

# Apache PDFBox Tutorial – Learn to create, edit and process PDFs

Learn to create, edit and process PDFs using [Java](#) by following this informative Apache PDFBox Tutorial.

## APACHE PDFBox



### Apache PDFBox Tutorial

#### About Apache PDFBox

Apache PDFBox is an open source from Apache Software Foundation. The tool is built in Java to work with Pdf documents. The tool is used to create, process and modify (or edit) pdf documents. It also contains command-line utilities.

[Setup a Java project with pdfbox libraries](#) to start working on pdf files.

### Features of Apache PDFBox

Following are the features and possibilities feasible with the tool :

#### ▪ Extract Text and Images

1. Extract text from PDF file.
2. Extract position and size of characters in the PDF file.
3. Extract words from PDF document.
4. Extract text line by line from PDF document.
5. Get position and size of images in the PDF.
6. Extract images from PDF.

#### ▪ Split and Merge

1. Split a PDF file into multiple PDF files.
2. Merge multiple PDF files into a single PDF file.

#### ▪ Fill Forms

1. Extract data from PDF form.
2. Fill a PDF form.

#### ▪ filePreflight

1. Validate PDF file against PDF/A-1b standard.

## ■ [Print](#)

1. Print a PDF file programmatically.

## ■ [Save as Image](#)

1. Save pages in PDF file as images.

## ■ [Create PDFs](#)

1. [Create a new PDF file and write text to it.](#)

## ■ [Signing](#)

1. Digitally sign PDF file.

### PDFBox

▸ [PDFBox Tutorial](#)

▸ [Setup Java Project with PDFBox](#)

### Text Processing

▸ [Create a PDF file with Text](#)

▸ [Read all the text from PDF](#)

▸ [Extract coordinates or position of characters in PDF](#)

▸ [Extract words from PDF](#)

▸ [Read text line by line from PDF](#)

▸ [PDFBox - Split PDF Document](#)

▸ [PDFBox - Merge multiple PDFs](#)

### Image Processing

▸ [Get Location and Size of Images](#)

▸ [Extract Images from PDF](#)