How to find text in pdf with regex with pdf extractor sdk in VB.NET with ByteScout PDF Suite

Learning is essential in computer world and the tutorial below will demonstrate how to find text in pdf with regex with pdf extractor sdk in VB.NET

The sample source codes on this page shows how to find text in pdf with regex with pdf extractor sdk in VB.NET. ByteScout PDF Suite is the set that includes 6 SDK products to work with PDF from generating rich PDF reports to extracting data from PDF documents and converting them to HTML. This bundle includes PDF (Generator) SDK, PDF Renderer SDK, PDF Extractor SDK, PDF to HTML SDK, PDF Viewer SDK and PDF Generator SDK for Javascript. It can find text in pdf with regex with pdf extractor sdk in VB.NET.

This prolific sample source code in VB.NET for ByteScout PDF Suite contains various functions and other necessary options you should do calling the API to find text in pdf with regex with pdf extractor sdk. Follow the instructions from scratch to work and copy the VB.NET code. Further improvement of the code will make it more robust.

If you want to try other source code samples then the free trial version of ByteScout PDF Suite is available for download from our website. Just try other source code samples for VB.NET.

FOR MORE INFORMATION AND FREE TRIAL:

Download Free Trial SDK (on-premise version)

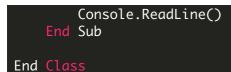Read more about ByteScout PDF Suite

Explore API Documentation

Get Free Training for ByteScout PDF Suite

Get Free API key for Web API

visit www.ByteScout.com

Source Code Files:

Program.vb

```vb
Imports System.Drawing
Imports Bytescout.PDFExtractor

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(".\Invoice.pdf")

        extractor.RegexSearch = True ' Enable the regular expressions

        Dim pageCount As Integer = extractor.GetPageCount()

        ' Search through pages
        For i As Integer = 0 To pageCount - 1

            ' Search dates in format 12/31/1999
            Dim regexPattern As String = "[0-9]{2}/[0-9]{2}/[0-9]{4}"
            ' See the complete regular expressions reference at https://msdn.microsoft

            ' Search each page for the pattern
            If extractor.Find(i, regexPattern, False) Then
                Do
                    Console.WriteLine("")
                    Console.WriteLine(("Found on page " & i & " at location ") + extra
                    Console.WriteLine("")

                    ' Iterate through each element in the found text
                    For Each element As ISearchResultElement In extractor.FoundText.Ele
                        Console.WriteLine("   Text: " + element.Text)
                        Console.WriteLine("   Font is bold: " + element.FontIsBold.ToSt
                        Console.WriteLine("   Font is italic:" + element.FontIsItalic.T
                        Console.WriteLine("   Font name: " + element.FontName)
                        Console.WriteLine("   Font size:" + element.FontSize.ToString()
                        Console.WriteLine("   Font color:" + element.FontColor.ToString
                        Console.WriteLine()
                    Next

                Loop While extractor.FindNext()

            End If
        Next

        ' Cleanup
            extractor.Dispose()

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

```
        Console.ReadLine()
    End Sub

End Class
```

VIDEO

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

ON-PREMISE OFFLINE SDK

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