www.bytescout.com

PDF to JSON in Powershell using ByteScout PDF Extractor SDK

Tutorial: how to do PDF to JSON in Powershell

Writing of the code to PDF to JSON in Powershell can be done by developers of any level using ByteScout PDF Extractor SDK. PDF to JSON in Powershell can be implemented with ByteScout PDF Extractor SDK. ByteScout PDF Extractor SDK is the Software Development Kit (SDK) that is designed to help developers with 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.

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. In order to implement this functionality, you should copy and paste code below into your app using code editor. Then compile and run your application. Use of ByteScout PDF Extractor SDK in Powershell is also explained in the documentation included along with the product.

On our website you may get trial version of ByteScout PDF Extractor SDK for free. Source code samples are included to help you with your Powershell application.

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

Source Code Files:

pdf-to-json.bat

```
@echo off

if "%~1"=="" (
    echo ------------------------------------------------------
    echo Invalid parameter!
    echo ------------------------------------------------------
    echo Usage: pdf-to-json.bat folder_path
    echo Example: pdf-to-json.bat "c:\documents"
    echo ------------------------------------------------------
    if not "%NOPAUSE%"=="1" pause
    exit /b 1
)

powershell -NoProfile -ExecutionPolicy Bypass -Command "& .\pdf-to-json.ps1" "%1"
echo Script finished with errorlevel=%errorlevel%

pause
```

pdf-to-json.ps1

```
Param(
    [Parameter(Mandatory = $true)]
    [string] $InputFolder = ""
)

# Add reference to Bytescout.PDFExtractor.dll assembly
Add-Type -Path "c:\Program Files\Bytescout PDF Extractor SDK\net4.00\Bytescout.PDFExtr

# Check input folder exists
if ((Test-Path $InputFolder) -eq $false) {
    Write-Host "Target folder does not exist." -ForegroundColor Red
    exit 0
}

# Create and activate JSONExtractor instance
$jsonExtractor = New-Object Bytescout.PDFExtractor.JSONExtractor
$jsonExtractor.RegistrationName = "demo"
$jsonExtractor.RegistrationKey = "demo"

try {
    # Get PDF files from input folder
    $files = Get-ChildItem -Path $InputFolder -Recurse -Include "*.pdf"
    foreach ($file in $files) {
        Write-Host "Input file" $file.FullName
        # Construct output file name
        $jsonFileName = [System.IO.Path]::ChangeExtension($file.FullName, "json")
        Write-Host "  Output file" $jsonFileName
```

```
        # Load PDF document
        $jsonExtractor.LoadDocumentFromFile($file.FullName)
        # Disable the formatting reconstruction
        $jsonExtractor.PreserveFormattingOnTextExtraction = $false
        # Extract first page to JSON
        $jsonExtractor.SaveJSONToFile(0, $jsonFileName)
        # Reset extractor
        $jsonExtractor.Reset()
    }
}
catch {
    Write-Host $_.Exception.Message
}

$jsonExtractor.Dispose()
```

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

visit www.PDF.co