



Processor



Key-Value Pairs



PROCESSOR &gt; GUIDES &gt; EXTRACTION &gt; KEY-VALUE PAIRS

# Streamline key-value pair extraction from documents



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](#).

PSPDFKit's key-value pair (KVP) extraction engine enables you to recognize related data items in a document and export them to an external destination like a spreadsheet.

## Key-Value Pair Extraction

Key-value pairs are two related data items: a key and a value. Depending on the type of document, the key-value pairs are different. For example, key-value pairs on invoices can be the following:

KEY	VALUE
Invoice Number	No 00162
Billing Date	20/09/2022
Total	1,165.10€

Key-value pair fields on government forms can be the following:



ASK AI

KEY	VALUE
Company Name	Nutrient GmbH
Registration Number	FN 548939p
Date of Incorporation	04/10/2013

It's easy to get key-value pairs from structured documents like Excel files because the values are all named. However, 90 percent of all documents have unstructured data. For these documents, you need a KVP extraction tool to retrieve the information. Intelligent document processing (IDP) extracts data from unstructured and semi-structured documents using OCR and artificial intelligence (AI) technologies.

By automatically recognizing the document type, PSPDFKit adapts to the context and determines the extraction approach that makes the best use of available resources. PSPDFKit recognizes the document type based on adaptive layout understanding and natural language processing (NLP) technologies. This hybrid approach includes heuristics, mathematics, and machine learning (ML) capabilities that address the usual weaknesses of traditional OCR and pure ML engines.

## How to Use Key-Value Pair Extraction

This example retrieves key-value pairs, such as the invoice number and billing date, from the scanned invoice below.

# Invoice

## No 00162

Customer	Vandelay Industries Around the Corner 13 NBC City
Delivery address	Vandelay Industries Around the Corner 13 NBC City
Invoice number	No 00162
Billing date	20/09/2022
Ref. number	34751
Order date	20/09/2022

Reference	Description	VAT	Price per unit (excl. VAT)	Quantity	Total (excl. VAT)
P00201	<b>Lake Mirror</b> Mirror with brass frame	5.5%	320.00€	1	320.00€
P00222	<b>Mont Blanc Chair</b> Black steel frame with black shell	5.5%	450.00€	2	900.00€
Subtotal					1,220.00€
Discount (10%)					-122.00€
VAT (5.5%)					+67.10€
Shipping cost					0.00€
<b>TOTAL</b>					<b>1,165.10€</b>

David Bricklane, CEO



Upper Arc Warp 70  
Summerfield  
Republic of PDF

+100 847 738 227  
info@bruuuk.com  
www.bruuuk.com

Company No. 400890w  
Tax No. AT98765321  
IBAN AT13 2060 4236 6111 5994



KEY	VALUE
Billing date	20/09/2022
Order date	20/09/2022
Republic of PDF	+100 847 738 227
IBAN	AT13 2060 4236 6111 5994
Customer	Vandelay Industries Around the Corner 13 NBC City
Delivery address	Vandelay Industries Around the Corner 13 NBC City
Invoice number	No 00162
Ref. number	34751
No	00162
Reference	P00201
Quantity Total (excl. VAT)	320.00€
Subtotal	1,220.00€
Discount (10%)	-122.00€
VAT (5.5%)	+6710€
Shipping cost	0.00€
TOTAL	1,165.10€
Description	Lake Mirror
VAT	5.5%
Price per unit (excl. VAT)	320.00€
Tax No.	AT98765321
#	<a href="mailto:infe@bruuuk.com">infe@bruuuk.com</a>

KEY	VALUE
#	<a href="http://www.bruuuk.com">www.bruuuk.com</a>

This table also contains information about the data type and the confidence level for each key-value pair:

- ✧ The data type describes the nature of the content. In this example, the engine recognizes the value `infe@bruuuk.com` as an email and the value `+100 847 738 227` as a phone number.
- ✧ The confidence level describes how confident the KVP engine is in the accuracy of the data extraction.

In this example, the KVP engine automatically detected all key-value pairs in the document with minimal code and without any preconfiguration. The engine supports many formats and languages, and it has no dependencies to external models, resources, and databases.

## Next Steps

- ✧ [Data Model](#)
- ✧ [Confidence Score](#)
- ✧ [Using the Data Extraction API](#)
- ✧ [Extract Data from Bank Statements](#)
- ✧ [Extract Data from Tables](#)
- ✧ [Extract Text](#)

Was this helpful?

☒ YES

☐ NO

Questions? [Contact us](#)