

# History of Changes

## Table of contents

1	Introduction and explanation of symbols.....	4
2	Version 0.8.0 (first apache release).....	4
2.1	Changes to the Code Base.....	4
3	Version 0.7.4-dev (never released).....	8
3.1	Changes to the Code Base.....	8
4	Version 0.7.3 (10/12/2006).....	8
4.1	Changes to the Code Base.....	8
4.2	Changes to Documentation.....	10
5	Version 0.7.2 (09/11/2005).....	10
5.1	Changes to the Code Base.....	10
6	Version 0.7.1 (04/10/2005).....	11
6.1	Changes to the Code Base.....	11
7	Version 0.7.0 (1/22/2005).....	12
7.1	Changes to the Code Base.....	12
8	Version 0.6.7 (10/09/2004).....	13
8.1	Changes to the Code Base.....	13
9	Version 0.6.6 (07/20/2004).....	14
9.1	Changes to the Code Base.....	14
10	Version 0.6.5 (03/08/2004).....	15
10.1	Changes to the Code Base.....	15
11	Version 0.6.4 (11/02/2003).....	16
11.1	Changes to the Code Base.....	16
12	Version 0.6.3 (09/13/2003).....	17
12.1	Changes to the Code Base.....	17

13	Version 0.6.2 (4/18/2003)	17
13.1	Changes to the Code Base	17
14	Version 0.6.1 (3/9/2003)	18
14.1	Changes to the Code Base	18
15	Version 0.6.0 (3/5/2003)	18
15.1	Changes to the Code Base	18
16	Version 0.5.6 (11/28/2002)	18
16.1	Changes to the Code Base	18
17	Version 0.5.5 (10/03/2002)	19
17.1	Changes to the Code Base	19
18	Version 0.5.4 (09/17/2002)	19
18.1	Changes to the Code Base	19
19	Version 0.5.3 (09/13/2002)	19
19.1	Changes to the Code Base	19
20	Version 0.5.2 (09/06/2002)	19
20.1	Changes to the Code Base	19
21	Version 0.5.1 (09/04/2002)	19
21.1	Changes to the Code Base	20
22	Version 0.5.0 (08/31/2002)	20
22.1	Changes to the Code Base	20
23	Version 0.4.1 (07/25/2002)	20
23.1	Changes to the Code Base	20
24	Version 0.4.0 (07/23/2002)	20
24.1	Changes to the Code Base	20
25	Version 0.3.0 (07/09/2002)	20
25.1	Changes to the Code Base	20
26	Version 0.2.0 (06/03/2002)	21
26.1	Changes to the Code Base	21
27	Version 0.1.0 (05/25/2002)	21

27.1 Changes to the Code Base..... 21

## 1. Introduction and explanation of symbols

Changes are sorted by "false" and then chronologically with the most recent at the top. These symbols denote the various action types: =add, =fix, =remove, =update

## 2. Version 0.8.0 (first apache release)

### 2.1. Changes to the Code Base

- [PDFBOX-51] PDFToImage fails to render correctly Committed by BJJ.
- [PDFBOX-93] Error in FlateFilter? Committed by BJJ.
- [PDFBOX-94] Unexpected end of ZLIB input stream Committed by BJJ.
- [PDFBOX-107] viewer crashed Committed by BJJ.
- [PDFBOX-110] bad font data with TrueTypeFont Committed by BJJ.
- [PDFBOX-141] PDF to image conversion can lead to mostly black area Committed by BJJ.
- [PDFBOX-148] Error getting pdf version (NumberFormatException) Committed by BJJ.
- [PDFBOX-152] Merge Landscape and Portrait PDFs does not keep orientation Committed by BJJ.
- [PDFBOX-162] font spacing Committed by BJJ.
- [PDFBOX-173] Some suggested COSString improvements Committed by BJJ.
- [PDFBOX-178] splitting some words randomly Committed by BJJ.
- [PDFBOX-183] java.lang.NullPointerException in highlighter.generateXMLHig Committed by BJJ.
- [PDFBOX-187] Error in parsing CMap file Committed by BJJ.
- [PDFBOX-211] Regression: ArrayIndexOutOfBoundsException in PDFBox 0.7.3 Committed by BJJ.
- [PDFBOX-221] NPE on convertToImage Committed by BJJ.
- [PDFBOX-223] CurrentColor in PageDrawer Doesn't Restore Properly Committed by BJJ.
- [PDFBOX-224] Printing Rectangles on rotated pages Committed by BJJ.
- [PDFBOX-227] ArrayIndexOutOfBoundsException:4 Committed by BJJ.
- [PDFBOX-234] spaces lost Committed by BJJ.
- [PDFBOX-249] Imbricated XObjects with the same name Committed by BJJ.
- [PDFBOX-250] Table borders not printing correctly Committed by BJJ.
- [PDFBOX-286] PDF document renders incorrectly Committed by BJJ.
- [PDFBOX-290] java.lang.NoSuchMethodError in fontbox Committed by BJJ.
- [PDFBOX-292] Text Extraction strips 1 char when extracting a twin pair Committed by BJJ.

- [PDFBOX-296] Extreme memory usage while extracting text from one pdf Committed by BJJ.
- [PDFBOX-313] OutOfMemoryError for larger PDF text extraction Committed by BJJ.
- [PDFBOX-318] Error getting pdf version Committed by BJJ.
- [PDFBOX-321] PDF printing or conversion : lines are too thick - SOLVED ? Committed by BJJ.
- [PDFBOX-324] One rectangle missing when converting PDF to image Committed by BJJ.
- [PDFBOX-330] Watermarks aren't correctly showed Committed by BJJ.
- [PDFBOX-335] Version incompatibility with Lucene? Committed by BJJ.
- [PDFBOX-343] java.lang.ClassCastException: org.pdfbox.cos.COSArray cannot Committed by BJJ.
- [PDFBOX-348] java.lang.NoClassDefFoundError: org/fontbox/afm/AFMParser Committed by BJJ.
- [PDFBOX-349] Spaces between words ignored in scanned pdf files Committed by BJJ.
- [PDFBOX-361] NullPointerException in PDPageNode.getAllKids Committed by BJJ.
- [PDFBOX-364] Latest trunk uses Java 5 autoboxing Committed by BJJ.
- [PDFBOX-373] (null) printed when characters cannot be decoded during text extraction Committed by BJJ.
- [PDFBOX-374] text areas not properly being sorted because of page rotation Committed by BJJ.
- [PDFBOX-377] Incorrect direction of extracted Arabic Text Committed by BJJ.
- [PDFBOX-379] PDType1Font uses the Java 5 constant Font.TYPE1 Committed by BJJ.
- [PDFBOX-385] ClassCastException when call parseCOSArray in BaseParser.java Committed by BJJ.
- [PDFBOX-390] org.pdfbox.filter.ASCIIHexFilter does not skip Whitespace Committed by BJJ.
- [PDFBOX-393] Maven files in jempbox do not work in Eclipse. Committed by BJJ.
- [PDFBOX-395] NPE on public key encryption of an unencrypted document Committed by BJJ.
- [PDFBOX-396] Incorrect permissions after decryption Committed by BJJ.
- [PDFBOX-401] setStrokingColorSpace and setNonStrokingColorSpace in PDPageContentStream doesn't work correct Committed by BJJ.
- [PDFBOX-404] ClassCastException in COSDictionaryMap Committed by BJJ.
- [PDFBOX-407] PDLineDashPattern missing call to super.clone() Committed by BJJ.
- [PDFBOX-409] Small hashcode issue, The code invokes hashCode on an array. Committed by BJJ.
- [PDFBOX-415] Errors when decomposing Arabic Ligatures Committed by BJJ.
- [PDFBOX-418] PDFStreamParser reads incorrect number (patch provided) Committed by BJJ.

- [PDFBOX-421] Unit tests are failing Committed by BJJ.
- [PDFBOX-425] Silent print ignores passed PrintJob Committed by BJJ.
- [PDFBOX-426] Class StrokePath has the wrong superclass Committed by BJJ.
- [PDFBOX-428] Error Printing: dash lengths all zero Committed by BJJ.
- [PDFBOX-436] PDFFontFactory.createFont returns null if the given parameter fontCache is null Committed by BJJ.
- [PDFBOX-438] FlateFilter: endless loop because of missing length check (for encrypted pdfs) Committed by BJJ.
- [PDFBOX-442] race condition in PdfFont Committed by BJJ.
- [PDFBOX-446] A empty page produces a NPE Committed by BJJ.
- [PDFBOX-450] PDFTextStripper CAN NOT extract correct font information for some early produced PDF documents Committed by BJJ.
- [PDFBOX-452] [patch] maven build errors in current trunk Committed by BJJ.
- [PDFBOX-453] FlateFilter decode() throwing OutOfMemoryError Committed by BJJ.
- [PDFBOX-454] IOException upon opening a PDF Committed by BJJ.
- [PDFBOX-455] java.lang.ClassCastException: org.pdfbox.cos.COSString cannot be cast to org.pdfbox.cos.COSName Committed by BJJ.
- [PDFBOX-456] PDFTextStripperByArea never finds any text (pageNo check in PDFTextStripper always returns false) Committed by BJJ.
- [PDFBOX-458] Wrong implementation of COSArray.getInt() Committed by BJJ.
- [PDFBOX-459] Trailer Dictionary object labeled "Size" is overwritten when there are 2 xref table objects Committed by BJJ.
- [PDFBOX-466] error parsing files generated by crystal reports Committed by BJJ.
- [PDFBOX-468] index out of bounds exception Committed by BJJ.
- [PDFBOX-470] corrupt zip stream causes document to not parse Committed by BJJ.
- [PDFBOX-471] invalid dictionary crashes parser Committed by BJJ.
- [PDFBOX-473] attempt to push back when content read Committed by BJJ.
- [PDFBOX-474] invalid xref entry causes parser to fail Committed by BJJ.
- [PDFBOX-477] extra spaces added to rotated text Committed by BJJ.
- [PDFBOX-478] PDFToImage don't render text in iText generated PDF Committed by BJJ.
- [PDFBOX-482] DeviceCMYK support in PDColorSpaceFactory Committed by BJJ.
- [PDFBOX-483] rendering issues during clipping (W/W\*-operator) Committed by BJJ.
- [PDFBOX-485] Fonts not printed on HP laserjet (1320 & 8150) when having landscape orientation Committed by BJJ.
- [PDFBOX-487] Font size not rendered with the needed precision Committed by BJJ.
- [PDFBOX-496] PDDocument.load hangs when loading zero-length file Committed by BJJ.
- [PDFBOX-498] some pdf-files have no newline after endobj, pdfbox fails on that Committed by BJJ.

- [PDFBOX-503] PDF loader causes infinite loop on non-PDF inputs Committed by BJJ.
- [PDFBOX-511] JVM crash in PDColorSpaceInstance.createColor() Committed by BJJ.
- [PDFBOX-512] org.apache.pdfbox.pdmodel.PDDocument.getPageMap() always returns null Committed by BJJ.
- [PDFBOX-518] ArrayIndexOutOfBoundsException in PDInlinedImage.java Committed by BJJ.
- [PDFBOX-523] IndexOutOfBoundsException in SetNonStrokingICCBasedColor Committed by BJJ.
- [PDFBOX-528] Inlined images with a ColorSpace in the resource dictionary don't work Committed by BJJ.
- [PDFBOX-529] Inline images can be drawn in the wrong place Committed by BJJ.
- [PDFBOX-302] Improve font handling (was: layout print problem) Committed by BJJ.
- [PDFBOX-319] Implementation of PDDeviceCMYK.createColorModel() Committed by BJJ.
- [PDFBOX-358] Vertical text extraction splitting text Committed by BJJ.
- [PDFBOX-363] Fixed Page rotation Committed by BJJ.
- [PDFBOX-365] Updating Lucene version (was: Error in LucenPDFDocument class) Committed by BJJ.
- [PDFBOX-368] Use the Maven standard directory layout Committed by BJJ.
- [PDFBOX-376] Remove the js.jar file Committed by BJJ.
- [PDFBOX-380] Limited support for SC and SCN operator Committed by BJJ.
- [PDFBOX-381] Remove direct JAI dependency Committed by BJJ.
- [PDFBOX-387] new Maven pom.xml files for pdfbox, fontbox, and jempbox Committed by BJJ.
- [PDFBOX-389] Support for b\*, B\*, d, i, j and J operator Committed by BJJ.
- [PDFBOX-405] Not a bug, but definitely incorrect code in PDPageContentStream Committed by BJJ.
- [PDFBOX-437] Prepare JempBox and FontBox for release Committed by BJJ.
- [PDFBOX-460] [PATCH] Improvements for bitmap production (resolution and color depth) Committed by BJJ.
- [PDFBOX-461] Disable javadoc creation timestamp Committed by BJJ.
- [PDFBOX-472] use commons logging Committed by BJJ.
- [PDFBOX-507] [PATCH] Option to disable close warning in finalizer of COSDocument. Committed by BJJ.
- [PDFBOX-517] PD Fonts do not attempt to load from the environment if an embed font cannot be loaded Committed by BJJ.
- [PDFBOX-524] Drop LoggingObject Committed by BJJ.
- [PDFBOX-530] Forward test logs to a log file Committed by BJJ.
- [PDFBOX-98] Print PDF Committed by BJJ.
- [PDFBOX-264] colorspace as an array entry Committed by BJJ.

- [PDFBOX-272] Identify text rotation angle in TextPosition Committed by BJJ.
- [PDFBOX-338] pdf page extraction Committed by BJJ.
- [PDFBOX-493] Ability to get page number for bookmarks Committed by BJJ.

### 3. Version 0.7.4-dev (never released)

#### 3.1. Changes to the Code Base

- [ 1648374 ] Unexpected end of ZLIB input stream Committed by BJJ.
- [ 1806912 ] Access to stream arrays Committed by BJJ.
- [ 1793593 ] PDFStreamEngine doesn't handle rotation of -90. Committed by BJJ.
- [ 1792084 ] spot colors' getTintTransform needs quick fix Committed by BJJ.
- Added dependency on JAI Committed by BJJ.
- [ 1666960 ] Patch for any kind of tiff based pdf. Committed by BJJ.
- Fixed issue with "Could not read all parameters to decode image" Committed by BJJ.
- [ 1726446 ] Fixing merge w&w/o acroform Committed by BJJ.
- [ 1745133 ] OutOfMemoryError exception during text Committed by BJJ.
- Upgraded to lucene 2.2.0 Committed by BJJ.
- [ 1672991 ] Specify non default printer in silentPrint mode. Committed by BJJ.
- [ 1607074 ] Portrait and Landscape PageFormat Fix Committed by BJJ.
- \*\*\*API Change:Renamed PDPageContentStream.drawImage to drawXObject\*\*\* Committed by BJJ.
- [ 1600672 ] Painting Form XObjects Committed by BJJ.
- Added integration with JempBox library Committed by BJJ.
- Upgraded to IKVM 0.32.0.0 Committed by BJJ.
- [ 1674903 ] InvalidCastException in write2file for images Committed by BJJ.
- [ 1057560 ] Overlay incorrectly orders overlay pages Committed by BJJ.
- Added ability to get/set document level JavaScript. Committed by BJJ.
- Fixed issue with PNG encoding when using LZW/Flate filters Committed by BJJ.
- [ 1562539 ] Implemented "B" operator for PageDrawer Committed by BJJ.
- [ 1562539 ] PageDrawer now uses the correct color to display text Committed by BJJ.
- Upgraded to junit4.3 Committed by BJJ.
- Upgraded to Lucene-2.1.0 Committed by BJJ.

### 4. Version 0.7.3 (10/12/2006)

#### 4.1. Changes to the Code Base

- Upgraded to Checkstyle 4.2 Committed by BJJ.
- Upgraded to IKVM 0.30.0.0 Committed by BJJ.

- [ 1546399 ] Use get/set functions for separators in PDFTextStripper Committed by BJL.
- PDDocument.silentPrint() to print without prompting for a printer Committed by BJL.
- [ 1544118 ] Bug in PDFont.getCodeFromArray Committed by BJL.
- [ 1529835 ] Add COSFloat.setValue() Committed by BJL.
- [ 1492555 ] PDChoiceField dead loop Committed by BJL.
- [ 1499521 ] NPE PDAppearance.convertToMultiLine Committed by BJL.
- [ 1522007 ] Error converting date Committed by BJL.
- Upgraded to Lucene 2.0.0 Committed by BJL.
- [ 1451164 ] Problems filling combo and radio form fields Committed by BJL.
- Upgraded to lucene 1.9.1 Committed by BJL.
- [ 1023133 ] Support PDF Functions Committed by BJL.
- Added command line org.pdfbox.PDFMerger Committed by BJL.
- \*\*\*API Change\*\*\* Promoted AppendDoc from example to util package, renamed to PDFMergerUtility. Committed by BJL.
- Upgraded to IKVM-0.24.0.1 Committed by BJL.
- [ 1391952 ] Problem extracting embedded attachments Committed by BJL.
- [ 1249607 ] Fixed issue with broken PDFs that contain multiple endobj Committed by BJL.
- [ 1153174 ] Added documentation for PDFHighlighter Committed by BJL.
- Removed log4j dependency Committed by BJL.
- [ 974661 ] getKids() Null Pointer Exception when parsing pdf Committed by BJL.
- Added better support for CJK encoding Committed by BJL.
- Changed signature of PDFPageContentStream.drawImage to take float arguments instead of int Committed by BJL.
- Fixed issue where form xobjects were not being drawn in the viewer Committed by BJL.
- Changed signature to PDDocumentCatalog.OpenAction to be an PDDestinationOrAction instead of just action. Committed by BJL.
- Added tolerance to text extraction sorting where text on a line was not at the same exact y coordinate but very close Committed by BJL.
- [ 1327133 ] Printing with form data Committed by BJL.
- Fixed issue with DateConverter that was trying to parse an empty string Committed by BJL.
- [ 1324846 ] appending text to PDPageContentStream messes up fonts Committed by BJL.
- Added new example ReplaceURLs to show how to replace a clickable URL in a PDF Committed by BJL.
- Implemented annotation drawing Committed by BJL.
- Implemented EndPath and StrokeAndClosePath operators Committed by BJL.
- Move text extraction permission checking from PDFTextStripper to ExtractText

Committed by BJJ.

- Added support for more annotations, thanks to a contribution from Paul King Committed by BJJ.
- Created new FontBox project to hold all font library code Committed by BJJ.
- Fixed issue where only the first page was sent to the printer Committed by BJJ.
- Now automatically sets the page orientation when printing Committed by BJJ.

## 4.2. Changes to Documentation

- Upgraded to Apache Forrest 0.8-dev Committed by BJJ.

## 5. Version 0.7.2 (09/11/2005)

### 5.1. Changes to the Code Base

- Upgraded to IKVM-0.20.0.0 Committed by BJJ.
- Added support to get annotations from a page and to create a RubberStamp annotation Committed by BJJ.
- Added PDDocument.print() to send the PDF to a printer. Committed by BJJ.
- [ 1276623 ] NullPointerException in PageDrawer:241 when extractin images Committed by BJJ.
- Allow creation of PDJpeg from a BufferedImage, thanks to contribution from Paul King Committed by BJJ.
- Removed PDTiff in favor of PDCcitt Committed by BJJ.
- PDFBox no longer requires log4j!! Committed by BJJ.
- New class to allow you to specify 'named' regions where text is to be extracted. Committed by BJJ.
- [ 1261555 ] Unexpected end of ZLIB input stream when stream has a zero length Committed by BJJ.
- [ 1226665 ] ImportXFDF giving NPE error Committed by BJJ.
- renamed COSDictionary.setItem( String, boolean ) to COSDictionary.setBoolean( String, boolean ) Committed by BJJ.
- Added sorting parameter to PDFTextStripper Committed by BJJ.
- Fixed issues with PDF encryption Committed by BJJ.
- Better date support, added support for PDFs that use non standard dates, support for time zone offsets Committed by BJJ.
- FlateFilter-class now supports PNG-Predictors for decoding the imagedata, thanks to a contribution from Marcel Kammer Committed by BJJ.
- Added support for extracting tiff images, thanks to a contribution from Marcel Kammer Committed by BJJ.

- Added PDDocument.removePage to remove PDF pages Committed by BJJ.
- Fixed issue when creating a COSString with a UTF 16 string Committed by BJJ.
- Committed patch for type 1 PFB font parser(special thanks to Michael Niedermair) Committed by BJJ.
- Committed patch for PNG predictors (special thanks Erik Martino) Committed by BJJ.
- [ 1227428 ] failure of getMediaBox Committed by BJJ.
- [ 1227426 ] null pointer in PDFToImage(ColorModel is null) Committed by BJJ.
- [ 1207113 ] Enhancement: runtime accessible version Committed by BJJ.
- [ 1213320 ] setFfFlag() of PDFField not working correctly Committed by BJJ.
- [ 1215945 ] Error in COSString.writePDF() - fixed escape sequences Committed by BJJ.
- [ 1198912 ] COSName with escaped characters not parsed correctly Committed by BJJ.
- Fixed issue where resources were not being cleared in PDFStreamEngine Committed by BJJ.
- [ 1165686 ] Expected int type parse error Committed by BJJ.
- [ 1182825 ] Wrong handling of signed/unsigned byte/int in TTF parsing Committed by BJJ.
- [ 1182892 ] PDFHighlight.setHighlightColor was removed because it is not implemented by adobe Committed by BJJ.

## 6. Version 0.7.1 (04/10/2005)

### 6.1. Changes to the Code Base

- [ 1170068 ] text field is not found Committed by BJJ.
- fixed NPE issue where an image did not have any applied filters Committed by BJJ.
- Fixed issue where extra spaces were being added during text extraction for type3 fonts Committed by BJJ.
- [ 1119420 ] Extract and Update the Meta-Information as XML Committed by BJJ.
- [ 1119410 ] Extract text in/between bookmarks Committed by BJJ.
- [ 1164476 ] XFDFImport should fail with non XFDF document Committed by BJJ.
- [ 1119408 ] Support named target for Bookmark extraction. Committed by BJJ.
- Created Resources/PDFBox\_External\_Fonts.properties to create a mapping for non-embedded fonts Committed by BJJ.
- **\*\*API Change\*\*** Renamed PDFField.getName() to PDFField.getPartialName(), added method getFullyQualifiedName() Committed by BJJ.
- **\*\*API Change\*\*** Renamed PDWidget to PDAnnotationWidget for naming consistency Committed by BJJ.
- Text is now extracted from embedded form xobjects. Committed by BJJ.
- Deployed site to new hosting vendor. Committed by BJJ.
- committed code for PDFHighlighter to highlight words in a PDF document. Committed

- by BJL.
- Added command line application org.pdfbox.PDFToImage Committed by BJL.
- Implemented runlength decoding Committed by BJL.
- Added patch from Jorge Hernández Sellés to append content streams to existing page. Committed by BJL.
- **\*\*API Change\*\***renamed package from pdmodel.graphics.image to pdmodel.graphics.xobject Committed by BJL.
- **\*\*API Change\*\***Removed PDRadioButton, should use PDCheckbox instead Committed by BJL.
- **\*\*API Change\*\***COSStream now extends COSDictionary instead of containing a dictionary Committed by BJL.
- [ 1021241 ] Text extraction should follow PDF article divisions Committed by BJL.
- Added implementation for PDF page articles Committed by BJL.
- Created TextToPDF command line application Committed by BJL.
- Created ImageToPDF example Committed by BJL.
- fixed parsing of header where a trailing % exists Committed by BJL.
- [ 1110029 ] Character ">" not quoted in COSName::writePDF Committed by BJL.

## 7. Version 0.7.0 (1/22/2005)

### 7.1. Changes to the Code Base

- committed [ 1097913 ] Enhance LucenePDFDocument streams(thanks to Olivier Parent) Committed by BJL.
- Added implementation for PDF Bookmarks Committed by BJL.
- Added implementation for PDF Destinations Committed by BJL.
- Updated website for better format for documentation Committed by BJL.
- Now ExportFDF and ExportXFDF will default output files to pdfname.fdf and pdfname.xfdf Committed by BJL.
- [ 1046278 ] ClassCastException when doing FDF/XFDF Committed by BJL.
- ExtractText now allows you to extract text if you decrypt with the owner password Committed by BJL.
- Added PDF 1.5 Object Stream support Committed by BJL.
- Added pdmodel.common.PDStream to represent COSStream Committed by BJL.
- changed PDPage.getContents to use PDStream instead of COSStream Committed by BJL.
- Updated LucenePDFDocument Javadoc to tell which Lucene fields it populates Committed by BJL.
- moved HelloWorld example from persistence to pdmodel and updated to use new PD Model features Committed by BJL.

- Refactored PDFStreamEngine based on contributions from Christophe Huault Committed by BJJ.
- This class no longer uses a gigantic if/else statement for all of the operators they are defined as properties when instantiating the class Committed by BJJ.
- Updated AFM resources to be ones released on Adobe's site, include AFM license as well Committed by BJJ.
- Added ability to embed TTF fonts, only WinAnsiEncoding is supported at this time Committed by BJJ.
- Added ability to extract images, thanks to contributions by Brigitte Mathiak Committed by BJJ.
- COSWriter now generates the document id if it does not already exist Committed by BJJ.
- improved performance for text extraction Committed by BJJ.
- [ 1058693 ] TextPosition does not take account of tz operator Committed by BJJ.
- upgraded to log4j-1.2.9 Committed by BJJ.
- include package-list for javadocs Committed by BJJ.
- [ 1037145 ] Infinite loop in PDFParser.parseObject Committed by BJJ.
- fixed error where spaces before integers was causing parse errors Committed by BJJ.

## 8. Version 0.6.7 (10/09/2004)

### 8.1. Changes to the Code Base

- Revamped the way character spacing and font information is obtained Committed by BJJ.
- Improved location information about a character drawn on the screen. Committed by BJJ.
- Changed the PDFStreamEngine.showString to showCharacter to support the newly improved location information. This will now only show one character at a time. Committed by BJJ.
- Fixed bug in PDDocument.isOwnerPassword and isUserPassword that was using the wrong length for the encryption key Committed by BJJ.
- Upgraded to ant 1.6.2 Committed by BJJ.
- Upgraded to checkstyle-3.4 Committed by BJJ.
- Upgraded to JUnit-3.8.1 Committed by BJJ.
- Upgraded to lucene-1.4.2 Committed by BJJ.
- Integrated patch(1016603) for issue 943319 to fix parsing of open office documents Committed by BJJ.
- Patch:985347 No longer throw exception for "No 'ToUnicode' and no 'Encoding' for Font" Committed by BJJ.

- Patch:996191 Fixed case statement with missing break Committed by BJJ.
- Patch:996781 Fixed null pointer exception in acroform fields Committed by BJJ.
- Renamed DecryptDocument to DocumentEncryption to support encryption and decryption Committed by BJJ.
- Added load/save/encrypt/decrypt convenience methods on the PDDocument class Committed by BJJ.
- COSWriter now attempts to keep object numbers from parsed documents and writes 'free' entries in the xref if necessary Committed by BJJ.
- Added the ability to set the word separator on the PDFTextStripper Committed by BJJ.
- Fixed issue where PDFBox would throw an IOException if a PDF was incorrectly missing an endobj tag Committed by BJJ.
- Fixed 918220 where PDFBox would freeze when parsing certain cmap files Committed by BJJ.
- Added initial colorspace support Committed by BJJ.
- Fixed issue where AppendDoc was throwing ClassCastException Committed by BJJ.
- Fixed 1013163 Can't parse filters that use filter abbreviation Committed by BJJ.
- Fixed 1011244 Where encrypting then decrypting was causing a problem Committed by BJJ.
- renamed TextPosition.getWidth to TextPosition.getCombinedHorizontalDisplacement to better reflect its actual value Committed by BJJ.
- Fixed 919215 PDFBox now support stream replacement Committed by BJJ.
- Fixed 955043 Added support for 'ETenms-B5-H' encoding Committed by BJJ.
- Fixed 996050 Class Cast exception when importing Committed by BJJ.
- Added support for Font descriptors Committed by BJJ.
- Fixed spacing issues when doing textfield FDF import Committed by BJJ.
- Fixed 1017175 Large number converted when re-written Committed by BJJ.
- Fixed 1029873 PDFBox now allows for multiple xref sections Committed by BJJ.
- Added support for document Viewer Preferences Committed by BJJ.
- Made currentDocument and pdfDocument protected in util.Splitter to allow easier subclassing Committed by BJJ.
- Fixed 1034427 After Splitting page orientation is lost Committed by BJJ.
- Added the following command line applications Committed by BJJ.

## 9. Version 0.6.6 (07/20/2004)

### 9.1. Changes to the Code Base

- Improved support for setting of checkbox fields(FDF import) Committed by BJJ.
- Added the org.pdfbox.PDFSplit utility to split a single document into many documents Committed by BJJ.

- PDFBox now ignore the Length field that is associated with a stream, it has been found to be wrong in some documents Committed by BJL.
- Fixed bug when writing out PDF documents and the document contained an non alphabetic character such as ( or ) Committed by BJL.
- Fixed bug in PDFont where dictionary encodings where not being processed correctly Committed by BJL.
- Fixed bug in COSDocument.isEncrypted which was comparing COSNull to the wrong object Committed by BJL.
- Integrated patch for supporting multiple lines in the appearance stream Committed by BJL.
- Upgraded to lucene-1.4-final Committed by BJL.
- org.pdfbox.ExtractText now uses the system encoding as the default encoding instead of ISO-8859-1 Committed by BJL.

## 10. Version 0.6.5 (03/08/2004)

### 10.1. Changes to the Code Base

- Fixed bug in revision 3 encryption algorithm Committed by BJL.
- added support for CIDFontType0 glyph widths, which fixed issue with spaces being during text extraction Committed by BJL.
- Fixed infinite loop when parsing a corrupt content stream Committed by BJL.
- Add characterspacing + wordspacing when determining the width of a space character Committed by BJL.
- Added support for more font types Committed by BJL.
- refactored the pdmodel.interactive package, form fields use object delegation instead of inheritance for the widget, see PDFField.getWidget and PDFField.getKids Committed by BJL.
- Fixed bug where an inheritable cropbox would cause stackoverflow exception Committed by BJL.
- Changed usage of PDFField/PDWidget to look like object delegation instead of inheritance by adding a PDFField.getWidget instead of extending PDWidget Committed by BJL.
- refactored interactive package, this will break any existing code that uses the PDFField/PDAnnotation classes. You will need to adjust your package names!! Committed by BJL.
- Now uses StandardEncoding as the default encoding Committed by BJL.
- Bug in AppendDoc example that did not take into account groups of pages Committed by BJL.
- PDFont now also tries the bootstrap classloader when loading AFM resources Committed

- by BJL.
- added -startPage and -endPage command line options to org.pdfbox.ExtractText Committed by BJL.
- Added support for corrupt PDFs with garbage before the header Committed by BJL.
- Fixed bug where there was whitespace instead of garbage characters in front of the first object Committed by BJL.
- performance improvements for the Matrix implementation Committed by BJL.
- upgraded to lucene 1.3 Committed by BJL.
- fixed bug in cmap parser for cmap files that all ended in 'def' Committed by BJL.
- Removed createObject method from COSDocument, COSWriter will handle all object references for you Committed by BJL.
- Updated AppendDoc to use PDDocument instead of COSDocument and a couple bug fixes Committed by BJL.
- PDFParser now closes the document if there were parse errors Committed by BJL.
- TextPosition now has the PDFFont that is associated with the piece of text Committed by BJL.
- Added initial version of org.pdfbox.PDFViewer, a GUI application to view the internal structure of a PDF document. This can be used for debugging purposes at this time but may end up being a Adobe Reader like application if there is enough interest Committed by BJL.
- Changed COSNumber/COSInteger/COSFloat interface to have both intValue and longValue Committed by BJL.
- Added methods isUserPassword & isOwnerPassword to PDDocument Committed by BJL.
- Added cmap files for CJK languages, please give me some feedback Committed by BJL.

## 11. Version 0.6.4 (11/02/2003)

### 11.1. Changes to the Code Base

- Fixed bug which caused infinite loop Committed by BJL.
- Fixed bug in encoding where DictionaryEncoding kept a reference instead of making a copy leading to encoding problems Committed by BJL.
- Added PDFTextStripper.(get|set)PageSeparator, which will allow the user to output a string after every page Committed by BJL.
- refactored text stripping code to separate the logic processing of PDF operators and the logic of extracting text Committed by BJL.
- ran findbugs on source code and fixed a couple minor issues Committed by BJL.
- Refactored font functionality to PDFFont, some API methods are no longer available in COSObject Committed by BJL.

- changed name of org.pdfbox.Main to org.pdfbox.ExtractText Committed by BJL.
- added contribution of org.pdfbox.Overlay from Mario Ivankovits Committed by BJL.
- added log.isDebugEnabled checks to log4j calls Committed by BJL.
- added better escaping when writing COSNames Committed by BJL.
- fixed bug where encryption dictionary is sometimes set to COSNull instead of not being present Committed by BJL.

## **12. Version 0.6.3 (09/13/2003)**

### **12.1. Changes to the Code Base**

- Now contains the ability to import/set FDF data thanks to a contribution from Stefan Uldum Grinsted Committed by BJL.
- No longer throw an error when stream is not followed by 0A or 0D0A to allow more PDFs to be parsed Committed by BJL.
- Added -encoding argument to org.pdfbox.Main to control the encoding of the output Committed by BJL.
- Remove Prev entry from trailer if it exists because PDFBox automatically clears all old entries, only an issue when modifying/saving an existing PDF document Committed by BJL.
- Fixed bug in master password encryption algorithm for Revision 3 encrypted documents Committed by BJL.
- COSString no longer uses UTF-8 when encoding the byte array Committed by BJL.
- Added PDDocument.getPageCount() Committed by BJL.
- Fixed bug in PDFEncryption where Committed by BJL.
- Now enforces text extraction permissions Committed by BJL.

## **13. Version 0.6.2 (4/18/2003)**

### **13.1. Changes to the Code Base**

- Modified build so that build.properties settings are no longer required Committed by BJL.
- Added required libraries to CVS Committed by BJL.
- Added log4j logging Committed by BJL.
- Significant text extraction work Committed by BJL.
- Added automatic handling of files encrypted with the empty password Committed by BJL.
- Added automated tests and test data for text extraction Committed by BJL.
- Removed unimplemented decoders from filters test Committed by BJL.

- Fixed several LZW decode bugs introduced after 0.5.6 Committed by B JL.
- Fixed bugs relating to processing out of spec PDF's with bad # escaping in the name ("java.io.IOException: Error: expected hex number" bug) Committed by B JL.
- Fixed Lucene UID generation bug Committed by B JL.
- Fixed GetFontWidths null pointer exception bug Committed by B JL.

## **14. Version 0.6.1 (3/9/2003)**

### **14.1. Changes to the Code Base**

- Fixed bug in parsing stream objects which led to "Unexpected end of ZLIB input stream" Committed by B JL.
- Changed license from LGPL to BSD to allow pdfbox to be used easily in Apache projects Committed by B JL.

## **15. Version 0.6.0 (3/5/2003)**

### **15.1. Changes to the Code Base**

- Massive improvements to memory footprint Committed by B JL.
- Must call close() on the COSDocument(LucenePDFDocument does this for you) Committed by B JL.
- Really fixed the bug where small documents were not being indexed Committed by B JL.
- Fixed bug where no whitespace existed between obj and start of object. Exception in thread "main" java.io.IOException: expected='obj' actual='obj<</Pro Committed by B JL.
- Fixed issue with spacing where textLineMatrix was not being copied properly Committed by B JL.
- Fixed 'bug' where parsing would fail with some pdfs with double endobj definitions Committed by B JL.
- Added PDF document summary fields to the lucene document Committed by B JL.

## **16. Version 0.5.6 (11/28/2002)**

### **16.1. Changes to the Code Base**

- Fixed bug in LucenePDFDocument where stream was not being closed and small documents were not being indexed Committed by B JL.
- Fixed a spacing issue for some PDF documents Committed by B JL.
- Fixed error while parsing the version number Committed by B JL.
- Fixed NullPointerException in persistence example Committed by B JL.

- Create example lucene IndexFiles class which models the demo from lucene Committed by BJL.
- Fixed bug where garbage at the end of file caused an infinite loop Committed by BJL.
- Fixed bug in parsing boolean values with stuff at the end like "true>>" Committed by BJL.

## **17. Version 0.5.5 (10/03/2002)**

### **17.1. Changes to the Code Base**

- Added example of printing document signature Committed by BJL.
- Added example to print out form fields values Committed by BJL.
- Fixed bug when appending documents Committed by BJL.
- Various other bug fixes Committed by BJL.

## **18. Version 0.5.4 (09/17/2002)**

### **18.1. Changes to the Code Base**

- Fixed bug in text output where '?' instead of the proper character Committed by BJL.
- Fixed bug where sections of text were not being output at all Committed by BJL.

## **19. Version 0.5.3 (09/13/2002)**

### **19.1. Changes to the Code Base**

- Fixed bug in 128 bit encryption Committed by BJL.

## **20. Version 0.5.2 (09/06/2002)**

### **20.1. Changes to the Code Base**

- Fixed bug where FDF documents could not be appended to PDF Documents Committed by BJL.
- Catch all NumberFormatExceptions and wrap them with IOExceptions Committed by BJL.

## **21. Version 0.5.1 (09/04/2002)**

### **21.1. Changes to the Code Base**

- Now supports unicode for the document summary Committed by BJL.
- Better support for Type0 fonts Committed by BJL.
- Fixed bug with an empty LZW stream Committed by BJL.
- Fixed parsing error for ID operator Committed by BJL.

### **22. Version 0.5.0 (08/31/2002)**

#### **22.1. Changes to the Code Base**

- Now supports unicode for the document summary Committed by BJL.
- Better support for Type0 fonts Committed by BJL.
- Fixed bug with an empty LZW stream Committed by BJL.
- Fixed parsing error for ID operator Committed by BJL.

### **23. Version 0.4.1 (07/25/2002)**

#### **23.1. Changes to the Code Base**

- Fixed bug where .notdef was being output as document text Committed by BJL.

### **24. Version 0.4.0 (07/23/2002)**

#### **24.1. Changes to the Code Base**

- Added extract text ant task Committed by BJL.
- Implemented AFM(Adobe Font Metrics) resource loading Committed by BJL.
- Fixed numerous bugs submitted by users Committed by BJL.
- Changed project from pdfparser to pdfbox to better reflect future needs Committed by BJL.

### **25. Version 0.3.0 (07/09/2002)**

#### **25.1. Changes to the Code Base**

- Added indexer for the lucene project Committed by BJL.
- Initial implementation of PDF encryption(not working yet) Committed by BJL.

## **26. Version 0.2.0 (06/03/2002)**

### **26.1. Changes to the Code Base**

- Added support for the various encodings Committed by B JL.
- Improved the accuracy of the text output Committed by B JL.

## **27. Version 0.1.0 (05/25/2002)**

### **27.1. Changes to the Code Base**

- Initial Version Committed by B JL.