

Robots

What is a robots.txt file?

A robots.txt file is a file placed at the root of a website and contains instructions to web robots. In order to understand information from and about your site, search engines like Google and Bing use programs called “robots” (also known as bots or spiders) to retrieve web documents, index content from those documents, and follow hyperlinks in order to discover new documents. This process is called “crawling” a site. It’s a completely automated process and is necessary in order for search engines to be able to display your site’s content in their results pages.

Webmasters can instruct search engines to *not* crawl specific pages or directories by listing them in their website’s robots.txt file. Common exclusions are admin pages, includes, non-public file libraries, e-commerce checkout pages, duplicates (for example “print versions” of content) and any type of page that robots should not visit or display in search results.

This section demonstrates how to assign a robots.txt file to your Bento pages.

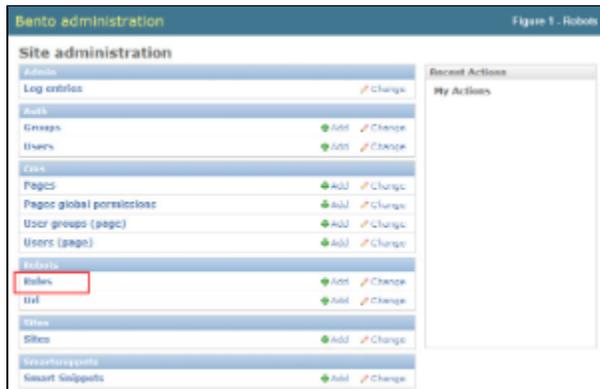
How to add rules to your Bento pages



Station Bento has a default robots.txt file that disallows the /admin/ folder. If this file is already assigned to your site, you only need to edit it if additional pages need to be blocked.

- From the main Bento administration dashboard, in the Robots section, click **Rules** (Figure 1).

Figure 1



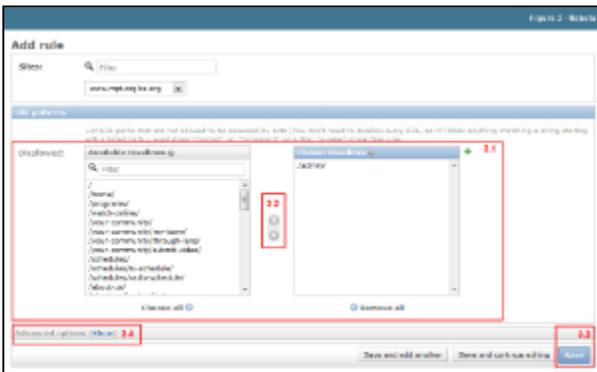
- If you have access to more than one Bento site, click the **site** to which you want to add a rule (Figure 2.1).
- If you have multiple sites, you can use the **search textbox** at the top of the page to search for the site you want (Figure 2.2).

Figure 2



- All of the pages on your site appear in the Available Disallows selectbox (Figure 3.1). Click the **folder(s)** or **page(s)** you want to disallow. Hold your Control or Command button down to select more than one.
- Click the **right facing arrow** to move the page to the Chosen Disallows selectbox (Figure 3.2). All pages in the selected folder will be disallowed so you do not need to select and move each page. For example, if you select /contact/, the following pages will also be disallowed: /contact /directions, /contact/directory, /contact/office-hours.
- Click **Save** (Figure 3.3).
- If you need to adjust the crawler access time, click **Show** next to *Advanced options* and type the **time in seconds** that you want your pages to be accessed by crawlers (Figure 3.4).

Figure 3



✔ Your website now contains a robots.txt file and there is nothing more you need to do.