



Essay Mantra

Work Sample

Configuration guide covering robots.txt crawler directives and .htaccess server rules within a SaaS platform's Settings application.

Table of Contents

Table of Contents	1
Overview	3
Main Dashboard	3
Feature List	3
Field Page	3
Accessing the 'Web Settings' Tab in the Settings Application.....	4
Understanding the 'robots.txt' File	6
Setting Up the 'robots.txt' File for a Website.....	7
Understanding the '.htaccess' File	8
Setting Up the '.htaccess' File for a Website	9

Robots.txt and .htaccess Configuration Guide

Overview

This user guide details the setup process for the `'robots.txt'` and `'htaccess'` files on a website using the `'robots.txt'` tab and the `'htaccess'` tab under 'Web Settings' in the 'Settings' application.

The `'robots.txt'` file communicates crawling permissions to search engines, while the `'htaccess'` file controls how web servers handle requests to a website, including redirects, access restrictions, and related behavior.

The `'robots.txt'` tab and the `'htaccess'` tab are discussed in detail below.

Main Dashboard

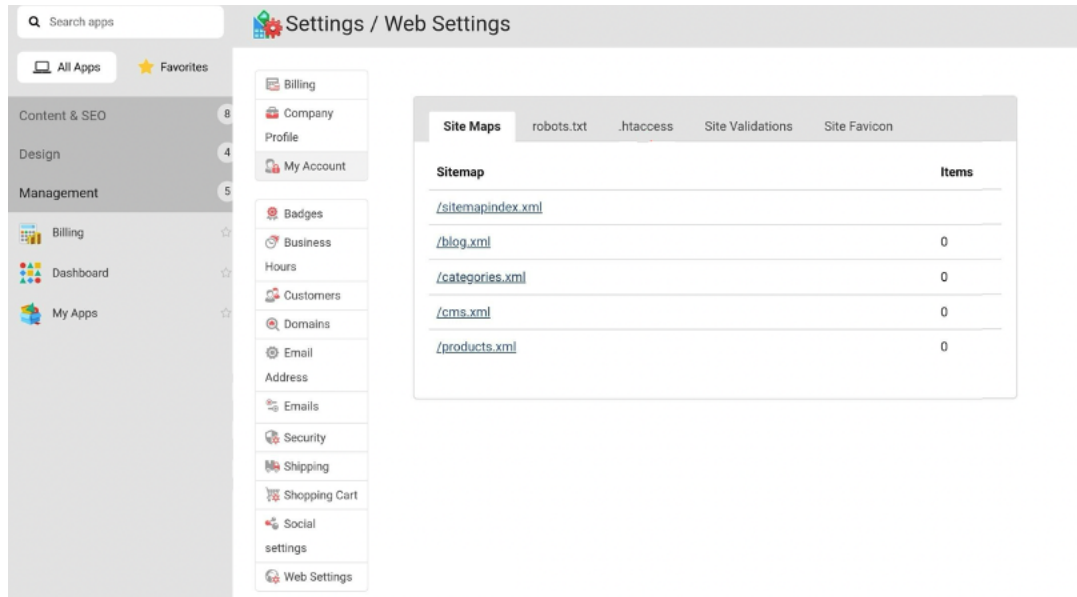
- The 'Settings' application's main dashboard opens upon accessing the application.
- It includes a list of features displayed on the left side, and a field page on the right.

Feature List

- All the features of the 'Settings' application are listed on the left side of the app's main dashboard.
- Clicking on any of the listed features displays the corresponding fields on the field page.
- For example, users can click on the 'Web Settings' feature to access the feature's field page.

Field Page

- All the fields that correspond to a selected feature are displayed on the right side of the main dashboard, i.e., the field page.
- Clicking on the 'Web Settings' feature will reveal a set of tabs, including `'robots.txt'` and `'htaccess'`, along with buttons and fields that allow users to edit how a website is indexed, accessed, monitored, validated, and audited.

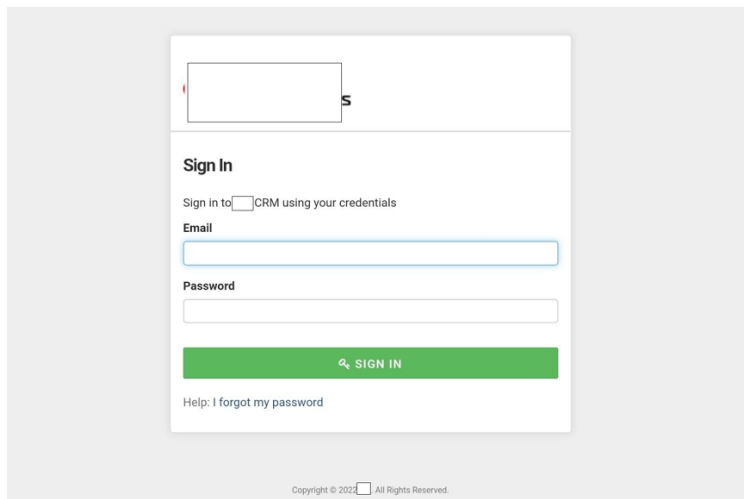


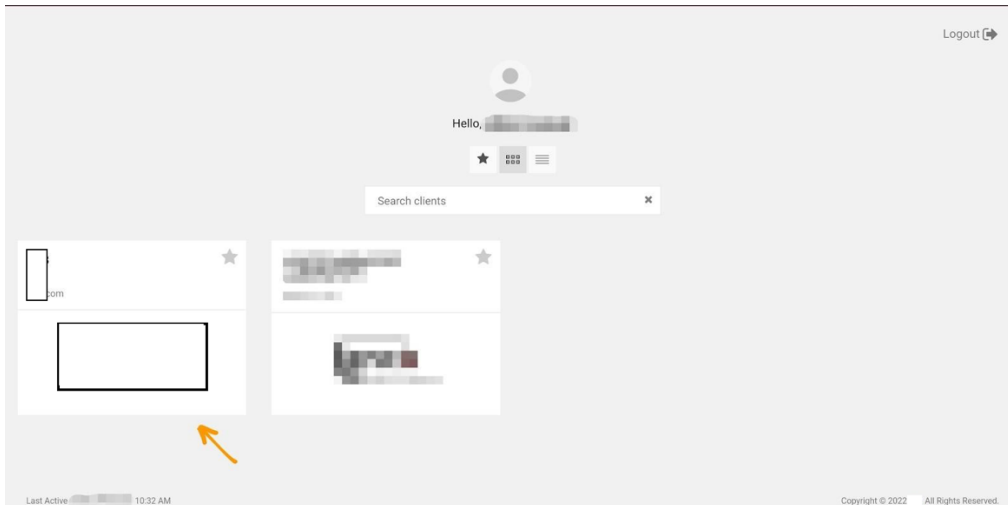
Accessing the 'Web Settings' Tab in the Settings Application

The steps below apply identically to the '`robots.txt`' and '`.htaccess`' configuration process.

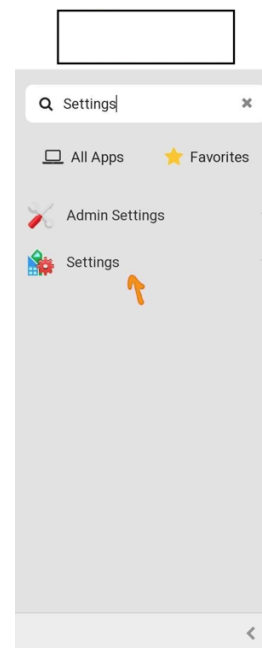
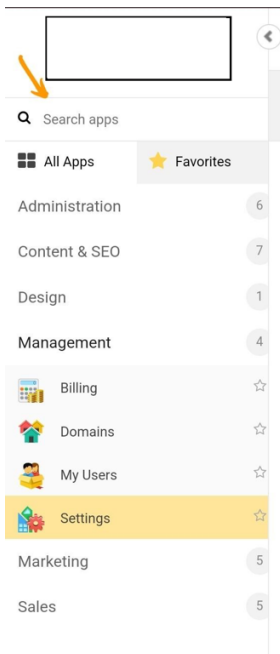
Step 1: Accessing the Settings application

1. Log in to the CRM application (the company's internal tools platform) and open the main page.

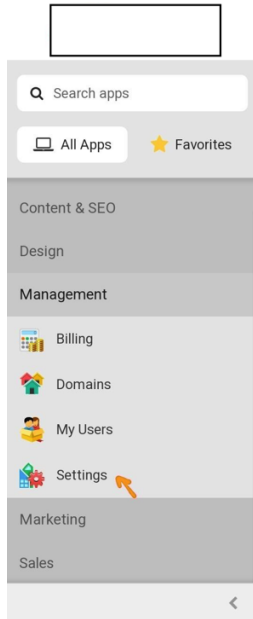




2. Use the search bar to search for 'Settings' and click on the application from the search results.

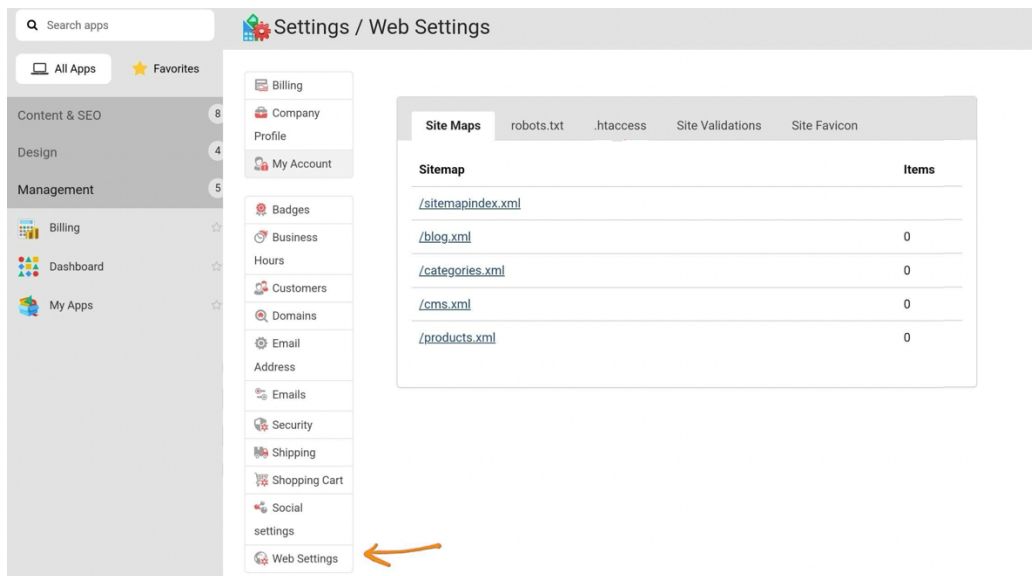


3. Alternatively, users can also click on the app under the 'All apps' menu.



Step 2: Access the 'Web Settings' feature and its corresponding field page on the 'Settings' application's main dashboard

1. Find the 'Web Settings' feature from the feature list displayed on the 'Settings' application's main dashboard and click on it to pull up its field page and its various tabs.



Understanding the 'robots.txt' File

- The 'robots.txt' file or the 'Robots Exclusion Standard' contains a set of soft restrictions that the website uses to communicate with search engine bots and crawlers.

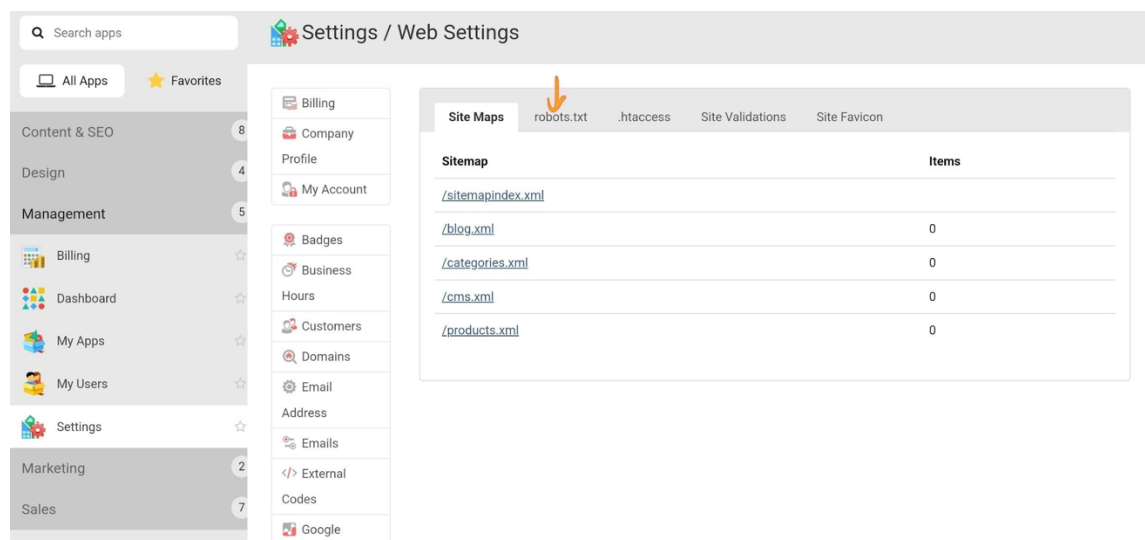
The file relays information about the URLs and areas of the website that are accessible, and those that are not.

- The file also helps manage traffic and prevents the website from becoming overloaded with automated requests when facing heavy traffic. It also contains the location of the website's sitemap.
- In the Settings application, the `'robots.txt'` tab is the second tab in the 'Web Settings' field page.
- The `'robots.txt'` tab contains the following fields and buttons:
 - **Text space:** This field allows the user to edit the content and the directives of the `'robots.txt'` file for the website.
 - **Save:** The 'Save' button allows users to save the changes they make to the `'robots.txt'` file and tab.
- The most commonly used directives in `'robots.txt'` files are as follows:
 - `'User-agent: *'`: This directive targets all search engine crawlers. A specific crawler (e.g., Googlebot) can be targeted by replacing the asterisk with its name.
 - `'Disallow: /admin/'`: This directive blocks crawlers from indexing the specified directory (in this case, the 'admin' directory). However, an empty 'Disallow' line permits access to the entire site.
 - `'Allow: /public/'`: This directive grants access to a specific path. It is typically used to specify an exception within a broader disallowed directory.
 - `'Sitemap: https://www.example.com/sitemap.xml'`: This directive points crawlers to the site's sitemap by providing its URL, helping search engines discover and index pages more efficiently.

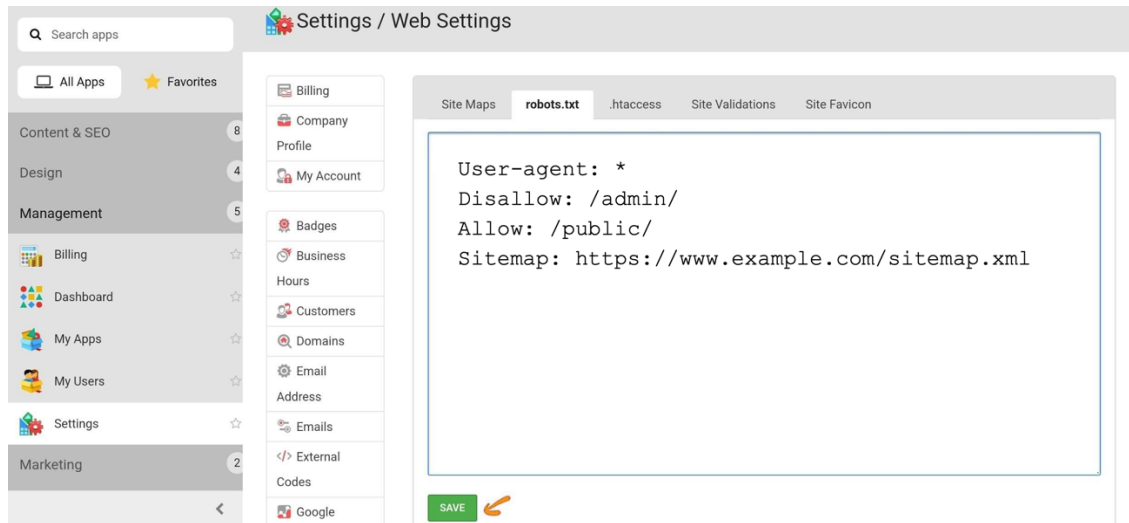
Setting Up the 'robots.txt' File for a Website

Step 1: Access the `'robots.txt'` tab on the 'Web Settings' feature's field page

1. Click on `'robots.txt'` to open the `'robots.txt'` tab.



2. Edit the `'robots.txt'` file using the provided text space.
3. Click on the 'Save' button to save all the edits made to the `'robots.txt'` file after adding the necessary directives and commands. All saved edits will be reflected in the website's `'robots.txt'` file automatically.



Note: A misplaced '/' after 'Disallow:' can deindex the entire site rather than one folder. Please confirm changes with the SEO lead before publishing to production.

Understanding the '.htaccess' File

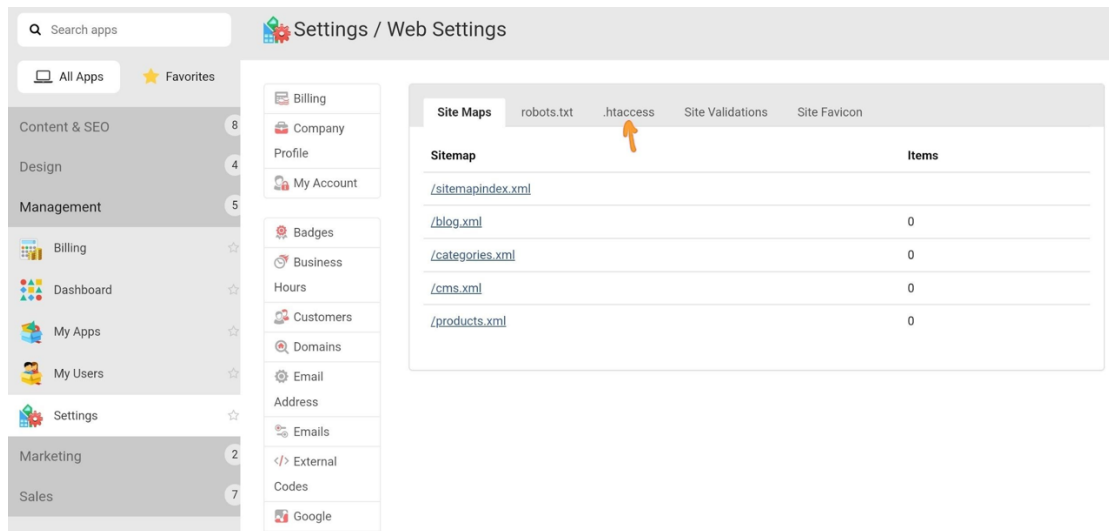
- The `'robots.txt'` file or the 'Hypertext Access' file allows users to monitor and handle various aspects of how a website operates and interacts with requests.
- Each line of code in an `'robots.txt'` file contains instructions that relate to the website's behavior. The file also contains hard restrictions for search engine crawlers and prevents them from accessing specified parts of the website.
- With the `'robots.txt'` file, users can:
 - redirect customers to another page,
 - direct all traffic to the secure 'HTTPS' protocol from the 'HTTP' protocol,
 - password-protect their website, and
 - block or rate-limit specific crawlers outright, preventing the website from being overloaded with requests.
- In the Settings application, the `'robots.txt'` tab is the third tab in the 'Web Settings' field page.
- The `'robots.txt'` tab contains the following fields and buttons:
 - **Text space:** This field allows the user to edit the content and the directives of the `'robots.txt'` file for the website.
 - **Save:** The 'Save' button allows users to save the changes they make to the `'robots.txt'` file and tab.
- The most commonly used directives in `'robots.txt'` files are as follows:

- `'RewriteEngine On'` : This directive enables the rewrite engine, and is required for any `RewriteRule` directive to function.
- `'RewriteCond %{HTTPS} off
RewriteRule ^(.*)$ https://%{HTTP_HOST}%{REQUEST_URI}
[L,R=301]'` : This directive is a rewrite rule that redirects all HTTP traffic to HTTPS. The first line of the code checks whether the connection is not secure before the rule applies, preventing a redirect loop.
- `'<Files "config.php"> Require all denied </Files>'` : This directive blocks direct access to a specific file. It is commonly used to protect sensitive configuration files from being requested directly by a browser.
- `'ErrorDocument 404 /custom-404.html'` : This directive displays a specified custom page (`'custom-404.html'` in the above example) when a visitor requests a URL that doesn't exist, instead of the server's default error page.

Setting Up the '.htaccess' File for a Website

Step 1: Access the '.htaccess' tab on the 'Web Settings' feature's field page

1. Click on '.htaccess' to open the '.htaccess' tab.



2. Edit the '.htaccess' file using the provided text space.
3. Click on the 'Save' button to save all the edits made to the '.htaccess' file on the tab after adding the necessary directives and commands. All saved edits will be reflected in the website's '.htaccess' file automatically.

Q Search apps

All Apps Favorites

Content & SEO8

Design4

Management5

Billing☆

Dashboard☆

My Apps☆

My Users☆

Settings☆

Marketing2

<

Settings / Web Settings

Billing

Company Profile

My Account

Badges

Business Hours

Customers

Domains

Email Address

Emails

External Codes

Google

Site Maps

robots.txt

.htaccess

Site Validations

Site Favicon

```
RewriteEngine On
RewriteCond %{HTTPS} off
RewriteRule ^(.*)$ https://%{HTTP_HOST}%{REQUEST_URI} [L,R=301]

<Files "config.php">
Require all denied
</Files>
```

SAVE

Note: An incorrect `.htaccess` directive can block legitimate traffic or make the site inaccessible. Please make sure to test your code in staging before publishing to production.