

How to extract line items from tables on multiple pages for document parser API in Java using PDF.co Web API

PDF.co Web API: the flexible Web API that includes full set of functions from e-signature requests to data extraction, OCR, images recognition, pdf splitting and pdf splitting. Can also generate barcodes and read barcodes from images, scans and pdf.

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about PDF.co Web API](#)

[Explore API Documentation](#)

[Get Free Training for PDF.co Web API](#)

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

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

Source Code Files:

ByteScoutWebApiExample.iml

```
<?xml version="1.0" encoding="UTF-8"?>
<module type="JAVA_MODULE" version="4">
  <component name="NewModuleRootManager" inherit-compiler-output="true">
    <exclude-output />
    <content url="file://$MODULE_DIR{code}quot;>
      <sourceFolder url="file://$MODULE_DIR$/src" isTestSource="false" />
    </content>
    <orderEntry type="inheritedJdk" />
    <orderEntry type="sourceFolder" forTests="false" />
    <orderEntry type="library" name="com.google.code.gson:gson:2.8.1" level="project" />
    <orderEntry type="library" name="com.squareup.okhttp3:okhttp:3.8.1" level="project" />
  </component>
</module>
```

MultiPageTable-template1.yml

```
---
# Template that demonstrates parsing of multi-page table using only
# macro expressions for the table start, end, and rows.
# If macro expression cannot be written for every table row (for example,
# if the table contains empty cells), try the second method demonstrated
# in `MultiPageTable-template2.yml` template.
templateVersion: 3
templatePriority: 0
sourceId: Multipage Table Test
detectionRules:
  keywords:
    - Sample document with multi-page table
  fields:
    total:
      expression: 'TOTAL{{Spaces}}{{Number}}'
      dataType: decimal
  tables:
    - name: table1
      start:
        # macro expression to find the table start in document
        expression: 'Item{{Spaces}}Description{{Spaces}}Price'
      end:
        # macro expression to find the table end in document
        expression: 'TOTAL{{Spaces}}{{Number}}'
      row:
        # macro expression to find table rows
        expression: '{{LineStart}}{{Spaces}}(?<itemNo>{{Digits}}){{Spaces}}(?<description>{{SentenceWithSingleSpaces}})'
      # output data types for columns
      columns:
        - name: itemNo
          type: integer
        - name: description
          type: string
        - name: price
          type: decimal
        - name: qty
          type: integer
        - name: extPrice
          type: decimal
      multipage: true
```

MultiPageTable-template2.yml

```
---
# Template that demonstrates parsing of multi-page table without using
# macro expression for table rows. Rows and cells are extracted automatically
# by specified column coordinates. Use `Template Editor` app to find the coordinates
# (coordinates of the mouse cursor are displayed in the toolbar).
templateVersion: 3
templatePriority: 0
sourceId: Multipage Table Test
detectionRules:
  keywords:
    - Sample document with multi-page table
  fields:
    total:
```

```
type: regex
expression: 'TOTAL{{Spaces}}>{{Number}}'
dataType: decimal
tables:
- name: table1
  # coordinate OR macro expression to find the table start on each document page
  start:
    #y: 136
    expression: 'Item{{Spaces}}Description{{Spaces}}Price'
  # coordinate OR macro expression to find the table end on each document page
  end:
    #y: 787
    expression: '(Page {{Digits}} of {{Digits}})|(TOTAL{{Spaces}}>{{Number}})'
  # left coordinate of the table (optional)
  left: 51
  # right coordinate of the table (optional)
  right: 528
  # column names, output data types and coordinates (left column edge)
  columns:
  - name: itemNo
    x: 51
    type: integer
  - name: description
    x: 102
    type: string
  - name: price
    x: 324
    type: decimal
  - name: qty
    x: 396
    type: integer
  - name: extPrice
    x: 441
    type: decimal
  multipage: true
```

VIDEO

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

ON-PREMISE OFFLINE SDK

[60 Day Free Trial](#) or [Visit PDF.co Web API Home Page](#)
[Explore PDF.co Web API Documentation](#)
[Explore Samples](#)
[Sign Up for PDF.co Web API 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