engine feels about it. We even served on a panel where a question came up about cloaking and about responsible cloaking guidelines.

As I mentioned to him, from our position as SEO's, we see the issues from opposite perspectives. Of course, we don't have a front row seat to the parade of spam but, even so, our viewpoint is dictated by that of legitimate businesses trying to ethically compete in arenas that are often very competitive and sometimes dominated by nefarious position jockeying. Therefore, on the issue of cloaking, I pointed out instances where it's the only tool available that prevents our work from being stolen. My thinking is there's no harm done provided that a person follows all of the engine's guidelines and does not spam in any way.

Baker's response: "Unfortunately, the 20% of sites that use redirects maliciously ruin it for the rest. Redirects and cloaks have become such a hassle that we can't afford to risk indexing them. We do work through PartnerSite IV customers to index cloaks. However, they are sent through a rigorous spam-detection process."

When I asked if we could show we're not trying to hide anything - perhaps form a partnership aimed at quality control within the framework of cloaking? His reply, "I agree with that. That is the best option."

The rest of the our dialogue went like this...

Robin: Does it have to be PartnerSite IV -- some small companies couldn't quite afford that option.

Baker: Currently, it would be through PartnerSite IV. However, we do plan to extend that offering to the other PartnerSite versions in the future.

Robin: Since redirects are frowned upon, what should you do if you've moved your entire site, for example, and you need to direct traffic to the new domain?

Baker: It is always best to use META REFRESH tags rather than JavaScript to implement redirects if you want search engines to know about it. Most search engines do not fully parse JavaScripts to find out what they actually do. Using a META REFRESH tag tells us in a much better way what you are trying to achieve so that we can do whatever we think is most sensible with the page.

How to score at the top of FAST . . .

Now let's get down to the nitty gritty. What does it take to achieve a top ranked page with FAST?

Here are some tips from the FAST rep himself: Baker suggests,

"Good content that is unique, not general, always has the 'authoritative' effect and people begin to link to that site. This achieves good scores in the two areas we care most about:

- 1) Static Rank, which is link analysis, and
- 2) Dynamic Rank, which is keywords and content."

Also . . .

"Get a handful of authoritative sites linking to you. Link popularity plays a large factor in determining rank. If your site is diverse, with multiple focuses, make sure the pages are well written, keywords are appropriately placed, and have a handful of authoritative sites linking to the site."

Baker added, "The trick, in my experience, is to get a handful of really good sites to link to a site that is non-thematic."

How important is link popularity with FAST? As with most engines, link popularity is an extremely important factor in determining relevancy.

Baker explains as follows, "Be concerned with links in, not links out. Link score is part of the relevancy calculation and has a formulaic impact on ranking. However, if a site has no link score, ranking is negatively impacted."

I asked him, "Is link popularity is based on the entire site, or

on individual pages? In other words, if one page has a high link popularity, will all the pages of a site be boosted, or only that one page?"

His reply, "Just that one page."

So, if your site has no links pointing to it, you'll likely rank poorly but the site can still be indexed provided the FAST spider can find it.

Robin: Does FAST make allowances for brand new sites?

Baker: "Not currently, but we are beginning to work with some of the new domains that are popping up."

Robin: Do links from pages on your own site count toward building link popularity?

Baker: "No . . . that would be too easy." ...he said with a smile.

Robin: How does FAST deal with asp and database-generated pages?

Baker: "As long as they are not forms and they are linked to, we can crawl them. Of course, the PartnerSite service has the ability to index dynamic pages as well."

Conclusive remarks . . .

As with most engines FAST sees spam as public enemy number one. They've identified two-thirds of the servers in the world as spam servers. The best way to gain entrance into the FAST index is to make certain that your site does not fall into any of their spam categories.

We suggest you take a few minutes to review their spam policy.

For guaranteed indexing and respidering every 48 hours, give their paid inclusion a test. The primary benefit being that you can experiment with keyword factors that boost relevancy and see the results of your tests in 48 hours or less.

Remember, paid inclusion guarantees your space in the directory and provides you the opportunity to experiment endlessly with relevancy factors within your page for a whole year.

Focus your attention on gaining high quality incoming links. Doing so will increase what FAST calls your "link score" (aka, link popularity). Remember also that sites with no incoming links are actually negatively impacted.

All other factors being equal, the freshest (most recent) page wins by an ever-so-slight boost in relevancy. There are many ways

to maintain site "freshness," such as using server side includes.

For shades of things to come, keep an eye on alltheweb.com -- we'll be watching to see how query analysis develops as a determiner of page relevancy in the near future.

Robin Nobles, Director of Training, Academy of Web Specialists, has trained several thousand people in her online search engine marketing (<http://www.academywebspecialists.com>) training courses. Visit the Academy's training site to learn more about their online search engine marketing training programs (<http://www.onlinewebtraining.com>) and search engine optimization software (<http://www.se-optimizer.com>). She also teaches 3-day hands-on search engine marketing workshops in locations across the globe with Search Engine Workshops (<http://www.searchengineworkshops.com>).

Copyright 2002 Robin Nobles. All rights reserved.

[Get-Articles.com : 1000's of reprintable business and internet marketing-related articles.](#)

[Submit your article for reprint.](#)