Apache PDFBox | A Java PDF Library

(/index.html)

Dependencies (/dependencies.html)
References (/references.html)

COMMUNITY

Support (/support.html)

Mailing Lists (/mailinglists.html)

Project Team (/team.html)

DOCUMENTATION

Architecture (/architecture.html)

Command Line Tools (/commandline/)

PDFBox Cookbook ▼

API Docs (/apidocs/)

FOR DEVELOPERS

FAQ (/userguide/faq.html)

Building PDFBox (/building.html)

Ideas (/ideas.html)

Coding Conventions (/codingconventions.html)

APACHE SOFTWARE FOUNDATION

Apache Software Foundation (http://www.apache.org/)

PDF Library

The Apache PDFBox™ library is an open source Java tool for working with PDF documents. This project allows creation of new PDF documents, manipulation of existing documents and the ability to extract content from documents. Apache PDFBox also includes several command line utilities. Apache PDFBox is published under the Apache License v2.0.

To get help on using PDFBox, please subscribe to the Users Mailing List and post your questions there. We're happy to help.

The project is a volunteer effort and we're always looking for interested people to help us improve PDFBox. There are a multitude of ways that you can help us depending on your skills. Subscribe to the Mailing Lists (/mailinglists.html) and find out how you can help.

Features

☐ **Text Extraction**
Extract from within a PDF for usage in other applications.

☐ **Merging & Splitting**
Merge multiple PDFs into one or split a single PDF into multiple PDFs.

☐ **Forms Filling**
Extract forms data from PDF forms or prefill a PDF form.

☐ **PDF/A Validation**
Validate PDFs against the PDF/A ISO standard.

☐ **PDF Printing**
Print a PDF file to printers supported by the Java printing API.

☐ **PDF to Image Conversion**
Convert you PDFs to Image files.

☐ **PDF Creation**
Create a PDF from scratch.

☐ **Lucene Search Engine Integration**
Integrate PDF indexing with the Lucene search engine.

...End of Fax