www.bytescout.com

ocr with best dataset with pdf extractor sdk in VBScript and ByteScout PDF Suite

ocr with best dataset with pdf extractor sdk in VBScript

An easy to understand guide to learn how to ocr with best dataset with pdf extractor sdk in VBScript. ByteScout PDF Suite was made to help with ocr with best dataset with pdf extractor sdk in VBScript. ByteScout PDF Suite is the bundle that provides six different SDK libraries to work with PDF from generating rich PDF reports to extracting data from PDF documents and converting them to HTML. This bundle includes PDF (Generator) SDK, PDF Renderer SDK, PDF Extractor SDK, PDF to HTML SDK, PDF Viewer SDK and PDF Generator SDK for Javascript.

If you want to quickly learn then these fast application programming interfaces of ByteScout PDF Suite for VBScript plus the guideline and the VBScript code below will help you quickly learn ocr with best dataset with pdf extractor sdk. If you want to implement this functionality, you should copy and paste code below into your app using code editor. Then compile and run your application. VBScript application implementation mostly involves various stages of the software development so even if the functionality works please check it with your data and the production environment.

On our website you may get trial version of ByteScout PDF Suite for free. Source code samples are included to help you with your VBScript application.

FOR MORE INFORMATION AND FREE TRIAL:

Download Free Trial SDK (on-premise version)

Read more about ByteScout PDF Suite

Explore API Documentation

Get Free Training for ByteScout PDF Suite

Get Free API key for Web API

visit www.ByteScout.com

Source Code Files:

ExtractTextFromImageOrScannedPdfUsingBestOcrDataset.vbs

```
' This example demonstrates the use of Optical Character Recognition (OCR) to extract t
' 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("..\..\sample_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

' Set OCR language
' "eng" for english, "deu" for German, "fra" for French, "spa" for Spanish etc - accord
extractor.OCRLanguage = "eng"
' Find more language files at https://github.com/bytescout/ocrdata/tree/master/ocrdata_

' Set PDF document rendering resolution
extractor.OCRResolution = 300


' You can also apply various preprocessing filters to improve the recognition on low-qu
' 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 segme
'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=NEwNs2b9YN8

ON-PREMISE OFFLINE SDK

60 Day Free Trial or Visit ByteScout PDF Suite Home Page
Explore ByteScout PDF Suite Documentation
Explore Samples
Sign Up for ByteScout PDF Suite Online Training

ON-DEMAND REST WEB API

Get Your API Key
Explore Web API Docs
Explore Web API Samples

visit www.ByteScout.com

visit www.PDF.co