# Warning

Techniques (code/configuration) given in this document are ignored by malicious crawlers such as those that harvest data and email addresses or those that scan for security vulnerabilities. Do not rely on these techniques to hide content that should be secured.

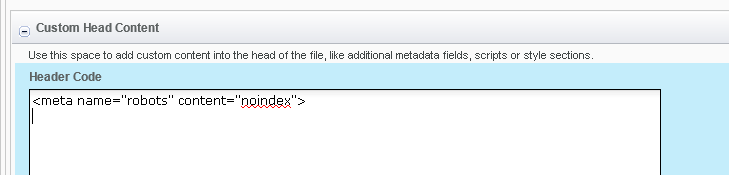
# How can I hide a page on our external site from search engines? How can I hide a page from showing up in search results on Google, Bing, Yahoo, etc.?

## Using HTML meta tag

We can tell search engines not to index a page by using robots meta tag in head section of the HTML page. The value of this tag should be either noindex or none (which is equivalent to noindex, nofollow). No Follow will tell search engines not to follow and index links on that page.

<meta name="robots" content="noindex">

For pages generated using page\_v2, popup\_v2 and the new responsive TeamSite templates, robots meta tag should go in the Custom Head Content -> Header Code section as:



## Using robots.txt file

Robots.txt is a text file used by web sites to instruct web robots (typically search engine robots) how to crawl pages on their website. It indicates whether certain user agents (web-crawling software) can or cannot crawl parts of a website.

Pages, files and directories can be disallowed for search engines from robots.txt file. We maintain robots file at the document root level for all domains. Contact Sherry or the AD team to update the robots file for specific requests.

## Using both meta tag and robots.txt file

In most cases, it makes sense to update both the page robots meta tag and the robots file for hiding a page as this will keep the page hidden in case it is renamed after initial deployment without an update to robots file. Also, this will also cover situations where content deployment and robots file updates are on different dates.

A page, image or document can also be referenced from other sites. Also, we may copy content, such as corporate and Multifamily blogs, across branches/directory and deploy to alternate locations. Robots meta tag in source content will keep content hidden from search engines in such cases.

## Using web server configuration

Another method to disallow search engines to index pages is using X-Robots-Tag in HTTP response codes.

# How do I hide a non-HTML file (PDF, image, Word, Excel, etc.) from search engines?

## Using robots.txt file

Non-HTML files such as PDF, Word, Excel documents, images, XML, scripts and other static files do not have HTML meta tags in them. These files can be hidden from search engines by disallowing them in robots.txt file.

We can also prevent search engines from indexing directories through robots.txt file. We already exclude cgi-bin, errors, include, etc. directories form indexing this way.

# How do I hide a page, site section or entire site from search engines?

## Using web server configuration

We can use web server (Apache in our case) configuration to hide a page, site section, directory or the entire site from search engine. This may not be the easiest approach for existing pages, directory or site sections but it could be very useful for building, testing, retiring or migrating site content.

In addition to regular HTTP response content, web servers can send additional X-Robots-Tag headers in their response. X-Robots-Tag headers tell search engines whether to index content; it can even tell content availability and expiration dates.

We have not tried using X-Robots-Tag for our web sites yet.

# Does “hiding a page/file from Search engines” make it invisible on the web?

No, search engines may still find our content and rank it. However, most search engines will not display it in their search results.

Top search engines, Google, Bing and Yahoo serve about 95% of web searches. They respect meta tags, robots file and HTTP headers. However other search engines and crawlers may completely ignore above directives. Protecting web content through password is the only way of preventing it from rouge search bots.

# References

Information on meta tags, robot file, indexing and how to remove indexed pages from search results:

<https://support.google.com/webmasters/answer/93710?hl=en&ref_topic=4598466>  
<https://developers.google.com/search/reference/robots_meta_tag>

<https://www.bing.com/webmaster/help/how-can-i-remove-a-url-or-page-from-the-bing-index-37c07477>