

Simsage Confluence Crawler

Overview

This document details the steps required to configure a Crawler to search across some or all Spaces inside a Confluence Cloud installation.

1. Create API token for the crawler

To be able to access the spaces inside Confluence the Crawler will need to have a user and an API token configured with access rights to all spaces that will be crawled. We recommend to use an account in Confluence with the sole responsibility as the SimSage crawler.

- Create an internal User with an email address to use for the crawler in your organisation
- Invite the user to the Confluence site
- Assure the user has SiteAdmin rights
- Log in as the user and create an API token

<https://support.atlassian.com/atlassian-account/docs/manage-api-tokens-for-your-atlassian-account/>

2. Configure Crawler

- Add a new Crawler and select 'Confluence Crawler' from the drop down menu
- Configure the crawler:

The screenshot shows the SimSage web application interface. On the left is a sidebar with navigation links: organisations, knowledge bases, user manager, group manager, inventory, document sources (highlighted), bot, synonyms, semantics, syn-sets, categorization, text to search, reports, and log viewer. The main content area displays the 'Edit Crawler' modal for a crawler named 'Confluence'. The modal has tabs for 'general', 'confluence crawler', 'metadata', 'ACLs', 'processors', and 'schedule'. The 'confluence crawler' tab is active, showing fields for 'Confluence Base Url' (https://simsagedev.atlassian.net), 'Confluence User' (ingo@simsage.co.uk), and 'Access Token' (masked with dots). Below these are fields for 'Categories to crawl' (knowledge-bases), 'Spaces to crawl' (~557058daa63b02e622449e96c6c0eeb68956fe,SD), and 'Spaces to exclude' (a list of space keys separated by commas to exclude from crawling (leave empty not to exclude any spaces)). At the bottom right of the modal are 'cancel' and 'save' buttons.

■ **Confluence Base Url**

This is the domain part of the URL used to access Confluence or other Atlassian cloud products.

- Confluence User
This is the user email configured above under which the crawler will run and which created the API token
- Access Token
The access token created for this user-name
- Categories to crawl
Comma separated list of categories assigned to spaces such as 'knowledge-bases'
If empty, all categories will be included, otherwise only spaces with these categories will be crawled
- Spaces to crawl (Please add the key of the space and not the space name)
Comma separated list of Spaces to crawl.
If empty all spaces (subject to the other filters) will be crawled, otherwise these spaces will be added to the list of spaces to crawl. **Note, this setting is additive to categories, e.g. if categories are entered, all of these plus the added spaces will be crawled.**
- Spaces to exclude
Comma separated list of spaces to exclude.
Any spaces added will not be crawled, even if they match any of the above filters.