#### comparepdfcmd version 2.5.0-pe ("perpectual edition")

#### Mark Summerfield

Copyright © Qtrac Ltd. 2018-2025. All Rights Reserved.



#### Contents

| T . 1                         |     |     |     |     |     |     |     |     |     |     |     |     |     | 0  |
|-------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| Introduction                  | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 3  |
| Examples                      | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 3  |
| Installing                    | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 5  |
| System Requirements .         | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 5  |
| Obtaining and Installing .    |     |     | ••• |     |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• |     | 5  |
| How to View the License       |     |     |     |     |     | ••• | ••• | ••• | ••• |     |     | ••• | ••• | 5  |
| How to Get Support            |     | ••• |     |     |     |     |     | ••• | ••• |     |     |     |     | 6  |
| How to Uninstall              |     |     |     |     |     |     |     | ••• | ••• |     |     | ••• |     | 6  |
| Tips                          |     |     |     |     |     |     |     | ••• |     |     |     |     |     | 6  |
|                               |     |     | ••• |     |     | ••• |     |     |     |     |     |     |     | 7  |
|                               |     |     |     |     |     |     |     |     |     |     |     |     |     | 7  |
| Setting Your Own Defaults     |     |     |     |     |     |     |     |     |     |     |     |     |     | 8  |
| Option Names and Values       |     |     |     |     |     |     |     |     |     |     |     |     | 8   |    |
|                               |     |     |     |     |     |     |     |     | ••• |     |     |     |     | 9  |
| Appearance Mode Options       | ••• |     |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 9  |
|                               |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 9  |
|                               | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | -  |
| New Renderer Option           |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 10 |
| 1 1                           | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 10 |
| 1                             | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 11 |
| Comparison Mode Option        |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 11 |
| Configuration File Option     |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 12 |
| Highlight Transparency Opt    |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 13  |    |
| Exclude Options               | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 13  |    |
| <b>Exclude</b> Corners Option | S   |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• |     |     | ••• | ••• | 14 |
| <b>Exclude Margins Option</b> | ıs  | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• |     | 14 |
|                               |     |     |     |     |     |     |     |     |     |     |     |     |     |    |

| <b>Exclude Rectangles Options</b>  | •••   | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 15 |
|------------------------------------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| Pages Options                      | •••   | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 15 |
| Password Options                   | •••   | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• |     | ••• | 16 |
| Report Options                     |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 16 |
| Report Compact Option              |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 17 |
| Report CSV Option                  |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 17 |
| Report Decimals Option             |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 18 |
| Report Filename Option             |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 18 |
| Report Format Option               |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 19 |
| Report Output Option               |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 20 |
| Report Path Option                 |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 20 |
| <b>Report Path in Title Option</b> |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 21 |
| Report Scale Option                | •••   | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 21 |
| Report Scalable PDF Option         |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 21 |
| Report Show Option                 |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 22 |
| Report Show Field Content Op       | tion  |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 22 |
| Report Summary Option              |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 22 |
| Text Mode Options                  |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 23 |
| Change Bar Option                  |       | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 23 |
| <b>Column Tolerance Option</b>     |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 23 |
| Ignore Hyphens Option              |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 23 |
| Line Tolerance Option              |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 24 |
| Normalize Hyphens Option           | •••   | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 24 |
| Normalize Ligatures Option         |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 24 |
| Text Comparison Algorithm Op       | otion |     | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 25 |
| Verbose Option                     | •••   | ••• |     | ••• | ••• | ••• | ••• | ••• | ••• |     | ••• | 25 |
| Maximum Directory Depth Option     |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 25 |
| The PDF Files or Folders to Compa  | re    | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 26 |
| Troubleshooting                    |       | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 26 |
| Invalid PDF Report Output          | •••   | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | ••• | 26 |
| Failed to Merge Error              | •••   |     | ••• | ••• |     | ••• |     | ••• | ••• |     |     | 26 |
| False Positives                    | •••   | ••• |     | ••• | ••• | ••• | ••• | ••• |     |     | ••• | 26 |
| Appendix: Color Names              | •••   | ••• | ••• | ••• | ••• | ••• |     | ••• | ••• | ••• |     | 27 |

# Introduction

The *comparepdfcmd* application is a command line program for comparing two PDF files or two folders of PDF files that runs in a console (i.e., in a Command Prompt window opened by running cmd.exe).\* The program can report the differences in several ways—for example, by showing the differences visually in PDF or PNG format, or by outputting a CSV, JSON, or XML file identifying words that have been inserted, deleted, or replaced.

The *comparepdfcmd* program provides three comparison modes: Words (word-by-word), Charactesrs (character-by-character), and Appearance. The first two—the text modes—are concerned with the words or characters only, and ignore line wrapping and margins; the appearance mode compares the appearance of each pair of pages and so detects differences in layout, differences in graphics, and differences in colors and fonts.

For comparisons, *comparepdfcmd* uses sensible defaults for ease of use. In addition it also supports many options to support customization of the output it produces, as well as options to influence how comparisons are done.

## Examples

These examples (and those throughout this manual) make three assumptions:

- 1. That you are using *comparepdfcmd* version 2.5.0-pe ("perpetual edition")
- 2. That you replace the paths used in the examples, e.g., C:\Users\me for home, E:\pdfs for a folder of PDF files, and the E:\diff folder for outputting difference reports, with your own equivalents.
- 3. That you have added *comparepdfcmd*'s folder to your %PATH% (see <u>Add *comparepdfcmd* to the PATH</u>), *or* your current directory is *comparepdfcmd*'s folder, *or* where you see "comparepdfcmd" in the examples, you actually precede this with its full path, e.g., C:\Users\me\comparepdfcmd\comparepdfcmd.exe.

Here's how to return a value (%errorlevel%) to the operating system (e.g., for use in a .bat file):

C:\Users\mark\>comparepdfcmd E:\pdfs\a1.pdf E:\pdfs\a2.pdf

To get some human-readable feedback use the verbose option (either -v or --verbose), e.g.:

```
C:\Users\mark\>comparepdfcmd -v E:\pdfs\b1.pdf E:\pdfs\b2.pdf
Different
```

This will return a value to the operating system as described above, and will also output either "Same" or "Different" depending on whether the two PDFs have the same words or not. (Note that colors and font effects shown in the examples, have no significance—they are used purely to aid understanding.)

<sup>\*</sup> If you need a GUI PDF comparison application, try DiffPDF: <u>www.diffpdf.com/diffpdf.html</u>.

If you want to see what the differences are, tell *comparepdfcmd* to output a difference report using the -r or --report option. For example:

C:\Users\mark\>comparepdfcmd -v -s -r E:\diff\report.pdf E:\pdfs\c1.pdf E:\pdfs\c2.pdf Wrote "E:\diff\report.pdf"

The report.pdf file will contain the different pages side-by-side and with the differences highlighted. In addition, due to the -s or --show option, the report will be shown in your PDF reader.

Incidentally, if c1.pdf and c2.pdf were the same, no report would be produced and the program would display "Same" rather than "Wrote "E:\diff\report.pdf"". And, of course, no report would be shown in your PDF reader.

If you want to compare two *folders* of PDFs which have the same names, you can give the folder names instead of the names of two PDFs, e.g.:

```
C:\Users\mark\>comparepdfcmd -f pdf -R E:\diff E:\pdfs\old E:\pdfs\new
"Result","PDF#1","PDF#2","Report or Error"
"Missing#1",,"E:\pdfs\new\only_in_new.pdf",
"Missing#2","E:\pdfs\old\only_in_old.pdf",,
"Same","E:\pdfs\old\one.pdf","E:\pdfs\new\one.pdf",
"Different","E:\pdfs\old\two.pdf","E:\pdfs\new\two.pdf","E:\diff\two.pdf"
```

The -f or --format option specifies the format of the difference report files; here it is PDF format (see the <u>Report Format Option</u>). The -R or --reportpath option specifies where the reports should go (see the <u>Report Path Option</u>). These options are followed by the two folders to search for PDF files to compare. The program shows the results in CSV format on the console, although you can tell it to save the summary into a file instead using the -S or --summary option (see the <u>Report Summary Option</u>).

When, comparing two folders of PDFs, by default, *comparepdfcmd* will look in subfolders ers (down to a depth of 1000 subfolders). However, it is possible to limit the depth to just the folders specified by using the -D or --maxdirdepth option with a value of 1. (See the Maximum Directory Depth Option.)

Many more command line options are supported, including using passwords to access password-protected PDFs, as well as options to exclude margins from comparisons (e.g., to ignore timestamps in a header or footer), amongst others. Even more options can be set in .ini configuration files. All this is explained in the <u>Configuration</u> section.

# Installing

## System Requirements

The *comparepdfcmd* program runs desktops and laptops running 64-bit Windows 7, 10, or later. The machine must have an x86-compatible processor (i.e., most desktop and laptop computers).

Since PDF comparisons are computationally expensive, it is best to use a machine with a fast multicore processor, ideally with a processor speed of at least 2GHz that has at least two cores. *comparepdfcmd* will take advantage of all available cores, so in general will run twice as fast on a quad core machine than on a dual core machine with the same clock speed—and faster still on machines with more and faster cores.

The *comparepdfcmd* program makes use of the system's default PDF reader, e.g., to show this manual, to show the license, and to show PDF difference reports when told to. So, naturally, a PDF reader must be installed.\*

## **Obtaining and Installing**

The program is provided in a single .zip file, e.g., comparepdfcmd-2.5.0-pe.zip which is provided once payment is received.

Once the .zip has been downloaded, unzip it in any convenient folder. This is all that is required for installation—*comparepdfcmd* does not touch the Windows registry, and does not change the %PATH%. However, having *comparepdfcmd* in your %PATH% can be very convenient, so we recommend that you permanently add it. (See the <u>Add *comparepdfcmd* to the PATH</u> sidebar.) Once the software has been installed it will work as-is with no license key required. The software has no expiry date and is not tied to any particular physical or virtual computer.

### How to View the License

The program's license can be viewed with this command:

#### C:\Users\mark\>comparepdfcmd license

This will show the license in your PDF reader, or if it can't, it will output the license to the console. Or, you can read the license PDF from *comparepdfcmd*'s folder since it is included in the .zip file. The license file is comparepdfcmd-license.pdf.

<sup>\*</sup> We prefer SumatraPDF www.sumatrapdfreader.org but Acroread, Acrobat, Evince, or any other should work fine.

#### Add comparepdfcmd to the PATH

Click the Start button, then click Control Panel. In the Search box, enter "path". This should give you a link to "Edit environment variables for your account". Click this link. This should pop up the Environment Variables dialog.

Click the Path variable in the top list of user variables, then click the Edit button. Click the Variable value's line editor, then press End to move the cursor to the very end. Now enter a semi-colon ";" followed by the drive and path to the folder you have unzipped *comparepdfcmd* into. For example, ;"C:\Users\me\comparepdfcmd". Then click OK, then close the Control Panel window.

## How to Get Support

It is worth reading this manual once through to become familiar with what *comparepdfcmd* can do. This manual covers every feature and configuration option that *comparepdfcmd* provides. However, if this manual doesn't answer your question, then we'll happily try to answer.

Email support is provided on a "best efforts" basis with no formal time limit, so long as Qtrac Ltd. is trading. The support email address is support@qtrac.eu — be sure to tell us what operating system you are using (e.g., "Windows 11") and especially which version of *comparepdfcmd* you are using.

## How to Uninstall

To uninstall *comparepdfcmd*:

- 1. Delete *comparepdfcmd*'s folder;
- 2. Delete comparepdfcmd.ini from your home folder (if it exists);
- 3. Remove *comparepdfcmd* from your %PATH% (if you added it).

No other steps are needed since *comparepdfcmd* does not touch the Windows registry.

# Tips

By default, *comparepdfcmd* will compare PDFs using words mode. This will produce the best results for most Latin-based languages like English or Spanish. If you want to compare text written in a CJK—Chinese, Japanese, Korean—language, it is best to use chars (characters) mode. And if you want to compare colors, fonts, diagrams, images, or text layout, you should use appearance mode. See the <u>Comparison Mode Option</u>.

If your PDFs have different numbers of pages or inserted or deleted pages, you can use page ranges to specify those pages from each PDF that correspond to each other. See the <u>Pages Options</u>.

The default text comparison algorithm should produce good results for single column PDFs, but for multicolumn PDFs it is usually best to use the multicolumn comparison algorithm: see the <u>Text Comparison Algorithm Option</u>.

It is also possible to tweak the comparison algorithm itself by setting the <u>Line Tolerance Option</u>, and for the multicolumn algorithm, both the <u>Line Tolerance Option</u> and the <u>Column Tolerance Option</u>. In particular, reducing the <u>Line Tolerance Option</u> from its default of 10 to 9 or 8 may help improve results for PDFs whose inter-line spacing is narrower than usual.

If any problems occur, have a look at the <u>Troubleshooting</u> section (which includes how to get support).

## Configuration

When *comparepdfcmd* runs it must be told which PDF files or folders to compare. And in most cases you'll want to specify extra options, for example, to say what kind of comparison you want done, and how to report differences. You may also want to change some of the defaults it uses. This can be done by creating a comparepdfcmd.ini configuration file and putting it in one of the locations that *comparepdfcmd* searches for configurations; or by creating a configuration file with any name you like and using the -C or --config option to tell *comparepdfcmd* where to find it (see the <u>Configuration File Option</u>). Here's how the configuration process works:

- 1. *comparepdfcmd* starts with every option set to a sensible default wherever possible;
- 2. Options are read from the file COMPAREPDFCMD\comparepdfcmd.ini if it exists (where COM-PAREPDFCMD is the folder where comparepdfcmd.exe is located);
- 3. Options are read from the file %USERPROFILE%\comparepdfcmd.ini if it exists (i.e., in your home folder);
- 4. Options are read from the file comparepdfcmd.ini (i.e., in the current folder) if it exists;
- 5. If the -C or --config command line option is given, options are read from the file it specifies;
- 6. Any options given on the command line (apart from -C or --config) are read.

Once the options have been read and a final configuration is reached, *comparepdfcmd* then performs the comparison in accordance with the configuration.

### **Exit Status Code**

During the comparison, and when the comparison is complete, *comparepdfcmd* may output messages depending on the options that have been set. And at the end the program returns a value to the operating system (accessible as %errorlevel% in .bat files). The value will be 0 if there were no errors and no differences were found, 1 if there were no errors and differences were detected, or 2 if an error occurred.

### **Setting Your Own Defaults**

If you want to change any of *comparepdfcmd*'s defaults, create a plain text file called comparepdfcmd.ini and save it in the same folder as *comparepdfcmd*'s executable or in your home folder (%USERPROFILE%). For any default you want to change, put your preferred default in this file. For example, if you want *comparepdfcmd* to default to doing appearance mode comparisons, add these lines:

```
[general]
mode=appearance
```

Examples of how to change other defaults are given wherever applicable in the explanations of *comparepdfcmd*'s options that follow. You can also copy parts of the example.ini file that you'll find in *comparepdfcmd*'s folder and paste them into your own comparepdfcmd.ini file, editing them to suit your needs.

### **Option Names and Values**

Command line options are always specified Unix-style, e.g., -v or --verbose, not Windows-style (/v). Command line option names are case-sensitive.

In configuration files, blank lines are ignored, and so is any text that begins with a semi-colon. This makes it possible to add comments if you want. (For an example, see the example.ini file in comparepdfcmd's folder.) Configuration file section names and key names are not case-sensitive.

Most option values are simple pieces of text or numbers, e.g., "words" or "30". Numbers that represent lengths or coordinates are assumed to be points  $(\frac{1}{72})$  or about 0.353mm). The top-left of the page is point (0, 0), with positive y coordinates moving down, and positive x coordinates moving right.

Some option values are used to turn a feature on or off. For these, a value of "1", "true", "t", "ves", "y", or "on", will turn the feature on; and any other value will turn it off. (And no value at all will leave it at its default.)

Option values which represent colors can be specified by name or using HTML-style colors: see the Color Options section for details.

Page numbers are always counted from 1, regardless of the page numbers used in the PDF itself (e.g., i, ii, iii, ..., 1, 2, 3, ...).

The most important option is the <u>Comparison Mode Option</u>. This defaults to words mode, so doesn't even need to be set if that's the kind of comparison you want (and providing you haven't changed this default in a configuration file). If you want to see the differences, then you'll need to use the <u>Report Options</u>. If the pages don't exactly correspond (e.g., due to inserted or deleted pages), or you only want to compare some specific pages, then you'll need to use the Pages Options. And, of course, you must specify The PDF Files or Folders to Compare.

### **Alternative Options**

It is possible to tell *comparepdfcmd* to provide some information or perform a special operation, without doing any comparisons at all. This involves using one of the alternative command line options. The most commonly used is -h or --help, e.g.,

E:\pdfs\>comparepdfcmd -h

This outputs to the console a summary of the standard and alternative command line options. Another useful alternative option, is -m or --manual or manual:

E:\pdfs\>comparepdfcmd -m

This shows *comparepdfcmd*'s manual in your PDF reader. Especially when requesting support, the -V (*capital-V*) or --version option is useful:

E:\pdfs\>comparepdfcmd -V

This shows comparepdfcmd's version.

The other alternative option, used individually, is license or --license. This was covered earlier in the <u>Installing</u> section and its subsections.

### **Appearance Mode Options**

These options only apply when appearance mode comparisons are made. (See also the <u>Comparison Mode Option</u>.) To control the highlighting color used in this mode, see the <u>2.0.0</u> <u>Color Options' highlight option and the Highlight Transparency Option</u>.

#### **Accuracy Option**

This tells *comparepdfcmd* how accurate appearance mode comparisons should be; it is ignored for text mode comparisons.

The default is 100 (i.e., 100%) which means that each pair of pages that are compared are only considered the same if they match pixel for pixel. The minimum accuracy is 80 (80%), and the maximum, 100.

Appearance mode comparisons are made by dividing each pair of pages into little squares and comparing each pair of corresponding squares. In some situations it is helpful to accept small differences (often ones that humans wouldn't even notice). This can be achieved by lowering the accuracy. For example, reducing the accuracy to 98% means that every pair of squares that are compared can be up to 2% different and still be considered "the same".

This option can be specified on the command line using the syntax -A *number* or --accuracy=*num*-*ber*. Here are two equivalent examples.

E:\pdfs\>comparepdfcmd -a -A90 -r diff.pdf file1.pdf file2.pdf

E:\pdfs\>comparepdfcmd --appearance --accuracy=90 --report diff.pdf file1.pdf file2.pdf

Both specify appearance mode, and an accuracy of 90%, and tell *comparepdfcmd* that if there are differences it should create the difference report file, diff.pdf, containing each pair of different pages with their differences highlighted. (See the <u>Comparison Mode Option</u> and the <u>Report Filename Option</u>.)

To set the default accuracy to, say, 95% for appearance mode comparisons, put these lines in a configuration file:

[appearance] accuracy=95

Whatever the default is, it can always be overridden on the command line using the -A or --accuracy option.

See also the <u>Square Size Option</u>.

#### **New Renderer Option**

This is an advanced option that should not normally need to be changed.

Whether to use the "new" renderer for appearance mode comparisons.

The "old" renderer is fast and accurate for almost all appearance mode comparisons. However, in some rare cases it may fail to correctly detect differences in some gradient fills or textures. In such cases, setting this option to true will solve the problem.

[appearance] new-renderer=true

#### **Square Size Option**

This is an advanced option that should not normally need to be changed.

The square size to use for appearance mode comparisons.

When a pair of pages is compared in appearance mode, i.e., pixel for pixel, the pages are divided into squares, and each corresponding square is compared. The default square size is 10 points, i.e., squares of size  $10 \times 10$  (100 pixels) are compared.

To change the default, put these lines in a configuration file, replacing 10 with your preferred default (minimum 4 points, maximum, 144 points):

```
[appearance] square-size=10
```

See also the <u>Accuracy Option</u>.

2.1.7

### **Color Options**

For reports output in PDF or PNG format, differences are indicated by highlighting them as if you'd drawn over them with a highlighter pen. This makes the differences stand out and at the same time allows the highlighted text or graphics to show through. The show through is controlled by the <u>Highlight Transparency Option</u>.

If corners, margins, or rectangles are excluded, their outlines are drawn to indicate the excluded areas: see the <u>Exclude Options</u>.

For text mode comparisons a colored vertical change bar is drawn in the left margin wherever changes are detected: see the <u>Change Bar Option</u>.

Sensible defaults are used for all the colors, but you can override them by adding lines to a configuration file using this syntax:

[colors] what=color

The *what* can be change-bar, exclusion-line, highlight, delete, insert, or replace. The highlight color is used for Appearance mode comparisons, and the delete, insert, and replace colors are used for text mode comparisons.

The *color* can be specified using any of the names listed in the <u>Appendix: Color Names</u>, or an HTML-style hexadecimal color name with the form *#hhh* or *#hhhhhh*, where each h is a hexadecimal digit.

For example:

```
[colors]
highlight=Teal
exclusion-line=#FF007F
replace=#EB5
```

Note that color names are *not* case-sensitive, so Teal, teal, and TEAL are all the same color.

### **Comparison Mode Option**

The comparison mode tells *comparepdfcmd* how to compare the PDF files. It can be specified on the command line or in a configuration file. The default is words mode unless you change this using a configuration file.

To specify appearance mode on the command line, you can use the syntax -a or --appearance. This will override the default. Here are two equivalent examples:

```
E:\pdfs\>comparepdfcmd -a file1.pdf file2.pdf
E:\pdfs\>comparepdfcmd --appearance file1.pdf file2.pdf
```

To make appearance mode the default, put these lines in a configuration file:

[general] mode=appearance

To specify chars (characters) mode on the command line, you can use the syntax -c or --chars. This will override the default. Here are two equivalent examples:

E:\pdfs\>comparepdfcmd -c file1.pdf file2.pdf
E:\pdfs\>comparepdfcmd --chars file1.pdf file2.pdf

To make chars mode the default, put these lines in a configuration file:

[general] mode=chars

The default is words mode, so if this is what you want to use there is normally no need to specify it at all. However, if a different mode has been set as the default in a comparepdfcmd.ini file, you can override it on the command line, e.g., using the syntax –w or –-words to specify words mode:

E:\pdfs\>comparepdfcmd -w file1.pdf file2.pdf

Appearance comparisons (appearance mode) works by in effect taking a "photograph" of each pair of pages being compared, and then comparing these images pixel by pixel. By default, even a single pixel difference is considered significant. However, it is possible to use the <u>Accuracy Option</u> to make comparisons tolerant of small differences. Another option that affects appearance mode comparisons is the <u>Square Size Option</u>.

Word-by-word comparison (words mode) is best for comparing documents' texts—this mode is also the fastest.\* However, in some situations (perhaps when comparing documents written in CJK—Chinese, Japanese, Korean—languages), character-by-character comparisons (chars mode), may produce better results, but at the cost of being slower. Both words and chars modes (the "text" modes) only consider text: they ignore images, fonts, and layout (apart from the order of the characters or words). For complete comparisons—of text (including fonts and layout), and of images—use appearance mode. Text mode comparisons are affected by the various <u>Text Mode Options</u>.

In addition, *all* comparisons are affected by the <u>Exclude Options</u> and the <u>Pages Options</u>.

### **Configuration File Option**

This option accepts the name of a configuration file given on the command line. The options in this file override any defaults, but themselves may be overridden on the command line.

This option may only be specified on the command line using the syntax -C *configfile* or --config*=configfile*, where *configfile* is a filename and may include a path. Here are two equivalent examples:

<sup>\*</sup> In tests *comparepdfcmd*'s words mode comparisons were  $2 \times -4 \times$  faster than *DiffPDF* or *diffpdfc*. Furthermore, *comparepdfcmd* makes better use of processor cores, and uses less memory.

```
comparepdfcmd • 13
```

```
E:\pdfs\>comparepdfcmd -C E:\myconfig.ini file1.pdf file2.pdf
E:\pdfs\>comparepdfcmd --config E:\myconfig.ini file1.pdf file2.pdf
```

This is useful for overriding defaults that cannot be set on the command line, or for using a specific set of options for particular comparison tasks without changing the overall defaults.

To permanently set your own default options, see Setting Your Own Defaults.

For an example configuration file, open example.ini in *comparepdfcmd*'s folder using a plain text editor.

### **Highlight Transparency Option**

Differences are indicated by highlighting them as if you'd drawn over them with a highlighter pen. This makes the differences stand out and at the same time allows the highlighted text or graphics to show through.

The default is 80 (i.e., 80%). The minimum is 10 (10%) and the maximum is 90 (90%).

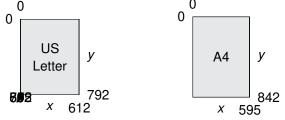
You can change the default if you want. For example, by reducing the transparency to, say, 70%, you can make the highlighting stronger, while still allowing the differences to be visible. This can be done by putting these lines in a configuration file:

```
[general]
highlight-transparency=70
```

You can change the underlying highlighting colors used by setting custom <u>Color Options</u>.

## **Exclude Options**

It is sometimes convenient to exclude a region of each page from comparisons. For example, you might want to exclude a header or footer that has a timestamp that is different for every PDF you compare, but which you don't want to affect the comparison.



The program can exclude corners, margins, and arbitrary rectangles.

Any text wholly within an excluded corner, margin, or rectangle, is ignored for text mode comparisons, and any part of the page within an excluded corner, margin, or rectangle, is ignored for appearance mode comparisons.

Coordinate (0,0) is the top-left corner; coordinate (612,792) is the bottom-right corner of a US-Letter size page, and coordinate (595,842) is the bottom-right of an A4 page, as illustrated above.

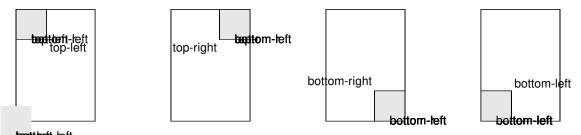
1.1.4

2.0.0

For PDF and PNG reports, exclusions are shown by dashed (PDF) or solid (PNG) red lines. To 2.0.0 change the color used, see the <u>Color Options</u>.

#### **Exclude Corners Options**

Excluded corners, if set at all, are set individually, and are specified as an x, y coordinate pair relative to their nearest corner.



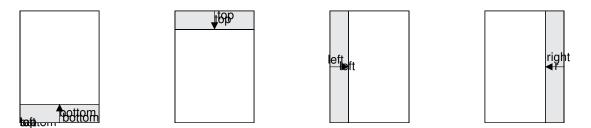
To exclude corners, add a corner exclusion section to a configuration file. For example, to exclude a one inch (25.4 mm) corner from the top-right of the PDFs being compared, add these lines:

```
[corners]
top-right=72,72
```

Corner coordinates may range from 0 to 300. You can specify up to four corners in a corners section, bottom-left, bottom-right, top-left, and top-right. The coordinates don't have to be the same—so you can exclude corner rectangles, not just squares.

#### **Exclude Margins Options**

Excluded margins, if set at all, are set individually, and are specified as a length from their nearest edge.



Margins can be excluded using the command line with the syntax --margin-which=number, where which is left, right, top, or bottom. For example, --margin-bottom=54. You can use none, one, two, three, or all four of these options on the command line.

It is also possible to exclude margins using a configuration file. To do this, add a margin exclusion section to a configuration file. For example, to exclude a <sup>3</sup>/<sub>4</sub> inch (19 mm) margin from the bottom of the PDFs being compared, add these lines:

[margins]
bottom=54

Margin coordinates may range from 0 to 300. You can specify up to four margins in a margins section, bottom, left, right, and top.

#### **Exclude Rectangles Options**

Excluded rectangles, if set at all, are set individually, and are specified as coordinates x1, y1, x2, y2, where (x1, y1) is a rectangle's top-left corner, and (x2, y2) is its bottom right corner.

To exclude rectangles, add a rectangles exclusion section to a configuration file. For example, to exclude the three rectangles illustrated, add these lines:

[rectangles]
rectangle1=155,79,217,158
rectangle2=300,400,350,450
rectangle3=61,533,558,594

Each rectangle must be individually numbered (from rectangle1 to rectangle20). The numbers need not be consecutive and there can be gaps (e.g., rectangle9, then rectangle15); nor does their order within the rectangles section matter.

Although rectangles can be used to exclude corners and margins, as well as arbitrary rectangles, it is usually easier to specify corners and margins using the <u>Exclude Corners Options</u> and the <u>Exclude Margins Options</u>.

### **Pages Options**

By default, *comparepdfcmd* compares all the pages in the first PDF with their corresponding pages in the second PDF.

For example, if you compare two PDF files, by default *comparepdfcmd* will compare both their page 1's with each other, both their pages 2's, both their page 3's, and so on.

In the case of PDF files that have different numbers of pages, or where you only want to compare particular pages, you can specify exactly which pages *comparepdfcmd* should compare using the command line --pages1 and --pages2 options. If neither option is given, *comparepdfcmd* will compare all the pages. If --pages1 is given, but not --pages2, then the page set for --pages1 will also be used for --pages2. And if both are specified, then *comparepdfcmd* will compare as specified.

Pages can be specified using the syntaxes --pages1=pages and --pages2=pages where pages is a set of comma-separated page numbers, or page ranges of the form *startpage-endpage*, or a mixture of both. Here are some examples:

```
E:\pdfs\>comparepdfcmd -r diff.pdf file1.pdf file2.pdf
E:\pdfs\>comparepdfcmd -r diff.pdf --pages1=1,3,5,7,11-18 file1.pdf file2.pdf
```

```
comparepdfcmd • 16
```

E:\pdfs\>comparepdfcmd -r diff.pdf --pages1=3-19 --pages2=5-21 file1.pdf file2.pdf

In the first example, all the pages are compared. In the second example pages 1, 3, 5, 7, 11, 12, 13, 14, 15, 16, 17, and 18 are compared. And in the third example file1.pdf's pages 3–19 are compared with file2.pdf's pages 5–21, i.e., 3 vs. 5, 4 vs. 6, 5 vs. 7, ..., 18 vs. 20, 19 vs. 21. In each case a difference report is written to the file diff.pdf; see the <u>Report Filename Option</u>.

### **Password Options**

If one or both of the PDFs you want to compare is password-protected, you must tell *comparepdfcmd* what password(s) to use so that it can read the PDFs. Note that if you tell *comparepdfcmd* to provide a PDF differences report, the report will *not* be password-protected.

To specify the password for the first PDF to compare use the syntax --password1=*password1* and similarly, to specify the password for the second PDF use the syntax --password2=*password2*. For example:

```
E:\pdfs\>comparepdfcmd -r diff.pdf --password1=Abc5 pwd1.pdf file2.pdf
E:\pdfs\>comparepdfcmd -r diff.pdf --password2=XyZ6 file1.pdf pwd2.pdf
E:\pdfs\>comparepdfcmd -r diff.pdf --password1=Abc5 --password2=XyZ6 pwd1.pdf pwd2.pdf
```

In the first example, pwd1.pdf is password-protected with password "Abc5", and compared with the unprotected file2.pdf. In the second example, unprotected file1.pdf is compared with pwd2.pdf which is password-protected with password "XyZ6". And in the third example the two password-protected PDFs are compared to each other. Note that in every case, if the PDFs are different, the diff.pdf report will be *unprotected*.

### **Report Options**

To get a difference report when comparing two PDFs, use the -r or --report command line option (see <u>Report Filename Option</u>). To have *comparepdfcmd* show this report in your PDF reader, use the -s or --show option (see <u>Report Show Option</u>).

To get difference reports when comparing two *folders* of PDFs, use the -R or --reportpath option (see <u>Report Path Option</u>) *and* the -f or --format option (see <u>Report Format Option</u>). The other options described in this section are used to configure some aspects of the reports.

For PDF and PNG reports you can override the default colors used for the change bar, 2.0.0 exclusions, and highlighting: see the <u>Color Options</u>.

#### **Report Compact Option**

If this option is false (the default), then JSON and XML reports will be output in a humanreadable form with newlines and indentation. If this option is true, then JSON and XML reports will be output as compactly as possible.

To change the default to true, put these lines in a configuration file:

[report]
compact=true

See also the <u>Report Decimals Option</u>.

#### **Report CSV Option**

CSV reports can be output in one of two different formats. The default format is normal:

[report] csv=normal

The normal format produces reports using one of these formats:

```
"File", "Page", "X1", "Y1", "X2", "Y2", "Text", "Change"
"File", "Page", "X1", "Y1", "X2", "Y2"
```

For CSV normal format, the differences in File #1 are shown, followed by the differences in File #2.

The first format is used for text mode comparisons and the second for appearance mode comparisons.

Alternatively, the special format can be used.

[report] csv=special

The special format produces reports using one of these formats:

```
"File1","Page","X1","Y1","X2","Y2","Text","Change","File2","Page","X1","Y1","X2","Y2","Text","Change"
"File1","Page","X1","Y1","X2","Y2","File2","Page","X1","Y1","X2","Y2"
```

For CSV special format, all the differences in File #1 are shown, side-by-side with the differences in File #2.

Again, the first format is used for text mode comparisons and the second for appearance mode comparisons.

See also the <u>Report Filename Option</u>.

2.0.0

2.4.2

#### **Report Decimals Option**

This option is used to set how many digits to show after the decimal point in CSV, JSON, and XML reports. The default is 2.

To change the default, put these lines in a configuration file, replacing 2 with your preferred default (minimum 0, maximum, 4):

[report]
decimals=2

See also the <u>Report Compact Option</u>, the <u>Report Filename Option</u>, and the <u>Report Format Option</u>.

#### **Report Filename Option**

This option is used when comparing two PDFs to specify the filename of a difference report.

(If you are comparing two *folders* of PDFs, don't use this option, use the <u>Report Path Option</u> and the <u>Report Format Option</u> instead.)

There is no default for this option, so *comparepdfcmd* will not output any reports by default.

Five output formats are supported:

- CSV (comma-separated variable; .csv)—suitable for importing into a spreadsheet and for regression testing;
- JSON (JavaScript Object Notation; .jsn or .json)—suitable for regression testing;
- PDF (portable document format; .pdf)—containing each pair of differing pages with differences highlighted;
- PNG (portable network graphics; .png)—an image for each pair of pages that differ, with differences highlighted;
- XML (extensible markup language; .xml)—suitable for post-processing and regression 2.0.0 testing.

If you output a PDF report, you can tell *comparepdfcmd* to show it in your PDF reader using the <u>Report Show Option</u>.

This option can only be specified on the command line using the syntax -r filename or --report=filename where filename is the filename for the difference report. The filename may include a path, and *must* end with .csv, .jsn, .json, .pdf, .png, or .xml, depending on which report format you want. If you want reports in two or more formats, simply repeat this option once for each format.

Here is an example:

```
E:\pdfs\>comparepdfcmd -v -r E:\diff\diff.pdf E:\pdfs\f1.pdf E:\pdfs\f2.pdf
Wrote "E:\diff\diff.pdf"
```

1.1.5

Here, two PDF files were compared and differences found which are shown in the diff.pdf file. If you give a report filename that includes a path, then what you have specified will be used. However, if you give a report filename without a path, the path specified by the <u>Report Path Option</u> will be used. The <u>Report Path Option</u> has no default, so if no path is given in the report filename and the <u>Report Path Option</u> isn't set, the report will be written to the current folder.\*

In the following example we assume that the <u>Report Path Option</u> has been set in a configuration file to, say, E:\diff.

```
E:\pdfs\>comparepdfcmd -v -r x.csv -r y.json -r z.pdf E:\pdfs\g1.pdf E:\pdfs\g2.pdf
Wrote "E:\diff\x.csv"
Wrote "E:\diff\y.json"
Wrote "E:\diff\z.pdf"
```

Because the <u>Report Path Option</u> has been set, *and* no path is specified by the -r or --report options, the <u>Report Path Option</u> has been used.

Here's one final example:

```
E:\pdfs\>comparepdfcmd -v -r E:\diff\pagesimage-.png E:\pdfs\h1.pdf E:\pdfs\h2.pdf
Wrote "E:\diff\pagesimage-09-09.png"
Wrote "E:\diff\pagesimage-11-11.png"
Wrote "E:\diff\pagesimage-12-12.png"
```

In this example, the two PDFs differ on pages 9, 11, and 12, and each pair of pages has been saved into its own PNG file using the report filename as the *pattern* which *comparepdfcmd* fills in as *name-page1-page2*.png. Naturally, the path could have been omitted if you wanted to save the images to the current folder, or to a folder specified by the <u>Report Path Option</u>.

#### **Report Format Option**

This option is used when comparing two *folders* of PDFs to specify the format to use for difference reports.

(If you are comparing just two PDFs, don't use this option, use the <u>Report Filename Option</u> instead.)

There is no default for this option, so *comparepdfcmd* will not output any reports by default when comparing two folders of PDFs.

The same five output formats as the <u>Report Filename Option</u> are supported. (JSON support was added in version 1.1.5; XML support was added in version 2.0.0) 2.0.0

This option can be set on the command line using the syntax -f format or --format=format where format is csv, jsn, json, pdf, png, or xml. For example:

<sup>\*</sup> Note that if the current folder is C:\, Windows will normally *silently ignore* requests to write to it.

```
E:\pdfs\>comparepdfcmd -f pdf -R E:\diff E:\pdfs\old E:\pdfs\new
"Result","PDF#1","PDF#2","Report or Error"
"Missing#1",,"E:\pdfs\new\only_in_new.pdf",
"Missing#2","E:\pdfs\old\only_in_old.pdf",,
"Same","E:\pdfs\old\one.pdf","E:\pdfs\new\one.pdf",
"Different","E:\pdfs\old\two.pdf","E:\pdfs\new\two.pdf","E:\diff\two.pdf"
```

Here the report format has been set to PDF using -f pdf and a path for the reports has been specified using -R E:\diff (i.e., using the <u>Report Path Option</u>).

In general we recommend always using this option and the <u>Report Path Option</u> together, unless they are both set in configuration files.

To change the default of no format (i.e., no reports output when comparing folders), put these lines in a configuration file:

```
[report]
format=pdf
```

Naturally, you could set the format to csv, json, png, or xml, if you preferred one of them to PDF format for reports of PDFs that differ when comparing folders.

#### **Report Output Option**

This is an advanced option that should not normally need to be changed.

By default PDF and PNG reports show each pair of differing pages side-by-side with differences highlighted. However, if you only want to see the pages in the first PDF (or first folder's PDFs when comparing folders of PDFs), you can tell *comparepdfcmd* to only output these pages with differences highlighted. Similarly, you can tell *comparepdfcmd* to only output the pages of the second PDF (or second folder's PDFs).

To change the default of showing pairs of pages side-by-side, put these lines in a configuration file:

```
[report]
ouput=one
```

This will only output the first PDF's different pages (or the first folder's PDFs). If you prefer the second, then use output=two instead. To get the default, just delete this line or use output=both.

#### **Report Path Option**

Use this option to specify the folder to use for reports when comparing two PDFs using the <u>Report Filename Option</u> (in which case the report filenames do not need to have paths since they'll be given the path specified by this option). And use this option when comparing two *folders* of PDFs—in conjunction with the <u>Report Format Option</u>.

This option can be set on the command line using the syntax -R *path* or --reportpath=*path*. For an example, see the <u>Report Format Option</u>, page 20.

To change the default of no default report path, put lines like these in a configuration file:

[report]
path=E:\diff

Naturally, you would use the path relevant to your system rather than the  $E:\$  here.

#### **Report Path in Title Option**

Use this option to specify whether to include the full path to the compared PDFs in report titles.

To change the default of including the full path of compared PDFs in report titles, put lines like these in a configuration file:

[report]
path-in-title=false

#### **Report Scale Option**

Use this option to scale PNG reports.

By default PNG reports are output at 100% scale (i.e., at 72 DPI). However, it can sometimes be useful to use higher resolutions, especially for very small PDFs.

To change the default of 100 (100%), put lines like these in a configuration file:

```
[report]
scale=150
```

The minimum is 100 (the default), and the maximum is 800 (800%). To find the best scale, try 150, then 200, and so on.

Higher resolution output takes more time (although this may not be noticable for small PDFs).

#### **Report Scalable PDF Option**

Use this option to switch off scalable PDF reports.

By default, PDF reports are output in scalable form.

However, in some very rare cases this fails to work correctly. In such cases set this option to false in a configuration file:

[report]
scalablepdf=false

1.1.7

2.3.0

If the resultant PDF report is too grainy, increase the report scale (e.g., to 150% or 200%; see <u>Report Scale Option</u>).

#### **Report Show Option**

Use this option to show *comparepdfcmd*'s PDF report if two PDFs are different.

This option can only be set on the command line using the syntax -s or --show. It only applies when comparing two PDF files (*not* two folders), *and* when the <u>Report Filename Option</u> is used to specify a PDF report.

For example:

```
E:\pdfs\>comparepdfcmd -v -s -r E:\diff\report.pdf E:\pdfs\k1.pdf E:\pdfs\k2.pdf
Wrote "E:\diff\report.pdf"
```

If the two PDFs are the same, the output will be the text "Same" on the console and no further action will be taken.

However, if the two PDFs are different (as in the example), the text "Wrote "E:\diff\report.pdf" will be output, *and*, due to the -s or --show option, the report will be shown in your PDF reader application.

#### **Report Show Field Content Option**

Use this option to specify whether to include the content of form fields in PDF reports. (This is done automatically for other report formats.)

The default is not to do this since it involves extra processing. To change the default, put lines like these in a configuration file:

```
[report]
show-field-content=true
```

#### **Report Summary Option**

When comparing two *folders* of PDFs, *comparepdfcmd* will output a summary of the results to the console. For an example, see the <u>Report Format Option</u>, page 20. When the summary is output to the console, the <u>Verbose Option</u> is ignored.

The summary can be sent to a file instead of the console (in which case the <u>Verbose Option</u> is respected), using the syntax -S *filename.csv* or --summary=*filename.csv*. For example:

```
E:\pdfs\>comparepdfcmd -v -S E:\summary.csv -f pdf -R E:\diff E:\pdfs\old E:\pdfs\new
Same
Wrote "E:\diff\two.pdf"
Wrote "E:\diff\three.pdf"
Wrote "E:\summary.csv"
```

## **Text Mode Options**

These options only apply when text mode comparisons are made, i.e., words and chars modes. (See also the <u>Comparison Mode Option</u>.) To control the highlighting colors used in these modes, see the <u>Color Options</u>' change-bar, delete, insert, and replace options, and the <u>Highlight Transparency Option</u>.

### **Change Bar Option**

This tells *comparepdfcmd* whether to show a red vertical change bar in pages' left margins whereever there are changes highlighted in the text. This option only applies to text mode comparisons and is ignored for appearance mode comparisons.

The default is true, i.e., to show change bars.

To change the default, i.e., to switch them off, put these lines in a configuration file:

[text] bar=**false** 

To change the change bar's color, see the <u>Color Options</u>.

#### **Column Tolerance Option**

This is an advanced option that should not normally need to be changed.

This option only applies to text mode comparisons that use the multicolumn algorithm. (See the <u>Comparison Mode Option</u> and the <u>Text Comparison Algorithm Option</u>.)

The default is 108. The allowed range is 10–300. If the default isn't satisfactory, try these values: 72, then 36, then 144, then 176.

To change the default, put these lines in a configuration file (with 108 changed to the default you want):

[text]
column-tolerance=108

#### **Ignore Hyphens Option**

By default, hyphens are treated like regular characters. If this option is used, hyphens are ignored. When hyphens are ignored it means that, for example, *one-two* is treated as *onetwo*. This works even if *two* is on the next line.

The default is not to ignore hyphens, but this can be overridden for any particular comparison on the command line using the syntax --ignore-hyphens. Or, you can make ignoring hyphens the default by adding these lines to a configuration file:

If you use this option, we recommend that you also normalize hyphens (which is the default behavior). See the <u>Normalize Hyphens Option</u>.

#### **Line Tolerance Option**

This is an advanced option that should not normally need to be changed.

To do sensible text comparisons *comparepdfcmd* needs to work out where the lines of text are on each page. This can be tricky when the lines are very close together. This option can be used to fine-tune how *comparepdfcmd* works out the lines.

The default is 10, but for PDFs that have lines very close together a setting of 9 or even 8 might produce better results. Changing the default can be done by adding lines like these to a configuration file:

[text]
line-tolerance=9

The minimum is 1 and the maximum is 27, but only 8, 9, and 10 are recommended.

#### **Normalize Hyphens Option**

This is an advanced option that should not normally need to be changed.

By default *comparepdfcmd* normalizes hyphens and a few other characters. This means that all kinds of hyphens (i.e., all hyphens in the Unicode "punctuation dash" character category, such as plain hyphens, minus signs, non-breaking hyphens, and en-dashes and em-dashes) are treated as if they were the same. In addition to normalizing hyphens, non-standard double quotes (Unicode code points U+0093 and U+0094 that are used in some PDF files) are considered to be identical to U+201C (") and U+201D (") for the purposes of comparison.\*

If you don't want hyphen normalization, you can turn it off on the command line using the syntax --no-normalize-hyphens. Or, you can make not normalizing hyphens the default by adding these lines to a configuration file:

[text]
normalize-hyphens=false

#### **Normalize Ligatures Option**

By default, a ligature is considered to be different from its spelled out form, for example,  $\mathbf{fi} \neq \mathbf{fi}$ , and  $\mathbf{fl} \neq \mathbf{fl}$ . If you want ligatures and their spelled out forms to be considered the same during comparisons, you can tell *comparepdfcmd* to normalize ligatures.

Ligature normalization can be turned on, on the command line, using the syntax --normalizeligatures. Or, you can make normalizing ligatures the default by adding these lines to a configuration file:

<sup>\*</sup>Note that *comparepdfcmd* always normalizes whitespace, that is, any kind of whitespace (space, non-breaking space, etc.), is treated as a space (i.e., as a word separator).

[text]
normalize-ligatures=true

### **Text Comparison Algorithm Option**

Three text mode comparison algorithms are provided. The default is the standard algorithm which gives good results in almost every case. For multicolumn PDFs using the multicolumn algorithm may produce better results. And in rare cases the special algorithm may be best.

You can change the default by adding lines like these to a configuration file:

[text]
algorithm=multicolumn

Naturally, if you prefer to use the special algorithm, you can set algorithm=special. And to restore the default either delete these lines or set algorithm=standard.

All three algorithms can be tuned using the <u>Line Tolerance Option</u>. And the The multicolumn algorithm can be further tuned using the <u>Column Tolerance Option</u>.

### **Verbose Option**

By default *comparepdfcmd* does not output anything to the console when comparing two PDF files. And when comparing two *folders* of PDFs *comparepdfcmd* outputs a summary to the console unless the <u>Report Summary Option</u> is used. (Of course, if errors occur, these are always reported, no matter what this option is set to.)

If you want *comparepdfcmd* to tell you the outcome of comparisons and what reports it has written, you can tell it to be verbose. This option can only be specified on the command line, using the syntax -v or --verbose.

### **Maximum Directory Depth Option**

When, comparing two folders of PDFs, by default, *comparepdfcmd* will look in subfolders (down to a depth of 1000 subfolders). However, it is possible to limit the depth to just the folders specified by using the -D or --maxdirdepth option with a value of 1.

You can change the default by adding lines like these to a configuration file:

```
[general]
maxdirdepth=1
```

You can set other depths, for example, use 2 to look in the folders specified and their immediate subfolders.

### The PDF Files or Folders to Compare

After any command line options have been specified, two PDF filenames (with paths if needed) or two folders *must* be given. If two PDF filenames are specified, these two PDFs will be compared. If two folders are specified, then *comparepdfcmd* will look inside each folder. For every PDF file it finds in both folders (and their subfolders, and so on), *comparepdfcmd* will compare those PDF files. For every PDF file that is in one folder but not the other, *comparepdfcmd* will report the nonexistent PDF as "missing". And any non-PDF files will be ignored.

See also the Maximum Directory Depth Option.

## Troubleshooting

Note that <u>www.qtrac.eu</u> and <u>www.diffpdf.com</u> are synonymous, as are support@qtrac.eu and support@qtrac.eu.

### **Invalid PDF Report Output**

In some very rare cases *comparepdfcmd*'s high quality scalable PDF output fails to work correctly. If this occurs, switch off scalability using the <u>Report Scalable PDF Option</u>.

### **Failed to Merge Error**

In some rare cases *comparepdfcmd* cannot produce a PDF report and instead outputs a "failed to merge" error message. *comparepdfcmd* produces PDF reports by copying each pair of pages from the original PDFs into a new PDF, but unfortunately, in these rare cases, copying the original pages doesn't work.

The solution is to output .PNG report files since these are created by in effect *photographing* the page pairs rather than copying them.

### **False Positives**

In some circumstances, *comparepdfcmd* will report differences that are not visible to the human eye. In some cases these are genuine differences (for example, two different kinds of hyphens that look the same), and in some cases these are false positives.

By default *comparepdfcmd* normalizes hyphens (i.e., treats all the different hyphen kinds as the same), but does not normalize ligatures (although it will if you tell it to). To control these, see the <u>Normalize Hyphens Option</u> and the <u>Normalize Ligatures Option</u>.

False positives are usually caused by very narrow inter-line spacing, or by text which mixes Latin text (e.g., English) with non-Latin text (e.g., Japanese), and in some cases by tables. In many cases these false positives can be eliminated—or minimized—by reducing the Line Tolerance Option. For example, try reducing the line tolerance from the default of 10 to 9 or 8, or try the special algorithm with these line tolerances (see the Text Comparison Algorithm Option).

## **Appendix: Color Names**

To use custom colors, see the <u>Color Options</u>. Note that the colors highlight, delete, insert, and replace will appear lighter than the colors shown here due to the <u>Highlight Transparency Option</u>. Note that very light colors may be difficult to see.



#### Dark0range Coral LightCoral 0range DarkRed Tomato **OrangeRed** Crimson Red HotPink DeepPink Pink LightPink PaleVioletRed Maroon WebMaroon MediumVioletRed VioletRed DarkMagenta Magenta Violet $\bigcap$ Plum Indigo 0rchid Medium0rchid DarkOrchid DarkViolet BlueViolet RebeccaPurple Purple WebPurple MediumPurple Thistle Snow2 Snow3 Snow4 Seashell2 Seashell3 Seashell4 AntiqueWhite1 AntiqueWhite2 AntiqueWhite3 AntiqueWhite4 Bisque2 Bisque3 Bisque4 PeachPuff2 PeachPuff3 PeachPuff4 NavajoWhite2 NavajoWhite3 NavajoWhite4 LemonChiffon2 LemonChiffon3 LemonChiffon4 Cornsilk2 Cornsilk3 Cornsilk4 Ivory2 Ivory3 $\left( \right)$ Ivory4 Honeydew2 Honeydew3 Honeydew4 LavenderBlush2 LavenderBlush4 LavenderBlush3 MistyRose2 MistyRose3 MistyRose4 Azure2 Azure3 $\square$ Azure4 SlateBlue1 SlateBlue2 SlateBlue3 SlateBlue4 RoyalBlue1 RoyalBlue2 RoyalBlue3 RoyalBlue4 Blue2 DodgerBlue2 DodgerBlue3 DodgerBlue4 SteelBlue1 SteelBlue2 SteelBlue3 SteelBlue4 DeepSkyBlue2 DeepSkyBlue3 DeepSkyBlue4 SkyBlue1 SkyBlue2 SkyBlue4 SkyBlue3 🔵 LightSkyBlue1 LightSkyBlue2 LightSkyBlue3 LightSkyBlue4 SlateGray1 SlateGray2 SlateGray3 SlateGray4 LightSteelBlue1 LightSteelBlue2 LightSteelBlue3 LightSteelBlue4 $\bigcap$ LightBlue1 LightBlue2 LightBlue3 LightBlue4 LightCyan2 LightCyan3 LightCyan4 PaleTurquoise1 PaleTurquoise2 PaleTurguoise3 PaleTurquoise4 CadetBlue1 CadetBlue2 CadetBlue3 CadetBlue4 Turquoise1 Turquoise2 Turquoise3 Turquoise4 Cyan2 DarkSlateGray1 DarkSlateGray2 DarkSlateGray3 Cyan3 DarkSlateGray4 Aquamarine2 Aquamarine4 DarkSeaGreen1 DarkSeaGreen2 DarkSeaGreen3 DarkSeaGreen4 SeaGreen1

| SeaGreen2       | SeaGreen3       | PaleGreen1      | PaleGreen3      |
|-----------------|-----------------|-----------------|-----------------|
| PaleGreen4      | SpringGreen2    | SpringGreen3    | SpringGreen4    |
| Green2          | Green3          | Green4          | Chartreuse2     |
| Chartreuse3     | Chartreuse4     | OliveDrab1      | OliveDrab2      |
| ) OliveDrab4    | DarkOliveGreen1 | DarkOliveGreen2 | DarkOliveGreen3 |
| DarkOliveGreen4 | Khaki1          | Khaki2          | Khaki3          |
| ) Khaki4        | LightGoldenrod1 | LightGoldenrod2 | LightGoldenrod3 |
| LightGoldenrod4 | LightYellow2    | LightYellow3    | LightYellow4    |
| Yellow2         | Yellow3         | Yellow4         | Gold2           |
| Gold3           | Gold4           | Goldenrod1      | Goldenrod2      |
| Goldenrod3      | Goldenrod4      | DarkGoldenrod1  | DarkGoldenrod2  |
| DarkGoldenrod3  | DarkGoldenrod4  | RosyBrown1      | RosyBrown2      |
| ) RosyBrown3    | RosyBrown4      | IndianRed1      | IndianRed2      |
| IndianRed3      | IndianRed4      | Sienna1         | Sienna2         |
| )Sienna3        | Sienna4         | Burlywood1      | Burlywood2      |
| )Burlywood3     | Burlywood4      | Wheat1          | Wheat2          |
| )Wheat3         | Wheat4          | Tan1            | Tan2            |
| Tan4            | Chocolate1      | Chocolate2      | Chocolate3      |
| Firebrick1      | Firebrick2      | Firebrick3      | Firebrick4      |
| Brown1          | Brown2          | Brown3          | Brown4          |
| )Salmon1        | Salmon2         | Salmon3         | Salmon4         |
| )LightSalmon2   | LightSalmon3    | LightSalmon4    | 0range2         |
| 0range3         | Orange4         | Dark0range1     | Dark0range2     |
| DarkOrange3     | DarkOrange4     | Coral1          | Coral2          |
| Coral3          | Coral4          | Tomato2         | Tomato3         |
| Tomato4         | 0rangeRed2      | 0rangeRed3      | OrangeRed4      |
| Red2            | Red3            | DeepPink2       | DeepPink3       |
| DeepPink4       | HotPink1        | HotPink2        | HotPink3        |
| )HotPink4       | Pink1           | Pink2           | Pink3           |
| ) Pink4         | LightPink1      | LightPink2      | LightPink3      |
| )LightPink4     | PaleVioletRed1  | PaleVioletRed2  | PaleVioletRed3  |
| PaleVioletRed4  | Maroon1         | Maroon2         | Maroon3         |
| Maroon4         | VioletRed1      | VioletRed2      | VioletRed3      |
| VioletRed4      | Magenta2        | Magenta3        | Orchid1         |
| )Orchid2        | Orchid3         | Orchid4         | Plum1           |
|                 |                 |                 |                 |

