User Manual

3-Heights®
PDF Analysis & Repair Shell

Version 6.27.6

pdftools

Contents

1 Introduction

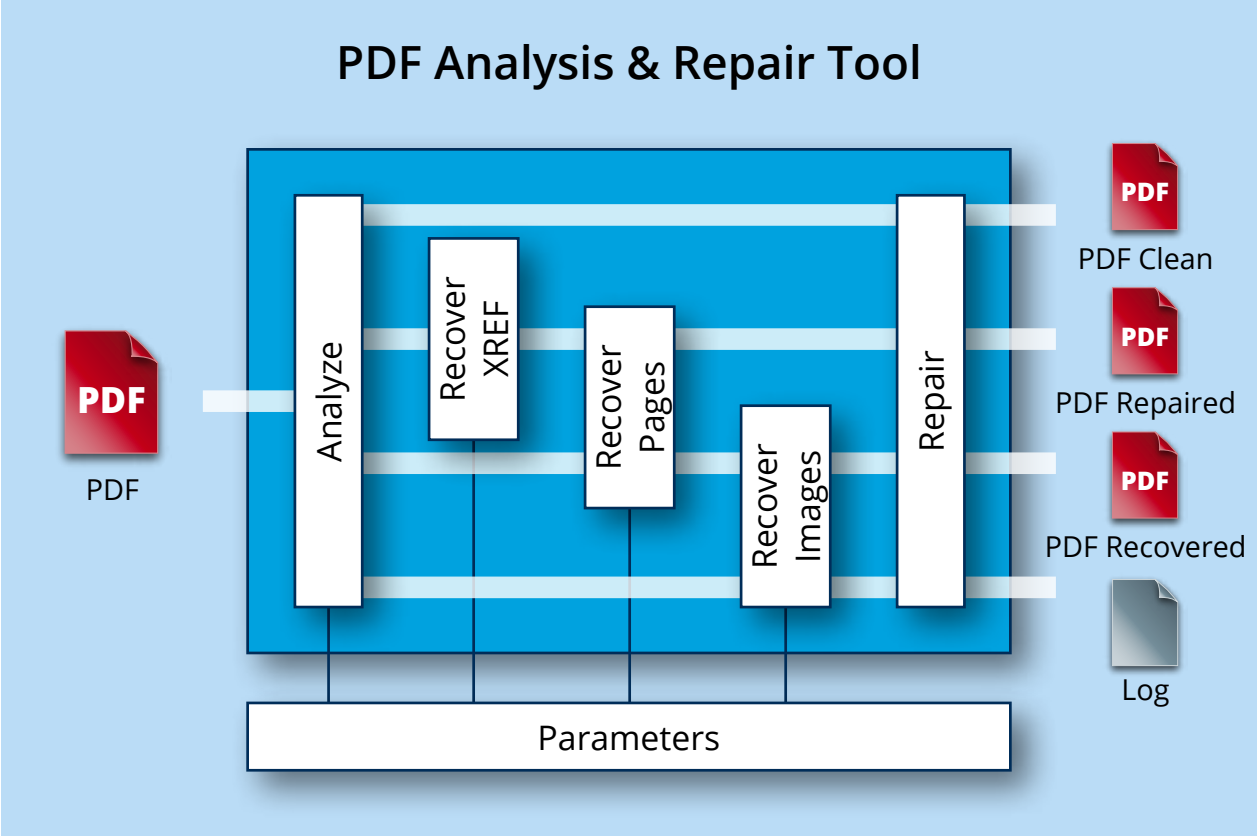
1.1 Description

The 3-Heights® PDF Analysis & Repair Shell 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, and defective copying on optical media.

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 Shell 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 Shell 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.

1.2.1 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 permission flags
- Set error level to identify whether errors, warnings or merely information occur
- Set reporting level to determine the messages to be issued (errors, warnings, information)
- Differentiate between “Repair” (corrects the errors in the document) and “Restore” (recreates the document based on the remaining legible information)

1.2.2 Formats

Input formats

- PDF 1.x (PDF 1.0, …, PDF 1.7)
- PDF 2.0
- PDF/A-1, PDF/A-2, PDF/A-3

Output formats

- PDF 1.x (PDF 1.0, …, PDF 1.7)
- PDF 2.0

1.2.3 Conformance

Standards:

- ISO 32000-1 (PDF 1.7)
- ISO 32000-2 (PDF 2.0)

1.3 Operating systems

The 3-Heights® PDF Analysis & Repair Shell is available for the following operating systems:

- Windows Client 7+ | x86 and x64
- Windows Server 2008, 2008 R2, 2012, 2012 R2, 2016, 2019, 2022 | x86 and x64

- Linux:
  - Red Hat, CentOS, Oracle Linux 7+ | x64
  - Fedora 29+ | x64
  - Debian 8+ | x64
  - Other: Linux kernel 2.6+, GCC toolset 4.8+ | x64
- macOS 10.10+ | x64

'+' indicates the minimum supported version.

2 Installation

2.1 Windows

The 3-Heights® PDF Analysis & Repair Shell comes as a ZIP archive or as an MSI installer.

To install the software, proceed as follows:

1. You need administrator rights to install this software.
2. Log in to your download account at https://www.pdf-tools.com. Select the product "PDF Analysis & Repair Shell". If you have no active downloads available or cannot log in, please contact pdfsales@pdf-tools.com for assistance.
   You can find different versions of the product available. Download the version that is selected by default. You can select a different version.
   There is an MSI (`*.msi`) package and a ZIP (`*.zip`) archive available. The MSI (Microsoft Installer) package provides an installation routine that installs and uninstalls the product for you. The ZIP archive allows you to select and install everything manually.
   There is a 32 and a 64-bit version of the product available. While the 32-bit version runs on both 32 and 64-bit platforms, the 64-bit version runs on 64-bit platforms only. The MSI installs the 64-bit version, whereas the ZIP archive contains both the 32-bit and the 64-bit version of the product. Therefore, on 32-bit systems, the ZIP archive must be used.
3. If you select an MSI package, start it and follow the steps in the installation routine.
4. If you are using the ZIP archive, uU nzip the archive to a local folder, e.g. `C:\Program Files\PDF Tools AG\`.
   This creates the following subdirectories:

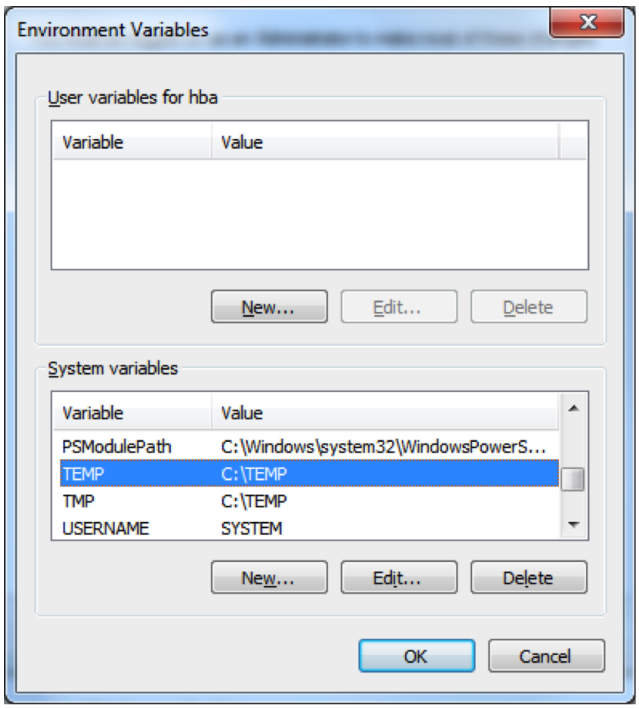
Subdirectory	Description
`bin`	Runtime executable binaries
`doc`	Documentation

5. (Optional) To easily use the 3-Heights® PDF Analysis & Repair Shell from a shell, the directory needs to be included in the "Path" environment variable.
6. (Optional) Register your license key using the License management.

2.1.1 How to set the environment variable "Path"

To set the environment variable "Path" in Windows, go to Start → Control Panel (classic view) → System → Advanced → Environment Variables.

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



2.2 Linux and macOS

This section describes installation steps required on Linux or macOS.

Here is an overview of the files that come with the 3-Heights® PDF Analysis & Repair Shell:

File description

Name	Description
`bin/x64/pdrepair`	Main executable
`doc/*.*`	Documentation

2.2.1 Linux

1. Unpack the archive in an installation directory, e.g. `/opt/pdf-tools.com/`
2. Verify that the GNU shared libraries required by the product are available on your system:

```
ldd pdrepair
```

   If the previous step reports any missing libraries, you have two options:
   a. Download an archive that is linked to a different version of the GNU shared libraries and verify whether they are available on your system. Use any version whose requirements are met. Note that this option is not available for all platforms.
   b. Use your system's package manager to install the missing libraries. It usually suffices to install the package `libstdc++6`.
3. Create a link to the executable from one of the standard executable directories, e.g.

```
ln -s /opt/pdf-tools.com/bin/x64/pdrepair /usr/bin
```

4. Optionally, register your license key using the license manager.

2.3 Uninstall

If you have used the MSI for the installation, go to Start → 3-Heights® PDF Analysis & Repair Shell… → Uninstall …

If you have used the ZIP file for the installation, undo all the steps done during installation.

2.4 Note about the evaluation license

With the evaluation license, the 3-Heights® PDF Analysis & Repair Shell automatically adds a watermark to the output files.

3 License management

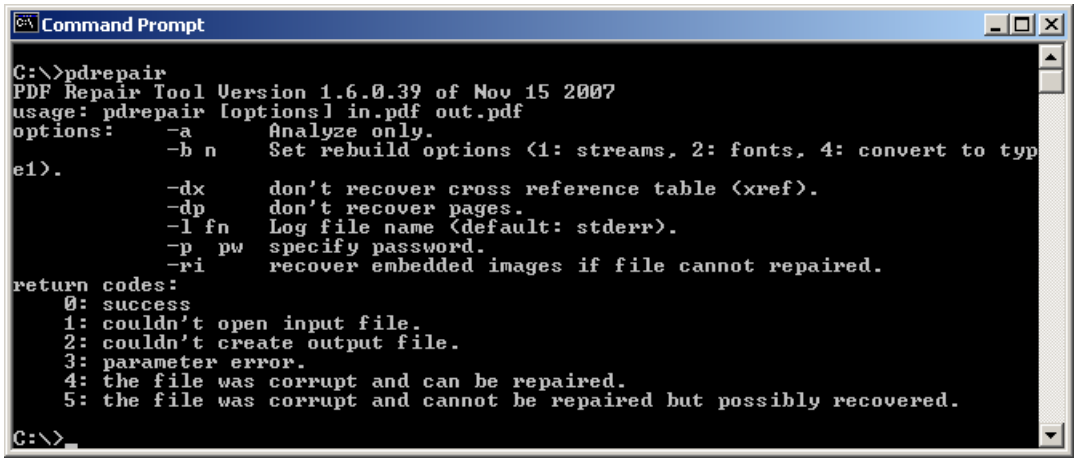
The 3-Heights® PDF Analysis & Repair Shell requires a valid license in order to run correctly. If no license key is set or the license is not valid, then the executable will fail and the return code is set to 10.

More information about license management is available in the license key technote.

4 Getting started

4.1 Usage

By typing `pdrepair` without parameters, the usage, the version, and a list of available options is returned.



4.2 Repair a file

The 3-Heights® PDF Analysis & Repair Shell 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 `input.pdf`. Repair the file and save the result in a new file called `output_rep.pdf`.

```
pdrepair input.pdf output_rep.pdf
```

4.3 Specify the folder of the output file

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

```
pdrepair input.pdf myfolder\output.pdf
```

or absolute (Windows):

```
pdrepair input.pdf C:\myfolder\output.pdf
```

4.4 Repairing all files in a directory

The 3-Heights® PDF Analysis & Repair Shell 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, use a variable to name the output files. Here is an example using the `for` command of the CMD shell:

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

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

You 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, you should proceed as follows:

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.

4.4.1 Difference between repairable and recoverable

Repairable The input PDF file contains errors that are repairable. For example, these can be 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 was 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 in such a way that the output PDF is valid according to the PDF specification.

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

- Somehow the file becomes 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.

5 Interface reference

5.1 Configuration options

5.1.1 -a Analyze only

Analyze only `-a`

When using this option, the processed input files are only analyzed and a log file is generated. There is no output created.

5.1.2 -b Set the rebuild options

Set the rebuild options `-b ‹n›`

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. CFF fonts can be converted to Type1 fonts only if fonts are rebuilt, i.e. 4 can be used together with 2 only.

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

Example: Repair the file and rebuild all

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

5.1.3 -dp Do not recover pages

Do not recover pages `-dp`

If pages are not part of the page tree (loose pages), they are recovered and added to the end of the document. If the option `-dp` is selected, loose pages are not recovered and are left out of the recovered document.

5.1.4 -dx Do not repair cross-reference table

Do not repair cross-reference table `-dx`

With this option, a corrupt XREF table is not repaired or recovered. This option is useful if it is taking too long to process a document, since repairing the cross-reference table is very time-consuming.

5.1.5 -l Set the log file

Set the log file `-l ‹fn›`

Using the `-l` switch 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).

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

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

5.1.6 -lk Set license key

Set license key `-lk ‹key›`

Pass a license key to the application at runtime, instead of using one that is installed on the system.

```
pdrepair -lk X-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX ...
```

This is required in an OEM scenario only.

5.1.7 -p Read an encrypted PDF file

Read an encrypted PDF file `-p ‹password›`

A PDF document that has a user password (the password to open the document) can only be processed when either the user or the owner password is provided. The password can be provided using the option `-p` followed by the password.

Example: The input PDF document is encrypted with a user password. Either the user or the owner password of the input PDF is “mypassword”. The command to process such an encrypted file is:

```
pdrepair -p mypassword input.pdf output.pdf
```

When a PDF is encrypted with a user password and the password is not provided or is incorrect, the 3-Heights® PDF Analysis & Repair Shell cannot read and process the file. Instead it generates the following error message:

```
Password wasn't correct.
```

5.2 Return codes

All return codes other than `0` indicate an error in the processing.

Return codes

Value	Description
`0`	Success.
`1`	Couldn't open input file.
`2`	PDF output file could not be created.
`3`	Error with given options, e.g. too many parameters.
`4`	PDF input file is corrupt and can be repaired.
`5`	PDF input file is corrupt and cannot be repaired but possibly recovered.
`10`	License error, e.g. invalid license key.

6 Troubleshooting

6.1 Repair time takes too long

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

6.2 File cannot be repaired

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

7 Version history

7.1 Changes in versions 6.19–6.27

- **Update** license agreement to version 2.9

7.2 Changes in versions 6.13–6.18

No functional changes.

7.3 Changes in versions 6.1–6.12

No functional changes.

7.4 Changes in version 5

- **Changed** error reporting behavior: Errors in the XMP metadata are no longer reported when saving a recovered document.
- **New** additional supported operating system: Windows Server 2019.

7.5 Changes in version 4.12

- **New** HTTP proxy setting in the GUI license manager.

7.6 Changes in version 4.11

- **New** support for reading and writing PDF 2.0 documents.
- **New** support for the creation of output files larger than 10GB (not PDF/A-1).
- **New** treatment of the `DocumentID`. In contrast to the `InstanceID` the `DocumentID` of the output document is inherited from the input document.

7.7 Changes in version 4.10

- **Improved** robustness against corrupt embedded font files.
- **Improved** robustness against corrupt input PDF documents.

7.8 Changes in version 4.9

- **Improved** support for recovering certain corruption types.
- **Improved** support for and robustness against corrupt input PDF documents.
- **Improved** repair of embedded font programs that are corrupt.
- **New** support for OpenType font collections in installed font collection.
- **Improved** metadata generation for standard PDF properties.

7.9 Changes in version 4.8

- **Added** repair functionality for TrueType font programs whose glyphs are not ordered correctly.

8 Licensing, copyright, and contact

Pdftools (PDF Tools AG) is a world leader in PDF software, delivering reliable PDF products to international customers in all market segments.

Pdftools 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.

Contact

PDF Tools AG
Brown-Boveri-Strasse 5
8050 Zürich
Switzerland
https://www.pdf-tools.com
pdfsales@pdf-tools.com