NETVIBES | Exalead CloudView

CloudView CV23

Proxem Connector

Table of Contents

Proxem Connector

The Exalead CloudView Proxem connector crawls Proxem objects, offering a full search experience.

Use one Exalead CloudView instance of the Proxem connector for one Proxem project.

Audience

This document explains how to configure and deploy Exalead CloudView Proxem connector. We assume that the reader has working knowledge of:

- The operating system on which the Exalead CloudView server and the connectors are installed.
- The Proxem source structure.

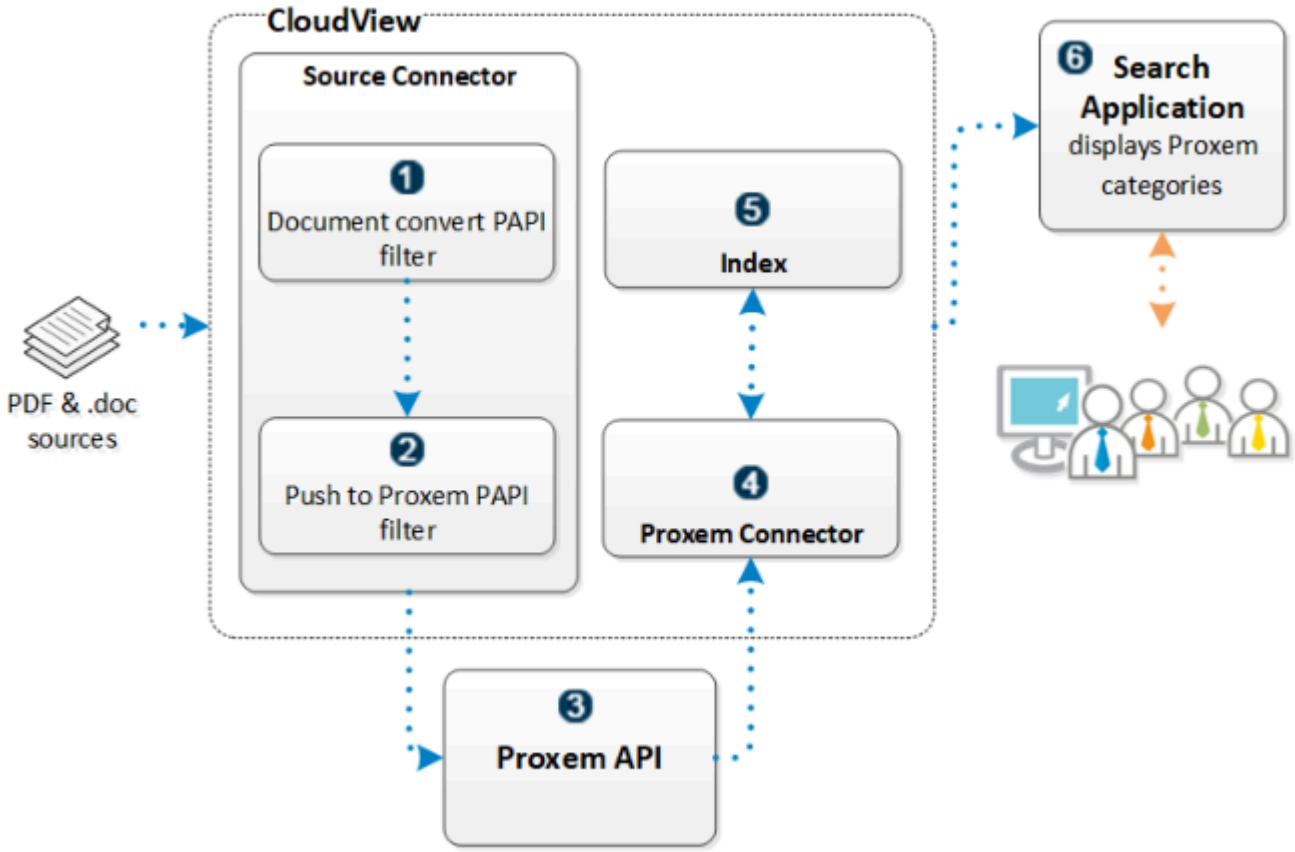
This document assumes that the reader is familiar with the Exalead CloudView Administration Console.

Compatibility

This connector is available for Exalead CloudView V6R2021x.FD04 and higher.

Standard Workflow

Standard Workflow



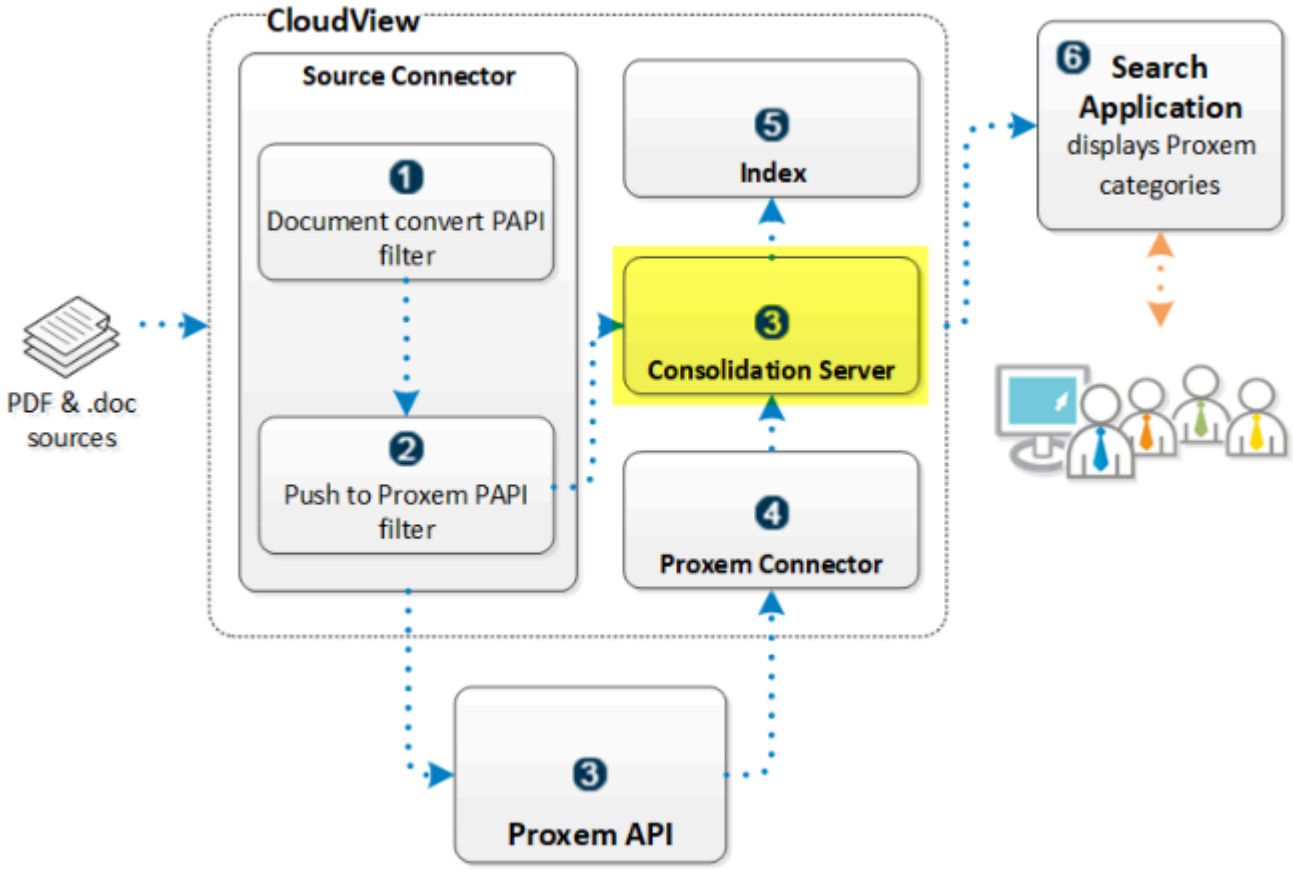
1	In the source connector, add a **Document convert PushAPI Filter** to convert binary files into text. For more information, see Configuring the Source Connector .
2	In the source connector, add a **Push to Proxem** PushAPI filter (installed by the first plugin) to specify the Proxem project for which you want to index analyzed documents. **Note:** Use the **Discard Document** document processor to prevent Exalead CloudView from indexing documents. You only want to index documents analyzed by Proxem. For more information, see Discarding Source Connector Documents .

3	Proxem analyzes the documents pushed by the source connector **Push to Proxem** PushAPI filter.
4	Configure the Proxem connector. For more information, see Configuring the Proxem Connector .
5	Configure the Exalead CloudView data model, to create the equivalent of the Proxem categories. For more information, see Configuring the Exalead CloudView Data Model .
6	In your search UI, verify that Proxem categories are displayed as expected.

Worflow with Consolidation Server

The Proxem analysis may take some time to complete. This workflow shows the use of the Consolidation Server to retrieve documents directly from the Source Connector, and replace these documents by the analyzed documents when they are ready.

Consolidation Server Workflow



3	The Consolidation Server receives documents directly from the Source connector. It allows you to avoid waiting for the end of the Proxem analysis, which takes more time. **Note:** For this step, the source connector pushes text only, and not the binaries. When the Proxem analysis is complete, the Proxem Connector sends analyzed documents to the Consolidation Server. The Consolidation Server then keeps the newly arrived documents, and deletes the documents first pushed after step 2. For more information, see Configuring the Consolidation Server.

Supported Features

The Proxem connector supports the following features:

Crash Recovery

The connector is always able to restart from the last successful scan. It uses checkpoints to know which object has been completely scanned and when. If a scan failed, it cannot restart from the point where it stopped. It always restarts from the last successful scan date. As of now, the connector creates checkpoints every 100 documents.

Incremental Indexing

The Proxem connector supports incremental indexing for:

- New or modified documents.
- Deleted documents.
- Re-analyzed documents (if the categorization rules are changed in Proxem Studio).

Logging

The connector logs any error (exceptions in Proxem, invalid configuration, etc.) which occurred during the scan.

Security Management

The Proxem connector supports the security defined of the source connector it is mapped with.

Installing the Connector in Exalead CloudView

The Proxem connector installation requires 3 plugins.

There are:

- One plugin to install a specific PushAPI filter on the source connector that fetches binary documents.
- One plugin for the Proxem connector that retrieves the documents and the categories extracted by Proxem, and makes them available in Exalead CloudView.
- One plugin to use the Consolidation Server and retrieve documents directly. This avoids waiting for the Proxem analysis to complete before pushing documents.

1. Open the Exalead CloudView Administration Console.
2. Go to **Plugins** and click **Upload plugin**.
3. Select the Proxem Push API filter plugin `proxem-papi-filter.zip` and click **Upload**.
4. Select the Proxem connector plugin `proxem-connector.zip` and click **Upload**.
5. Select the Consolidation Server aggregation processor plugin `proxem-aggregation-processor.zip` and click **Upload**.

All plugins are added to the Installed custom plugins list. You are now ready to create Proxem connectors in the Administration Console.

Configuring the Connector in Exalead CloudView

This section describes how to prepare and configure the Proxem connector in Exalead CloudView, to use the Proxem categories as refinement.

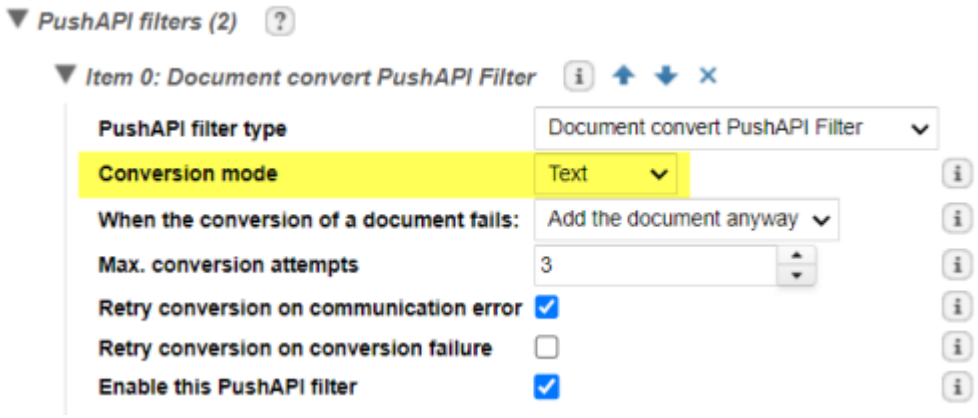
Configuring the Source Connector

You first have to configure a source connector to fetch, prepare, and push documents to the Proxem API.

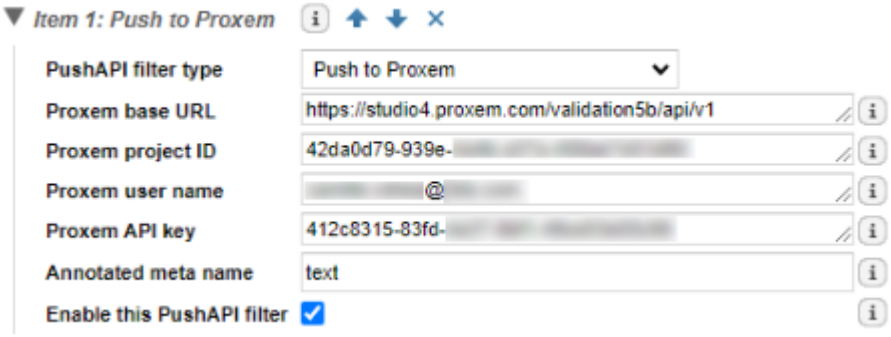
Note: As the Proxem analysis can take time, you can optionally push documents directly to a Consolidation Server to get a quicker access to documents.

1. Open the Exalead CloudView Administration Console.
2. Add a connector instance:
   a. Go to **Connectors** and click **Add connector**.
   b. For **Name**, enter a descriptive name for the connector, for example, `filesToProxem`.
   c. In **Creation mode**, select either **new** or **copy** if you want to copy an existing connector.
   d. In the **Type** list, select a source connector like **Files**.
   e. Click **Accept**.
3. Configure the source connector. For example, for a Files connector, you have to define the paths to crawl in the **Configuration** tab.
4. Go to the **Advanced** tab.
5. Add a **Document convert PushAPI Filter**.

   Important: For the **Conversion mode** property, select **Text**.



6. Add a **Push to Proxem** PushAPI filter.



Property	Description
Proxem project ID	Specify the URL of your Proxem studio.
Proxem project ID	ID of the Proxem project for which you want to index documents.
Proxem user name	User name to connect to the Proxem server.
Proxem API key	Specify the key of the Proxem external API.
Annotated meta name	Specify the name of the meta that the Proxem service must send and annotate. The default meta is `text`.

7. Click **Apply**.

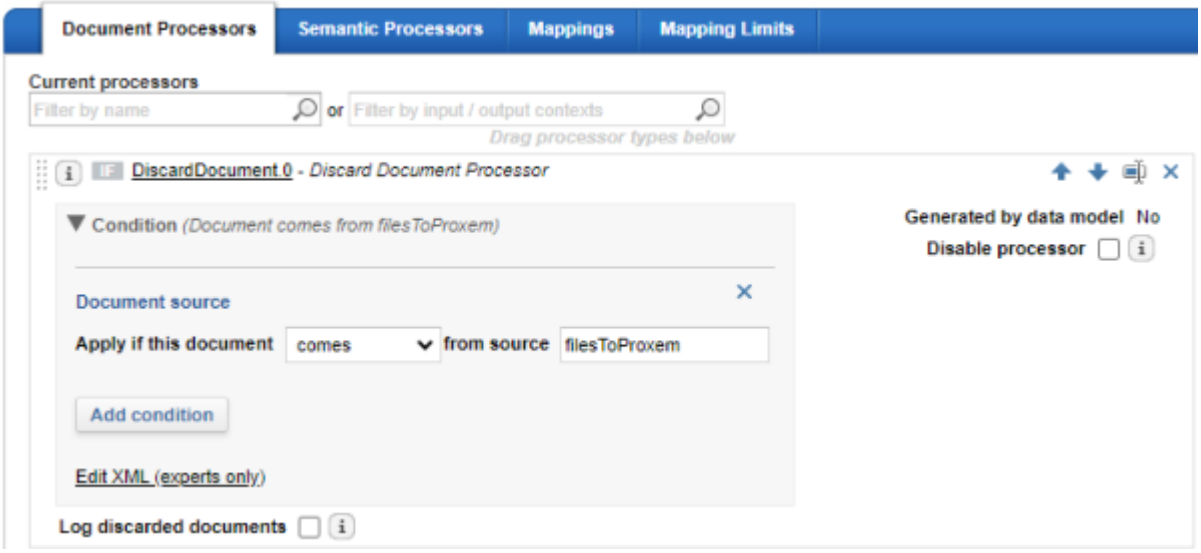
For the standard workflow, you can now discard source connector documents to make sure that Exalead CloudView does not index them.

Discarding Source Connector Documents

You can discard documents coming from the source connector to avoid indexing and seeing them in the Exalead CloudView interface.

You do not want to search on these documents but only on the documents analyzed by Proxem.

1. Go to **Data Processing** and open your Analysis pipeline, for example, `ap0`.
2. In the **Document Processors** tab, add a **Discard Document Processor** at the top of the **Current processors** list.
3. Expand **Condition** tab, and configure for documents coming from the source connector (for example, `filesToProxem`).



4. Click **Apply**.
5. Go to the **Home** page and click **Scan** for the source connector.

You can now configure the Proxem connector.

Configuring the Proxem Connector

You can now configure a Proxem connector to index the documents analyzed by the Proxem API.

1. Open the Exalead CloudView Administration Console.
2. Add a connector instance:
   a. Go to **Connectors** and click **Add connector**.
   b. For **Name**, enter a descriptive name for the connector, for example, `proxem`.
   c. In **Creation mode**, select either **new** or **copy** if you want to copy an existing connector.
   d. In the **Type** list, select **Proxem**.
   e. Click **Accept**.
3. In the **Configuration** tab, configure your Proxem connector as required.

Property	Description
Proxem base url	URL of the Proxem environment. It must point to the API. For example, `https://studio3.proxem.com/production4d/api/v1`
Proxem project ID	ID of the Proxem project for which you want to index documents.
Proxem user name	User name to connect to the Proxem server.
Proxem API key	Specify the key of the Proxem external API.
Source connector name	Specify the name of the source connector that fetches documents to push them into the Proxem connector.
Proxem results page size	Specify the number of Proxem documents to show per page. This option has an impact on performance.

4. Go to the **Operation** tab and schedule a full scan every 5 minutes.

   Note: This step is to make sure that the connector has up-to-date annotations.

5. Click **Apply**.
6. Go to the **Home** page, and click **Scan** for the Proxem connector.

   Documents will be indexed only after the first Proxem analysis. This can take time.

You can now configure the Index Data Model.

Configuring the Exalead CloudView Data Model

You can now configure the Exalead CloudView data model to create the categories generated by Proxem.

See your Proxem Explore interface, to find the categories that you want to retrieve.

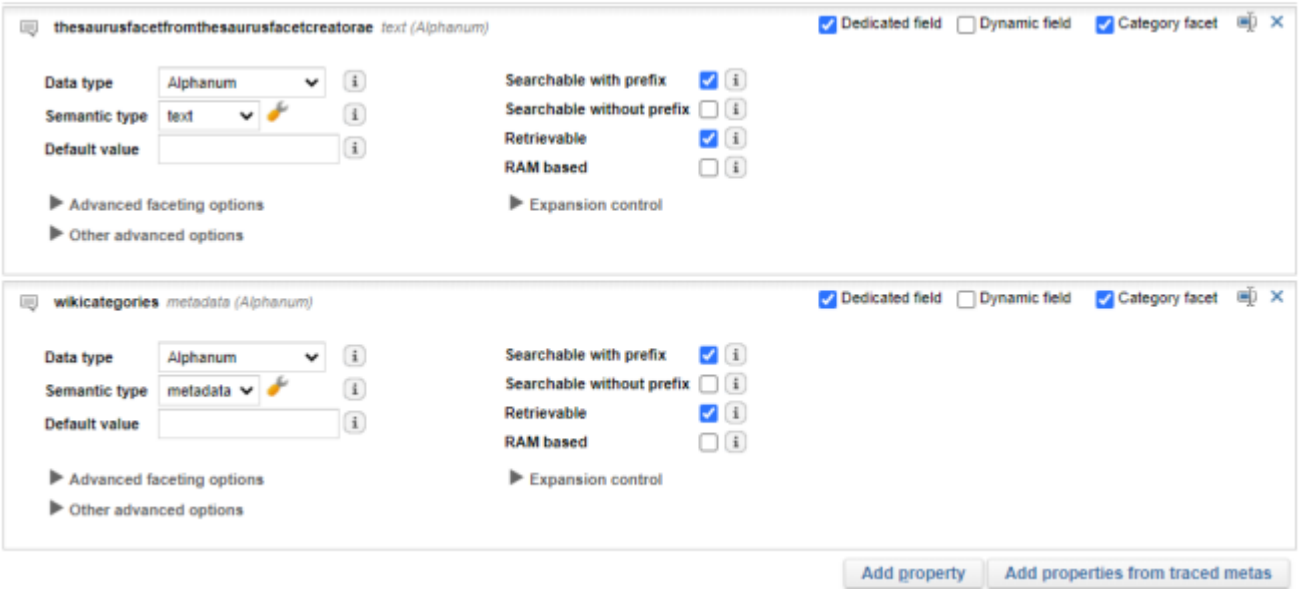
1. In the Administration Console, go to **Index > Data Model**.
2. Under **Data model options**, select **Trace all metas**
3. Click **Apply**.
4. Go to the **Home** page, click **Scan** for the connector.

   This saves all scanned metas to an internal database.

5. Go back to **Index > Data Model > Classes**, click **Add properties from traced metas**.

   Exalead CloudView enables you to select multiple metas to save as properties.

6. Select the Proxem categories (their technical names) that you want to index, and in their property configuration, select **Category facet**.



7. Click **Generate properties**.
8. Click **Apply**.
9. As you have changed the Index Schema by adding new properties, you need to clear the documents in the build group and reindex your data:
   a. On the **Home** page, under **Indexing**, click **Clear**. Wait for the index to clear its documents.
   b. Under **Connectors**, click **Scan** for your Proxem connector.

   Important: In **Index > Data Model** , clear the **Trace all metas** option.

You can now test your Search Application or configure a Consolidation Server.

Configuring the Consolidation Server

Optionally, you can configure Exalead CloudView to use a Consolidation Server, that will receive documents directly from the Source Connector, and then from the Proxem Connector when the Proxem analysis is complete.

This is useful to save time and avoid waiting for the end of the analysis.

1. Prepare both the Source Connector and the Proxem Connector to use the Consolidation Server:
   a. Open the connector, and go to the **Deployment** tab.
   b. For **Push to PAPI server**, select the Consolidation Server instance on which the connector must push its documents.
   c. Apply the configuration.

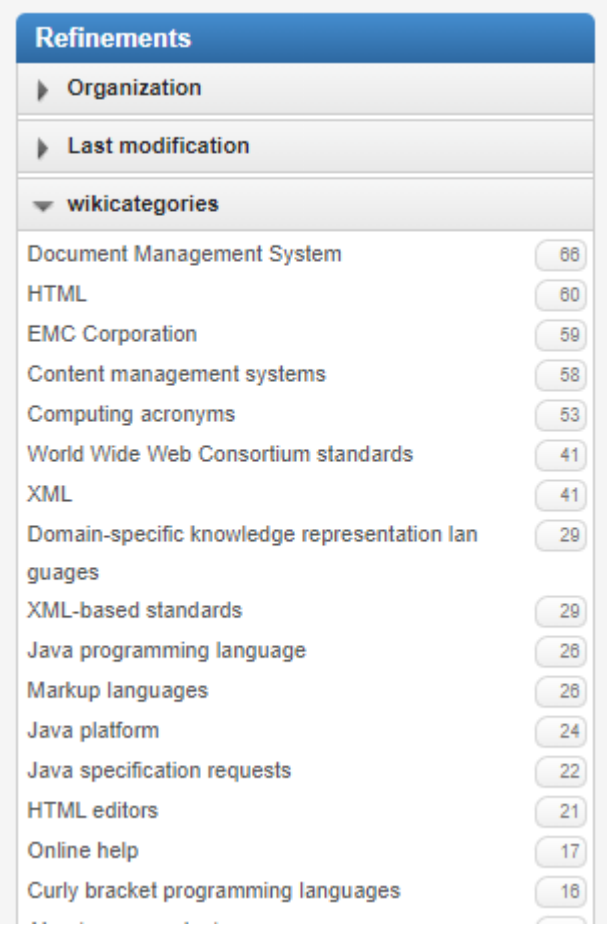
2. Disable the **Discard Document Processor** configured in Discarding Source Connector Documents .
   a. Go to **Data Processing** and open your Analysis pipeline, for example, `ap0`.
   b. In the **Document Processors** tab, for the **Discard Document Processor**, select **Disable processor**.
   c. Click **Apply**.
3. Configure the Consolidation Server.
   a. Go to **Index > Consolidation**.
   b. Under **Aggregation Processors**, click **Add processor**.
   c. Select **com.dassault_systemes.exalead.processor.proxem.ProxemAggregationProcessor ("Proxem Aggregation Processor")**
   d. Click **Apply**.
4. Rescan the impacted sources using the Consolidation Server:
   a. Go to the **Home** page.
   b. Under **Indexing**, click **Clear** to clear the index (as you have changed the **Push to PAPI server** property from a build group configuration to a Consolidation Server configuration).
   c. Scan your connectors.

You can now test your Search Application.

Testing Results in the Search Application

Test the use of Proxem categories in your Search interface.

You can now use the Proxem categories in your Search interface.



Important: When you modify one of the configuration properties, you must clear documents and start a Full Scan to synchronize all documents.