www.bytescout.com

ocr with fast dataset with pdf extractor sdk in VBScript using ByteScout Premium Suite

Build ocr with fast dataset with pdf extractor sdk in VBScript

:

Step-by-step instructions on how to do ocr with fast dataset with pdf extractor sdk in VBScript

These source code samples are assembled by their programming language and functions they use. Ocr with fast dataset with pdf extractor sdk in VBScript can be applied with ByteScout Premium Suite. ByteScout Premium Suite is the bundle that includes twelve SDK products from ByteScout including tools and components for PDF, barcodes, spreadsheets, screen video recording.

This rich and prolific sample source code in VBScript for ByteScout Premium Suite contains various functions and options you should do calling the API to implement ocr with fast dataset with pdf extractor sdk. Follow the steps-by-step instructions from the scratch to work and copy and paste code for VBScript into your editor. These VBScript sample examples can be used in one or many applications.

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

Source Code Files:

ExtractTextFromImageOrScannedPdfUsingFastOcrDataset.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 Premium Suite Home Page
Explore ByteScout Premium Suite Documentation
Explore Samples
Sign Up for ByteScout Premium 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