

# Search Engine Optimization

## Organic Definitions

### Current Position

---

The web is flooded with billions of pages, over 100 languages, inconsistent correct language usage per character set, inconsistent bandwidth for technologies used (rich, java, javascripting), and enter mobile media. Education levels, competitiveness and relative newness of the technologies all create havoc for the search engines to accurately represent on any consistent basis the relative importance of a given page or site on the Internet.

As the engines' primary reason for existence is the [somewhat] accurate representation of the worlds knowledge is a correctly ordered index, it is easy to see where definitions and interpretations of the definitions of each engine of what constitutes proper organic optimization and where artificially and negatively influencing the engines lie.

**Organic Search Engine Optimization-** Any activity on an off-page or on-page code or other that will affect the scoring algorithm that creates web contents' natural positioning, based on measured relevancy of a page or domain, on a given search engine.

**From Wikipedia-** Search engine optimization (SEO) is the process of improving the volume or quality of traffic to a web site from search engines via "natural" ("organic" or "algorithmic") search results. Typically, the earlier a site appears in the search results list, the more visitors it will receive from the search engine. SEO may target different kinds of search, including image search, local search, and industry-specific vertical search engines.

As an Internet marketing strategy, SEO considers how search engines work and what people search for. Optimizing a website primarily involves editing its content and HTML coding to both increase its relevance to specific keywords and to remove barriers to the indexing activities of search engines.

The acronym "SEO" can also refer to "search engine optimizers," a term adopted by an industry of consultants who carry out optimization projects on behalf of clients, and by employees who perform SEO services in-house. Search engine optimizers may offer SEO as a stand-alone service or as a part of a broader marketing campaign. Because effective SEO may require changes to the HTML source code of a site, SEO tactics may be incorporated into web site development and design. The term "search engine friendly" may be used to describe web site designs, menus, content management systems and shopping carts that are easy to optimize.

Another class of techniques, known as black hat SEO or Spamdexing, use methods such as link farms and keyword stuffing that degrade both the relevance of search results and the user-experience of search engines. Search engines look for sites that employ these techniques in order to remove them from their indices.

This is a tough one. We know that the engines prefer text links over images with javascript in displaying mouse hover effects on a navigation system. By using a CSS-driven navigation system, would the site then be targeted as unfairly influencing an engines decision to crawl a site over the invisible javascript version?

Based on the work produced by SEMPA members, it appears that CSS is allowed as a white hat technique. But, as another associate commented, *“Building a good website is presenting an unfair advantage over the vast majority of garbage out there, so just building a good web site could be considered as a form of cheating”*.

## Current Organic Search Engine Levels

---



**White Hat, Lily-** This version of site optimization is the most preferred, and mandated in becoming a member of SEMPA (Search Engine Marketing Professionals Association).

The first core code used by engine robots is `<meta name="revisit-after" content="15 days" />`. This is considered to be “gray-to-black hat” by this organization, and cannot appear on any site to be considered for membership. With just this one line, copied from staticworx.com. This is considered “lily white”. ADA-compliant code, text and meta content unique and professionally crafted. No deviation from the hierarchal DTDs offered by the W3C; H1 always proceeds H2, which always proceeds H3, etc. No htaccess 301/302 redirects.

In examining their client lists and reading the forums and past customer feedback, one quickly realizes SEMPA can afford to stand their ground. Annual fees to access their membership as a client starts at \$100,000. So the hours and average rate allows for dedicated R&D departments, high wages to employ doctorates or steal staff from the engines, and extremely high profile links equivalent to our sitemap-g program (yes, even they rely on this tactic). And if you look carefully on some of the SEMPA client sites NOT linked from SEMPA member portfolio pages, you may see YOUR old code for a particularly tough industry and low client budget...



**White Hat, Real World-** This more realistic approach to coding assumes that nothing is being done to artificially influence or affect search engine robots. All code is clean, and is reviewed and aligned at time of launch. Common practice strict DTDs from the W3Cs HTML 4.0 implementation is used. Transitional can be used with extraordinary documentation, as Bobby-compliant (US ADA) visibility is also required. No links are or may be paid for, other than industry-known search engine PPC programs.

Even PPC to a primary domain is questionable, due to aggregator PPC link campaigns like go.com, cnet. This is where a secondary educational site, blog or Social 2.0 page may be built for PPC, which then has corresponding product offerings on the educational page. An extraordinarily expensive way to develop sites. Very few sites on the web comply to this standard. It is not financially compatible with standard corporate budgets, and is typically seen in state and federal websites targeting consumers (tourism sites).



**Gray Hat, Light [90%-99%]-** Transitional HTML code is allowed, as the expansion of user experience is greatly enhanced. Javascript, rich media, Meta Robots, Dublin Core and other meta level code is now allowed. This code remains in place on Staticworx. A cross-domain links page is allowed, but only if strict

commonality can be proven and the link structure does not solicit via outbound email or PPC for new links on a paid-for basis.

Low value mutual links between domains are accepted as part of the process to create the inclusive 'web'. No hidden text that influences an engine is allowed, using CSS or other code, unless directly related to a users navigational ease. (Staples, products description page or our Home page are two accepted practices). No key phrase stuffing, no duplicate titles sitewide, no title stuffing, ALT tags used on all images, unique for each image, must be documented.

The above two types of sites do gain consideration by SEMPA and, I believe, will emerge within SEMPA into a pure unified 'white hat' category. All three levels reflect that the sites are or have been built by professional or academically educated coders, that know how to code and code properly, link correctly, and are building sites for ease of crawlability and content being added to the respective indexes. No 'get rich quick' code can be seen, other than an occasional snippet from what may have been an older developer or grandfathered code on an old domain.



**Gray Hat, Medium-** This is where the vast majority of web properties exist, based on the site builders available knowledge and resources. Most people don't study what they are allowed and not allowed to do, so these sites generally appear with one or more areas of code as being highly suspect. The beginnings of intentional cheating is seen, or a mark of a too aggressive approach without enough resources allocated, or using vendors that are unaware of the risks (another indication of limited resources).

Based on the sheer number of these sites that otherwise represent honest industries, companies, organizations or individuals, requests to remove these sites from the indices I believe will mostly go ignored. However, other punitive steps may be taken if seen as blatantly intentional; keyword stuffing would fall into this category. Into the sandbox for 6 months, following by automatically re-scoring the domain to a PRO with a delayed re-indexing may and probably does occur.

The offending code is examined and modifications made to the algorithm, to identify and reduce page ranks on other pages and domains using similar metrics. Links from these domains are reduced or eliminated in value. Some unrelated inbound links which may or may not be paid, may also be found by the robots. If found to be paid and are from dark neighborhoods, additional relevancy penalties would be applied. Site code may be inconsistent hierarchically, DTDs ignored. CSS inconsistent or non-consistent, which may reduce penalties; the person building the site was obviously unaware of 'best practice', so no unduly harsh penalties would be assessed.



**Gray Hat, Dark-** These sites are 'pushing the envelope' at more than one element of the code, and usually heavily at the meta level elements. This is being done intentionally, but typically the infractions remain at the code level. If reported, or found and identified as egregious by a robot through the algorithm, these domains are subject to a long sandbox hibernation or potential permanent exclusion from the index.

Examples include multiple iterations of the same phrases replicated identically at all levels of the metas, three or more iterations per phrase block.

Key phrase stuffing at this level and other element levels; the bottom of visible text blocks, ALT stuffing, link title stuffing and cross-link phrase stuffing. Dozens to hundreds of duplicate domains with identical landing page content and no attempt at individualizing, with automatic redirects or meta refreshes to other visible domains. Paid-for inbound links on free-for-all Denver link farms, where no commonality can be traced by a robot.



**Black Hat-** Any vestiges of flying below the radar are removed. These sites are egregious in the lack of content value to the web user, which is the search engine customer. This is where the greatest resources are applied from the engines to counter bad code, and where the greatest resources are being applied from those industries benefiting the most from manipulation.

During the mid-to-late 1990s, we studied this code on the best performing sites to determine what elements may and should not be used, as well as weighting and prominence of phrases. The software available on a subscription basis surpassed this need back in 1999/2000. The industries that are noted in using the most aggressive techniques focus on base human needs at the bottom of Maslow's Hierarchy and thus, are extremely in demand. This creates the competitive environment that fosters cheating mentality; money is the only goal.

This is where black text on a black background is first seen. Meta refresh, first used to forward users to an updated page on a corporate site, is now used to force users to an external domain. All the tactics found in Dark Gray Hat. Automatic content generators flooding dynamically generated pages to automatically created domains and sub-domains, all pointing to yet another forwarding domain. "Hide behind the complexity of hundreds of concurrent strategies", is the site owners complete thought and reasoning. Link structures include highly-ranked legitimate sites, or hijacking external PR pages for backlinks to bad domains, to their benefit.

The tactics become increasingly sophisticated and automated, with doctorate-level R&D specialists guiding the ongoing development and viral population.