

## M-Files OCR

M-Files OCR (Optical Character Recognition) interfaces directly with virtually any scanner to produce searchable PDF files from paper documents. In other words, M-Files converts imported files into searchable PDFs so that a full-text search of scanned documents is possible.

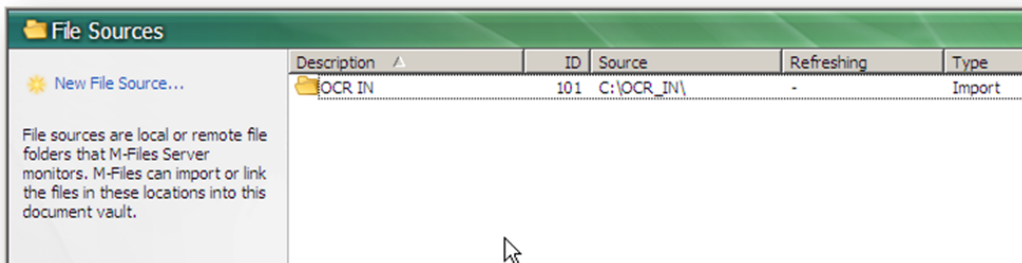
M-Files OCR also provides "ID-based" scanning for automating tasks such as file naming, defining properties and metadata, initiating workflows, and processing a signed contract based on identification of the contract's signature.

**Based on the popular I.R.I.S. OCR technology, the M-Files OCR module provides extensive support for connecting M-Files directly to scanners and eliminates the need for additional third-party scanning and OCR software. In addition to the text recognition, you can also define OCR zones in which certain characters are recognized and saved automatically as metadata of a document.**

*The M-Files OCR module is an M-Files add-on product available for extra fee. It can be activated with a license code. After that, the OCR related functions will be available in M-Files Server Administrator and M-Files Client.*

### 1. Scanning and text recognition (OCR) from a network scanner

By using *connections to external sources*, you can bring scanned documents automatically into M-Files from external sources. This option has been included in previous M-Files versions, but M-Files 6.0 offers the possibility to use text recognition and define OCR zones. Activation of these functions requires the M-Files OCR module.



*You can define File Sources / Connection Properties in M-Files Server Administrator.*

### Text recognition (searchable PDF)

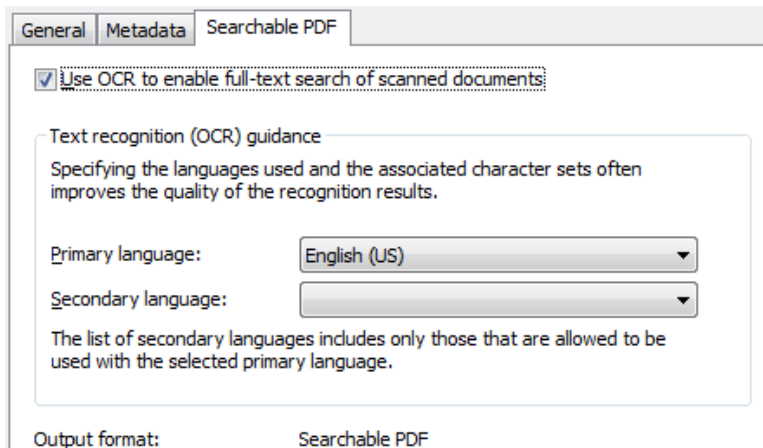
To perform automatic text recognition in M-Files of files imported from the scanner, select the option *Use OCR to enable full-text search of scanned documents*. M-Files then converts the imported image files into *searchable PDFs* so that a full-text search of the scanned document is possible. After conversion, you can find PDF format document by searching using text from the document, e.g., in the case of an agreement document by searching using the names of the contracting parties or other text included in the agreement.

Converting the file to a searchable PDF does not affect the outward appearance of the document when viewing it. The users still see the original scanned image. M-Files stores the automatic text recognition results in the PDF as invisible text, which is used when searching the file. Possible text recognition inaccuracies will not affect the appearance of the scanned document in any way when viewed on screen or printed.

Text recognition can be performed on the following file formats: TIFF, JPEG, BMP, PNG, and GIF.

### Text recognition (OCR) guidance

Although the OCR automatically recognizes all Western languages and Cyrillic character sets, specifying a language selection often improves the quality of the text recognition results. In ambiguous cases, a problematic recognition result may be resolved by a language-specific factor, such as recognition of the letter 'ä' in Finnish. The list of secondary languages only includes languages that are allowed to be used together with the selected primary language.



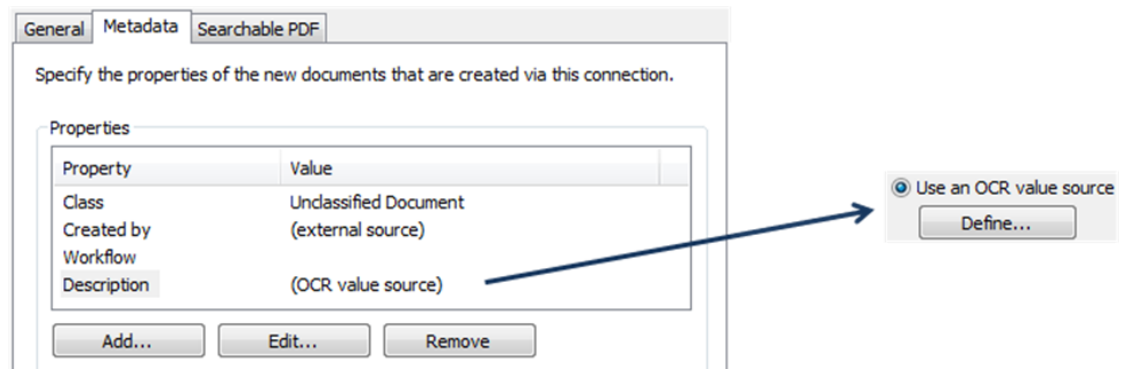
Use the *Searchable PDF* tab to enable full-text search of scanned documents.

Text recognition can also be performed in M-Files Client using local scanning or conversion function. If you wish to use text recognition using external sources through the M-Files Server Administrator only, this limitation can be set by changing the registry settings. The registry settings can be used to set other limitations as well. For more information on registry settings, contact technical support at: [support@m-files.com](mailto:support@m-files.com).

### Automatically saved metadata from scanners (use an OCR value source)

For files imported from an external source, you can define an automatically added property that uses the OCR value source. The OCR value source is a zone defined on the page. Using OCR, the zone gives a value for the selected property. This value will be saved automatically to the document's metadata.

Use of the OCR value source is only possible when using an external source. The OCR value source cannot be defined in M-Files Client.



First add the new property from the *Metadata* tab for the connection. Then define the property that uses the OCR value source (*Zone OCR*).

## Text recognition (OCR) guidance

Although the OCR automatically recognizes all Western languages and Cyrillic character sets, specifying a language selection often improves the quality of the text recognition results. In ambiguous cases, a problematic recognition result may be resolved by a language-specific factor, such as recognition of the letter 'ä' in Finnish. The list of secondary languages only includes languages that are allowed to be used together with the selected primary language.

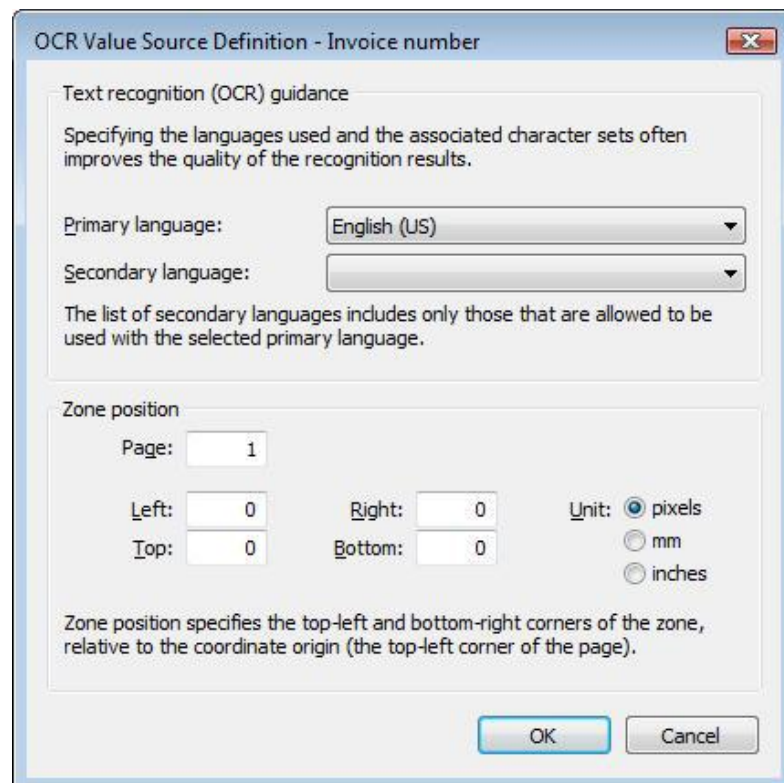
## Define an OCR zone for OCR value source

You can define a zone in which certain characters are recognized as values of a defined property. The characters may include any letters, numbers, or punctuation marks. For example, an invoice number shown on a page can be added as the *Invoice number* property value for the scanned document. This enables you to automate, e.g., the scanning and storing of specified documents in M-Files with metadata that is always correct.

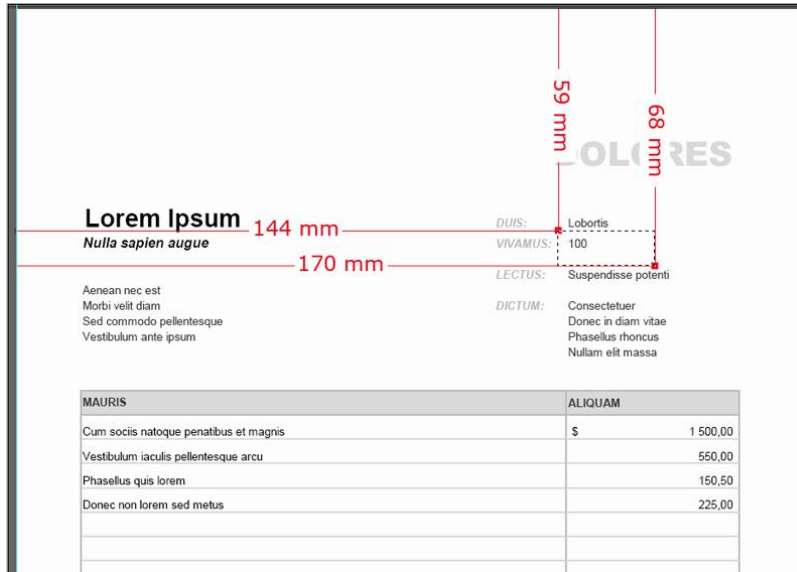
In most cases, to ensure that the defined zone is correctly positioned, the document to be scanned should be placed onto the scanner's glass plate by hand rather than fed via an automatic sheet feeder.

In some cases, the OCR may give an incorrect recognition result of the text: for example, depending on the font type or size, the number 1 may be interpreted as the letter l. To ensure that the characters are added correctly to the document metadata, you can check the property values with event handlers and VBScript. You can then use VBScript to check, e.g., that all added characters are numbers. For more information about event handlers, VBScript code, and M-Files API, please contact [support@m-files.com](mailto:support@m-files.com).

Note: You can use the OCR value source without selecting the *Use OCR to enable full-text search of scanned documents* function in the *Searchable PDF* tab.



Define the zone in which characters are to be recognized as values of the defined property.



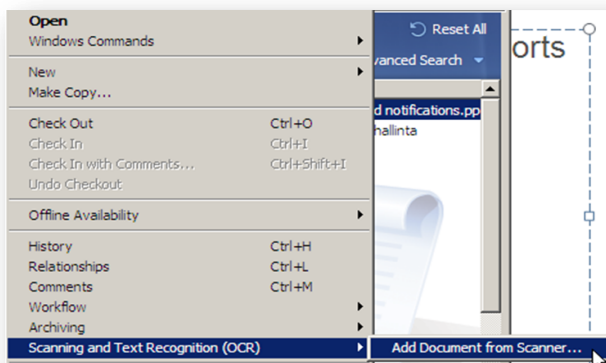
Zone position specifies the top-left and bottom-right corners of the zone, relative to the coordinate origin (the top-left corner of the page). In this example, the following values are used: left 144 mm, top 59 mm, right 170 mm, and bottom 68 mm.

## 2. Scanning and text recognition (OCR) from a local scanner

Scannable paper documents can be added to M-Files by using external connections (network scanning) or a local scanner.

Scanner integration in M-Files Client uses the TWAIN and WIA technologies. Scanners that can be equipped with a TWAIN or WIA driver are supported.

When using local scanning, the scanner must be directly connected to the computer that will be used to add the scanned file to M-Files. The scanning functions can be found in the *M-Files Client / Operations menu* or using the *shortcut menu*.



You can find the scanning functions using the *Operations* or *shortcut menu* in M-Files Client.

### Add Document from Scanner

You can add new documents from the scanner. M-Files then creates a new document from the scanned file.

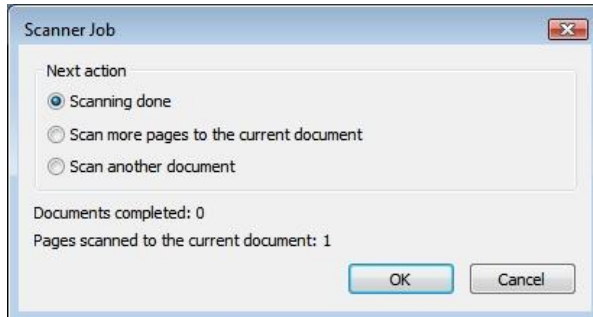
### Add File from Scanner

If you want to add files to a multi-file document, select *Add File from Scanner*. M-Files then adds a new file from the scanned file to the multi-file document. Note: the multi-file document

only has one set of metadata, i.e. files belonging to the multi-file do not have any separate metadata.

### Replace with File from Scanner

You can replace an individual file with a file from the scanner. This function can be used to replace a file in a single-file or multi-file document.



After scanning, M-Files asks if you want to complete the scanning job or add more pages or add another document.

### Text recognition after scanning

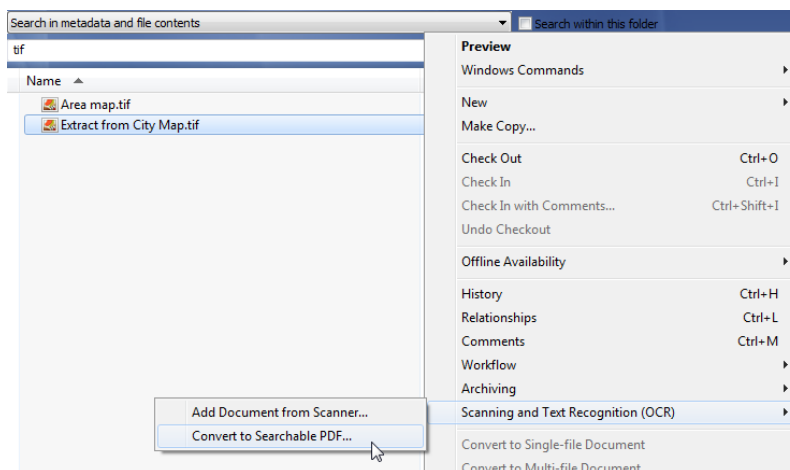
If the M-Files OCR module has been enabled, M-Files suggests that the file should be converted to a searchable PDF by text recognition when scanning is completed. You can choose the text recognition or ignore it. You can also define advanced settings for the text recognition (for more information, see chapter 3).



If the text recognition is performed on an image file which was not saved and returned to M-Files, the file will only be saved as a PDF. Otherwise, the original image file can be found in the document version history.

## 3. Convert to searchable PDF

If the M-Files OCR module has been enabled, you can convert an image file into a searchable PDF file. This makes full-text search of the scanned document possible.



You can find the OCR functions using the Operations or shortcut menu in M-Files Client.

After conversion, you can find this PDF format document by searching using text from the document, e.g., in the case of an agreement document by searching using the names of the contracting parties or other text included in the agreement.

Converting the file to a searchable PDF does not affect the outward appearance of the document when viewing it. The users still see the original scanned image. M-Files stores the automatic text recognition results in the PDF as invisible text, which is used when searching the file. Possible text recognition inaccuracies will not affect the appearance of the scanned document in any way when viewed on screen or printed.

Text recognition can be performed on the following file formats: TIFF, JPEG, BMP, PNG, and GIF.

If the text recognition is performed on an image file which was not saved and returned to M-Files, the file will only be saved as a PDF. Otherwise, the original image file can be found in the document version history.

Note: You cannot perform text recognition on a file which has already been saved as a PDF in M-Files. If you want to convert already existing PDF files to searchable PDF files, please contact our consulting services ([support@m-files.com](mailto:support@m-files.com)).

#### **4. Skip Conversion**

If the M-Files OCR module has been enabled, M-Files suggests that an image file can be converted to a searchable PDF when dragging it with your mouse to M-Files. You can choose the text recognition or skip conversion. If you skip the conversion the file will be saved in the original image file format to M-Files.

If you wish to use text recognition using external sources through the M-Files Server Administrator only, this limitation can be set by changing the registry settings. The registry settings can be used to set other limitations as well.

---

*The M-Files OCR module is an M-Files add-on product available for extra fee. It can be activated with a license code. The old license code must be replaced by the license code that enables use of OCR. After that, the OCR related functions will be available in M-Files Server Administrator and M-Files Client. For further details, please contact your M-Files representative or our sales department ([sales@m-files.com](mailto:sales@m-files.com)).*