PDF2DATA

What is iText pdf2Data?

iText pdf2Data is a solution to easily recognize and extract data from documents in a structured, reusable format. It is available for Java and C# (.NET), and as a CLI version.
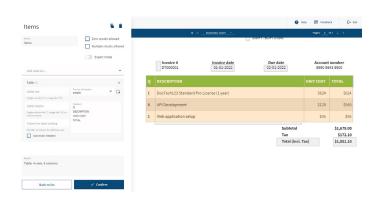
It offers a framework to intelligently recognize data inside PDF documents, based on selection rules that you define in a template. This offers significant advantages over AI-based alternatives which need extensive training to recognize documents.

Thanks to the intuitive browser-based pdf2Data Editor (also available as a Docker image), anyone, from marketers to information managers to HR staff, can create and update templates. You don't need to be a developer to benefit from using iText pdfData.

How does iText pdf2Data work?

Using the user-friendly browser-based pdf2Data Editor, you define extraction rules for the areas, fonts, patterns, or tables of interest in a template that is used for all PDFs created in the same format, such as an invoice or other commercial documents.

Then, using the Java or .NET pdf2Data SDK you parse your documents against the pre-defined template with just a few lines of code.

```
Template template = Pdf2DataExtractor.parseTemplateFromPDF(pathToPdfTemplate);
Pdf2DataExtractor extractor = new Pdf2DataExtractor(template);
ParsingResult result = extractor.recognize(pathToFileToParse);
result.saveToXML(pathToOutXmlFile);
```

Java code example

```
Template template = Pdf2DataExtractor.ParseTemplateFromPDF(pathToPdfTemplate);
Pdf2DataExtractor extractor = new Pdf2DataExtractor(template);
ParsingResult result = extractor.Recognize(pathToFileToParse);
result.SaveToXML(pathToOutXmlFile);
```

.NET code example

Smart tip: Combine iText pdf2Data with pdfOCR to recognize and extract data from scanned documents.

PDF2DATA

Features

Automate PDF data extraction from PDF invoices, forms and other documents

Extract and process data from small or large volumes of PDFs by defining the information that is important for your data processes in a template.

Define which specific data you want to target for PDF data extraction

Easily define the desired information you want to extract in a template with the pdf2Data Editor. Data extraction works with all PDF documents with repeatable and predictable formats, such as invoices, forms, reports etc. and makes PDF data processing a highly efficient part of your workflow.

Integrate automated PDF data extraction into your existing document process

iText pdf2Data exports its data as XML or JSON, making it easy to integrate into existing workflows. In addition to the easy to use pdf2Data Editor, it also includes developer-focused SDKs for Java and .NET (C#) as well as a command line interface. PDF data processing for the 21st century.

Better than AI-based alternatives?

Since the content recognition is based on selectors you define in the template, iText pdf2Data requires no prior training to recognize and extract data. The data recognition uses a number of rules, which need to be defined in advance per each data field. Typical rules use all details from the PDF document, and can be combined for correct data extraction.

> **iText pdf2Data only needs a single example document to begin automated data extraction.**

We're also developers!

In our over 20 years of code, we know how important it is to have good documentation, and good processes in place. You can always find comprehensive documentation and code examples online. Commercial customers can also count on our world-class in-house customer support, offering direct access to expert assistance from our development team.

Get iText pdf2Data

Scan the QR-code or go to the link below:

https://itextpdf.com/products/itext-pdf2data/extract-your-content-from-pdf