

pdf2Data

an iText 7 add-on



Unlock Your Data

Data is an important commodity, and you may have more than you realize locked inside your PDF documents. **pdf2Data** is an iText 7 add on that enables you to extract and process PDF data by defining the information that is important to you and pulling it out, programmatically. It works best by creating a template PDF to set the fields that you will extract from. This way all of your data can be reliably extracted within your set fields. This allows you to extract information from thousands of PDFs quickly and easily.

Selection Process

Choosing your data is simple. Use the selectors to pick your desired fields and pdf2Data will extract it. Using selectors gives you the power to choose the data you want pulled out in your existing document process. Keeping your process fast, and now, more efficient.

Key Advantages

- Unlock data in your PDF documents for use
- Define desired information in a template
- Automate data extraction from your PDF invoices and documents



How Does It Work?

pdf2Data works by defining the areas, fonts, patterns, or tables of interest in a template that is used for all PDFs created in the same format. You then can define the areas of interest with selectors. Each selector uses a different way of identifying the information that is important and can be used in conjunction or alone to meet your needs. For example, if you wish to extract an address that is always located within the same area on an invoice PDF, then the boundary selector can be used to define that area. To get a more specific piece of information, you can use both the font family selector and the regular expression selector together. This allows you to extract the data you want quickly within your document workflow.



Example

Some Selectors Include:

Font family	Font size	Font style
Regular expression (user mode, expert mode) - Allows for specific text to be selected by matching a pattern	Boundary	Align (GUI option - pick all text lines left/right aligned with the reference one)
Paragraphs	Lines	Table
Picker	Pattern	Frequency table (Expert mode)
Date	IBAN	Special keyword
VAT	Font	Price