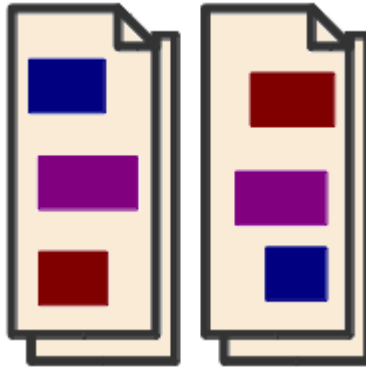


diffpdfc (windows console edition) v 5

Mark Summerfield

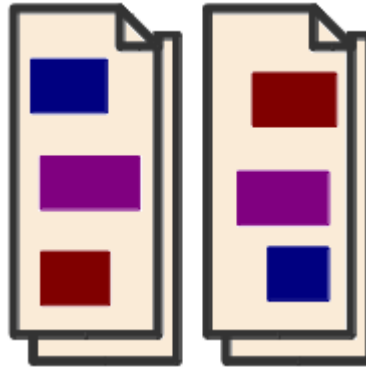
Copyright © Qtrac Ltd. 2013-17. All Rights Reserved.



diffpdfc (windows console edition) v 5

Mark Summerfield

Copyright © Qtrac Ltd. 2013-17. All Rights Reserved.



Contents

Introduction	...	4
Installing	...	5
System Requirements	...	5
Obtaining and Installing <i>diffpdfc</i>	...	5
How to Register a License Key	...	6
Permanently Adding <i>diffpdfc</i> to the Path	...	7
Uninstalling	...	8
Upgrading	...	8
Licensing	...	8
Support	...	9
Examples	...	9
Running <i>diffpdfc</i> Within Other Programs	...	10
How Best to Compare PDFs	...	11
Option Values	...	11
Command Line Options	...	12
Accuracy	...	12
Comparison Mode	...	13
Configuration File (Ad Hoc)	...	13
Cores	...	14
First (Left) PDF File or Directory	...	14
First (Left) PDF File's Pages	...	14

Format	...	15
Help	...	15
Highlighting Mode	...	15
Ignore Hyphens	...	15
License	...	16
Manual	...	16
Normalize Hyphens	...	16
Normalize Ligatures	...	17
Pairs of Pages to Compare as a Unit	...	17
Passwords	...	17
Quiet	...	18
Register a License Key	...	18
Reports	...	18
Report Files	...	18
Report New Renderer	...	19
Report Pretty	...	20
Report Scale	...	20
Report Directory	...	20
Second (Right) PDF File or Directory	...	21
Second (Right) PDF File's Pages	...	21
Summary	...	21
Verbose	...	22
Version	...	22
Configuration File Options	...	22
accuracy	...	22
bar	...	22
bar.color	...	23
bar.extend	...	23
bar.offset	...	23
bar.width	...	23
column.tolerance	...	23
compare	...	23
comparison.algorithm	...	24
cores	...	25
corners	...	25
corner.bottom.left	...	25
corner.bottom.right	...	25
corner.color	...	25
corner.style	...	26
corner.top.left	...	26
corner.top.right	...	26
format	...	26

Register Window “Invalid key code” Error	34
False Positives	35

Introduction

The *diffpdfc* (*windows console edition*) application is a command line tool for comparing two PDF files. The tool can report the differences in several ways—for example, by showing the differences visually or by outputting a file identifying words that have been inserted, deleted, or replaced.

The *diffpdfc* program can be used by archivers, engineers, journalists, marketers, package designers, publishers, researchers, and others who want to compare one version of a document, report, book, or label with another. It can also be used by software testers who want to verify that software generated PDF files have the correct text or appearance and no unexpected differences. Numerous other uses are possible—for example, a translator might compare two PDFs to see what has changed so that they need only update their translation for the changed parts.

Comparing PDF files is not necessarily straightforward: for example, do we consider two PDFs the same if they have the same words in the same order—but laid out differently (e.g., due to slightly different margins)? The *diffpdfc* program uses the same comparison engine as the GUI (graphical user interface) *DiffPDF* (*Windows edition*) application—see www.qtrac.eu/diffpdf.html. This engine provides three comparison modes: character by character, word by word, and appearance. The first two (text modes) are concerned only with the characters or words and ignore the layout; the appearance mode compares the appearance of each pair of pages and so detects differences in layout, differences in graphics, and differences in fonts.

The most common and important options for controlling *diffpdfc*'s configuration can be specified on the command line. The way *diffpdfc* gets its configuration is as follows:

1. Start with a configuration where every option is set to a sensible default where possible
2. Read options from file %USERPROFILE%\diffpdfc.dpc if it exists, and update the configuration accordingly
3. Read options from file diffpdfc.dpc (in the current directory) if it exists, and update the configuration accordingly
4. If the `-C` or `--config` command line option is given, read options from the file it specifies, and update the configuration accordingly
5. Read any options given on the command line (apart from `-C` or `--config`), and update the configuration accordingly

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

When the program finishes running it returns a status value to the operating system (which can be used in shell scripts/batch files). And if the `--quiet` option has *not* been used, it also outputs a message indicating the results.

<i>Status</i>	<i>Message</i>	<i>Note</i>
0	“same text” or “same appearance”	Depends on the comparison mode. No message if --quiet is used.
1	“invalid license key” or “expired trial license key”	Register a valid full license key.
2	<i>varies</i>	A usage error message, i.e., an incorrect command line option.
3	“appearance different”	No message if --quiet is used.
4	“text different”	No message if --quiet is used.
5	<i>varies</i>	An error message if the comparison is canceled using Ctrl+C.
6	<i>varies</i>	An error message for an “expected” error (e.g., an invalid/unreadable PDF).
7	<i>varies</i>	An error message for an “unexpected” error.

Installing

System Requirements

diffpdfc can be installed on all modern versions of Windows, providing the machine uses an x86-compatible processor (i.e., most desktop and laptop computers). However, 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. *diffpdfc* 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.

Obtaining and Installing *diffpdfc*

diffpdfc is provided with both 32-bit (e.g., *diffpdfc-5.6.0-win32.msi*) and 64-bit (e.g., *diffpdfc-5.6.0-amd64.msi*) Microsoft Installer files. The 32-bit version will run on modern 32-bit *and* 64-bit versions of Windows. The 64-bit version will run only on a 64-bit Windows, but does not have the 32-bit 2GB memory limit. In most cases the 32-bit version is fine to use, but doing Appearance mode comparisons on huge PDFs may need a 64-bit version (and lots of memory). The installers (and equivalent .zip files) can be downloaded from www.qtrac.eu/diffpdfc.html. Once an installer has been downloaded, double-click it to start the standard Windows installer program and follow the on-screen instructions.*

Once the software has been installed, a valid license key must be registered for it to work. (This may not be necessary if upgrading from one minor version to another—see [Upgrading](#).)

* If both *diffpdfc* and *DiffPDF* are installed, they should be installed into separate directories. Their default installation directories are different, so this will automatically happen if the defaults are accepted.

Tiered price discounts are applied if you buy at least 10 license keys. When you're ready to buy, click the Register window's Buy Now button or visit www.qtrac.eu/buy?app=diffpdfc either of which will take your web browser to the secure MyCommerce purchase page which shows the price tiers.

How to Register a License Key

Make sure that the person logged in is the person who is to use *diffpdfc*. Then, run the program from the command line with the `--register` option:

```
C:\Users\mark\>diffpdfc --register
```

3.1.1

From version 3.1.1 the installer automatically tries to add *diffpdfc* to the PATH. If this fails, or if using an older version, you must either add *diffpdfc* to the PATH manually (see [Permanently Adding diffpdfc to the Path](#)), or specify the full path.

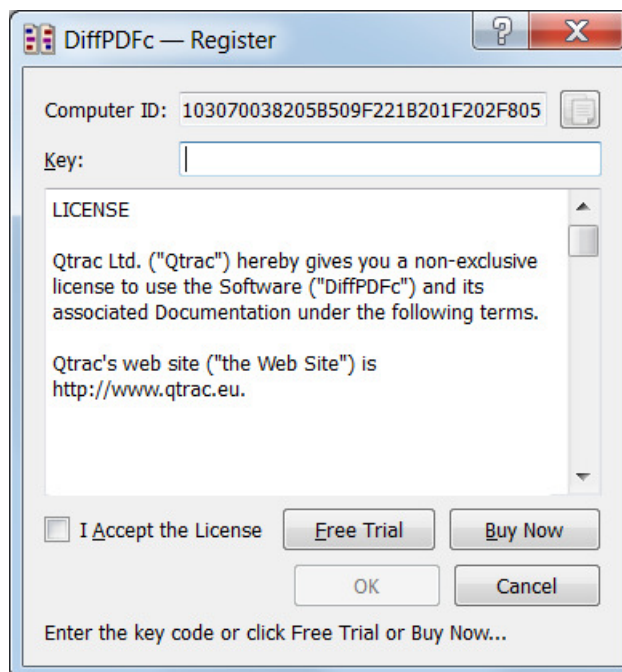
Here we assume that *diffpdfc.exe* is in a particular directory; if it isn't in the one shown, specify the path you installed it into instead:

≤3.1.0

```
C:\Users\mark\>"C:\Program Files (x86)\diffpdfc\diffpdfc" --register
```

Running *diffpdfc* with the `--register` option will pop-up the Register window:

4.0



To obtain a free trial license key click the Free Trial button—this should open your web browser at the free trial page. Alternatively, if you want to buy a full license key, click the Buy Now button—this should open your web browser at the MyCommerce ecommerce website where you can make the payment. If your security settings prevent *diffpdfc* from accessing the internet, these buttons won't work. In this case, open your web browser and visit www.qtrac.eu/diffpdfc.html and obtain your license key from there.

Free trials are anonymous, but you are only allowed one trial per application per major version per computer. To get a free trial license key you must enter your *Computer ID* into

the web form. The Computer ID is shown in the Register window—simply click the “Copy” icon beside it to copy it to the clipboard, then paste it into the web form, and then click the Issue Trial License Key button. The trial license key will then appear: copy and paste the key into the Register window’s Key line edit and then press OK.

If you choose to buy, once payment has been accepted you will be shown the MyCommerce payment confirmation page. This page will show the text “Loading...” and then this will be replaced by a web page to visit to get your license key(s) from. You can go directly to the “Get Key” page at www.qtrac.eu/getkey but it is even more convenient to click the link in the confirmation email you are sent, since this includes your *Order ID*, which is needed to obtain your license key(s). To get a license enter the Order ID (unless it is already displayed) and the Computer ID (as shown in the Register window), and then click the web page’s Get License Key button. The license key will then appear—copy and paste it into the Register window’s Key line edit and then press OK. This process should be done on each computer for which a license has been purchased—or once if you are using a server (see the license).

For a full license key you only need to do this once (per program, per major version, per computer). For a trial license key you will need to register the trial license key, and later on, a full license if your trial license expires and you buy a full license. To register, just run the program using the `--register` option, and enter the full license key from the www.qtrac.eu/getkey page. Because the key is long it is best to copy and paste it if possible.

Each license key is tied to the computer it is acquired on and is valid for the particular program it was bought for at its major version. For example, *diffpdfc*, all 5.x versions, Windows Console on the machine the license was acquired on. Note that license keys are valid for both 32-bit and 64-bit editions so either can be used.

If you want to use a different product, such as the graphical user interface *DiffPDF (Windows edition)*, or to upgrade to a new major version (e.g., 6.x), or to use the program on a different computer, you must buy a license for that program, major version, and computer. (However, on request, we will provide a free upgrade to anyone who has bought *diffpdfc* within 60 days of a new major version being released. Also, one license transfer is allowed within the first year—see the license for details.)

Permanently Adding *diffpdfc* to the Path

From version 3.1.1, the installer automatically tries to permanently add *diffpdfc* to the PATH and no further action is required.

3.1.1

If the installer is unable to add *diffpdfc* to the PATH (e.g., due to lack of permissions), or if you are using version 3.1.0 or older, you can take one of three different approaches.

≤3.1.0

The first approach is simple but inconvenient: enter the full path each time. For example:

```
C:\Users\mark\>"C:\Program Files (x86)\diffpdfc\diffpdfc" file1.pdf file2.pdf
```

The second approach is slightly more convenient: add *diffpdfc* to the PATH once per session:

```
C:\Users\mark\>set path=%path%;"C:\Program Files (x86)\diffpdfc"
```

Here we’ve assumed a particular directory where *diffpdfc* is usually installed to: naturally, you should change this to the actual directory if it is different from the one shown here.

Now that *diffpdfc* has been added to the PATH it can be run without specifying its path. However, the `set path=...` command must be issued every time you open a Windows console.

The third and most convenient approach is to permanently add *diffpdfc* to the PATH. However, on some versions of Windows, adding to the PATH requires Administrator permissions.

To add *diffpdfc* to the PATH, click Start (the Windows logo), then click Control Panel. Use the Search to find “environment variables” and click through to the dialog that shows them. Click the PATH environment variable and add the following text (including the leading semi-colon and the double-quotes exactly as shown) at the *end* of the existing path:

```
;"C:\Program Files (x86)\diffpdfc"
```

Note that the path entered here must be the one that *diffpdfc* was actually installed to, so enter the actual path you installed to if it is different.

Uninstalling

If you want to uninstall the program, you can do so in the usual way by going to the Windows Control Panel, choosing Uninstall Program, and then choosing *diffpdfc* and clicking Uninstall.

Upgrading

To upgrade from one minor version to another—say, from version 5.1.2 to 5.3.1—do the following. First, uninstall the old version as described in the preceding subsection ([Uninstalling](#)). Second, install the new version as described earlier ([Obtaining and Installing diffpdfc](#)). There is no need to re-register the license key—the new version will detect and use the one that is already registered.

To upgrade from one major version to another—say, from version 4.2.3 to 5.5, uninstall the old version and then install the new version as described above. Note though, that license keys are only valid for particular major versions, for example, for all 5.x versions or all 6.x versions, so you will need to register the new version with a new trial or full license key as described above ([Register a License Key](#)).

Licensing

The *diffpdfc* program can be freely downloaded, but won't work unless a paid for full license key or a time-limited free trial license key has been registered with it. A full license is valid for a major version and all that version's minor and patch releases. For example, a full license for *diffpdfc* will work with *diffpdfc* 5.0, 5.1, 5.1.2, 5.2, and so on for all 5.x.y versions—but not for *DiffPDF (Windows edition)* or *diffpdfc* 6.x or later.

Trial license keys are free, although only one per product per major version per computer is allowed, and they are time limited. Full license keys must be paid for—see www.qtrac.eu/diffpdfc.html for prices and to buy. The license text itself is shown when a licence key is registered and can be viewed at any time by running the program with the `--licence` command line option.

The same executable is used for the trial and full license keys, so to switch from a trial to a fully licensed product simply buy a full license key and register the new full key with the

existing executable by running the program with the `--register` option. However, if a minor or patch release has been issued in the meantime, feel free to download and use it—or any subsequent minor or patch release—since the key will remain valid.

Licenses must be bought through the web site, with discounts available if 10 or more licenses (for a single product, i.e., *diffpdfc*), are bought at the same time.

Support

It is worth reading this manual once through to become familiar with what *diffpdfc* can do. (For example, to learn about the `pairs` setting.) If the manual doesn't answer your question, then we'll happily try to answer.

If you believe you have found a bug, first check which version you are using:

```
C:\Users\mark>diffpdfc -v
```

Then visit *diffpdfc*'s web site, www.qtrac.eu/diffpdfc.html, to see if there is a newer version. If a new version is available, upgrade to it (see [Upgrading](#)), and then see if you still have the bug. If the problem persists, contact us and we will try to help.

Note that changes between versions are listed at www.qtrac.eu/diffpdfc-changes.html.

Email support is formally available for the duration of a trial license key and for the first 60 days of a full license key. However, in practice we will respond to a reasonable number of emails for a fully licensed product indefinitely. The support email address is `support@qtrac.eu` — be sure to contact us using the email address you used to obtain your license key, and to tell us what operating system you are using (e.g., “Windows 7”) and especially which version of *diffpdfc* you are using.

Examples

These examples assume that *diffpdfc* has had a full or unexpired trial license key registered (see [How to Register a License Key](#)), and that *diffpdfc* is in the PATH (see [Permanently Adding diffpdfc to the Path](#)). The examples also assume a home directory of `C:\Users\mark`; naturally you must adjust this to the directory you want to use.

```
C:\Users\mark>diffpdfc file1.pdf file2.pdf
```

This will return a value to the operating system and output a message indicating whether the PDFs have the same text or not.

```
C:\Users\mark>diffpdfc --quiet -r report-a.jsn a1.pdf a2.pdf
```

```
C:\Users\mark>diffpdfc --quiet -r report-b.jsn b1.pdf b2.pdf
```

This is similar to the previous example, but produces no messages. This is useful when *diffpdfc* is run within a shell script or batch program. For each run, only if the PDFs have differences is the report written. (See [Quiet](#) and [Reports](#).)

```
C:\Users\mark>diffpdfc -r report.csv -r report.xml -r report.pdf -Hc c1.pdf c2.pdf
```

If the PDFs have differences in their text this will produce two reports on text differences, one in CSV format and the other in XML. It will also produce a PDF showing the differences visually. Thanks to the `-H` (`--highlight`) option being set to `c` (changes), all the reports will show insertions, deletions, and replacements, rather than simply indicating differences. (See [Highlighting Mode](#).)

Note that since no paths have been given all the reports shown above will be output to the current directory, i.e., `C:\Users\mark\report.csv` and so on, or to the `report.path` directory if one has been specified in a configuration file. If you want them to go elsewhere, give the report filenames a path, e.g., `-r C:\reports\report.xml`.

It is possible to set preferred defaults so that they don't need to be entered on the command line every time. This is done by creating a plain text file called `diffpdfc.dpc` in the `%USERPROFILE%` directory, and entering the new defaults in this file. Here is an example:

```
# diffpdfc.dpc
highlight: changes
bar.width: 1.5
report.decimals: 3
```

This file will change three defaults: the highlighting mode will default to `changes` rather than `plain`; the vertical change bar width will be 1.5 points wide rather than the default of 1 point wide; and text format reports will output numbers (e.g., coordinates) using three decimal places rather than the default of 2. (This example configuration file is in *diffpdfc*'s `doc` directory, usually, `C:\Program Files (x86)\diffpdfc\doc\diffpdfc.dpc`. Simply copy it to the `%USERPROFILE%` directory or the local directory where you want it to apply, and edit it to suit your requirements.)

Incidentally, blank lines and lines beginning with `#` are ignored in configuration files.

```
C:\Users\mark\>diffpdfc -Hc -f .pdf -R reportdir olddir newdir
```

This run will compare each PDF file in the `olddir` directory with a PDF file in the `newdir` directory that has the same name (if there is one). For each pair of PDFs that differ, it produces a PDF report (due to the `-f .pdf` option; see [Format](#)) with the same name as the compared PDFs and saved into the `reportdir` directory (see [Report Directory](#); see also [report.pdfprefix](#)).

4.2

4.3

It is also possible to get a summary using the `-S` or `--summary` option; see [Summary](#).

Running *diffpdfc* Within Other Programs

It is possible to run *diffpdfc* from within another program. A tiny Visual Basic example function is shown in the file `rundiffpdfc.bas` which is in *diffpdfc*'s `doc` directory, usually, `C:\Program Files (x86)\diffpdfc\doc\rundiffpdfc.bas`.

How Best to Compare PDFs

We saw in the [Examples](#) above how to use *diffpdfc*. But that didn't tell us the best choices to make when performing a comparison.

By default, *diffpdfc* will compare PDFs using words mode with a [pairs](#) setting of 1. This will produce the best results in a wide range of situations. If we want to compare text written in a CJK language, we might be better off using chars (characters) mode. And if we want to compare colors, fonts, diagrams, images, or text layout, we should use appearance mode. (See [Comparison Mode](#).)

If we want to see changes that span page boundaries we can set [pairs](#) to the page count of the longest PDF. One disadvantage of doing this is that comparisons take much longer because it is much more computationally expensive when [pairs](#) is greater than one. Another disadvantage is that for some complex PDFs, comparing across page boundaries can produce unsatisfactory results.

But what if our PDFs have different numbers of pages or inserted or deleted pages? There are two approaches we can take. One is to use page ranges—these can be set using [First \(Left\) PDF File's Pages](#) and [Second \(Right\) PDF File's Pages](#). The advantage of using page ranges is that we can specify exactly which pages to match against each other and ignore any of the inserted pages that we don't need to consider.

Another approach is to set the [pairs](#) to the page count of the longest PDF. This pushes the burden of detecting inserted or deleted pages onto *diffpdfc*—at the price of taking a lot longer. We recommend using page ranges wherever possible, but using [pairs](#) may be necessary in some cases. If we use [pairs](#) then we may be able to avoid having inserted or deleted pages appear in our reports. See [show_ins_del_pages](#) for how to do this and for an explanation of why it might be useful. 4.0

The default comparison algorithm should produce good results for single column PDFs, but for multi-column PDFs it is usually much better to use the multi-column comparison algorithm: see [comparison.algorithm](#). 5.5

It is also possible to tweak the comparison algorithm itself by setting the [line.tolerance](#), and for the multi-column algorithm, [column.tolerance](#). 5.5

Option Values

Here are some notes on the values that can be given to command line options and to configuration file options.

Boolean (true/false) values

These are taken to be *true* if given as 1 or t or true or y or yes or on, and *false* otherwise.

Color values

These may be specified using an X11 color name (e.g., red or orange), or using an R,G,B triple where each component is 0-255 and comma separated (e.g., 255,102,0), or using an HTML-style six digit hex color (e.g., #FFCC00). (A bug prevented R,G,B triples validating correction in versions 3.0.y.) 3.1

Length, offset, height, and width values

These are assumed to be in points ($\frac{1}{72}$ "), unless another unit is specified. For example, 36 is 36 points, 12.7mm is 12.7 millimeters, and 0.5" and 0.5in are both 0.5 inches.

Line style values

Three line styles are supported: dashed (-----), dotted (.....), and solid (———).

Page number values

All page numbers are counted from 1 (the first page)—these are sometimes called logical page numbers. Documents' own page numbers (e.g., i, ii, ..., 1, 2, 3, ...) are not used.

Blank lines in configuration files are ignored—and so are lines beginning with #.

Multi-part option names (e.g., report.decimals or highlight.insert.color) may be specified in configuration files using periods or hyphens (e.g., report-decimals or highlight-insert-color).

Command line options must be given in Unix-style (-w or --words) *not* Windows-style (/foo), and multi-part names must always be given using hyphens (e.g., --no-report-pretty). Command line options that take values may be specified in all the standard ways, e.g., --scale=150, --scale 150, -s150, -s=150, -s 150.

Note that some options only apply to appearance mode comparisons and some only to words or chars (characters) mode comparisons. In such cases, the applicable mode is indicated in the margin beside the option's name. (See [Comparison Mode](#).)

Command Line Options

There are many ways to customize how *diffpdfc* performs the comparison and how it reports any differences. Here we list every command line option with cross-references for those that can be set in configuration files. The options are listed in alphabetical order for ease of reference.

See [Option Values](#) for how to specify option values (e.g., colors, lengths, and line styles).

All the examples of command line usage assume that *diffpdfc* is in the PATH (see [Permanently Adding *diffpdfc* to the Path](#)).

Accuracy

```
C:\Users\mark>diffpdfc -A 98 ...
```

```
C:\Users\mark>diffpdfc --accuracy 98 ...
```

How accurate (as a percentage) appearance mode comparisons should be. The default is 100%.

For appearance mode comparisons, each pair of corresponding squares is compared. If accuracy is set to 100%, the squares are considered the same only if every pixel is identical. This is the default behavior.

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”.

The accuracy can also be set in a configuration file; see [accuracy](#).

See also [Comparison Mode](#) and [square.size](#).

Comparison Mode

```
C:\Users\mark\>diffpdfc -a ...
C:\Users\mark\>diffpdfc --appearance ...
C:\Users\mark\>diffpdfc -c ...
C:\Users\mark\>diffpdfc --chars ...
C:\Users\mark\>diffpdfc -w ...
C:\Users\mark\>diffpdfc --words ...
```

The comparison mode to use.

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 languages), character by character comparisons (chars mode), may work better. 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.

The default is words mode. To set the comparison mode on the command line use *one* of these options: `-a` or `--appearance` for appearance mode; `-c` or `--chars` for characters mode; or `-w` or `--words` (or no comparison mode option at all) for words mode.

Of the text comparison modes, chars mode is much slower than words mode, but is also more precise, especially when highlighting changes. Both text comparison modes can produce false positives—see [line.tolerance](#), [column.tolerance](#), and [comparison.algorithm](#) for details and how to fine tune to avoid these.

When appearance mode is used the pairs setting is ignored (pages are compared one pair at a time regardless), and differences are highlighted using plain highlighting. (See also [Accuracy](#).)

The comparison mode can also be set in a configuration file; see [compare](#).

Configuration File (Ad Hoc)

```
C:\Users\mark\>diffpdfc -C file ...
C:\Users\mark\>diffpdfc --config file ...
```

This option accepts the name of a configuration file given on the command line.

See the [Introduction](#) for details of how the program’s configuration is built up.

Cores

```
C:\Users\mark>diffpdfc --cores N
```

This option is used to set the number of cores that *diffpdfc* should use. The number *N* should be the number of cores that the machine has.

This option should not normally be needed, since *diffpdfc* automatically detects how many cores the machine has, and spreads the work over all of them so as to minimize the runtime.

When running *diffpdfc* from within another process or on some Windows server operating systems, it is possible that the core auto-detection will fail. This option allows the number of cores to be specified manually when auto-detection fails.

The cores can also be set in a configuration file; see [cores](#).

First (Left) PDF File or Directory

```
C:\Users\mark>diffpdfc file1.pdf file2.pdf
```

```
C:\Users\mark>diffpdfc oldpdfsdir newpdfsdir
```

The first (left) PDF file to compare (*file1.pdf*), or the first directory of PDF files to compare (*oldpdfsdir*).

4.2

There is no default value and a value is required.

This filename can be set in a configuration file (see [pdf1.filename](#)); but it is more conventional to pass it on the command line after all the options. If a filename is specified then the second (right) PDF file or directory must also be a filename. If a directory is specified then the second must also be a directory.

If a directory is specified, *diffpdfc* will only compare PDF files found in the first directory (and its subdirectories) with those of the same name found in the second directory (and its subdirectories). PDF files that are in one directory but not the other, and non-PDF files, are safely ignored.

See also [Second \(Right\) PDF File or Directory](#), [Format](#), [Report Directory](#), and [report.pdfprefix](#).

First (Left) PDF File's Pages

```
C:\Users\mark>diffpdfc --pages1 1-5,7,11-18,19 ...
```

Which pages in the first (left) PDF file to include in the comparison. (This option is allowed when comparing directories of PDFs but rarely makes sense in that context.)

The default is no value which means that every page is compared. (For example, this is the same as using `--pages1=1-20` for a 20 page PDF.)

This can also be set in a configuration file; see [pages1](#).

See also [Second \(Right\) PDF File's Pages](#).

Format

```
C:\Users\mark>diffpdfc -f .pdf ...
```

```
C:\Users\mark>diffpdfc --format .pdf ...
```

This option causes *diffpdfc* to write reports in the given format. The reports are written to the `--reportdir` directory and the name of each report is that of the compared file with the given format suffix. Use this option multiple times to get reports in more than one format on the same run.

This option only makes sense when comparing two directories of PDF files and will only work if a report directory has been specified. See [First \(Left\) PDF File or Directory](#), [Second \(Right\) PDF File or Directory](#), [Report Directory](#), and [report.pdfprefix](#).

This can also be set in a configuration file; see [format](#).

Help

```
C:\Users\mark>diffpdfc -h
```

```
C:\Users\mark>diffpdfc --help
```

This option causes *diffpdfc* to print a summary of its command line options, and then exit. (The help text also reports how many cores *diffpdfc* has detected. If this is fewer than the actual number, use the [Cores](#) option to set the correct value.)

Highlighting Mode

```
C:\Users\mark>diffpdfc -Hc ...
```

```
C:\Users\mark>diffpdfc --highlight changes ...
```

```
C:\Users\mark>diffpdfc -Hp ...
```

```
C:\Users\mark>diffpdfc --highlight plain ...
```

What kind of highlighting to provide when differences are shown visually, and how differences are reported.

The default is `plain` which simply highlights differences rather like a highlighter pen for visual output and indicates different rectangles and text for reports. The alternative is `changes` which also works like a highlighter pen, only it uses different colors for insertions, deletions, and replacements for visual output and indicates not only different rectangles and text, but also the kind of change for reports.

The highlighting mode can also be set in a configuration file, as can the highlighting colors, opacity, and extents. See [highlight](#).

Ignore Hyphens

```
C:\Users\mark>diffpdfc --ignore-hyphens ...
```


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.

This option was introduced in version 5.7.8. It can also be set in a configuration file; see [ignore.hyphens](#).

If you use this option, we recommend that you also normalize hyphens (which is the default behavior). See [Normalize Hyphens](#).

License

```
C:\Users\mark>diffpdfc --license
```

```
C:\Users\mark>diffpdfc --licence
```

This option causes *diffpdfc* to print the license and then exit. (The license is also shown in the Register window and must be accepted to register a trial or full license key.)

This option will work even if the program hasn't yet had a license key registered.

Manual

```
C:\Users\mark>diffpdfc -m
```

```
C:\Users\mark>diffpdfc --manual
```

This option causes *diffpdfc* to open this manual in a PDF viewer (if one is installed), and then exit.

This option will work even if the program hasn't yet had a license key registered.

Normalize Hyphens

```
C:\Users\mark>diffpdfc --no-normalize-hyphens ...
```

By default, hyphens are normalized; this command line option can be used to turn normalization off.

When normalize hyphens is in force all the different kinds of hyphens (i.e., all hyphens in the Unicode “punctuation dash” character category) are treated as if they were the same.

When this option is in force, in addition to normalizing hyphens, non-standard double quotes (Unicode code points U+0093 and U+0094) are considered to be identical to U+201C (“) and U+201D (”) for the purposes of comparison.

This option can also be set in a configuration file; see [normalize.hyphens](#).

Note that whitespace is always normalized, that is, any kind of whitespace (space, non-breaking space, etc.), is treated as a space (i.e., as a word separator).

Hyphens can also be ignored; see [Ignore Hyphens](#).

Normalize Ligatures

```
C:\Users\mark>diffpdfc --normalize-ligatures ...
```

By default, a ligature is considered to be different from its spelled out form, for example, **fi** ≠ **fi**, and **fl** ≠ **fl**. This command line option can be used to turn normalization on, in which case ligatures and their spelled out forms are considered to be the same.

This option can also be set in a configuration file; see [normalize.ligatures](#).

Char-
acters &
Words
modes

5.5

Pairs of Pages to Compare as a Unit

```
C:\Users\mark>diffpdfc -p integer
```

```
C:\Users\mark>diffpdfc --pairs integer ...
```

How many pairs of pages to compare as a unit.

The default is 1 which means that both PDFs' page 1s are compared then their page 2s and so on. A value of 2 means that two pairs of pages are compared as a unit, i.e., both PDFs' page 1s and 2s as if they were a single long page, then their page 3s and 4s, then 5s and 6s and so on. Similarly a value of 3 means that three pairs are compared as a unit, pages 1, 2, and 3; then 4, 5, and 6, and so on. The advantage of comparing two or more pairs at a time is that differences that span page boundaries are handled more gracefully.

Comparing more than one pair of pages as a unit is computationally expensive. Furthermore, chars mode is typically 5× more expensive than words mode. Nonetheless, from version 3.6.0 the maximum has been increased to 1000, and from 5.2.0 to 5000.

If you have a large PDF (e.g., a book or long report), we recommend setting pairs to the number of pages and letting the program run for ten minutes. After this time the percentage done should be a good guide to how long it will take to complete. And you can always cancel (e.g., press Ctrl+C), and then try again with a lower number of pairs (e.g., one chapter's worth of pages). Be aware that in some cases using a pairs value greater than 1 can produce unsatisfactory results.

(Note that this option is concerned with how many pages are compared as a unit; this is a separate matter from concurrency. The *diffpdfc* program is smart enough to automatically utilize all the processors/cores available whenever concurrent processing is possible and leads to performance improvements.)

When the comparison mode is appearance, pages are compared one pair at a time and this setting is ignored.

See also [show_ins_del_pages](#) for some additional information relevant when the PDFs have inserted or deleted pages.

This option can also be set in a configuration file; see [pairs](#).

Char-
acters &
Words
modes

3.6

5.2

4.0

Passwords

```
C:\Users\mark>diffpdfc --password1=abc123 ...
```

```
C:\Users\mark>diffpdfc --password2=xyz987 ...
```

```
C:\Users\mark>diffpdfc --password1=abc123 --password2=xyz987...
```

5.6

The password for one or both PDFs if needed to open them.

Most PDFs are not password protected and can be read and processed by *diffpdfc* just by giving their filename. However, some PDFs are password protected and for these it is possible to specify the password on the command line for either or both PDFs as required.

Note that using these options only makes sense when comparing two PDFs directly—not when comparing two directories of PDFs.

Quiet

```
C:\Users\mark\>diffpdfc -q ...
C:\Users\mark\>diffpdfc --quiet ...
```

Whether to suppress non-error messages.

The default behavior is for *diffpdfc* to print on the console which configuration files it has read, what the outcome of the comparison was, and what files it has written, as well as any error messages. All output except error messages can be suppressed by setting this option.

This option can also be set in a configuration file; see [quiet](#). (See also [verbose](#).)

Register a License Key

```
C:\Users\mark\>diffpdfc --register
```

This option causes *diffpdfc* to pop up the Register window so that you can obtain and register a license key. (See [How to Register a License Key](#) for details.)

Naturally, this option will work even if the program hasn't yet had a license key registered.

Reports

Visual and textual reports of PDF differences can be output in various formats using the `-r` or `--report` command line option.

Alternatively, when comparing two directories of PDFs it is possible to create reports that have the same name as the original PDFs and which are saved to a separate directory. See the [Format](#), and [Report Directory](#) command line options and the [format](#), [reportdir](#), and [report.pdfprefix](#) configuration file options.

In addition to specifying the reports to be output it is also possible to influence some aspects of the output using the [Report Pretty](#), [Report Scale](#) and [Report New Renderer](#) command line options, and the [report.decimals](#), [report.path](#), [report.path_in_title](#), [report.pretty](#), [report.scale](#), and [report.new.renderer](#) configuration file options.

4.3

Report Files

```
C:\Users\mark\>diffpdfc -r report.jsn -r report.csv -r report.pdf ...
C:\Users\mark\>diffpdfc --report report-.png --report report.xml ...
```

The names of one or more files to report differences to.

The supported text formats are CSV (Comma-Separated Values, suitable for import into a spreadsheet), JSON (JavaScript Object Notation), and XML (eXtensible Markup Language). The supported visual formats are PDF (Portable Document Format) and PNG (Portable Network Graphics), in the latter case specified as a file *pattern* rather than a file name. A summary report that just provides basic information with no details is also available in LOG (plain text log file) format, e.g., `-r summary.log`.

4.1.2

For text reports, when doing a text comparison (chars or words), if the highlight mode is plain (the default), the report will list the filename, page number (counting from 1), bounding box (as $x1, y1, x2, y2$ coordinates), and text of each difference. If the highlight mode is changes, the report will also list the kind of change (insertion, deletion, replacement). When doing an appearance comparison the report will list the filename, page number and bounding box of each difference.

For visual reports the differences are shown in the PDF or PNG output. If the highlight mode is plain differences are shown highlighted in a single color; if the highlighted mode is changes, insertions, deletions, and replacements are highlighted in different colors. In the case of a PDF report a single PDF file is produced with each page in landscape format showing the differences for each corresponding pair of pages from the original PDFs. In the case of a PNG report, one PNG is output per pair of different pages per PDF. For example, if the filename pattern is given as `report-.png` and pages 3 and 15 of a 20 page PDF have changes, two PNG files will be produced: `report-003-003.png` and `report-015-015.png`. (The page numbers may not be the same if page ranges are used; see [First \(Left\) PDF File's Pages and Second \(Right\) PDF File's Pages](#).)

There is no default value, and without a value no reports are output.

It is possible to produce multiple reports at the same time by using the `-r` or `--report` option multiple times.

If the report filename doesn't include an absolute path, the report will be output to the current directory—or to the `report.path` directory if one has been set in a configuration file. To output a report to a specific path, simply include the path with the filename, e.g., `-r C:\reports\report.xml`.

Note that for most versions of Windows, trying to write a file (e.g., a report) to the root directory (`C:\`) will silently fail.

Although report files are normally set on the command line, it is possible to set them in a configuration file; see [report.fileNames](#).

Note that this option doesn't make sense when comparing two directories of PDF files. In that case use the `Format` option (and optionally the `report.pdfprefix` option), along with the `Report Directory` option instead.

4.3

Report New Renderer

3.6.1

```
C:\Users\mark\>diffpdfc -o ...
```

```
C:\Users\mark\>diffpdfc --use-old-renderer ...
```

Visual reports (i.e., to PNG or PDF files) are normally output using the new high quality renderer. However, in some rare cases the old renderer produces better results, hence the availability of this option.

This option can also be set in a configuration file; see [report.new.renderer](#).

Report Pretty

```
C:\Users\mark>diffpdfc --no-report-pretty ...
```

Normally textual reports are “pretty printed”, that is, output with newlines and spaces to make them human-readable. This option can be used to ensure that textual reports are output with as little whitespace as possible.

This option only applies if `report_filenames` specifies one or more files, and then only for text reports (e.g., JSON and XML formats).

This option can also be set in a configuration file; see [report.pretty](#).

Report Scale

```
C:\Users\mark>diffpdfc -s integer ...
```

```
C:\Users\mark>diffpdfc --scale integer ...
```

The percentage scale to use for visual reports.

This option only applies to visual reports (if any have been specified to be output). The default is 100%. The allowed range is 25-800%. (Up to version 5.6.7 the limit was 400%.) Scaling beyond 400% is only recommended on a fast computer with lots of RAM and using a 64-bit version of *diffpdfc*.

The “scaling” is achieved by increasing the resolution (i.e., the image quality), so increasing the scale (e.g., to 150%) improves the quality of the images (in PNG and PDF reports). A scale of 150% doubles the resolution, and a scale of 200% quadruples the resolution. The trade-off for improved image quality is slower report generation.

This option can also be set in a configuration file; see [report.scale](#).

Report Directory

4.3

```
C:\Users\mark>diffpdfc -R reportdir ...
```

```
C:\Users\mark>diffpdfc --reportdir reportdir ...
```

This option causes *diffpdfc* to write reports in the given directory. The name of each report is that of the compared file with the given format suffix (see [Format](#)), although PDF reports can have a prefix (see [report.pdfprefix](#)).

This option only makes sense when comparing two directories of PDF files, and only has an effect when one or more format suffixes are specified. See [First \(Left\) PDF File or Directory](#), [Second \(Right\) PDF File or Directory](#), [Format](#), and [report.pdfprefix](#).

This can also be set in a configuration file; see [reportdir](#).

Second (Right) PDF File or Directory

```
C:\Users\mark>diffpdfc file1.pdf file2.pdf
```

```
C:\Users\mark>diffpdfc oldpdfsdir newpdfsdir
```

The second (right) PDF file to compare (*file2.pdf*), or the second directory of PDF files to compare (*newpdfsdir*). 4.2

There is no default value and a value is required.

This filename can be set in a configuration file (see [pdf2.filename](#)); but it is more conventional to pass it on the command line after all the options and after the first (left) PDF file. If a filename is specified then the first (left) PDF file or directory must also be a filename. If a directory is specified then the first must also be a directory.

If a directory is specified, *diffpdfc* will only compare PDF files found in the first directory (and its subdirectories) with those of the same name found in the second directory (and its subdirectories). PDF files that are in one directory but not the other, and non-PDF files, are safely ignored.

See also [First \(Left\) PDF File or Directory](#), [Format, Report Directory](#), and [report.pdfprefix](#).

Second (Right) PDF File's Pages

```
C:\Users\mark>diffpdfc --pages2 1-6,12-20 ...
```

Which pages in the second (right) PDF file to include in the comparison. (This option is allowed when comparing directories of PDFs but rarely makes sense in that context.)

The default is no value which means that every page is compared. Using `--pages2=same` means use the same numbers as `--pages1`.

This can also be set in a configuration file; see [pages2](#).

See also [First \(Left\) PDF File's Pages](#).

Summary

```
C:\Users\mark>diffpdfc -S summary.csv ...
```

```
C:\Users\mark>diffpdfc --summary summary.csv ...
```

When two directories of PDF files are compared a summary of each comparison is output to the console. This output can be sent to a file using redirection or using the `-S` or `--summary` option. This option only makes sense when comparing directories of PDF files. (See [First \(Left\) PDF File or Directory](#) and [Second \(Right\) PDF File or Directory](#).)

Note that this option is not influenced by the [report.path](#) if one is set, so a full path should normally be given.

See also [Format, Report Directory](#), and [report.pdfprefix](#).

Verbose

3.2

```
C:\Users\mark>diffpdfc -V ...
```

```
C:\Users\mark>diffpdfc --verbose ...
```

This option causes *diffpdfc* to show its progress as an approximate percentage during the comparison phase. Use this option *twice*, i.e., `diffpdfc -V -V ...` to produce a list of the pages that differ and the approximate percentage differences at the end.

4.3

If reports have been requested (see [Reports](#)), they are produced after the comparison has finished.

This option should not be used in batch files or shell scripts. This option can also be set in a configuration file; see [verbose](#). (See also [quiet](#).)

Version

```
C:\Users\mark>diffpdfc -v
```

```
C:\Users\mark>diffpdfc --version
```

This option causes *diffpdfc* to print its version number and then exit.

This option will work even if the program hasn't yet had a license key registered.

Configuration File Options

Most command line options, and many other options, can be set in configuration files. See the [Introduction](#) for details of how the program's configuration is built up.

See [Option Values](#) for how to specify option values (e.g., colors, lengths, and line styles).

accuracy

```
accuracy: 100
```

How accurate (as a percentage) appearance mode comparisons should be. This only applies to appearance mode comparisons (see [Comparison Mode](#)).

See [Accuracy](#) for details and how to set the accuracy on the command line.

Appear-
ance
mode

4.3

bar

```
bar: true
```

Whether to show a vertical change bar in the left margin parallel with differences.


The default is `true`. This may only be set in a configuration file. Vertical change bars are not produced for appearance mode comparisons.

Char-
acters &
Words
modes

bar.color

```
bar.color: red
```

The color of the vertical change bar.

This option only applies if `bar` is true. The default is red . This may only be set in a configuration file.

Char-
acters &
Words
modes

bar.extend

```
bar.extend: 3.5
```

The length that the vertical change bar extends above and below the top and bottom of the associated differences.

This option only applies if `bar` is true. The default of 3.5 points normally makes vertically nearby changes (e.g., on two lines) have a single vertical change bar. Allowed range 0-15 points. This may only be set in a configuration file.

Char-
acters &
Words
modes

bar.offset

```
bar.offset: 10
```

The offset from the left edge where the vertical change bar should appear.

This option only applies if `bar` is true. The default is 10.0 points. Allowed range 0-72 points. This may only be set in a configuration file.

Char-
acters &
Words
modes

bar.width

```
bar.width: 1
```

The width of the vertical change bar.

This option only applies if `bar` is true. The default is 1.0 points. Allowed range 0-15 points. This may only be set in a configuration file.

Char-
acters &
Words
modes

column.tolerance

```
column.tolerance: 108
```

The horizontal difference in position to allow while considering characters or text to be in the same column. (Has no effect in appearance mode and only has an effect when the multi-column [comparison.algorithm](#)—algorithm 7—is set.)

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

See [comparison.algorithm](#) for more details; see also [line.tolerance](#).

Char-
acters &
Words
modes

5.5

compare

```
compare: words
```

The comparison mode to use.

To set the comparison mode add *one* of these lines: `compare: appearance` or `compare: chars` or `compare: words` to the configuration file.

See [Comparison Mode](#) for details and how to set the comparison mode on the command line.

comparison.algorithm

```
comparison.algorithm: 5
```

Which comparison.algorithm to use. (Has no effect in appearance mode.)

From version 3.5.0 the default is algorithm 5. This may only be set in a configuration file. For single-column PDFs we recommend trying the following in order:

- *Algorithm 5 with a [line.tolerance](#) of 10 points.* This produces the best results (few or no false positives) for most situations. These are the default settings. 3.4
- *Algorithm 5 with a [line.tolerance](#) of 9 or 8 or even fewer points.* Try these as alternatives that can produce fewer false positives in some cases where lines are close together (e.g., inside tables), or when different sized fonts are used. 3.5
- *Algorithm 6 with a [line.tolerance](#) of 10 points.* Try this as an alternative that can produce fewer false positives in very rare cases. 3.7
- *Algorithm 7 with a [line.tolerance](#) of 9 or 10 points and a [column.tolerance](#) of 108 points.* This algorithm is a variation of *Algorithm 5* designed for multi-column PDFs. 5.5

For multi-column PDFs we recommend *Algorithm 7*.

Summaries of the algorithms:

5. This is the default algorithm from version 3.5.0. It gives the best results in general (i.e., the fewest false positives over the widest range of PDFs). This algorithm works best with a [line.tolerance](#) of 10 points, although in some rare cases a [line.tolerance](#) of 9 or 8 or fewer points will produce better results. This is the Standard algorithm used by the GUI (graphical user interface) *DiffPDF*. 3.5
6. This is a refinement of algorithm 5 that in theory should produce better results. In practice, algorithm 5 is almost always better—and also runs faster—so this algorithm is made available purely for use in those very rare cases when it is better. This is the Special algorithm used by the GUI (graphical user interface) *DiffPDF*. 3.7
7. This is a variation of algorithm 5 that should produce much better results for multi-column PDFs. This algorithm works best with a [line.tolerance](#) of 9, 10, or 8 points and with a [column.tolerance](#) of 108 points. This is the Multi-Column algorithm used by the GUI (graphical user interface) *DiffPDF*. 5.5

When comparing PDFs using one of the text modes (chars or words), *diffpdfc* must read the text from each corresponding pair of pages and compare it. In general the best results are achieved when there is sufficient leading—that is, where there is sufficient vertical space between lines.

cores

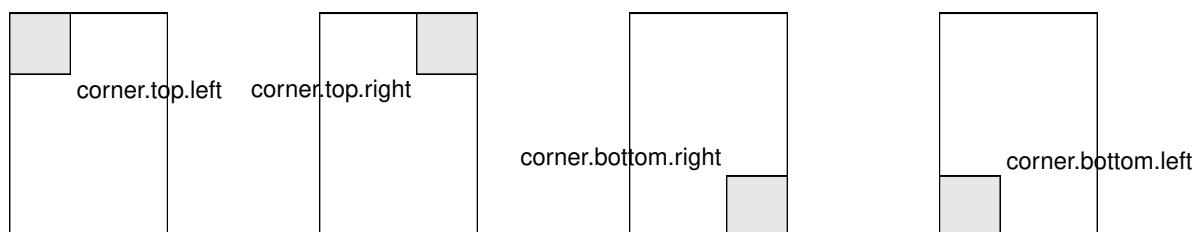
```
cores: 4
```

The number of cores to use, i.e., the number of cores the machine has.

See [Cores](#) for details and how to set the number of cores on the command line.

corners

Corners, if set at all, are set individually, and are specified as an x, y coordinate pair relative to their nearest corner. Each coordinate is specified in points unless a unit is given. For example, `corner.top.left: 36,72`, `corner.top.left: 0.5",1in`, and `corner.top.left: 12.7mm,25.4mm`, are all equivalent. Any text wholly within a corner is ignored for text comparisons, and any part of the page within a corner is ignored for appearance mode comparisons. (See also [margins](#) and [rectangle](#).)



corner.bottom.left

```
corner.bottom.left: 0,0
```

The x, y position of the bottom left corner's furthest point measured from the bottom left of the page.

Anything within the corner is ignored for the purposes of comparison. The default of `0,0` means that nothing in the bottom left corner of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

corner.bottom.right

```
corner.bottom.right: 0,0
```

The x, y position of the bottom right corner's furthest point measured from the bottom right of the page.

Anything within the corner is ignored for the purposes of comparison. The default of `0,0` means that nothing in the bottom right corner of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

corner.color

```
corner.color: darkred
```

The color used for drawing corner lines.

The default is darkred ■. This may only be set in a configuration file. (See also [corner.style](#).)

corner.style

3.1

```
corner.style: dotted
```

The line style used for drawing corner lines.

The default is dotted (.....). This may only be set in a configuration file. (See also [corner.color](#).)

corner.top.left

```
corner.top.left: 0,0
```

The x, y position of the top left corner's furthest point measured from the top left of the page.

Anything within the corner is ignored for the purposes of comparison. The default of $0,0$ means that nothing in the top left corner of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

corner.top.right

```
corner.top.right: 0,0
```

The x, y position of the top right corner's furthest point measured from the top right of the page.

Anything within the corner is ignored for the purposes of comparison. The default of $0,0$ means that nothing in the top right corner of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

format

4.3

```
format: .pdf .csv
```

The space-separated names of one or more report formats to use when reporting the differences between two directories of PDF files. (See [Report Directory](#) and [report.pdfprefix](#).)

See [Format](#) for details and how to set these on the command line.

highlight

```
highlight: plain
```

What kind of highlighting to provide when differences are shown visually, and how differences are reported.


See [Highlighting Mode](#) for details of the highlighting modes and how to set the highlighting mode on the command line. See the following subsections for how to set highlighting colors, opacity, and extents.

highlight.color

```
highlight.color: yellow
```

The highlighting color used for plain highlighting in the chars and words comparison modes, and for highlighting in the appearance comparison mode. (See [Comparison Mode](#).)


Char-
acters &
Words
modes

The default is yellow . This may only be set in a configuration file.

highlight.delete.color

```
highlight.delete.color: red
```

The highlighting color used for deleted text with changes highlighting.


The default is red . This may only be set in a configuration file.

Char-
acters &
Words
modes

highlight.insert.color

```
highlight.insert.color: cyan
```

The highlighting color used for inserted text with changes highlighting.

The default is cyan . This may only be set in a configuration file.

Char-
acters &
Words
modes

highlight.opacity

```
highlight.opacity: 50
```

How opaque (transparent) to make the highlighting colors (to allow the highlighted text to show through).


The default is 50 which is fairly transparent. The valid range is 10-255 with 10 being almost completely transparent and 255 being solid (so the highlighted text won't be visible). This may only be set in a configuration file.

Char-
acters &
Words
modes

highlight.replace.color

```
highlight.replace.color: magenta
```

The highlighting color used for replaced text with changes highlighting.

The default is magenta . This may only be set in a configuration file.

Char-
acters &
Words
modes

highlight.x.extend

```
highlight.x.extend: 4.0
```

The length by which highlighting extends left and right of a difference.

This is particularly useful for difference highlighting in chars comparison mode, since it effectively combines the highlighting of adjacent different characters. The default is 4.0 points. Allowed range 0-15 points. This may only be set in a configuration file.

Char-
acters &
Words
modes

highlight.y.extend

```
highlight.y.extend: 0.5
```

The length by which highlighting extends above and below a difference.

This makes highlights easier to see. The default is 0.5 points. Allowed range 0-15 points. This may only be set in a configuration file.

Char-
acters &
Words
modes

ignore.hyphens

```
ignore.hyphens: false
```

Whether to ignore hyphens.

See [Ignore Hyphens](#) for details and how to set this option on the command line.

Char-
acters &
Words
modes

line.tolerance

```
line.tolerance: 10
```

The vertical difference in position to allow while considering characters or text to be on the same line. (Has no effect in appearance mode.)

The default is 10 points. Allowed range 1-27 points, although 8, 9, and 10 are normally the only values worth using.

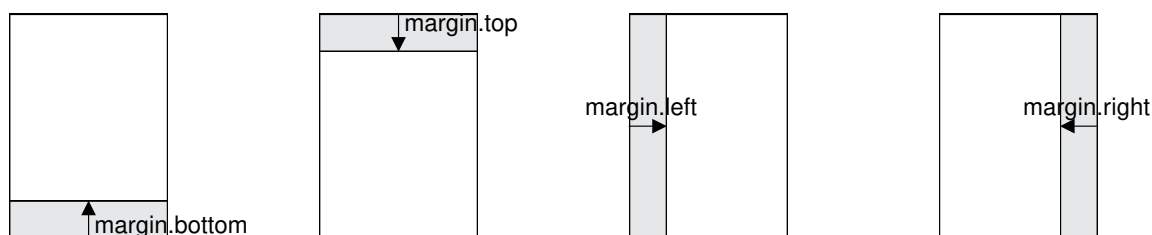
See [comparison.algorithm](#) for more details; see also [column.tolerance](#).

Char-
acters &
Words
modes

3.3

margins

Margins, if set at all, are set individually, and are specified as a length from their nearest edge. Any text wholly within a margin is ignored for text comparisons, and any part of the page within a margin is ignored for appearance mode comparisons. (See also [corners](#) and [rectangle](#).)



margin.bottom

```
margin.bottom: 0.0
```

The height of the bottom margin measured up from a page's bottom edge.

Anything below the bottom margin is ignored for the purposes of comparison. The default of 0.0 points means that nothing at the bottom of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

margin.color

```
margin.color: darkred
```

The color used for drawing margin lines.

The default is darkred . This may only be set in a configuration file. (See also [margin.style](#).)

3.1

margin.left

```
margin.left: 0.0
```

The width of the left margin measured in from a page's left edge.

Anything left of the left margin is ignored for the purposes of comparison. The default of 0.0 points means that nothing on the left of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

margin.right

```
margin.right: 0.0
```

The width of the right margin measured in from a page's right edge.

Anything right of the right margin is ignored for the purposes of comparison. The default of 0.0 points means that nothing on the right of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

margin.style

```
margin.style: dashed
```

The line style used for drawing margin lines.

The default is dashed (-.-.-.-). This may only be set in a configuration file. (See also [margin.color](#).)

3.1

margin.top

```
margin.top: 0.0
```

The height of the top margin measured down from a page's top edge.

Anything above the top margin is ignored for the purposes of comparison. The default of 0.0 points means that nothing at the top of each page is ignored. Allowed range 0-300 points. This may only be set in a configuration file.

normalize.hyphens

```
normalize.hyphens: true
```

Whether to normalize hyphens.

See [Normalize Hyphens](#) for details and how to set this option on the command line.

Char-
acters &
Words
modes

normalize.ligatures

```
normalize.ligatures: false
```

Whether to normalize ligatures.

See [Normalize Ligatures](#) for details and how to set this option on the command line.

Char-
acters &
Words
modes

5.5

pages1

```
pages1: 1-5,7,11-18,19
```

Which pages in the first (left) PDF file to include in the comparison. (This option is allowed when comparing directories of PDFs but rarely makes sense in that context.)

The default is no value which means that every page is compared. (For example, this is the same as using `pages1: 1-20` for a 20 page PDF.)

See [First \(Left\) PDF File's Pages](#) for how to set this option on the command line.

pages2

```
pages2: 1-6,12-20
```

Which pages in the second (right) PDF file to include in the comparison. (This option is allowed when comparing directories of PDFs but rarely makes sense in that context.)

The default is no value which means that every page is compared. A value of same means use the same numbers as `--pages1`.

See [Second \(Right\) PDF File's Pages](#) for how to set this option on the command line.

pairs

```
pairs: 1
```

How many pairs of pages to compare as a unit.

See [Pairs of Pages to Compare as a Unit](#) for details and how to set this option on the command line. See also [show_ins_del_pages](#) for some additional information relevant when the PDFs have inserted or deleted pages.

Char-
acters &
Words
modes

4.0

pdf1.filename

```
pdf1.filename: \path\to\file1.pdf
```

```
pdf1.filename: \path\to\oldpdfsdir
```

The first (left) PDF file to compare (*file1.pdf*), or the first directory of PDF files to compare (*oldpdfsdir*).

This filename or directory is normally given on the command line; see [First \(Left\) PDF File or Directory](#).

4.2

pdf2.filename

```
pdf2.filename: \path\to\file2.pdf
```

```
pdf2.filename: \path\to\newpdfsdir
```

The second (right) PDF file to compare (*file2.pdf*), or the second directory of PDF files to compare (*newpdfsdir*).

This filename or directory is normally given on the command line; see [Second \(Right\) PDF File or Directory](#).

4.2

quiet

```
quiet: false
```

Whether to suppress non-error messages.

See [Quiet](#) for details and how to set this option on the command line.

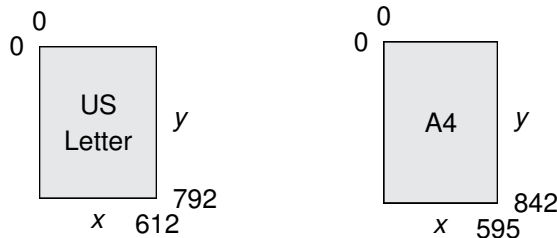
rectangle

5.1

```
rectangle: x1 y1 x2 y2
```

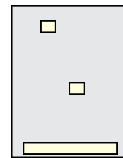
Rectangles on the pages to exclude from the comparison. Any number of rectangles may be specified in terms of their top-left and bottom-right coordinates.

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 below.



Here is an example with all units in points (see [Option Values](#) for specifying in other units):

```
rectangle: 155 79 217 158
rectangle: 300 400 350 450
rectangle: 61 533 558 594
```



Although the rectangle option can be used to exclude corners and margins as well as arbitrary rectangles, it is usually easier to specify corners and margins using the [corners](#) and [margins](#) options.

report

A report is produced for every filename specified in `report.filesnames` or given on the command line using the `-r` or `--report` command line option (if any). See [Reports](#) for details. See also [Format](#), [Report Directory](#), and [report.pdfprefix](#) for how to produce reports when comparing two directories of PDF files.

4.3

report.decimals

```
report.decimals: 2
```

How many digits after the decimal place to output in reports.

This option only applies if `report.filesnames` specifies one or more files, and then only for text reports. This option controls how many digits after the decimal place are given in reports for

floating-point numbers (e.g., $x1, y1, x2, y2$ coordinates). The default is 2. The allowed range is 0-9; we recommend using 4 or fewer. This may only be set in a configuration file.

report.filenamees

```
report.filenamees: \path\to\report.csv \path\to\report-.png
```

The names of one or more files to report differences to.

Multiple reports can be produced at the same time by setting this option to multiple space-separated names—use double-quotes to include filenames with spaces. Each name should have a unique suffix which is used to determine the text or visual report file's format.

See [Report Files](#) for details and how to set these on the command line.

report.new.renderer

```
report.new.renderer: true
```

Whether to output visual reports (i.e., PNG and PDF reports) using the new high quality renderer or the old renderer.

See [Report New Renderer](#) for details and how to set this on the command line.

report.path

3.1

```
report.path: C:\reports
```

The directory to write reports to for those reports whose filenames don't include an absolute path (i.e., those with no path or with a relative path).

If this is not set (the default), then such reports are output to the current directory. If a report's filename includes an absolute path, that path is respected and this option is ignored. (Note that this option has no effect on the [Summary](#) filename's path, which should be given in full.)

report.path_in_title

5.6.5

```
report.path_in_title: true
```

Whether to include the full paths of the PDF filenames in report titles (for those kinds of reports which have titles).

If this is true (the default), filenames in report titles will include their full absolute paths. If this is set to false, then report titles will include filenames without paths.

report.pdfprefix

4.3

```
report.pdfprefix:
```

The prefix to insert in front of PDF report filenames when comparing two directories of PDF files.

If this is not set (the default), no prefix is used, so the report PDFs have the same names as the compared PDFs. This only has an effect if a `.pdf` report format has been specified (see [Format](#)) and a report directory has been specified (see [Report Directory](#)).

report.pretty

```
report.pretty: true
```

Whether to pretty-print reports (i.e., include newlines and indentation).

See [Report Pretty](#) for details and how to set this on the command line.

report.scale

```
report.scale: 100
```

The percentage scale to use for visual reports.

See [Report Scale](#) for details and how to set this on the command line.

reportdir

```
reportdir: C:\Users\mark\reports
```

The directory to use for reports when comparing two directories of PDF files. This only has an effect if one or more report formats have been specified; see also [Format](#) and [report.pdfprefix](#).

See [Report Directory](#) for details and how to set this on the command line.

4.3

save_memory

```
save_memory: false
```

Tell *diffpdfc* to save memory (the default is not to save memory).

If you set this to true, *diffpdfc* will only read each PDF's meta-data and the page data as needed, so as to minimize memory use. In some rare cases this can cause comparisons to fail, which is why the default is false.

If you routinely compare very large PDFs (thousands of pages) it is probably best to set this option to true, and only set it to false if *diffpdfc* cannot compare some PDFs, since when false, *diffpdfc* uses a lot more memory. If only small- and medium-sized PDFs are ever compared, it is best to leave this at its default of false.

5.6.1

5.6.2

show_ins_del_pages

```
show_ins_del_pages: true
```

This setting only applies in text modes when [pairs](#) is greater than one.

If true (the default), when comparing PDFs with different numbers of pages or with inserted or deleted pages, any inserted pages are shown with every word (or character) highlighted as inserted. Similarly, any deleted pages are shown with every word (or character) highlighted as deleted.

If set to false, deleted or inserted pages are not shown at all, so long as *diffpdfc* can detect them. In general, *diffpdfc* can spot inserted or deleted pages if their text is very different from the corresponding page's text, but can't detect them if their text is similar.

Not showing deleted or inserted pages is useful when you know that pages have been inserted or deleted and just want to see the pages that have changed. If deleted or inserted

Char-
acters &
Words
modes

4.0

pages are shown then a changed page might be shown opposite a deleted or inserted page with the corresponding changed page ahead or behind.

We recommend using page ranges where possible, but if `pairs` is used, it might be helpful to compare twice, once with this option true (the default), and again with this option set to false to see which produces the most useful results for the particular PDFs being compared.

This option may only be set in a configuration file.

square.size

```
square.size: 10
```

The square size to use for appearance mode comparisons.

If a pair of pages differ visually, they are compared square by square to identify the exact differences. The square size defaults to 10 points (minimum 4 points, maximum, 144 points). The bigger the square size the faster the comparison, and the coarser the result. This option may only be set in a configuration file.

See also [Accuracy](#).

Appear-
ance
mode

text.missing_page

```
text.missing_page: Missing Page
```

The text to use for missing pages.

When two PDFs are compared that have a different number of pages, or when page ranges are used that compare different numbers of pages, pages missing from the end of the shorter PDF are represented by blank pages with the text “*diffpdfc* Missing Page”. The “Missing Page” part of the text can be replaced by setting this option in a configuration file to any non-empty text.

5.1

verbose

```
verbose: 0
```

How much progress to show during the comparison phase. A value of 0 (the default) means don't show progress, 1 means show an overall summary at the end, and 2 means show a summary of each differing pair of pages as well as the overall summary at the end.

See [Verbose](#) for details and how to set this option on the command line.

3.2

4.3

Troubleshooting

Register Window “Invalid key code” Error

When a license key is acquired from the “Get Key” page at www.qtrac.eu/getkey, the license key that is issued is tied to the Computer ID that's used. In other words, each license key is tied to the computer it is acquired on, and in particular, to the computer's motherboard and CPU.

If—within one year of purchase—a license key is acquired for the wrong computer by mistake, or if the computer it was acquired for is scrapped or replaced or has its motherboard or CPU changed, the license key can be transferred to another computer. Contact support@qtrac.eu with the original Order ID if a transfer is required.

If a license key is acquired as one class of user (e.g., Administrator), and then used by another class of user (e.g., User), even though they are on the *same* machine, the acquired license key may not be visible to *diffpdfc* and so result in an “Invalid key code” error. If this is the case, in the User account, use regedit.exe to change (or add) the registry key: Computer\HKEY_CURRENT_USER\Software\Qtrac Ltd.\diffpdfc\key5 and give it the value of the license key that was acquired by the Administrator for this computer.

False Positives

In some circumstances, *diffpdfc* 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 *diffpdfc* normalizes hyphens (i.e., treats all the different hyphen kinds as the same), but does not normalize ligatures (although it can if told to). To control these, see [Normalize Hyphens](#) and [normalize.ligatures](#).

False positives are usually caused by very narrow interline 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](#). For example, try reducing the line tolerance from the default of 10 points to 9 points or to 8 points.
