PDFBox - References

Table of contents

1. PDFBox References

This page lists projects that utilize PDFBox and articles that have been written about PDFBox. Send me an e-mail if your article or project is missing.

2. Projects

Project Name	License	Project Description
Alfresco	LGPL - commercial services/support/training is available	Alfresco is an open source, open-standards content repository built by the most experienced content management team that includes the co-founder of Documentum.
Centric CRM	Free To Use But Restricted/Commercial	The Most Advanced Open Source CRM Software.
Canoo Webtest	BSD Like	Free OpenSource tool for XP-style acceptance testing of Java-based Web applications.
Jahia	collaborative source license	The Jahia product is currently the most powerful, ready-to-use and affordable integrated midrange Java Content Management and Corporate Portal Server.
jLibrary	BSD	jLibrary is a Document Management System, oriented for personal and enterprise use.
Jomic	GPL	Jomic is a viewer for comic book archives.
JpdfUnit	Apache License V2.0	JpdfUnit is a framework for testing a generated pdf document with the JUnit Test Framework.
Liferay Portal	MIT	Liferay Portal is an open source portal that helps organizations collaborate more efficiently by

		providing a consolidated view of disparate applications.
[LIUS](#)	GPL	LIUS is an indexing Java framework based on the Jakarta Lucene project. The LIUS framework adds to Lucene many files format indexing fonctionalities as: Ms World, Ms Excel, Ms PowerPoint, RTF, PDF, XML, HTML, TXT, Open Office suite and JavaBeans.
[LuceGene](#)	Artistic License	LuceGene is an open-source document/object search and retrieval system specially tuned for bioinformatics text databases and documents.
[MMBase Lucene Module](#)	MPL	Lucenemodule is a plugin (module) for the MMBase content management system that enables Lucene full text search through it's content, and thanks to PDFBox also PDF content.
[Nutch](#)	ASL	Nutch is open source web-search software. It builds on Lucene, adding web-specifics, such as a crawler, a link-graph database, parsers for HTML and other document formats, etc.
[OpenCms](#)	[Custom](#)	OpenCms is a professional level Open Source Website Content Management System.
[Orbeon PresentationServer](#)	LGPL	Orbeon PresentationServer (OPS) is an open source J2EE-based platform for XML-centric web applications. OPS is built around XHTML, XForms, XSLT, XML pipelines, and Web Services, which makes it ideal for applications that capture, process and

		present XML data. Commercial consulting/training/support is available through [orbeon]().
[PDFcat]()	LGPL	PDFcat is multi-platform catalog manager that provides searching capability over documents among virtual catalogs.
[PodReader]()	GPL	PodReader is an application that facilitates making electronic documents like eBooks readable on your iPod.
[SearchBlox]()	Commercial	SearchBlox is a high-performance corporate search software designed for the Java 2 Enterprise Edition (J2EE) platform.
[Terrier]()	MPL	Terrier is software for the rapid development of Web, intranet and desktop search engines.
[Triboni GinkGO]()	Commercial	Triboni GinkGO is a highly scalable J2EE services platform that is based on a simple XML business object defintion and scripting language. Toghether with XSLT content centric web applications can be configured in a very short time.
[Zilverline]()	[Collaborative Source License]()	Zilverline is a search engine that offers web access to your personal or intranet content.

3. Articles/Books

Article Name	Article Abstract
Build an eDoc Reader for your iPod [Part 1]() - User Interface [Part 2]() - Document Reading Engine [Part 3]() - *Integration with PDFBox*	A three part article that discusses the implementation of the PodReader application. PodReader is Cocoa application written in Objective-C and article discusses how to use the Cocoa-Java bridge to integrate with the Java

	version of PDFBox.
[Lucene In Action](#)	A book that discusses integrating with the lucene search engine. One chapter discusses how to index various file formats and highlights PDFBox for indexing PDF documents.
[Java Developers Journal - March 2005](#)	An article written by the lead developer of PDFBox discussing text extraction and AcroForm integration using PDFBox functionality.
[Refactoring trends across N versions of N Java open source systems: an empirical study](#)	This article describes an empirical study of multiple versions of a range of open source Java systems in an attempt to understand whether refactoring occur and, if so, which types of refactoring were most (and least) common. PDFBox is used as a case study.