eRoom How To Configure

Step 1. Launch Aspire / Open the Content Source Management Page

Launch Aspire (if it's not already running). See:

- Launch Control
- Browse to: http://localhost:50505. For details on using the Aspire Content Source Management page, please refer to Admin UI

Step 2. Add a new eRoom Content Source

To specify exactly what shared folder to crawl, we will need to create a new "Content Source".

To create a new content source:

1. From the Content Source , click on "Add Source" button.
2. Click on "eRoom Connector".

Step 2a. Specify Basic Information

In the "General" tab in the Content Source Configuration window, specify basic information for the content source:

1. Enter a content source name in the "Name" field.
    a. This is any useful name which you decide is a good name for the source. It will be displayed in the content source page, in error messages, etc.
2. Click on the **Scheduled** pulldown list and select one of the following: *Manually, Periodically*, *Daily, Weekly or Advanced*.
    a. Aspire can automatically schedule content sources to be crawled on a set schedule, such as once a day, several times a week, or periodically (every N minutes or hours).For the purposes of this tutorial, you may want to select Manually and then set up a regular crawling schedule later.
3. Click on the **Action** pulldown list to select one of the following: *Start, Stop*, *Pause,* or *Resume*.
    a. This is the action that will be performed for that specific schedule.
4. Click on the **Crawl** pulldown list and select one of the following: *Incremental, Full*, *Real Time,* or *Cache Groups.*
    a. This will be the type of crawl to execute for that specific schedule.

After selecting a Scheduled, specify the details, if applicable:

- *Manually*: No additional options.
- *Periodically*: Specify the "Run every:" options by entering the number of "hours" and "minutes."
- *Daily*: Specify the "Start time:" by clicking on the hours and minutes drop-down lists and selecting options.

- *Weekly*: Specify the "Start time:" by clicking on the hours and minutes drop-down lists and selecting options, then clicking on the day checkboxes to specify days of the week to run the crawl.
- *Advanced*: Enter a custom CRON Expression (e.g. 0 0 0 ? * *)

> (i) You can add more schedules by clicking in the **Add New** option, and rearrange the order of the schedules.

> (i) If you want to disable the content source just unselect the the "Enable" checkbox. This is useful if the folder will be under maintenance and no crawls are wanted during that period of time.

> ⚠ Real Time and Cache Groups crawl will be available depending of the connector.

Step 2b. Specify the Connector Information

In the "Connector" tab, specify the connection information to crawl the eRoom.

1. Enter the eRoom URL you want to crawl.(check eRoom Urls)
2. Enter the account info for the crawl user (username and password).
3. Check on the other options as needed:
   a. Include/Exclude patterns: Enter regex patterns to include or exclude items.
   b. Scan Recursively: Scan through container's child nodes.
   c. Index Containers: index sites, lists and folders. If unchecked, only list items will be indexed.
   d. Use Group Expansion: Specify the eRoom Url, username, password and the schedule for the group expansion process.
   e. Use SSL: Specify the path of the keystore file and the password when you try to crawl a https url.

eRoom URLs

A eRoom URL is needed to tell the connector application what to crawl. This should be the URL of one eRoom in the eRoom server.

It must have the following format: http://**<server>**/*eRoom*/**<facility_name>**/**<eroom_name>**/

Examples:

- http://searchTech/eRoom/Test_Facility/TEST_EROOM/ will crawl the complete eRoom named TEST_EROOM
- http://corp.emc.com/eRoom/emc1/eRoomGuide/ will crawl the complete eroom named eRoomGuide

---

> ⊘ An url of an item like this:
>
>   http://win-cca7ctfki2n/eRoom/Test_Facility/TEST_EROOM/0_99b/
>
> > ⚠ *Dev Comment: Does not work for the crawl. Do we need the complete ACL hierarchy of the parent items or the specific item in order to generate the item ACLs. (We don't have a way to get the parent of an item.)*

Step 2c. Specify Workflow Information

In the "Workflow" tab, specify the workflow steps for the jobs that come out of the crawl. Drag and drop rules to determine which steps should an item follow after being crawled. This rules could be where to publish the document or transformations needed on the data before sending it to a search engine. See W orkflow for more information.

1. For the purpose of this tutorial, drag and drop the *Publish To File* rule found under the *Publishers* tab to the **onPublish** Workfl ow tree.
   a. Specify a *Name* and *Description* for the Publisher.
   b. Click *Add*.
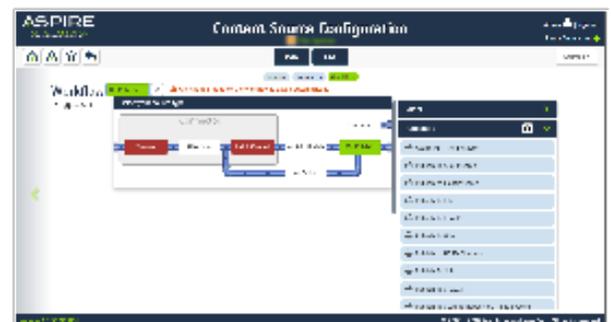
After completing this steps click on the **Save** then **Done** and you'll be sent back to the Home Page.

Step 3: Initiate a Full Crawl

Now that the content source is set up, the crawl can be initiated.

1. Click on the crawl type option to set it as "Full" (is set as "Incremental" by default and the first time it'll work like a full crawl. After the first crawl, set it to "Incremental" to crawl for any changes done in the repository).
2. Click on the Start button.

During the Crawl

During the crawl, you can do the following:

- Click on the "Refresh" button on the Content Sources page to view the latest status of the crawl.
  The status will show **RUNNING** while the crawl is going, and **CRAWLED** when it is finished.
- Click on "Complete" to view the number of documents crawled so far, the number of documents submitted, and the number of documents with errors.

If there are errors, you will get a clickable "Error" flag that will take you to a detailed error message page.

Step 4: Initiate an Incremental Crawl

If you only want to process content updates from the eRoom (documents which are added, modified, or removed), then click on the "Incremental" button instead of the "Full" button. The eRoom connector will automatically identify only changes which have occurred since the last crawl.

If this is the first time that the connector has crawled, the action of the "Incremental" button depends on the exact method of *change* discovery. It may perform the same action as a "Full" crawl crawling everything, or it may not crawl anything. Thereafter, the Incremental button will only crawl updates.

> ⓘ Statistics are reset for every crawl.

Group Expansion

Group expansion configuration is done on the Connector tab.

1. Scroll down to Group Expansion and click the checkbox.
2. Add a new source for each repository you want to expands groups from (you'll need administrator rights on all of them to be able to do this).
3. Set the default domain, user name and password of the crawl account.
4. Set an schedule for group expansion refresh and cleanup.
5. As an optional setting click on the "Use external Group Expansion" checkbox to select an LDAP Cache component for LDAP group expansion. See more info on the LDAP Cache component on LDAP Cache