



3-Heights™ PDF Analysis & Repair Shell

User's Manual

Version 1.91

Contact: pdfsupport@pdf-tools.com

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

June 25, 2010

Table of Contents

1	Introduction.....	3
1.1	Description	3
1.2	Functions	3
	Features.....	4
	Formats.....	4
	Compliance.....	4
1.3	Operating Systems	4
2	Installation	6
2.1	Installing the 3-Heights™ PDF Analysis & Repair Shell.....	6
	How to set the Environment Variable "Path"	6
3	Getting Started and User's Guide	7
3.1	Usage	7
3.2	Repair a File	7
3.3	Specify the Folder of the Output File	7
3.4	Repairing all Files in a Directory	7
3.5	Difference between Repairable and Recoverable.....	8
4	Reference Manual	9
4.1	Switches	9
	-a Analyze only	9
	-b Set the Rebuild Options.....	9
	-dp Do Not Recover Pages.....	9
	-dx Do Not Repair Cross-Reference Table.....	9
	-l Set the Log File	9
	-p Specify Password.....	10
	-ri Recover Embedded Images.....	10
4.2	Return Codes.....	10
5	Troubleshooting.....	11
5.1	The Repair Time Takes Too Long	11
5.2	The File Cannot be Repaired	11

1 Introduction

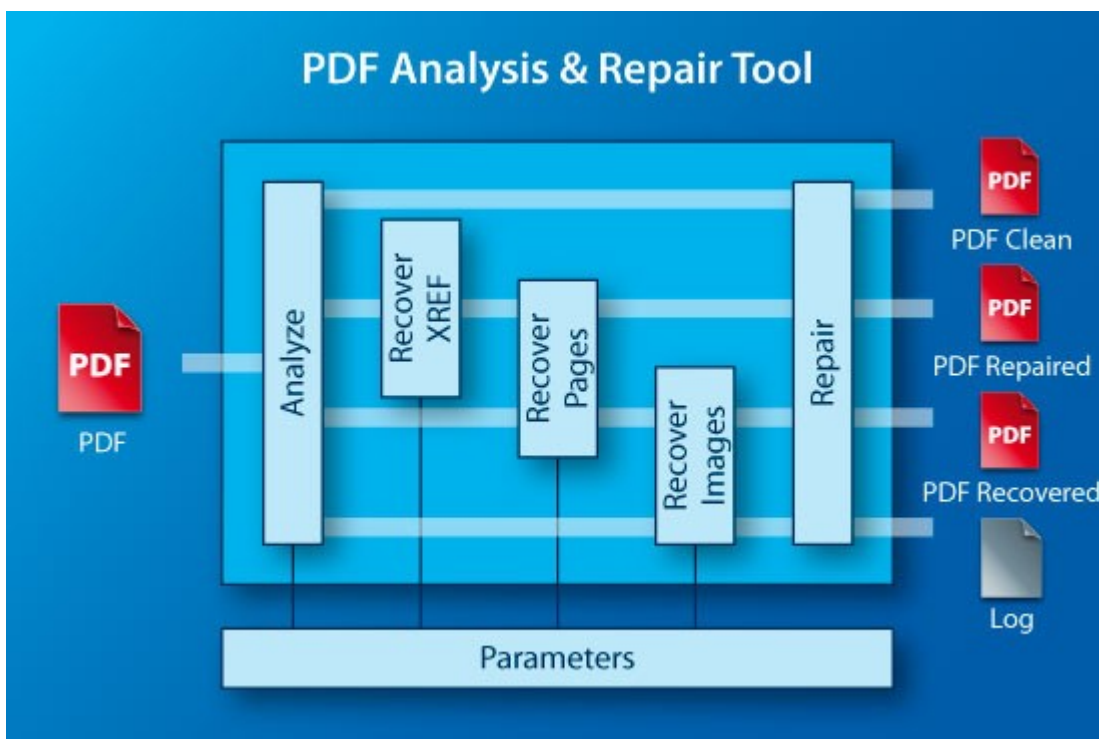
1.1 Description

The 3-Heights™ PDF Analysis & Repair tool is used to analyze, repair and restore the content of corrupt PDF documents.

Unfortunately, the number of corrupt PDF documents is incredibly huge. The cause is usually down to defective generating tools, converters and other influences such as attempts at manual editing, copying via FTP without correct settings, system crashes during PDF creation, network interruptions, defective copying on optical media, etc.

The result leads to an enormous loss of important information and to production downtimes caused by corrupt PDF documents.

The 3-Heights™ PDF Analysis & Repair analyzes PDF documents with regard to PDF specifications. Defective files are automatically repaired as far as possible and unreadable data is restored.



1.2 Functions

PDF Analysis & Repair is used to check and, where indicated, repair PDF documents. Users can determine customized profiles from a broad range of analysis and repair options. An exact and detailed description is issued for each reported error. The tool is also capable of reading and processing encrypted PDF files without any problems.

June 25, 2010

Features

Analyze and/or repair one or more PDF documents

Set analysis options, including:

- Objects
- Page tree
- Content stream

Set repair options, including:

- Restore data streams
- Restore fonts
- Restore XRef table
- Restore pages
- Restore images if pages cannot be restored

Display error description for every message, including:

- Type (errors, warnings, information)
- Error code
- Text-based description
- Page number
- Number of events

Write error messages to log file

Read encrypted PDF files

Encrypt restored file and set user authorizations

Differentiate between Repair (corrects the errors in the document) and Restore (recreates the document based on the remaining legible information)

Formats

Input Formats

- PDF 1.x (e.g. PDF 1.4, PDF 1.5, etc.), PDF/A

Target Formats

- PDF 1.x (e.g. PDF 1.4, PDF 1.5, etc.)

Compliance

Standards: ISO 32000 (PDF 1.7)

1.3 Operating Systems

- Windows 2000, XP, 2003, Vista, 2008, Windows 7 – 32 and 64 bit

June 25, 2010

- FreeBSD 4.7 for Intel
- HP-UX 11.0 – 32 bit
- IBM AIX (4.3: 32 Bit, 5.1: 64 bit)
- Linux (SuSE and Red Hat on Intel)
- Mac OS X
- Sun Solaris (2.7 and higher)

2 Installation

2.1 Installing the 3-Heights™ PDF Analysis & Repair Shell

1. Download the ZIP archive of the product from your download account at www.pdf-tools.com.
2. Open the ZIP archive.
3. Check the appropriate option to preserve file paths (folder names) and unzip the archive to a local folder (e.g. *C:\program files\pdf-tools*).
4. The unzip process now creates the following subdirectories:
 - *Bin*: Contains the runtime executable binary code
 - *Doc*: Contains documentation files
5. To start the 3-Heights™ PDF Analysis & Repair Tool from a shell, the directory needs to be included in the "Path" environment variable.

How to set the Environment Variable "Path"

To set the environment variable "Path" on Windows 2000: Go to **Start -> Settings -> Control Panel -> System -> Advanced -> Environment Variables**

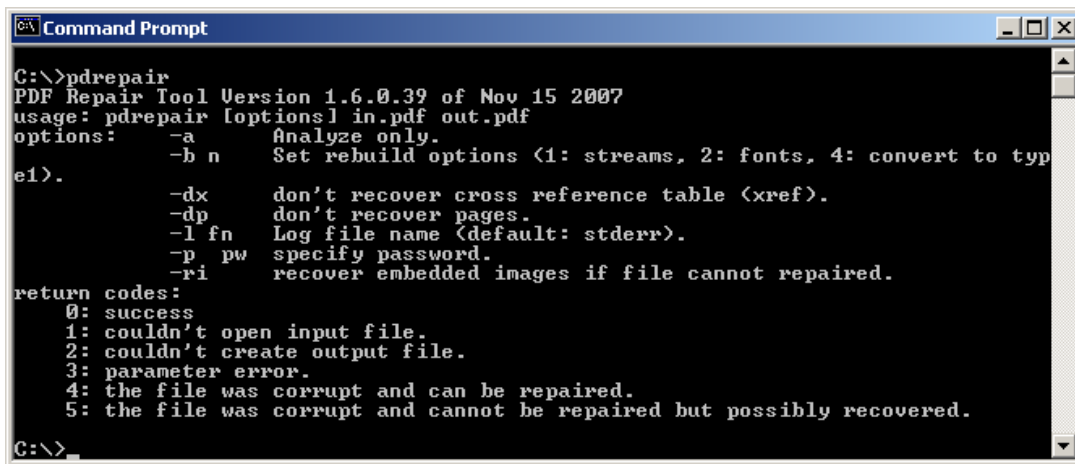
Windows XP: Go to **Start -> Control Panel (classic view) -> System -> Advanced -> Environment Variables**.

Select "Path" and **Edit**, then add the directory where *pdfrepair.exe* is located to the "Path". If the environment variable "Path" does not exist, create it.

3 Getting Started and User's Guide

3.1 Usage

Type `pdrepair` in the shell, and the usage is returned.



```
Command Prompt
C:\>pdrepair
PDF Repair Tool Version 1.6.0.39 of Nov 15 2007
usage: pdrepair [options] in.pdf out.pdf
options:  -a      Analyze only.
          -b n    Set rebuild options (1: streams, 2: fonts, 4: convert to typ
e1).
          -dx    don't recover cross reference table (xref).
          -dp    don't recover pages.
          -l fn  Log file name (default: stderr).
          -p pw  specify password.
          -ri    recover embedded images if file cannot be repaired.
return codes:
0: success
1: couldn't open input file.
2: couldn't create output file.
3: parameter error.
4: the file was corrupt and can be repaired.
5: the file was corrupt and cannot be repaired but possibly recovered.
C:\>
```

3.2 Repair a File

The 3-Heights™ PDF Analysis & Repair Tool requires at least two parameters: A name of an existing PDF input file and the desired name for the repaired PDF output file

Example: Read the damaged input file "myfile.pdf". Repair the file and save the result in a new file called "myfile_rep.pdf".

```
pdrepair myfile.pdf myfile_rep.pdf
```

3.3 Specify the Folder of the Output File

The output folder can simply be added in front of the output file name

```
pdrepair myfile.pdf myfolder\myfile_rep.pdf
```

or absolute (Windows):

```
pdrepair myfile.pdf C:\myfolder\myfile_rep.pdf
```

3.4 Repairing all Files in a Directory

The 3-Heights™ PDF Repair Shell Tool reads the input file while it already writes on the output file. For this reason, it is not possible to directly overwrite the input file.

If you would like to repair all PDF documents in a directory, it is required to use a variable to name the output files. Here is an example using the FOR command of the CMD shell:

June 25, 2010

Example: Process and repair all files in the current directory.

```
for %i in (*.pdf) do pdrepair %i %~ni_rep.pdf
```

Of course, one can adjust the paths, or use a different output name.

If you would like the repaired file to have the same name as the original file, it is suggested to use the following process:

1. Repair the file and create a repaired copy of the original file (like in the for loop above).
2. Ensure the repaired documents are generated correctly. For example check the return code of the repair tool and require it to be 0 or 4, or ensure the files are not empty (i.e. just a few bytes in size).
3. When you are sure the repaired file is okay and you do not need the original file anymore, delete the original file and rename the recovered file.

3.5 Difference between Repairable and Recoverable

Repairable: The input PDF file contains errors which are repairable. These are for example PDF syntax errors. In the repaired PDF output file, these errors are fixed.

Recoverable: The input PDF file is missing data. For example data of an embedded image, which got lost/overwritten when copying the file from one location to another failed partially. When relevant data is lost, the file is irreparable, however the file can be recovered, such that the output PDF is valid according to the PDF Specification.

Example: Assuming an original PDF (file O) is valid.

- Somehow the file gets corrupted (-> file C) and file O gets lost.
- If file O can be rebuilt based on file C, then file C is repairable.
- If a new file N can be built based on file C, and file N is a valid PDF and contains (part of) the content of file O, then file C is recoverable.

4 Reference Manual

4.1 Switches

-a Analyze only

By default, documents will be analyzed and invalid PDF files will be repaired or recovered. With the switch **-a**, only the analysis process will be executed. The results of the analysis will be displayed on screen.

Example: Analyze the file "myfile.pdf".

```
pdrepair -a myfile.pdf
```

-b Set the Rebuild Options

This options controls what parts of the PDF are to be repaired. Available options are:

- 1: Rebuild streams
- 2: Rebuild fonts
- 4: Convert CFF fonts to Type1 fonts.

If 4 is applied, the compressed fonts are decompressed, this can potentially lead to an increase of the file size.

If multiple options are to be selected, add the values.

Example: Repair file and rebuild all.

```
pdrepair -b 7 myfile.pdf myfile_rep.pdf
```

-dp Do Not Recover Pages

By default, if pages are not part of the page tree (loose pages), they will be recovered and added at the end of the document. With the switch **-dp**, these pages are recovered, but removed from the document.

-dx Do Not Repair Cross-Reference Table

This option repairs the PDF document except for the cross-reference table. This option is useful if it is taking too long to process a document, since repairing the cross-reference table is very time consuming.

-l Set the Log File

Using the switch **-l** followed by a file name, log messages during the repair process are written into the specified log file. If no log file is specified, the log messages are written to standard error (stderr).

June 25, 2010

Example: Write messages to the log file "log.txt" instead of standard error.

```
pdrepair -l log.txt myfile.pdf myfile_rep.pdf
```

-p Specify Password

If the input file is encrypted with a user password, a password needs to be provided to read the input PDF document file. This can be either the user or owner password.

Example: Process a PDF file which is encrypted. Either the owner or the user password of the encrypted PDF file is "password".

```
pdrepair -p password myfile.pdf myfile_rep.pdf
```

-ri Recover Embedded Images

This option allows you to recover imbedded images if the file cannot be repaired.

Example: Images are extracted and stored as TIFF or JPEG files. The following command recovers all images that it can find in the resources of "myfile.pdf".

```
pdrepair -ri myfile.pdf myfile_rep.pdf
```

It extracts the images and names them as:

```
img{objno}.tif or
```

```
img{objno}.jpg.
```

where objno is the PDF object number that originally contained the image.

4.2 Return Codes

All return codes other than "0" indicate a warning or error in the processing.

Table: Return Codes	
Value	Description
0	Success
1	PDF Input File could not be opened or invalid parameters
2	PDF Output File could not be created
3	Invalid option or option values were entered
4	PDF input file is corrupt and can be repaired
5	PDF input file is corrupt and cannot be repaired but possibly be recovered

5 Troubleshooting

5.1 The Repair Time Takes Too Long

Try excluding the cross-reference table (switch **-dx**). This will speed up the repair time.

5.2 The File Cannot be Repaired

If the switch **-dx** has been selected and the cross-reference table is corrupt, the file cannot be repaired. Ensure the switch **-dx** is not set.