

OCR analyser for PDF in VBScript using ByteScout PDF Extractor SDK

Tutorial: how to do OCR analyser for PDF in VBScript

Writing of the code to OCR analyser for PDF in VBScript can be done by developers of any level using ByteScout PDF Extractor SDK. ByteScout PDF Extractor SDK was made to help with OCR analyser for PDF in VBScript. 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.

This rich sample source code in VBScript for ByteScout PDF Extractor SDK includes the number of functions and options you should do calling the API to implement OCR analyser for PDF. This VBScript sample code should be copied and pasted into your application's code editor. Then just compile and run it to see how it works. Code testing will allow the function to be tested and work properly with your data.

Trial version can be obtained from our website for free. It includes this and other source code samples for VBScript.

VBScript - PerformOcrAnalysisAndExtractText.vbs

```
' This example demonstrates the use of OCR Analyser - a tooling class for analysis of
scanned documents
' in PDF or raster image formats to find best parameters for Optical Character
Recognition (OCR) that
' provide highest recognition quality.

' Input document
inputDocument = "..\..\sample_ocr.pdf"

' Document page index
pageIndex = 0

' Location of OCR language data files
ocrLanguageDataFolder = "c:\Program Files\Bytescout PDF Extractor SDK\ocrdata\"

' OCR language
ocrLanguage = "eng" ' "eng" for english, "deu" for German, "fra" for French, "spa"
for Spanish etc - according to files in "ocrdata" folder
' Find more language files at https://github.com/bytescout/ocrdata
```

```
' Create OCRAnalyzer object and activate it with your registration information
Set ocrAnalyzer = CreateObject("Bytescout.PDFExtractor.OCRAnalyzer")
ocrAnalyzer.RegistrationName = "demo"
ocrAnalyzer.RegistrationKey = "demo"

' Load document to OCRAnalyzer
ocrAnalyzer.LoadDocumentFromFile(inputDocument)

' Setup OCRAnalyzer
ocrAnalyzer.OCRLanguage = ocrLanguage
ocrAnalyzer.OCRLanguageDataFolder = ocrLanguageDataFolder

WScript.Echo "Starting the OCR analysis. Click OK and wait, it may last long." &
outputDocument

' Perform analysis and get results
Set analysisResults = ocrAnalyzer.AnalyzeByOCRConfidence(pageIndex)

' Now extract page text using detected OCR parameters

outputDocument = ".\result.txt"

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

' Load document to TextExtractor
textExtractor.LoadDocumentFromFile(inputDocument)

' Setup TextExtractor
textExtractor.OCRMode = 1 ' OCRMode.Auto
textExtractor.OCRLanguageDataFolder = ocrLanguageDataFolder
textExtractor.OCRLanguage = ocrLanguage

' Apply analysis results to TextExtractor instance
ocrAnalyzer.ApplyResults (analysisResults), (textExtractor)

' Save extracted text to file
textExtractor.SaveTextToFile(outputDocument)

WScript.Echo "Extracted text saved to " & outputDocument

Set textExtractor = Nothing
Set ocrAnalyzer = Nothing
```

FOR MORE INFORMATION AND FREE TRIAL:

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

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

[Explore documentation](#)

[Visit \[www.ByteScout.com\]\(http://www.ByteScout.com\)](#)

or

[Get Your Free API Key for \[www.PDF.co\]\(http://www.PDF.co\) Web API](#)