

Search Engines

How They Work:

1. Pages are gathered off the Web using a process called spidering.
 - To be placed in a Search Engine, the URL is either submitted or is pulled from a list in the domain name registry.
 - Search engines use programs called spiders, robots or crawlers to collect pages of the Web for indexing.
2. Collected pages are indexed.
 - Even the largest search engines only index's about a third of the documents online. No one knows the size of the Web, undoubtedly increasing.
 - Saves pages or portions of pages for analysis and follows links.
 - Indexing: determine what the page is about, methods vary but basically looks at various components: <title>, <meta>, comment text, link titles, text in headings and body text.
 - Once the pages are analyzed for keywords, it's ranked relative to other pages with similar keywords.
3. On a search page the user can enter queries and see a list of pages that are related.
 - Ranking is the very secret part of search engine operations.
 - Final part is the search page itself (where query is made). Once queried, the search engine will retrieve the pages that meet the criteria and present them on the results page.

Robot Exclusion

It isn't always a good idea to have a robot index your entire site. Some things don't need to be indexed, it can overwhelm the server hosting the site being indexed, it may index transitory pages or it may do what's called "deep-linking" allowing a user to enter the site deep within. To deal with limiting robot access, Robot Exclusion protocol was adopted.

- Use file called robots.txt found in root directory of the Web site.
- Alternative Method is through <meta> tags:
<meta name="robots" contents="noindex">
 - You can inform spiders to not follow any links coming out.
<meta name="robots" contents="noindex, nofollow">
 - Downside: fewer public search engines support <meta> tag than robot.txt.

Optimizing

Overview: search engines look at the page and make educated guesses about what pages are about by following a set of rules to try and understand using things like word frequency and number of linking sites.

<META> Tag Keywords

- Many search engines look here for keywords or descriptions of page content.
- Start with the most specific keywords and end with generic keywords
- Some place redundant keywords in <meta> to improve ranking, but many search engines consider this to be keyword loading and drop the page from indexing. Combine with words and phrases to make less obvious and vary the order. Be aware search engines may still notice and consider this spamming which reduces the page's ranking or drops it from indexing.

- Ex: `<meta name="keywords" contents="Butler-1000, Robot butler, Metallic Man Servant, Robot butler, Butler-1000, robot, robot butler, Democompany, metallic man servant, butler-1000, robot, butler, Robot butler, Butler-1000">`

<META> Tag Description

- Keep the description to a sentence or two (suggested: 25 words or less) at most three to four sentences.
- Provide valuable info in the description that will help the user determine if they want to visit your site.
- Ex: `<meta name="description" contents="The DemoCompany Robot Butler is the most outstanding metallic man servant on the market. The Butler-1000 comes complete with multiple personalities and voice modules including the ever-popular faux-British accent">`

Titles and File Naming

- A good, descriptive `<title>` is a very important aspect.
 - Ex: `<title>Butler 1000 Robot Butler from Demo Company</title>`
 - Be aware that users see title too and they are used for bookmarking; so really long titles are only for the search engines.
- Name of file can be as important
 - Ex: `butler.htm` vs. `butler1000_robot_butler.htm`
- Good domain name and directory structure allow you to create URL that almost make sense.
 - `http://democompany.com/products/robots/butler1000_robot_butler.htm`
 - Includes the same info as title, secondary way of knowing where you are in site.

Relevant Text Content

- Best ways to have keywords and phrases actually within the content of the page like `<H1>`, `<H2>`. May also look at contents of link text `Blah`
- Use ALT attribute for `` to provide extra info
- With pages that are large images some put hidden and or small text so search engines see it but not users (be careful with this, some consider this spamming and may cause you to be blacklisted).

Links and Entry Pages

- Another aspect of ranking is the number of links leaving a page as well as a number of pages that link to a page.
- Landmark pages such as home pages tend to have a lot of outgoing and incoming links. Search engines prefer to rank these pages highly, so it's important that key pages have links to them from nearly every page.
- Submit many pages of the Web site - all of these entry pages or doorway pages point to important content in the site

Submit Your Site to Search Engines

- These are the three main search engines and you will want to submit your site to at least these three sites when you make major changes to your site.
 - Bing/ MSN: <http://www.bing.com/docs/submit.aspx>
 - Google: <http://www.google.com/submityourcontent/index.html>
 - Yahoo!: <http://search.yahoo.com/info/submit.html>

Tricks and Tips

- Don't waste your time trying to trick search engines as the effects are only temporary. Instead build your pages for your users and the rest will follow.

- Here's more about what the engines consider high quality vs. low quality, according to what they recommend in their guidelines.

High Quality

Yahoo

- Original and unique content of genuine value.
- Pages designed primarily for humans, with search engine considerations secondary.
- Hyperlinks intended to help people find interesting, related content, when applicable.
- Metadata (including title and description) that accurately describes the contents of a Web page.
- Good Web design in general.

Google

- Make pages for users, not for search engines. Don't deceive your users, or present different content to search engines than you display to users.
- Avoid tricks intended to improve search engine rankings. A good rule of thumb is whether you'd feel comfortable explaining what you've done to a Website that competes with you. Another useful test is to ask, "Does this help my users? Would I do this if search engines didn't exist?"
- Don't participate in link schemes designed to increase your site's ranking or PageRank. In particular, avoid links to Web spammers or "bad neighborhoods" on the Web as your own ranking may be affected adversely by those links.
- Don't use unauthorized computer programs to submit pages, check rankings, etc. Such programs consume computing resources and violate our terms of service. Google does not recommend the use of products such as WebPosition Gold™ that send automatic or programmatic queries to Google.

Low Quality

Yahoo

- Pages that harm accuracy, diversity or relevance of search results.
- Pages dedicated to directing the user to another page.
- Pages that have substantially the same content as other pages.
- Sites with numerous, unnecessary virtual hostnames.
- Pages in great quantity, automatically generated or of little value.
- Pages using methods to artificially inflate search engine ranking.
- The use of text that is hidden from the user.
- Pages that give the search engine different content than what the end-user sees.
- Excessively cross-linking sites to inflate a site's apparent popularity.
- Pages built primarily for the search engines.
- Misuse of competitor names.
- Multiple sites offering the same content.
- Pages that use excessive pop-ups, interfering with user navigation.
- Pages that seem deceptive, fraudulent or provide a poor user experience.

Google

- Avoid hidden text or hidden links.
- Don't employ cloaking or sneaky redirects.
- Don't send automated queries to Google.
- Don't load pages with irrelevant words.
- Don't create multiple pages, subdomains or domains with substantially duplicate content.
- Avoid "doorway" pages created just for search engines, or other "cookie cutter" approaches such as affiliate programs with little or no original content.

Frameset Problem

Search engines have a tough time with frames. Using frames can prevent them from finding pages within a Web site or cause them to send visitors into a site without the proper frame "context" being established. Both problems can be corrected, with a little foresight by Webmasters.

Search Engines and Frames: <http://www.searchenginewatch.com/Webmasters/article.php/2167901>

Keyword Tools

Use free tools online to find out what keywords you should use or help you generate a list of keywords. Analyze the amount of keywords you are using in each of your Web pages. You want to aim for a keyword density on your site from 3-20%. This will give you a good range. You can find and analyze your keywords by using these online tools:

1. Keywords Selector: <http://www.keyword-selector-tool.com/>
2. Keywords Generator: <http://freekeywords.wordtracker.com/>
3. Keyword Density: <http://www.keyworddensity.com/3>

Don't repeat your keywords more than 3 times. Use different keywords for each Web page.

Related Glossary Terms

- Backlinks
- Bad Neighborhood
- Cloaking
- Crawler
- Keyword (Key Phrase)
- Link Farm
- Link Popularity
- META data
- Natural Search
- Optimization
- Organic Search
- Page Rank (PR)
- Pay-Per-Click (PPC)
- Pay-Per-Click Search Engines
- Query
- Ranking
- Robot
- Sandboxed
- Search Engine
- Searching
- Search Engine Optimization (SEO)
- Spider

Continued Reading

- SEO Tools 101, Part 1: <http://searchenginewatch.com/3633455>
- Search Engine Submission 101, Part 2: <http://searchenginewatch.com/3633773>
- SEO 101 - Basic Optimization Techniques: <http://www.searchengineguide.com/jim-hedger/seo-101-basic.php>
- Yahoo! Help: <http://help.yahoo.com/us/yahoo/smallbusiness/Webhosting/index.html>
- Google Webmaster Guidelines: <http://www.google.com/support/Webmasters/bin/answer.py?answer=35769>

References:

This information was gathered from the following sources:

- "Web Design: The Complete Reference" published by Osborne/McGraw Hill © 2000
- <http://www.searchenginewatch.com>
- <http://www.Webmasterworld.com>
- <http://Websitehelpers.com/seo/>