

Fundamental Principles of Search Engine Optimization

The traditional objective of search engine optimization is to create a favorable visual impression upon random searchers that entices them to click through search results listings to specific destinations. It is generally assumed that the first position in search results listings is the most beneficial in that the majority of user click-throughs historically occur on the first position. However, knowledgeable searchers will click on the most apparently relevant listing in the search results or on more than one listing.

The search optimizer's goal should thus be to create a compelling visibility in search results that drives traffic to a specific destination. However, for a destination that requires more than just random traffic (that is, some specific action is required), the search optimizer should be driving pre-qualified, convertible traffic to the destination. That is, the people who click on the listing in the search results should be reasonably persuaded they will find what they are seeking before they click on the link.

There are four factors that directly affect your Web content's performance in search results:

1. What you do with your page
2. What other people do on their pages
3. What the search engines do with the data they collect about Web pages
4. What people search for

Although you cannot control what other people do with their pages or what people search for, you do control what you do on your own pages and the search engines encourage Webmasters to provide some guidance for how to index their pages. Commercial search engine optimization must address four specific areas in order to achieve success: keyword research, content organization, search visibility, and linking relationships. This paper discusses each topic in more detail below.

Keyword Research

One very common mistake many business Web site operators and designers make is to proceed on the assumption that favored keywords are used by all people to find specific types of content. For example, an insurance professional may assume that people are using the word "insurance" to find Web sites like his own. However, his insurance business most likely specializes in one of several branches of insurance coverage. And it is also more than likely that his practice is limited to a small area.

Hence, it would behoove the insurance professional to learn how people search the Internet for providers of insurance in his specialty and location. Keyword research is generally conducted with the aid of tools that draw upon pools of data provided by one or more search engines for recent user queries. Query data provides insight into how people look for information on the Internet, and it often turns out that people who are unfamiliar with a specific industry's jargon use words that professionals with years of experience would never think to use.

The search optimizer must evaluate the probable benefit of using words that knowledgeable searchers would use versus the words that naïve searchers use. Does the Web site being promoted offer value to both types of visitors? Which visitor is more likely to be converted to a transaction-taking user/customer? In the best of all environments, the search optimizer can identify usable keywords that address multiple markets.

However, query data does not usually indicate whether search patterns are cyclical (seasonal or annual), continuous, or spike/event-driven. The best keyword research compares historical trends in query data over the course of 6 to 24 months to determine which keywords are most likely to be used by the most people. Historical trend analysis is an often-overlooked aspect of keyword research, which should be conducted on at least a semi-annual if not quarterly basis after initial search engine optimization is performed.

Once the optimal keywords have been identified in user query patterns, a best fit must be found with respect to the content that will actually be uploaded to the Web. Exceptions to the “best fit” category should be considered for pay-per-click or other forms of advertising, or perhaps for secondary content to be prepared at a later time.

Matching performance keywords to existing or planned copy ensures that an optimal relevance between the Web copy and search patterns is achievable.

Content Organization

Web site designers need to understand what is necessary to make pages not only optimally indexable but also prominently visible in search engines. Every page on the Web site should have unique title and meta description data that accurately describes the page’s content. The keywords meta tag may also be useful but is not really necessary.

Page content should be organized in a tiered structure, emphasizing the page’s main topic and dividing content into sub-headed secondary topics. On-site navigation should be included in the content organization, both in so-called “navigation structures” and as part of the original Web copy, which can be used to effectively promote and co-brand related copy located on other pages.

The keywords for which a page is most relevant should be emphasized through reasonable use of bold and italics, underline, and/or quotation marks. They can and should also be incorporated in section headers, page headers, image alt= text, and other user-beneficial structures. If an HTML structure provides no benefit to users, then it should not be adapted to optimization.

Web copy should be presented in a free-flowing, streamlined format with as little in-copy formatting and embedding as possible. The long-running debate over the use of div/span formatting versus table-formatting for text is irrelevant to the search engine optimization process. Either div/span-driven or table-driven layout can impede the indexability and relevance of on-page copy. In fact, many non-tabular layouts relying specifically on div/span formatting have provided minimal on-page relevance and thus were ineffective for search engine optimization.

Generally speaking, the more usable and accessible a Web page is, the more easily a search engine will be able to index the content and determine what that content is relevant to. That is, the less page layout and design obstructs or mingles with indexable text, the easier it is for the text to perform to its optimal potential in search results.

Large content Web sites should implemented tiered, multi-directory structures to improve both crawling and user navigation. A common misconception is that “deep” content located several directories deep inside a site’s hierarchy have a poor chance of being found, indexed, and ranked by search engines. In fact, the location of a page and its depth within a Web site are irrelevant to the search engines’ ability to find the content and index it. Clean, easy-to-use, static HTML on-site navigation is one of the most vital elements required for successful search engine optimization. Javascript, drop-down boxes, Flash, and other non-text navigational links impede search engine crawling and indexing. Every page should have user-visible static text navigation links.

A large content site also benefits from cross-promotion between sections and levels within the directory tree. Every page on a large content site should have at least two internal links pointing to it to help crawlers find the page. The more internal links that point to any given page, the more easily it will be found, crawled, and indexed by search engines. And the more easily it will be found and visited by people who just navigate through the Web site.

Search Visibility

Search visibility begins with the page title. This is usually one of the first elements coded into the heading section of a typical Web page and it is the first part of the page’s content that is displayed to searchers in search results. Page titles should be compelling and informative. They should not be unnecessarily repetitive. The most effective page titles emphasize the value of the content rather than the keywords the page owner wants to be found for.

After the title, the page meta description should be robust and informative, incorporating a call-to-action (or reinforcing a call-to-action employed in the title). Some search engines may use static directory titles and descriptions (from Yahoo! or DMOZ) and these should be disabled through use of the “robots” meta tag (use NOODP,NOYDIR in the meta tag “content”) so that the page controls its visibility in search results.

Page URLs can also influence both search engine relevance scores and searcher decisions to click on the URLs. Calls-to-action should actually be avoided in page URLs. Instead, the best performing page URLs tend to be those which concisely describe the page’s 2-3 most important keywords.

Search visibility also encompasses the various queries for which a page may potentially be found in search results. Just because you optimize a page for “some specific keyword expression” does not mean it won’t be found for other keyword expressions. Broader search visibility can be influenced by use of secondary optimization techniques and strategies. One common technique is to point link anchor text from other pages to a destination page. However, simply mentioning a page URL in copy can help improve that page’s visibility in search results.

Managing search visibility requires a very broad strategic approach that assumes every indexable word on a page contributes to its ability to appear in search engine results.

Linking Relationships

Linking relationships between Web pages help search engines find new content and determine (to a limited extent) how reputable that content may be. Unfortunately, a great deal of misinformation has been shared through the years regarding the value and importance of linking relationships in search engine optimization.

Linking relationships provide value beyond search engine optimization. An SEO campaign should be only one aspect of a much broader online marketing strategy. A very useful linking relationship provides brand visibility and traffic regardless of whether search engines discover and utilize the linking relationship. It is thus very important to evaluate linking relationships on the basis of how they help a Web site *as if search engines either do not exist or do not evaluate links*.

Where search engines do look at links, any link may or may not pass value. Some Web sites cannot pass value through their links because those pages have not yet been indexed. Some pages may be filtered or penalized for violating search engine guidelines. Some pages may be constructed so as to pass as little value as possible.

The value links pass can be described as:

1. Visibility – If a real person sees a link, they see a pathway to your content
2. Anchor text – If a search engine allows links to pass anchor text, links enhance relevance
3. Traffic – Links on high traffic pages tend to send traffic to their destinations
4. Trust – The fact that a trusted page links to another shows the destination should be trusted
5. Link Weight – Called “PageRank” by Google, Link Weight is used by search engines in several ways

Most people mistakenly believe that Link Weight (*PageRank*) is important to search engine optimization. All other things being equal, Link Weight may very well be the deciding factor in more than one choice, but quite often Link Weight has considerably less impact on search engine visibility and results than people in general have been led to believe.

Links are classified as *inbound* or *external links*, *internal links*, or *outbound links*. Inbound links are often confused with *backlinks*, which include both external and internal links. Outbound links are sometimes called forward links, but forward links include both internal links and outbound links. Internal links are the links you use for intrasite navigation. Outbound links are the links you point to pages you do not control. Every forward link on a page is a backlink for another page.

An inordinate amount of attention is given to backlinks. Most backlinks do not actually help with search engine results rankings for a variety of reasons. They may not pass anchor text, or they may not be trusted, or their anchor text may be irrelevant to specific queries. The Link Weight a page accrues from a large number of backlinks may help with some search results rankings, but most Link Weight really makes little to no difference in search results.

Where Link Weight proves to be useful is in helping search engines crawl pages. High Link Weight pages tend to be crawled more often than low Link Weight pages. It therefore helps new content considerably to be pointed to by pages with high Link Weight, as that increases the chances of the new content being found and indexed quickly.

Many Webmasters overlook their own internal linking relationships and devote too much time and effort to seeking external links when they already have sufficient resources to ensure their new content pages are found and indexed quickly. It is also possible to use internal link anchor text to help boost the relevance of pages for specific keywords.

SEO Myths

Search engine optimization resources such as Web forums, blogs, newsletters, and tutorials provide a variety of opinions and conclusions about what works best with search engines. Because each search engine develops its own algorithms and weighting schema, no one concept works exactly the same way for all search engines. And as the search engines update their indexing and filtering technologies (or other technologies), techniques or ideas that were once useful to know become outdated. The following are popular SEO myths that have been debunked. Nonetheless, they persist in the popular imagination and impede successful search engine optimization.

The PageRank reported by the Google Toolbar is useful information.

Nothing could be further from the truth. Google's Toolbar does not report the actual PageRank that Google calculates internally for its indexing and crawling. The real PageRank values assigned to page are values somewhere between 0 and 1. The numbers reported by the Toolbar (ranging from 0 to 10) are proxy values determined according to an undisclosed proprietary scale. But Google representatives have repeatedly emphasized that these numbers are only published 3-4 times a year and that when they are published they have no effect or impact on search engine results.

The numbers represent values determined at some point in the past that have already been factored into Google's index by the time the Toolbar PageRank numbers are published. In other words, if a page is shown as a PR 4 value today, it may already have incurred several internal PageRank adjustments subsequent to the valuation that was converted into the PR 4. It is therefore impossible to use the Google Toolbar PageRank value to determine anything useful for search engine optimization about a Web page.

A page is optimized if you insert your keywords into the title and meta tags.

A page is "optimized" if you do everything possible to show both your visitors and search engines which words are most important throughout all the content of your page. If you only insert keywords into the title tag and keywords meta tag, your page is NOT optimized. In fact, simply inserting keywords into the title and keywords meta tag is rarely beneficial. On-page optimization remains the most important part of the search engine optimization process. Your page copy should be unique, substantial, well-organized, and presented in a user-beneficial way.

Search engines, especially Google, rely on links more than anything else to determine rankings.

This has never been true. What is true is that many people, particularly in the search engine optimization community, have to come to rely almost solely upon links to build their search engine traffic. This is a very inefficient approach but it is the easiest method to learn and many people have mistaken the ease with which pages can be “link bombed” to the top of search results for true search engine optimization.

The vast majority of the several billion queries that people enter into search engines every month are resolved on the basis of relevance, and the search engines determine relevance by looking at the content on the pages. Search engines that allow links to pass anchor text also look at anchor text to determine relevance, but Google in particular now ignores approximately 80% of all links.

The best links come from the .EDU and .GOV top-level domains.

Statistically, links from Web sites located in these top-level domains tend to be more trustworthy and reliable, but the search engines claim they do not favor these top-level domains. Rather, as many people outside the .EDU and .GOV domains point to those types of pages, those pages accrue value from the large numbers of inbound links. But many .EDU and .GOV sites also have thousands of Web pages. Such large content sites are often able to concentrate much of their value in very important, “key” pages.

You lose Link Weight when you link out to other sites.

Link Weight (often called PageRank) is calculated on the basis of linking relationships. But the amount of Link Weight a page accrues is only determined by the links pointing to that page. Some people have accepted the false argument that a Web site can lose Link Weight if its pages link out to other sites. In fact, Link Weight is constantly being recalculated as search engines find new pages and discard old pages from their indexes. Every page’s Link Weight therefore changes constantly regardless of whether a site links out.

Furthermore, sites that attempt to “preserve PageRank” are most likely treated as if they link to every other page on the Web. Hence, the Link Weight their pages can pass on is distributed anyway, but they don’t lose any Link Weight by doing so. The “bleeding PageRank” point of view is based on a misunderstanding of how PageRank is actually calculated.

#