

Coveo Enterprise Search

How to Setup Coveo's OCR Module Open Converter Script

Applies to CES 5 and later

The following document explains how to setup Coveo's OCR module open converter script.

Adding an Open Converter

1. Create a directory entitled *C:\CES6\Scripts (%IndexPath%\Scripts)*.
2. Copy *OCRConversion6.js* from *C:\Program Files\Coveo Enterprise Search 6\OCR Module (%InstallationPath%)* to *C:\CES6\Scripts (%IndexPath%\Scripts)*.
3. Create a source to index the OCR documents (refer to the [How to Add a Source](#) help topic in the Administration Tool help).
4. Indicate the presence of conversion scripts (document converters) to the indexer by adding an open converter in the Administration Tool.
 - a. Access the **Converter Managers** page (Configuration > Converters).
 - b. In the left navigation pane, click **Open Converters**. The **Open Converters** page is displayed.
 - c. Click **Add**. The **Add an Open Converter** page is displayed.

Save Cancel

- d. In the **Name** field, enter a name for the converter (e.g. OCR).
- e. In the **Script File** field, enter the path of the converter script file (e.g. *C:\CES6\Scripts\OCRConversion6.js*).
- f. In the **Script Language** drop-down list, select **JScript**.
- g. Click **Save**.


Associating a Document Type to the Open Converter Script

















After adding an open converter to the system, a document type must be associated to the open converter script in order to trigger OCR processing. A typical configuration could include *Adobe Acrobat* documents (PDF) and images.


Note: To ensure the highest level of performance and quality, it is recommended to separate the source files requiring OCR from those that do not. It is possible to do so by moving files in different file system

folders or through the use of source filters (for more information on filters, refer to the [How to Add and Modify Filters](#) help entry in the Administration Tool help).

Important: The documents supported by OCR are *.tiff*, *.tiff-fx*, *.pcx*, *.dcx*, *.bmp*, *.jpeg*, *.png*, *.max*, *.gif*, *.pbm*, and *.pdf*.

1. In the **Sources and Collections** page (Index > Sources and Collections), select the source on which the open converter script is used. The **Status** page is displayed.
2. In the left navigation pane, click **Document Types**.
3. Click  **Edit**. The **Document Types** page is displayed.
4. Click the document type on which the converter script will be used, e.g. **Adobe Acrobat Documents**. The following is displayed.

Name	Adobe Acrobat Documents
File Extensions	<input type="text" value=".pdf"/>  Use semicolons to separate entries.
Action	<input type="text" value="Index entire document"/> 
Indexing Failure Action 	<input type="radio"/> Reject document <input checked="" type="radio"/> Index file information only
Converter 	<input checked="" type="radio"/> Use a default converter <input type="text" value="Detect"/> <input type="radio"/> Use an open converter <input type="text"/>
Content Types 	<input type="text"/> One entry per line.
sysfiletype Field Value 	<input checked="" type="radio"/> Use the value set by the converter <input type="radio"/> Use this value <input type="text"/>
Option	<input checked="" type="checkbox"/> When indexing attachments, index the parent document <input checked="" type="checkbox"/> Inherit source options  <input type="checkbox"/> Disable document summarization  <input type="checkbox"/> Open results with cached version 
Title Selection Sequence	Use the title extracted by the converter   Automatically detect the title of documents   Use the filename  
Title Metadata Name	<input type="text" value="Title"/> 

5. In the **Action** drop-down list, select **Index entire document**.
6. In the **Converter** section, select **Use an open converter** and the name of the converter script to use from the drop-down list, e.g. **OCR**.
7. Click  **Apply Changes**.
8. Repeat the steps above for each source and document type that uses the OCR converter.
9. Open the **CES Console** (*Windows Start* menu > All Programs > Coveo Enterprise Search 6 > CES Console).
10. Rebuild your sources in the Administration Tool and observe the activity in the CES Console. Documents are crawled, converted and transactions are applied to index.

Note: Once transactions are applied to the index, the user can search for document content.

Modifying the OCR Module License

The installation kit for the OCR module copies the licenses entered during the setup into the file named *License.txt* in the installation folder (usually in *C:\Program Files\Coveo Enterprise Search 6\OCR Module*). The OCR module does not work unless the license contained in this file is valid. If necessary, the user can edit *License.txt* to modify the license text.