www.bytescout.com

ocr (optical character recognition) and pdf with pdf extractor sdk in VB.NET and ByteScout Data Extraction Suite

Learn ocr (optical character recognition) and pdf with pdf extractor sdk in VB.NET

This page explains the steps and algorithm of implementing ocr (optical character recognition) and pdf with pdf extractor sdk and how to make it work in your application. ByteScout Data Extraction Suite was created to assist ocr (optical character recognition) and pdf with pdf extractor sdk in VB.NET. ByteScout Data Extraction Suite is the bundle that includes three SDK tools for data extraction from PDF, scans, images and from spreadsheets: PDF Extractor SDK, Data Extraction SDK, Barcode Reader SDK.

The below SDK samples describe how to quickly make your application do ocr (optical character recognition) and pdf with pdf extractor sdk in VB.NET with the help of ByteScout Data Extraction Suite. Just copy and paste this VB.NET sample code to your VB.NET application's code editor, add a reference to ByteScout Data Extraction Suite (if you haven't added yet) and you are ready to go! Want to see how it works with your data then code testing will allow the function to be tested and work properly.

Trial version along with the source code samples for VB.NET can be downloaded from our website

Source Code Files:

Program.vb

```
Imports Bytescout.PDFExtractor

' This example demonstrates the use of Optical Character Recognition (OCR) to extract t
' from scanned PDF documents and raster images.

' To make OCR work you should add the following references to your project:
' "Bytescout.PDFExtractor.dll", "Bytescout.PDFExtractor.OCRExtension.dll".

Class Program

    Friend Shared Sub Main(args As String())

        ' Create Bytescout.PDFExtractor.TextExtractor instance
        Dim extractor As New TextExtractor()
        extractor.RegistrationName = "demo"
        extractor.RegistrationKey = "demo"

        ' Load sample PDF document
        extractor.LoadDocumentFromFile("sample_ocr.pdf")

        ' Enable Optical Character Recognition (OCR)
        ' in .Auto mode (SDK automatically checks if needs to use OCR or not)
        extractor.OCRMode = OCRMode.Auto

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

        ' Set OCR language
        extractor.OCRLanguage = "eng"  ' "eng" for english, "deu" for German, "fra" for
        ' 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.

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

        ' Remove vertical or horizontal lines (sometimes helps to avoid OCR engine's pa
        '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 preproc
                ' filters for your specific document.
                ' See "OCR Analyser" example.


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

        ' Cleanup
                extractor.Dispose()

        ' Open output file in default associated application
        System.Diagnostics.Process.Start("output.txt")

    End Sub

End Class
```

VIDEO

https://www.youtube.com/watch?v=NEwNs2b9YN8

ON-PREMISE OFFLINE SDK

60 Day Free Trial or Visit ByteScout Data Extraction Suite Home Page
Explore ByteScout Data Extraction Suite Documentation
Explore Samples
Sign Up for ByteScout Data Extraction 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

www.bytescout.com