

Overview:

Search engines strive to deliver “preferred” content to their users. They match a user’s keyword query to internet pages. Although many of the ranking factors search engines use are off-page factors, on-page factors are still critical for success. Considerations include:

- Head Tags
 - Title tag
 - Meta tags
- Body Content
 - Readable by Search Engines
 - 250-350 Words
 - Unique Content
 - Valuable Information
 - Page Relevancy
 - Keyword Density
 - SPAM
- PageRank
- HTML errors



Head Tags

Head tags are placed in the beginning of a website page. They serve to inform the search engine of various page attributes. The most common tags include the Title Tag and Meta Tags.

Meta Tags are further segmented into:

- Meta description
- Meta keyword
- Meta robot
- Meta refresh

Title Tag – The Title Tag displays in the title bar of a browser window. It serves as a short website page description.

Of all the head tags, the Title Tag is the most important for organic optimization. Search Engines and directories use a website page Title Tag in their listings. Because the Title Tag is short (classically less than 67 characters), the words used must be concise and specific to that page. It is this characteristic that makes the Title Tag so important to search engines.

Search engines like Google examine the Title Tag from a keyword density point of view. More words in the title reduce the keyword density and have a negative affect on keyword phrase rankings.

Meta Description – The Meta Description tag is a long description of the page. Although this tag is not shown anywhere on a website page, search engines and directories commonly use the Meta Description tag as a more detailed and verbose page description.

Because of its length, Google and other search engines place only minor organic optimization emphasis on the Meta Description. A strong description tag, however, has other marketing advantages. Keywords matching a search query will have BOLD formatting. As such, users will be more likely to click on a search engine listing containing their keyword.

Meta Description tags should be less than 30 words (approximately 160 characters).

Meta Keyword – This tag is a list of keyword phrases separated by comas. Although this tag at one time played a significant role in early search engines algorithms, all major search engines have now stated that this tag has minimal ranking influence.

The Meta Keyword tag may be relatively long. However, best practices suggest the tag be limited to 10-15 keywords.

Meta Robots Tag – This tag is instructional only. It instructs robots or search engine crawlers to either stop or proceed. The two most common instructions are:

```
<...CONTENT="NOINDEX, NOFOLLOW">
and
<...CONTENT="INDEX, FOLLOW">
```

The first instruction requests that a search engine exclude a page from its index. Likewise, the second instruction informs the robot to continue. It should be noted that search engines are not required to obey these instructions. They may choose to crawl a website page even with “NOINDEX, NOFOLLOW” instructions.

Meta Refresh Tag – This tag can be very dangerous. A Meta Refresh tag direct a visitor to another website page. If the re-direction is set too fast, visitors never have a chance to see the contents of the first page. A search engine crawls and ranks content that visitors never see. In this case, a search engine may consider the page as SPAM.

There are, however, legitimate uses of the Meta Refresh tag. In these cases, Google has suggested a time of 5 or more seconds to avoid potential SPAM penalties.

Body Content

A search engine must be able to determine what a page is about before it can apply appropriate keyword ranking. Search engines do this by crawling or spidering a page.

Readable by Search Engines - During the crawling process, search engines examine the code that is used to generate a page. However, search engines may not always be able to read all the words that may appear on a page. Some websites use graphics or images for text. Others may use JavaScript print commands. This kind of content is not readable by search engines and should be avoided. Rather, content should be coded in a manner that is easy for the search engines to read.

250-350 Words - A common question is: “how much text is enough”. The answer comes from statistics sampling methods. Search engines remove all punctuation (except apostrophe marks and hyphens) and all “stop” words. Stop words are commonly used words (e.g. the, is, a, it, an, with, etc.) that typically account for 40-50% of all content. After stop words are removed, search engines will examine the remaining words.

Consider a website page of 300 words. If 50% are stop words, only 150 remain. Statistically speaking, a sample size of 96 words provides the confidence interval of 95% (approx.). A sample size of 150 provides a confidence level of approximately 99%. In other words, an analysis of a website page with 150 words (excluding stop words) will produce a strong understanding of what the page is about.

Fewer than 250 words (stop words included) in paragraph format will start to compromise the statistical confidence. More than 350 words will add little value from a search engine perspective because the statistical confidence level does not increase much more.

Unique Content - Some have speculated that as much as 30% of the internet is duplicate content. Technically speaking, duplicate content is an exact copy of a website page. Since each page has a calculated "hash" value that acts like a unique fingerprint, search engine detection of exact duplicate content is relatively easy.

But modifying even one thing on a page will change the hash value. After making a small change, the content is no longer an exact duplicate – it is near-duplicate.

Google has developed special technology to counter this condition. According to Google's patent, pages found "to be in the same cluster" (pages sharing a portion of the same content) will not appear on the same SERP (Search Engine Results Page). Therefore, a website page worthy of ranking on page one may be forced to page two if there is another website page of slightly greater strength that shares some duplicate content. From this point of view, the page forced to page two is penalized. If a website is found to have substantial near-duplicate content, the entire site may be penalized.

Valuable Information – Internet pages that have unique content but little or no value may avoid search engine penalties. Unfortunately, these types of pages don't satisfy the public. Pages that offer real value tend to attract greater public visibility. And search engines measure this visibility by monitoring the number and quality of links to these pages (backlinks).

The concept of valued content extends beyond search engines. Strong ranking may produce many visitors. But strong content with a compelling message helps convince visitors to take a desired action:

- Fill in a request for more information
- Call, chat, or email
- Download article, white paper, case study, or resource sheet.
- Sign up for Webinar, RSS feed, or newsletter
- Request samples
- Enter Store to shop for specific items

The choice and form of an appropriate "call to action" is very dependant on the purpose of the website, market and the purchase cycle.

Page Relevancy – Any page can be optimized for any keyword. But an optimized page may not read well. In some cases, current content doesn't properly support the keyword(s). Optimization efforts, although technically correct, violates the Valuable Information concept above. In these cases, there are 2 options:

1. Abandon the page and create fresh content (new or existing URL)
2. Modify the current content to better suit the keywords

Keyword Density – Search engines use variations of a probabilistic term vector analysis to determine what a website page is about. Although it is not the same as keyword density, the methods have similarities. Since an index of billions of website pages is required to calculate term vectors and this information is not publicly available, the SEO community relies on keyword density analysis. Keyword density is a ratio of keyword phrases to the number of words found on a website page. The method of calculation varies depending upon whether "stop words", link text, and other tags are included. For this reason, various online keyword density tools produce different values.

Ideal keyword density values depend on the densities found among top ranking websites. The objective is to create a density that is among the top range of websites currently ranking well. If a keyword density is too high, it may trigger a search engine flag as "unnatural"; too low of a keyword density will produce lower rankings. As a standard rule of thumb, 3-4% density, without "stop word" or tag exclusion, is typical (3-4 mentions of the keyword for every 100 words).

Keyword density in the Title Tag is nearly as important as densities in the body. Since the page Title Tag (a tag found in the head section of the HTML page) has limited space, the wording is particularly important. Long titles will have lower keyword densities, which will have a negative affect on rankings. It should be noted that the Title Tag, more than any other condition, limits the number of keywords that may be effectively optimized for a specific page.

SPAM – The definition of SPAM has taken on several definitions. From a search engine perspective, SPAM is a violation of published search engine guidelines.

Engine	Published Webmaster Guidelines
Google	http://www.google.com/support/webmasters/bin/answer.py?answer=35769
Yahoo	http://help.yahoo.com/l/us/yahoo/search/basics/basics-18.html
MSN	http://www.seoconsultants.com/search-engines/guidelines/

All SPAM techniques either:

1. Hide content from the search engines that is visible to the public,
2. Or, serve content to the search engine that is not visible to the public.

The most common SPAM techniques include:

- Hidden Text / links
 - Hidden Div Tags
 - Cascading Style Sheets (CSS)
 - Text same or near color as background color
- Very small text
- Repetitive keywords
- JavaScript and Meta Refresh redirects
- Selling text links without rel="nofollow" attributes

Spam is still a big issue with search engines. Technology does exist to automatically detect all these techniques. However, the required computing resources to detect these techniques are very intensive. For this reason, search

engines rely heavily on SPAM reports from other webmasters (competitors) who report offending websites. Google's SPAM reporting tool is located here:

<https://www.google.com/webmasters/tools/spamreport>.

PageRank

Google defines PageRank as the "importance of a website page". It is passed from one website page to another through backlinks. More backlinks typically translates into higher PageRank values.



In most cases, the home page of a website has the highest PageRank value. This means the home page has the greatest optimization potential (assuming all other ranking factors are equal).

1st level interior pages, that have direct links from the home page, typically have slightly lower PageRank values. PageRank values decline for deeper interior pages. Lower PageRank values typically mean lower ranking potential.

E-commerce store pages that are deep within a website often have very low PageRank values. For this reason alone, they are poor candidates for all but the weakest competitive keywords.

HTML errors

Conservative estimates suggest that more than 30% of all website pages have some HTML coding errors. Internet browsers are designed to 'forgive' many of the common errors and render internet pages correctly.



Search engines, however, are not as forgiving and more sensitive to poor coding practices because browsers on cell phones and PDAs are not as robust. If incorrect HTML code causes an error, a search engines may abandon the page before it can be downloaded for analysis.

Pages that have excessive broken links may experience ranking demotions because a search engine does not want to be indirectly associated with a poorly maintained website.

Testing HTML code is a simple task of code validation. Online tools offered by W3C are particularly helpful: <http://validator.w3.org/>.

Broken links may be tested using desk-top applications like Xenu: <http://home.snafu.de/tilman/xenulink.html>.

Conclusion

In the final analysis, search engines define the perfect page. Their Webmaster guidelines serve as strong guidance.

Tactic	Reason
Targeted Head Tags	A strong Title tag is among the most influential organic optimization factors.
Strong Body Content optimization	To determine with good confidence what a page is really about.
Valued content produces great visibility	Visibility as measured by the number and quality of backlinks to pages (external backlinks preferred).
Minimal HTML errors	Pages resolve correctly and reflect well on search engine's capabilities to deliver good content across a broad range of browsers (cell phones, PDAs, and PCs)
High PageRank Value	Home page or 1 st level interior pages offer the strongest ranking opportunities.

When on-page factors, combined with a strong backlink structure, meet or exceed conditions of website pages currently ranking for specific keywords, the chances of ranking success are dramatically improved.