

parallel processing of PDF files in VB.NET with ByteScout PDF Extractor SDK

Make parallel processing of PDF files in VB.NET

:

Tutorial on how to do parallel processing of PDF files in VB.NET

The coding tutorials are designed to help you test the features without need to write your own code. ByteScout PDF Extractor SDK was made to help with parallel processing of PDF files in VB.NET. ByteScout PDF Extractor SDK is the SDK is designed to help developers with pdf tables and pdf data extraction from unstructured documents like pdf, tiff, scans, images, scanned and electronic forms. The library is powered by OCR, computer vision and AI to provide unique functionality like table detection, automatic table structure extraction, data restoration, data restructuring and reconstruction. Supports PDF, TIFF, PNG, JPG images as input and can output CSV, XML, JSON formatted data. Includes full set of utilities like pdf splitter, pdf merger, searchable pdf maker and other utilities.

You will save a lot of time on writing and testing code as you may just take the code below and use it in your application. To do parallel processing of PDF files in your VB.NET project or application you may simply copy & paste the code and then run your app! Code testing will allow the function to be tested and work properly with your data.

ByteScout PDF Extractor SDK is available as 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 PDF Extractor SDK](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout PDF Extractor SDK](#)

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

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

Source Code Files:

Module1.vb

```
Imports System.IO
Imports System.Threading
Imports Bytestcout.PDFExtractor

Module Module1

    ' Limit to 4 threads in queue.
    ' Set this value to number of your processor cores for max performance.
    ReadOnly ThreadLimiter As Semaphore = New Semaphore(4, 4)

    <MTAThread(>
    Sub Main()

        ' Get all PDF files in a folder
        Dim files = Directory.GetFiles("..\\..\\..\\..\\", "*.pdf")
        ' Array of events to wait
        Dim doneEvents(files.Length - 1) As ManualResetEvent

        For i As Integer = 0 To files.Length - 1

            ' Wait for the queue
            ThreadLimiter.WaitOne()

            ' Start thread with filename and event in params
            doneEvents(i) = New ManualResetEvent(False)
            Dim threadData As Object = New Object() {files(i), doneEvents(i)}
            ThreadPool.QueueUserWorkItem(AddressOf ConvertPdfToTxt, threadData)

        Next

        ' Wait until all threads finish
        WaitHandle.WaitAll(doneEvents)

        Console.WriteLine()
        Console.WriteLine("All is done.")
        Console.WriteLine()
        Console.WriteLine("Press any key to exit...")
        Console.ReadKey()

    End Sub

    Private Sub ConvertPdfToTxt(ByVal state As Object)

        ' Get filename and event from params
        Dim file As String = state(0)
        Dim doneEvent As ManualResetEvent = state(1)

        Dim resultFileName = Path.GetFileName(file) + ".txt"

        Try
```

```
        Console.WriteLine("Converting " + file)

        Using extractor As New TextExtractor("demo", "demo")

            extractor.LoadDocumentFromFile(file)
            extractor.SaveTextToFile(resultFileName)

        End Using

        Console.WriteLine("Finished " + resultFileName)

    Finally

        ' Signal the thread is finished
        doneEvent.Set()
        ' Release semaphore
        ThreadLimiter.Release()

    End Try

End Sub

End Module
```

VIDEO

https://www.youtube.com/watch?v=s28W3_KMraU

ON-PREMISE OFFLINE SDK

[60 Day Free Trial](#) or [Visit ByteScout PDF Extractor SDK Home Page](#)
[Explore ByteScout PDF Extractor SDK Documentation](#)
[Explore Samples](#)
[Sign Up for ByteScout PDF Extractor SDK 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)

www.bytescout.com