

## ocr with mean dataset with pdf extractor sdk in VB.NET using ByteScout Premium Suite

ocr with mean dataset with pdf extractor sdk in VB.NET

This page explains the steps and algorithm of implementing ocr with mean dataset with pdf extractor sdk and how to make it work in your application. ByteScout Premium Suite was made to help with ocr with mean dataset with pdf extractor sdk in VB.NET. ByteScout Premium Suite is the bundle that includes twelve SDK products from ByteScout including tools and components for PDF, barcodes, spreadsheets, screen video recording.

The below SDK samples describe how to quickly make your application do ocr with mean dataset with pdf extractor sdk in VB.NET with the help of ByteScout Premium Suite. If you want to know how it works, then this VB.NET sample code should be copied and pasted into your application's code editor. Then just compile and run it. Use of ByteScout Premium Suite in VB.NET is also described in the documentation included along with the product.

ByteScout Premium Suite is available as a free trial. You may get it from our website along with all other source code samples for VB.NET applications.

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about ByteScout Premium Suite](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout Premium Suite](#)

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

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

Source Code Files:

```
Imports Bytescout.PDFExtractor
```

```
' This example demonstrates the use of Optical Character Recognition (OCR) to extract text
' 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 French
        ' Find more language files at https://github.com/bytescout/ocrdata/tree/master

        ' 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 problems)
        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 prepro
' 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 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](http://www.ByteScout.com)

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

