



Design and content guidelines

- Make a site with a clear hierarchy and text links. Every page should be reachable from at least one static text link.
- Offer a site map to your users with links that point to the important parts of your site. If the site map is larger than 100 or so links, you may want to break the site map into separate pages.
- Create a useful, information-rich site, and write pages that clearly and accurately describe your content.
- Think about the words users would type to find your pages, and make sure that your site actually includes those words within it.
- Try to use text instead of images to display important names, content, or links. The Google crawler doesn't recognize text contained in images.
- Make sure that your TITLE and ALT tags are descriptive and accurate.
- Check for broken links and correct HTML.
- If you decide to use dynamic pages (i.e., the URL contains a "?" character), be aware that not every search engine spider crawls dynamic pages as well as static pages. It helps to keep the parameters short and the number of them few.
- Keep the links on a given page to a reasonable number (fewer than 100).

Technical guidelines

- Use a text browser such as Lynx to examine your site, because most search engine spiders see your site much as Lynx would. If fancy features such as JavaScript, cookies, session IDs, frames, DHTML, or Flash keep you from seeing all of your site in a text browser, then search engine spiders may have trouble crawling your site.
- Allow search bots to crawl your sites without session IDs or arguments that track their path through the site. These techniques are useful for tracking individual user behavior, but the access pattern of bots is entirely different. Using these techniques may result in incomplete indexing of your site, as bots may not be able to eliminate URLs that look different but actually point to the same page.
- Make sure your web server supports the If-Modified-Since HTTP header. This feature allows your web server to tell Google whether your content has changed since we last crawled your site. Supporting this feature saves you bandwidth and overhead.
- Make use of the robots.txt file on your web server. This file tells crawlers which directories can or cannot be crawled. Make sure it's current for your site so that you don't accidentally block the Googlebot crawler. Visit <http://www.robotstxt.org/wc/faq.html> to learn how to instruct robots when they visit your site.
- If your company buys a content management system, make sure that the system can export your content so that search engine spiders can crawl your site.
- Don't use "&id=" as a parameter in your URLs, as we don't include these pages in our index.

Quality guidelines - basic principles

- 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, which is commonly referred to as "cloaking."

- 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.

Quality guidelines - specific guidelines

- 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.
- Don't create pages that install viruses, trojans, or other badware.
- Avoid "doorway" pages created just for search engines, or other "cookie cutter" approaches such as affiliate programs with little or no original content.



Yahoo! strives to provide the best search experience on the Web by directing searchers to high-quality and relevant web content in response to a search query. Pages Yahoo! Wants Included in Its Index

- 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

Unfortunately, not all web pages contain information that is valuable to a user. Some pages are created deliberately to trick the search engine into offering inappropriate, redundant or poor-quality search results; this is often called "spam." Yahoo! does not want these pages in the index.

What Yahoo! Considers Unwanted

Some, but not all, examples of the more common types of pages that Yahoo! does not want include:

- 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



Guidelines for Successful Indexing

Here are some recommendations that may help MSNBot and other Web crawlers effectively index and rank your site. We've also provided a list of items and techniques that MSN Search discourages.

- Technical recommendations for your website
- Content guidelines for your website
- Items and techniques discouraged by MSN Search

Technical Recommendations for your Website

- Use only well-formed HTML code in your pages. Ensure that all tags are closed, and that all links function properly. If your site contains broken links, MSNBot may not be able to index your site effectively, and people may not be able to reach all of your pages.
- If you move a page, set up the page's original URL to direct people to the new page, and tell them whether the move is permanent or temporary.
- Make sure MSNBot is allowed to crawl your site, and is not on your list of web crawlers that are prohibited from indexing your site.
- Use a robots.txt file or meta tags to control how MSNBot and other web crawlers index your site. The robots.txt file tells web crawlers which files and folders it is not allowed to crawl. The Web Robots Pages provide detailed information on the robots.txt Robots Exclusion standard. This site may be available in English only.
- Keep your URLs simple and static. Complicated or frequently changed URLs are difficult to use as link destinations. For example, the URL www.example.com/mypage is easier for MSNBot to crawl and for people to type than a long URL with multiple extensions. Also, a URL that doesn't change is easier for people to remember, which makes it a more likely link destination from other sites.

Content Guidelines for your Website

The best way to attract people to your site, and keep them coming back, is to design your pages with valuable content that your target audience is interested in.

- In the visible page text, include words users might choose as search query terms to find the information on your site.
- Limit all pages to a reasonable size. We recommend one topic per page. An HTML page with no pictures should be under 150 KB.
- Make sure that each page is accessible by at least one static text link.

- Create a site map that is fairly flat (i.e., each page is only one to three clicks away from the home page). Links embedded in menus, list boxes, and similar elements are not accessible to web crawlers unless they appear in your site map.
- Keep the text that you want indexed outside of images. For example, if you want your company name or address to be indexed, make sure it is displayed on your page outside of a company logo.

Items and Techniques Discouraged by MSN Search

The following items and techniques are not appropriate uses of the index. Use of these items and techniques may affect how your site is ranked within MSN Search and may result in the removal of your site from the MSN Search index.

- Loading pages with irrelevant words in an attempt to increase a page's keyword density. This includes stuffing ALT tags that users are unlikely to view.
- Using hidden text or links. You should use only text and links that are visible to users.
- Using techniques to artificially increase the number of links to your page, such as link farms.