



Release Notes

3-Heights™ PDF Tools

Version 1.80

Contact: pdfsupport@pdf-tools.com

Owner: **PDF Tools AG**
Geerenstrasse 33
CH-8185 Winkel
Switzerland
www.pdf-tools.com

April 15, 2009

Table of Contents

1 Overview	3
2 Technical Support	3
3 How to Download	3
4 System Requirements	3
5 New Products	4
5.1 OCR Add-On.....	4
5.2 Make a PDF searchable	4
6 New Features to all Products	4
6.1 Enhancements to all 3-Heights™ Products	4
6.2 Enhancements to the 3-Heights™ PDF Rendering Engine	5
7 New Features to Specific Products	6
7.1 3-Heights™ PDF to PDF/A Converter	6
7.2 3-Heights™ PDF Optimization Tool	7
7.3 3-Heights™ PDF Extract Tool	8
7.4 3-Heights™ PDF Printer, Printer Pro and Printer Service.....	9
7.5 3-Heights™ PDF Producer and TIFF Producer	10
7.6 3-Heights™ PDF Security Tool.....	12
7.7 3-Heights™ Image to PDF Converter.....	13
7.8 3-Heights™ PDF to Image Converter.....	14
7.9 3-Heights™ PDF Validator	14
7.10 3-Heights™ PDF Repair Tool	14
7.11 3-Heights™ PDF Viewer and PDF Viewer PRO	15
8 About PDF Tools AG	16

April 15, 2009

1 Overview

The 3-Heights™ PDF tools represent the newest product line from PDF Tools AG. The 3-Heights™ PDF tools are available as programming libraries (APIs), command line tools and Windows Services. The tools allow for a wide variety of manipulation of PDF files including viewing, printing, extraction of information, conversion, validation, repairing, and optimization.

2 Technical Support

Please report problems by contacting our support department by mail:

pdfsupport@pdf-tools.com

3 How to Download

The 3-Heights™ PDF tools can be downloaded from the product description pages on our website. There is no charge for downloading evaluation versions (valid for a 30 day period).

<http://www.pdf-tools.com/asp/products.asp>

4 System Requirements

All of the 3-Heights™ PDF tools are available for:

- Windows Vista
- Windows XP
- Windows 2008, Windows 2003
- Windows 2000

Some of the tools are also available for:

- IBM AIX
- Sun Solaris (2.7 and later)
- HP-UX 11.0
- Mac OS X
- Linux
- Free BSD

Other platforms are available on request. Please refer to the individual product pages to obtain information on supported operating system platforms.

April 15, 2009

5 New Products

5.1 *3-Heights™ OCR Enterprise Add-On*

The new 3-Heights™ OCR Enterprise Add-On is a customized runtime license of the well known ABBYY FineReader 8.1 OCR engine. It is designed to cooperate with other 3-Heights™ products, such as the 3-Heights™ PDF to PDF/A Converter, the 3-Heights™ Image to PDF Converter and the 3-Heights™ Document Converter Service. It can be purchased separately, in addition to the mentioned base products.

The OCR add-on and the corresponding base product is connected through an extensible interface technology. The advantage of such a technology is that new engines can seamlessly be integrated in existing production environments. Furthermore, integrators and OEM partners can develop their own connectors based on the open interface definition.

Currently there is an interface DLL available which operates with the OCR add-on. It is part of the distribution kit of the Image to PDF Converter and the PDF to PDF/A Converter products.

5.2 *Make a PDF searchable*

A **new shell tool 'pdfocr'** has been added 3-Heights™ Image to PDF Converter Shell which is capable to locate all raster images in a PDF document, run them through an OCR engine and embed the recognized text as a hidden layer on top of the images to make the PDF document fully text searchable. The tool does not modify other objects such as vector graphics and font based text (i.e. it does not rasterize the page).

The existing tools **3-Heights™ PDF to PDF/A Converter** and **3-Heights™ Image to PDF Converter** have been enhanced to be able to interact with an OCR engine during the conversion of the documents. The recognition process is implemented the same way as it is described above.

6 New Features to all Products

6.1 *Enhancements to all 3-Heights™ Products*

- **New XMP library:** The PDF Kernel has been enhanced by a new implementation of the XMP library. This was necessary to improve the validation of XMP meta data in the PDF Validator, to synchronize the document information dictionary with the XMP meta data and to merge XMP meta data from different sources. All products which create and process PDF files such as the PDF Producer, the PDF to PDF/A Converter, the PDF Optimizer, etc. profit from this improvement.
- **Support for localized file names on UNIX:** The file names on UNIX are encoded in a character set that depends on the locale setting in the particular

April 15, 2009

installation. The PDF Kernel now fully supports these character sets. Earlier versions supported file names using the ANSI character set only.

- Document collections: The open function of the PDF Kernel now can open a document that is contained in a collection if the open action instructs the reading application to do so.
- Improved color management: The Color Management Systems (CMS) of the PDF Kernel have been adapted to support the black point compensation (BPC) feature. This produces better visual results mainly when printing black text.
- Improved performance of decompression filters: The “Flate” and “CCITT” decode filters have been optimized to minimize buffer re-allocations and generally increase the decompression performance.
- Improved performance of function execution: The execution performance of Type4 functions (PostScript Calculator), which are mainly used in the transformation of “DeviceN” to device colors has significantly been improved, resulting in faster decoding of large images with “DeviceN” and “Separation” colors.
- Improved processing of non-conforming files: The PDF kernel now accepts a wider range of files which do conform to the specification. Examples are:
 - A file can be opened even if the cross reference table is corrupt but can be recovered, if the end-of-file marker is followed by garbage (max. 4k bytes).
 - the implementation limits are exceeded (e.g. integer > 32767),
 - the integrity check of an embedded font fails
 - and many more.
- JBIG2: The JBIG2 compression and decompression library has been updated to version 1.08 (November, 2008).

6.2 Enhancements to the 3-Heights™ PDF Rendering Engine

- New Option ‘eOptionDrawPopups’: This option flag controls the way annotations are rendered. If it is set and an annotation dictionary contains a “Popup” entry which is in the “Open” state then the associated popup annotation will be rendered. This option is set by default for the PDF Viewer and clear for the PDF Printer and PDF to Image Converter.
- Improved font replacement: If a font is not embedded and a font with the same name is not installed in the operating system the rendering engine has to select an appropriate replacement font. This replacement algorithm has

April 15, 2009

been enhanced by using a “tailoring” algorithm. The tailoring chooses the nearest font and adapts the glyph widths of the actual font to the widths of the one referenced in the PDF document. Furthermore if the non-embedded font is a CID font without any Unicode information then “Bold” and “Italic” style fonts will be replaced by their “Regular” counter parts.

- Improved accuracy of coordinate transformations: If anti-aliasing in the accurate mode is enabled then the coordinate transform of objects are now computed with sub-pixel coordinates. In earlier versions they were rounded to integer coordinates.
- Detection of rendering problems: If there are any rendering problems during the drawing of a page an error flag is stored and the rendering process is completed and the stored flag is returned after the last object has been rendered. In earlier versions rendering problems were ignored.
- Improved text output performance: The rendering of text objects has been improved by merging text fragments into larger text objects. This not only increases the rendering performance but also reduces the size of the spool files if the text in a PDF document is heavily fragmented.
- Improved bitmap filtering: The filtering of bitmaps in the border zone has been improved producing better visual results if images are divided into slices.
- Conditional back buffer allocation: The allocation of a back buffer to support raster operations that access the pixels of the device surface is now only allocated if it is required by the specific content of a page. This significantly saves memory space and increases the rendering performance.

7 New Features to Specific Products

7.1 *3-Heights™ PDF to PDF/A Converter*

- New switch ‘ocr’: If a PDF document has to be made fully text searchable even if the text is part of a raster image then the images which are contained in the PDF document must be run through an OCR engine. With this switch the user can select an OCR engine, e.g. “Abbyy”, and instruct the tool to embed the recognized text as a hidden layer on top of the image. If the OCR interface DLL is not found or the engine cannot be initialized (because it is not installed or the license key is not valid) then an error message is issued.
- New switch ‘ocl’: The OCR engine can be given hints about what languages should be recognized. This switch receives a list of language names such as “English” and “German”. A list of languages can be found in the OCR add-on manual.

April 15, 2009

- New switch 'ocp': In order to optimize the performance of the OCR process a number of parameters (key / value pairs) can be passed to the engine, e.g. "DetectBold=false". A list of parameters can be found in the OCR add-on manual.
- New switch 'le': The available OCR engines can be listed.
- New switch 'ofn': With this switch the fonts in a document are replaced by the full (non subsetted) font program regardless of whether they are already embedded or not. If this switch is absent then correctly embedded fonts will not be replaced and newly embedded fonts will be subsetted.
- New switch 'ow': This switch instructs the converter to optimize the output file for fast web-viewing (linearization).
- New return codes: The shell tool now returns error code 6 when post-analysis error occur and error code 7 if there are problems during the linearization of the output file.
- New method 'SetOCREngine': see switch 'ocr'.
- New method 'SetOCRLanguages': see switch 'ocl'.
- New method 'SetOCRParams': see switch 'ocp'.
- New property 'SubsetFonts': see switch 'ofn'.
- New method 'GetOCREngine': see switch 'le'
- New method 'GetOCREngineCount': see switch 'le'
- Improved conversion and repair function: The converter has been significantly improved to convert non-conforming files. Examples are the conversion of simple fonts to CID fonts, the re-encoding of character strings. Furthermore, some commonly known problems are fixed on-the-fly such as invalid page objects in destinations and the like.

7.2 3-Heights™ PDF Optimization Tool

- Indexed to gray conversion: If the lookup table of an indexed image indicates a gray scale image then its color space is converted to a real gray scale image.
- Down-sampling and image interpolation: If an image is down-sampled the 'Interpolate' entry is not set to 'true' any more which was the case in earlier versions.

April 15, 2009

7.3 3-Heights™ PDF Extract Tool

- Improved text extraction: The heuristics to identify the Unicode of a character from the font's encoding has been extended yielding better text extraction results.
- Double character removal. Double characters (characters with the same Unicode and at the same position) are removed now in order to increase the readability of the extracted text.
- Transparent images: Images with a soft mask can now be exported as TIFF containing an alpha channel (extra samples).
- Improved extraction of JPEG streams: If an image in a PDF file is DCT encoded and the requested output format is a .jpg file then the JPEG stream is saved without recompressing it.
- New switch 'laf': This switch lists the form fields in a document. Since form fields are also annotations they may also be listed using the 'la' switch. The difference however is, that form fields may be hierarchically nested (parents / children) and that the listing contains fields that are more related to form fields than annotations. Furthermore, annotations that are not form fields, e.g. link annotations, are not listed with this switch.
- Changed switch 'pg': The switch is used to specify the range of pages to be listed. If it is missing the range defaults to all pages of the document. In previous versions the presence or absence of the switch controlled the way how resources were listed. In the actual version this side effect has been removed and the desired behavior is explicitly controlled by the new switch 'r'.
- New switch 'r': This switch modifies the way how resources (fonts, images, etc.) are listed. If it is specified then resources are listed that are marked as such regardless whether they are used on a page or not. If this switch is not specified only those resources are listed that are actually used including inline images.
- Changed switch 'li': This switch is used to list images. Now the listing replaces the fields 'IsMask' and 'HasMask' with the new field 'Mask' which can have the values 'Null', 'Stencil', 'Explicit' and 'Soft'. The field 'ColorSpace' is now set to 'Null' for stencil mask images.
- Changed switch 'la': This switch is used to list annotations. It now lists the destination of Goto actions and the URI of link annotations too.
- Changed switch 'ld': This switch is used to list the properties of a document. If it used in conjunction with 'x' then the XMP metadata of the document is now extracted.

April 15, 2009

- New property 'GetSMask': With this property the soft mask of an image can be extracted now.
- Changed property 'StrokeColor' and 'FillColor': If a color doesn't exist (e.g. with an uncolored pattern) then -1 is now returned.

7.4 3-Heights™ PDF Printer, Printer Pro and Printer Service

- New switch 'pri': With this switch the priority of the print job can be set. The priority controls the order in which the print spooler sends the jobs to the print device. It does not have any effect on the operation of the tool itself.
- New switch 'rm': With this switch the page rotation can be set to 'Portrait' or 'Landscape'. In contrast to the 'o' switch, the page is rotated before being sent to the printer whereas the 'o' switch specifies the orientation of the physical paper.
- Changed switch 'cm' : The copy mode is used in conjunction with the 'c' and 'cl' switches has been revised to function as follows. If it is set to 0 then the printer tool delegates the copy operation to the printer driver. Otherwise, if it is set to 1, then the tool performs the copy operation based on the number of copies and the collate mode. This switch is useful if the driver does not implement all collate and copy modes.
- Changed switch 'od': The effect of this switch has been slightly changed. To simulate transparent (translucent) graphics a back buffer is used. In earlier version this buffer has always been allocated. With the actual version it is only allocated if the page contains transparent objects. The switch is used to disable the back buffer to speed up the rendering operation and save memory.
- Changed switch 'g': The grouping of files into print jobs is now performed by the grouping function of the spooler. In earlier versions it was performed by the tool itself. The function does not work with the 'file' switch.
- New switch '-ou': The glyphs of embedded fonts are selected using a font internal number. Some application scenarios such as spool file post-processors require that the glyphs are selected using the character's Unicode. This switch was introduced to control the rendering engine to use Unicodes instead of glyph numbers for embedded fonts. However, this switch has to be used with caution and exclusively for the said applications.
- New switch 'pjl': This switch allows to specify a file which contains a number of PjL commands that are merged with the spool file before it is sent to the printer. It is useful in situations where commands have to be sent to the printer which are not supported through the device mode. This switch does not work if the 'file' switch is used. (Undocumented Feature)

April 15, 2009

- New method 'BeginGroup': This method is used to mark the beginning of a job group. See switch 'g'.
- New method 'EndGroup': This method is used to mark the end of a job group. See switch 'g'.
- New method 'SetCMSEngine': This method allows to specify the name of the Color Management System (CMS) engine used to convert colors from one color space to another. Valid names are 'none', 'neugebauer', 'msicm' which is the default on Windows and 'lcms' which is the default on Unix systems. The name can also be a file name containing a set of customized Neugebauer parameters.
- New property 'ReportingLevel': With this property an error level can be set and retrieved. The error level is a threshold to control the type of errors that a reported. With level 0 the reporting is disabled. With level 1 only errors are reported, with level 2 errors and warnings are reported and with level 3 errors, warnings and informational messages are reported.
- Changed property 'Handle': The handle which is defined by the 'OpenPrinter' call is now opened with the PRINTER_ALL_ACCESS access rights.

7.5 3-Heights™ PDF Producer and TIFF Producer

- New Office Add-In: An add-in for the Microsoft Office Word, Excel and PowerPoint applications is provided with the 3-Heights™ PDF Desktop Producer. After the installation a "Save as PDF/A" button appears in the tool bar of the MS Office application with which a PDF/A conforming document can directly be created. The add-in also collects document information such as the author, the creation date, the application name etc. and embeds them as an XMP meta data package in the saved document.
- New printing option "Rotate Landscape Pages": None, Clockwise, Counter-clockwise. If this option is set then landscape pages are automatically rotated in the specified direction.
- New printing option "Fast web view (linearize)": Yes / No. If this option is set then the output file is optimized for fast web viewing.
- New port configuration 'Remove file name prefixes': With this box checked in the file name some predefined prefixes such as "Microsoft Word –" etc. are removed from the resulting file name. This configuration parameter is set by default.
- Changed configuration parameter 'Execute Command': In previous versions the command was validated when the configuration was updated. Now the command is validated before it is executed. This allows for configuring a command which is later installed.

April 15, 2009

- Changed command execution: In previous version the command was executed in the user context of the spooler process. Now the command is executed in the context of the user that submitted the print job.
- Support for per-user output folders: If the port name specifies a relative path, e.g. ".\PDF Outbox" then the output folder is stored in "My Documents" folder of the user that submitted the print job. However only users that have such a folder on the target system may submit print jobs otherwise the operation will fail.
- XMP metadata interface: An add-in, such as the Microsoft Office Add-In, now can pass an XMP metadata stream to the PDF Producer which is merged and embedded in the output file.
- Support for transparent bitmaps: The PDF Producer now supports commonly used raster operation sequences to render transparent bitmaps, even if the back buffer is disabled. This enhancement helps to avoid using the back buffer in environments where speed and memory usage is critical.
- Support for all bitmap formats: All bitmap formats, including some seldom used formats, such as run-length encoded (RLE) and bit-field bitmaps, are now supported.
- Support for 'Delete Job': If a job is deleted or aborted either through a user action or programmatically through 'AbortDoc' then the resulting file is deleted now from the output folder.
- Support for global and per-user device mode: The PDF and TIFF Producer have been enhanced to save the printer settings in the global and per-user device mode portion of the registry. These are used by some applications as an alternative to directly retrieving the device mode from the printer driver.
- Enhanced TIFF Producer settings: The TIFF producer now offers more predefined classes including 'Fax N', 'Bitonal', 'Grayscale', 'RGB', 'CMYK'. A class is a set of parameters such as color space, compression algorithm, resolution, etc. which are grouped into a meaningful set. The default class is 'Bitonal'.
- Enhanced TIFF compression: The TIFF producer now automatically uses a predictor in conjunction with the LZW compression. This increases the compression rate for grayscale, RGB and CMYK images significantly.
- Automatic resolution restriction: The TIFF Producer (and the PDF Producer, if the back buffer is enabled) automatically reduces the resolution until there is enough memory to allocate the back buffer.
- New method 'SetEnvironment': The installer now provides a method to specify the printer driver environment, e.g. "W32x68" or "x64". This allows for

April 15, 2009

installing the PDF and TIFF Producer on printer servers which provide access to clients with different platform environments.

- New method 'GetDefaultPort': The method returns the recommended default port name to the installer program.
- Improved error logging: The installer writes a detailed log if an error occurs during the installation process.

7.6 3-Heights™ PDF Security Tool

- Changed method 'SaveInMemory': The method now allows for specifying encryption parameters in order to create an encrypted PDF file in memory.
- New method 'AddInfo': The method adds a key / value pair to the documents info object. Values of predefined key are also stored in the XMP metadata package.
- New method 'AddDocMDPSignature': With this method a modification detection and prevention (MDP) signature can be added to the document. In addition a set of access rights can be specified. A document can only have one such signature.
- New property 'RevisionCount': This property returns the number of revisions of the document (the number of incremental updates). Although a linearized file looks like an incrementally updated file it only counts as one revision.
- New method 'Revision': This method allows to retrieve the PDF file for a given revision. All revisions which were applied after the given revision are ignored.
- New method 'RevisionLength': This method returns the length of the PDF file for a given revision. It is used to allocate a memory buffer prior to calling the Revision() method.
- New property 'Revision': The signature interface now provides a new property returning the revision number associated with the particular signature. It is useful to retrieve the state of the PDF file at the time it has been signed.
- New property 'Filter': This property returns the name of the filter with which the signature has been created.
- New property 'SubFilter': This property returns a name which is used to check the interoperability among signature filters.
- New method 'PdfReleaseSignature': The signature plug-in interface now provides this new method to release resources in the signature object which were previously allocated by the plug-in.

April 15, 2009

7.7 **3-Heights™ Image to PDF Converter**

- New switch 'ocr': If a PDF document has to be made fully text searchable even if the text is part of a raster image then the images which are contained in the PDF document must be run through an OCR engine. With this switch the user can select an OCR add-on, e.g. "Abbyy", and instruct the tool to embed the recognized text as a hidden layer on top of the image. If the add-on is not found or the engine cannot be initialized (because it is not installed or the license key is not valid) then an error message is issued.
- New switch 'ocl': The OCR engine can be given hints about what languages should be recognized. This switch receives a list of language names such as "English" and "German". A list of languages can be found in the OCR add-on manual.
- New switch 'ocp': In order to optimize the performance of the OCR process a number of parameters (key / value pairs) can be passed to the engine, e.g. "DetectBarcodes=false". A list of parameters can be found in the OCR add-on manual.
- New switch 'bc': It is used to specify a crop box coordinates (left, bottom, width, height). The crop box is used to clip the imported image. If the crop box is not specified it defaults to the media box which is computed from the images dimensions.
- New property 'ImageQuality': With this property the image quality for the JPEG, JPEG2000 and JBIG2 compression filters can be set. A value of 100 denotes the best possible quality and lossless compression for JBIG2. The default quality is 80.
- New method 'SetOCREngine': see switch 'ocr'.
- New method 'SetOCRLanguages': see switch 'ocl'.
- New method 'SetOCRParams': see switch 'ocp'.
- Enhanced image import codecs: Various enhancements have been made to the image codecs such as the possibility to inhibit the re-compression of JPEG2000 images, support for the JPEG2000 YCC color space, the skipping of garbage bytes in front of the SOI marker in JPEG streams, the support for various sub-sampling formats in TIFF V6 JPEG streams and the substitution of a default resolution and default orientation if it is missing in the image stream.
- The tool now sets the 'Creator' entry in the documents info object to an appropriate value.

April 15, 2009

7.8 3-Heights™ PDF to Image Converter

- Enhanced switch 'h': The switch now accepts four values (0: no dithering, 1: Floyd Steinberg, 2: Halftone block, 3: Halftone continuous) instead of only the first two.
- Changed / new switch 'oq': The switch now disables anti-aliasing only instead of disabling high-quality rendering. The switch is now also available for the Converter Service.
- Changed rendering behavior: If an error is detected while rendering a page the conversion process is stopped and an error code is returned (Shell, API ConvertFile).
- New switch 'cs' (pdf2pdfimg): The pdf2pdfimg tool now accepts this switch in the same way as it is implemented for the pdf2img tool. The switch allows for specifying the output color space of the document including Gray, RGB and CMYK.
- New property 'ColorSpace' (Pdf2PdfImgOCX): See switch 'cs'.
- Changed behavior of switch 'cs' and property 'ColorSpace' (pdf2img): The value of the color space now is implicitly changed if one of the 'BitsPerPixel' or 'Compression' properties is changed.

7.9 3-Heights™ PDF Validator

- Isartor compliance: The PDF/A validation has been enhanced such that the PDF Validator now passes the Isartor test suite which was published by the PDF/A Competence Center (see www.pdfa.org).
- New switch 'pw': This switch allows for specifying a user password to open the file to be validated.
- New field object number in error report: With each reported error the tool provides now the object number as an additional error context information.

7.10 3-Heights™ PDF Repair Tool

- New property 'ReportingLevel': With this property an error level can be set and retrieved. The error level is a threshold to control the type of errors that are reported. With level 0 the reporting is disabled. With level 1 only errors are reported, with level 2 errors and warnings are reported and with level 3 errors, warnings and informational messages are reported.
- New property 'ErrorCode': This property returns the error code of the last operation. The list of error codes can be seen in the file 'pdferror.h'.

April 15, 2009

7.11 3-Heights™ PDF Viewer and PDF Viewer PRO

- **Popup annotations:** Annotations which have an associated popup in the open state are now rendered by default.
- **Case insensitive search:** The search function has been extended to handle case insensitivity for languages other than Latin (e.g. Russian).
- **Printing in Viewer PRO:** The PRO variant of the PDF Viewer now has an own rendering object to print the currently viewed document. This change reduces unwanted interferences between viewing and printing. And, the default rendering mode for the printing operation is now different from the viewing operation. It is now set to 'Fast' by default.

April 15, 2009

8 About PDF Tools AG

PDF Tools AG (www.pdf-tools.com) is a world leader in PDF (Portable Document Format) software, delivering reliable PDF products to international customers in all market segments.

PDF Tools AG provides server-based software products designed specifically for developers, integrators, consultants, customizing specialists and IT-departments. Thousands of companies worldwide use our products directly and hundreds of thousands of users benefit from the technology indirectly via a global network of OEM partners. The tools can be easily embedded into application programs and are available for a multitude of operating system platforms.