OCR Action

Optical Character Recognition (OCR) converts PDF files that include scanned or photographed images of typed or printed text into files containing machine-encoded text. The converted files can then be searched for keywords and copied and pasted. The OCR Action runs on a PDF attached in a file field and creates a new text-based version of the PDF, which is then attached to the same or another file field.

Any file that users likely need to search can benefit from OCR. For example, if you have lengthy, paper-based amendments that accompany a contract, you can use an OCR action on the scanned files to make them searchable for anyone viewing the original Contract record.

> ⓘ **Prerequisites**
>
> In the table where you'll create the OCR action, you must have at least one File with Versioning field to hold the source file and, if desired, a second File with Versioning field to save the newly converted file.
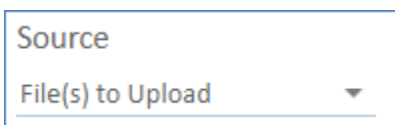
Create an OCR Action

You can access the Actions wizard in several ways, but the easiest way is to select Setup [Table] from the table where you want to create the action.

1. From the top nav bar, expand the table's drop-down and select Setup [Table].
2. Select the Actions tab in the Table wizard.
3. Click Create OCR Action.
4. On the General tab, name your action and give it a description.

   > ⚠ Once your action is saved, the system automatically adds an O: before the given title in order to distinguish the action as an OCR action.

5. Navigate to the Options tab and choose a source file field in the current table. This field holds the documents that the OCR action uses.

6. Choose where and how to store the OCR-converted files:
   - **In the original field:** Adds the converted file to the same field as the original file.
     - **Replace original files:** Replaces the original file with the converted file.
     - **Append OCR files:** Appends the converted file to the field and retains the original. You can only choose this option if the file field allows multiple files to be attached.
   - **In File field:** Allows you to select from a drop-down menu of file fields in the current table where to save the converted file. If the source field allows multiple files to be attached, this field must also allow multiple files.

- Overwrite/Update existing files: Replaces existing attachments with the converted file.
- Append OCR file(s): Appends the converted file to the selected field and retains the original file. You can only choose this option if the file field allows multiple files to be attached.

7. Define any text to be appended to the converted file name. This can include field variables, such as `$id` or `$company_name`.

Append to file name:

ocr$id

8. Choose whether you want to run Document Quality Evaluation on the document. This option runs an algorithm to check if the document is a good candidate for AI models. Select one of the options if you want to run the algorithm:
   - Evaluate document quality and run OCR only if OCR is needed. This option evaluates document quality before spending system resources on running OCR. The evaluation must return "OCR Needed" in order for OCR to run. If there are other problems with the file, OCR does not run, and you can configure additional automation to flag someone to review the issue. If the file is already in a readable format without running OCR, OCR also does not run, and you can configure additional automation to pass the file along to an AI model.
   - Evaluate document quality after OCR. This option evaluates the quality of the OCR'd document. The system then evaluates the quality of the final OCR file result.
     - Store the OCR'd file only if it passes the quality evaluation. Use this option to automatically discard files if OCR does not produce a document of sufficient quality.
9. If you selected an option to run Document Quality Evaluation, complete the subsequent fields to configure it.
   a. Under Evaluation Result, select a Choice field. Make sure the field includes all of the following choice options. Most of these choices are errors, but if the document is high enough quality to use with AI models, it receives a Check Passed result.
      - OCR Needed: flagged when a document contains less than 10 lines of text, or if the lines are empty.
      - No Language Detected: flagged if no language is detected for over 50% of the lines.
      - Multilanguage Document: flagged if percentage of English sentences in the document is lower than 50%.
      - Bad Document: flagged if the average number of words per line is lower than three.
      - Check Passed: flagged if number of lines is more than 10, the lines are not empty, a language is detected, and the average number of words per line is more than three.
   b. Under Language, select a Multi-choice field. This is used to store detected languages. If the contract contains more than one language, it will not pass the evaluation. The choice list should consist of ISO-639-1 language abbreviations.
   c. Under Evaluation Output, select a text field.
10. Select the output file format. The available options are Text or PDF.

11. Select the expected language of the original document to improve OCR accuracy. Hold Ctrl if you need to select more than one.
12. If you plan to use the action in a rule rather than an action button, select the "Ignore errors on corrupted PDF (s)" option. This allows rules to proceed even if the OCR action fails due to a corrupted document. If the action is run by an action button instead, this option has no effect, and corrupted documents stop the execution of actions.
13. Choose whether to queue OCR actions or run them immediately. OCR can take some time to run, and it affects system performance while it is running, so depending on the frequency and volume of OCR actions you run in your system, you might want to queue them to run in the background.
14. Click Finish to save the action.