

How to parse with OCR in VB.NET and ByteScout Document Parser SDK

This code in VB.NET shows how to parse with OCR with this how to tutorial

These sample source codes on this page below are demonstrating how to parse with OCR in VB.NET. ByteScout Document Parser SDK is the customizable data extraction platform for batch data extraction from documents. Relies on special templates that can be created with no special technical skills required. Supports millions of documents as input and designed to handle multiple threads. Can output data as JSON, CSV, XML or custom format and you can use it to parse with OCR with VB.NET.

This code snippet below for ByteScout Document Parser SDK works best when you need to quickly parse with OCR in your VB.NET application. Just copy and paste the code into your VB.NET application's code and follow the instruction. Enjoy writing a code with ready-to-use sample VB.NET codes.

Trial version of ByteScout Document Parser SDK can be downloaded for free from our website. It also includes source code samples for VB.NET and other programming languages.

VB.NET - Module1.vb

```
Imports ByteScout.DocumentParser

' This example demonstrates parsing of scanned documents
' using the Optical Character Recognition (OCR).

Module Module1

    Sub Main()

        Dim template As String = ".\DigitalOcean.yml"
        Dim inputDocument As String = ".\DigitalOcean-scanned.jpg"

        ' Create and activate DocumentParser instance
        Using documentParser As New DocumentParser("demo", "demo")

            ' Enable Optical Character Recognition (OCR) in Auto mode
            ' (DocumentParser automatically detects if OCR Is required).
            documentParser.OCRMode = OCRMode.Auto

            ' Set PDF document rendering resolution
            documentParser.OCRResolution = 300

            ' Set the location of OCR language data files
            documentParser.OCRLanguageDataFolder = "c:\Program Files\ByteScout
Document Parser SDK\ocrdata"
```

```

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

' Note: The OCRLanguage can be overridden in a template.
' See the Template Creation Guide.

' You can also apply various preprocessing filters
' to improve the recognition on low-quality scans.

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

' Remove vertical or horizontal lines (sometimes helps to avoid OCR
engine's page segmentation errors)
'documentParser.OCRImagePreprocessingFilters.AddVerticalLinesRemover()
'documentParser.OCRImagePreprocessingFilters.AddHorizontalLinesRemover()

' Repair broken letters
'documentParser.OCRImagePreprocessingFilters.AddDilate()

' Remove noise
'documentParser.OCRImagePreprocessingFilters.AddMedian()

' Apply Gamma Correction
'documentParser.OCRImagePreprocessingFilters.AddGammaCorrection(1.4)

' Add Contrast
'documentParser.OCRImagePreprocessingFilters.AddContrast(20)

' Load template
documentParser.AddTemplate(template)

Console.WriteLine("Template loaded.")
Console.WriteLine()

Console.WriteLine($"Parsing "{inputDocument}" with OCR...")
Console.WriteLine()

' Parse document data to JSON format
Dim jsonString As String = documentParser.ParseDocument(inputDocument,
OutputFormat.JSON)

' Display parsed data in console
Console.WriteLine("Parsed data in JSON format:")
Console.WriteLine()
Console.WriteLine(jsonString)

End Using

Console.WriteLine()
Console.WriteLine("Press any key to continue...")
Console.ReadLine()

End Sub

```

End Sub

End Module

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about ByteScout Document Parser 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](#)