

IBM Connections Introduction

The IBM Connections connector will crawl content from IBM Connections server.

On this page

- [Features](#)
- [Content Retrieved](#)
 - [Include](#)
- [Operation Mode](#)
- [Limitations](#)

Related pages

- [IBM Connections Prerequisites](#)
- [IBM Connections How To Configure](#)
- [IBM Connections FAQ & Troubleshooting](#)
- [IBM Connections For Developers](#)

Features

The IBM Connections connector will crawl content from any IBM Connections applications (Activities, Blogs, Bookmarks, Communities, Files, Forums, Profiles, Wikis), also will user's status and Events.

Note that the IBM Connections connector is not part of Aspire Enterprise bundle, however it may be purchased separately.

Some of the features of the IBM Connections connector include:

- Performs incremental crawling (so that only new/updated/deleted documents are indexed)
- Metadata extraction.
- Fetches access control lists (ACLs) for document-level security (For this feature you need to configure the [LDAP Cache service](#) - Make sure you enable Return user's GUID option).
- Is search engine independent
- Runs from any machine with access to the given IBM Connections Site
- Filter the crawled documents by file names using regex patterns.

Content Retrieved

The IBM Connections connector retrieves several types of documents, listed below are the inclusions and exclusions of these documents.

Include

 **Activities**

 **Blogs**

 **Bookmarks**

 **Files**

Forums

Wikis

Profiles

Communities

Communities Events

User Status

Operation Mode

The connector will use the Seedlist service provider interface (SPI) provided with IBM Connections to integrate a search engine with IBM Connections content over HTTP or HTTPS to communicate. The connector acquires content by doing the following:

- Send a GET request to the seedlist feed for all application whose data you want to crawl.

`http://<servername>/activities/seedlist/myserver`

`http://<servername>/blogs/seedlist/myserver`

`http://<servername>/dogear/seedlist/myserver`

`http://<servername>/communities/seedlist/myserver`

`http://<servername>/files/seedlist/myserver`

`http://<servername>/forums/seedlist/myserver`

`http://<servername>/profiles/seedlist/myserver`

`http://<servername>/wikis/seedlist/myserver`

`http://<servername>/dm/seedlist/myserver`

`http://<servername>/news/seedlist/myserver`

`http://<servername>/communities/calendar/seedlist/myserver`

- Process the returned feed. Find the rel=next link and send a GET request to the web address specified by its href attribute.
- Repeat the previous two steps until the response includes a <wplc:timestamp> element in its body.
- Store the value of the <wplc:timestamp> element; you must pass that value as a parameter when you perform a subsequent crawl of the data.

Limitations

Due to API limitations, the IBM Connections connector has the following limitations:

- The SPI list used to crawl does not report comments or likes as changes during incremental crawls for Files application.
- When a user's profile is deleted, IBM SPI reports the change as an update, and not a delete.