Grant
COMMUNICATIONS LLC

Site Assessment™

What Google Sees

Study Methodology

Submitted: February 25, 2014

Upgrading report to "Graded" or "Raw Data w/ Recommendations" versions are valid for 60 days from this date. After this date we will need to re-measure all elements.

Get Results

Introduction

We've been bombarded with questions the last two years as to what Google has been doing, and how we plan to address the changes. In Late 2012, sites that we use for testing purposes started to show up, even though they were employing tactics that were borderline 'black hat' or cheating. Other sites, and some of our clients, had nearly perfect profiles, but slipped from their top positions while their competitors who were blatantly cheating climbed to the top spots. This is exactly what Google was supposed to be penalizing. Yet there are the results that continue to this day. So, there had to be some factors we weren't yet considering. The good news- unless budgets didn't exist for search engine optimization activities, our client sites have continued to perform within expectations to the hours spent on SEO activities.

Study Breakdown

The goal was to have as widespread of a picture of the web, while also recognizing that the majority of internet marketing spend rates would be in the larger, national retail sector. So we used our own client base to begin the benchmarking, then added their competitors. Finally, the national retailers and their top competitors were added for an inclusive snapshot.

By Traffic		
To Date, TOTAL	922 Sites	
Larger Operations	183	
Small Businesses	629	
National Reach	412	
Local/Regional Customer Base	400	
Higher Daily Site Traffic	416	
Niche Traffic Volume	304	

By Vertical Market

Professionals	82
Manufacturing	236
Corporate/Retail Service Companies	95
Construction/Real Estate	118
Retail Trade	311
Software Companies	34
Internet-Related	19
Affiliate Resellers	17

Legend:
- Professionals
- Manufacturing
- Corporate/Retail Service Companies
- Construction/Real Estate
- Retail Trade
- Software Companies
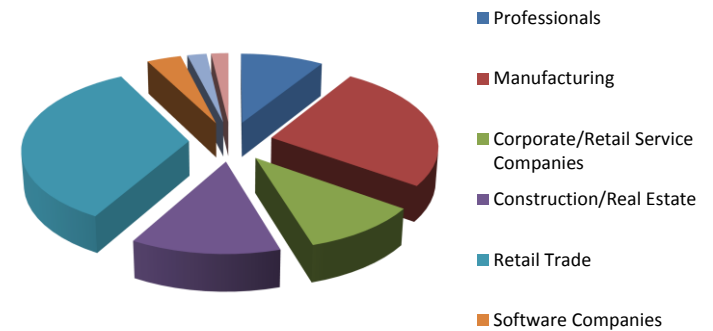
Reverse Engineering Google

Using the benchmark elements we had already revised sites with, we undertook this as a scientific exercise with no "Sacred Cows", preconceived notions or conditions. All aspects of site development were considered. Those elements that scored relatively high in probability of greatest impact were then measured from up to 3 different sources, to ensure the most accurate data was being returned.

Get Results

We utilized 30 different software packages and subscriptions to rate over 400 specific website-related criteria. The tests ranged from on-site specifics like code, hierarchy of content, internal linking and use of basic styling. Off-site factors were also measured, including a number of different scorecard-style programs that evaluated the page or site based on their own rules and tests; some with as many as 20 tests leading to the resulting score. Each of the services and credibility of the source files were also determined, ensuring an objective "black and white" picture was achieved.

What Google Sees

In its simplest form, a website is comprised of pages and images hosted on a web server. Technically, it is more involved, each component of which can affect the total results clients are looking for. We used our general knowledge acquired from 1995 to the present in building an algorithm that focused on domain value, creating an overview from all major aspects. The assumptions were tested on forum research and studying the tool test results in arriving at the following breakdown.

Data Structure

- We broke each aspect down to its logical subset "Area".
- Then, we identified the tools needed to measure each subset.
- Finally, we reassembled all 420+ data points and ran them through our algorithms, determining overall grading.
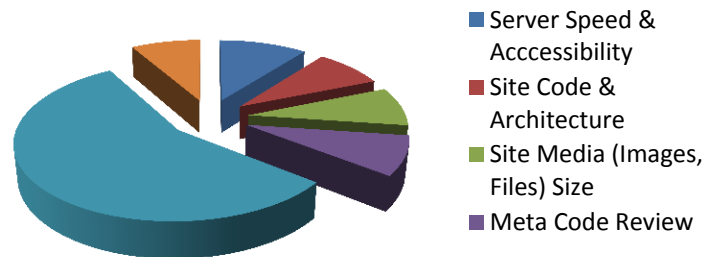- Then adjusted any graded criteria weighting to match the 900 top actual results as ranked on Google

The next page shows the proportional breakdown for the main areas Google reviews.

Area Breakdown

On-Domain

Any aspect of the site that is directly related to the site and performance. Much of this can be adjusted or influenced directly. Traditional SEO firms have used the meta, body content and internal link strategies solely for ranking improvements.

% of Total Site

- Server Speed & Acccessibility
- Site Code & Architecture
- Site Media (Images, Files) Size
- Meta Code Review

Area	# Tests	# Graded
Server Speed & Accessibility	35	20
Site Code & Architecture	44	30
Site Media (Images, Files) Size	54	18
Meta Code Review	40	18
Body Page Content Review	92	26
Internal Link Structures	45	11
Total	**310**	**123**

Off- Domain

Refers to the inbound links and overall domain visibility from other domains to pages on the ranked and measured page. These can also be influenced, but requires additional effort and in some cases a decreased positive indicator to the financial resources or labor applied.

- Inbound Links & Overall Domain Valuation
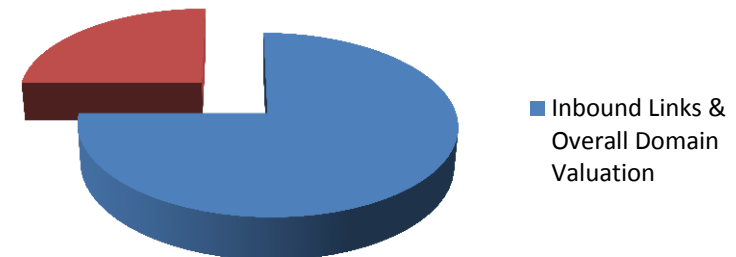
Area	# Tests	# Graded
Inbound Links & Overall Domain Valuation	78	20
Social Web 2.0 Visibility & Activity	42	14
Total	**120**	**34**

Get Results

Study Methodology

The logical starting point was our client sites where we had access to Google analytics and Webmaster Tools.
- This identified stable, declining and increasing traffic.
- Next, we located a key phrase from our client research, that the site was NOT performing well on. (This was a bit more difficult than it sounds.)
    o Then applied the body content-related tools to the measured pages and logged all results.
- Then, we identified the Top 10 Competitors for the phrase.
    o Then applied the tools and methodology to these sites and logged all results.

To ensure that no false positives, or self-perpetuating data was being injected into the scoring, we identified what the top 10 phrases were, for the month prior to our instituting the study.

- We measured all criteria for the #1 sites for each phrase.
- For sites coming in at #2 to #11 positions, we dissected the best performing page from all content areas.
- Other critical scores were also logged for cross-reference purposes.

Scoring and grading used two different conditions;
- Client-specific results as a ratio of the entire range for the measured category, and
- Client-specific results as a ratio of "Best/Empirical" quantity.
    o If this was exceeded, the ratio to the entire range was used.

This approach approximates Google results very closely in a reward-based fashion. Even very small sites with far less resources can thus also rank well for certain phrase criteria. The scope of this study was both comprehensive and technically challenging. Our goal was to reproduce Google results without the availability of 3D Latent Semantic Indexing overlays that Google employs. Over 900 sites were analyzed in varied detail, and any actionable items are flagged by color codes identifying best to worse case scenarios. The recommendations included are also automated for ease of internal task tracking, if we are selected for implementation of the corrective measures.

[1] *After reviewing the domains where phrases slipped in rankings and comparing these to the KEI of the phrase and the# of Monthly Searches web users typed in the phrase, we can see the effects of LSI first-hand. Phrases with a KEI of 0.0 and traffic of less than 200 per month used a completely different dynamic, as Google views the target phrase as nearly worthless. This is even more pronounced where Google sees no traffic, and Compete.com also shows no traffic. Very limited updating is done to the index even during major algorithm updates. We do not know when this practice will be reversed.*

Area Definitions

Here are descriptions of each of the main areas or facets that are reviewed by Google and other search engines. The data acquired comes from both their unique robots, as well as companies that sell their data pre-weighted to the search engines. These other companies, called "Aggregators", are a cost effective way for the search engines to continually explore and add new content from the web into their indexes. At over 48 trillion documents processed by Google to date, you can quickly see why efficiencies were needed in obtaining the freshest data.

On Domain / Off-Page

Server Speed & Accessibility

We had figured that this component would be more critical to overall site performance, based on Google introducing the simplified "Page Speed" test from the more complex WebPage Test organization it funds. Google's goal is to reduce operating costs by reducing the amount it needs to store in cache for every domain in its index, including current and historical cache for every image, Flash file, and any and all other files associated and linked to on a domain. We have included both test results, broken out into the corresponding measured areas. The largest number of the Google tests are found in measuring server responsiveness and cascading server requests. Ultimately, only a few of these measurements are considered as very important.

On Domain / On-Page

Code and Architecture

In conjunction with the server, this was also an initial target for review and retrofits for our clients. Based on the overall results, identifying each page as the most authoritative content for the key phrase focus is paramount in placement.

Meta Code

Considered irrelevant by most Internet Marketing and SEO firms, this is actually a prequalifying area for the search engines. Why? Because any properly formatted APA document fully defines itself in this block

Body Content

This is the most important criteria of all areas measured, and is the reason for a site in the first place. Dissemination of knowledge following APA rules of properly formatted content is key. Using Document Type Definitions (DTDs) consistently and in the correct cascade ensures maximum credibility to the bots. Therefore, weighting of elements should be an immediate consideration during this rehabilitation. For instance, only ONE H1 (Heading 1) tag should appear per page, with a maximum of 6-8 words, and the key phrase assigned to this page appearing within the tag.

Internal Link Structures

Based on forum community commentary, empirical knowledge and statements from all aggregators, this 'critical' component was split out from the Body area. However, based on internal reviews, no single measurement reached the Top 30 in terms of causing any major changes in overall rankings. Many of the measurements can, however, influence in a small way final placement.

Media Size and Compression

Based again on the Google PageSpeed reporting, this area was originally scheduled for inclusion with Server Speed. However, the files themselves are on-page assets. As we identified new factors that influenced final rankings, we split this out into its own measured area. Compression at both the server level and assets level is clearly key in performance. Based on overall results and Google's swing toward Social and mobile, it became apparent that Flash, of all the identified and graded criteria, was actually considered a penalty. If video display via Flash is currently being used, see if the production company can supply an MP4 file type for alternative embedding.

Off Domain

Inbound Links and Aggregator Domain Visibility

Inbound links, or links from other sites to your site, is the basis for Google's unique PageRank (PR) Base-10 algorithm. This was the foundation for how Google ranked websites through 2012. At that time, Google stated that its primary goal was to start lessening PR as the dominant factor in site performance and introduced the 3D Latent Semantic Indexing (What we have named SiteRank) as the primary filter. With the Hummingbird algorithm update added to the mix in August of 2013 and fully rolled out in October of that year, Google has clearly stated that mobile technology and user preferences towards Social media would be king moving forward. Much of the Google index is fueled by outside content aggregators through their spiders and 'bots, which still use inbound link measurements as the unifying element in reviewing domain and site visibility. No matter how you examine this, inbound links are still critical to focus on moving forward. More on the importance of links as Google sees them is described here, with examples of PageRank to visible links needed for top rankings.

Social Media Visibility and Activity

Enter Google's Hummingbird algorithm in August and October 2013. What does a 'Humming' BIRD do? It chirps cheeps… and '**tweets**'. Thus Google once again tells us what the major portion of an algorithm shift is. From primary measurements, however, Twitter isn't the target for placement, other than as an indicator of overall activity levels. Facebook is, after the integration of the Facebook OpenGraph algorithm into the Google algorithm and API.

Glossary

Click here for a comprehensive listing and definitions for all of the criteria measured.