

OCR with multiple languages in VBScript with ByteScout PDF Extractor SDK

ByteScout PDF Extractor SDK is the SDK is designed to help developers with pdf tables and pdf data extraction from unstructured documents like pdf, tiff, scans, images, scanned and electronic forms. The library is powered by OCR, computer vision and AI to provide unique functionality like table detection, automatic table structure extraction, data restoration, data restructuring and reconstruction. Supports PDF, TIFF, PNG, JPG images as input and can output CSV, XML, JSON formatted data. Includes full set of utilities like pdf splitter, pdf merger, searchable pdf maker and other utilities.

FOR MORE INFORMATION AND FREE TRIAL:

[Download Free Trial SDK \(on-premise version\)](#)

[Read more about ByteScout PDF Extractor SDK](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout PDF Extractor SDK](#)

[Get Free API key for Web API](#)

[visit www.Bytescout.com](http://www.Bytescout.com)

Source Code Files:

OcrWithMultipleLanguages.vbs

```
' This example demonstrates the use of Optical Character Recognition (OCR) to extract text  
' from scanned PDF documents and raster images.
```

```
' Create TextExtractor object  
Set extractor = CreateObject("Bytescout.PDFExtractor.TextExtractor")  
extractor.RegistrationName = "demo"  
extractor.RegistrationKey = "demo"
```

```
' Load sample PDF document  
extractor.LoadDocumentFromFile("../multi_language_ocr.pdf")
```

```
' Enable Optical Character Recognition (OCR)
extractor.OCRMode = 1 ' OCRMode.Auto = 1

' Set the location of OCR language data files
extractor.OCRLanguageDataFolder = "c:\Program Files\Bytescout PDF Extractor SDK\ocrdata_best"

' Set OCR language targeting multiple languages english and french
' "eng" for english, "deu" for German, "fra" for French, "spa" for Spanish etc - according to files in "ocrdata" folder.
extractor.OCRLanguage = "eng+fra"
' Find more language files at https://github.com/bytescout/ocrdata

' Set PDF document rendering resolution
extractor.OCRResolution = 300

' You can also apply various preprocessing filters to improve the recognition on low-quality scans.
' But they significantly hit the performance, so do not enable them by default.

' Automatically deskew skewed scans
extractor.OCRImagePreprocessingFilters.AddDeskew()

' Remove vertical or horizontal lines (sometimes helps to avoid OCR engine's page segmentation errors)
extractor.OCRImagePreprocessingFilters.AddVerticalLinesRemover()
extractor.OCRImagePreprocessingFilters.AddHorizontalLinesRemover()

' Repair broken letters
extractor.OCRImagePreprocessingFilters.AddDilate()

' Remove noise
extractor.OCRImagePreprocessingFilters.AddMedian()

' Apply Gamma Correction
extractor.OCRImagePreprocessingFilters.AddGammaCorrection()

' Add Contrast
extractor.OCRImagePreprocessingFilters.AddContrast(20)

' (!) You can use new OCRAnalyzer class to find an optimal set of image preprocessing
' filters for your specific document.
' See "OCR Analyser" example.

' Save extracted text to file
extractor.SaveTextToFile("output.txt")

WScript.Echo "Extracted text saved as 'output.txt'."

Set extractor = Nothing
```

VIDEO

https://www.youtube.com/watch?v=s28W3_KMraU

ON-PREMISE OFFLINE SDK

[60 Day Free Trial](#) or [Visit ByteScout PDF Extractor SDK Home Page](#)
[Explore ByteScout PDF Extractor SDK Documentation](#)
[Explore Samples](#)
[Sign Up for ByteScout PDF Extractor SDK Online Training](#)

ON-DEMAND REST WEB API

[Get Your API Key](#)
[Explore Web API Docs](#)
[Explore Web API Samples](#)

[visit www.ByteScout.com](http://www.ByteScout.com)

[visit www.PDF.co](http://www.PDF.co)

www.bytescout.com