



Processor



Extraction



PROCESSOR > GUIDES > EXTRACTION

Extract tables from PDF documents easily



PSPDFKit Processor has been deprecated and replaced by [Document Engine](#). To migrate to Document Engine and unlock advanced document processing capabilities, refer to our migration guide. Learn more about these enhancements on our [blog](#).

This guide explains how to extract table information from PDF documents using Processor.

Before you get started, make sure [Processor is up and running](#).

You can download and use either of the following sample documents for the examples in this guide:

✧ [Example eight-page PDF](#)

✧ [Example four-page PDF](#)

You'll be sending [multipart POST requests with instructions](#) to Processor's `/build` endpoint. To learn more about multipart requests, refer to our blog post on the topic, [A Brief Tour of Multipart Requests](#).

Check out the [API Reference](#) to learn more about the `/build` endpoint and all the actions you can perform on PDFs with PSPDFKit Processor.

Sending the Request to Extract Data

To extract table data from a document, post a multipart request to the `/build` API endpoint. In your request instructions, specify the following output parameters:



ASK AI

- ❖ `type` specifies the output type. Set this to `json-content`.
- ❖ `tables` is a Boolean value that determines whether to extract table data.
- ❖ `language` specifies the language used for recognizing text with optical character recognition (OCR). Sometimes, text is stored in a PDF or an image in a way that makes it so you cannot search or copy it. PSPDFKit's OCR engine allows you to recognize text and save it in a separate file where you can both search and copy and paste the text.

SHELL

HTTP

```
1 curl -X POST http://localhost:5000/api/build \  
2   -F document=@/path/to/example-document.pdf \  
3   -F instructions='{  
4     "parts": [  
5       {  
6         "file": "document"  
7       }  
8     ],  
9     "output": {  
10      "type": "json-content",  
11      "tables": true,  
12      "language": "english"  
13    }  
14  }' \  
15   -o result.pdf
```



For more information on the `/build` instructions, refer to the [API Reference](#).

Example Data Extraction Response

```
1 {  
2   "pages": [  
3     {  
4       "pageIndex": 0,  
5       "tables": [  
6         {  
7           "confidence": 95.4,  
8           "bbox": {  
9             "left": 0,  
10            "top": 0,  
11            "width": 100,  
12            "height": 100  
13          },  
14          "cells": [  
15            {
```



```
15         {
16             "bbox": {
17                 "left": 0,
18                 "top": 0,
19                 "width": 100,
20                 "height": 100
21             },
22             "rowIndex": 0,
23             "columnIndex": 0,
24             "isHeader": true,
25             "text": "Invoice number"
26         }
27     ],
28     "columns": [
29         {
30             "bbox": {
31                 "left": 0,
32                 "top": 0,
33                 "width": 100,
34                 "height": 100
35             }
36         }
37     ],
38     "lines": [
39         {
40             "bbox": {
41                 "left": 0,
42                 "top": 0,
43                 "width": 100,
44                 "height": 100
45             },
46             "isVertical": false,
47             "thickness": 0
48         }
49     ],
50     "rows": [
51         {
52             "bbox": {
53                 "left": 0,
54                 "top": 0,
55                 "width": 100,
56                 "height": 100
57             }
58         }
59     ]
60 }
61 ]
62 }
63 ]
64 }
```

Was this helpful?

✓ YES

✗ NO

Questions? [Contact us](#)

