

## **Google Goes Back to Basics: Site Maps, Relevance, and AdWords**

The information in this article will appear sometime in 2005 in Information World Review. Mr. Arnold will deliver his Google Seminar as a pre-conference event at the International Online Show in London, U.K. in November 2005. Updated June 9, 2005.

The bulk of the material in this section comes from the The Google Legacy: An Analysis of the Google Platform, to be published in August 2005 by Infonortics, Ltd. ([www.infonortics.com](http://www.infonortics.com)), Tetbury, Glou. This material is provided for reference only and may not be redistributed or republished without the written permission of the copyright holder, Stephen E. Arnold, Postal Box 320, Harrod's Creek, Kentucky USA 40027.

Consider the situation of U.S. software development company Avidian ([www.avidian.com](http://www.avidian.com)). The company went from the peak of Google rankings on queries for sales management software to the middle of the stack of about 30,000 hits. The vice president of the company, Robert Thurmond, said, "We went from the top to nowhere. What can we do?"

Another different challenge tests the senior management of a \$200 million financial services firm. Their site [www.theleasinggroup.com](http://www.theleasinggroup.com) is in the Google index four times. The screenshot with this article shows what the Google cache has stored for this Web site. Robert Callendar, the president of The Leasing Group said, "We pay a company to optimise our pages. How can this happen?"

These are two different problems and represent the opposite ends of the challenge facing those who want a high Google ranking.

On one hand, Avidian has followed all of the SEO (search engine optimization rules). These are the tips and tactics hosed out in online fora, conferences, books, seminars, and Webinars constantly.

On the other hand, The Leasing Group has a Web site that Google cannot see clearly. The notion of epitomizing the site to make it indexable by Google has not turned into Google-friendly Web pages.

Between these two extremes are most of the estimated 10 to 15 billion publicly-accessible Web pages.

In June 2004, Google introduced its automated site map function. This service requires a snippet of code on the Web site. The code calls the Google-generated site map. The idea is that something as basic as site maps is not available for many sites. (For information about the Google Sitemap Protocol, see. [www.google.com/webmasters/sitem-](http://www.google.com/webmasters/sitem-)

Some Information from *The Google Legacy*  
Published by Infonortics, Ltd. [www.infonortics.com](http://www.infonortics.com)

[sps/docs/en/protocol.html](http://sps/docs/en/protocol.html))

</html

PHONE 502-456-2800  
FAX 502-456-1188  
[www.theleasinggroup.com](http://www.theleasinggroup.com)

**LEASE APPLICATION  
(PLEASE PRINT)**

**BUSINESS INFORMATION**

FULL BUSINESS NAME		PHONE#		
BUSINESS ADDRESS (Street, City, State, ZIP)		FAX#		
<input type="checkbox"/> PROPRIETOR/SUP <input type="checkbox"/> PARTNER/INDV LLC	<input type="checkbox"/> CORPORATION <input type="checkbox"/> TAX EXEMPT	YES, IN BUS.	TYPE OF BUSINESS	TAX ID #
NAME OF BASIC		OFFICER		PHONE#
DATE	<input type="checkbox"/> CHECKED NO.	<input type="checkbox"/> CHECKED NO.	<input type="checkbox"/> CHECKED NO.	
	<input type="checkbox"/> LOAN	<input type="checkbox"/> LOAN	<input type="checkbox"/> LOAN	
	NAME, ADDRESS		PHONE#	HIGH CREDIT
TRADE OR BUSINESS REFERENCES	NAME, ADDRESS		PHONE#	HIGH CREDIT
	NAME, ADDRESS		PHONE#	HIGH CREDIT
ADVERT	ADDRESS		PHONE#	
DISBURSMENT				

**PERSONAL INFORMATION**

This is from the Google cache for the LeasingGroup.com. What's immediately evident is that the Googlebot spider has little substantive content to index. Content management systems can generate Web pages that defeat even the most robust indexing algorithms. The fix is to craft pages that present meaningful content, properly tagged, and refreshed with meaningful updates. What's this mean? Double work. The CMS generates Web pages and then a person must create the "flat ASCII" pages that the spiders can index.

Google operates a saloon-door business. Go through the door one way, and Google will index a Web site without charge. Offer rich content and observe some common sense rules, and Google will display hits on that site higher in the results for a query.

Go through the saloon door the other way, and Google will sell AdWords to the Web site owner. An AdWord is a text advertisement that appears alongside the list of Google search results. Click through rates vary by message, but in general, AdWords' customer report traffic ranging from several dozen visitors to thousands. The gotcha is that AdWords are not free. Customers bid on a word, deposit money in an AdWords account, and manage the details.

The question arises, "How does Google balance the need to index objectively with the rapidly-growing pay-for-traffic business?" At this time, Google provides no definite answer. Since the initial public offering, Google provides an abundance of technical information but modest amounts about their business processes and strategies. Expect Google to waffle with regard to relevance helping to drive AdWord sales and AdWord sales relevance to the PageRank algorithm. These are separate sides of Google's house, but like, certain mysterious subatomic particles apparently linked in inexplicable

ways.

The best way to handle the swinging door set up is to go back and forth. Web masters will want to set up Web pages to provide the Google indexing system or Googlebot with the data it needs to index a site, rank it appropriately, and update the index when the site changes.

The SEO industry will adjust to help its customers adjust to Google's "back to basics" initiative. The site map is a starting point. Google will almost certainly adjust its indexing algorithm, tune its fraud detection routines, and provide more pointers for Web masters.

Some Web sites are, at least to the Googlebot, invisible. For example, sites generated by some content management systems make it computationally-expensive for Google to figure out if a page is a new page or the same old page with a different session identification number. Other sites exist only when a user clicks a link. Some of the soon-to-be-released Dot Net 2.0 content management systems present this type of hurdle to the Googlebot. Spiders require specialized scripts to handle certain types of dynamic sites.

The first order of business is to make some sites visible to the Googlebot and other spiders.

SEO experts, however, have been advising clients on ways to take advantage of Google's PageRank algorithm. Google appears to penalize Web sites that over-optimize. "Over-optimization" means a systematic practice of tweaking a Web site so that it appears as high in a list of hits as possible.

What are over-optimization sins? Some are obvious like packing too many words in metatags. Others are devious like creating what appears to be rich content but is content snagged from another Web site. Poking one's head into a search engine optimization teleconference will yield other ideas.

The Google site map push is a signal that Web sites need to do some basic things well. Here's a list of six tips that seem to match closely with Google's suggestions to Web page developers:

1. Frames. Google uses these in Google Maps, Google Print, Google Scholar. You aren't Google. Switch to tables.
2. Clean code. Use validator.w3.org or buy a validation tool. These are built in to some editors like Namo Version 6 or spend \$200 for Web Position Gold or IBP8.
3. Fresh is best. Update content. Stale content translates to ranking downchecks.
4. Solid content without copyright headaches. Useful, relevant, factual, related content means a straight semantic vector to Google. Unrelated content, linguistic spoofs to fool Google, and copyright violations that Google detects can send a site to the bottom of the results list or worse, removed from the Google index.

Some Information from *The Google Legacy*  
Published by Infonortics, Ltd. [www.infonortics.com](http://www.infonortics.com)

5. Can you say it to your mother? If there is anything on your site you can't say to your mother, remove it.

6. In-bound links. Work to get other high traffic sites to link to you. A good link is a listing in the Yahoo! or DMOZ service. A bad link is anything that has an XXX on its site.

Google is an interesting company. It warrants close attention from developers, users, and publishers. Watching swinging doors closely is the first step in getting through them without getting a knock on the nose.

June 9, 2005