

NAME

`latexdiff` – determine and markup differences between two latex files

SYNOPSIS

`latexdiff` [**OPTIONS**] *old.tex new.tex* > *diff.tex*

DESCRIPTION

Briefly, *latexdiff* is a utility program to aid in the management of revisions of latex documents. It compares two valid latex files, here called `old.tex` and `new.tex`, finds significant differences between them (i.e., ignoring the number of white spaces and position of line breaks), and adds special commands to highlight the differences. Where visual highlighting is not possible, e.g. for changes in the formatting, the differences are nevertheless marked up in the source.

The program treats the preamble differently from the main document. Differences between the preambles are found using line-based differencing (similarly to the Unix `diff` command, but ignoring white spaces). A comment, "`%DIF >`" is appended to each added line, i.e. a line present in `new.tex` but not in `old.tex`. Discarded lines

are deactivated by prepending "`%DIF <`". Changed blocks are preceded by comment lines giving information about line numbers in the original files. Where there are insignificant differences, the resulting file `diff.tex` will be similar to `new.tex`. At the end of the preamble, the definitions for *latexdiff* markup commands are inserted. In differencing the main body of the text, *latexdiff* attempts to satisfy the following guidelines (in order of priority):

- 1 If both `old.tex` and `new.tex` are valid LaTeX, then the resulting `diff.tex` should also be valid LaTeX. (NB If a few plain TeX commands are used within `old.tex` or `new.tex` then `diff.tex` is not guaranteed to work but usually will).
- 2 Significant differences are determined on the level of individual words. All significant differences, including differences between comments should be clearly marked in the resulting source code `diff.tex`.
- 3 If a changed passage contains text or text-producing commands, then running `diff.tex` through LaTeX should produce output where added and discarded passages are highlighted.
- 4 Where there are insignificant differences, e.g. in the positioning of line breaks, `diff.tex` should follow the formatting of `new.tex`

For differencing the same algorithm as *diff* is used but words instead of lines are compared. An attempt is made to recognize blocks which are completely changed such that they can be marked up as a unit. Comments are differenced line by line but the number of spaces within comments is ignored. Commands including all their arguments are generally compared as one unit, i.e., no mark-up is inserted into the arguments of commands. However, for a selected number of commands (for example, `\caption` and all sectioning commands) the last argument is known to be text. This text is split into words and differenced just as ordinary text (use options to show and change the list of text commands, see below). As the algorithm has no detailed knowledge of LaTeX, it assumes all pairs of curly braces immediately following a command (i.e. a sequence of letters beginning with a backslash) are arguments for that command. As a restriction to condition 1 above it is thus necessary to surround all arguments with curly braces, and to not insert extraneous spaces. For example, write

```
\section{\textem{This is an emphasized section title}}
```

and not

```
\section {\textem{This is an emphasized section title}}
```

or

```
\section\textem{This is an emphasized section title}
```

even though all varieties are the same to LaTeX (but see **--allow-spaces** option which allows the second variety).

For environments whose content does not conform to standard LaTeX or where graphical markup does not make sense all markup commands can be removed by setting the PICTUREENV configuration variable, set by default to `picture` and `DIFnomarkup` environments; see **--config** option). The latter environment (`DIFnomarkup`) can be used to protect parts of the latex file where the markup results in illegal markup. You have to surround the offending passage in both the old and new file by `\begin{DIFnomarkup}` and `\end{DIFnomarkup}`. You must define the environment in the preambles of both old and new documents. I prefer to define it as a null-environment,

```
\newenvironment{DIFnomarkup}{}{}
```

but the choice is yours. Any markup within the environment will be removed, and generally everything within the environment will just be taken from the new file.

It is also possible to difference files which do not have a preamble.

In this case, the file is processed in the main document mode, but the definitions of the markup commands are not inserted.

All markup commands inserted by *latexdiff* begin with `"\DIF"`. Added blocks containing words, commands or comments which are in `new.tex` but not in `old.tex` are marked by `\DIFaddbegin` and `\DIFaddend`. Discarded blocks are marked by `\DIFdelbegin` and `\DIFdelend`. Within added blocks all text is highlighted with `\DIFadd` like this: `\DIFadd{Added text block}` Selected ‘safe’ commands can be contained in these text blocks as well (use options to show and change the list of safe commands, see below). All other commands as well as braces `"{"` and `"}"` are never put within the scope of `\DIFadd`. Added comments are marked by prepending `"%DIF > "`.

Within deleted blocks text is highlighted with `\DIFdel`. Deleted comments are marked by prepending `"%DIF < "`. Non-safe command and curly braces within deleted blocks are commented out with `"%DIFDEL CMD < "`.

OPTIONS

Preamble

The following options determine the visual markup style by adding the appropriate command definitions to the preamble. See the end of this section for a description of available styles.

--type=markupstyle or **-t markupstyle**

Add code to preamble for selected markup style. This option defines `\DIFadd` and `\DIFdel` commands. Available styles:

```
UNDERLINE CTRADITIONAL TRADITIONAL CFONT FONTSTRIKE INVISIBLE
CHANGEBAR CCHANGEBAR CULINECHBAR CFONTCBHBAR
```

[Default: UNDERLINE]

--subtype=markstyle or **-s markstyle**

Add code to preamble for selected style for bracketing commands (e.g. to mark changes in margin). This option defines `\DIFaddbegin`, `\DIFaddend`, `\DIFdelbegin` and `\DIFdelend` commands. Available styles: `SAFE MARGINAL DVIPSCOL`

[Default: `SAFE`]

--floattype=markstyle or **-f markstyle**

Add code to preamble for selected style which replace standard marking and markup commands within floats (e.g., marginal remarks cause an error within floats so marginal marking can be disabled thus). This option defines all `\DIF...FL` commands. Available styles: `FLOATSAFE TRADITIONALSAFE IDENTICAL`

[Default: `FLOATSAFE`]

--encoding=enc or **-e enc**

Specify encoding of `old.tex` and `new.tex`. Typical encodings are `ascii`, `utf8`, `latin1`, `latin9`. A list of available encodings can be obtained by executing

```
perl -MEncode -e 'print join ("\n", Encode-encodings( ":all" ));' >
```

[Default encoding is `utf8` unless the first few lines of the preamble contain an invocation `\usepackage[...]{inputenc}` in which case the encoding chosen by this command is assumed. Note that ASCII (standard latex) is a subset of `utf8`]

--preamble=file or **-p file**

Insert file at end of preamble instead of generating preamble. The preamble must define the following commands `\DIFaddbegin`, `\DIFaddend`, `\DIFadd{..}`, `\DIFdelbegin`, `\DIFdelend`, `\DIFdel{..}`, and varieties for use within floats `\DIFaddbeginFL`, `\DIFaddendFL`, `\DIFaddFL{..}`, `\DIFdelbeginFL`, `\DIFdelendFL`, `\DIFdelFL{..}` (If this option is set `-t`, `-s`, and `-f` options are ignored.)

--packages=pkg1,pkg2,..

Tell latexdiff that `.tex` file is processed with the packages in list loaded. This is normally not necessary if the `.tex` file includes the preamble, as the preamble is automatically scanned for `\usepackage` commands. Use of the **--packages** option disables automatic scanning, so if for any reason package specific parsing needs to be switched off, use **--packages=none**. The following packages trigger special behaviour:

`amsmath`

Configuration variable `amsmath` is set to `align*` (Default: `eqnarray*`)

`endfloat`

Ensure that `\begin{figure}` and `\end{figure}` always appear by themselves on a line.

`hyperref`

Change name of `\DIFadd` and `\DIFdel` commands to `\DIFaddtex` and `\DIFdeltex` and define new `\DIFadd` and `\DIFdel` commands, which provide a wrapper for these commands, using them for the text but not for the link defining command (where any markup would cause errors).

[Default: scan the preamble for `\usepackage` commands to determine loaded packages.]

---show-preamble

Print generated or included preamble commands to stdout.

Configuration

---exclude-safecmd=exclude-file or **-A exclude-file** or

---exclude-safecmd="cmd1,cmd2,..."

---replace-safecmd=replace-file

---append-safecmd=append-file or **-a append-file** or **---append-safecmd="cmd1,cmd2,..."**

Exclude from, replace or append to the list of regular expressions (RegEx) matching commands which are safe to use within the scope of a `\DIFadd` or `\DIFdel` command. The file must contain one Perl-RegEx per line (Comment lines beginning with `#` or `%` are ignored). Note that the RegEx needs to match the whole of the token, i.e., `^regex$` is implied and that the initial `"\"` of the command is not included. The **---exclude-safecmd** and **---append-safecmd** options can be combined with the **---replace-safecmd** option and can be used repeatedly to add cumulatively to the lists.

---exclude-safecmd and **---append-safecmd** can also take a comma separated list as input. If a comma for one of the regex is required, escape it thus `"\,"`. In most cases it will be necessary to protect the comma-separated list from the shell by putting it in quotation marks.

---exclude-textcmd=exclude-file or **-X exclude-file** or **---exclude-textcmd="cmd1,cmd2,..."**

---replace-textcmd=replace-file

---append-textcmd=append-file or **-x append-file** or **---append-textcmd="cmd1,cmd2,..."**

Exclude from, replace or append to the list of regular expressions matching commands whose last argument is text. See entry for **---exclude-safecmd** directly above for further details.

---replace-context1cmd=replace-file

---append-context1cmd=append-file or `=item` **---append-context1cmd="cmd1,cmd2,..."**

Replace or append to the list of regex matching commands whose last argument is text but which require a particular context to work, e.g. `\caption` will only work within a figure or table. These commands behave like text commands, except when they occur in a deleted section, when they are disabled, but their argument is shown as deleted text.

---replace-context1cmd=replace-file

---append-context2cmd=append-file or `=item` **---append-context2cmd="cmd1,cmd2,..."**

As corresponding commands for context1. The only difference is that context2 commands are completely disabled in deleted sections, including their arguments.

---config var1=val1,var2=val2,... or **-c var1=val1,..**

-c configfile

Set configuration variables. The option can be repeated to set different variables (as an alternative to the comma-separated list). Available variables (see below for further explanations):

MINWORDSBLOCK (integer)

FLOATENV (RegEx)

PICTUREENV (RegEx)

MATHENV (RegEx)

MATHREPL (String)

MATHARRENV (RegEx)

MATHARRREPL (String)

ARRENV (RegEx)

COUNTERCMD (RegEx)

---show--safecmd

Print list of RegEx matching and excluding safe commands.

---show--textcmd

Print list of RegEx matching and excluding commands with text argument.

---show--config

Show values of configuration variables.

---show--all

Combine all ---show commands.

NB For all ---show commands, no `old.tex` or `new.tex` file needs to be specified, and no differencing takes place.

Miscellaneous

---verbose or -V

Output various status information to stderr during processing. Default is to work silently.

---ignore--warnings

Suppress warnings about inconsistencies in length between input and parsed strings and missing characters. These warning messages are often related to non-standard latex or latex constructions with a syntax unknown to `latexdiff` but the resulting difference argument is often fully functional anyway, particularly if the non-standard latex only occurs in parts of the text which have not changed.

---allow--spaces

Allow spaces between bracketed or braced arguments to commands. Note that this option might have undesirable side effects (unrelated scope might get lumped with preceding commands) so should only be used if the default produces erroneous results. (Default requires arguments to directly follow each other without intervening spaces).

---flatten

Replace `\input` and `\include` commands within body by the content of the files in their argument. If `\includeonly` is present in the preamble, only those files are expanded into the document. However, no recursion is done, i.e. `\input` and `\include` commands within included sections are not expanded. The included files are assumed to be located in the same directories as the old and new master files, respectively, making it possible to organise files into old and new directories.

Use of this option is not recommended primarily the processing time for the large documents is prohibitive, and the resulting difference document no longer reflects the structure of the input documents.

--help or **-h**
Show help text

--version
Show version number

Predefined styles

Major types

The major type determine the markup of plain text and some selected latex commands outside floats by defining the markup commands `\DIFadd{...}` and `\DIFdel{...}`.

UNDERLINE

Added text is wavy-underlined and blue, discarded text is struck out and red (Requires color and ulem packages). Overstriking does not work in displayed math equations such that deleted parts of equation are underlined, not struck out (this is a shortcoming inherent to the ulem package).

CTRADITIONAL

Added text is blue and set in sans-serif, and a red footnote is created for each discarded piece of text. (Requires color package)

TRADITIONAL

Like CTRADITIONAL but without the use of color.

CFONT Added text is blue and set in sans-serif, and discarded text is red and very small size.

FONTSTRIKE

Added tex is set in sans-serif, discarded text small and struck out

CCHANGEBAR

Added text is blue, and discarded text is red. Additionally, the changed text is marked with a bar in the margin (Requires color and changebar packages).

CFONTCHBAR

Like CFONT but with additional changebars (Requires color and changebar packages).

CULINECHBAR

Like UNDERLINE but with additional changebars (Requires color, ulem and changebar packages).

CHANGEBAR

No mark up of text, but mark margins with changebars (Requires changebar package).

INVISIBLE

No visible markup (but generic markup commands will still be inserted).

Subtypes

The subtype defines the commands that are inserted at the begin and end of added or discarded blocks, irrespectively of whether these blocks contain text or commands (Defined commands: `\DIFaddbegin`, `\DIFaddend`, `\DIFdelbegin`, `\DIFdelend`)

SAFE No additional markup (Recommended choice)

MARGIN Mark beginning and end of changed blocks with symbols in the margin nearby (using the standard `\marginpar` command – note that this sometimes moves somewhat from the intended position).

DVIPSCOL

An alternative way of marking added passages in blue, and deleted ones in red. Note that DVIPSCOL only works with the dvips converter, e.g. not pdflatex. (it is recommended to use instead the main types to effect colored markup, although in some cases coloring with dvipscol can be more complete).

Float Types

Some of the markup used in the main text might cause problems when used within floats (e.g. figures or tables). For this reason alternative versions of all markup commands are used within floats. The float type defines these alternative commands.

FLOATSAFE

Use identical markup for text as in the main body, but set all commands marking the begin and end of changed blocks to null-commands. You have to choose this float type if your subtype is MARGIN as `\marginpar` does not work properly within floats.

TRADITIONALSAFE

Mark additions the same way as in the main text. Deleted environments are marked by angular brackets `[` and `]` and the deleted text is set in scriptscript size. This float type should always be used with the `TRADITIONAL` and `CTRADITIONAL` markup types as the `\footnote` command does not work properly in floating environments.

IDENTICAL

Make no difference between the main text and floats.

Configuration Variables**MINWORDSBLOCK**

Minimum number of tokens required to form an independent block. This value is used in the algorithm to detect changes of complete blocks by merging identical text parts of less than MINWORDSBLOCK to the preceding added and discarded parts.

[Default: 3]

FLOATENV

Environments whose name matches the regular expression in FLOATENV are considered floats. Within these environments, the *latexdiff* markup commands are replaced by their FL varieties.

[Default: (?:figure|table|plate)[\w\d*@]*]

PICTUREENV

Within environments whose name matches the regular expression in PICTUREENV all latexdiff markup is removed (in pathologic cases this might lead to inconsistent markup but this situation should be rare).

[Default: (?:picture|DIFnomarkup)[\w\d*@]*]

MATHENV,MATHREPL

If both `\begin` and `\end` for a math environment (environment name matching `MATHENV` or `[` and `]`) are within the same deleted block, they are replaced by a `\begin` and `\end` commands for `MATHREPL` rather than being commented out.

[Default: `MATHENV=(?:displaymath|equation)` , `MATHREPL=display-math`]

MATHARRENV,MATHARRREPL

as `MATHENV,MATHREPL` but for equation arrays

[Default: `MATHARRENV=eqnarray*?` , `MATHREPL=eqnarray`]

ARRENV If a match to `ARRENV` is found within an inline math environment within a deleted or added block, then the inlined math is surrounded by `\mbox{...}`. This is necessary as underlining does not work within inlined array environments.

[Default: `ARRENV=(?:array|[pbvBV]matrix)`]

COUNTERCMD

If a command in a deleted block which is also in the `textcmd` list matches `COUNTERCMD` then an additional command `\addtocounter{cntcmd}{-1}`, where `cntcmd` is the matching command, is appended in the diff file such that the numbering in the diff file remains synchronized with the numbering in the new file.

[Default: `COUNTERCMD=(?:footnote|part|section|subsection...|subsubsection|paragraph|subparagraph)`]

BUGS

UTF-8 support requires a relatively new version of perl (5.8.0 is sufficient but 5.6.2 up would work OK, too).

Option `allow-spaces` not implemented entirely consistently. It breaks the rules that number and type of white space does not matter, as different numbers of inter-argument spaces are treated as significant.

Please send bug reports to tilmann@esc.cam.ac.uk. Include the serial number of *latexdiff* (from comments at the top of the source or use `--version`). If you come across latex files that are error-free and conform to the specifications set out above, and whose differencing still does not result in error-free latex, please send me those files, ideally edited to only contain the offending passage as long as that still reproduces the problem.

SEE ALSO

`latexrevise`

PORTABILITY

latexdiff does not make use of external commands and thus should run on any platform supporting Perl 5.6 or higher. If files with encodings other than ASCII or UTF-8 are processed, Perl 5.8 or higher is required.

The standard version of *latexdiff* requires installation of the Perl package `Algorithm::Diff` (available from www.cpan.org - <http://search.cpan.org/~nedkonz/Algorithm-Diff-1.15>) but a stand-alone version, *latexdiff-so*, which has this package inlined, is available, too. *latexdiff-fast* requires the *diff* command to be present.

AUTHOR

Copyright (C) 2004–2007 Frederik Tilmann

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License Version 2

Contributors of fixes and additions: V. Kuhlmann, J. Paisley, N. Becker, T. Doerges, K. Huebner
Thanks to everyone who send in bug reports.