

www.bytescout.com

How to parse simple document for document parser API in Java using PDF.co Web API

What is PDF.co Web API? It is 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:

AmazonAWS.yml

```
---
templateVersion: 3
templatePriority: 0
sourceId: Amazon Web Services Invoice
detectionRules:
  keywords:
  - Amazon Web Services
  - ATTN
  - Invoice
fields:
  total:
    type: macros
    expression: 'TOTAL AMOUNT DUE ON{{Anything}}{{Dollar}}{{Number}}'
    dataType: decimal
  subTotal:
    type: macros
```

```

expression: '{{LineStart}}{{Spaces}}Charges{{Spaces}}{{Dollar}}{{Number}}'
dataType: decimal
dateIssued:
  type: macros
  expression: 'Invoice Date:{{Spaces}}{{Anything}}{{LineEnd}}'
  dataType: date
  dateFormat: MMMM d , yyyy
invoiceId:
  type: macros
  expression: 'Invoice Number:{{Spaces}}{{Digits}}'
companyName:
  type: static
  expression: Amazon Web Services, Inc.
companyWebsite:
  type: static
  expression: aws.amazon.com
billTo:
  type: rectangle
  expression: 'Bill to Address:{{ToggleSingleLineMode}}{{AnythingGreedy}}'
  rectangle:
    - 33
    - 115.5
    - 213.75
    - 72.75
pageIndex: 0
currency:
  type: static
  expression: USD
tables:
- name: table1
  start:
    expression: '{{LineStart}}{{Spaces}}Detail{{LineEnd}}'
  end:
    expression: '{{EndOfPage}}'
  row:
    expression: '{{LineStart}}{{Spaces}}(?<description>{{SentenceWithSingleSpaces}}){{Spaces}}{{Dollar}}(?<unitPrice>'
  columns:
    - name: unitPrice
      type: decimal

```

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>

```

DigitalOcean.yml

```
---
templateVersion: 3
templatePriority: 0
sourceId: DigitalOcean Invoice
detectionRules:
  keywords:
    # Template will match documents containing the following phrases:
    - DigitalOcean
    - 101 Avenue of the Americas
    - Invoice Number
fields:
  # Static field that will "DigitalOcean" to the result
  companyName:
    type: static
    expression: DigitalOcean
  # Macro field that will find the text "Invoice Number: 1234567" and return "1234567" to the result
  invoiceId:
    type: macros
    expression: 'Invoice Number: {{{Digits}}}'
  # Macro field that will find the text "Date Issued: February 1, 2016" and return the date "February 1, 2016" in ISO format
  dateIssued:
    type: macros
    expression: 'Date Issued: {{{SmartDate}}}'
    dataType: date
    dateFormat: auto-mdy
  # Macro field that will find the text "Total:
```

```
{codeFileName}
```

```
{code}
```

```
10.00" and return "110.00" to the result
total:
  type: macros
  expression: 'Total: {{{Dollar}}>{{Number}}'
  dataType: decimal
  # Static field that will "USD" to the result
  currency:
    type: static
    expression: USD
tables:
  - name: table1
    # The table will start after the text "Description    Hours"
    start:
      expression: 'Description{{Spaces}}Hours'
    # The table will end before the text "Total:"
    end:
      expression: 'Total:'
    # Macro expression that will find table rows "Website-Dev (1GB) 744 01-01 00:00 01-31 23:59"
```

```
{codeFileName}
```

```
{code}
```

0.00", etc.

row:

Groups <description>, <hours>, <start>, <end> and <unitPrice> will become columns in the result table.

expression: '{{LineStart}}{{Spaces}}(<description>{{SentenceWithSingleSpaces}}){{Spaces}}(<hours>{{Digits}}){{S

Suggest data types for table columns (missing columns will have the default "string" type):

columns:

- name: hours

type: integer

- name: unitPrice

type: decimal

Google.yml

```
---
templateVersion: 3
templatePriority: 0
sourceId: Google Invoice
detectionRules:
  keywords:
    - Google
    - 77-0493581
    - Invoice
fields:
  invoiceId:
    expression: 'Invoice number:{{Spaces}}{{Digits}}'
  dateIssued:
    expression: 'Issue date:{{Spaces}}{{SmartDate}}'
    dataType: date
    dateFormat: MMM d, yyyy
  total:
    expression: 'Amount due in USD:{{Spaces}}{{Number}}'
    dataType: decimal
  subTotal:
    expression: 'Subtotal in USD:{{Spaces}}{{Number}}'
    dataType: decimal
  taxRate:
    expression: 'State sales tax {{OpeningParenthesis}}{{Digits}}{{Percent}}{{ClosingParenthesis}}'
    dataType: integer
  tax:
    expression: 'State sales tax{{Anything}}{{Number}}{{LineEnd}}'
    dataType: decimal
  companyName:
    type: static
    expression: 'Google LLC'
  billTo:
    type: rectangle
    rectangle:
      - 0
      - 152
      - 280
      - 72
    pageIndex: 0
  billingId:
    expression: 'Billing ID:{{Spaces}}{{DigitsOrSymbols}}'
  currency:
    type: static
    expression: 'USD'
tables:
- name: table1
  start:
    expression: 'Description{{Spaces}}Interval{{Spaces}}Quantity{{Spaces}}Amount'
```

```
end:  
  expression: 'Subtotal in USD'  
row:  
  expression: '{{LineStart}}{{Spaces}}(?<description>{{SentenceWithSingleSpaces}}){{Spaces}}(?<interval>{{3Letters}}'  
columns:  
  - name: quantity  
    type: integer  
  - name: amount  
    type: decimal
```

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