**Meta tags that Google understands**

Meta tags are a great way for webmasters to provide search engines with information about their sites. Meta tags can be used to provide information to all sorts of clients, and each system processes only the meta tags they understand and ignores the rest. Meta tags are added to the <head> section of your HTML page and generally look like this:

<!DOCTYPE html> <html> <head> <meta charset="utf-8"> <meta name="Description" CONTENT="Author: A.N. Author, Illustrator: P. Picture, Category: Books, Price: £9.24, Length: 784 pages"> <meta name="google-site-verification" content="+nxGUDJ4QpAZ5l9Bsjdi102tLVC21AIh5d1Nl23908vVuFHs34="/> <title>Example Books - high-quality used books for children</title> <meta name="robots" content="noindex,nofollow">

Google understands the following meta tags (and related items):

|  |  |
| --- | --- |
| <meta name="description" content="A description of the page" /> | This tag provides a short description of the page. In some situations this description is used as a part of the snippet shown in the search results. [More information](https://support.google.com/webmasters/answer/answer.py?answer=35264) |
| <title>The Title of the Page</title> | While technically not a meta tag, this tag is often used together with the "description". The contents of this tag are generally shown as the title in search results (and of course in the user's browser). [More information](https://support.google.com/webmasters/answer/answer.py?answer=35264) |
| <meta name="robots" content="..., ..." /> <meta name="googlebot" content="..., ..." /> | These meta tags can control the behavior of search engine crawling and indexing. The **robots** meta tag applies to all search engines, while the "googlebot" meta tag is specific to Google. The default values are "index, follow" (the same as "all") and do not need to be specified. We understand the following values (when specifying multiple values, separate them with a comma):   * [noindex](https://support.google.com/webmasters/answer/answer.py?answer=61050): prevents the page from being indexed * [nofollow](https://support.google.com/webmasters/answer/answer.py?answer=96569): prevents the Googlebot from following links from this page * [nosnippet](https://support.google.com/webmasters/answer/answer.py?answer=35304): prevents a snippet from being shown in the search results * [noodp](https://support.google.com/webmasters/answer/answer.py?answer=35264): prevents the alternative description from the ODP/DMOZ from being used * [noarchive](https://support.google.com/webmasters/answer/answer.py?answer=35306): prevents Google from showing the **Cached** link for a page. * unavailable\_after:[date]: lets you specify the exact time and date you want to stop crawling and indexing of this page * [noimageindex](https://support.google.com/webmasters/answer/answer.py?answer=61050): lets you specify that you do not want your page to appear as the referring page for an image that appears in Google search results. * none: is equivalent to noindex, nofollow.   You can now also specify this information in the header of your pages using the "X-Robots-Tag" HTTP header directive. This is particularly useful if you wish to limit indexing of non-HTML files like graphics or other kinds of documents. [More information about robots meta tags](http://code.google.com/web/controlcrawlindex/docs/robots_meta_tag.html" \t "_blank) |
| <meta name="google" content="nositelinkssearchbox" /> | When users search for your site, Google Search results sometimes display a search box specific to your site, along with other direct links to your site. This meta tag tells Google not to show the sitelinks search box. Learn more about [sitelinks search box](https://developers.google.com/webmasters/richsnippets/sitelinkssearch). |
| <meta name="google" content="notranslate" /> | When we recognize that the contents of a page are not in the language that the user is likely to want to read, we often provide a link to a translation in the search results. In general, this gives you the chance to provide your unique and compelling content to a much larger group of users. However, there may be situations where this is not desired. This meta tag tells Google that you don't want us to provide a translation for this page. |
| <meta name="google-site-verification" content="..." /> | You can use this tag on the top-level page of your site to verify ownership for Webmaster Tools. Please note that while the values of the "name" and "content" attributes must match exactly what is provided to you (including upper and lower case), it doesn't matter if you change the tag from XHTML to HTML or if the format of the tag matches the format of your page. [More information](http://www.google.com/support/webmasters/bin/answer.py?answer=35659) |
| <meta http-equiv="Content-Type" content="...; charset=..." /> <meta charset="..." > | This defines the page's content type and character set. Make sure that you surround the value of the content attribute with quotes - otherwise the charset attribute may be interpreted incorrectly. We recommend using Unicode/UTF-8 where possible. [More information](http://code.google.com/webstats/2005-12/metadata.html) |
| <meta http-equiv="refresh" content="...;url=..." /> | This meta tag sends the user to a new URL after a certain amount of time, and is sometimes used as a simple form of redirection. However, it is not supported by all browsers and can be confusing to the user. The W3C [recommends that this tag not be used](http://www.w3.org/TR/WCAG10-HTML-TECHS/#meta-element). We recommend using a server-side 301 redirect instead. |

Other points to note:

* Google can read both HTML and XHTML-style meta tags, regardless of the code used on the page.
* With the exception of verify, case is generally not important in meta tags.

This is not an exclusive list of available meta tags, and you should feel free to use unlisted meta tags if they are important to your site. Just remember that Google will ignore meta tags it doesn't know.

# Keep a simple URL structure

A site's URL structure should be as simple as possible. Consider organizing your content so that URLs are constructed logically and in a manner that is most intelligible to humans (when possible, readable words rather than long ID numbers). For example, if you're searching for information about aviation, a URL like http://en.wikipedia.org/wiki/Aviation will help you decide whether to click that link. A URL like http://www.example.com/index.php?id\_sezione=360&sid=3a5ebc944f41daa6f849f730f1, is much less appealing to users.

Consider using punctuation in your URLs. The URL **http://www.example.com/green-dress.html** is much more useful to us than **http://www.example.com/greendress.html**. We recommend that you use hyphens (-) instead of underscores (\_) in your URLs.

Overly complex URLs, especially those containing multiple parameters, can cause a problems for crawlers by creating unnecessarily high numbers of URLs that point to identical or similar content on your site. As a result, Googlebot may consume much more bandwidth than necessary, or may be unable to completely index all the content on your site.

**Common causes of this problem**

Unnecessarily high numbers of URLs can be caused by a number of issues. These include:

* **Additive filtering of a set of items** Many sites provide different views of the same set of items or search results, often allowing the user to filter this set using defined criteria (for example: show me hotels on the beach). When filters can be combined in a additive manner (for example: hotels on the beach and with a fitness center), the number of URLs (views of data) in the sites explodes. Creating a large number of slightly different lists of hotels is redundant, because Googlebot needs to see only a small number of lists from which it can reach the page for each hotel. For example:
  + Hotel properties at "value rates":

http://www.example.com/hotel-search-results.jsp?Ne=292&N=461

* + Hotel properties at "value rates" on the beach:

http://www.example.com/hotel-search-results.jsp?Ne=292&N=461+4294967240

* + Hotel properties at "value rates" on the beach and with a fitness center:

http://www.example.com/hotel-search-results.jsp?Ne=292&N=461+4294967240+4294967270

* **Dynamic generation of documents**. This can result in small changes because of counters, timestamps, or advertisements.
* **Problematic parameters in the URL.** Session IDs, for example, can create massive amounts of duplication and a greater number of URLs.
* **Sorting parameters.** Some large shopping sites provide multiple ways to sort the same items, resulting in a much greater number of URLs. For example:

http://www.example.com/results?search\_type=search\_videos&search\_query=tpb&search\_sort=relevance &search\_category=25

* **Irrelevant parameters in the URL, such as referral parameters.** For example:

http://www.example.com/search/noheaders?click=6EE2BF1AF6A3D705D5561B7C3564D9C2&clickPage= OPD+Product+Page&cat=79

http://www.example.com/discuss/showthread.php?referrerid=249406&threadid=535913

http://www.example.com/products/products.asp?N=200063&Ne=500955&ref=foo%2Cbar&Cn=Accessories.

* **Calendar issues.** A dynamically generated calendar might generate links to future and previous dates with no restrictions on start of end dates. For example:

http://www.example.com/calendar.php?d=13&m=8&y=2011

http://www.example.com/calendar/cgi?2008&month=jan

* **Broken relative links.** Broken relative links can often cause infinite spaces. Frequently, this problem arises because of repeated path elements. For example:

http://www.example.com/index.shtml/discuss/category/school/061121/html/interview/ category/health/070223/html/category/business/070302/html/category/community/070413/html/FAQ.htm

**Steps to resolve this problem**

To avoid potential problems with URL structure, we recommend the following:

* Consider using a robots.txt file to block Googlebot's access to problematic URLs. Typically, you should consider blocking dynamic URLs, such as URLs that generate search results, or URLs that can create infinite spaces, such as calendars. Using regular expressions in your robots.txt file can allow you to easily block large numbers of URLs.
* Wherever possible, avoid the use of session IDs in URLs. Consider using cookies instead. Check our [Webmaster Guidelines](https://support.google.com/webmasters/answer/answer.py?answer=35769) for additional information.
* Whenever possible, shorten URLs by trimming unnecessary parameters.
* If your site has an infinite calendar, add a [nofollow](https://support.google.com/webmasters/answer/answer.py?answer=96569) attribute to links to dynamically created future calendar pages.
* Check your site for broken relative links.

**Use canonical URLs**

Set a preferred URL for your content

This article describes how you can use canonical URLs to improve link and ranking signals for content available through multiple URL structures or via syndication.

In the world of content management and online shopping systems, it's common for the same content to be accessed through multiple URLs. With content syndication, it's also easy for content to be distributed to different URLs and domains entirely. For example:

|  |  |
| --- | --- |
| The same product page has **dynamic URLs** as a result of user session and/or search preference. | https://www.example.com/products?category=dresses&color=green https://example.com/dresses/cocktail?gclid=ABCD https://www.example.com/dresses/green/greendress.html |
| Your blog system automatically saves **multiple URLs** as you position the same post under multiple sections. | https://blog.example.com/dresses/green-dresses-are-awesome/ https://blog.example.com/green-things/green-dresses-are-awesome/ |
| Your server is configured to serve the **same content for the www subdomain or the http** protocol. | http://example.com/green-dresses https://example.com/green-dresses http://www.example.com/green-dresses |
| Content you provide on that blog for **syndication** to other sites is replicated in part or in full on those domains. | https://news.example.com/green-dresses-for-every-day-155672.html (syndicated post) https://blog.example.com/dresses/green-dresses-are-awesome/3245/ (original post) |

While these systems make it more convenient to develop and distribute content, they cause some challenges when people use search engines to reach your page. For instance:

* **Consolidating link signals for the duplicate or similar content**. It helps search engines to be able to consolidate the information they have for the individual URLs (such as links to them) on a single, preferred URL. This means that links from other sites to http://example.com/dresses/cocktail?gclid=ABCD get consolidated with links to https://www.example.com/dresses/green/greendress.html.
* **Tracking metrics for a single product/topic**. With a variety of URLs, it's more challenging to get consolidated metrics for a specific piece of content.
* **Determining the URL you want people to see.**You prefer people reach your green dresses product page via https://www.example.com/dresses/green/greendress.html rather than https://example.com/dresses/cocktail?gclid=ABCD.
* **Addressing syndicated content.** If you syndicate your content for publication on other domains, you want to consolidate page ranking to your preferred URL.

To address these issues, we recommend you define a canonical URL for content (or equivalent content) available through multiple URLs. You can do this for Google Search in a number of ways:

* [Set your preferred domain](https://support.google.com/webmasters/answer/139066?hl=en#1)
* [Indicate the preferred URL with the rel="canonical" link element](https://support.google.com/webmasters/answer/139066?hl=en#2)
* [Use a sitemap to set preferred URLs for the same content](https://support.google.com/webmasters/answer/139066?hl=en#3)
* [Use 301 redirects for URLs that are not canonical](https://support.google.com/webmasters/answer/139066?hl=en#4)
* [Indicate how to handle dynamic parameters](https://support.google.com/webmasters/answer/139066?hl=en#5)
* [Specify a canonical link in your HTTP header](https://support.google.com/webmasters/answer/139066?hl=en#6)
* [Prefer HTTPS over HTTP for canonical links](https://support.google.com/webmasters/answer/139066?hl=en#https)

While we encourage you to use any of these methods, none of them are required. If you don't indicate a canonical URL, we'll identify what we think is the best version or URL.

**Don't** use the robots.txt file for canonicalization purposes.  
**Don't** use the URL removal tool for canonicalization: it removes all versions of a URL from search.  
**Don't** specify different URLs as canonical for the same page (e.g. one URL in a sitemap and a different URL for that same page using rel="canonical").

### Set your preferred domain

Tell Google which version of your site's URL you prefer for your domain:

* https://www.example.com
* https://example.com

If you set your preferred domain as https://example.com, Google treats links to https://www.example.com exactly the same as links to https://example.com.

Read [Set your preferred domain](https://support.google.com/webmasters/answer/44231) for details.

### Indicate the preferred URL with the rel="canonical" link element

Suppose you want https://blog.example.com/dresses/green-dresses-are-awesome/ to be the preferred URL, even though a variety of URLs can access this content. You can indicate this to search engines as follows:

* **Mark up the canonical page and any other variants with a rel="canonical" link element**.  
  Add a <link> element with the attribute rel="canonical" to the <head> section of these pages:

<link rel="canonical" href="https://blog.example.com/dresses/green-dresses-are-awesome" />

This indicates the preferred URL to use to access the green dress post, so that the search results will be more likely to show users that URL structure. (Note: We attempt to respect this, but cannot guarantee this in all cases.)

**Avoid errors**: use absolute paths rather than relative paths with the rel="canonical" link element.  
  
Use this structure:  https://www.example.com/dresses/green/greendresss.html  
Not this structure: /dresses/green/greendress.html).

### Use a sitemap to indicate preferred URLs for the same content

Pick a canonical (preferred) URL for each of your pages, and tell us about your preference by submitting these canonical URLs in a [sitemap](https://support.google.com/webmasters/answer/156184).

We don't guarantee that we'll use the URLs you submit in a sitemap, but submitting one is a useful way to tell Google about the pages on your site you consider most important.

### Use 301 redirects for URLs that are not canonical

Suppose your page can be reached in multiple ways:

* https://example.com/home
* https://home.example.com
* https://www.example.com

It's a good idea to pick one of those URLs as your preferred (canonical) destination, and use 301 redirects to send traffic from the other URLs to your preferred URL. A server-side 301 redirect is the best way to ensure that users and search engines are directed to the correct page. The 301 status code means that a page has permanently moved to a new location.

### Indicate how to handle dynamic parameters

Use [Parameter Handling](https://support.google.com/webmasters/answer/1235687) to tell Google about any parameters you would like ignored. Ignoring certain parameters can reduce duplicate content in Google's index, and make your site more crawlable. For example, if you specify that the parameter sessionid should be ignored, Google will consider https://www.example.com/dresses/green.php?sessionid=273749 to be the same as https://www.example.com/dresses/green.php.

### Specify a canonical link in your HTTP header

If you can configure your server, you can use rel="canonical" [HTTP headers](http://en.wikipedia.org/wiki/List_of_HTTP_header_fields) to indicate the canonical URL for HTML documents and other files such as PDFs. Say your site makes the same PDF available via different URLs (for example, for tracking purposes), like this:

https://www.example.com/downloads/white-paper.pdf https://www.example.com/downloads/partner-1/white-paper.pdf https://www.example.com/downloads/partner-2/white-paper.pdf https://www.example.com/downloads/partner-3/white-paper.pdf

In this case, you can use a rel="canonical" HTTP header to specify to Google the canonical URL for the PDF file, as follows:

Link: <http://www.example.com/downloads/white-paper.pdf>; rel="canonical"

Google currently supports these link header elements for Web Search only.

### Prefer HTTPS over HTTP for canonical URLs

Google prefers HTTPS pages over equivalent HTTP pages as canonical, except when there are conflicting signals such as the following:

* The HTTPS page has an invalid SSL certificate.
* The HTTPS page contains insecure dependencies.
* The HTTPS page is roboted (and the HTTP page is not).
* The HTTPS page redirects users to or through an HTTP page.
* The HTTPS page has a rel="canonical" link to the HTTP page.
* The HTTPS page contains a noindex robots meta tag

Although our systems prefer HTTPS pages over HTTP pages by default, you can ensure this behavior by taking any of the following actions:

* Add 301 or 302 redirects from the HTTP page to the HTTPS page.
* Add a rel="canonical" link from the HTTP page to the HTTPS page.
* Implement [HSTS](http://en.wikipedia.org/wiki/HTTP_Strict_Transport_Security" \t "_blank).

To prevent Google from incorrectly making the HTTP page canonical, you should avoid the following practices:

* Bad SSL certificates and HTTPS-to-HTTP redirects cause us to prefer HTTP very strongly. Implementing HSTS cannot override this strong preference.
* Including the HTTP page in your sitemap or [hreflang entries](https://support.google.com/webmasters/answer/189077?hl=en" \t "_blank) rather than the HTTPS version.
* Implementing your SSL/TLS certificafe for the wrong host-variant: for example, example.com serving the certificate for www.example.com.  The certificate must match your complete site URL, or be a wildcard certificate that can be used for multiple subdomains on a domain.

If you block a resource with a robots.txt file, be sure to block both the HTTP and HTTPS versions of the resource.

**Use rel=”nofollow” for specific links**

"Nofollow" provides a way for webmasters to tell search engines "Don't follow links on this page" or "Don't follow this specific link."

Originally, the nofollow attribute appeared in the page-level meta tag, and instructed search engines not to follow (i.e., crawl) any outgoing links on the page. For example:

<meta name="robots" content="nofollow" />

Before nofollow was used on individual links, preventing robots from following individual links on a page required a great deal of effort (for example, redirecting the link to a URL blocked in robots.txt). That's why the nofollow attribute value of the rel attribute was created. This gives webmasters more granular control: instead of telling search engines and bots not to follow any links on the page, it lets you easily instruct robots not to crawl a specific link. For example:

<a href="signin.php" rel="nofollow">sign in</a>

#### How does Google handle nofollowed links?

In general, we don't follow them. This means that Google does not transfer PageRank or anchor text across these links. Essentially, using nofollow causes us to drop the target links from our overall graph of the web. However, the target pages may still appear in our index if other sites link to them without using nofollow, or if the URLs are submitted to Google in a Sitemap. Also, it's important to note that other search engines may handle nofollow in slightly different ways.

#### What are Google's policies and some specific examples of nofollow usage?

Here are some cases in which you might want to consider using nofollow:

* **Untrusted content**: If you can't or don't want to vouch for the content of pages you link to from your site — for example, untrusted user comments or guestbook entries — you should nofollow those links. This can discourage spammers from targeting your site, and will help keep your site from inadvertently passing PageRank to bad neighborhoods on the web. In particular, [comment spammers](https://support.google.com/webmasters/answer/answer.py?answer=81749) may decide not to target a specific content management system or blog service if they can see that untrusted links in that service are nofollowed. If you want to recognize and reward trustworthy contributors, you could decide to automatically or manually remove the nofollow attribute on links posted by members or users who have consistently made high-quality contributions over time.
* **Paid links**: A site's ranking in Google search results is partly based on analysis of those sites that link to it. In order to prevent paid links from influencing search results and negatively impacting users, we urge webmasters use nofollow on such links. Search engine guidelines require machine-readable disclosure of paid links in the same way that consumers online and offline appreciate disclosure of paid relationships (for example, a full-page newspaper ad may be headed by the word "Advertisement"). [More information on Google's stance on paid links.](https://support.google.com/webmasters/answer/answer.py?answer=66736)
* **Crawl prioritization**: Search engine robots can't sign in or register as a member on your forum, so there's no reason to invite Googlebot to follow "register here" or "sign in" links. Using nofollow on these links enables Googlebot to crawl other pages you'd prefer to see in Google's index. However, a solid information architecture — intuitive navigation, user- and search-engine-friendly URLs, and so on — is likely to be a far more productive use of resources than focusing on crawl prioritization via nofollowed links.

**Indicate paginated content**

Sites paginate content in various ways. For example:

* News and/or publishing sites often divide a long article into several shorter pages.
* Retail sites may divide the list of items in a large product category into multiple pages.
* Discussion forums often break threads into sequential URLs.

If you paginate content on your site, and you want that content to appear in search results, we recommend one of the following three options.

* **Do nothing.** Paginated content is very common, and Google does a good job returning the most relevant results to users, regardless of whether content is divided into multiple pages.
* **Specify a View All page.** Searchers commonly prefer to view a whole article or category on a single page. Therefore, if we think this is what the searcher is looking for, we try to show the View All page in search results. You can also add a [rel="canonical"](https://support.google.com/webmasters/answer/answer.py?answer=139394) link to the component pages to tell Google that the View All version is the version you want to appear in search results.
* **Use rel="next" and rel="prev" links** to indicate the relationship between component URLs. This markup provides a strong hint to Google that you would like us to treat these pages as a logical sequence, thus consolidating their linking properties and usually sending searchers to the first page.

## Using rel="next" and rel="prev"

You can use the HTML attributes rel="next" and rel="prev" to indicate the relationship between individual URLs. Using these attributes is a strong hint to Google that you want us to treat these pages as a logical sequence.

Let's say you have content paginated into the following URLs:

http://www.example.com/article-part1.html http://www.example.com/article-part2.html http://www.example.com/article-part3.html http://www.example.com/article-part4.html

1. In the <head> section of the first page (http://www.example.com/article-part1.html), add a link tag pointing to the next page in the sequence, like this:

<link rel="next" href="http://www.example.com/article-part2.html">

Because this is the first URL in the sequence, there’s no need to add markup for rel="prev".

1. On the second and third pages, add links pointing to the previous and next URLs in the sequence. For example, you could add the following to the second page of the sequence:

<link rel="prev" href="http://www.example.com/article-part1.html"> <link rel="next" href="http://www.example.com/article-part3.html">

1. On the final page of the sequence (http://www.example.com/article-part4.html>), add a link pointing to the previous URL, like this:

<link rel="prev" href="http://www.example.com/article-part3.html">

Because this is the final URL in the sequence, there’s no need to add a rel="next" link.

Google treats rel="previous" as a syntactic variant of rel="prev". Values can be either relative or absolute URLs (as allowed by the <link> tag). And, if you include a <base> link in your document, relative paths will resolve according to the base URL.

Some things to note:

* rel="prev" and rel="next" act as hints to Google, not absolute directives.
* If a component page within a series includes parameters that don't change the page's content, such as session IDs, then the rel="prev" and rel="next" values should also contain the same parameters. This helps our linking process better match corresponding rel="prev" and rel="next" values. For example, the page http://www.example.com/article?story=abc&page=2&sessionid=123 should contain the following:

<link rel="prev" href="http://www.example.com/article?story=abc&page=1&sessionid=123" /> <link rel="next" href="http://www.example.com/article?story=abc&page=3&sessionid=123" />

* rel="next" and rel="prev" are orthogonal concepts to rel="canonical". You can include both declarations. For example, http://www.example.com/article?story=abc&page=2&sessionid=123 may contain:

<link rel="canonical" href="http://www.example.com/article?story=abc&page=2"/> <link rel="prev" href="http://www.example.com/article?story=abc&page=1&sessionid=123" /> <link rel="next" href="http://www.example.com/article?story=abc&page=3&sessionid=123" />

* If Google finds mistakes in your implementation (for example, if an expected rel="prev" or rel="next" designation is missing), we'll continue to index the page(s), and rely on our own heuristics to understand your content.

**Tag site for cild-directed treatment**

Visit the [Tag for Child Directed Treatment](https://www.google.com/webmasters/tools/coppa" \t "_blank) page to tag a site or service that you would like Google to treat as child-directed in whole or in part for the purposes of the [Children’s Online Privacy Protection Act](http://business.ftc.gov/privacy-and-security/childrens-privacy) (COPPA). If you have not already added a site to Webmaster Tools, you must first [add the site and verify ownership](https://support.google.com/webmasters/bin/answer.py?answer=34592).

Keep in mind the following:

* You can tag an entire domain or portions of a domain (subdomain or subdirectory) for treatment as child-directed
* Any pages beneath a domain or directory are also covered by the tag.
* It may take some time for this designation to take effect in applicable Google services.
* Google may limit the number of domains or sub-domains you may include at any time.

For finer control over how your content is treated, you can also tag individual ad units for treatment as child-directed. See the following pages for specific product information: [AdSense](https://support.google.com/adsense/answer/3248194?hl=en" \t "_blank), [AdX](https://support.google.com/adxseller/answer/4442399?hl=en" \t "_blank), and [DoubleClick](https://support.google.com/dfp_sb/answer/3721907" \t "_blank).

**Note:**Tagging individual ad units has immediate effect on the ad served. An ad unit tag takes precedence over any applicable site-level settings.