

Confessions of a Google Tracker – Increase Web Site Traffic

Summary: *Here you can find a series of concise information about Google Search Results, Google Dance and Googlebots...*

Google is one of the most important search engines to be listed under right now. It accounts for more than 90% of visitors for most of the sites. So it is mandatory for a web master or a search engine optimization expert to have full information of the way Google works, so as to stay one up on the search engine regularly.

So here you would find updated information about various little, but non-ignorable, aspects of Google. You can use this information free of cost to try out various strategies in your Google optimization process. I do not claim that all this information is discovered by me, so please don't cry foul, if you find the same information elsewhere on the web! I also don't claim the entire authenticity and periodicity of this information, so please bear with me while I unleash the secret ways of Google onto you....!

Google Dance GoogleBot, Google DeepBot and Google FreshBot Confessions of a Google Tracker

Google Dance

Google Dance is a word which is used to depict the time period when GoogleBot (the indexing spider of Google) visits sites across the world to re-index the content. The Google Update happens within a duration of 15 days. This Google update is known as Google Dance because Google is supposed to be dancing all over the web at this time.

Why is Google Dance so significant to monitor? Well, whenever you make any changes to your pages, it's better to make them just before a Google Dance so that within a short time, you would be able to see the effectiveness of your changes. This is very useful when you are doing some experiments to get your page listed fast, and/or under some specific search terms.

How do you track the Google Dance? Following are two good sites which you can regularly use and understand the phenomenon:

Google Dance - Check your site in all the 3 servers of Google and notice the difference. I will explain it in detail under "Google Bot, Google Deepbot and Google FreshBot" below.

Dance EFactory - Information about Google's 8 Data Centers and the order in which they received the new index for the last Google Dance.

There is also a fun page (though they don't claim it to be!) about the co-relation between Google Dance and the Phases of Moon at the moment the dance begins ! Check it out here:

<http://www.webmasterworld.com/forum3/2657.htm>

GoogleBot, Google Deepbot and Google Freshbot - The Three Musketeers of Google

Contrary to popular opinion, Google seems to have 3 main spiders for indexing content. Each main spider may have hundreds of other spiders for help, but mostly perform similar functions.

GoogleBot is the main spider which indexes the entire content of the World Wide Web (WWW) fast. Its main function with regard to Google DeepBot is to pick up new content to be indexed and give it to the DeepBot to dig in to retrieve that information. Its main function with regard to Google FreshBot, is to note changes in the content of a page and keep it in memory. If it does happen regularly, it allocates that particular page to Google FreshBot to index the page more frequently. It's one very important function is to also find how many pages link to a particular page, on what search terms ; which helps it in the calculation of the Google PageRank™

Google DeepBot is the deep rooted spider which goes into each page and indexes the content on that page. So if your regular page is picked up by Google Bot and has started being displayed in Google with the new content, but the new links added to that page are not being displayed yet; there's no need to panic. It's the work of Google DeepBot to recognize these new links on the regular page and index them to be displayed in the Google Search Results as soon as possible. Deepbot indexes almost all types of content including PDF documents, MS Office documents, Zip files apart from HTML and Text content.

Google Freshbot is a spider which is the most active of them all. It indexes a select number of pages, almost daily or at least once in 3 days or a week. FreshBot is the one which recognizes that a site is being updated regularly, once GoogleBot allots it the page; and decides to index the page regularly to let Google's visitors have the latest information at their finger tips. It follows the simple logic that if the content is being updated regularly and/or if a page is getting a huge number of visitors from Google, then obviously it's a very important page and needs to be indexed regularly. FreshBot indexes News, Blogs and Resources sites very regularly. Unlike DeepBot, Freshbot indexes only HTML and Text content.

As I had indicated above I would let you know, here is some information on the three servers of Google: <http://www.google.com> is the main server of Google (of course there are in reality hundreds of similar servers all across the world which act as mirrors to the main site) . This information is updated only when a Google Dance is finished (which takes around 15 days and happens monthly). So you wouldn't be able to find the latest information here unless it is being served by Google FreshBot.

<http://www1.google.com> and <http://www2.google.com> are the two other servers of Google, apart from the 8 data centers which index content on the WWW in a systematic way. One of the data centers indexes the new content on your page and then sends it to the main google to add to its index and display it after it finishes its entire dance. Google DeepBot provides the content for these data centers and so, you would be able to see the new changes to your page more sooner on these data centers than the main Google site.

Confessions of a Google Tracker (What happens when you hurry to conclusions)

Tracking Google and its changes in indexing methods, is like going on a wild goose chase! You do some change, expect a result and when that result seems to have appeared and you seem to have established the proof of concept, something emerges which sends your proof to the dogs. Something similar happened to me recently.

We had decided to put some pages of our sister site under login in order to allow visitors to provide any information about themselves that they didn't mind giving us. This would have helped us in many ways. This would have given us some information on who our visitors are, and help us know whether we are reaching our target audience or not. So we decided to put the following page under login:

<http://www.web-enable.com/business/thinkebusiness.asp>

We programmed the page in such a way that, when people clicked on this link, a javascript redirected them to the following login page:

<http://www.web-enable.com/member/login.asp>

This page would then ask them to login (if they have an account) or register for a new account (which is free). Only then would they get to see the page

Here it would be necessary to mention that our intention to put Java Script on the page was only to check for Login, so that if a visitor has already logged in the site, he is not prompted again to provide his access details; when he wishes to view this page.

So we were happy to see people actually registering and reading the article and we getting the information what we wanted. Only thing we were skeptical about was that Google would be not be able to index this content as it is under a login.

So imagine the pleasant shock I got when I saw that the page had actually been indexed by Google with its entire content. I was surprised as to why this happened and I thought maybe I was witnessing a new phenomenon where Google indexes password-protected content. So I became excited and put it on some Search Engine Forums talking about my discovery that maybe the Google spider is doing automated registrations to get to password protected content, if the registration is free. I was pretty excited because we could continue to attract visitors to our site through our content being indexed by Google, but let them read the content only if they register with us. This would give us an opportunity to know more about our visitors.

Many were agreeable to my discovery and pointed me to sites like Washingtonpost.com and NewYorkTimes.com whose content, though under a login, gets indexed by Google very regularly.

But my happiness was short lived. One person on the Web Master World Google Forum replied to me thus:

Reply begins here

The JavaScript of your page is as follows:

```
<script language = "JavaScript">
function abc()
{
document.frmRedirect.submit();
}
</script>
<form name="frmRedirect" method="post" action="/member/login.asp">
<input type="hidden" name="hPageName" value="/business/thinkebusiness.asp">
<script language = "JavaScript">
abc();
</script>
</form>
```

This page when opened with it's URL first returns the full content, then instantly redirects a human user to the login page. This happens too fast for the naked eye to register, but the GoogleBot notices this and that's why it has indexed your page.

You can check this yourself by disabling JavaScript and visiting the page.

Reply ends here

So there goes my new Google discovery down the drain! And Google again escapes from my clutches and I'm back to square one. I need to again start my experiments to try and tame Google for displaying my pages for my search terms. Ah, there goes my vacation!